Tech Tonics

TIMSCDR Research Journal

Volume 3 2015-2016

AIM

The research and development has been transforming the computing paradigms and technology in multidimensional directions. Tech Tonics aims to inculcate research culture among post graduate students and make them aware of new innovational happenings in the field of information technology.

TECH TONICS
TIMSCDR Research Journal

Volume 3
Academic Year: 2015-2016

Vision

Thakur Institute of Management Studies, Career Development and Research will become a premier Institute renowned internationally for providing education in Software Application to graduates from various disciplines.

Mission

To achieve excellence in providing software education so that students can grasp existing as well as emerging technologies and to inculcate leadership and managerial qualities in them so that they can deliver results in the organization they join.

Quality Policy

We, the staff, faculty and the management of Thakur Institute of Management Studies, Career Development and Research are committed to offer excellence in software education, conducive academic environment and state of-the-art infrastructure to our students

We will work as a team and interact with students in pro-active manner to achieve our Quality Objectives and fulfill all academic, statutory and regulatory requirements to entire satisfaction of our students as well as for continual improvement of QMS.

PROGRAM EDUCATIONAL OBJECTIVES (PEOs)

PEO-1 : To enable students to gain knowledge across all domains of Information Technology with in-depth understanding of their applications.

PEO-2 : To enable students to analyze problems and to design and develop software solutions using emerging tools and technologies.

PEO-3 : To enable students to continue Life-long learning, Research and Entrepreneurial pursuit in their chosen fields.

PEO-4 : To develop communication, teamwork, and leadership skills necessary to manage multidisciplinary projects and serve the society as responsible and ethical software professionals.

PROGRAM OUTCOMES (POs)

1. Apply domain specific knowledge of computing and mathematics for designing of software solutions for defined problems and requirements.

2. Understand and analyze a problem and suggest feasible solutions.

3. Design, evaluate, and develop effective solutions for complex computing problems to meet desired needs.

4. Design and conduct experiments and use research-based methods to investigate complex computing problems.

5. Use appropriate techniques and software tools for computing activities.

6. Understand and commit to professional norms, regulations and ethics.

7. Recognize the need for and have the ability to engage in independent learning for continual professional development.

8. Understand and apply project management principles, as a member or leader in multidisciplinary environments.

9. Effectively communicate technical information, both oral and written with range of audience.

10. Analyze societal, environmental, cultural and legal issues within local and global contexts when providing software solutions.

11. Work as a member or leader in diverse teams in multidisciplinary environments.

12. Use Innovation and Entrepreneurship for creation of value and wealth.

Contents

Editorial

Research

Editorial

It gives us immense pleasure to present the third volume of our Research Journal, Tech Tonics – TIMSCDR Research Journal. It is a collection of scholarly research papers written by students of MCA (Masters in Computer Applications) course, comprising of research work in the domains of Computer Science, Information Technology and Applications.

The Journal showcases the research endeavors of Post Graduate level students. This helps them to understand IT industry problems either analytically or practically. These efforts will inculcate ability to think and elaborate new ideas amongst the students in the dynamic field of Information Technology.

The Journal represents research work in various specializations in Information Technology like E-business, Human Computer Interface and Advanced Database Theory and Applications. E-business includes a wider range of business processes, such as supply chain management, electronic order processing and customer relationship management. E-business processes help organizations to operate more effectively and efficiently in the field of sales and marketing. Human Computer Interface is the study and planned design of human and computer activities. HCI uses productivity, safety and entertainment to support and fulfill human-computer activities and is applied to various types of computer systems, including air traffic control, nuclear processing, offices and computer gaming. Advanced Database Theory and Application allows gaining an awareness of the basic issues in parallel and distributed data organizations. This subject elaborates the emerging database models including multimedia, spatial and temporal databases, Web-DBMS integration technology with XML for Internet database applications, data-warehousing and data-mining techniques and its applications.

To ensure the originality of research work, the research papers are thoroughly checked and selected for publication. Students of TIMSCDR get opportunity of sound exposure to the field and relevance of standard research work through this academic exercise of performing research and representing the same through research papers.

Finally, this Research Journal is a modest effort to encourage the young, enthusiastic and resourceful minds of the students to do research using latest techniques, and innovate and pen down emerging ideas in the field of Information Technology and Applications.

Editor
Dr. Vinita Gaikwad
I/c Director, TIMSCDR

Abandoning ACID in Favor of BASE
Changes to the long-held relational model

Guided by: Mr. Shirshendu Maitra, *Assistant Professor, TIMSCDR*

Aarti Kori	Abhishek Pandey	Rameez Khan
Student, TIMSCDR	*Student, TIMSCDR*	*Student, TIMSCDR*
Kishan Mallah	Hurais Ansari	Adeel Ahmed
Student, TIMSCDR	*Student, TIMSCDR*	*Student, TIMSCDR*

Abstract - **Relational databases are designed with reliability and consistency at their core. The engineers who developed them focused on a transactional model that ensures that the four principles of the ACID model will always be preserved. However, the advent of a new unstructured database model is turning ACID on its head. The model eschews the highly structured relational model in favor of a flexible key/value store approach. This unstructured approach to data requires an alternative to the ACID model: the BASE model.**

Index Terms – Transactions, Atomicity, Consistency, Durability, Basic Availability, Soft State, Eventual Consistency,CAP.

I. INTRODUCTION

There are four basic tenets to the ACID model:

(1) *Atomicity:* The atomicity of transactions ensures that each database transaction is a single unit that adopts an "all or nothing" approach to execution. If any statement in the transaction fails, the entire transaction is rolled back.

(2) *Consistency:* Relational databases also ensure the consistency of each transaction with the database's business rules. If any element of an atomic transaction would disrupt the consistency of the database, the entire transaction fails.

(3) *Isolation:* The database engine enforces isolation between multiple transactions occurring at or near the same time. Each transaction occurs either before or after every other transaction and the view of the database that a transaction sees at its beginning is only altered by the transaction itself before its conclusion. No transaction should ever see the intermediate product of another transaction. .

(4) *Durability:* The final ACID principle, durability, ensures that once a transaction is committed to the database, it is permanently preserved through the use of backups and transaction logs. In the event of a failure, these mechanisms may be used to restore committed transactions.

NoSQL databases, on the other hand, embrace situations where the ACID model is overkill or would, in fact, hinder the operation of the database. Instead, NoSQL relies upon a softer model known, appropriately, as the BASE model. This model accommodates the flexibility offered by NoSQL and similar approaches to the management and curation of unstructured data. BASE consists of three principles:

(1) *Basic Availability*: The NoSQL database approach focuses on availability of data even in the presence of multiple failures. It achieves this by using a highly distributed approach to database management. Instead of maintaining a single large data store and focusing on the fault tolerance of that store, NoSQL databases spread data across many storage systems with a high degree of replication. In the unlikely event that a failure disrupts access to a segment of data, this does not necessarily result in a complete database outage.

(2) *Soft State:* BASE databases abandon the consistency requirements of the ACID model pretty much completely. One of the basic concepts behind BASE is that data consistency is the developer's problem and should not be handled by the database.

(3) *Eventual Consistency*: The only requirement that NoSQL databases have regarding consistency is to require that at some point in the future, data will converge to a consistent state. No guarantees are made, however, about when this will occur. That is a complete departure from the immediate consistency requirement of ACID that prohibits a transaction from executing until the prior transaction has completed and the database has converge to a consistent state.

The BASE model isn't appropriate for every situation, but it is certainly a flexible alternative to the ACID model for databases that don't require strict adherence to a relational model.

II. . FUNCTIONAL PARTITIONING

Functional partitioning is important for achieving high degrees of scalability. Any good database architecture will decompose the schema into tables grouped by functionality. Users, products, transactions, and communication are examples of functional areas. Leveraging database concepts such as foreign keys is a common approach for maintaining consistency across these functional areas.

Relying on database constraints to ensure consistency across functional groups creates a coupling of the schema to a database deployment strategy. For constraints to be applied, the tables must reside on a single database server, precluding horizontal scaling as transaction rates grow. In many cases, the easiest scale-out opportunity is moving functional groups of data onto discrete database servers.

Schemas that can scale to very high transaction volumes will place functionally distinct data on different database

servers. This requires moving data constraints out of the database and into the application. This also introduces several challenges that are addressed later in this article.

III. CAP THEOREM

Eric Brewer, a professor at the University of California, Berkeley, and cofounder and chief scientist at Inktomi, made the conjecture that Web services cannot ensure all three of the following properties at once (signified by the acronym CAP):

Consistency: The client perceives that a set of operations has occurred all at once.

Availability: Every operation must terminate in an intended response.

Partition tolerance: Operations will complete, even if individual components are unavailable.

Specifically, a Web application can support, at most, only two of these properties with any database design. Obviously, any horizontal scaling strategy is based on data partitioning; therefore, designers are forced to decide between consistency and availability.

IV. ACID SOLUTIONS

ACID database transactions greatly simplify the job of the application developer. Database vendors long ago recognized the need for partitioning databases and introduced a technique known as 2PC (two-phase commit) for providing ACID guarantees across multiple database instances. The protocol is broken into two phases:

- First, the transaction coordinator asks each database involved to precommit the operation and indicate whether commit is possible. If all databases agree the commit can proceed, then phase 2 begins.
- The transaction coordinator asks each database to commit the data.

If any database vetoes the commit, then all databases are asked to roll back their portions of the transaction. What is the shortcoming? We are getting consistency across partitions.

The availability of any system is the product of the availability of the components required for operation. The last part of that statement is the most important. Components that may be used by the system but are not required do not reduce system availability. A transaction involving two databases in a 2PC commit will have the availability of the product of the availability of each database. For example, if we assume each database has 99.9 percent availability, then the availability of the transaction becomes 99.8 percent, or an additional downtime of 43 minutes per month.

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar.

ACID Alternative

If ACID provides the consistency choice for partitioned databases, then how do you achieve availability instead? One answer is BASE (basically available, soft state, eventually consistent).

BASE is diametrically opposed to ACID. Where ACID is pessimistic and forces consistency at the end of every operation, BASE is optimistic and accepts that the database consistency will be in a state of flux. Although this sounds impossible to cope with, in reality it is quite manageable and leads to levels of scalability that cannot be obtained with ACID.

The availability of BASE is achieved through supporting partial failures without total system failure. Here is a simple example: if users are partitioned across five database servers, BASE design encourages crafting operations in such a way that a user database failure impacts only the 20 percent of the users on that particular host. There is no magic involved, but this does lead to higher perceived availability of the system.

So, now that you have decomposed your data into functional groups and partitioned the busiest groups across multiple databases, how do you incorporate BASE into your application? BASE requires a more in-depth analysis of the operations within a logical transaction than is typically applied to ACID. What should you be looking for? The following sections provide some direction.

Consistency Patterns

Following Brewer's conjecture, if BASE allows for availability in a partitioned database, then opportunities to relax consistency have to be identified. This is often difficult because the tendency of both business stakeholders and developers is to assert that consistency is paramount to the success of the application. Temporal inconsistency cannot be hidden from the end user, so both engineering and product owners must be involved in picking the opportunities for relaxing consistency.

Figure 2 is a simple schema that illustrates consistency considerations for BASE. The user table holds user information including the total amount sold and bought. These are running totals. The transaction table holds each transaction, relating the seller and buyer and the amount of the transaction. These are gross oversimplifications of real tables but contain the necessary elements for illustrating several aspects of consistency.

Fig. 1

In general, consistency across functional groups is easier to relax than within functional groups. The example schema has two functional groups: users and transactions. Each time an item is sold, a row is added to the transaction table and the counters for the buyer and seller are updated. Using an ACID-style transaction, the SQL would be as shown in figure 3.

```
Begin transaction
  Insert into transaction(xid, seller_id, buyer_id, amount);
  Update user set amt_sold=amt_sold+$amount where id=$seller_id;
  Update user set amt_bought=amount_bought+$amount where id=$buyer_id;
End transaction
```

Fig. 2

The total bought and sold columns in the user table can be considered a cache of the transaction table. It is present for efficiency of the system. Given this, the constraint on consistency could be relaxed. The buyer and seller expectations can be set so their running balances do not reflect the result of a transaction immediately. This is not uncommon, and in fact people encounter this delay between a transaction and their running balance regularly (e.g., ATM withdrawals and cellphone calls).

How the SQL statements are modified to relax consistency depends upon how the running balances are defined. If they are simply estimates, meaning that some transactions can be missed, the changes are quite simple, as shown in figure 4.

```
Begin transaction
  Insert into transaction(id, seller_id, buyer_id, amount);
End transaction
Begin transaction
  Update user set amt_sold=amt_sold+$amount where id=$seller_id;
  Update user set amt_bought=amount_bought+$amount
       where id=$buyer_id;
End transaction
```

Fig. 3

We've now decoupled the updates to the user and transaction tables. Consistency between the tables is not guaranteed. In fact, a failure between the first and second transaction will result in the user table being permanently inconsistent, but if the contract stipulates that the running totals are estimates, this may be adequate.

What if estimates are not acceptable, though? How can you still decouple the user and transaction updates? Introducing a persistent message queue solves the problem. There are several choices for implementing persistent messages. The most critical factor in implementing the queue, however, is ensuring that the backing persistence is on the same resource as the database. This is necessary to allow the queue to be transactionally committed without involving a 2PC. Now the SQL operations look a bit different, as shown in figure 5.

```
Begin transaction
  Insert into transaction(id, seller_id, buyer_id, amount);
  Queue message "update user("seller", seller_id, amount)";
  Queue message "update user("buyer", buyer_id, amount)";
End transaction
For each message in queue
  Begin transaction
    Dequeue message
    If message.balance == "seller"
      Update user set amt_sold=amt_sold + message.amount
          where id=message.id;
    Else
      Update user set amt_bought=amt_bought + message.amount
          where id=message.id;
    End if
  End transaction
End for
```

Fig. 4

This example takes some liberties with syntax and oversimplifying the logic to illustrate the concept. By queuing a persistent message within the same transaction as the insert, the information needed to update the running balances on the user has been captured. The transaction is contained on a single

database instance and therefore will not impact system availability.

A separate message-processing component will dequeue each message and apply the information to the user table. The example appears to solve all of the issues, but there is a problem. The message persistence is on the transaction host to avoid a 2PC during queuing. If the message is dequeued inside a transaction involving the user host, we still have a 2PC situation.

One solution to the 2PC in the message-processing component is to do nothing. By decoupling the update into a separate back-end component, you preserve the availability of your customer-facing component. The lower availability of the message processor may be acceptable for business requirements.

Suppose, however, that 2PC is simply never acceptable in your system. How can this problem be solved? First, you need to understand the concept of idempotence. An operation is considered idempotent if it can be applied one time or multiple times with the same result. Idempotent operations are useful in that they permit partial failures, as applying them repeatedly does not change the final state of the system.

The selected example is problematic when looking for idempotence. Update operations are rarely idempotent. The example increments balance columns in place. Applying this operation more than once obviously will result in an incorrect balance. Even update operations that simply set a value, however, are not idempotent with regard to order of operations. If the system cannot guarantee that updates will be applied in the order they are received, the final state of the system will be incorrect.

Update Table

updates_applied
trans_id
balance
user_id

Fig. 5

In the case of balance updates, you need a way to track which updates have been applied successfully and which are still outstanding. One technique is to use a table that records the transaction identifiers that have been applied. The table shown in figure 6 tracks the transaction ID, which balance has been updated, and the user ID where the balance was applied. Now our sample pseudocode is as shown in figure 7.

```
Begin transaction
  Insert into transaction(id, seller_id, buyer_id, amount);
  Queue message "update user("seller", seller_id, amount)";
  Queue message "update user("buyer", buyer_id, amount)";
End transaction
For each message in queue
  Peek message
  Begin transaction
    Select count(*) as processed where trans_id=message.trans_id
      and balance=message.balance and user_id=message.user_id
    If processed == 0
      If message.balance == "seller"
        Update user set amt_sold=amt_sold + message.amount
            where id=message.id;
      Else
        Update user set amt_bought=amt_bought + message.amount
            where id=message.id;
      End if
      Insert into updates_applied
        (message.trans_id, message.balance, message.user_id);
    End if
  End transaction
  If transaction successful
    Remove message from queue
  End if
End for
```

Fig. 6

This example depends upon being able to peek a message in the queue and remove it once successfully processed. This can be done with two independent transactions if necessary: one on the message queue and one on the user database. Queue operations are not committed unless database operations successfully commit. The algorithm now supports partial failures and still provides transactional guarantees without resorting to 2PC.

There is a simpler technique for assuring idempotent updates if the only concern is ordering. Let's change our sample schema just a bit to illustrate the challenge and the solution (see figure 8). Suppose you also want to track the last date of sale and purchase for the user. You can rely on a similar scheme of updating the date with a message, but there is one problem.
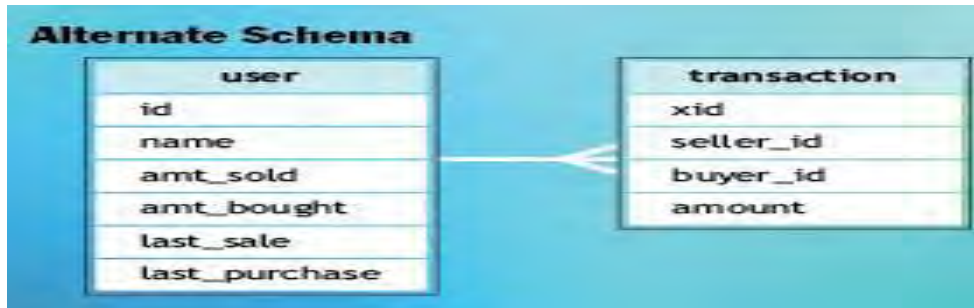
Fig. 7

Suppose two purchases occur within a short time window, and our message system doesn't ensure ordered operations. You now have a situation where, depending upon which order the messages are processed in, you will have an incorrect value for last_purchase. Fortunately, this kind of update can be handled with a minor modification to the SQL, as illustrated in figure.

```
For each message in queue:
    Peek message
    Begin transaction
        Update user set last_purchase=message.trans_date where id=message.buyer_id and last_purchase<message.trans_date;
    End transaction
    If transaction successful
        Remove message from queue
End for
```

Fig. 8

By simply not allowing the last_purchase time to go backward in time, you have made the update operations order independent. You can also use this approach to protect any update from out-of-order updates. As an alternative to using time, you can also try a monotonically increasing transaction ID.

Ordering of Message Queues: A short side note on ordered message delivery is relevant. Message systems offer the ability to ensure that messages are delivered in the order they are received. This can be expensive to support and is often unnecessary, and, in fact, at times gives a false sense of security. The examples provided here illustrate how message ordering can be relaxed and still provide a consistent view of the database, eventually. The overhead required to relax the ordering is nominal and in most cases is significantly less than enforcing ordering in the message system. Further, a Web application is semantically an event-driven system regardless of the style of interaction. The client requests arrive to the system in arbitrary order. Processing time required per request varies. Request scheduling throughout the components of the systems is nondeterministic, resulting in nondeterministic queuing of messages. Requiring the order to be preserved gives a false sense of security. The simple reality is that nondeterministic inputs will lead to nondeterministic outputs.

Soft State/Eventually Consistent: Up to this point, the focus has been on trading consistency for availability. The other side of the coin 0understands the influence that soft state and eventual consistency has on application design. As software engineers we tend to look at our systems as closed loops. We think about the predictability of their behavior in terms of predictable inputs producing predictable outputs. This is a necessity for creating correct software systems. The good news in many cases is that using BASE doesn't change the predictability of a system as a closed loop, but it does require looking at the

behavior in total. A simple example can help illustrate the point. Consider a system where users can transfer assets to other users. The type of asset is irrelevant—it could be money or objects in a game. For this example, we will assume that we have decoupled the two operations of taking the asset from one user and giving it to the other with a message queue used to provide the decoupling.

Immediately, this system feels nondeterministic and problematic. There is a period of time where the asset has left one user and has not arrived at the other. The size of this time window can be determined by the messaging system design. Regardless, there is a lag between the begin and end states where neither user appears to have the asset. If we consider this from the user's perspective, however, this lag may not be relevant or even known. Neither the receiving user nor the sending user may know when the asset arrived. If the lag between sending and receiving is a few seconds, it will be invisible or certainly tolerable to users who are directly communicating about the asset transfer.

V. EVENT-DRIVEN ARCHITECTURE

What if you do need to know when state has become consistent? You may have algorithms that need to be applied to the state but only when it has reached a consistent state relevant to an incoming request. The simple approach is to rely on events that are generated as state becomes consistent. Continuing with the previous example, what if you need to notify the user that the asset has arrived? Creating an event within the transaction that commits the asset to the receiving user provides a mechanism for performing further processing once a known state has been reached. EDA (event-driven architecture) can provide dramatic improvements in scalability and architectural decoupling. Further discussion about the application of EDA is beyond the scope of this article.

VI. CONCLUSION

Scaling systems to dramatic transaction rates requires a new way of thinking about managing resources. The traditional transactional models are problematic when loads need to be spread across a large number of components. Decoupling the operations and performing them in turn provides for improved availability and scale at the cost of consistency. BASE provides a model for thinking about this decoupling.

REFERENCES

[1] www.johndcook.com/blog

[2] www.dataversity.net/acid-vs-base
databases.about.com/od/otherdatabases/a/Abandoning-Acid-In-Favor-Of-Baseit.toolbox.com/blogs/oracle-guide/acid-vs-base

[3] www.techopedia.com/definition/23949/atomicity-consistency-isolation-durability-acid

[4] http://searchsqlserver.techtarget.com/definition/ACID

[5] www.service-architecture.com/articles/database/acid_properties

[6] Image source: goggle/images

Applications of Data Mining

Rashmi Vipat
Assistant Professor,
TIMSCDR

Kanaka Hattiangadi
Student, TIMSCDR

Roidon Rodrigues
Student, TIMSCDR

Abstract -With the wide application of business intelligence in corporate, the demand for data mining software increases daily. As we progress into a more integrated world where technology has become an integral part of the business processes, the process of transfer of information has become more complicated. Data mining fondly called patterns analysis on large sets of data uses tools like association, clustering, segmentation and classification for helping better manipulation of the data help in various fields where data is of greater importance. This paper proposes various fields where application of data mining is important.

Index Terms—Data mining, Applications of Data Mining, Data Mining methods.

I. INTRODUCTION

In the 21st century human beings are used in the different technologies to adequate in the society. Each and every day the human beings are using the vast data and these data are in the different fields .It may be in the form of documents, may be graphical formats ,may be the video ,may be records .As the data are available in the different formats so that the proper action to be taken. Not only to analyze these data but also take a good decision and maintain the data .As and when the customer will required the data should be retrieved from the database and make the better decision .This technique is actually we called as a data mining or Knowledge Hub or simply KDD(Knowledge Discovery Process).The important reason that attracted a great deal of attention in information technology the discovery of useful information from large collections of data industry towards field of "Data mining" is due to the perception of "we are data rich but information poor". There is huge volume of data but we hardly able to turn them in to useful information and knowledge for managerial decision making in business. To generate information it requires massive collection of data. It may be different formats like audio/video, numbers, text, figures, Hypertext formats .
To take complete advantage of data; the data retrieval is simply not enough, it requires a tool for automatic summarization of data, extraction of the essence of information stored, and the discovery of patterns in raw data. With the enormous amount of data stored in files, databases, and other repositories, it is increasingly important, to develop powerful tool for analysis and interpretation of such data and for the extraction of interesting knowledge that could help in decision-making. The only answer to all above is „Data Mining". Data mining is the extraction of hidden predictive information from large databases; it is a powerful technology with great potential to help organizations focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, helps organizations to make proactive knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by prospective tools typical of decision support systems. Data mining tools can answer the questions that traditionally were too time consuming to resolve. They prepare databases for finding hidden patterns, finding predictive information that experts may miss because it lies outside their expectations. Data mining, popularly known as Knowledge Discovery in Databases (KDD), it is the nontrivial extraction of implicit, previously unknown and potentially useful information from data in databases. It is actually the process of finding the hidden information/pattern of the repositories.

II. DATA MINING TECHNIQUES

Industries rely on decision oriented, systemic selection models that enable the decision maker to evaluate the payoff that is expected to result from the implementation of a proposed selection program. Such models go beyond an examination of the size of the validity coefficient and take a host of issues such as capital budgeting and strategic outcomes at the group and organizational levels. Many organizations generate mountains of data about their new learning"s discovered and its performance reports, etc. This data is a strategic resource. Now, making use of most of these strategic resources will lead to improving the quality of their respective industries.

The following are six important steps in the Data Mining process as
 1. Problem Definition.

 2. Knowledge acquisition.

 3. Data selection.

 4. Data Preprocessing.

 5. Analysis and Interpretation.

 6. Reporting and Use.

(A) Associations, Mining Frequent Patterns
These methods identify rules of affinities among the collections. (Hand, Mannila and Smyth, 2001) mention that patterns occur frequently during Data Mining process. The applications of association rules include market basket analysis, attached mailing in direct marketing, fraud detection, department store floor/shelf planning etc.

(B) Classification and Prediction

The classification and prediction models are two data analysis techniques that are used to describe data classes and predict future data classes. A credit card company whose customer credit history is known can classify its customer record as Good, Medium, or Poor. Similarly, the income levels of the customer can be classified as High, Low, and Medium. (Adriaans Peiter and Zantinge Dolf, 2005) explain that if we have records containing customer behavior and we want to classify the data or make prediction, we will find that the tasks of classification and prediction are very closely linked. The models of decision trees, neural networks based classifications schemes are very much useful in pharma industry. Classification works on discrete and unordered data, while prediction works on continuous data. Regression is often used as it is a statistical method used for numeric prediction. Primary emphasis should be made on the selection measurement accuracy and predicative efficiency of any new drug discovery. Simple or multiple regressions is the basic prediction model that enables a decision maker to forecast each criterion status based on predictor information. (Smith and Gupta, 2002) show through case studies how neural network technology is useful from different areas of business. We limited our discussion on algorithms and proof here.

(C) Clustering

It is a method by which similar records are grouped together. Clustering is usually used to mean segmentation. An organization can take the hierarchy of classes that group similar events. Using clustering, employees can be grouped based on income, age, occupation, housing etc. In business, clustering helps identify groups of similarities; characterize customer groups based on purchasing patterns, etc.

III. RESEARCH AND OBSERVATION

Data mining is being used in a wide variety of applications.Areas like prediction and description, relationship marketing, customer profiling customer segmentation, fraud detection and so on.It is even used in the pharmaceutical industry.We would be understaning the application of data mining in the pharmaceutical industry in detail so that we can understand the concept better. Almost two decades ago, the information flow in the pharmaceutical industry was relatively simple and the application of technology was limited. However, as we progress into a more integrated world where technology has become an integral part of the business processes, the process of transfer of information has become more complicated. Today increasingly technology is being used to help the pharmaceutical firms manage their inventories and to develop new product and services. The implications are such that by a simple process of merging the drug usage and cost of medicines (after completing the legal requirements) with the patient care records of doctors and hospitals helping firms to conduct nation wide trials for its new drugs. Other possible uses of information technology in the field of pharmaceuticals include pricing (two-tier pricing strategy) and exchange of information between vertically integrated drug

companies for mutual benefit. Nevertheless, the challenge remains though data collection methods have improved data manipulation techniques are yet to keep pace with them. Most healthcare institutions lack the appropriate information systems to produce reliable reports with respect to other information than purely financial and volume related statements (Prins & Stegwee, 2000). The management of pharma industry starts to recognize the relevance of the definition of drugs and products in relation to management information. In the turmoil between costs, care-results and patient satisfaction the right balance is needed and can be found in upcoming information and Communication technology.

The delivery of healthcare has always been information intensive, and there are signs that the industry is recognizing the increasing importance of information processing in the new managed care environment (Morrisey, 1995). Most automated systems are used as a tool for daily work: they are focused on „production" (daily registration). All the data, which are used to keep the organization running, operational data, are in these automated systems. These systems are also called legacy systems. There is a growing need to do more with the data of an organization than to use them for administration only. A lot of information is hidden in the legacy systems. This information can easily be extracted. Most of the times this can not be done directly from the legacy systems, because these are not build to answer questions that are unpredictable. Research shows that (Zuckerman and Alan, 2006); Armoni, 2002; Rada, 2002) that successful decision systems enriched with analytical solutions are necessary for healthcare information systems.

Given the size of the databases being queried, there is likely to be a trade-off in accuracy of information and processing time. Sampling techniques and tests of significance may be satisfactory to identify some of the more common relationships; however, uncommon relationships may require substantial search time. The thoroughness of the search depends on the importance of the query (e.g., life threatening vs. "curious to know"), the indexing structures used, and the level of detail supplied in the query. Of course, the real data mining challenge comes when the user supplies only a minimal amount of information. For example: find possible serious side effects (not necessarily reported in the manufacturer's product literature) involving food and any type or brand of antacid. A user-interface may be designed to accept all kinds of information from the user (e.g., weight, sex, age, foods consumed, reactions reported, dosage, length of usage). Then, based upon the information in the databases and the relevant data entered by the user, a list of warnings or known reactions (accompanied by probabilities) should be reported. Note that user profiles can contain large amounts of information, and efficient and effective data mining tools need to be developed to probe the databases for relevant information. Secondly, the patient's (anonymous) profile should be recorded along with any adverse reactions reported by the patient, so that future correlations can be reported. Over time, the databases will become much larger, and interaction

data for existing medicines will become more complete. The amount of existing pharmaceutical information (pharmacological properties, dosages, contraindications, warnings, etc.) is enormous; however, this fact reflects the number of medicines on the market, rather than an abundance of detailed information about each product.

One of the major problems with pharmaceutical data is actually a lack of information. For example, an food and drug administration department estimated that only about 1% of serious events are reported to the food and drug administration department. Fear of litigation may be a contributing factor; however, most health care providers simply don't have the time to fill out reports of possible adverse drug reactions. Furthermore, it is expensive and timeconsuming for pharmaceutical companies to perform a thorough job of data collection, especially when most of the information is not required by law. Finally, one should note that the food and drug administration department does not require manufacturers to test new medicines for potential interactions.

There are in general three stages of drug development namely finding of new drugs, development tests and predicts drug behavior, clinical trials test the drug in humans and commercialization takes drug and sells it to likely consumers (doctors and patients).

IV. CONCLUSION

Data mining is a field where research is still being carried on. Data being an important part of any business acts as the vitality of any business. New techniques and algorithms are created so as to use the data efficiently.

REFERENCES

[1]„Novartis Business Intelligence report" Cognos press 2004, www.cognos.com

[2]. Adriaans Peiter, Zantinge Dolf. (2005) Data Mining, Pearson Education, pp.6971.

[3]. Armoni,A. (2002) „Effective Healthcare information systems", IRM Press.

[4]. Berson A and Smith, S.J (2005) Data Warehousing, Data Mining & OLAP, Tata McGraw-Hill Edition, pp.5-6.

[5]. Berson A., Smith, S. and Thearling, K. (1999) Building Data Mining Applications for CRM. McGraw-Hill Professional.

[6]. Berthold Michael and Hand David, J. (1999) Intelligent Data Analysis: An Introduction. Springer, pp.3-10.

[7]. Cooman De Frankey. (2005)   „Data mining in a Pharmaceutical Environment", Belgium press.

[8]. Cosper Nate. (2003) „Insightful Strategies For Increasing Revenues In The Pharmaceuticals Industry: Data Mining for Successful Drugs", retrieved 21 august from http://www.frost.com/prod/servlet/market -insight-top.pag?docid=5701902.

[9] Ben-Gal I., Outlier detection, In: Maimon O. and Rockach L. (Eds.) Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers," Kluwer Academic Publishers, 2005, ISBN 0-387-24435-2.

Are Predictive typing features in Keypad Applications improving our Language?

Dr. Vinita Gaikwad
Associate Professor
TIMSCDR

Jyoti H S Mani
Student, TIMSCDR

Parth Desai
Student, TIMSCDR

Dhananjay K.
Student, TIMSCDR

Aarti Kori
Student, TIMSCDR

Abstract **- In this paper, we aim particularly to deal with cases where if it's possible for people to be able to learn a particular language and improve their grammatical mistakes through the predictive typing features like auto-correct and auto-complete in Smartphone Keypad Applications. Also, if they are aware of any such application and just more comfortable making mistakes while typing without such features. We have laid a survey from a number of random working and studying crowd and tried to understand if these applications helped build their language or not.**

Index Terms - predictive typing, auto-correct, auto-complete, Smartphone Keypad applications, language learning, human linguistic behavior

I. INTRODUCTION

Technology often mediates, and thus influences, patterns of human communication. Spelling errors that happen to result in a real word in the lexicon cannot be detected by a conventional spelling checker.

With the passing days technology is touching every aspect of our life. Technology has effected in almost every single respect one can think of, be it communication, social efforts, education, industry, business, entertainment, comfort, luxury, efficiency, etc. to name a few in good ways one thinks.

Mobile phones have recently improved, the auto-correct and auto-complete features in Smartphone keypad applications. There are different Android Keypad apps which you can install that change layout, languages, and add special features that the default Keypad such as the Google Keypad doesn't have. Some of them such as the SwiftKey Flow app, the Swype app, the Xperia Keypad all provide predictive typing and multinational support. [1] These features help us type efficiently and effectively without any error.

Because of the domination of technology, many people have allowed their socialization skills to slip. Things, such as appearance, common courtesy and our ability to network with others, have all become progressively worse in the workplace. We may be effective in communicating technologically, but we are becoming complete failures in communicating socially. [5] If we cannot communicate effectively in this field, none of the technologies will be able to alter the perceptions of our socialites like, our co-workers, our managers, our customers, our vendors, or even our friends and family. As the use of technology is increasing, our social skills are decreasing.

II. PREDICTIVE TYPING

Predictive typing in Keypad applications improves our response rate with the correct language. The features included are as follows:

A. Auto-Correct

Automatic correction or Auto-Correct is a kind of spell checker that corrects our common typos. It also formats our text message by recognizing our character usage with the inbuilt algorithm, and thus saving us from having to use more tedious functions. [2]

B. Auto-Complete

Automatic complete or Auto-Complete speeds up the human computer interactions by correctly predicting the rest of the word that we are typing. [3]

III. PROBLEMS

We have become so much used to the technology that it becomes almost impossible to find out our errors in the first place. Even if we find out, we forget to learn from our mistakes. A day without these features is like paralyzing the human brain. But why do we rely so much on keypad applications? Many problems may arise due to the same if not prevented on time.

A. Language Learning

How far are these features helping us understand the correct use of language? They are just covering up our lack in language understandability and implementation. Some of the traits of the problem can be found in our day to day life. People find it difficult to even type an easy, grammatically correct sentence.

The reason is lack of proper knowledge in language, and thus, we take external help of keypad applications to accomplish the purpose by typing smoothly and communicating effectively and efficiently. We try to protect ourselves from the embarrassment of not knowing a particular language in front of others through the app. We do not try to learn from our mistakes as that's a tedious work rather such apps serve our purpose.

With the growing technology, the people find ways and means to use time productively. Such applications save our time and work, so we make use of them.

B. Effects on Human Behavior

These applications make us prone to impulsive and thoughtless behavior in our everyday life. Mistyping and grammatical mistakes are easily avoided during an emotional turmoil with the use of predictive typing. Such technologies are training people to be fast but inaccurate. [4]

Having said that, some people are too sloppy to even make use of the keypad applications, rather just type in haze and send the text without realizing if the recipient understands it or not.

IV. SURVEY METHOD

A. Procedure and Design:

An online survey was done for which we sent a link of the form created on Google drive, on Facebook and WhatsApp. Open ended questions were designed. Few questions were extended by text field for input from participant side. After submitting the form the data was saved in a spreadsheet template in Google drive. Our basic objective was to find out the liability and the severity of the problems. There was a positive reply from the respondents.

We let participate think their view, both in objective and subjective way because we were interested in getting participants' reactions to the level where they being real user give their current input as per the experience they had till now and future scope with more betterment in the application with the latest features.

B. Measures:

Both objective and subjective measures were used. One objective measure was, for the basic set of questions, whether participants gave their yes/no input as per the function of the keypad application feedback. For the productive set of questions we measured levels of usage and advantages participate achieved from predictive typing features in their Keypad application.

The primary subjective variables in the questionnaire were the responses to the individual items in the question and the answers to the questions posed by us

V. RESULTS

A. Questionnaires along with Pie Charts and Tables:

1) Occupation:
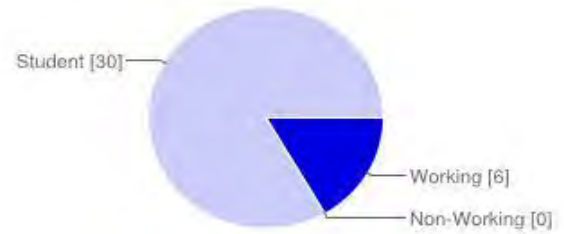
Fig. 1 Respondent Demographics.
.

TABLE I:
RESPONDENT DEMOGRAPHICS

Working	6	16.7%
Non-Working	0	0%
Student	30	83.3%

2) Do you have a smart phone?:

Fig 2:Respondents having a Smartphone.

TABLE II:

NO. OF RESPONDENTS HAVING A SMARTPHONE

Yes	35	97.2%
No	1	2.8%

3) Do you use any Keypad Application?:
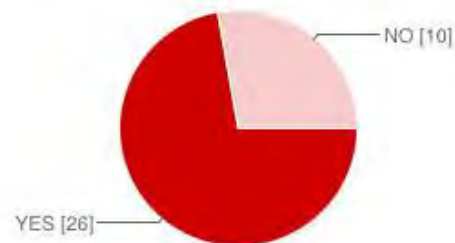
Fig 3. Respondents using a Keypad Application.

11

TABLE III

NO. OF RESPONDENTS USING A KEYPAD APPLICATION

Yes	26	72.2%
No	10	27.8%

4) *Which keypad app do you use most often?:*
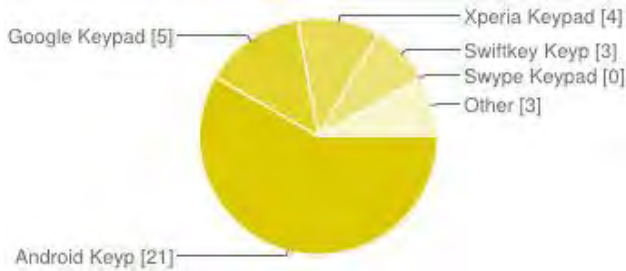
Fig 4. Respondents using various keypad apps.

TABLE IV:

NO. OF RESPONDENTS USING DIFFERENT KEYPAD APPS

Android Keypad(Default Keypad)	21	58.3%
Google Keypad	5	13.9%
Xperia Keypad	4	11.1%
Swiftkey Keypad	3	8.3%
Swype Keypad	0	0%
Other	3	8.3%

5) *Which is your default keypad app currently?:*
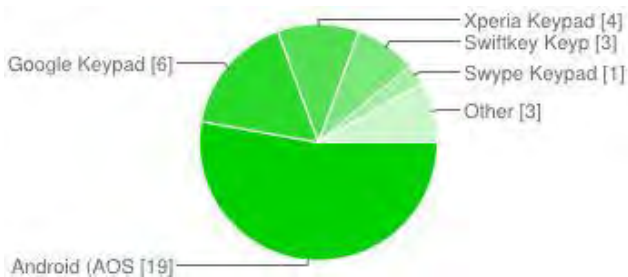
Fig 5. Default Keypad App of the Respondents.

TABLE V. DEFAULT KEYPAD APP OF THE RESPONDENTS

Android Keypad(Default Keypad)	19	52.8%
Google Keypad	6	16.7%
Xperia Keypad	4	11.1%
Swiftkey Keypad	3	8.3%
Swype Keypad	1	2.8%
Other	3	8.3%

6) *How do you enter text using keypad?:*
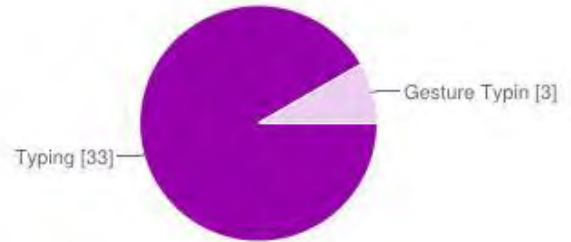
Fig 6. Respondents' mode of entering text.

TABLE VI.

RESPONDENTS' MODE OF ENTERING TEXT

Typing	33	91.7%
Gesture	3	8.3%

7) *A feature in which an application predicts the rest of the word a user is typing is known as _____?:*
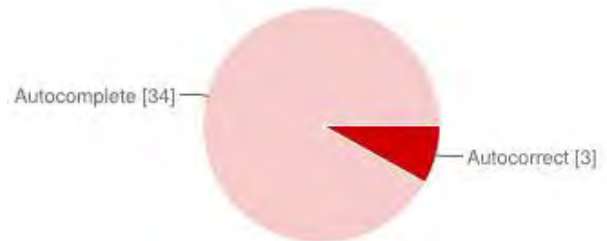
Fig .7 Feature predicting rest of the word.

TABLE VII

FEATURE PREDICTING REST OF THE WORD

Autocorrect	3	8.3%
Autocomplete	34	94.4%

8) A feature to correct common spelling or typing errors, saving time for the user is known as _____?:
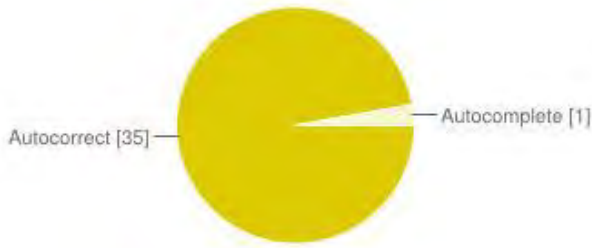
Fig 8. Feature correcting common typos.

TABLE VIII

FEATURE CORRECTING COMMON TYPOS

Autocorrect	*35*	97.2%
Autocomplete	*1*	2.8%

9) How many times does it happen that you want to type something, but something else is typed?:
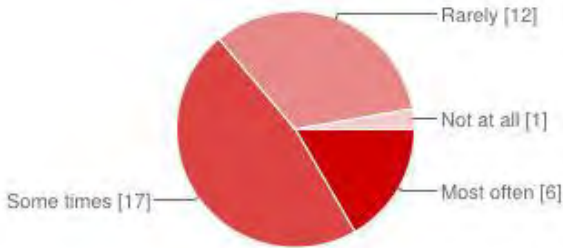
Fig 9.Occurrence of Mistyping

TABLE IX

OCCURRENCE OF MISTYPING

Most often	*6*	16.7%
Sometimes	*17*	47.2%
Rarely	*12*	33.3%
Not at all	*1*	2.8%

10) How do you feel when you witness the above scenario?:
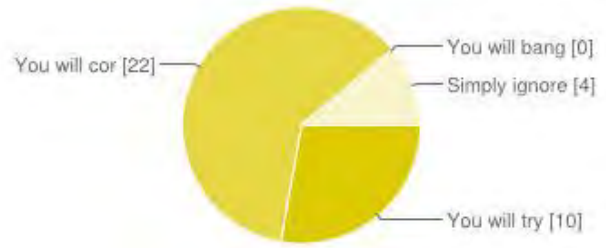
Fig 10. Respondents' behavior after Mistyping.

TABLE X

RESPONDENTS' BEHAVIOR AFTER MISTYPING

You will try again and see what went wrong?	*10*	27.8%
You will correct your mistake regardless of how it happened?	*22*	61.1%
You will bang your phone and stop using the Keypad app	*0*	0%
Simply ignore it	*4*	11.1%

11) Find the correct word and write its meaning in 'Other' tab:
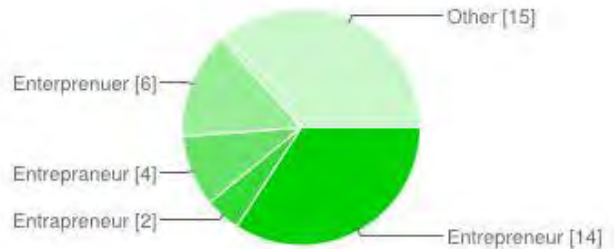
Fig 11 Writing the meaning of the correct word.

TABLE XI

TABLE WRITING THE MEANING OF THE CORRECT WORD

Entrepreneur	*14*	38.9%
Entrapreneur	*2*	5.6%
Entrepraneur	*4*	11.1%
Enterprenuer	*6*	16.7%
Other	*15*	41.7%

The correct word was Entrepreneur and it was seen that 38.8% respondents could crack it."

13

12) You need not come unless you want to":
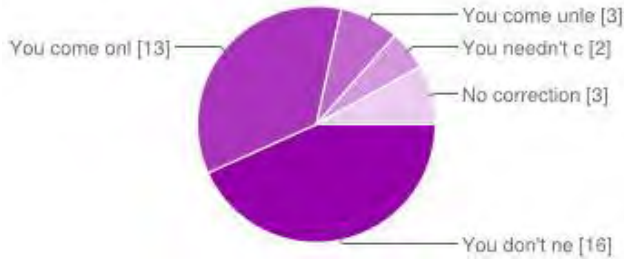
Fig 12. Choosing the nearest sentence to the given idiom.

TABLE XII

CHOOSING THE NEAREST SENTENCE TO THE GIVEN IDIOM

You don't need to come unless you want to	16	44.4%
You come only when you want to	13	36.1%
You come unless you don't want to	3	8.3%
You needn't come until you don't want to	2	5.6%
No correction is required	3	8.3%

The nearest meaning was You don't need to come unless you want to and it was seen that 44.4% respondents could crack it.

13) 1) I 2) help 3) not 4) you 5) did:
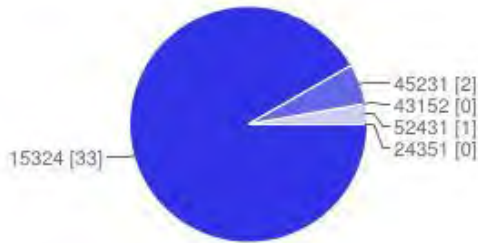
Fig 13. Choosing the correct order of words.

TABLE XIII

CHOOSING THE CORRECT ORDER OF WORDS

24351	0	0%
15324	33	91.7%
45321	2	5.6%
43152	0	0%
52431	1	2.8%

The correct arrangement was 15324 and it was seen that 91.7% respondents could crack it.

14) Do you think Keypad apps help in improving your Writing skills?:
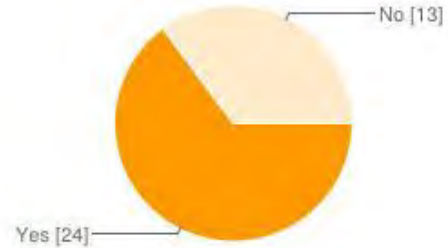
Fig 14. No. of Respondents thinking Keypad apps improve Writing Skills.

TABLE XIV

NO. OF RESPONDENTS THINKING KEYPAD APPS IMPROVE WRITING SKILLS

Yes	24	66.7%
No	13	36.1%

15) What is the impact of Keypad Apps on your Verbal Communication
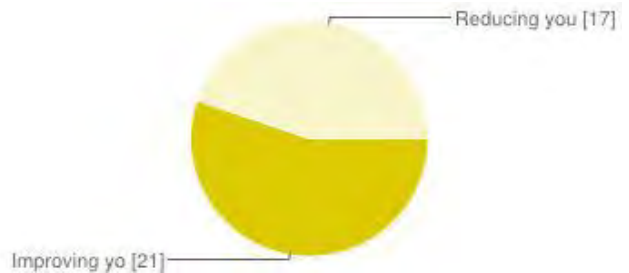
Fig 15. Impact of Keypad Apps on Respondents' Verbal Communication.

TABLE XV

IMPACT OF KEYPAD APPS ON RESPONDENTS' VERBAL COMMUNICATION

Improving your verbal communication	21	58.3%
Reducing your verbal communication	17	47.2%

B. Overall Analysis

The given graph shows the profile of the male and the female respondents. They are further classified into Students and Working class. The analysis shows that 61% of the total are

14

male respondents, while remaining 39% are female respondents. The 50% male respondents are Working and 11% male respondents are Students. The 31% female respondents are Working and 8% female respondents are Students.
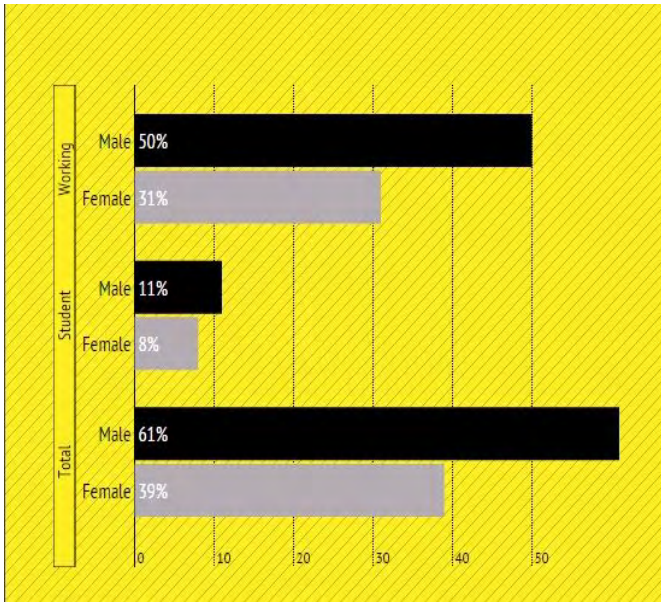
Fig 16 .Respondent's Profile.

The graph below shows the Factors affecting human typing. 45% of the respondents lack knowledge of language, 35% due to human behavior (Ex. Stress), 10% due to mistyping, 5% due to switching between languages and remaining 5% due to dependence on technology.
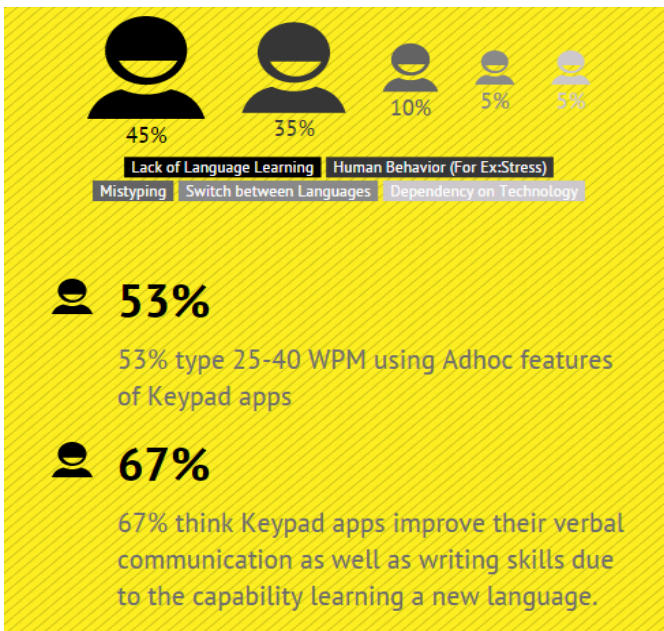
Fig 17.Factors affecting Typing.

The following graph shows the current trends of Keypad Usability wherein 58.3% respondents use Google Keypad (Jelly Bean and lesser versions), 13.9% use Google Keypad (Kitkat and upper versions), 11.1% use Xperia Keypad, 8.3% use Swift Keypad and remaining 8.3% use others.
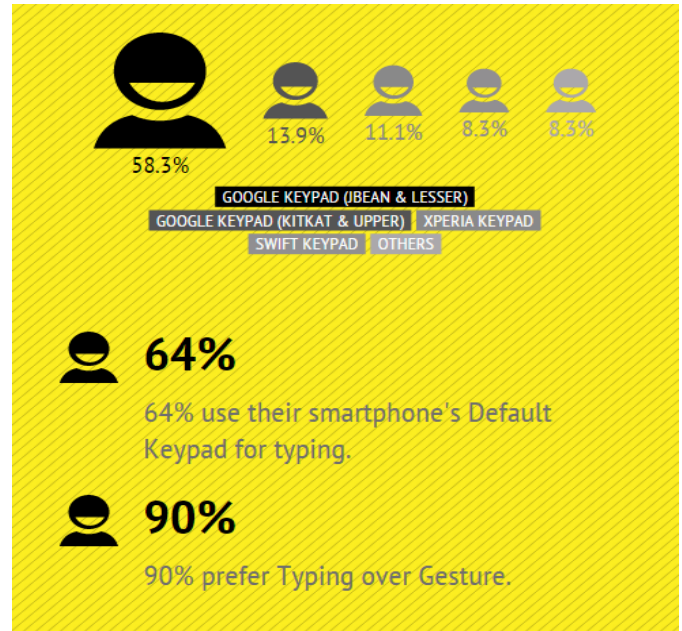
Fig 18.Current Trends of Keypad Usability.

The above analysis showed that the anticipated problems exist in the real world. And thus, their solutions were found as follows.

VI.   SOLUTION

The problems could be solved in many ways, either we could tackle it or we could simply avoid it altogether. Some of the solutions we could track down were as follows:

1) *Education:* We found out that the people didn't have much knowledge of language, so they are needed to be educated, but that was not much feasible.

2) *Multilingual Feature:* Adding a multilingual feature in the keypad application, so if one doesn't know about the required language can still communicate smoothly, this solution was comparatively feasible for the first but yet not applicable in learning a language in its entirety.

3) *E-learning feature:* With this feature on, would pinpoint the user on making mistakes by alerting the user with a sound or highlighting the text and correcting it before sending to the recipients. Here, the response rate is given lesser priority over language learning.

4) *Use of text-to-speech*: Using this technology in keypad application would certainly avoid the problems altogether.

15

These were the temporary solutions which would be effective until we find out the real nature of the problem. On in all we also found out that there were still many people who were either unaware of these latest apps or even after knowing were not using them. There were few who were even fine in sending misspelled texts over to the recipients, as they felt that now everyone is used to such communication.

With our limited research, we found mixed reaction from the respondents; some said Keypad Applications helped them increase their knowledge of oral and written communication, while some said the other way round.

All we can say is that our life has got great effects of technology and it all depends on how the technology is being used. If we use the technology in the negative way, then it is the most harmful way of human life. [5] And if we use it for the welfare of the people, then it can prove to be a boon to mankind. According to us technology is an endless field of creations and inventions.

Technology has helped in increasing the speed of communication and decreasing its cost. However, at the same time it has even caused people to become more impersonal with each other.

The over dependency we have developed on our technology and is the cause for an alarm. We are being driven by technology as opposed to the other way around. And by unplugging our technology, we are unplugging the human-being. [5]

REFERENCES

[1] http://www.makeuseof.com/tag/how-to-choose-the-best-android-Keypad-for-your-own-needs/

[2] http://en.wikipedia.org/wiki/Autocorrection

[3] http://en.wikipedia.org/wiki/Autocomplete

[4] http://www.dailymail.co.uk/sciencetech/article-1205578/How-predictive-texting-takes-toll-childs-brain.html

[5] http://www.civilserviceindia.com/subject/Essay/effects-of-technology1.html

Audit Trail in an Oracle

Rashmi Vipat
Assistant Professor
TIMSCDR

Atul Singh
Student, TIMSCDR

Reshma Tiwari
Student, TIMSCDR

Tripti Tiwari
Student, TIMSCDR

Abstract - **Sarbanes-Oxley's section 404 requires company's key systems be audited. However, many companies have 'auditable' systems and don't even know it. This paper explores methods by which companies can create an auditable system by implementing various levels of audit trails in Oracle Applications.**

I. INTRODUCTION

The technology revolution of the past ten years has been staggering. We have witnessed Monumental and historic events such as Y2K and the dot-com boom/bust. As part of this revolution, we have seen many companies migrate from using mainframesystems to client-server and, most recently, to web-enabled applications. We have also observed the explosion of ERP applications that run on multi-tier architectures. With new technology brings new challenges.

Alongside the revolution of technology, there has been a revolution of corporate accountability. Legislation such as Sarbanes-Oxley (SOX), California Senate Bill 1386, HIPAA, BaselII, and the Gramm-Leach-Bliley Act (among others) has resulted in greater scrutiny and larger penalties for poor corporate governance and security. Much of the risk from this legislation rests on IT systems.

Sarbanes-Oxley's section 404 requires company's key systems to be audited. However, many companies have 'auditable' systems and don't even know it. There is no magic solution or silver bullet in creating an auditable system that satisfies all the requirements for proper corporate governance; rather a layered approach of auditing and logging is required to provide a comprehensive and thorough audit trail. This paper explores methods and degrees by which companies can create an auditable system by implementing various levels of audit trails in Oracle Applications.

There are five primary ways to develop an audit trail:

1. Standard Application Auditing
2. Application Level Audit Trail
3. Database Event Auditing
4. Database Trigger Auditing
5. External Auditing

Standard Application Auditing comes with the 'vanilla' install and provides the last update who and last update when information. However, throughout most of the application, detailed tracking of the changes made between the initial creation of a record and the last change made doesn't exist, but is necessary to show to an auditor. Therefore, this method, alone, is not a sufficient audit trail. Unfortunately, for most companies that run OracleApplications, this is the only audit trail they have (although some applications and processes do have an audit trail built into their development).

Application level audit trail allows a company to track changes at the detailed level. When enabled at the table level through the Sys Admin function, the application works with the database to build detailed records on changes, additions, and deletions at the database level. This audit trail is stored in a separate table from the production data, which still maintains the Standard Application Auditing discussed above. This is part of the solution for companies, but is rarely enabled by companies due to the concern about system performance.

Database event auditing tracks activity at the event level. An example of an event is the user logging into the applications. Database event auditing is necessary to track certain events, but is rarely enabled by companies.

Database trigger auditing is the alter ego of the Application Level Audit Trail as that feature relies on database triggers to form the audit trail. Database trigger auditing may be necessary to record changes made at the database level, not made through the application login. Some form of database trigger auditing is necessary, but is rarely enabled by companies.

External auditing allows an audit trail to build through the 'mining' of certain information contained in the redo logs. This approach gives access to the greatest amount of data in that it allows for the mining of any change or access of the data. However, it is next to impossible to build internally and very expensive to purchase the technology from a third party vendor.

The net result is that a company's need to develop a comprehensive strategy using various technologies crafted to their circumstances. Absent an audit trail, a company's applications may not be 'auditable' and an auditor could assess a material weakness or significant deficiency in the context of a company's section 404 internal controls report.

The technology revolution of the past ten years has been staggering. We have witnessed Monumental and historic events such as Y2K and the dot-com boom/bust. As part of this revolution, we have seen many companies migrate from using mainframe systems to client-server and, most recently, to Web-enabled applications. We have also observed The explosion of (Enterprise Resource Planning) ERP applications that run on multi-tier architectures. With new technology brings new challenges. Along side the revolution of technology, there has been a revolution of corporate accountability. Legislation such as Sarbanes-Oxley (SOX), California SenateBill 1386, HIPAA, BaselII, and the Gramm-Leach-Bliley Act (among others) has resulted in greater scrutiny and larger penalties for poor corporate governance and security. Much of the risk from this legislation rests on IT systems.

In this whitepaper, we will explore the challenges companies face in developing an auditable system given these new technologies and the increased emphasis on corporate governance and security. We will also explore methods and degrees by which companies can address these challenges by implementing various levels of audit trails in their ERP applications.

II. WHY IS AN AUDIT TRAIL SO IMPORTANT – DEVELOPING CONTEXT?

Sarbanes-Oxley (SOX) legislation passed in 2002has had the most far-reaching impact of all the corporate governance initiatives. More recently auditors who are auditing under SOX's 404 provisions are placing greater emphasis on general IT controls surrounding the 'key systems' identified as part of company's controls assessment. The audit of company's controls has turned into a "de facto audit of IT" (Source: CFOIT, summer 2005). Auditors are changing their approach to the audit of company's change management process from only testing a few change requests to testing for unauthorized changes. To do so auditors are requesting (and sometimes requiring) companies to prove that all system changes have been properly authorized and processed according to their change management plan. To facilitate this depth of a review, companies need to be able to identify all 'system' changes to their applications, at least forth systems being Audited under SOX's 404provisions. Unfortunately, implementers and IT management have had little focus on developing strong general IT controls in conjunction with the implementation and development of these new technologies. This leaves many companies exposed to the *possibility* of having an 'auditable' system because many companies cannot identify all changes made to system. Some applications and processes do have standard audit trail. An example of an application is HRMS that allows for date tracking of changes. An example of a process that has a standard audit trail is purchasing which can be implemented to record the approval hierarchy for a given purchase order.

Having an auditable system will lead to various challenges. One example is the automation of controls. Many companies have put a lot of effort into the automation of controls. If the general IT controls cannot be relied upon (i.e. change management process), the various application controls may not be able to be relied upon or may require significant testing to validate them. If proper audit trail for testing of company's change management process cannot be presented to your auditor, they will doubt, at least to some extent, the effectiveness of your change management process. At the extreme, they may disallow the use of your company's application controls.

III. WHAT ARE 'SYSTEM' CHANGES?

In an Oracle Applications environment there are several layers to the system. There is 'Code' stored at the OS level, database level, and at the middle tier. A change to any of the code (also referred to as objects) should be subject to a company's change management plan, regardless how small or large the change. Code changes also include the change of some of the setups within the application that can be done through the application (user interface/forms). Many setups have the effect of changing code because they cause the application to process/react differently based on the different configurations. A sample list of 'system' changes that would need to have an audit trail would include:

- **Changes** to the database structure
- **Addition**, deletion, or change to database triggers
- **Changes** to programs, libraries, or scripts at the OS level
- **Changes** to objects or packages at the database level
- **Changes** to the setups or profile options at the application level

How then is a company to develop an audit trail of all changes made at each of these three levels? This paper will focus primarily on the changes made at the application and database levels. (CREATED_BY and CREATION_DATE) and updating(LAST_UPDATED_BY, LAST_UPDATED (datefunction)). The under lying tables supporting the application store who/when the record was created and who/when the record was last updated. However, in most cases, the database doesn't store WHAT was changed on the record. There are some exceptions to this in certain applications like HR/Payroll where'date-tracking' is critical, but this is predominantly the case throughout most of the applications. Without detailed tracking of the changes made to a record, there is not a sufficient audit-trail for an auditor to review.

Having an audit trail only at the application level does not provide the critical 'what' component required by proper audit trail. If an auditor wants to review a change to setups, for example and all that is stored in the database is the when/who of the record creation and the when/who of the record last update, it

As we will see as we explore this topic further, a strategy to develop an audit trail will require a company to thoughtfully put together a strategy with people from various skill sets throughout its' organization. Development of the strategy often requires input from the internal audit department, your external auditor's orSarbanes-Oxleyadvisors, the IT security team, functional users in Finance and other departments, the application DBAs, and application system administrators.

Unfortunately, there is no magic solution or silver bullet in creating an auditable system that satisfies all the requirements for proper corporate governance. Rather layered approach of auditing and logging is required to provide a comprehensive and thorough audit trail. There are five primary ways to develop an audit trail:
1. Standard Application Auditing
2. Application Level Audit Trail
3. Database Event Auditing
4. Database Trigger Auditing
5. External Auditing

IV. STANDARD APPLICATION AUDITING

Oracle Applications by default provides only basic auditing functions. The application Automatically tracks standard Who/When for most records as it relates to creating doesn't provide any visibility to the changes that could have been made been the first time it was created and the last time it was updated. Furthermore, it doesn't provide the visibility to what column, in particular, was changed. 'What' was changed has significant impact on the risk of the change.

Also, the standard application audit trail allows you to audit when a user logs in, what responsibilities or roles they use and what forms they access; however, this auditing is not enabled by default. To enable logging of all user, responsibility, and forms accesses, set the system profile option 'Sign-On: Audit Level' to FORMS (default is NONE).

Most commonly in the implementation of Oracle Applications, only the default application level of audit trail is the extent of most companies' audit trail. To

the degree that this is still the case for most companies, significant audit risk still exists.

Companies need to remedy this as soon as possible.

Pros of Application Level auditing:
1. Standard part of the application configuration
2. No performance impact

Cons of Application Level auditing
1. Does not provide detail on what gets changed in most cases – just who and when a record is updated
2. Does not provide the level of detail required by an auditor
3. Does not track enough detail to be able to reconstruct activity in the case of fraud

V. APPLICATION LEVEL AUDITING

Oracle Applications has the ability to build a complete audit trail of changes made Through the application.

When a user logs into the application, certain information is maintained by the application about that session, including the application login used.

Application level auditing works with the database to record information about a transaction as the application interacts with the database. For those tables/columns that are configured to do so, as records are changed added, deleted, or updated.

VI. DATABASE TRIGGER AUDITING

Database Trigger Auditing is similar to Application Level Auditing and uses the same Core technology– database triggers to develop the audit trail. However, database level auditing can audit some transactions that the application cannot. Employees (and unwelcome on-employees) can log directly into the database and use SQL tools to query the data directly instead of accessing the data through the applications. ODBC applications also access the database directly to interact with the database. All interaction with the database requires a login and password just like an application does. So, Database level auditing captures changes to the database through logins other than from the application.

VII. DATABASE EVENT AUDITNG

The Oracle Database has sophisticated built-in auditing capabilities. Many databases Operations can be audited including database logins, key SQL statements, and data access. The audit trail information can be stored either in the database or toa file. Since the auditing is enabled at the database level, there is no application user session information captured. All Oracle Applications users' sessions and most maintenance programs connect to the database using a single database account (APPS); therefore, the value of the auditing information is diminished.

Database Event Auditing only captures WHO and WHEN information, not WHAT was changed for many of the audited database operations. Auditing SELECT, INSERT, UPDATE, and DELETESQL statements provide only information on the database object (table or view) that was accessed, not what data was manipulated.

Database Event Auditing is most useful in capturing accesses to the database and changes in the structure and privileges, rather than auditing changes in individual tables. Since
This type of auditing is built-in functionality, it is easy to configure and there is minimal performance impact. The most significant downsides that the audit trail can be easily manipulated by the DBA.

Pros of Database Event Auditing:
1. Part of standard database technology
2. Easy to configure and implement

Cons of Database Event Auditing:
1. Those with sufficient understanding of, and access to, the audit tables can alter the audit trail
2. Application user session information is not captured

There is one other note to mention on Database Trigger Auditing and the difference from Application Level Auditing regarding the interaction between the database and the applications – if auditing is enabled only at the database level, the application session information (e.g., application username) is not available. All Oracle Applications users' sessions connect to the database using a single database account (APPS). When the database trigger auditing is enabled through the application, the application specific database trigger stores the application user

session information in the shadow table. Since the applications connect to the database through the APPS login, if database level auditing alone was enabled the audit trail would show that the APPS user made all the changes in the application. Therefore, as it relates to changes made at the application level, the Application Level Auditing is superior to just using the Database Trigger Auditing.

However, since users and other applications can also directly access the database without logging into the application, it may also necessary to audit certain information at the database level.

Pros of Database Trigger Auditing:
1. Can provide meaningful auditing of low-volume high-risk for database sessions activity on tables such a score application set ups and application security tables (like those contained in FND and SYS schemes, etc.)
2. Part of standard application/database technology

Cons of Database Trigger Auditing:
1. Many companies and DBAs have-not enabled application auditing and are unfamiliar with it's' deployment. As such, strict best practice project methodology needs to be followed.

2. Auditing high volume transactional tables is impractical, if not impossible, due to the impact on performance caused by the database triggers

3. Those with sufficient understanding of, and access to, the audit tables can alter the audit trail

VIII. EXTERNAL AUDITING

The final method we will review is the ability to develop an audit trail external to the Database. There are primarily two methods used to capture the audit trail information externally: databases redo logs and network traffic.

We will focus solely on database redo logs as there do logs provide much more complete auditing solution and this capability is delivered with the Oracle database, nevertheless commercial solutions exist that will audit an Oracle database by analysing network traffic. Redo logs are used by the database to allow for rolling back of transactions and to allow for recovery of the database in event of a hardware or software failure.

The redo logs are a record of all changes in the database, therefore, are the ideal audit trail as every

insert, update and delete to the database are captured. Since the redo logs area core function of the database and are always enabled, there is not the overhead cost of having to create triggers on the tables as with Application Level Auditing and Database Trigger Auditing. The downside of the redo logs is that the format is cryptic and difficult to understand even furthermost expert DBA. There are several solutions to access and query there do logs to allow for the building of a secure audit trail – Oracle provides the Log Miner tool with the database. Log Miner has limited abilities in version 8i and more functionality in newer database versions such as 9i and 10g (see Metalink note291686.1). LogMiner is not an auditing solution; therefore, a custom auditing solution must be built if the LogMiner tool is used. Otherwise, several commercial database auditing tools that use the redo logs are available.

Pros of External Auditing:
1. Provides greatest ability to track activity for those that have database access
2. Depending on the architecture of the storage of the data, this audit trail typically cannot be altered by high-risk employees
3. Use of third party tools to extract such data can virtually eliminate risk of corruption of audit trail.

Depending on the database version and abilities of staff, extracting the correct data via Log Miner can be challenging. Purchase of third party tool or use of consultants to develop may be required. Additional hardware may be required to support the auditing

IX. CONCLUSION

Most companies running Oracle Applications have significant audit risk because of the Inability to produce the detailed level of audit-trail auditors are coming to expect. With greater scrutiny on IT controls relating to financial systems, companies must take action now by putting in place the technology necessary to build such audit trails. Unfortunately, there is no magic solution or silver bullet in creating an auditable system. That satisfies all the requirements for proper corporate governance. Rather layered approach of auditing and logging is required to provide a comprehensive and thorough audit trail.

REFERENCES

[1] Jeffrey T. Hare, CPA, Jeffrey T. Hare, CPA is the founder of ERP Seminars (www.erpseminars.com) and the Oracle User Best Practices

Automatic Database Tuning and Administration

Shirshendu Maitra
Assistant Professor
TIMSCDR

Priyanka K.
Student, TIMSCDR

Devendra Gupta
Student, TIMSCDR

JangBahadur G.
Student, TIMSCDR

Chaitali Mistry
Student, TIMSCDR

Punit Kanojia
Student, TIMSCDR

Abstract- **The study on database technologies, or more generally, the technologies of data and information management, is an important and active research field. Recently, many exciting results have been reported. In this fast growing field, researchers play more and more active roles. Research papers from scholars, both in India and abroad, appear in prestigious academic forums.**
Although the paper covers only a small number of topics and the selection of the topics is far from balanced, we hope that such an effort would attract more and more researchers, to enter the frontiers of database research and promote collaboration.

I. INTRODUCTION

In the current scenario the manual work (Done by Human) cost more to an organization than the automatic work (Done by Machine) and the ratio is increasing day by day as per the increment in Machine (Hardware + Software) Intelligence. We are moving towards the world where the Machines will be able to perform better than today by their own intelligence. They will adjust themselves as per the customer's performance need. But to make this dream true, lots of human efforts (Theoretical and Practical) are needed to increase the capability of Machines to take their own decision and make the future free from manual work and reduce the working cost. For example, building indexes and materialized views often dramatically improve the performance on a given query workload, but it is very difficult to select the necessary indexes and views because such decision depends on how these queries are executed. On the other hand, the cost of hardware has dropped dramatically. Thus the cost for human to tune and manage the database systems often dominates the cost of ownership. To reduce such cost, it is desirable to automate database tuning and administration.
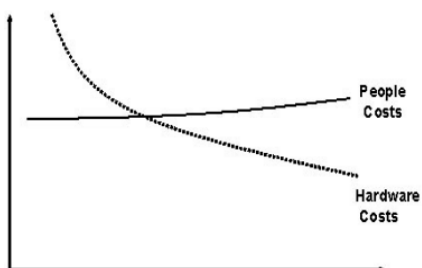
Fig 1: database cost estimation

II. PERFORMANCE TUNING

Most systems will respond to increased load with some degree of decreasing performance. A system's ability to accept higher load is called scalability, and modifying a system to handle a higher load is synonymous to performance tuning. Systematic tuning follows these steps:

1. Assess the problem and establish numeric values that categorize acceptable behavior.
2. Measure the performance of the system before modification.
3. Identify the part of the system that is critical for improving the performance. This is called the bottleneck.
4. Modify that part of the system to remove the bottleneck.

A performance problem may be identified by slow or unresponsive systems. This usually occurs because high system loading, causing some part of the system to reach a limit in its ability to respond. This limit within the system is referred to as a bottleneck. A handful of techniques are used to improve performance. Data drives today's businesses, and managing databases often involves complex
Planning, time management and system wide routine task implementation. Database automation helps enterprises better manage their database operations, reducing downtimes as well as the overall time taken in.

III. MANUAL TUNING

Manual Tuning Framework Database Administrator Is responsible for enhancing the performance of database system. The detection of performance degradation is achieved by continuously monitoring system performance parameters. Several methods including the usage of materialized views and indexes, pruning table and column sets, usage of self healing Techniques, usage of physical design tuning etc. have been proposed that proactively monitor the system performance indicators, analyze the symptoms and auto tune the DBMS to deliver enhanced performance. The performance degradation is due to increased workload on the system.
This increased load has to be minimized to enhance the response rate of the system. In order to achieve this objective, either the administrator decreases some amount of load by closing some files or he may increase the RAM. The administrator has to check continuously or we can say, at regular intervals the Buffer Cache Hit Ratio (BCHR) Based

on this hit ratio, the database administrator determines if more amount of RAM has to be allocated. This task of load reduction by increasing RAM requires manual intervention and thus may take even days & Years to complete. However, Oracle manages RAM memory demands according to the demands of each task by using sophisticated algorithms to improve the speed of RAM intensive tasks. Oracle DBA can dynamically de-allocate RAM memory as well as re-allocate it. But since database administrator is a normal human being, he cannot calculate the actual amount of RAM memory required by an application.
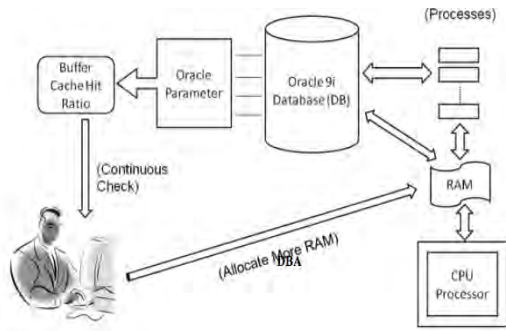
Fig 2: Manual Database Design.

Sometimes, more amount of RAM is allocated than needed which wastes the extra portion of RAM. Thus, there is a great need of dynamic memory allocation features to create a self-tuning database. In Oracle Database 10g, a self-tuning feature such as Automatic Memory Management (AMM) allows the database system to detect shortages and adjusts the main memory regions according to the changing demands on the Oracle environment. Therefore, researchers are now focusing on theDevelopment of self-tuning techniques such as the COMFORT automatic tuning project or the MAPE approach given by IBM for a continuous adaptation. Ranking of various tuning parameters based on statistical analysis is presented in the ranking of parameters is based on the amount of impact they produce on the system performance for a given workload. A formal knowledge framework for self tuning database system is presented in that define several knowledge components which include Policy knowledge, Workload knowledge, Problem diagnosis knowledge,Problem Resolution Knowledge, Effectors knowledge, and Dependency knowledge.

The architecture presented in this paper involves extracting useful information from the system log and also from the DBMS using system related queries. This information gathered over a period of time is then used to run the SQL scripting for a desired output response time. The application framework would then estimate the extent of correction to be applied to the key system parameters that help scale up the System performance. The classical control is modified and a three-stage control involving Monitor, Analyze and Tune is

employed to ensure system stability. The architecture presented in for self-healing database forms the basis for the new architecture presented in this paper. This paper presents a new DBMS architecture based on modular approach, where in each functional module can be monitored by set of monitoring hooks. These monitoring hooks are responsible for saving the current status information or a snapshot of the server to the log. This architecture has high monitoring overhead, due to the fact that when large number of parameters to be monitored, almost every module's status information has to be stored on to the log and if done frequently may eat up a lot of CPU time. Moreover, this architecture focuses more on healing the system and does not consider tuning the DBMS for performance improvement.

IV. AUTOMATED SYSTEM ARCHITECTURE

Many business applications demand the use of complex database systems, which should be administered and optimized for better performance. As suggested in physical tuning should be avoided, as it is expensive. As the physical design of database suffers from various limitations, a new script based automated database architecture is proposed in order to achieve high grade of performance. The architecture as shown in figure 3 is employed for identifying the symptoms and altering key system parameters. The DBMS system log file will be the primary source of information checks the current status of the system. The data miner tool compresses the data into smaller information base since the log file may contain huge amount of data. The architecture has three basic building blocks comprising of Data Miner, Script and Tuner. After the extraction of meaningful information, the proposed script and algorithms estimate the extent of correction required.
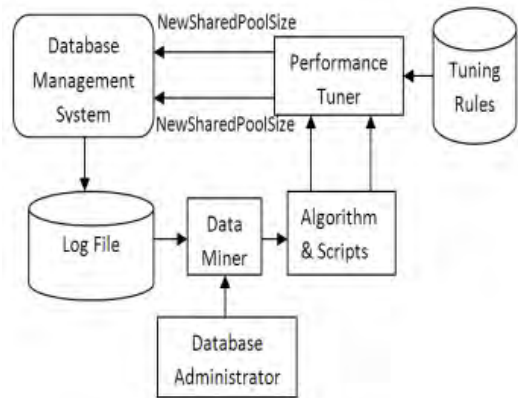
Fig 3: Script Based Tuning Architecture

These algorithms and scripts would tune the database using various tuning rules as well as system parameters. However, several parameters can be altered simultaneously for better performance gain. The algorithm estimates the required buffer size based on the current DBMS input parameters and the

tuner applies the necessary correction to the buffer size based on the tuning rules. Most importantly the internal corrective measure such as altering the buffer size of the DBMS used in query processing is explored in this architecture. In this research , we provided a self-tuned database system architecture as shown in fig 4 in order to enhance system performance. Since DBA is responsible for administration and optimization of various tasks, he can either increase RAM or can decrease the amount of load on CPU for the purpose of performance optimization. But this would be time-consuming technique asDBA is a normal human being who cannot perform complex calculations within seconds like a computer system. DBA may not know exactly how much RAM is to be allocated for enhancing system performance. So, we propose an approach to automate this optimization task of DBA as shown in fig i.e. the taskWhich DBA has to do for performance enhancement would now be done by theComputer system within small timelines.
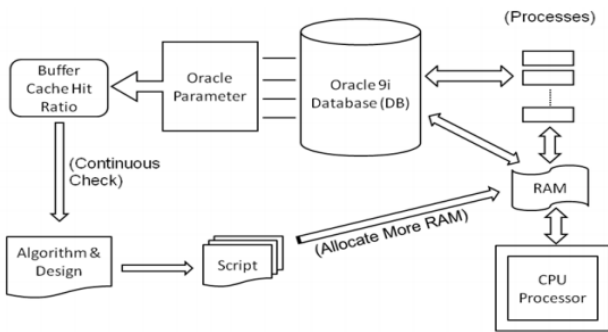
Fig 4: Automated Database Design

V. ALGORITHM & FLOW CHART

The algorithm defines three variables: ΔRT abbreviates for change in response time, BUFFER_SIZE denotes the current size of buffer, and CACHE_SIZE corresponds to the size of cache memory.

ALGORITHM :

1. dbTuner (ESTIMATED_CACHE_SIZE)
2. Begin
3. Run application, algorithm and process
4. Calculate the change in response time (ΔRT)
5. If ($\Delta RT>0$)
    {
    Run Script { BUFFER_SIZE = BUFFER_SIZE + 1
    Allocate more RAM and update CACHE_SIZE
    }
    Else IF ($\Delta RT<0$)
    {
    Run Script

    {
    BUFFER_SIZE = BUFFER_SIZE - 1
    Reduce RAM and update CACHE_SIZE
    }
    }
6. Go To Step 4
7. Stop application, algorithm and process
8. End

Flow Chart:
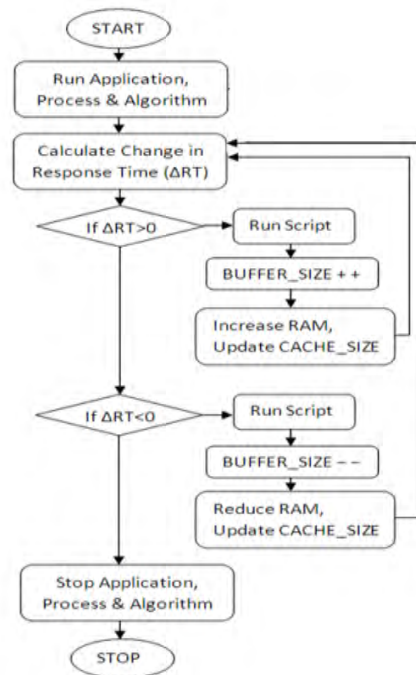
Fig 5: Automated Tuning Flow Chart.

VI. CONCLUSION

Tuning the database can become quite complex, but modern databases offer the administrator an unparalleled ability to control the PGA and SGA. Until old databases evolve into a completely self-tuning architecture, the DBA was responsible for adjusting the dynamic configuration of the system RAM. Automated SGA adjustment scripts can be used to allow the DBA to grow and shrink the SGA regions. Manual tuning cost more for an organization but it is one of the major need for an organization to attract the customer. So in this research work we have found an algorithmic solution to fulfill the need of an organization in the shape of this Automation Framework. This framework will not take any cost and it will give faster result compare to manual tuning. The future work is to implement this framework and test in a working environment. There are many research problems unsolved in this area. First, very little work has been done in automatically tuning system parameters, and it is challenging to predict the system

performance after changing such parameters. Second, little is known on how to adjust the system to changes of the workload. Ideally, the database system shall be able to automatically adjust to such changes. Third, given the numerous features to tune, it remains challenging to identify the system bottleneck as well as to tune all these together. As the work is still in progress we hope to see better results in coming years.

REFERENCES

[1] Sanjay Agrawal, Surajit Chaudhuri, Abhinandan Das, and VivekNarasayya. Automating layout of relationaldatabases. In ICDE'2003, 2003.

[2] Sanjay Agrawal, Surajit Chaudhuri, and Vivek R. Narasayya. Automated selection of materialized views and indexesin SQL databases.In VLDB'2000, pages 496–505, 2000.

[3] Philippe Bonnet and Dennis Elliott Shasha. Database Tuning: Principles, Experiments, and Troubleshooting Techniques.Morgan Kaufman, 2002.

[4] Surajit Chaudhuri and Vivek R. Narasayya. An efficient cost-driven index selection tool for microsoft SQL server.In VLDB'1997, pages 146–155, 1997.

[5] SurajitChaudhuri and Vivek R. Narasayya. Index merging. In ICDE'1999, pages 296–303, 1999.

[6] Jun Rao, Chun Zhang, Nimrod Megiddo, and Guy M. Lohman. Automating physical database design in a paralleldatabase. In SIGMOD'2002, pages 558–569, 2002.

[7] Steve Rozen and Dennis Shasha. A framework for automating physical database design.In VLDB'1991, pages401–411, 1991.

[8] Gary Valentin, Michael Zuliani, Daniel C. Zilio, Guy M. Lohman, and Alan Skelley. DB2 advisor: An optimizersmart enough to recommend its own indexes. In ICDE'2000, pages 101–11,2000.

Big Data and Its Applications

Rashmi Vipat
Assistant Professor
TIMSCDR

Raajas Rane
Student, TIMSCDR

Rahul Sharma
Student, TIMSCDR

Abstract-**Analyzing and Managing data have always offered the greatest benefits and the greatest challenges for organizations of all sizes and across all industries.Businesses have long struggled with finding a pragmatic approach to capturing information about their customers, products, and services.Indeed, we are dealing with a lot of complexity when it comes to data. Some data is structured and stored in a traditional relational database, while otherdata, including documents, customer service records, and even pictures and Videos, is unstructured. Our world is generating data at a speed faster than ever before. In 2010, 5 Exabyte (1 billion gigabytes) of data were createdevery two days, exceeding the totalamount of information that was created by human beingsfrom the dawn of civilization to 2003.Till 2020, over 40 Zettabytes (10^21 bytes) of data would be created, replicated,and consumed.With the overwhelming amount of data pouringinto our lives, from anywhere, anytime, and any device, we have undoubtedly entered the era of Big Data. Big data brings big value. With advanced big data analyzing technologies, insights can be acquired to enable better decisionmaking for critical development areas such as healthcare, economic productivity, energy, and natural disaster prediction,to name but a few. This Research paper is aimed at highlighting the omnipresence of Big Data in Modern Digital world & its diverse applications.**

Keywords- Big Data Characteristics; Big Data Architecture ; Big Data Platforms; MapReduce; Hadoop, Big Table ; Big Data Applications

I. INTRODUCTION

Data explosion is an inevitable trend as the world is connected more than ever. This speed of data generation will continue in the coming years and is expected to increase at an exponential level [4]. The above fact gives birth to the widely circulated concept *Big Data*. Big data is data that exceeds the processing capacity of conventionaldatabase systems.In 2012, *Gartner* updated its definition as follows: "Big data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization. The data is too big, moves too fast, or doesn't fit the structures of the current database architecture.Is big data really new or is it an evolution in the data management journey? The answer is yes — it is actually both.

Big data is not a single technology but a combination of old and new technologies that helps companies gain actionable insight [1].Therefore, big data isthe capability to manage a huge volume of disparate data, at the right speed, and within the right time frame to allow real-time analysis and reaction.The value of big data to an organization falls into two categories: analytical use and enabling new products. Big data analytics can revealinsights hidden previously by data too costly to process, such as peerinfluence among customers, revealed by analyzing shoppers'transactionsand social and geographical data. Being able to process everyitem of data in reasonable time removes the troublesome need forsampling and promotes an investigative approach to data, in contrastto the somewhat static nature of running predetermined reports. The past decade's successful web startups are prime examples of big data used as an enabler of new products and services. For example, bycombining a large number of signals from a user's actions and thoseof their friends, Facebook has been able to craft a highly personalizeduser experience and create a new kind of advertising business. It is nocoincidence that the lion's share of ideas and tools underpinning bigdata have emerged from Google, Yahoo, Amazon, and Facebook [2].The emergence of big data into the enterprise brings with it a necessary counterpart: agility. Successfully exploiting the value in big data requiresexperimentation and exploration. Whether creating new productsor looking for ways to gain competitive advantage, the job callsfor curiosity and an entrepreneurial outlook.

In information technology, big data has emerged as a widely recognized trend, attracting attentions from government, industry and academia. Big data are high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making,insight discovery and process optimization. The leading IT companies, such as SAG, Oracle, IBM, Microsoft,SAP and HP, have spent more than $15 billion on buyingsoftware firms specializing in data management and analytics.This industry on its own is worth more than $100 billion andgrowing at almost 10% a year, which is roughly twice as fastas the software business as a whole.

II. CHARACTERISTICS

Big data can be described by the following characteristics [4]:

Volume – The quantity of data that is generated is very important in this context. It is the size of the data which determines the value and potential of the data under consideration and whether it can actually be considered Big Data or not. The name ‗Big Data' itself contains a term which is related to size and hence the characteristic.

Variety - The next aspect of Big Data is its variety. This means that the category to which Big Data belongs to is also a very essential fact that needs to be known by the data analysts. This helps the people, who are closely analyzing the data and are associated with it, to effectively use the data to their advantage and thus upholding the importance of the Big Data.

Velocity - The term ‗velocity' in the context refers to the

speed of generation of data or how fast the data is generated and processed to meet the demands and the challenges which lie ahead in the path of growth and development.

Variability - This is a factor which can be a problem for those who analyze the data. This refers to the inconsistency which can be shown by the data at times, thus hampering the process of being able to handle and manage the data effectively.

Veracity - The quality of the data being captured can vary greatly. Accuracy of analysis depends on the veracity of the source data.

Complexity - Data management can become a very complex process, especially when large volumes of data come from multiple sources. These data need to be linked, connected and correlated in order to be able to grasp the information that is supposed to be conveyed by these data. This situation, is therefore, termed as the _complexity' of Big Data.

Big data analytics consists of 6 Cs in the integrated industry 4.0 and Cyber Physical Systems environment. 6C system, that is, consist of connection (sensor and networks), Cloud (computing and data on demand), Cyber (model and memory), content/context (meaning and correlation), community (sharing and collaboration), and customization (personalization and value). In this scenario and in order to provide useful insight to the factory management and gain correct content, data has to be processed with advanced tools (analytics and algorithms) to generate meaningful information. Considering the presence of visible and invisible issues in an industrial factory, the information generation algorithm has to capable of detecting and addressing invisible issues such as machine degradation, component wear, etc. in the factory floor.

III. BIG DATA ARCHITECTURE

Before we go into the architecture, it is important to take into account the functional requirements for bigdata.Figure 1-1 illustrates that data must first be captured, and organized and integrated.After this phase issuccessfully implemented, data can be analyzed based on the problem being addressed [1].

Finally, management takes action based on the outcome of thatanalysis. For example, Amazon.com might recommend a book based on apast purchase or a customer might receive a coupon for a discount for afuture purchase of a related product to one that was just purchased.

Fig 1: The cycle of Big Data management

To understand big data, it helps to lay out the components of the architecture.A big data management architecture must include a variety of servicesthat enable companies to make use of myriad data sources in a fast and effectivemanner. To help make sense of this, the components are put into adiagram (see Figure 1-2) that will help see what is there and the relationshipbetween the components.

Fig 2: The Big Data Architecture

A. Interfaces and Feeds:

As seen in Figure 1.2 above, on either side of the diagram are indications of interfacesand feeds into and out of both internally managed data and data feedsfrom external sources. To understand how big data works in the real world,it is important to start by understanding this necessity. In fact, what makesbig data big is the fact that it relies on picking up lots of data from lots ofsources.

Therefore, open application programming interfaces (APIs) will becore to any big data architecture. In addition, keep in mind that interfacesexist at every level and between every layer of the stack. Without integration services, big data cannot happen.

B. Redundant Physical Infrastructure:

The supporting physical infrastructure is fundamental to the operation andscalability of a big data architecture. In fact, without the availability of robustphysical infrastructures, big data would probably not have emerged as suchan important trend. To support an unanticipated or unpredictable volumeof data, a physical infrastructure for big data has to be different than thatfor traditional data. The physical infrastructure is based on a distributedcomputing model. This means that data may be physically stored in manydifferent locations and can be linked together through networks, the use of adistributed file system, and various big data analytic tools and applications.

Redundancy is important because we are dealing with so much data fromso many different sources. Redundancy comes in many forms. If the companyhas created a private cloud, there has to be a redundancy builtwithin the private environment so that it can scale out to support changing workloads. If the company wants to contain internal IT growth, it may useexternal cloud services to augment its internal resources. In some cases, thisredundancy may come in the

form of a Software as a Service (SaaS) offeringthat allows companies to do sophisticated data analysis as a service [7]. TheSaaS approach offers lower costs, quicker startup, and seamless evolution ofthe underlying technology.

C. Security Infrastructure:

The more important big data analysis becomes to companies, the moreimportant it will be to secure that data. For example, for a healthcarecompany, they would probably want to use big data applications to determinechanges in demographics or shifts in patient needs. This data about yourconstituents needs to be protected both to meet compliance requirementsand to protect the patients' privacy. It is necessary to take into account whois allowed to see the data and under what circumstances they are allowed todo so. You will need to be able to verify the identity of users as well as protectthe identity of patients. These types of security requirements need to bepart of the big data fabric from the outset and not an afterthought.

D. Operational data sources:

When we think about big data, it is important to understand havingto incorporate all the data sources that will give a complete picture ofthe business and see how the data impacts the way the business is operated [1].Traditionally, an operational data source consisted of highly structureddata managed by the line of business in a relational database. But as theworld changes, it is important to understand that operational data now hasto encompass a broader set of data sources, including unstructured sourcessuch as customer and social media data in all its forms. We find new emerging approaches to data management in the big dataworld, including document, graph, columnar, and geospatial databasearchitectures. Collectively, these are referred to as *NoSQL,* or not only SQL,databases. In essence, we need to map the data architectures to the types of transactions.

We also need data architectures that support complex unstructuredcontent. We need to include both relational databases and non-relationaldatabases in the approach to harnessing big data. It is also necessary toinclude unstructured data sources, such as content management systems, sothat we can get closer to the 360-degree business view.All these operational data sources have several characteristics in common:

- They represent systems of record that keep track of the critical datarequired for real-time, day-to-day operation of the business.

- They are continually updated based on transactions happening withinbusiness units and from the web.

- For these sources to provide an accurate representation of the business,they must blend structured and unstructured data.

- These systems also must be able to scale to support thousands of userson a consistent basis. These might include transactional e-commercesystems, customer relationship management systems, or call centerapplications.

E. Organizing data services & tools:

Not all the data that organizations use is operational. A growing amountof data comes from a variety of sources that are not as quite as organized orstraightforward, including data that comes from machines or sensors, andmassive public and private data sources. In the past, most companies were not able to either capture or store this vast amount of data. It was simply tooexpensive or too overwhelming. Even if companies were able to capturethe data, they did not have the tools to do anything about it. Very few toolscould make sense of these vast amounts of data. The tools that did exist werecomplex to use and did not produce results in a reasonable time frame. Inthe end, those who really wanted to go to the enormous effort of analyzingthis data were forced to work with snapshots of data. This has the undesirableeffect of missing important events because they were not in a particularsnapshot [3].

F. Analytical data warehouses and data marts:

After a company sorts through the massive amounts of data available, it is often pragmatic to take the subset of data that reveals patterns and put it into a form that's available to the business. These warehouses and marts provide compression, multilevel partitioning, and a massively parallel processing architecture.

G. Big data analytics:

The capability to manage and analyze petabytes of data enables companiesto deal with clusters of information that could have an impact on the business.This requires analytical engines that can manage this highly distributeddata and provide results that can be optimized to solve a business problem. Analytics can get quite complex with big data. For example, some organizationsare using predictive models that couple structured and unstructureddata together to predict fraud. Social media analytics, text analytics, and newkinds of analytics are being utilized by organizations looking to gain insightinto big data.

H. Reporting and visualization:

Organizations have always relied on the capability to create reports to give them an understanding of what the data tells them about everything from monthly sales figures to projections of growth. Big data changes the way that data is managed and used. If a company can collect, manage, and analyze enough data, it can use a new generation of tools to help management truly understand the impact not just of a collection of data elements but also how these data elements offer context based on the business problem being addressed. With big data, reporting and data visualization become tools for looking at the context of how data is related and the impact of those relationships on the future.

I. Big data applications:

Traditionally, the business expected that data would be used to answer questions about what to do and when to do it. Data was often integrated as fields into general-purpose business applications. With the advent of big data, this is changing. Now, we are seeing the development of applications that are designed specifically to take advantage of the unique characteristics of big data.

Some of the emerging applications are in areas such as healthcare, manufacturing management, traffic management, and so on. What do all these big data applications have in common? They rely on huge volumes, velocities, and varieties of data to transform the behavior of a market. In healthcare, a

big data application might be able to monitor premature infants to determine when data indicates when intervention is needed. In manufacturing, a big data application can be used to prevent a machine from shutting down during a production run. A big data traffic management application can reduce the number of traffic jams on busy city highways to decrease accidents, save fuel, and reduce pollution.
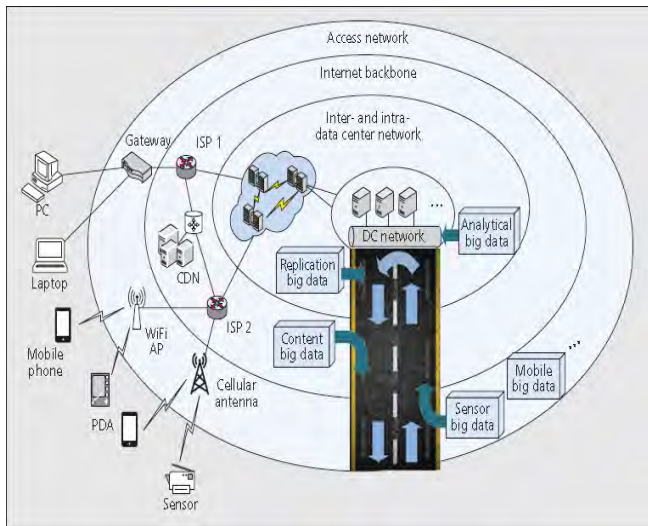
Fig 3: Layered network architecture from the perspective of big data applications

IV. BIGDATA MANAGEMENT PLATFORMS

Not all the data that organizations use is operational. A growing amountof data comes from a variety of sources that aren't quite as organized orstraightforward, including data that comes from machines or sensors, andmassive public and private data sources. In the past, most companies were not able to either capture or store this vast amount of data. It was simply tooexpensive or too overwhelming.With the evolution of computing technology, it is now possible to manage immense volumes of data that previously could have only been handled bysupercomputers at great expense [1].

The realbreakthrough in big data happened as companies like Yahoo!, Google, andFacebook came to the realization that they needed help in monetizing themassive amounts of data their offerings were creating.Their resulting solutions are transforming the data management market. In particular, the innovations MapReduce, Hadoop,and Big Table proved to be the sparks that led to a new generation of datamanagement.

A. MapReduce:

MapReduce was designed by Google as a way of efficiently executing a set offunctions against a large amount of data in batch mode. The ―map‖ componentdistributes the programming problem or tasks across a large number ofsystems and handles the placement of the tasks in a way that balances the load and manages recovery from failures. After the distributed computationis completed, another function called ―reduce‖ aggregates all the elementsback together to provide a result. An example of MapReduce usage would beto determine how many pages of a book are written in each of 50

differentlanguages.

B. BigTable:

Big Table was developed by Google to be a distributed storage systemintended to manage highly scalable structured data. Data is organized intotables with rows and columns. Unlike a traditional relational database model,Big Table is a sparse, distributed, persistent multidimensional sorted map. Itis intended to store huge volumes of data across commodity servers.

C. Hadoop:

Hadoop is an Apache-managed software framework derived from MapReduceand Big Table. Hadoop allows applications based on MapReduce to run onlarge clusters of commodity hardware. The project is the foundation for thecomputing architecture supporting Yahoo!‗s business. Hadoop is designed toparallelize data processing across computing nodes to speed computationsand hide latency. Two major components of Hadoop exist: a massively scalabledistributed file system that can support petabytes of data and a massivelyscalable MapReduce engine that computes results in batch.

V. BIGDATA APPLICATIONS

Big data has increased the demand of information management specialists in that Software AG, Oracle Corporation, IBM, Microsoft, SAP, EMC, HP and Dell have spent more than $15 billion on software firms specializing in data management and analytics. In 2010, this industry was worth more than $100 billion and was growing at almost 10 percent a year: about twice as fast as the software business as a whole.

Developed economies make increasing use of data-intensive technologies. There are 4.6 billion mobile-phone subscriptions worldwide and between 1 billion and 2 billion people accessing the internet. Between 1990 and 2005, more than 1 billion people worldwide entered the middle class which means more and more people who gain money will become more literate which in turn leads to information growth. The world's effective capacity to exchange information through telecommunication networks was 281 petabytes in 1986, 471 petabytes in 1993, 2.2 Exabyte in 2000, 65 Exabyte in 2007 and it is predicted that the amount of traffic flowing over the internet will reach 667 Exabyte annually by 2014. It is estimated that one third of the globally stored information is in the form of alphanumeric text and still image data, which is the format most useful for most big data applications.

A. . Government:

The use and adoption of Big Data, within governmental processes, is beneficial and allows efficiencies in terms of cost, productivity and innovation. Data analysis often requires multiple parts of government (central and local) to work in collaboration and create new and innovative processes to deliver the desired outcome [4].

INDIA: Big data analysis was, in parts, responsible for the BJP and its allies to win a highly successful Indian General Election 2014. The Indian Government utilizes numerous techniques to ascertain how the Indian electorate is responding to government action, as well as ideas for policy augmentation.

USA: In 2012, the Obama administration announced the Big Data Research and Development Initiative, to explore how big data could be used to address important problems faced by the government. The initiative is composed of 84 different big data programs spread across six departments.Big data analysis played a large role in Barack Obama's successful 2012 re-election campaign.The United States Federal Government owns six of the ten most powerful supercomputers in the world.

The Utah Data Center is a data center currently being constructed by the United States National Security Agency. When finished, the facility will be able to handle a large amount of information collected by the NSA over the Internet. The exact amount of storage space is unknown, but more recent sources claim it will be on the order of a few Exabyte.

Project	Begin time	Department	Goal
1000 Genomes Project	1/2008	National Institutes of Health	To produce an extensive public catalog of human genetic variatio including SNPs and structural variants, and their haplotype conte
ARM Project	3/2012	Department of Energy	To collect and process climate data from all over the world to und stand Earth's climate and come up with answers to climate chang issues.
XDATA	3/2012	Defense Advanced Research Projects Agency (DARPA)	To develop new computational techniques and software program that can analyze structured and unstructured big data sets faster more efficiently.
BioSense 2.0	3/2012	Center for Disease Control and Prevention	To track public health problems and make data instantly accessib end users across government departments.
The Open Science Grid	3/2012	National Science Foundation (NSF) and Department of Energy	To provide advanced fabric of services for data transfer and analy scientists worldwide for collaboration in science discovery.
Big Data for Earth System Science	3/2012	U.S. Geological Survey	To provide scientists with state-of-the-art computing capabilities collaborative tools to make sense of huge data sets and better un stand the earth.
Human Brain Project	2/2013	European Commission	To simulate the human brain and model everything that scientist know about the human mind using a supercomputer.
Unique Identification Authority	2/2009	The Indian Planning Commission	To create a biometric database of fingerprints, photographs, and scan images of all 1.2 billion people for efficient resident identifi tion in welfare service delivery.

Fig 4: Illustration of Big Data Applications used by Governments across the world.

B. Manufacturing:

Based on TCS 2013 Global Trend Study, improvements in supply planning and product quality provide the greatest benefit of big data for manufacturing. Big data provides an infrastructure for transparency in manufacturing industry, which is the ability to unravel uncertainties such as inconsistent component performance and availability. Predictive manufacturing as an applicable approach toward near-zero downtime and transparency requires vast amount of data and advanced prediction tools for a systematic process of data into useful information. A conceptual framework of predictive manufacturing begins with data acquisition where different type of sensory data is available to acquire such as acoustics, vibration, pressure, current, voltage and controller data. Vast amount of sensory data in addition to historical data construct the big data in manufacturing. The generated big data acts as the input into predictive tools and preventive strategies such as Prognostics and Health Management (PHM).

C. Media& Advertising:

*Internet of Things (IoT):*In order to hone into the manner in which the media utilizes Big Data, it is first necessary to provide some context into the mechanism used for media process. It has been suggested by Nick Couldry and Joseph Turow that Practitioners in Advertising and Media approach Big Data as many actionable points of information about millions of individuals. The industry appears to be moving away from the traditional approach of using specific media environments such as newspapers, magazines, or television shows and instead tap into consumers with technologies that reach targeted people at optimal times in optimal locations. The media industries process Big Data in a dual, interconnected manner:
- Targeting of consumers (for advertising by marketers)
- Data-capture

Big Data and the IoT work in conjunction. From a media perspective, Data is the key derivative of device inter connectivity, whilst being pivotal in allowing clearer accuracy in targeting. The Internet of Things, with the help of big data, therefore transforms the media industry, companies and even governments, opening up a new era of economic growth and competitiveness. The wealth of data generated by this industry (i.e. Big Data) allows Practitioners in Advertising and Media to gain an elaborate layer on the present targeting mechanisms utilized by the industry.

Technology: *ebay.com* uses two data warehouses at 7.5 petabytes and 40PB as well as a 40PB Hadoop cluster for search, consumer recommendations, and merchandising. Inside eBay's 90PB data warehouse.

Amazon.com handles millions of back-end operations every day, as well as queries from more than half a million third-party sellers. The core technology that keeps Amazon running is Linux-based and as of 2005 they had the world's three largest Linux databases, with capacities of 7.8 TB, 18.5 TB, and 24.7 TB.

Facebook handles 50 billion photos from its user base.

D. Science & Research:

The *Large Hadron Collider* experiments represent about 150 million sensors delivering data 40 million times per second. There are nearly 600 million collisions per second. After filtering and refraining from recording more than 99.99995% of these streams, there are 100 collisions of interest per second [6].
As a result, only working with less than 0.001% of the sensor

stream data, the data flow from all four LHC experiments represents 25 petabytes annual rate before replication (as of 2012). This becomes nearly 200 petabytes after replication. If all sensor data were to be recorded in LHC, the data flow would be extremely hard to work with. The data flow would exceed 150 million petabytes annual rate, or nearly 500 Exabyte per day, before replication. To put the number in perspective, this is equivalent to 500 quintillion (5×10^{20}) bytes per day, almost 200 times more than all the other sources combined in the world.

The *Square Kilometer Array* is a telescope which consists of millions of antennas and is expected to be operational by 2024. Collectively, these antennas are expected to gather 14 Exabyte and store one petabyte per day.

When the *Sloan Digital Sky Survey* (SDSS) began collecting astronomical data in 2000, it amassed more in its first few weeks than all data collected in the history of astronomy. Continuing at a rate of about 200 GB per night, SDSS has amassed more than 140 terabytes of information. When the Large Synoptic Survey Telescope, successor to SDSS, comes online in 2016 it is anticipated to acquire that amount of data every five days.

The *Human Genome* originally took 10 years to process for decoding. Now it can be achieved in less than a day: the DNA sequencers have divided the sequencing cost by 10,000 in the last ten years, which is 100 times cheaper than the reduction in cost predicted by Moore's Law. The *NASA Center* for Climate Simulation (NCCS) stores 32 petabytes of climate observations and simulations on the Discover supercomputing cluster.

REFERENCES

[1] Judith Hurwitz, Alan Nugent, Dr. Fern Halper, Marcia Kaufman: ―Big Data for Dummies"; Wiley Publications; Year 2013

[2] O'Reilly Media Inc: ―Big Data Now, Current Perspectives from O'Reilly Media"; 2012 Edition

[3] CRC Press, Taylor & Francis Group: ―Big Data Computing" Edited by Rajendra Akerkar

[4] Wikipedia Publications; http://en.wikipedia.org/wiki/Big_data

[5] Xiaomeng Yi, Fangming Liu, Jiangchuan Liu, and Hai Jin: ―Building a Network Highway for Big Data: Architecture and Challenges"; IEEE Network July / August 2014

[6] https://www.linkedin.com/pulse/20131113065157-64875646-the-awesome-ways-big-data-is-used-today-to-change-our-world

[7] Zibin Zheng, Jieming Zhu, and Michael R. Lyu: ―Service-generated Big Data and Big Data-as-a-Service: An Overview".

Challenges in building a DBMS Resource Consultant

Shirshendu Maitra
Assistant Professor
TIMSCDR

Jayashree Jha
Student, TIMSCDR

Nikita Birid
Student, TIMSCDR

Neelam Jaipuria
Student, TIMSCDR

Mamta Mahadik
Student, TIMSCDR

Abstract - **Administration increasingly dominates the total cost of ownership of database management systems. A key task, and a very difficult one for an administrator, is to justify upgrades of CPU, memory and storage resources with quantitative predictions of the expected improvement in workload performance. We discuss the design issues and challenges involved in building a Resource Consultant.**

I. INTRODUCTION

Administering database management systems (DBMS) is a complex and increasingly expensive task. There is a pressing need to raise the level of abstraction at which database administrators (DBAs) interact with the system, by automating tasks which currently require substantial human effort and expertise.

This research paper focus on the task of resource (re)provisioning: determining the number, type, and configuration of hardware resources most appropriate to a given workload, hardware budget, and performance goals. The cost of experts is major for large enterprises and prohibitive for small ones. Even experts find it difficult to quantify the expected benefit of a resource upgrade. The net result is over-provisioned systems with no guarantees on performance. The key technical challenge in automating resource provisioning decisions is automated prediction of performance in hypothetical hardware configurations. In other words, we wish the system itself to provide accurate, quantitative answers to "what-if" questions such as "what would be the increase in throughput if the server's main memory were doubled?"

II. DESIGN PRINCIPLES

Large commercial databases are complex systems that depend on several physical resources such as the back end storage system, volatile main memory and CPUs. A database administrator (DBA) must decide on a good initial configuration of these resources, and then continuously monitor the system for new bottlenecks and changes in workload. To do this she must have an intimate understanding of the various database components, their interactions, and of the workload. Such skilled DBAs are expensive and even they do not have the tools to accurately and easily predict the performance effect of any resource provisioning decision. Consider a DBMS running multiple application workloads with different resource demands and performance requirements. OLTP workloads have high concurrency and require not only high throughput but also bounded response time. For any proposed resource provisioning the DBA must estimate the impact on the performance of each workload, taking into account the resource contention between them.

The most common approach to (re)provisioning such systems is to monitor the performance counters provided by most commercial systems. These counters measure aggregate load statistics for various resources, which is not always sufficient to find the global bottlenecks. They do not offer any insights into response time, as they do not track per-request resource usage or distinguish between critical-path and background resource usage.

We advocate a system architecture that addresses these problems by

1. Tracing per-request resource usage and control flow at fine granularity.

2. Modelling hardware resources and the algorithms that schedule or share them across multiple requests.

3. Predicting performance on hypothetical hardware by combining workload traces with hardware models.

The bottleneck in DBMS provisioning today is the human in the loop. CPU cycles are relatively abundant, allowing fine-grained yet low-overhead tracing of the live system, as well as offline trace processing using idle cycles.

III. EXPERIENCE

Based on the principles and high-level design described above, we have designed and implemented a Resource Consultant for SQL Server, which predicts the performance of a live workload under hypothetical hardware upgrades. Here we briefly describe our experiences with an early prototype based on analytic models. We then describe our current simulator-based approach.

A. Analytic modelling

Figure 1 shows the high-level design of the Resource Consultant. It relies on fine-grained, low-overhead event tracing from an instrumented DBMS.

The instrumentation points are chosen to enable end-to-end tracing of each request from the moment it enters the system to its completion. We record each use of system resources— CPU, memory, I/O — as well as virtual resources such as locks.
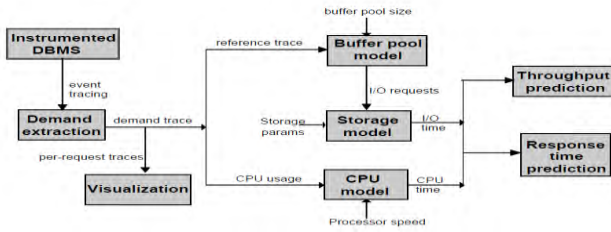
Fig 1: Resource Advisor Architecture

Fig 3: Predicting OLTP performance for hypothetical memory changes

Table 1 shows the set of events traced by our instrumentation. These events allow the Resource Consultant to reconstruct exactly the sequence of resource demands issued by the workload. Since this sequence is an aggregate of many concurrently executing requests, the Resource Consultant first separates it out into per-request demand traces.

	Event Type	Arguments	Description
Control Flow	StartRequest		SQL transaction begins
	EndRequest		SQL transaction ends
	EnterStoredProc	procname	Stored procedure invocation
	ExitStoredProc	procname	Stored procedure completion
CPU scheduling	SuspendTask	taskID	Suspend user-level thread
	ResumeTask	taskID	Resume user-level thread
	Thread/CSwitchIn	cpuID, sysTID	Schedule kernel thread
	Thread/CSwitchOut	cpuID, sysTID	Deschedule kernel thread
Buffer pool activity	BufferGet	pageID	Fetch a page (blocking)
	BufferAge	pageID	Reduce the "heat" of a page
	BufferTouch	pageID	Increase the "heat" of a page
	BufferDirty	pageID	Mark a page as dirty
	BufferReadAhead	startpage, numpages	Prefetch pages (non-blocking)
	BufferEvict	pageID	Evict and free page
	BufferNew	pageID	Create a new page
	BufferSteal	numpages	Allocate memory from free pool
	BufferFree	bufferID	Release memory to free pool
Disk I/O	DiskIO	startpage, numpages	Asynchronous read/write pages
	DiskIOComplete	startpage, numpages	Signal read/write completion
Locking	EnterLockAcquire	resourceID, mode, timeout	Attempt to lock a resource
	ExitLockAcquire	status	Success/failure of lock acquisition
	LockRelease	resourceID, mode	Release a held lock

Fig 2: Instrumentation events

This requires instrumentation of all context switches: points where a resource such as CPU stops working on one request and starts work on another. The raw event trace is transformed into a per-request demand trace, where each request is represented as to some extent ordered set of resource demands, each for a specific resource. The aggregate demand on the system is then the effect of concurrently executing these per-request demands. Subsequent steps in processing are parameterized by the characteristics of the hypothetical "what-if" hardware: The buffer cache memory size, the CPU clock speed, and disk parameters such as rotational speed. The buffer references are processed by a cache simulator to generate an I/O trace, and the I/O and CPU traces are fed to analytic models that predict the throughput and mean response time of each transaction type. Our analytical models are able to accurately predict the effect of changing the buffer cache memory on the throughput and response time of an OLTP workload. Figure 3 shows the prediction accuracy for two different types of "what-if" questions. Thus the models have good accuracy but restricted applicability. They make two major assumptions about The workload, which are valid for OLTP but not for other workloads such as DSS that buffer cache misses cause a random-access I/O pattern;_ that the throughput bottleneck remains the same throughout the workload execution, i.e. the workload does not have different phases with different bottlenecks.
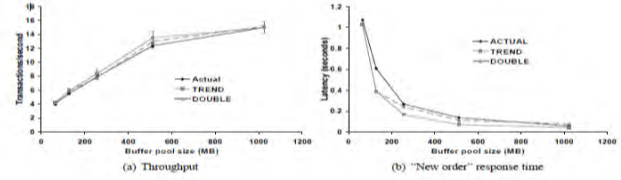
With analytic models based on operational analysis, it is easy to predict aggregate throughput, assuming Sufficient concurrency that the bottleneck resource is always busy. However, if there are multiple concurrent Users, each with a different workload then it is difficult to predict the throughput of each user individually. Analytic models also make it difficult to predict response time. Our models predict mean response time per transaction type but are specific to OLTP. They also assume that a request's response time is dominated by its resource demands rather than queuing and scheduling delays caused by concurrently executing requests. To correctly model queuing and scheduling delays, and to compute second-order metrics such as variance in response time, we need queuing models. However, analytic queuing models rely on assumptions about request arrival time distributions that are often unrealistic.

B. Simulation-based modelling

The current version of the Resource Consultant is based on event-driven simulation rather than analytical throughput and response time models. Live workload traces are decomposed into per-request demand traces as before, and the concurrent execution of the requests on hypothetical hardware is modelled by the simulator. The result of the simulation is an execution trace with the predicted timing information (including scheduling delays) of each resource demand within each request. This allows us to compute the predicted throughput, response time, or another performance metric of interest. Unlike the analytic models, the simulation approach is workload-agnostic and also enables a wider range and a finer granularity of performance metrics. Each request's demands are executed by the simulator according to the partial order specified in the demand Trace. Demands are executed by passing them to the appropriate resource model, which determines their Completion time by adding any queuing/scheduling delay as well as the predicted service time: The CPU model computes scheduling delay by simulating a non-preemptive FIFO scheduler. Service times computed by scaling, i.e. the speed of CPU execution is assumed to be proportional to the clock speed._ the buffer cache model simulates an LFU eviction policy. Cache misses generate disk demands which are handled by the disk model. Track and sector positions are inferred from LBNs (logical block

numbers) based on the known disk geometry, and seek and rotational times are inferred from these. Based on this, the disk model is able to simulate the SSTF (shortest seek time first) scheduling policy used by most disks. The lock model handles requests for locks at various granularities —page, record, etc. —and in different Modes — shared, exclusive, shared-intention, etc. It uses the same default policies as the DBMS to make Decisions on competing lock grant/upgrade requests.

Simulation has a higher overhead than analysis but is still typically much faster than real time: we simulate A CPU computation of arbitrary length in constant time and a disk access with a few cycles of computation. For in-memory, lock-bound workloads simulation is slower than real time, as the simulator's buffer and lock management are no faster than that of the DBMS.

IV. ONGOING CHALLENGES

There are many open questions on designing, building, and deploying systems such as the Resource Consultant.
Here we present some of these questions and our thoughts on answering them.

A. Granularity

What is the best granularity to represent resource demands? For example, we represent a CPU "Demand" as a single number: the number of cycles of computation. Including information such as L2 cache Misses and the integer/floating point instruction mix could allow "what-if" questions about different processor architectures rather than just different clock speeds. However, this finer granularity comes at the cost of increased complexity in instrumentation and modelling. We envision a need for models at multiple levels of complexity, with the DBA using a "drill-down" approach to increase complexity where needed. For example, crude CPU and disk models might suffice to indicate that a faster CPU would be more valuable than a faster disk. The DBA could then use a more refined CPU model to exactly quantify the performance benefits of different processor upgrade options.

B. Scope.

How much of the system should we model? The key insight that makes performance prediction feasible is that we only need to model those aspects of the system that affect performance and are affected by resource availability. Aspects which are essential to the correct functioning of the system but independent of resource availability can be ignored. Thus we must trace the system at a level above that of the resource manager but below that of any resource agnostic components, to avoid the complexity of modelling them. However, it limits us to modeling resource-agnostic query optimisers that are not adaptive to changes in resource availability but make decisions solely based on data statistics and cardinality estimates.

C. Evolution.

When building a Resource Consultant for a legacy DBMS, we chose to insert only passive instrumentation, while maintaining the simulation/analytic models separately. However, this introduces the additional burden of keeping the models consistent with the DBMS components as they evolve. For example, if the lock scheduling algorithm changes in the DBMS, a corresponding change must be made to the lock model. If the code itself is restructured, then the tracing instrumentation points may need to be changed; if these results in a change in the semantics of the traced events, this will cause a change in the models as well. We surmise that tighter integration of predictive models with DBMS components, i.e. making each component truly self-predictive, will help to alleviate this problem. However, we currently lack the programming tools and techniques for developers to maintain a performance model for each component in tandem with its functionality.

D. Hierarchical models.

The drill-down approach also requires us to ask and answer "what-if" questions at Different component granularities. For example, the storage component could be a file server with a network RAID back end. For the initial phase of resource planning, the DBA might simply ask "What if the entire storage subsystem was twice as fast?" If the predicted benefits of this look promising, she might investigate different ways to achieve this speedup, for example "What if I made the file server 4-way SMP" or "What if I moved from mirroring to RAID-5?"

Administrative boundaries:-In a typical 2- or 3-tier architecture, there are multiple components — application servers, database servers, networked storage —typically from different vendors and possibly with different administrators. We could hope that in the future each of these would be self-predicting, but it is likely that they will provide this prediction as a "black-box" functionality that does not expose model internals. Thus the tight integration of different predictive components that we use in the Resource Consultant may not be feasible. Rather than predict the performance of individual resource demands, we might have to process the entire workload trace with the DBMS to create a "storage access trace" and pass that to the storage model to get the timings of the I/Os generated. Since the I/O timings would affect the timings of the entire workload, we would have to iterate this process to converge on a solution.

E. Distributed modeling.

End-to-end performance prediction for large distributed systems is a significant challenge. Individual hosts can efficiently generate local trace information; however, a request in a multi-tiered or clustered configuration might trigger activity on multiple hosts. Backhauling all event traces to a centralised location is a simple but non-scalable solution, and hence we need distributed modelling and prediction algorithms.

V. Conclusion

This paper presented a design and implementation for a database Resource Consultant that predicts the impact of changing resource availability on workload performance. The Resource Consultant is based on fine-grained,low-overhead tracing; per-request demand extraction and simple, lightweight, workload-agnostic resource and performance models. We intend to tackle more challenging aspects of performance prediction, such as the effect of locking on response time, as well as performance.

References

[1] http://research.microsoft.com/apps/pubs/default.aspx?id=68106
[2] http://searchdatamanagement.techtarget.com/resources/Database-management-systems-DBMS
[3] http://effectivedatabase.com/resources/five-steps-to-ensure-the-success-of-your-database/
[4] http://effectivedatabase.com/resources/skills-required-of-a-database-manager/
[5] http://searchdatamanagement.techtarget.com/tutorial/Understanding-the-role-of-consultants-and-building-client-relationships
[6] http://www.tcs.com/resources/white_papers/Pages/Dynamics-Right-Database-Platform-Selection.aspx

Cultural Differences in Online Shopping Behavior: India and Australia

Pankaj Mudholkar
Assistant Professor
TIMSCDR

Tripti Tiwari
Student, TIMSCDR

Reshma Tiwari
Student, TIMSCDR

Atul Singh
Student, TIMSCDR

Abstract— **Internet usage and online shopping are increasing in popularity. Consumers from different cultures and with different consumption values are using online shopping due to benefits such as ease of search and order, and entertainment. The present study explores cultural differences in online shopping behavior and consumption value using data from a sample of consumers in India and the Australia. The findings reveal that there are differences in the online shopping behavior of Indian and Australian consumers.**

Index Terms — Internet shopping adoption; Intention to shop online; Culture in country of origin; Culture in country of residence; Indian women.

I. INTRODUCTION

The internet has been identified as the world's fastest growing market place with an enormous increase in the globalization environment for marketing products and services (Domains 1999). Projected estimates indicate that the worldwide internet population may reach 1.35 billion by the end of current year (Internet World Stats 2007; eTForecasts 2004; Gong et al., 2007). The present scenario depicts that globally Asia scores the highest global online percentage (34%), followed by Europe (29.2%) and North America (24.9%) (Internet World Stats 2007). Though the internet has been utilized widely through out the world, internet access is uneven across countries. Considering, the internet usage in Asia-Pacific region alone, variations do persist across different countries. Current internet usage with the population is forecasted to be 63.3% of South Korea, 35.9% of Malaysia, 12.8% of Thailand and 3.6% of India (Internet World Stats 2005). It has been projected that enormous growth of internet users in near future could be expected from populous countries such as China, India, Brazil and Indonesia (eTForecasts 2004). Internet shopping is becoming increasingly popular and internet sales are estimated to grow from $172 billion in 2005 specifications needed for preparing electronic versions of their to $329 billion in 2010. According to the worldwide statistics, there are 32 countries with more than 50% of internet penetration rate. The ability of the internet users has been increased from 16% to 32% .Internet shopping adoption has initiated the research in recent years, apart from attracting consumers either from a consumer oriented or technology oriented view. Reinforcement of these views focus on consumer's beliefs of internet purchases integrating with the technical specifications of these internet purchases. Consumer research is critical to the success of any

market. However there might be differential adoption patterns among the consumers. Intensification of competition of the e-commerce paves the way to understand the antecedents that influence the consumers to internet shopping as it is critical for building customer relationship. This practice has been identified as the key aspect of effective business strategy in obtaining success in the e-markets .As such internet marketers need to assess the varying receptivity levels of consumers in different countries and factors that affect the internet use and access across countries.

II. ORGANIZATION OF THE PAPER

A conceptual model has been presented from the existing theoretical background. Indian women residing in Australia, apart from India form a rich source of database for the study. Further factors influencing their intention to shop online and its subsequent effect on adoption of actual internet shopping are portrayed. The impact of prevailing culture in country of origin (CCOO) and culture in country of residence (CCOR) are compared and the effect of socio-demographics on adoption of internet shopping has been studied in detail. A brief relevant review of literature leading to a conceptual framework is presented. Results are presented and analyzed followed by discussion. Implications for managers are provided and limitations are addressed.

III. THEORETICAL FRAMEWORK

The theory of reasoned action (TRA), theory of planned behavior (TPB) and technology acceptance model (TAM) have been used extensively in the existing literature to explain and predict consumer behavior in an online environment. TAM explains how actual adoption is influenced by intention to use, which is in turn influenced by consumer's attitude towards usage. Perceived use and perceived ease of use directly affect the consumer's attitudes. Where as, diffusion of innovations (DOI) theory explains how adoption takes place within a social system. In the present study adoption of internet shopping is considered as an innovative method of shopping contrast to the traditional mall shopping.

The adoption rate of an innovation is further influenced by characteristics of the innovation itself; communication channels, time elapsed since the introduction of the innovation and the social system in which the diffusion of innovation

takes place. DOI research has widely focused on the perceived attributes of an innovation that affect the rate at which it is adopted within a social system.

The schema for evaluating the perceived attributes of an innovation involves five constructs relative advantage, compatibility, trial ability, observability and complexity. A large number of studies later incorporated the construct perceived risk which is of a particular relevance in a service context With reference to the DOI theory, studies can be conducted both at individual and aggregate levels.

The rate of adoption of different innovations and the extent to which an innovation is adopted within communities, countries or social units of varied demographic, economic and cultural situations could be studied.

The role of culture may be either implicit or explicit in these types of comparisons. Although TAM model has been used extensively to study different environments, characteristics that are specific to internet shopping are not captured by this model.

The generic information provided by the TAM model alone is not sufficient to motivate the consumer's to perform online shopping. Therefore in the present study an attempt has been made to integrate TAM and DOI models and an extended model is presented in the conceptual framework. Further study progresses by empirically studying the effects of the cultural dimensions prevailing in the country of origin and in country of residence on adoption of internet shopping by Indian women.

IV. CONCEPTUAL MODEL AND PROPOSITION DEVELOPMENT

Based on the theoretical framework the following conceptual model has been developed and hypotheses are stated to empirically test the expected relationships between the variables.
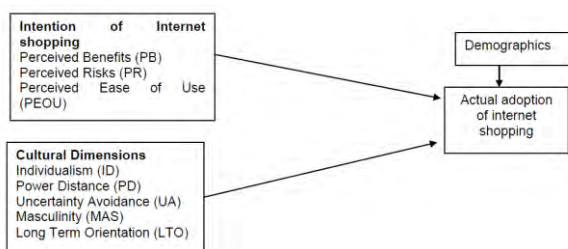
Figure 1: Conceptual Model

The perceived attributes of an internet shopping affect initially the intention to shop online and subsequently affect the actual adoption of internet shopping within a social system Perceived attributes according to the existing literature includes five constructs: relative advantage, compatibility, observability, trial ability and complexity. Later the construct perceived risk has been added in many studies.

The above mentioned attributes are incorporated as they have a profound influence in affecting the consumer's

intention to shop online. The cultural classification by Hofstede depicts five empirically derived cultural dimensions: individualism-collectivism, power distance, uncertainty avoidance, masculinity-femininity and long term-short term orientation (Hofstede 2001). Therefore in the present study the cultural framework proposed by Hofstede has been used. Though Hofstede's framework is based on the work related values, has been used Intention of Internet shopping Perceived Benefits (PB) Perceived Risks (PR) Perceived Ease of Use (PEOU) Demographics Actual adoption of internet shopping Cultural Dimensions Individualism (ID) Power Distance (PD) Uncertainty Avoidance (UA) Masculinity (MAS) Long Term Orientation (LTO).

V. INTENTION AND ADOPTION OF INTERNET SHOPPING

Various constructs of perceived attributes are explained as follows: ***Relative advantage*** is the degree to which an innovation (in the current study internet shopping is considered to be an innovation) is perceived by potential adopters as being better than the idea, product or service it supersedes (Rogers 1962). It is not the superior performance of an innovation, but rather the superiority of performance as perceived by the consumer (Szymigin and Bourne 1999). ***Compatibility*** is the degree to which an innovation is perceived as consistent with past values, experiences and needs of the potential adopter (Rogers 1962). This concept relates to the consumer's familiarity with the use of internet. ***Trial ability*** refers to the degree to which an innovation is perceived as being trial able on a limited basis prior to any decision to adopt (Rogers 1962). It is often identified as an opportunity to trial and is an effective means of reducing the perceived risk. ***Observability*** of an innovation is the degree to which an innovation is visible to the other members of a social system (Rogers 1962). The more visible is an innovation and its potential benefits, the more the rate of its adoption. The above mentioned constructs for the purpose of the present study are considered as the perceived benefits (PB).

Complexity is the degree to which an innovation is perceived as relatively difficult to understand. *Uncertainty* plays a key role in adoption decision in the form of perceived risk. ***Risk*** is often associated with outcome uncertainties such as lack of knowledge regarding distribution of potential outcomes and uncontrollability of outcome attainment. ***Security*** is the extent to which an individual believes that internet is secure for transmitting sensitive information. It is a relevant construct as internet shopping often involve transfer of secure information in the form of credit card numbers, bank account details etc., (Salisbury *et al.,* 2001). Another interesting construct is trust. ***Trust*** is the extent to which an individual trusts internet purchases as well as internet transactions. Trust is important for performing repeat transactions and often attracts a good number of potential consumers through word of mouth communication from satisfied existing consumers. These constructs in the current research are considered as the

perceived risks (PR) associated with the internet shopping.

Ease of use is defined as the degree to which an individual believes that using a particular system would be free of effort. Technology comfort level, length, intensity and type of internet use are treated as perceived ease of use (PEOU). ***Internet experience*** is likely important in understanding its influence on intention of internet purchases. Adoption of computer technology by internet users often creates a belief in their ability to use internet for commercial purposes. ***Type of internet usage*** depends upon the consumer's orientation towards internet purchases. These could be utilitarian (dealing with a specific information) or hedonic (for fun or pleasure). These constructs simplify the consumer's intention of adopting online shopping.

Perceived attributes play a key role in influencing the consumer's intention to shop online. Intention to shop online paves way for actual adoption of internet shopping. Positive associations are exhibited between intention and actual adoption of internet shopping. Intention of purchases on internet exhibited varied effects ranging from very strong to optimum on actual adoption behavior.

VI. Culture in Country of Origin (CCOO) and Culture in Country of Residence (CCOR)

Culture represents a shared set of values consisting of objective and subjective realities which comprise socialized predispositions and beliefs that guide individual's perceptions of observed events and personal interactions, and the selection of appropriate responses in social situations (Johansson 1997). Based on Hofstede's typology of culture, there are five cultural dimensions: individualism-collectivism, power distance, uncertainty avoidance, and masculinity-femininity and long-term versus short-term orientations (Hofstede 1984, 1991). Culture is defined as the collective programming of the mind which distinguishes the members of one group or category of people from another.

Individualism-collectivism dimension describes the relation between the individual and the group. Members from individualistic cultures tend to exhibit more favorable attitudes towards uniqueness and differentiation. Power distance relates to the extent to which the members of a society accept that power is distributed unequally. Cultures that exhibit large power distance are predicted to be less innovative. Uncertainty avoidance is the degree to which societies can tolerate uncertainty and ambiguity. Cultures with low uncertainty avoidance tend to more innovative due to greater tolerance for risk. Masculinity-femininity refers the extent to which a society is closely related to the gender roles. Masculinity is often related to the diffusion of technological product innovations and femininity is more related to diffusion of cosmetics and appare. Long-term versus short-term orientation measures people's consideration of the future.

While purchasing products, consumers may face many decisions with respect to its purchase and intended usage. The effect of country of origin and its conspicuousness on purchase of products is evident from the existing literature. However, as of now there are no studies relating to country of residence and its further impact on internet purchases. In the present study, country of residence is defined as "the country the individual is currently residing with a different country of origin (India)".

A. Demographics

Consumer demographics impact on adoption of internet shopping has been studied frequently since 1990's with reference to gender, age, education and income. From the existing literature, the effect of age on internet purchases is unclear. Some studies identified a positive relationship and others reported a negative relationship. However, few other studies depicted no relationship between age and adoption of internet shopping. Similarly, with regard to education also, mixed effects are identified. Some results indicate a positive relationship and others negative between the variables education and internet shopping. Income exhibited a positive relationship with the adoption of internet shopping.

B. Choice of Countries

Women of Indian origin, but currently residing in India and in Australia are selected. The traditional roles exhibited by Indian women are changing at a faster pace due to increased levels of education and changing societal situations. Also currently more number of women are entering into the information technology sector with technological advancements Therefore, it is worthwhile to examine the patterns exhibited by Indian women in the adoption of internet shopping. Cultural patterns exhibited by the people residing in these countries are different based on the Hofstede's typology of cultural dimensions. India as a developing country, culturally is collectivistic with masculine culture and exhibiting high levels of power distance, high uncertainty levels and with low long-term orientation. Australia is a developed country with prevalent individualistic and feminine culture, low levels of power distance, uncertainty avoidance and focusing more on long-term orientation .Apart from the prevailing cultural differences, differences across countries may also relate to gross domestic product (GDP) per capita, infrastructure of the country, literacy levels, openness, media access etc.. Strong associations between the cultural considerations in the country of residence and the penetration rates of internet as well as adoption of internet shopping are focused in the present research.

VII. RESEARCH METHODOLOGY

Data Collection and Analysis: The analysis for addressing the research question about the effect of culture on online shopping behavior reveals significant differences in all dimensions of online shopping behavior between the Indian and Australian samples

Online Shopping Frequency	13.40	0.00
Goods Purchased		
Apparel	6.81	0.01
Books/Magazines	15.68	0.00
Computer hardware	6.38	0.01
DVD/ Video	15.51	0.00
Electronics	6.78	0.01
Health/Beauty	2.47	0.00
Food/Drink	2.23	0.12
Home/Garden	12.85	0.00
Gift/Flower	27.64	0.14
Sport/Hobbies	16.68	0.00
Toys/Games	23.61	0.00
Online Site Type Used		
Manufacturer site	15.19	0.00
Mortar retailer site	14.43	0.00
Catalogue hybrid site	20.69	0.00
Pure dot.com site	13.50	0.00
Mall site	17.99	0.00
Broker site	3.63	0.06
Payment Method		
Online payment with giving credit card information	0.10	0.76
Online payment with virtual credit card	0.02	0.88
Online payment with direct money transfer	3.18	0.08
Cash on delivery	49.88	0.00

Shopping frequencies of Australian and Indian online shoppers are significantly different. Regarding the goods purchased dimension of online shopping behavior, the level of online buying amounts for goods purchased is significantly different for nine of the eleven product categories investigated In the food/drink and gift/flower product categories, the level of online buying amounts between the Australian and Indian consumers do not differ significantly.

The online shopping sites used for purchases also differ significantly for Indian and Australian consumers. The differences in the usage levels of manufacturer sites, mortar retailer sites, catalogue hybrid sites and pure.com sites are highly significant. Australian customers have a higher usage level for all types of online shopping sites relative to Indian consumers. Australian and Indian online shoppers' usage levels of payment methods for online purchases differ significantly for two of the four payment methods (direct money transfer and cash on delivery). These payment methods are used at higher levels by Indian consumers. The usage levels for credit cards and virtual credit cards for online payment are not significantly different between the two samples.

Effect of Intention of Internet Shopping, Cultural Dimensions and the demographics Income and Education in Country of Origin and Country of Residence on the Adoption of Internet Shopping

Study the impact of intention of internet shopping and cultural dimensions on the actual adoption of internet shopping in country of origin and country of residence multiple regression analyses has been performed and presented in Table III and Table IV in Model 1 without controlling for any other relevant variables in the study. Results reveal that the variables perceived benefits and perceived ease of use are associated positively with the adoption of internet shopping in country of origin and country of residence and the variable perceived risks is negatively associated with the adoption of internet shopping. Among the cultural dimensions, in India the variables individualism and long term orientation are associated negatively with the adoption of internet shopping and the dimensions power distance, uncertainty avoidance and masculinity are positively associated with the adoption of internet purchases. Where as in Australia, the variables individualism, masculinity and long term orientation are associated positively and the dimensions power distance and uncertainty avoidance are associated negatively with the adoption of internet shopping. Most of the results obtained are in conformity with the previous studies. However, Australia as a developed nation, the expected relationship of Indian women residing in Australia tend to exhibit increased levels of internet shopping is void as the results indicate positive association of the masculinity dimension with the adoption of internet shopping thus partially contradicting the earlier stated hypothesis.

VIII. DISCUSSION AND CONCLUSION

The present research explores the effect of culture on online shopping behavior and consumption value, as well as the effect of consumption value on online shopping behavior. The findings indicate that online shopping behavior significantly varies between the two distant cultures studied, and that culture does not have a significant differentiating effect on consumption value. Results obtained have implications on venture capitalists and e-commerce business strategists. As venture capitalists increasingly try to invest in overseas operations, it is important to consider prevailing cultural situations for diffusion of the internet as well as for the development of e-commerce. The role of women is constantly changing in today's competitive world due to increase in their level of education and working capabilities. The internet usage rate of the women is also increasing at a faster pace than in the past. Therefore it is important for marketers to understand their consumers and formulate effective marketing strategies. Marketers need to monitor and evaluate the cultural conditions both in the country of residence and in the country of origin and their further impact on the adoption of internet shopping.

IX. Acknowledgment

This research paper is made possible through the help and support from everyone, including: parents, teachers, family, friends, and in essence, all sentient beings. Especially, please allow us to dedicate our acknowledgment of gratitude toward the following significant advisors and contributors:

First and foremost, we would like to thank Mr. Pankaj Mudholkar for his most support and encouragement. He kindly read our paper and offered invaluable detailed advices on grammar, organization, and the theme of the paper.

We also thank Thakur College of Management CD & R for providing us withal the resources in our college to work on the paper.

Finally, we sincerely thank to my parents, family, and friends, who provide the advice and their support. The product of this research paper would not be possible without all of them.

References

[1] Aaker, J. L and Maheswaran, D (1997), "The Effect of Culture Orientation on Persuasion", *Journal of ConsumerResearch*, Vol. 24, pp. 315-28.

[2] Ajzen, I (1991), "The Theory of Planned Behavior", *Organizational Behavior and Human Decision Processes*, Vol. 50, pp. 179-211.

[3] S.J. Yoon, The antecedents and consequences of trust in online-purchase decisions. *Journal of Interactive Marketing,* 16(2), p47-63, 2002. http://dx.doi.org/10.1002/dir.10008.

Deadlock Detection and Prevention in Distributed Systems

Shirshendu Maitra	Juhi Naik	Tina Lohar	Sapna Jaiswar	Sharvari Barve
Assistant Professor *TIMSCDR*	*Student, TIMSCDR*	*Student, TIMSCDR*	*Student, TIMSCDR*	*Student, TIMSCDR*

Abstract - **In a distributed database environment, data is spread across several sites there are many concerns to deal with such as concurrency control and deadlock. Deadlocks affect the overall performance of the system. A deadlock is a condition in a system where a process cannot proceed because it needs to obtain a resource held by another process which it itself is holding a resource that the other process require. In this paper we have covered the deadlock detection and resolution techniques that are used.**

Index Terms - Wait-For-Graph, Deadlock, Transactional database, Distributed database system

I. INTRODUCTION TO DEADLOCKS

In Concurrent programming, a deadlock occurs when two or more competing actions are each waiting for the other to finish, and thus neither ever does.

In a Transactional database, a deadlock happens when two processes each within its own transaction updates two rows of information but in the opposite order. Example, process P1 updates row 1 then row 2 in the exact timeframe that process P2 updates row 2 then row 1. Process P1 can't finish updating row 2 until process P2 is finished, but process P2 cannot finish updating row 1 until process P1 is finished. No matter how much time is allowed to pass, this situation never resolve itself and because of this database management systems will typically kill the transaction of the process that has done the least amount of work.In an operating system, a deadlock situation occurs when a process or thread enters a waiting state because a resource requested is being held by another waiting process, which in turn is waiting for another resource held by another waiting process. If a process is unable to change its state indefinitely because the resources requested by it are being used by another waiting process, then the system is said to be in a deadlock.
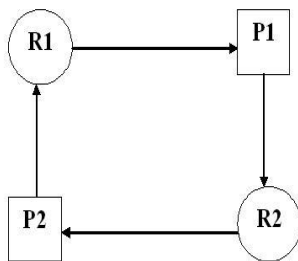
Fig. 1

II. INTRODUCTION TO DISTRIBUTED DATABASES

A distributed database is a database in which storage devices are not all attached to a common processing unit (CPU), controlled by a distributed database management system . It may be stored in multiple computers, located at same physical location; or may be scattered over a network of interconnected computers. In parallel systems, processors are tightly coupled and constitute a single database system whereas distributed database system consists of loosely-coupled sites that do not share physical components.

A distributed database can reside on network servers on the Internet, intranets, extranets, or on company networks. As they store data across multiple computers, distributed databases improve performance at end-user worksites by allowing transactions to be processed on many machines, instead of single machine.

Following processes ensure that the distributed databases remain up-to-date and current: replication and duplication Replication involves specialized software that looks for any changes in the distributive database. Once the changes are identified, the replication process makes all the databases look the same. The replication process can be complex, time consuming depending on the size and number of the distributed databases. This process also requires a lot of time and computer resources.

Duplication on the other hand, with less complexity, basically identifies one database as a master and then duplicates that database. The duplication process is normally done at a set time after hours. This is to ensure that each distributed location has same data. In duplication process, users may change only the master database. This ensures that local data will not be overwritten. Both replication and duplication keeps the data current in all distributive locations.

III. DEADLOCK DETECTION IN DISTRIBUTED DATABASES

Deadlock detection and prevention is one of the major components of a successful distributed database management system. A Distributed system consists of a collection of sites that are interconnected through a communication network each maintaining a local database system. Unfortunately, they are harder to detect, avoid, and prevent. Distributed deadlocks can occur in distributed systems when distributed transactions or concurrency control is being used.

A distributed system consists of a number of sites connected by network. Each site maintains some of the resources of the system. Processes with a globally unique identifier run on the distributed system. They make resource requests to a controller as there is one controller per site. If the resource is local, the process makes a request to the local controller. If the desired resource is at a remote site, the process sends a message. After a process makes a request, but before it is granted, it is blocked and said to be dependent on the process that holds the desired resource.

The controller at each site could maintain a Wait-For-Graph (WFG) for each process request. This is the local WFG. But, each site's WFG could be cycle free and yet the distributed system could be deadlocked. This is called global deadlock. This would occur in the following situation:

1. Process P1 at site 1 holds a lock on resource X.
2. Process P1 has requested, but has not been granted, resource Y at site 2.
3. Process P2 at site 2 holds a lock on resource Y.
4. Process P2 has requested, but has not been granted, resource X at site 1.

Both processes are blocked by the other one so there is a global deadlock. In a WFG, nodes are processes and there is a directed edge from node P1 to mode P2 if P1 is blocked and is waiting for P2 to release some resource. A system is deadlocked if and only if there exists a directed cycle or knot in the WFG.

Actual global wait-for-graph contains NO cycle

Local wait-for-graph at site A contains cycle

Local wait-for-graph at site B

Local wait-for-graph at site C
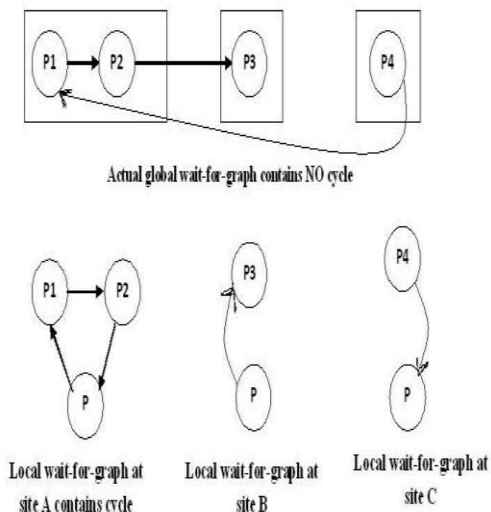
Fig. 2

IV. DISTRIBUTED DEADLOCKS PREVENTION

An alternative for detecting deadlocks is to design a system where deadlock situation is impossible. We examined the four conditions for deadlock. If we can deny at least one of these conditions then we will not have deadlock.

A. Mutual exclusion

To deny this means that we will allow a resource to be held by more than one process at a time. If a resource can be shared then there is no need for mutual exclusion and deadlock cannot occur. Too often, however, a process requires mutual exclusion for a resource because the resource is some object that will be modified by the process.

B. Hold and wait

Denying this means that processes that hold resources cannot wait for another resource. This typically implies that a process should grab all of its resources at once. This is not practical either since we cannot always predict what resources a process will need throughout its execution.

C. Non-preemption

A resource, once granted, cannot be taken away. In transactional systems, allowing preemption means that a transaction can come in and modify the resource that is being used by another transaction. This differs from mutual exclusion since the access is not concurrent but the same problem arises of having multiple transactions modify the same resource. We can support this with optimistic concurrency control algorithms that will check for out-of-order modifications at commit time and roll back (abort) if there are inconsistencies.

D. Circular wait

Avoiding circular wait means that we ensure that a cycle of waiting on resources does not occur. We can do this by enforcing an ordering on granting resources and aborting transactions or denying requests if an ordering cannot be granted.

One way of avoiding circular wait is to obtain a globally-unique timestamp for every transaction so that no two transactions get the same timestamp. When one process is about to block waiting for a resource that another process is using, check which of the two processes has a younger timestamp and give priority to the older process. If a younger process is using the resource, then the older process (that wants the resource) waits. If an older process is holding the resource, the younger process (that wants the resource) aborts itself. This forces the resource utilization graph to be directed from older to younger processes, making

cycles impossible. This algorithm is known as the **wait-die algorithm**.

An alternative, but similar, method by which resource request cycles may be avoided is to have an old process abort (kill) the younger process that holds a resource. If a younger process wants a resource that an older one is using, then it waits until the older process is done. In this case, the graph flows from young to old and cycles are again impossible. This variant is called the **wound-wait algorithm.**

REFERENCES

[1] Distributed Database Systems – by Chhanda Ray
[2] Database management systems-by Raghu Ramakrishnan
[3] Distributed Databases - Stefano CERI
[4] Tamer M. Ozsu and Patrick Valduriez, Principles of Distributed Database Systems, Second Edition, Prentice-Hall
[5] Elmasri Navathe, Database Concepts By Pearson Education
[6] Coronel Rob, Introduction to Database Concepts
[7] Chandy K. M., Hass L. M and Misra J, Distributed Deadlock Detection, ACM Transaction on Computer Systems, Vol.1, No.2, pp.144-56,
[8] Marsland T.A., and Isloor S.S. Detection of deadlocks in distributed database systems
[9] Bernstein P. and Goodman N., Concurrency Control in Distributed Database Systems, ACM Computing Surveys 13/2
[10]Singhal Mukesh "Deadlock Detection in Distributed System" IEEE transaction on Software Engineering, Vol. 4, No. 3, pp. 195-199.

E-Business Logistics: Supply Network View

Pankaj Mudholkar
Assistant Professor
TIMSCDR

Gaurav Paswan
Student, TIMSCDR

Sasmit Vaity
Student, TIMSCDR

Prajakta Pawar
Student, TIMSCDR

Rumana Shaikh
Student, TIMSCDR

Abstract- **In the present day world, India has witnessed an unprecedented index of Cyber crimes whether they pertain to Trojan attacks, salami attacks, e-mail bombing, DOS attacks, information theft, or the most common offence of hacking. Despite technological measures being adopted by corporate organizations and individuals, we have witnessed that the frequency of cyber crimes has increased over the last decade. Since users of computer system and internet are increasing worldwide in large number day by day, where it is easy to access any information easily within a few seconds by using internet which is the medium for huge information and a large base of communications around the world. Certain precautionary measures should be taken by all of us while using the internet which will assist in challenging this major threat Cyber Crime. In this paper, I have discussed various categories of cyber crime and cyber crime as a threat to person, property, government and society. In this paper I have suggested various preventive measures to be taken to snub the cyber crime.**

Index Terms - Cyber crime. Computer crime, hacking, cyber fraud, Prevention of cyber crime

I. INTRODUCTION

Cyber criminals take full advantage of the anonymity, secrecy, and interconnectedness provided by the Internet, therefore attacking the very foundations of our modern information society. Cyber crime can involve botnets, computer viruses, cyber bullying, cyber stalking, cyber terrorism, cyber pornography, Denial of Service attacks, hacktivism, identity theft, malware, and spam. Law enforcement officials have struggled to keep pace with cyber criminals, who cost the global economy billion annually. Police are attempting to use the same tools cyber criminals use to perpetrate crimes in an effort to prevent those crimes and bring the guilty parties to justice. This essay begins by defining cyber crime and then moves to a discussion of its economic and social impacts. It continues with detailed excursions into cyber bullying and cyber pornography, two especially representative examples of cyber crime, and concludes with a discussion of ways to curtail the spread of cyber crime.

Keywords Botnet: Computer Virus; Cyber Crime; Cyber bullying; Cyber stalking; Cyber terrorism; Cyber pornography; Denial of Service Attack; Hacktivism; Identity Theft; Information Society; Internet; Malware; Spam Social Impacts of Cyber Crime Overview Computer-related crime dates to the origins of computing, though the greater connectivity between computers through the Internet has brought the concept of cyber crime into the public consciousness of our information society.

In 1995, when the World Wide Web was in its early stages of development, futurist Dr. Gene Stephens wrote about the present and future reality of cyber crime and made several predictions: "Billions of dollars in losses have already been discovered. Billions more have gone undetected. Trillions will be stolen, most without detection, by the emerging master criminal of the twenty-first century—the cyberspace offender" (Stephens, 1995, p. 24).

II. OVERVIEW

There are many research questions that need to be address in detail, issues like what are we trying to protect? Our Assets, Network Infrastructure, Availability Network Resources, Confidential Personal Data can be classified into three categories. i. Information stored electronically or on transit in the network (digital) ii. Information stored physically (hard copies) iii. Information in human brain. There may be more to it that need to be researched into, all these and many more need maximum protection.

Why are intrusions so often successful? There is need for us to know why we often fall for these cybercrime activities, issues concerning our cybercrime detection, response, and scalation process needs to be reviewed, policies where applicable needs to be thoroughly looked into, sometimes, there may be no procedures for proactive auditing, or event management, focus have been may have been shifted to reactive measures to detect what happen which are often not successful. Use of authentication and/or authorization systems, Ignorance of logical and/or organizational boundaries within a network Infrastructure and many others will form the basis of research in this aspect.

What are the greatest challenges that we face today ? Environment complexity, new technologies, new threats and exploits, limited focus on security, limited budget for security, shortage of qualified security experts. There is no doubt that what some mobile devices like smart phones and others can do today, some desktop and even servers computers of the early day were not able to do. There must be a perfect trade off and an acceptable level of compromise between the security, functionality and ease of use. Why are most attacks gone undetected? Most companies don't even know if they have been a victim of cybercrime activities or not, this is very alarming, knowing fully well that been hacked is bad, but not knowing if you've been hacked is worst. This research will dig deep into why attacks often gone undetected, again focus

will also be on detection, response, escalation and monitoring mechanism. Although lack of skilled personnel in cyber security or incident response may be a cause, we'll find out! Why are attackers not getting appropriate jail terms?

Sometimes even when intrusion is detected the offenders are either not being detected or there is no enough evidence to convince the jury to convict a cybercriminal. Where there is evidence, the get light sentence. There is no international law that operates everywhere. Whereas cybercrime is a global crime, there should be no geographical boundaries in implementing cyber laws as the entire cyber space should be regarded as a single entity for a single jurisdiction. The research will see what is obtainable, desirable and applicable.

What are the motivations for the cybercriminals?

Although cyber criminals often carry out their activities for financial gains, there are some who does what they do for other reason's other than financial gains; it may sometimes be for personal reasons either as a retaliation for unlawful termination, for pride or political reasons, whatever the motivation maybe, it can also be a pointer to the type of evidence, where and how to collect evidence for a successful prosecution.Engr. Effiong Ndarake Effiong @ July 2013 Page no 6 What risk is the attacked willing to take in other to get the target ? The attacker must be willing to take certain amount of risk for the attack to be successful, this we must find out.How can we tired the loose ends and make our cyber space more secure that it is currently?

This very question will lead us to the aim of this result, making sure that our cyber space is secured, and to ensure that offenders are punished accordingly.

III. PREVENTION OF CYBER CRIME

Prevention is always better than cure. It is always better to take certain precautions while working on the net. One should make them a part of his cyber life. Sailesh Kumar Zarkar, technical advisor and network security consultant to the Mumbai Police Cyber crime Cell, advocates the 5P mantra for online security: Precaution, Prevention, Protection, Preservation and Perseverance. Identification of exposures through education will assist responsible companies and firms to meet these challenges. One should avoid disclosing any personal information to strangers, the person whom they don't know, via e-mail or while chatting or any social networking site. One must avoid sending any photograph to strangers by online as misusing or modification of photograph incidents increasing day by day.

An update Anti-virus software to guard against virus attacks should be used by all the netizens and should also keep back up volumes so that one may not suffer data loss in case of virus contamination. A person should never send his credit card number or debit card number to any site that is not secured, to guard against frauds. It is always the parents who have to keep a watch on the sites that their children are accessing, to prevent any kind of harassment or depravation in children. Web site owners should watch traffic and check any irregularity on the site. It is the responsibility of the web site owners to adopt some policy for preventing cyber crimes as number of internet users are growing day by day. Web servers running public sites must be physically separately protected from internal corporate network. It is better to use a security programs by the body corporate to control information on sites. Strict statutory laws need to be passed by the Legislatures keeping in mind the interest of netizens.IT department should pass certain guidelines and notifications for the protection of computer system and should also bring out with some more strict laws to breakdown the criminal activities relating to cyberspace. As Cyber Crime is the major threat to all the countries worldwide, certain steps should be taken at the international level for preventing the cybercrime. A complete justice must be provided to the victims of cyber crimes by way of compensatory remedy and offenders to be punished with highest type of punishment so that it will anticipate the criminals of cyber crime.

REFERENCES

[1] http://www.brainyquote.com/words/cr/crime149615.html#ixzz1h5KlT6pg
[2] http://www.crime-research.org/articles/joseph06/
[3] http://en.wikipedia.org/wiki/Computercrime
[4] http://www.webopedia.com/TERM/C/cybercrime.html

E-Commerce in India: Accelerating growth

Pankaj Mudholkar
Assistant Professor
TIMSCDR

Sundeep
Sharmadhikari
Student, TIMSCDR

Rohit Sonar
Student, TIMSCDR

Vivek Gariya
Student, TIMSCDR

Abstract -**The eCommerce sector has seen unprecedented growth in 2014. The growth was driven by rapid technology adoption led by the increasing use of devices such as smartphones and tablets, and access to the internet through broadband, 3G, etc, which led to an increased online consumer base. Furthermore, favoured demographics and a growing internet user base helped aid this growth. In terms of highlights, the growth shown by homegrown players such as Flipkart and Snapdeal and the huge investor interest around these companies displayed the immense potential of the market. With the entry of eCommerce behemoths such as Amazon and Alibaba, the competition is expected to further intensify. Both these international players come with deep pockets and the patience to drive the Indian eCommerce market. Also, their strong domain knowledge and best practices from their international experience give them an additional edge. Additionally, these companies have been part of markets where they have seen the eCommerce market evolve and are aware of the challenges and strategies to address issues there of. Indian companies realise this, and are therefore aiming to continue their focus on expanding sellers and selection on their platforms, innovating on multiple customer touch points, and providing seamless and rapid delivery services in order to compete with the international entities. Competition is expected to continue, with these eCommerce companies experimenting with different ways to attract customers and increase online traffic.**

I. INTRODUCTION

The Indian government's ambitious Digital India project and the modernisation of India Post will also affect the eCommerce sector. The Digital India project aims to offer a one stop shop for government services that will have the mobile phone as the backbone of its delivery mechanism. The programme will give a strong boost to the eCommerce market as bringing the internet and broadband to remote corners of the country will give rise to an increase in trade and efficient warehousing and will also present a potentially huge market for goods to be sold.

For India Post, the government is keen to develop its distribution channel and other eCommerce related services as a major revenue model going ahead, especially when India Post transacted business worth 280 crore INR in the cash on delivery (CoD) segment for firms such as Flipkart, Snapdeal and Amazon. Both these projects will have significant impact on increasing the reach of eCommerce players to generally non serviceable areas, thereby boosting growth.

India's overall retail opportunity is substantial, and coupled with a demographic dividend (young population, rising standards of living and upwardly mobile middle class) and rising internet penetration, strong growth in eCommerce is expected. From an investment perspective, the market is a primarily minority stake market, with maximum traction in early stage deals. Such early stage funding will help companies develop a strong foundation to start from. With such strong market prospects and an equally upbeat investor community.

II. INDUSTRY ON AN UPTURN

In 2013, Asia-Pacific emerged as the strongest business-to-consumer (B2C) eCommerce region in the world with sales of around 567.3 billion USD, a growth of 45% over 2012, ranking ahead of Europe (482.3 billion USD) and North America (452.4 billion USD). The top three were followed by Latin America, and the Middle East and North Africa (MENA) region, according to Ecommerce Europe1. Globally, B2C eCommerce sales increased by 24% over 2012.

This reflects the huge untapped potential of eCommerce by retail companies, both in their country of origin and across borders. eCommerce or electronic commerce, deals with the buying and selling of goods and services, or the transmitting of funds or data, over an electronic platform, mainly the internet. These business transactions are categorised into either business-to-business (B2B), business-to-consumer (B2C), consumer-to-consumer (C2C), consumer-to-business (C2B) or the recently evolved business-to-business-to-consumer (B2B2C). eCommerce processes are conducted using applications, such as email, fax, online catalogues and shopping carts, electronic data interchange (EDI), file transfer protocol and web services and e-newsletters to subscribers. eTravel is the most popular form of eCommerce, followed by eTail which essentially means selling of retail goods on the internet conducted by the B2C category.

According to Ecommerce Europe, country-wise, the US, UK and China together account for 57% of the world's total B2C eCommerce sales in 2013, with China having total sales of 328.4 billion USD. As against this, India had sales of only 10.7 billion USD, 3.3% of that of China in 2013 with fifth position in Asia-Pacific. This is despite the fact that India enjoys high demographic dividends just like China. India's internet penetration with total e-households at 46 million against China's 207 million is one of the reasons behind India's poor B2C sales growth.

According to Forrester Research, an independent technology and market research firm, only 16% of India's total population was online in 2013 and of the online users only 14% or 28 million were online buyers. India, therefore, was still in a nascent or immature stage of evolution of online retail spending. China was in ascending stage at 50%, whereas Japan (69%), Australia (57%) and South Korea (70%) were in mature stage.

III. INDIA'S GROWTH POTENTIAL

Since the eCommerce industry is fast rising, changes can be seen over a year. The sector in India has grown by 34%

(CAGR) since 2009 to touch 16.4 billion USD in 20142. The sector is expected to be in the range of 22 billion USD in 2015.

Currently, eTravel comprises 70% of the total eCommerce market. eTailing, which comprises of online retail and online market places, has become the fastest-growing segment in the larger market having grown at a CAGR of around 56% over 2009-2014. The size of the eTail market is pegged at 6 billion USD in 2015. Books, apparel and accessories and electronics are the largest selling products through eTailing, constituting around 80% of product distribution. The increasing use of smart phones, tablets and internet broadband and 3G has led to developing a strong consumer base likely to increase further. This, combined with a larger number of homegrown eTail companies with their innovative business models has led to a robust eTail market in India rearing to expand at high speed.

IV. FACTORS THAT WILL FUEL GROWTH

A significantly low (19%) but fast-growing internet population of 243 million in 2014 is an indicator of the sector's huge growth potential in India.

It is evident that in absolute terms India's internet users are short by only 36 million as compared with 279 million in the US and higher than that in Japan, Brazil and Russia. However, in relation with its population, only 19% Indians use the internet. This indicates the potential of internet use in India and as internet penetration increases, the potential of growth for the eCommerce industry will also increase.

An analysis of the demographic profile of internet users further testifies that eCommerce will rise rapidly in India in coming years. Around 75% of Indian internet users are in the age group of 15 to 34 years. This category shops more than the remaining population. Peer pressure, rising aspirations with career growth, fashion and trends encourage this segment to shop more than any other category and India, therefore, clearly enjoys a demographic dividend that favours the growth of eCommerce. In coming years, as internet presence increases in rural areas, rural India will yield more eCommerce business.

By 2020, eTail in India is expected to account for 3% of total retail. Further, orders per million are expected to more than double from five million in 2013 to 12 million by 2016, which will mean more opportunities for both consumers and eTail companies. While the share of online shopping in total retail has increased at a fast pace in the last few years, it is still miniscule compared to the figure in China, where the share is 8-10%.

Online business models To get the maximum benefit from eCommerce business, a large number of companies are adopting different innovative ideas and operating models including partnering with online marketplaces or setting up their own online stores. Some key operating models include the following:

• Marketplace and pick-up & drop is a model where sellers often partner with leading marketplaces to set up a dedicated online store on the latter's website. Here sellers play a key role of managing inventory and driving sales. They leverage on high traffic on the marketplaces' website and access their distribution network. However, the sellers have limited say on pricing and customer experience.

• Self-owned inventory is a model where the eCommerce player owns the inventory. The model provides better post-purchase customer experience and fulfilment. It provides smoother operations due to ready information on the inventory, location, supply chain and shipments, effectively leading to better control over inventory. On the flipside, however, there are risks of potential mark downs and working capital getting tied up in inventory.

• Private label reflects a business where an eCommerce company sets up its own brand goods, which it sells through its own website. This model offers a wide-ranging products and pricing to its customers and competes with branded labels. Here, margins are typically higher than third-party branded goods.

• White label involves the setting up of a branded online store managed by the eCommerce player or a third party. The brand takes the responsibility of generating website traffic and providing services by partnering with payment gateways. It helps build trust, customer affinity and loyalty and provides better control of brand and product experience.

V. INVESTMENTS IN THE ECOMMERCE SECTOR

With the new government being elected, business confidence has significantly improved. In 2014, investors aggressively funded the eCommerce sector due to strong growth prospects. Apart from the traditional online formats of retail and lifestyle, newer online business segments such as classifieds, real estate, grocery and healthcare were also tapped.

The eCommerce businesses will continue to attract investor interest. Several of India's blue-chip PE firms, which previously avoided investing in eCommerce, are now looking for opportunities in the sector. The focus is mainly on ancillary service providers—companies involved in support functions ranging from delivery, logistics and payments—with investments largely driven by the relatively lower valuations and smaller amounts of capital required. The last year has seen several developments that have given a fillip to the eCommerce industry.

VI. MOBILE TO BE THE MOST INFLUENTIAL ASPECT OF ECOMMERCE

With mobile apps being developed by most eCommerce websites, smartphones are increasingly replacing PCs for online shopping. In 2013, only 10% of the mobile users used smartphones, and only 5% of the eCommerce transactions were made through a mobile device. This figure has more than doubled, and more than 13% of all eCommerce transactions today happen via mobile3. According to some industry players, over 50% of the orders are being placed through

mobile apps, which is not only leading to substantial customer acquisition but also building customer loyalty for various brands. However, most mobile transactions so far are for entertainment, such as booking movie tickets and music downloads. This trend will change soon with more and more merchandise being ordered online.

VII. MORE BUSINESS COMING FROM SMALLER TOWNS

eCommerce is increasingly attracting customers from Tier 2 and 3 cities, where people have limited access to brands but have high aspirations. According to eCommerce companies, these cities have seen a 30% to 50% rise in transactions.

VIII. ENHANCED SHOPPING EXPERIENCE

Besides general online shopping, customers are also shopping online for weddings and festivals, thanks to wider range of products being offered and aggressive advertisements. The free and quick shipment and wider choice of products, along with the ease of shopping online as compared to in-store shopping, is also helping eCommerce gather momentum.4 Further, eCommerce companies are doing rapid business due to sales. New concepts such sales on weekends, holidays and festivals are attracting a lot of new customers and building customer loyalty among existing customers. Television and social media, particularly Facebook, are playing a proactive role in promoting eTailing through aggressive advertisements. This has helped several eCommerce companies build substantial brand image.

IX. EXCLUSIVE PARTNERSHIPS WITH LEADING BRANDS

Over the year or so, there has been a trend of exclusive tie-ups between eTailers and established boutiques, designers, and high-end lifestyle and fashion brands. For instance, in 2014, Jabong added international fashion brands such as Dorothy Perkins, River Island, Blue saint and Miss Selfridge, along with local fashion brands through Jabong Boutiques. Similarly, Myntra benefited from exclusive tie-ups with brands such as Harvard Lifestyle, Desigual and WROGN from Virat Kohli.

X. EXPANDING THE PRODUCTBASKET

There is a recent trend of relatively newer products such as grocery, hygiene, and healthcare products being purchased online. Similarly, lingerie and Indian jewellery has also been in great demand among customers outside India. Export comprises 95% of cross-border eCommerce, with the US, UK, Australia, Canada and Germany being the major markets.5

XI. KEY MARKET FACTORS TO BE EVALUATED BEFORE ENTERING A NEW E-COMMERCE BUSINESS

Entering a new eCommerce business To achieve their vision, eCommerce companies will need to understand the intricate landscape of new markets in addition to their own internal capabilities and limitations. The following factors must be considered:

- Market size: Before moving too aggressively into a new market, it is important to consider how sizable the overall opportunity is.
- eCommerce readiness: It is essential to fully understanding the payment and logistical infrastructure, consumer behaviour, retail opportunity and technological developments.
- Scope of growth: It is also important to look at the internet penetration, demographics of the online buying population and understand which phase of development each market is in.
- Barriers to entry: Players should understand the regulatory environment and connect with solution providers, content distribution networks, and digital agencies.
- Competition: There is also a need to do an in-depth assessment of what competitors are doing, their online strategy and the nature of each offering.

XII. INDUSTRY SPEAK

Certain eCommerce players and industry observers have raised concerns that deep discounts, free shipping, intense competition and higher rejection rates due to cash on delivery (CoD) have impacted online eTailing adversely. Some of these concerns are specific to India and are more difficult to overcome than issues such as internet penetration and getting more people to shop online.

Some of the key concerns are listed below:

• Generation and sustenance of traffic: Competition from established eCommerce players is making it difficult for private label brands to generate traffic on their white-label websites.

• High customer acquisition cost: The customer acquisition costs have been rising due to intense competition by the relatively better off companies with more funds.

• Last-mile delivery: Poor last-mile connectivity, especially in remote areas with larger population, is another problem faced by Indian eTailers.

• High payment cost: CoD services impose substantial financial cost. In India, unlike in developed markets, CoD continues to be a preferred route of payment.

• Low profitability: Profitability is negatively impacted by high customer acquisition costs, free shipping and high rejection rate of CoD orders.

• Regulatory barriers: Regulatory barriers in the Indian eCommerce market are higher as compared to more mature markets.

• Skilled manpower: Lack of talent availability and high attrition are causing manpower crunch, which is fast becoming a hurdle.

XIII. PERSPECTIVE

Product and market strategy Customer and digital experience Payments and transactions FulfilmentOrganisation scaling Tax and regulatory structuring Risk, fraud and cyber security Compliance framework eCommerce Drivers and Challenges Internal elementsExternal elements While the growth in this sector excites entrepreneurs and financial investors alike, some serious challenges are beginning to weigh down on the sector. eCommerce players in India need to address eight key aspects of their business, both internal and external.

XIV. EXTERNAL CHALLENGES

External forces impact how eCommerce companies plan their growth strategy and provide seamless customer experience onsite and post-transaction.

• Product and market strategy: eCommerce companies have to address issues pertaining to rapidly evolving customer segments and product portfolios; access information on market intelligence on growth, size and share; manage multiple customer engagement platforms; focus on expansion into new geographies, brands and products; and simultaneously tackle a hypercompetitive pricing environment.

• Customer and digital experience: Companies have to provide a rich, fresh and simple customer experience, not geared towards discovery; manage inconsistent brand experience across platforms; manage proliferation of technologies; and handle time-to-market pressure for new applications. In the recent past, social media has become more influential than paid marketing.

• Payments and transactions: eCommerce companies may face issues around security and privacy breach and controlling fictitious transactions. Further, RBI restrictions for prepaid instruments or eWallets act as impediments. From a transactions perspective, cross-border tax and regulatory issues, and backend service tax and withholding tax can have serious implications.

• Fulfilment: Companies will need to check if the physical infrastructure gets affected by the internet speed. Also, the lack of an integrated end-to-end logistics platform and innovation-focused fulfilment option could cause delivery issues. Challenges around reverse logistics management and third party logistics interactions could also act as barriers to growth.

XV. INTERNAL CHALLENGES

Internal forces impact how eCommerce companies can organise to drive and sustain growth.

• Organisation scaling: eCommerce companies will have to make sure organisation design keeps pace with the rapidly evolving business strategy, along with fluid governance, strong leadership and management development. From a growth perspective, identifying acquisition opportunities, fund raising and IPO readiness becomes necessary. From a technology perspective, it is important to transform IT as an innovation hub and address the lack of synergy between business, technology and operations functions of the enterprise.

• Tax and regulatory structuring: Companies will need to address issues around sub-optimal warehouse tax planning; imbalance between FDI norms vis-à-vis adequate entity controls; inefficient holding, IPR or entity structures; and international tax inefficiencies. Future challenges include the new Companies Act, policy on related-party transaction pricing, and the uncertainty around GST roadmap.

• Risk, fraud and cyber security: From a risk perspective, eCommerce companies could face issues around brand risk, insider threats and website uptime. Issues around employee-vendor nexus, bribery and corruption make companies vulnerable to fines. Cyber security also raises some concerns around website exploitation by external entities.

• Compliance framework: eCommerce companies have to comply with several laws, many of which are still evolving. Potential issues around cyber law compliance, inefficient anti-corruption framework, legal exposure in agreements or arrangements, indirect and direct tax compliance framework and FEMA contraventions and regularisation could pose problems. Also, uncertainty around VAT implications in different states due to peculiar business models could cause issues.

XVI. OUTLOOK 2015

In 2015, eCommerce players see mobile commerce as the most preferred route with mobile wallet as the preferred way of payment. With 4G services expected to be launched in 2015, internet penetration is likely to take a significant leap, which is likely to give another boost to mobile commerce. Changes in lifestyle and shopping choices will see buyers preferring online and mobile channel over physical channel to save time and seek wider range and possibly comparative pricing. For mobile wallets, improvements on the payments front with multiple payment instruments and increase in payment gateways aided by enhanced security with multiple authentication layers will help the consumers with a seamless mobile experience. Niche categories are also expected to gather momentum including cars and real estate. Premium and international brands are likely to join eTail, purchases from Tier 2 and 3 cities will continue to rise and differentiated products such as exclusive brands by designers will grow. Riding high on the growth prospects eCommerce companies will look at more ways to raise funds such as IPOs. Some consolidation will also happen with the leaders focusing on high growth and smaller players finding their own niche.

XVII. TOP 10 THINGS THE ECOMMERCE COMPANIES NEED TO DO TO ACCELERATE GROWTH

Customer Experience: As the customers progress from research to purchase to fulfilment stages, their expectations change fast. eCommerce companies need to understand these change drivers and adapt their proposition accordingly. Easy transitions between ordering on tablets, mobile phones or PCs will have to be facilitated. Besides, convenient multichannel returns and delivery options need to be developed along with

the provisions of touch and feel the product before buying. They should also ensure sufficient after sales service and support. Online product reviews and ratings, videos, more advanced sizing and fitting tools should be provided.

Convergence of online and offline channels: As the customers progress from research to purchase to fulfilment stages, their expectations change fast. eCommerce companies need to understand these change drivers and adapt their proposition accordingly. Easy transitions between ordering on tablets, mobile phones or PCs will have to be facilitated. Besides, convenient multichannel returns and delivery options need to be developed along with the provisions of touch and feel the product before buying. They should also ensure sufficient after sales service and support. Online product reviews and ratings, videos, more advanced sizing and fitting tools should be provided.

Technological advancements : eCommerce companies constantly have to upgrade their offerings with changing technology. For instance, shopping through mobiles have truly arrived, they need to devise easy to use mobile apps for their websites. They need to ensure that their websites have the required speed to do fast business, especially during sale, deals and discounts. Solutions enabling seamless integration of back-end and front-end infrastructure, customer experience enhancement initiatives, integrated inventory management and analytics would be crucial for the eCommerce firms.

Delivery experience : With lack of integrated end to end logistics platform, the eCommerce industry is facing issues related to procurement operations and transportation. Online purchases from Tier-2 and Tier-3 cities are expected to significantly increase, thanks to the emergence of low cost smartphones, however, poor lastmile connectivity could act as a deterrent. Keeping control on logistics and on ground fleet management, especially courier companies, is essential for growth.

Payments and transactions : India continues to be a cash-based society due to limited banking and credit card penetration. This, combined with a lack of consumer trust in online merchants, has forced companies to offer CoD services, which imposes significant financial cost for firms in the form of labour, cash handling and higher returns of purchased items. Data protection and the integrity of the system that handles the data and transactions are serious concerns. Companies should take necessary action for management even if this imposes a cost on them.

Tax and regulatory environment : Laws regulating eCommerce in India are still evolving and lack clarity. Favourable regulatory environment would be key towards unleashing the potential of eCommerce and help in efficiency in operations, creation of jobs, growth of the industry, and investments in back-end infrastructure.

Furthermore, the interpretation of intricate tax norms and complex inter-state taxation rules make eCommerce operations difficult to manage and to stay compliant to the laws. With the wide variety of audience the eCommerce companies cater to, compliance becomes a serious concern. Companies will need to have strong anti-corruption programs

for sourcing and vendor management, as well as robust compliance frameworks. It is important for the eCommerce companies to keep a check at every stage and adhere to the relevant laws, so as to avoid fines.

Operational framework : Business models have been evolving rapidly in the eCommerce sector largely due to heightened competition and the inability of players to sustain high costs. Companies in eCommerce will need to adapt and innovate constantly to sustain their businesses. Furthermore, several of these companies entered into the eCommerce industry as startups and have grown to a huge size aided by the continuous growth in the market but lack well defined capabilities and organisational structure. System building, financial and talent management become key.

Digital infrastructure : Digital disruption has driven change in the eCommerce industry with shoppers embracing multiple touch points in their purchase journeys. Companies should spend enough resources on technology development as also advertising and branding, especially because the younger population is demanding. In the journey toward digital business transformation, embedding SMAC technologies in the business becomes crucial.

Customer acquisition : The customer acquisition costs in Indian eCommerce have been climbing rapidly due to intense competition between multiple well-funded players. Only 2% of website visits currently result into transaction. Thus, there is a gap between potential and actual buyers. Coupled with high transaction costs, this area could pose serious problems. In the US, 75% of consumers have stated that they will usually switch between brands, and for the rest of the world, this rate is 60%, according to Ecommerce Foundation.7 This suggests companies should constantly work on their brand positioning.

Addressable markets: To grow their businesses, the Indian eCommerce sector needs to closely watch the growth of their markets in the Tier 2 and 3 cities. They need to improve their logistics and supply chain management in these cities, do an effective demand management to keep an eye on what products are being sought in these cities. With eCommerce largely being a borderless activity companies need to keep in mind that customers always have the option to buy across the border if they cannot fulfil customers' expectations.

XVIII. CONCLUSION

According to the report Future of India - The Winning Leap, emergence of new technologies, especially mobile, in India has sparked a social change that's difficult to quantify. While mobile, internet, and social media penetration and growth can be quantified, describing the changes in social values and lifestyles that have accompanied those trends is far more challenging.

New technologies such as virtual walls and virtual mirrors will further help improve the retail customer experience, thereby encouraging greater consumption. Virtual mirrors let shoppers 'try on' clothes and accessories virtually before making buying decisions. Virtual walls help customers scan barcodes for items on an electronic wall using their mobile phones and place orders with retailers. Tesco in South Korea

was an early adopter of this technology. In India, HomeShop18 has launched India's first virtual-shopping wall. Scan N Shop at New Delhi's international airport uses a similar technological interface.

A key outcome of the technology revolution in India has been connectivity, which has fuelled unprecedented access to information. Millions of people who had little means to join the national discourse can now gain new insights into the world around them. Farmers know crop prices. Consumers understand global standards of product and service quality. Rural Indians recognise the differences between the opportunities available to them and those available to their urban counterparts. And citizens have a mass forum for expressing their political opinions. The upshot of this connectivity revolution has been empowerment of Indians.

The number of mobile subscribers in India jumped from 261 million in 2007-2008 to 910 million in 2013-2014. Along with telephony, internet penetration is soaring in rural and urban India. Moreover, the number of rural internet users is growing by 58% annually. Increases in the number of smartphones and 3G subscriptions are further driving this growth. Indeed, the number of smartphone users is expected to grow at a CAGR 91% from 2012 through 2016, jumping from 29 million to 382 million. Similarly, the number of 3G subscribers could expand at a CAGR of 84%—from 23 million to 266 million—during the same period.

Thanks to rising internet penetration, the gross number of online users in India now exceeds the number of people who have completed primary education. This shift emphasises the increasing relevance of India's digital economy. The number of internet users soared from approximately 20 million in 2004 to nearly 250 million in 2014. By contrast, the number of people who have studied beyond the eighth standard is about 200 million, indicating that even uneducated people are accessing the internet. While increases in the use of traditional options for gaining knowledge, such as education, may be linear, the proliferation of knowledge through the use of new digital technologies appears exponential.

The eCommerce industry in India may currently be behind its counterparts in a number of developed countries and even some emerging markets. However, with India's GDP growth pegged at 6.4% by the International Monetary Fund and the World Bank, it is expected to grow rapidly. Moreover, the Indian eCommerce industry has access to funds from within the country and international investors. Overall, the eCommerce sector is maturing and a number of serious players are entering the market.

What differentiates the Indian eCommerce market from that of a country like China is that while market concentration in China is largely on account of Alibaba-owned Taobao and Tmall (with these players holding a higher percentage of market share than the top players in most of the other major markets), in India the market share is divided amongst several ecommerce companies, each coming up with its own business model. As a result, customers have a wide range of products and services to choose from.

In our view, there is humongous potential for eCommerce companies owing to the growing internet user base and advancements in technology. However, this will not be without its share of challenges, be it operational, regulatory, or digital. How a company prepares itself to meet these challenges will decide whether or not it succeeds.

REFERENCES

[1] Pwc india

[2] www.facebook.com

[3] www.flipkart.com

[4] www.snapdeal.com

E-Commerce Security

Pankaj Mudholkar
Assistant Professor
TIMSCDR

Sanjay Kumar
Student, TIMSCDR

Nitish Prajapati
Student, TIMSCDR

Vipin Patel
Student, TIMSCDR

Abstract - **Electronic commerce can help enterprises reducing costs, obtaining greater market and improving relationships between buyers and sellers. At the same time, new risks and threats have also occurred, such as, mutual trust, intellectual property, network attacks and so on. This paper analyzes the threat classification and control measures, and on this basis, a conceptual risk management framework is provided. Enterprises engaged in e-commerce can use the framework to improve their security.**

I. INTRODUCTION

E-Commerce is getting more and more popular with customers due to the ever-growing number of companies that provide business to consumer e-commerce services (electronic business transactions between businesses and individual consumers who are buyers).

One of the critical success factors of e-commerce is its security. Without the assurance of security, e-commerce may not work normally. And it is a complexity issue, because e-commerce security relates to the confidence between sellers and buyers, credit card and extremely sensitive personal information. Therefore, the security of e-commerce depends on a complex interrelationship among applications platforms, database management systems, software and network infrastructure and so on. Any single weakness can jeopardize the e-commerce security. A. Sengupta etc., viewed the security of e-commerce as an engineering management problem and examined the issues from a life cycle approach [1]. Fang-Yie Leu etc, studied e-commerce security based on cloud computing, wireless network and they paid more attention to technical approaches to prevent a system from being attacked [2]. Meng Xiangsong etc., focused on the security of mobile agent and the secure authentication infrastructure based on PKI was proposed [3]. Giannakis Antoniou etc. and Rhys Smith etc paid attention to the privacy in e-commerce transactions and provided some suggestions [4, 5]. In addition, the security technology application at multiple levels will slow down the system [6]. Therefore, it becomes critically important to analyze e-commerce security. This paper will analyze the security issues of e-commerce and then put forward a conceptual risk management framework.

II. SECURITY ISSUES OF ELECTRONIC COMMERCE

The rapid development of Internet has promoted electronic commerce explosion. However, at the same time, the internet businesses have brought large security issues such as mutual trust, intellectual property, and possible attacks to the network. And with the development of electronic commerce,

these issues have obtained more and more attentions.

A. Mutual Trust in Business

In the traditional commerce, participant can face to face, so there may be little distrust. However, there is difference in electronic commerce. For example, in electronic commerce, the location of the business and the goods are unknown. More important, there is not personal contact between the seller and the buyer. In addition, there is lack of a clear legal framework in electronic commerce. Therefore, how to enhance mutual trust is an important issue.

B. Intellectual Property

Intellectual property (IP) is an important legal term that refers to copyright and related rights. It is expected to play an increasing role in the future. Especially in e-commerce, IP is more important. E-commerce is more special than other business systems, because many products and services sold in internet are based on IP, such as music, software, pictures, photos, designs, etc. In these goods, IP is the main component of value. And this kind of goods is more suitable to be traded through e-commerce [7]. If there is lack of protection for IP, these goods may be stolen or pirated and even the worst thing, whole business can be destroyed.

In addition, IP is involved in the whole e-commerce work. Internet itself is a product of IP. The systems include software, networks, designs, chips, routers and switches, the user interface, etc. All of these are forms of IP and protected by IP rights. In other words, without IP e-commerce may be nonexistent. Similarly, branding and goodwill are essential elements of e-commerce. And they are also protected by trademarks and related laws. So IP management is the foundation of e-commerce.

C. Network Attacks

Networks attacks have become a general phenomenon, especially with the spread of e-commerce. And the types of attacks are more and more. Security for e-commerce broadly is related to security of networks and databases. The most common kinds of attacks include:

1. IP spoofing attacks IP spoofing attacks is the common security issue in the Internet business. That is to say, a hacker steals an authorized Internet Protocol (IP) address. And then the hacker pretends the authority to make a business. This behavior may bring the authority a big problem.

2. Stealing information In the transaction process, the hacker listens to Transmission Control Protocols/Internet

Protocol (TCP/IP) packets. Therefore, the hacker can steal the information of business or some important personal message, such as e-mail messages, credit card numbers and so on.

3. Password attacks Most users use passwords to control the system, such as e-mail, bank account *etc*. So password is the common target of hackers. Hackers generally find a user who has an easy password If a hacker obtained the system administrator password, the whole system is in dangerous.

4. Social engineering attacks: There are many users who have little understanding their computer system. And they are the aim of the attack. The hacker sends an e-mail message to the users and asks for their password is a typical attack.

5. Vulnerable technological attacks: The system typically the operating system allows a hacker to access the system normally is attacked. A typical one is for the user to gain access to a system. And the user may run an intensive program which can slows down the system.

6. Trust-access attacks: The typical attack is that a hacker adds its system to the list of systems. As a result, the system can allow others to enter into the system without a user password.

III. A MODEL FOR THREAT CLASSIFICATION AND CONTROL MEASURES OF E-COMMERCE

This part will provide a model to analyze the threat classification and control measures of e-commerce. Firstly, we consider threats from two points of view: threat agents and threat techniques. Then we analyze the security control measures.

A. Environmental Factors

Environmental factors are common sense. It is more prone to certain environmental influences and natural disasters than others in some areas. For example, fire is not geographically dependent. However, tornadoes and floods can be predicted in specific areas. In addition to the natural disasters, the danger of mechanical and electrical equipment failure should be paid to more attention. So is the interruption of electrical power.

B. Authorized users

There are some potential threats when authorized users and personnel are engaged in supporting operations. Especially they exceed their privileges and authorities. It may affect the ability of the system to perform its mission. Personnel should be considered as potential threats, when they have the access to a system or occupy positions of special trust. Because they have the capability or opportunity to abuse their access authorities, privileges or trusts. And it may bring danger to the system.

C. Unauthorized users

An unauthorized user can be anyone who is not engaged in the system. It can attempt to interrupt the operation of the system overtly or covertly. It may sabotage hardware and associated equipment. And it also could be accomplished through the manipulation of software.

IV THREAT TECHNIQUES

Techniques can be classified into 5 types: physical, personnel hardware, software, and procedure.

A. Physical

It implies to use a physical means to enter into restricted areas.

B. Personnel

Personnel are the people who have authority or privilege to access a system, either as users or operators. Penetration techniques and methods generally deal with them. Threat agent may recruit them to penetrate the system, operation or facility. They themselves can become motivated to make an attack.

C. Hardware

Attacks using the characteristics of the hardware may involve a physical attack against the equipment, a bug implanted within a hardware controller or an attack against the supporting utilities. The purpose of it is to subvert or deny use of the system [9]. In this category, hardware generally includes any kind of equipment of the system, such as power supplies, air conditioning systems and so on.

D. Software

Software attacks have a large scope from discreet alterations to less discreet changes. The discreet alterations are subtly imposed for the aim of compromising the system. And the less discreet changes intend to bring the result of destruction of data or other system features. Techniques can penetrate the software, application programs or utility routines to threat the system.

E. Procedural

If the system is lack of adequate controls or existing controls are failure, authorized or unauthorized users can penetrate the system [10]. For example, if former employees retained the used valid passwords, unauthorized personnel may pick up output. This is a procedural penetration

V. SECURITY CONTROL MEASURES

There are some detailed security control measures in the ISO 7498-2 Standard lists. For example, there are involving authentication, access Control, data confidentiality data integrity and non-repudiation. Computer security experts widely accept this classification. And they are also recommended by the authors good control measures [11]. The threat agent, threat technique and security measures are shown in fig.

Fig: 1

VI. QUANTIFYING THE COST OF SECURITY

To analyzing the e-commerce, it is better to quantify the cost of security. Firstly, enterprises should know the values of assets, especially the assets exposed to the threat [12]. We can classify the logical and physical assets into 5 categories, as following:

A. Information: It involves the documented data or intellectual property. They are used to meet the mission of an organization.

B. Software: Software is the important component of e-commerce system. It can process, store or transmit information.

C. Hardware: Hardware mainly includes information technology physical devices.

D. Personnel: People is the most valuable asset, especially who posses skills, knowledge, and experience. It is difficult to replace.

E. Systems: Systems are a combination of information, software, and hardware assets. In addition it includes any host, client, or server being considered. It can process and store information.

VII. A CONCEPTUAL RISK MANAGEMENT FRAMEWORK FOR E-COMMERCE

To contain the complexity and maintain focus and relevance, this paper will restrict to issues related to the security of database and information system of e-commerce And we put forward a conceptual risk management framework for e-commerce. According to the following five stages, we can firstly identify the vulnerabilities of a company, second evaluate the existing security measures, and then select the most appropriate and cost-effective countermeasuresName Stand. Abbrev., in press. Analyze Vulnerability and Risk: This analysis can be divided three parts. Firstly, identify vulnerabilities Companies must identify the weakness or flaws in the design, implementation or operation of the security controls of a facility or system.

VIII. CONCLUSION

This paper analyzed the security issues confronted by enterprises engaged in e-commerce. The paper highlighted the role of trust and intellectual property management in e-commerce. At the same time, a three dimensional view of threat agent, technique and control measures are provided to further illustrate the e-commerce security. And in the paper, we also pay attention to the cost of security and give an equal to quantify it. Finally, a conceptual framework is provided to deal with these issues. The framework is just a conceptual idea for enterprises to handle security issues, and further research will continue.
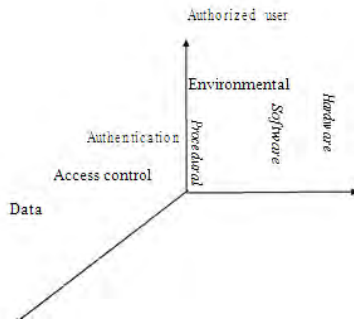
REFERENCES

[1] 1.C. Mazumdar Sengupta and M. S. Barik, E-commerce security-a life cycle approach", Sadhana, vol. 30, no. 2-3, (2005).

[2] 2.F.-Y. Leu, C.-H. Lin and A. Castiglione "Special issue on cloud, wireless and e-commerce security", Journal of Ambient Intelligence and Humanized Computing

[3] 3.M. Xiangsong and H. Fengwu, "Design on PKI-based anonymous mobile agent security in e-commerce

[4] 4.G. Antoniou and L. Battern, "E-commerce: protecting purchaser privacy to enforce trust"

[5] 5. R. Smith and J. Shao, "Privacy and e-commerce "

Facebook and their Data warehouse-Analytical Study

Shirshendu Maitra
Assistant Professor
TIMSCDR

Parth Desai
Student, TIMSCDR

Hitesh Paliwal
Student, TIMSCDR

Jyoti H S Mani
Student, TIMSCDR

Mh. Shaif Khan
Student, TIMSCDR

Yachna Mishra
Student, TIMSCDR

Abstract—**The size of data sets being collected and analyzed in the industry for business intelligence is growing fast, making traditional warehousing solutions prohibitively expensive. The culture of the data warehousing team of Facebook is very much similar to other engineering parts. The key observation here is that the company's core values are the Business Intelligence (BI) Consultant engrained in this team. In particular, "move fast" and "focus on impact" are at the root of every decision BI Consultant make. In this paper, we present an analytical study of the Hive, an open-source data warehousing solution built on top of Hadoop. It involves queries in a SQL-like declarative language – HiveQL (HQL), that are compiled into a map. They reduce jobs that are executed using Hadoop. The technology stack at Facebook is fairly unique as it has truly adopted Hadoop as an enterprise wide tool. Heavily leveraged Facebook Hive for exposing raw data logs and transforming them into usable formats.**

Keywords—*Data warehouse, BI Consultant, Hive, Hadoop, SQL, HiveQL (HQL), Map-reduce job*

I. INTRODUCTION

A) Facebook

Facebook is a popular free social networking website that allows registered users for creating profiles, uploading photos and video, sending messages and keeping in touch with friends, family and all colleagues. The site is available in 37 languages, including public features as follows:

- Marketplace - allows users to post, read and respond to classified ads.

- Groups - allows users who have common interests to find each other and interact.

- Events - allows users to publicize an event, invite guests and track who plans to attend.

- Pages - allows users to create and promote a public page built around a specific topic.

There are several key networking components within each user's profile. The Wall, most popular component is essentially a virtual bulletin board. Messages that are left on a user's Wall are text, video & photos. The Virtual Photo Album is another popular component. The photos can be directly from a smartphone camera or uploaded from the user's desktop.

The Facebook staff removes any inappropriate or copyrighted images from the user's profile. The user's contact (who are called generically called "friends") is allowed to comment on each other's photos and identify (tag) people in the photos due to an interactive album feature.

Another popular profile component,the status updates, a micro-blogging feature through which the users broadcast short Twitter-like announcements to their friends. All communications are posted in the news feed, distributed in real-time to the user's friends.

B) Data Warehouse

A data warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management decision making process**.**

1) *Subject-oriented:* A data warehouse can be used to analyze a particular subject area.

2) *Integrated:* A data warehouse integrates data from disparate data sources.

3) *Time-Variant:* Historical data are kept in a data warehouse. The data may be weeks, months or years older.

4) *Non-volatile:* Data in the data warehouse is not changed/updated. Hence, the historical data in a data warehouse should never be altered.

According to Ralph Kimball, "A data warehouse is a copy of transaction data specifically structured for query and analysis."

II. FACEBOOK AND DATA WAREHOUSE

The culture of the data warehousing team of Facebook is very similar to other engineering fields of Facebook. The company's core values are very BI Consultant engrained in this team. In particular, "move fast" and "focus on impact" are at the root of every decision BI Consultant make. Facebook has a culture built around getting things done and not letting any kind of red tape get in the way. The phrase "done is better than perfect" is thrown around here and it's a philosophy that has served the team BI Consultant.

The project work has typically involved leveraging on scraped log data from Hadoop and creating data pipelines to get the data into a reportable platform and format. This type of work is at the core of what the team does. They take the raw data gathered from the site and

organize it so Facebook can make decisions or highlight success/failures. Most projects are born out of a need to better analyze some part of Facebook and can include creating completely new pipelines or overtaking the present ones.

The ETL developer, data modeler, DBA, project manager, and business analyst all are roles played by the data engineer. In general terms, the team users who create reports and dashboards are not creating ETL pipelines (and vice versa) BI Consultant typically will work directly with the site engineers to understand what they're logging and how. BI consultant also works with data analysts and data scientists to understand how they need to analyze their data so BI Consultant can design something that works for them. Then they build it. Like a lot of other places BI Consultant also have to support the operational stuffs such as investigating failures and fixing data issues.

The technology stack at Facebook is fairly unique as they are one of the few companies that have truly adopted Hadoop as an enterprise wide tool. They heavily leverage Hive for exposing raw data logs and transforming them into utilizable formats. There are a few custom reporting tools that can read directly from Hive but because of the high latency of map-reduce; it isn't a very favorable platform for reporting.

They also have a Facebook developed tool called Facebook Presto which works somewhat like Cloudera Impala. Oracle Exadata is the reporting platform/EDW which MicroStrategy and Tableau point to, for the most part. Facebook recently made a significant investment in Vertica, which the team would be leveraging soon. SQL queries in the hundreds of lines of code per query are the norm, not the exception. Their personal best is about 1500 lines in a single HQL (Hive) statement.

All of the ETL pipelines are built using custom Python frameworks. It's fairly rare that Python is used for the data transformation. The frameworks are instead used for flow control, batch execution and data movement.

They are flexible because you can very easily create scripts that automatically generate ETL tasks with variable inputs. Many of their pipelines consist of programmatically generated tasks with generated SQL/HQL code based on different types of logic. This is quite unique from most other places as usually ETL developers don't need much of programming skills. But data engineers are regularly trying to mend cool new ways to do things in their Python frameworks.

Most of the other tools they leverage are built on top of the Facebook platform and are either Linux or web based.

III. Data Compression Technique

When you have a 300 PB data warehouse running on top of Hadoop, you have to do everything in order to keep adding another rack of disk drives to it.

While Facebook is not afraid to throw a lot of hardware at a scalability problem, its software engineers always look for ways to make the infrastructure software that underpins Facebook more efficient and therefore require less hardware to get more capacity and performance.
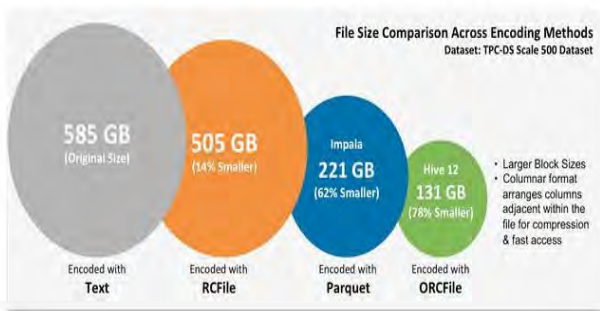
Such is the case with a new compression algorithm that Facebook has developed for its data warehouse, called ORCFile, which runs on top of the Hive relational layer on the data warehouse, and uses the Hadoop Distributed File System to store all kinds of log information from their site. Facebook has updated an earlier version of the program, called RCFile, which allows the data to be compressed into compact.

Facebook has an enormous analytics engine. The company has about 100,000 servers running at its various data centers. According to Ken Rudin, head of analytics at Facebook, the analytics engine and the advertising operations at the company run over tens of thousands of server nodes – a significant portion of the entire server infrastructure of the company. Facebook has its own clustered Hadoop, which is geographically distributed. It has a 300 PB of aggregate data, and Facebook adds about 600 TB of data daily to it.

The company runs MapReduce batch jobs on this data warehouse. It created the Hive relational layer because it wanted something that looked and felt more like a SQL-driven relational database to run queries against the data stored in HDFS. MapReduce is hooked by Hive. The huge data warehouse is also hooked into Presto, a SQL query engine that bypasses MapReduce and can run queries against HDFS directly. That techniques that Facebook's software engineers have come up with to improve compression on its data warehouse are available to all Hadoop users.

Facebook software engineers Pamela Vagata and Kevin Wilfong announced the new ORCFile compression technique in a blog post, offering up a technical discussion of how it works and what performance the new method delivered for Hive tables. The original Record-Columnar File (RCFile) format was created by Facebook, breaking data into row groups and then stacking them up into columnar chunks inside of HDFS blocks. A data compression algorithm such as Zib or Lzo is used for writing each column of data to the disk drives in the Hadoop cluster. When the queries are run against the data, the sniff describes each compressed chunk and only open and decompress the part of the column data store that are necessary for that particular query, which boosts performance.

ORCFile is part of the "Stinger" project at HortonWorks to boost the performance of queries in Hive by a factor of 100X. The HortonWorks was getting about a 4.5X compression ratio on the ORCFile embedded in Hive 12 compared to plain vanilla text encoding. Facebook couldn't see such a jump in its own data with ORCFile, and so thus made some refinements.

File Size Comparison Across Encoding Methods
Dataset: TPC-DS Scale 500 Dataset

585 GB (Original Size) — Encoded with **Text**
505 GB (14% Smaller) — Encoded with **RCFile**
Impala 221 GB (62% Smaller) — Encoded with **Parquet**
Hive 12 131 GB (78% Smaller) — Encoded with **ORCFile**

• Larger Block Sizes
• Columnar format arranges columns adjacent within the file for compression & fast access

With ORCFile, data is written to disk in 256 MB stripes, but there are other layers of encoding and metadata to make data skipping work better and to boost the performance. However, ORCFile caused some Hive tables to bloat up while others compressed well on Facebook's data, and there was not much improvement in the average. Facebook deeply made a lot of change in the code, which anyone running Hive might be interested in looking at the result. Below is the effect on query performance with the three different versions of RCFile:

And below is the effect on writing times across all queries running on the Facebook data warehouse:

Write Times (across all queries)
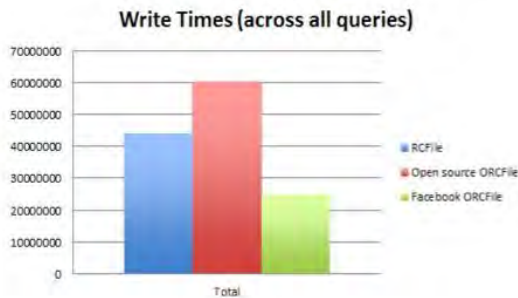
To summarize it all, the modified ORCFile format compresses Facebook's data by a factor of 8X over the plain text format, which is about 63 percent better than the prior RCFile method. The ORCFile writer is three times faster than the open source variant at HortonWorks.

Furthermore, Facebook is looking at a number of other areas to improve the modified ORCFile tool. These ideas include supporting additional codecs such as LZ4HC, using different codecs and compression levels for different columns when encoded differently, storing additional statistics and exposing these to query engines, and pulling in work from open sources, the software engineers write in the blog.

They also have several other ideas to improve storage efficiency in their warehouse further, by eliminating logical redundancies across source and derived tables, sampling cold data sets, and augmenting the Hive type system with additional commonly needed types that are currently stored as strings.

Query 1

Query 2

Query 3
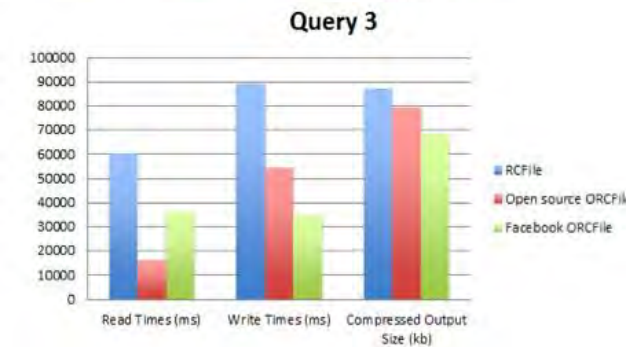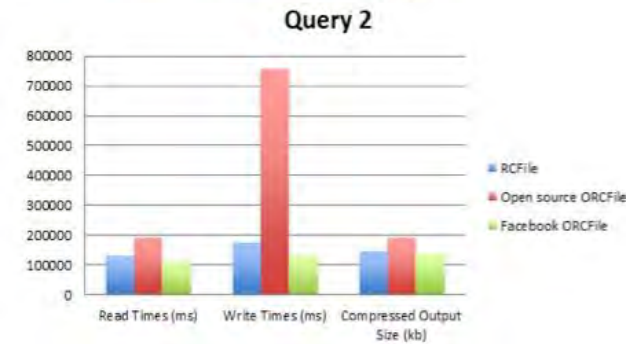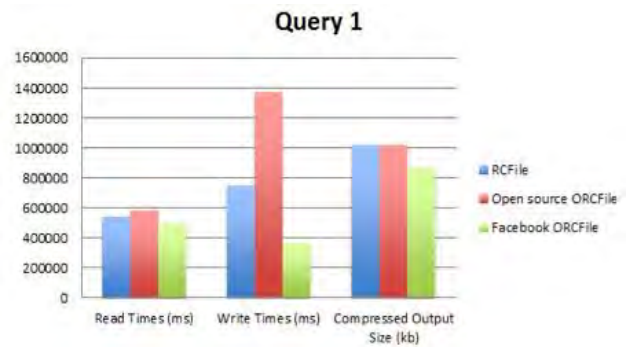
Thus far, Facebook has rolled this modified ORCFile compression onto tens of petabytes of tables its capacity in the data warehouse, and it has again claimed tens of petabytes of storage. Looking ahead, Facebook is looking at the possibility of adding different codecs to the tool and allowing different codecs to be used for compression and decompression on different parts of the data warehouse, depending on what method is the best.

REFERENCES

[1] http://www.enterprisetech.com/2014/04/11/facebook-compresses-300-pb-data-warehouse/

[2] http://whatis.techtarget.com/definition/Facebook

Human Behaviour Analysis on Interfaces

Dr. Vinita Gaikwad
Associate Professor
TIMSCDR

Unmesh Gaikwad	Roidon Rodrigues	Kaustubh Bhosekar	Poonam Dhoot
Student,TIMSCDR	*Student,TIMSCDR*	*Student,TIMSCDR*	*Student,TIMSCDR*

Abstract − **An interface is designed keeping various aspects in mind. One of the key aspect to keep in mind is human behaviour. Human Behaviour plays an important role when designing interfaces. To understand how human behaviour is affected, we would be carrying out our research with the help of the application Runtastic. Runtastic is a mobile fitness company. Using this application users can track their improvement and set goals. As with this app we will be conducting a survey for finding out the best learning for student as well as the transparencey needed between colleges and student for the attendence.**

Keywords:Fitness applications,Designing interfaces, Human behaviour.

I. INTRODUCTION

In the present generation people have a lot of health problems.A simple reason for this is, they have no time or they cannot monitor their health.Thus taking care of their well being becomes an issue.

The problem that we are going to putforth is about a persons health and how can an applicaton like runtastic help in improving it by affecting the behaviour of an individual.

Technology is becoming an important part of our lives.Humans use technology at various stages in their daily interactions.Thus technology should be built in such a manner that it syncs with human behaviour.Runtastic is an application which helps in achieving this sync.

Runtastic is a mobile application which helps an individual to build his/hers overall fitness.Runtastic provides services like online training logs,detailed data analysis, comparisons to other users to improve overall fitness.[1]

A field in psychology known as Applied behaviour analysis(ABA) speaks about various disciplines of psychology like human cognitive abilities, ergonomics,mental health etc.It is defined as a process of systematically applying interventions based upon the principles of learning theory to improve socially significant behaviours to a meaningful degree and to demonstrate that the interventions employed are responsible for the improvement of behaviour.[2] The intervention that can be used here is technology, and if applied systematically can be of great importance to us.Thus human behaviour can be modified and improved for better results.

II. RESEARCH AND OBSERVATON

Runtastic is a fitness application similar to the other applications available in the market.But what sets it apart is its easy an simple to use interface as well as the accuracy it provides while monitoring your fitness.The simplicity and accuracy that it provides helps a person to use it more often than the rest of the applications available.Researching about this application on various forums and youtube channels people have voted it the best in terms of design and simplicity.[3]

So linking this observation to human behaviour we understand that people like things to be simple.In this case, the design of the interface being simple attracted more users to it.Other researches show that if learning something new is simple and easy to grasp it remains for a longer period of time in the memory.Learning something new is closely associated with memory.The term memory is commonly used in reference to the ability to respond to or recount information that was experienced at an earlier time.

Speaking about the health problems that a person faces due to several reasons as mentioned earlier, one scenario being that the individual is not able to monitor his health efficiently.The solution to this problemis, he can use Runtastic since its simple to use and with minimum effort can monitor his health activities.So what we understand by this is humans like things to be as simple as possible.

There are several features in the Runtastic application which will help us understand human behaviour better.One such feature is the workout goal.Runtastic monitors several parameters like the distance travelled ,time and calories burnt.The workout goal feature helps in setting an individuals workout goal before initiating his workout regime.So after completing the workout goal the application prompts a card showing the target he achieved.This helps as a motivation booster for the individual thus feeling a sense on accomplishment.

Human behaviour gets a positive approach when given a good feedback or when motivated.This human behaviour is very well utlized in the application and hence makes it an application to be liked by all.Giving a feedback to an individual helps him/her realise where there is a deficiency of efforts and where he/she has to work upon to improve oneself.

When speaking about motivation researchers have categorized motivation into two categories known as push motivation and pull motivation.Push motivations are those where people push themselves towards their goals or to achieve something, such as the desire for escape, rest and relaxation, prestige, health and fitness, adventure, and social interaction.Pull motivation is the opposite of push. It is a type of motivation that is much stronger. [4] Some of the factors are those that emerge as a

result of the attractiveness of a destination as it is perceived by those with the propensity to travel.

Runtastic has a feature that motivates an individual during his workout regimes.Runtastics provides a functionality where the user can share his/hers workout history and can get motivation from their friends.This is one of the excellent feature that sets it apart from the other fitness application.[5]

Runtastic has given more importance to human behaviour while building its application.This is the reason why it is considered the best in the market.Thus human behaviour should be given vital importance when designing interfaces as well as building its functionality.[7]Runtastic has other features aswell, like its GPS tracking system etc. but our aim was to analyse how human behaviour can be modified with simple things like technology for better and efficient results.

Now we would be understanding a student's behaviour with respect to his/her attendance.The punctuality of a student rewards him/her with a certain percentage of marks.Taking this aspect into consideration, what if there existed an interface which would prompt the student about his/her attendance.Would this interface that provides information regarding the students attendance be beneficial?As mentioned earlier technology should be in sync with a human's behaviour for better results.To see if this really did affect a student's thinking, we carried out a survey with a few questions which would help us determine the end result.
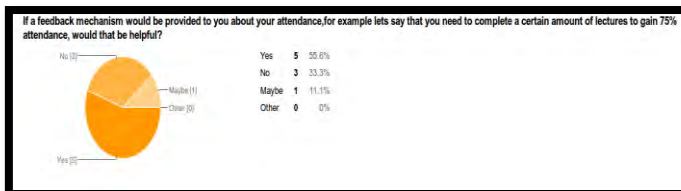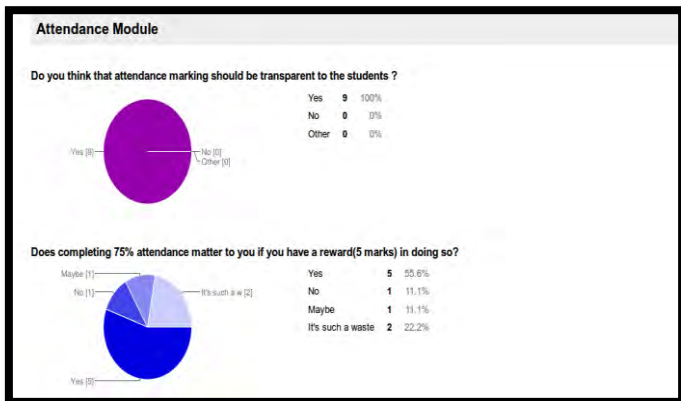
The survey was carried out using google survey forms as you can see above.The questionaires were precise and specific so as to get more accurate data.What we observed by this survey is that majority of the students would benefit if provided with such an interface.Thus our assumption about this concept proved successful.With more data and an actual experiment we can get better and relevant information.

In the second scenario we tried to analyse a student's learning process when provided with tutorial videos.These tutorials would be created by the respective tutor teaching a particular subject.A common interface would be designed for the students where they can access these tutorials.To see if this really proved beneficial for the students and in this case even for the teachers we carried out a questionaire.
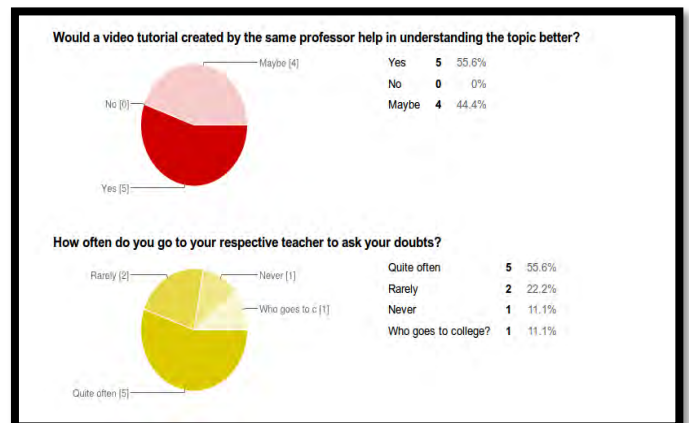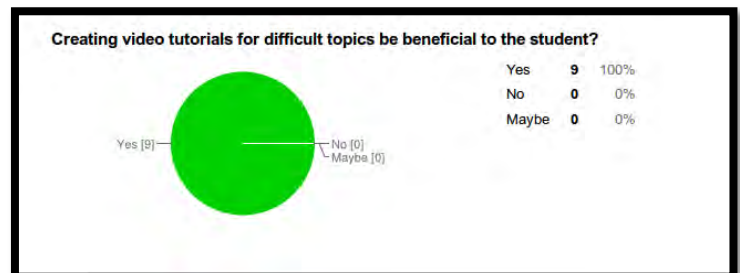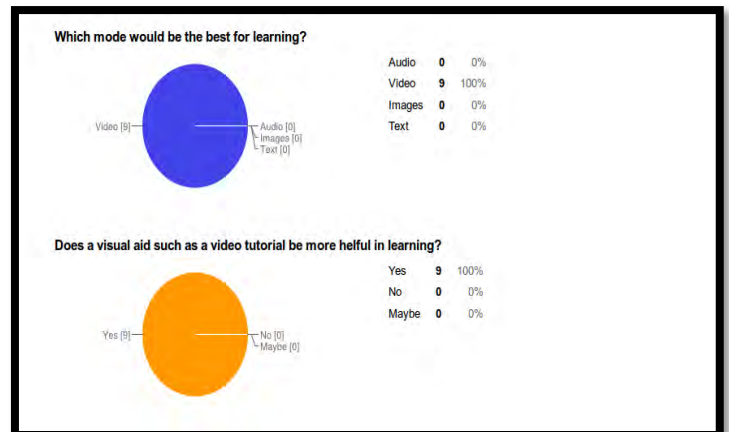
A study was carried out on the learning process of a human being.It was observed that a visual aid is more beneficial when it comes to learning a new skill or technique, thus enhancing the learning process.Tweaking this study further, we decided to combine this study with the above scenario.The above data was noted from the questionaires. Thus our assumption about this concept proved successful.With more data and an actual experiment we can get better and relevant information.

III. Conclusion

Despite all the UI toolkits available today it is still not easy to design good application interfaces.Various techniques and

concepts are being developed to find a solution to this.As seen above applicaton Runtastic, by taking into consideraton human behaviour a proper interface can be designed .

So in order develop better interfaces, considering human behaviour on a particular interface is very important.As of the survey it was detected that students found it easier to learn from the video much better then from the text or audio and also for the students the transparency was said to be of a high priority when it comes for the attendance .

Therefore a proper sync between human behaviour and technolgy has to be made for building better and effiicient application or devices.

This research was conducted to find out how human behaviour plays an important role when designing interfaces. We made use of the internet and a couple of PDFs to conduct our research.

REFERENCES

1. http://en.wikipedia.org/wiki/Runtastic
2. http://en.wikipedia.org/wiki/Behavior
3. http://www.pcmag.com/slideshow/story/292474/the-25-best-fitness-apps
4. http://en.wikipedia.org/wiki/Motivation
5. https://www.runtastic.com/
6. https://www.youtube.com/user/runtastic
7. The Principles of Learning and Behaviour - Domjan

Impact of E-Business on Corporate world of IT Sector

Pankaj Mudholkar
Assistant Professor, TIMSCDR

Yuvraj Pandey
Student, TIMSCDR

Ghanshyam Sharma
Student, TIMSCDR

Ritesh B Trivedi
Student, TIMSCDR

Abstract:- E-business is fast becoming an important initiative for companies to consider, one that impacts every aspect of how a business is run. This report investigates the impact of e-business implementation on various aspects of the organization including; strategy, human resources, customer relationship management, the IT department, technology, the business environment, trust, service management and performance metrics .Implementing e-business applications will require process redesign, organizational restructuring and alignment, new job descriptions and reviewed and revised policies. Organizations' will also have to examine tax, legal and security issues.

The Internet economy necessitates a fundamental transformation of traditional organizations. The true benefit of e-business is achieved through the digitization of the entire value chain. For a successful e-business implementation it is important that decision-makers understand the nature of these changes, their potential impact, plan for them and manage the entire process in such a way as to ensure buy-in of all the relevant stakeholders. The decision to implement an e-business initiative should not be undertaken lightly and the benefits that can be gained from such a venture must be investigated thoroughly before deciding to go ahead.

I. INTRODUCTION

The handling of internal and external business procedures in electronic networks has provided opportunities for process and product innovation in all industries. These opportunities have been taken up by firms in different industries with varying intensity The present paper will examine usage patterns and impacts of e-business in selected *service industries*. Particular attention will be paid to the heterogeneity within the service sector and to country differences.

Service industries have long been regarded as industries with little scope for the use of technical systems, poor productivity improvements over time and a mainly local orientation(see, for example, Miles 1996 and Miles and Boden 2000). This has changed dramatically with the introduction of information and communication technologies (ICT) and the new options available in globally oriented electronic networks (Barras 1986, Miles and Boden2000). Many services show high information intensity, and hence, tools that facilitate the proceeding and diffusion of information are likely to have an impact on the way services are generated and delivered (Miles 1993). Especially the options arising from internet-based e-business models may be used by service firms to streamline procedures or to develop new innovative offers. Services that have a high information content and those which assist other firms in realizing networked systems benefit from the diffusion of information technology and are drivers of technical change in their own business as well as in that of their customers (Preissl 1998). Indeed, computer services show

the highest share of innovators of all business services in Germany (Czarnitzki/Spielkamp 2003).However, it can be assumed that due to fundamental differences in the nature of services provided, not all services can benefit to the same extent from electronic systems. Furthermore, the use of e-business may differ between service sectors in different countries because of differences in business climate or market dynamics (Preissl1998). The impact of e-business on service activities and eventually on performance and market constellations will depend on the type of services considered and the drivers of market development. It can be assumed that services that have used information technology intensively over the last decade will also be intensive users and/or forerunners in the adoption of e-business.

To assess the economic impact of e-business in an organization, I will concentrate on the important sectors in the European union which are the Retail industry, Banking industry and the Travel and tourism industry.

According to GMT [9] and Dalton [6] using ICT technology businesses can achieve significant improvements and cost reductions in several key functions in various departments, including
• Human resource services such as benefits, retirement planning and job postings.
• Internal communication.
• An executive information system.
• Purchasing.
• Sales force automation and management.
• Product development teams.
• Knowledge management.
• Improving relationships with partners.
• Improving time to market.
• Reducing operational costs.
• Increasing employee communication and satisfaction.
• Maintaining a competitive edge.
• Improving customer satisfaction.

It will be particularly interesting to have a closer look at the situation in Germany, since this country is generally considered a 'service laggard' (see summaries of the debate in Cornetz/Schäfer 1999 and in Stille, Preissl, Schupp 2003). On the basis of the empirical Material on e-business available from the survey, it will be possible to check, if the general. Lack of service sector dynamics also holds for the introduction of innovative technology-based business models in service firms. If service sectors in Germany, compared with their main competitors in Europe are reluctant to introduce e-business adoptions, the service gap gains another dimension. The gap is usually identified in services in total employments .And Value added 'Soundness' of an economy (Stille, Preissl, Schupp 2003) which rather rests on the systemic context that makes a particular mix of industries successful. However, undoubtedly, the performance of services that provide essential inputs in many processes of production and efficient services to private households plays an important role in generating a modern services-manufacturing mix. Here the use of advanced network technology and its

61

implementation in successful business models is a key driver. This leads to two central research questions which will be discussed in this paper: (1) how is the heterogeneity of services reflected in their use of ICT and e-business adoptions across Europe? and (2) do German service industries differ from other European countries with respect to the use of e-business opportunities?

Electronic Business: It is commonly referred as "E business", it may be defined as the utilization of information and communication technologies (ICT) in support of all the activities of business.

Electronic Commerce: Electronic commerce or E commerce (E com) is a subset of E business. E commerce is the purchasing, selling, and exchanging of goods and services over computer networks (such as the Internet) through which transactions or terms of sale are performed electronically.

Types of Ecommerce: Ecommerce can be broken into four main categories:
- B2B (Business-to-Business): Companies doing business with each other such as manufacturers selling to distributors and wholesalers selling to retailers. Pricing is based on quantity of order and is often negotiable. Like alibaba.com is B2B website.
- B2C (Business-to-Consumer): Businesses selling to the general public typically through catalogs utilising shopping cart software. Google is the good example of B2C website.
- C2B (Consumer-to-Business): A consumer posts his project with a set budget online and within hours companies review the consumer's requirements and bid on the project. The consumer reviews the bids and selects the company that will complete the project. Ebay.com is C2B website.
- C2C (Consumer-to-Consumer): where individuals can buy and sell thanks to online payment systems like Paypal where people can send and receive money online with ease.

II. STRATEGY

The implementation of an e-business has a significant impact on a company's existing strategy as well as strategy formulation. The basis for an effective e business strategy is an understanding of the company's business goals – not its e-business goals. E-business is away to facilitate improvements in a company's processes, in addition to allowing a company to expand market penetration, geographical markets, etc. Therefore, the processes that will be integrated into the e-business infrastructure must be both known and ready for integration. To implement e-business without understanding the company's business goals or processes is likely to lead to e-business failure. Much like long-range business planning, e-business plans must be re-evaluated and adjusted on a regular basis. An e business strategy is not a one-time event. It is an iterative process. In today's e-economy traditional planning horizons tend to be too long for the very fluid state of business. Continuous planning with feedback has evolved as the strategy of choice for the fluid and volatile e-environment

In a survey conducted jointly by Information Week and Business Week, IT and business executives from all industries agree that generating new sources of revenue isn't the main goal of deploying e-business initiatives. In fact that ranked well behind creating or maintaining a competitive edge, improving customer satisfaction and keeping pace with the competition The ability to react and change direction is critical. Speed is everything. Grounding the organization with sound, winning strategies is key]. E-business enabled organizations must be ready and able to adjust their business and IT strategies rapidly, depending on unpredictable competitors and market pressures. Today's e-business climate requires the continuous optimization of an organization's business and IT strategies. Because IT now has such a significant impact on every business process – from order taking to inventory to billing – both business and IT strategies are now developed in parallel. The lack of a well-though-out e-business strategy has led to some disastrous missteps [18]. In the past, businesses had the luxury of developing business strategies in the boardroom and information technology (IT) strategies in the IT department. They then brought these strategies together to run the overall business. A company implementing an e-business strategy cannot afford this luxury. Technology is no longer an afterthought in formulating business strategy, but the actual cause and driver. According to Agarwal strategy and execution are key to developing and sustaining a successful e-business initiative. Only those organizations that successfully integrate key business strategy and processes dramatically increase their efficiencies. To be successful, organizations must also form the right strategic relationships and develop efficient business processes with robust back-end solutions that are able to meet users' demands for real-time service today and into the future. Considered once a company has decided to embark on an E-business initiative.

1. Industry and company readiness: Is our company or industry prepared for e-business? Should we rapidly develop e-business capabilities? How will our internal processes and skills need to be altered to make a successful transition to e-business?
2. Channel conflict: How will our channel partners react to our entrance to the e-business realm? How large a risk is channel conflict? What steps can we take to mitigate that risk?
3. Prioritizing e-business initiatives: Which elements of e-business need our immediate attention? Which projects should be funded and which need to be placed on hold?
4. Market models: Which models are successful for an organization such as ours? What considerations need to be made before adopting a market model?
5. Organizational design: What changes need to be made to our existing internal structure to facilitate our e-business initiative? Which organizational structures work and which do not? How can we ensure employee buy-in?

A. Retail sector

Retailing is one of the most important industry sector in the world, with no exception of the EU. In 2004 the EU retail sector, the industry comprised of approximately 17 million firms that employed people in the EU27 (e-business watch 2007) In the retail sector, e-business has had great impact on the value chain. It has made a great difference in terms of shelf life of products and stock rotation time. It also highlighted the relevance of inter-business exchanges. In fact, retailers, beside their relationship with suppliers and consumers, have significant impact on intercompany exchanges through

CEDI (centre of distribution). In essence, retail companies will want to rationalize and reduce costs of supply and management of the supply chain, starting with purchasing cost. They will also want to differentiate their own offers from the competitors, exploiting customer relation management techniques as to communication, sales and customers' loyalty (e-business Watch 2000).The UK figures released by the Office of National Statistics in February 2009 showed that internet retail sales, as a percentage of total retail sales, rose by 13.2% (average weekly value of £167 million). These figures closely match figures released by the IMRG/Capgemini Sales Index which showed that the e-business retail market rose by 17% in July compared to July 2008, this increase has been driven by the clothing, accessories and electrical goods online retail sales.

B. Banking Industry
The introduction of E-business (ICT) in the European banking system has had a significant impact on banks operating system and their operations within physical branches. The most important form of e-business used by banks is the online banking which has help cut down cost. E-business has enabled banks to redefine their boundaries and also gained competitive advantage through it. Internet banking is nowadays supported by advanced ICT solutions which enable most everyday banking services to be conducted online (www.ebusiness-watch.org). Studies also shows productivity growth rose in the EU from year 2000 onwards, while average working hours per employee has decreased subsequently, this study tells use e-business investment is largely substituting labour particularly in retail banking.

C. Travel and tourism Industry
In the last few years travel pattern has changed in the EU. The travel sector in the EU is so large that even a small share of it produces a major online market. Taking the internet for example, it has had a very high impact on the air travel industry. It can provide a direct connection between airlines and customers without the need to use travel agents or Computer Reservation Systems (CRS), thereby leading to cost savings in the distribution of air tickets (Law and Leug, 2000) the internet has become very extensive in the travel industries and its impact has been particularly significant on the distribution channel for air travels and the use of the Internet for searching and purchasing airline tickets has become common in travel markets. The introduction of e-business to the airline industry has enabled most travellers to bypass travel agents all together, with most agreeing that the most current air market will shift from traditional travel agents to internet based agents like www.opodo.com. A recent study shows that firms that have reinforced their e-business strategy in the airline industry has had an increase in sales volume for airlines' website. The internet has contributed strongly to the growth of most budget airlines taking easy jet as an example. Easy jet has never used the travel agent network; they rely on the internet because all their flights are booked online by customers

III. MAJOR SUCCESS FACTORS FOR E-BUSINESS

The following major success factors identified by plan must be incorporated into a company's strategy to ensure success:

- Internet technology fully integrated into the company's overall strategy.
- Competitive advantage maintained in both operational efficiency and distinctive strategic positioning.
- Basis of competition not shifted from traditional competitive advantage such as cost, profit, quality, service and features.
- Company's strategic positioning well maintained.
- Support from top management.
- Buyer behavior and customer personalization.
- Quick time to market.
- Right system infrastructure.
- Good cost control.
- Good e-business education and training to employees, management and customers.
- Customers' and partners' expectations well managed.
- Good products and services offered by e-business.
- Current e-business systems expanded to cover entire supply chain.
- New competitors and market shares tracked.
- Website of high quality that meets or succeeds user expectations.
- Company's virtual marketplace established.

IV. INFORMATION TECHNOLOGY AND E-BUSINESS IN SERVICE INDUSTRIES

E-business related changes in service provision and service markets E-business refers to the handling of business procedures in networked digital systems. The codification and documentation of activities in digital form and the use of electronic networks to manage flows of information mark the essential changes that lead to a new way of providing services. Information technology is a tool that enables companies to realize business models and production schemes that have been summarized as e-business. One of the most influential implications of e-business is that it changes the way services are generated and delivered to the customer. Services can, thus, be provided more efficiently, and entirely new types of services are developed (see also Licht and Moch 1997). Many features that have traditionally been used to distinguish services from manufacturing have, thus, become obsolete. Services traditionally made little use of technology. The introduction of ICT, however, has substantially increased investment in technology and made services more capital intensive (see National Research Council 1993). The implementation of e-business solutions requires investment in quite sophisticated information technology, in skills and in the maintenance of technical systems. In order to be able to make these investments, service firms do not only need a sufficient capital basis, but they must also be prepared 'to think in network terms', i.e., to accept the rationale of technology and network-based service provision. One consequence of this enlargement of competencies is that a new set of qualifications is required in service firms, in particular, technological skills, which enable personnel to operate electronic networks.Electronically supported processes of service provision change basic characteristics of service *delivery*. With tighter control of information flows and standardization of communication patterns, craft-like

processes are turned into mass-production and can easily be controlled and improved by introducing the strict logic of digital systems into organizational procedures. Automated processes of production require standardized inputs and produce standardized outputs. As a result, economies of scale are no longer alien to services.

E-business does not only change service provision in the supplier companies, it also affects the typical features of *service markets*. Enhanced possibilities of control and integration, greater flexibility with respect to location, as well as the need to handle huge amounts of information flexibly, have facilitated and supported the deregulation and privatisation of service industries. Economies of scale and improved management of large units lead to concentration processes and to an expansion of markets beyond regional boundaries (Preissl 2001). Small scale, family-run businesses are no longer the characteristic type of service firm in an e-business scenario. A common argument that explains the poor innovation dynamics of service firms is that many service firms are small enterprises with little investment potential and little scope for mass production technology (see Hepworth/Ryan 2000). Gobalised service markets are no longer characterised by small scale service operations, but by impressive merger activity and new co-ordination modes between small units, such as franchising. This is also an economically rational result of the newly emerged economies of scale in providing services. In an economy based on ICT, services take the form of information and can be stored, transported and consumed without the personal presence of the supplier. Many services mainly consist of the provision of information in different forms (see also Baumol 1986 on the implications of this phenomenon). This implies that information can be consumed or used many times without visible signs of consumption. Thus, one process of production potentially results in a large number of identical products that can be individually shaped according to a customer's needs with relatively little additional effort. This raises problems of property rights and exclusion, but also offers potential for productivity gains. Thus, in e-business models highly customized services can be delivered at relatively low cost. Individualization of service is, thus, one feature of electronically supported services. On the other hand, the tools made available by ICT enhance the standardization and automation of many service activities and make them more similar to mass production goods. The combination of these two tendencies leads to a wider range of options for the shaping of service quality. The possibility of storing services in the form of information removes the constraint of time in the provision of services. Services can be delivered via electronic networks without the physical movement of either supplier or client. As a consequence, a global reorganisation of service companies has been expected (see Bressand 1986; Daniels 1993).

E-business is the most conspicuous realization of a new customer interface in services. Personal encounters are substituted by virtual presence, and automated procedures; search engines and self-service information provision directly show the new quality of service provider-client interaction. Theoretically, time and location constraints have become insignificant. In most e-business models for services, the consumers' participation in the design and the delivery of the service is quite intense (see Klein/Totz 2003). The use of computer terminals for information services, where the customer provides the service in interaction with a machine, is a prominent example for this. One has to bear in mind, however, that the service sector is heterogeneous. The possibilities to use electronic networks to change customer-supplier-interfaces and to streamline internal procedures vary from industry to industry. Therefore, also the impact of e-business on competitiveness and market structures will vary considerably between sectors (see E-biz Market Watch Group 2003). E-business offers a set of options to companies. Whether its potential will be exploited, depends on the companies' policies and their organizational, technological and financial abilities. The quantitative dimension of the changes that e-business can induce in the service sector varies from industry to industry, and from country to country. It depends on

- The configuration of the services sector in each country
- Driving forces operating in the specific industry
- Access to essential equipment and infrastructures
- The speed of diffusion of the technology
- Organizational and strategic adjustments in companies and markets.

Given the heterogeneity of services, it is not surprising that network-based adoptions and are not used to the same extent in all service industries (see also Baumol 1986). If the internet and its use in supporting business is the basis for eminently important changes in the economy, those sectors that have a higher *absorption capacity* for information technology will show more dynamic changes than others (see. for example Ducatel 2000). In some industries the nature of the service,the configuration of customer groups and suppliers will make buying and selling online a more natural solution than in others. It can be shown that the drivers of e-business adoption and their varying relevance in a heterogeneous service sector leads to industry-specific and to country-specific implementation patterns. A typology based on information and its role in services Earlier case study research has lead to the conception of a scheme of four service categories with respect to the use of ICT (Preissl 1995). The criteria adopted here were *information content* in input and output of a service and the *use of ICT* either in core or in peripheral function of a service activity. The following scheme has emerged from this research. The question whether a service lends itself to the adoption of e-business follows similar criteria; however, the external – or e-commerce part of e-business requires considering additional criteria. Here, the possibility of digitization and codification of a service might be a central factor. This can be illustrated by the example of a piano teacher. She precedes a lot of information about playing techniques and sound; however, most of this information cannot easily be codified and digitized. Providing piano lessons over electronic networks, thus, seems illusory. Firemen do not have any significant share of information in their inputs and outputs (apart from knowledge accumulated in education and training), but they might use ICT intensively to co-ordinate their assignments. This intensive use of ICT, however, does not imply that their service can be provided electronically these examples lead to a criterion that includes information-content indicators and technological elements, as well as elements of the frequency and type of use of ICT. Only those services that consist of information can be completely digitized and provided via electronic networks, such as consulting or information services.

However, for other services Important parts of the transaction can be organized online. Going back to the fire department: remote control systems can initiate an offline service, and firemen can inform about their services (which usually go beyond extinguishing fires), collect funds and provide information about preventing fires and other accidents online. Hence, the relevant question is not whether services can be delivered online at all, but which part of the activity can be digitized and sold electronically. In this paper survey data will be used to construct combined indicators for e-business readiness, e-business usage and e-business impact. Apart from differences between services and counties, the indicator approach allows to observe interesting relations between readiness, usage and impact indicators.

V. THE SURVEY

As mentioned above, the data used here are taken from a survey which has been conducted in the context of a research project financed by the European Commission (DG Enterprise). Project partners are empirical, Bonn (project leader), Databank Consulting, Milan, Berlecon Research, Berlin and the German Institute for Economic Research (DIW), Berlin. The European ICT & e-Business Market Watch ("*e-Business W@tch*") monitors and assesses the maturity of e-business in 15 industry sectors across EU Member States, including eight service sectors[5]. All sectors have been covered in 4 countries, France, Germany, Italy and the UK. For each sector the sample size was 100 (or slightly above) in each country. A synopsis of the sectoral studies and a public use data file can be accessed on the project's home page (http://www.ebusiness-watch.org/marketwatch/).

Telephone interviews have been conducted with 'decision makers' in the target firms. A wide range of readiness, usage and impact indicators have been covered. However, the range of possible subjects to be included and the level of detail to be reached were clearly limited by time and budget constraints.

VI. SECTOR CHARACTERISTICS

Four service industries have been chosen for the present analysis: Business services (NACE 74), ICT services (NACE 64 and 72), retailing (mainly NACE 52.11, 52.12, 52.4) and real estate (NACE 70). These services are all market-oriented, and have shown a dynamic path of development in almost all countries over the last ten years. A short description of sectors will give some background information for the survey results presented later.

A. Business services

The industries included in this sector cover a wide range of activities with different characteristics. Therefore, some authors have preferred to single out a more homogeneous Category, knowledge-intensive business services, which comprises consultancy and Accounting services as well as advertising and engineering services (see, for example, Miles et al. 1994). However, in the data set used here, this is not possible, and – apart from the knowledge-intensive services mentioned above - the data set includes industries, such as industrial cleaning, investigation and security activities as well as a relatively large category of 'miscellaneous business services i.e.'. Nevertheless, it is safe to say that the sector as a whole is quite information-

intensive, since about two-thirds of their values added at factor cost are created in sub-sectors with high information shares in input and output. The business services sector has been one of the most dynamic sectors in the EU with respect to growth of output and employment. Despite this generally optimistic picture, some drawbacks have affected the industries in the recent past.

The slowdown of economic activity in most industrialized countries has led to a slump in the business services sector, because many firms cannot afford these services any more or have postponed new projects (see Berlecon 2002a). The end of the boom of the 'New Economy' has further calmed down activities in the sector, since fewer new companies that might need advice for the set-up of their business, are founded, and fewer companies are in an expansion phase. Furthermore, there are a series of challenges for business service providers which emphasize the need to activate efficiency potentials and to find ever new solutions for operative and marketing questions.

Most business services require highly qualified staff which implies relatively high costs and in some areas – intense competition for experts and substantial search costs. On the other hand, costs can be reduced by referring to the internet for the investigative parts of service provision. Firms can use e-business to access information as an essential input more efficiently than before, they can use internal data management to automate search and conception procedures, and, thus, to standardize services. On the other hand, clients expect individual attention and personal responses to their consulting or information problems. Thus, business services seem an ideal field to realize a new mix of automated service support and individual service provision.

B. Information and telecommunication services

In the data set this sector comprises two equally large industries. Addressing a mass customer market in an e-business model will require different features form advertising individual software services, both activities require their own back-office Support. In any case, the intense usage of technology, the fact that information technology itself is the rationale for existence for these services, makes them most likely to be very intense E-business adopters. However, as we will see, there is still variation with respect to certain features and with respect to the countries the firms are operating in. Information and Most business services require highly qualified staff which implies relatively high costs and – in some areas – intense competition for experts and substantial search costs.

On the other hand, costs can be reduced by referring to the internet for the investigative parts of service provision. Firms can use e-business to access information as an essential input more efficiently than before, they can use internal data management to automate search and conception procedures, and, thus, to standardize services. On the other hand, clients expect individual attention and personal responses to their consulting or information problems. Thus, business services seem an ideal field to realize a new mix of automated service support and individual service provision.

VII. SUMMING UP SECTOR AND COUNTRY RESULTS

A. Sector Differences

Sectors in which information is an important component of inputs and output, use e-business intensively in the business procedures. Here, business services show slightly higher indicator values than real estate services (with the exception of the sub-indicator for E-commerce). Obviously, this is due to the fact that the processing of information runs through all phases of the business services value chain, whereas it is only essential in certain parts of the value chain in real estate services. However, the difference is not as marked as that between these two sectors and the ICT sector. As mentioned before, the business service sector is quite heterogeneous, and the business models in some sub-sectors are not as easily transferable into electronic networks as in others. As expected, retailing as a sector with relatively low information content in service provision, also shows a poor diffusion of e-business. Formulated answer 'services do not lend themselves for selling online'. Table 6.1 gives the percentage of interview partners that 'completely agree' with this statement.

Surprisingly, the firms in the sector with the lowest e-business activity, retailing, less often hold that their business does not lend itself to selling online than others. Hence, other factors, like a predominance of small enterprises or organizational difficulties must be responsible the hesitation. In these cases it is likely that firms will start to engage in e-business with some delay.

Table 7.1 'Services do not lend themselves to selling online': per cent of responding firms

	Business	ICT	Retailing	Real Estate
EU-4	51.3	45.9	38.5	54.6
France	55.9	54.9	39.3	49.2
Germany	65.6	45.6	35.1	61.7
Italy	42.6	37.2	40.9	53.1
UK	46.7	48.2	36.6	60.1

The other astonishing finding is that firms in the same sector, but in different countries show quite different opinions concerning this barrier. The variation is particularly evident in business services, where less than 43% of the Italian and 65% of the German respondents consider this barrier as relevant. Again, this can be the result of variation in the composition of the sector or due to difference in the business climate in the single countries. It can also be related to difference in cultural attitudes towards doing business electronically.

Differences between countries are less pronounced, where skepticism in the sector is generally low. In countries where the level of e-business usage is generally not very high, firms lack demonstration effects and, thus, find it more difficult to envisage scenarios of useful e-business models. With a more widespread use, these demonstration effects can convince non-users that indeed there are meaningful adoptions in their business.

It is interesting to see that service sectors that show a high 'e-readiness' also normally have high usage indicator values. The two sets of indicators can be combined to form an exploitation factor. Since all indicators are normalized to produce a maximum value of 100, calculating the share of usage indicator values in readiness indicator values generates an indicator for the exploitation of existing equipment for e-business, or an *exploitation factor*. A factor of 1 would mean that all firms that possess the technical preconditions also use them to engage in e-business.

Table 7.2 E-business exploitation factors

	Business services	ICT services	Retailing	Real estate services
EU-4	0.56	0.83	0.47	0.5
France	0.38	0.69	0.16	0.5
Germany	0.57	0.86	0.66	0.52
Italy	0.50	0.73	0.37	0.37
UK	0.72	0.89	0.62	0.70

Source: E-biz Market Watch. Own calculations.

B. Country Differences

Indicator values show a wide range of e-business activity in different European countries. However, in general, sector specificity has a stronger influence than country specificity. However, there are some interesting country differences. E-readiness indicators show Italy ahead, followed by Germany, the France and the UK, if all four sectors are considered together. French firms rank highest in business and ICT services, and Italy in retailing and real estate services. Usage indicators show a completely different picture: Exploitation factors put the UK and Germany clearly above the European average, and Italy and France follow at some distance, and well below the average. The UK, thus, uses a relatively poor endowment with network infrastructures and connections to generate a high level of e-business activities. This holds for all sectors, (except retailing). Germany comes second with respect to usage indicators. In France low e-readiness is resulting in low e-business levels, and Italian firms do not use (sometimes remarkable) infrastructures to do e-business at the same level as firms in other countries. Hence, whereas Italy has higher e-readiness indicators than the other counties, it comes only third with respect to e-commerce and overall e-business usage. The e-readiness indicator in Italian real estate services, for example, is substantially higher than the European average, but it corresponds with a usage indicator that is lower than in any other country in this sector. German service providers rank second for all indicators, i.e., there is no significant discrepancy between infrastructure endowment and its use for e-business.

In an overall assessment of the advancement of use of internet related technologies and the realization of e-business schemes, English service firms rank higher than their German counterparts. Both are well ahead of firms in the other two reference countries. However, German firms show a higher level of e-readiness, i.e., they could catch-up with respect to usage in the near future. Hence, a German service gap can only be observed in comparison with the UK, and here only for usage and impact indicators.

However, differences in e-business diffusion in service sectors can be a consequence of an overall reluctance in a national economy to engage in electronic transactions. In this case, the same or similar configurations would be

found with respect to manufacturing industries as well. Therefore, indicators for four manufacturing sectors, chemicals, electronics, machine construction and transport equipment, have been calculated. These indicators can be seen as approximation to a manufacturing sector.

In Germany and Italy readiness indicators are higher in manufacturing than in services, in the UK and in France the contrary is the case. This means that German service firms lag behind the manufacturing sector. It can therefore concluded that the dominance of the German manufacturing sector is also visible in a more advanced e-readiness. However, rankings among countries are not dramatically affected by this phenomenon. France and the UK change places, whereas Italy and Germany occupy first and second positions in both groups.

Usage indicators show more pronounced differences: here services lie way ahead of manufacturing in all countries (hence, exploitation factors tend to be much lower in manufacturing than in service firms). This holds for the two sub-indicators as well, however, in e-commerce the differences are much more pronounced than for internal e-business variables. In the UK the wider diffusion of e-business usage in services with respect to manufacturing is most striking, while France and Germany show comparatively modest differences. In terms of country ranking, this puts the UK far ahead of the other countries in the service sector, Germany remains in second place. The positions are exactly opposite in manufacturing. Because of the relatively high usage indicator in German manufacturing, in Germany the difference between services and manufacturing remains further below the European average than for the readiness indicator, although in absolute terms, the German service sector usage indicator is considerably higher than the EU-4 average. This confirms the argument that service sector weakness often is not a weakness per se but mainly becomes evident in comparison with manufacturing, if the manufacturing sector is particularly strong.

For the impact indicator the low values in Germany's service sector correspond to low values in manufacturing. However, again, the difference is more pronounced here than in other countries, resulting in a particularly marked superiority of the manufacturing over the service indicator. The particular configuration of the German service sector and its particular position in the economy are also affecting e-business adoption in this sector. However, this should not lead to hasty judgments about the superiority of one configuration over the other. The relationship and interaction between the two sectors have given rise to a complex debate in which the results presented here cannot be more than a small pebble.
For the analysis as a whole, it can be concluded that e-readiness is a necessary condition to do e-business, but there is no automatism that leads from good infrastructures to e-business activity.

However, German firms see their e-business experience by far less positive than firms in other countries. This might lead to a slowing down of future expansion rates. The positive experience of Italian e-business adopters, on the other hand, gives rise to a more optimistic attitude and can induce other firms to engage in electronic business models.

VIII. STRUCTURE OF GLOBAL AND INDIAN IT INDUSTRY

Growth of Global IT Economy: The global IT industry has matured over the years and has emerged to be a chief contributor to the global economic growth. The global IT sector, constituted by the software and services, Information Technology Enabled Services (ITES) and the hardware segments, has been on a gradual growth trajectory with a steady rise in revenues as witnessed in the past few years. 2008 was a strong year as the number of contracts; the total value and the annualized contract values exceeded that of the preceding year. Among all users above average growth was witnessed in the government, healthcare and the manufacturing segments. The global software and services industry touched USD 967 billion, recording an above average growth of 6.3% over the past year. Worldwide ITES grew by 12%, the highest among all technology related segments. Hardware spend is estimated to have grown by 4% from USD 570 billion to nearly USD 594 billion in 2008.

Currently, the global IT industry is experiencing a slump with the recessions in the US and many industrial countries with the level of impact varying by country / market and industry. Forrester in its recent report has predicted that the US IT market will dip to 1.6% in 2009, down from 4.1% growth in 2008. The Asia Pacific region, using a weighted average of local currencies, will do a bit better in 2009, with 3.1% growth. The Western and Central Europe markets will have growth in local currency that is closer to 1%. By 2010, the US market will shift to 7.3% growth, not far behind the 9.5% growth in the other Americas, well ahead of the 5.5% growth in Asia Pacific and 5.3% growth in Western and Central Europe.

IX. GLOBAL SCENARIO - IT PURCHASES
As it stands, the US market accounts for majority of the global purchases of IT goods and services. The US market which represented 37% of the global market for IT goods and services in 2005 had shrunk to 33% share in 2008. Western and Central Europe would see its share of global IT purchases fluctuate between 26% and 28% between 2008 to 2010; Eastern Europe, the Middle East, and Africa and Asia Pacific are expected to hold their share positions. The global IT purchases are expected to plummet as strong dollar would hurt dollar-denominated growth rates for IT purchases going ahead. The British pound was 23% lower in Q4 2008 from the year-ago level, the Indian rupee is down 20%, the Canadian dollar is 19% weaker, and the euro is down 9%. Only the Japanese yen and the Chinese Yuan renminbi have gained in value against the US dollar. While these currency swings are likely to reverse in 2009 as the financial crisis fades, the dollar is still likely to remain above 2008 levels for most of the year. That will dampen global IT market growth measured in dollars and hurt the reported revenues of US vendors like Accenture, Hewlett-Packard (HP), and IBM with large overseas operations. With global tech market in US dollars likely to shrink, global IT vendors' revenues is expected to equal $ 1.66 trillion in 2009, declining by 3% after an 8% rise in 2008. The Asia Pacific region has been a major growth engine for the tech industry. Its total purchases of IT goods and services of $448 billion in 2008 were almost as large as Western and Central Europe's. Countries like Hong Kong, India, Malaysia, Singapore, South Korea, and Taiwan, have seen growth slow as exports to the US and Europe slowed. Asia / Pacific would experience a delayed

impact of the global financial crisis. Gross Domestic Product (GDP) growth is expected to slow in most countries / markets in 2009, which will affect IT spending. Asia pacific is still growing more aggressively than other regions in GDP and in IT. As a result, vendors would be looking to this region for growth and stability.
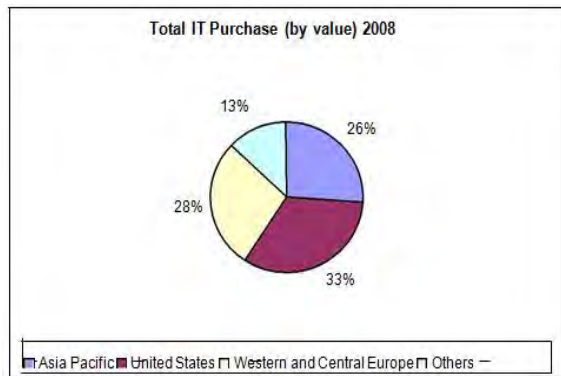

Fig 1 Total IT purchase

X. STRUCTURE OF INDIAN IT INDUSTRY

The IT-ITES industry in India has today become a growth engine for the economy, contributing substantially to increases in the GDP, urban employment and exports, to achieve the vision of a powerful and resilient India. While the Indian economy has been impacted by the global slowdown, the IT-ITES industry has displayed resilience and tenacity in countering the unpredictable conditions and reiterating the viability of India's fundamental value proposition.

Value proposition
The main reasons for the successful establishment of software companies in India and its strong performance can be attributed to the following:

Cost advantage
Given the labour market conditions in India, there exists substantial scope of cost arbitrage for performing services from India. This, along with a large pool of talented and English people labour force, was the genesis of the IT sector's dominance in the world IT services industry . Breadth of service offering and innovation Service offerings have evolved from low-end application development to high-end integrated IT solutions

Quality / maturity of process
Having made its mark as a centre of low-cost and wide range of service offerings, the Indian IT / ITES sector has also proved its mettle in the quality of the service offerings, as demonstrated by the fact that it hosts more than 55% of SEI CMM level five firms and the highest number of ISO certified companies

Ease of scalability
The vast and trained labour pool of technically competent, English speaking people has made it easy for the Indian companies to enter and exit this industry. Moreover, the ease with which a company can scale its operations (up or down) has been a great value driver for the success of the Indian IT / ITES service sectors growth story

XI. PERFORMANCE OF THE INDIAN IT-ITES INDUSTRY

The information technology sector has been playing a key role in fuelling the Indian economic performance which has been stellar with robust GDP growth. India's total IT industry's (including hardware) share in the global market stands at 7%; in the IT segment the share is 4% while in the ITES space the share is 2%. The industry is dominated by large integrated players consisting of both Indian and international service providers. During the year, the share of Indian providers went up to 65-70% due to the emerging trend of monetization of captives. MNCs however, continued to make deeper inroads into the industry and strengthened their Indian delivery centers during 2008. The continuing contribution of this sector to the Indian economy is evident from the fact that revenue generated from this sector has grown from 1.2% in FY 1998 to an estimated 5.8% in the FY 2009. The net value added by this sector to the economy is estimated at 3.5-4.1% for FY 2009. Some of the key highlights of the Indian IT / ITES industry for FY 2009 are as follows:

- The export revenues are estimated to gross USD 47.3 billion in FY 2009, accounting for 66% of the total IT-ITES industry revenues
- IT services exports grew substantially on account of increasing traction of the industry in emerging markets such as remote infrastructure management and traditional segments such as application management
- Domestic market continued to gain momentum, growing at 26% in INR terms on account of the overall positive economic climate, increased adoption of technology and outsourcing
- Engineering services and software product exports increased by 29% (USD)
- Direct employment reached nearly 2 million – with 1.5 million in the exports segment, increase of 26% in 2008. The indirect employment multiplier suggested that the industry created between 6-8 million additional jobs
- US and UK together constituted 79% of the global exports in FY 2008 thereby dominating the export markets
- BFSI remained the largest market followed by Hi-tech / Telecom which together accounted for more than 60% of exports

A. Impact of exchange rate on revenues
In IT sector, the margins are likely to be challenged on account of the slowing growth in the US. Rupee depreciation seems to be the only tailwind that the sector enjoys. This can be evident from the fact that the out of the increase in the IT export revenues for FY 2008 over FY 2007, almost half of the increase could be attributed to the rupee depreciation during the same period.

B. Pricing poised for decline in favour of volumes
Pricing has been difficult in this sector compared to other sectors: On an average, the US financial sector has driven bulk volumes through lower onsite pricing, higher off shoring and aggressive volume discounts. It is safe to infer that BFSI application business margins especially in the top companies are a few percentage points below the higher margin verticals like, say, energy. Hence, a replacement of financial services business with business from other verticals is likely to positively impact the

bottom line. A speedy replacement is however, easier said than done. Volumes are expected to remain weak over the next three quarters for most players forcing further price cuts. The reduction in pricing is expected to be lower in magnitude compared to FY 02-FY 03. This is because the current pricing has not touched the FY 02-FY 03 bubble proportions. Infosys has already reported 1.8% decline in blended pricing (constant currency) in Q3 FY 09 while HCL Tech announced free transitioning for deals amounting to $1billion bagged during the quarter as a strategy to garner volumes. TCS and Wipro too have acknowledged pricing pressures and the impact would be more visible in the coming quarters. Fitch Rating expects the sector to face margin pressures over 2009 and 2010 due to the intensified competition for new contracts, thereby putting pressure on billing rates. Competition even for smaller contracts has increased, as companies try to maintain utilisation levels. Customer cost pressures could also result in renegotiations of maturing contracts at lower terms. There could also be an increased shift from traditional hourly billings towards a new return on capita based price contracts providing tangible savings, while variable time / material contracts could be renegotiated at lower levels. Vendor consolidation will be the order of the day in the current environment, as this would result in cost savings for customers. Fitch believes that the large Indian IT players will gain market share. However, these risks to operating margins are partly offset by the fact that Indian IT services retain some flexibility in terms of their cost model. As the impact of the slowdown becomes more severe, companies will increasingly look at cutting costs in the form of overheads and reduction in variable pay / annual increments. The industry has also been reducing its hiring, as well as changing the hiring profile to ensure that operating costs are in control.

C. New technological innovation: E-business
Year 1830-1900 was a time of rapid economic growth and technological innovation which has been well documented and analyzed. Business historians Chandler (1977) and Porter and Livesay (1971) discuss changing patterns of production and distribution over this period. In 1830, almost all goods were generic. Production processes used relatively simple technology and were low volume, with nearly all products distributed within a small locality. The historical trend identified is the increasingly concentrated nature of the market, and the reduced costs of exchange. Changes in technology and market conditions created an economic situation that favored manufacturer-controlled distribution systems. However, that new technology (E business) is an important contributing factor to the changing organizational structure of firms. The four ways in which technological innovations can have impact: by necessitating new organizational structures to handle complexities, by encouraging administrative or other process changes that reduce costs or improve efficiency, by facilitating the development of new distribution channels, and by changing the power relations amongst organizations.

US Retail E-commerce sales ,2004-2009 in billion $

Year	Value
2009Q1	133.1
2008	133.6
2007	127.7
2006	106.6
2005	87.4
2004	70.9
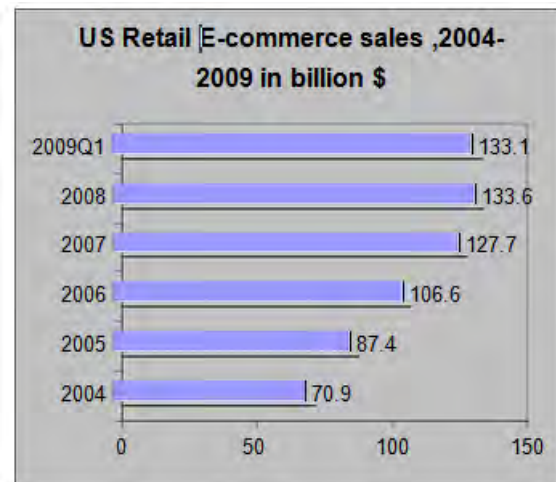
12. CONCLUSION

Incorporating electronic business in an organization invariably has a significant impact on various operations and aspects of the organization, causing changes in areas such as human resources, strategy planning, technological infrastructure and customer service. For a successful e-business implementation it is important that decision makers understand the nature of these changes, their potential impact, plan for them and manage the change process in such as way as to ensure buy-in of all the relevant stakeholders. The decision to implement an e-business initiative should not be undertaken lightly and the benefits that can be gained from such a venture must be investigated thoroughly before deciding to go ahead. E-business plans must be devised as part of the corporate strategy and must take into consideration the impact e-business will have on processes, governance and people. It is important that companies create a single coherent plan, formalize decision-making procedures and communicate e-business initiatives across the organization and integrate the e-business plan with corporate goals.

REFERENCES

1. http://mergermarketgroup.com/publication/inside-corporate-wolf-theiss-corporate-monitor-fy-2014/#.VQ7e7YuUcfF
2. http://www.answers.com/Q/What_is_meaning_of_corporate_world
3. http://www.mergermarket.com/info/category/press-release/2015/
4. http://seekdl.org/nm.php?id=489
5. http://www.academia.edu/1917796/The_Impact_of_Electronic_Business_on_the_Organisation
6. http://www.ukessays.com/essays/ecommerce/the-impact-of-e-business-on-the- economy.php

Impact of E-Business on the Retail Market

Pankaj Mudholkar
Assistant Professor, TIMSCDR

Nivedita Parab
Student, TIMSCDR

Abstract - Online retailers such as Flipkart and Snapdeal, despite their multi-billion dollar valuations, will not be able to challenge traditional retailers that will continue to dominate the domestic market, audit and consultancy firm.

Index Terms- E Business, Retail, Online shopping

I. INTRODUCTION

The e-business sector has seen unprecedented growth in 2014. The growth was driven by rapid technology adoption led by the increasing use of devices such as smart phones and tablets, and access to the internet through broadband, 3G, etc, which led to an increased online consumer base. Furthermore, favored demographics and a growing internet user base helped aid this growth. In terms of highlights, the growth shown by homegrown players such as Flipkart and Snapdeal and the huge investor interest around these companies displayed the immense potential of the market.

With the entry of e-business behemoths such as Amazon and Alibaba, the competition is expected to further intensify. Both these international players come with deep pockets and the patience to drive the Indian e-business market. Also, their strong domain knowledge and best practices from their international experience give them an additional edge. Additionally, these companies have been part of markets where they have seen the e-business market evolve and are aware of the challenges and strategies to address issues thereof. Indian companies realize this, and are therefore aiming to continue their focus on expanding sellers and selection on their platforms, innovating on multiple customer touch points, and providing seamless and rapid delivery services in order to compete with the international entities. Competition is expected to continue, with these e-business companies experimenting with different ways to attract customers and increase online traffic. The Indian government"s ambitious Digital India project and the modernization of India Post will also affect the e-business sector. The Digital India project aims to offer a one-stop shop for government services that will have the mobile phone as the backbone of its delivery mechanism. The program will give a strong boost to the e-business market as bringing the internet and broadband to remote corners of the country will give rise to an increase in trade and efficient warehousing and will also present a potentially huge market for goods to be sold.

For Indian Post, the government is keen to develop its distribution channel and other e-business related services as a major revenue model going ahead, especially when India Post transacted business worth 280 Cr. INR in the cash-on-delivery (CoD) segment for firms such as Flipkart, Snapdeal and Amazon. Both these projects will have significant impact on increasing the reach of e-business players to generally non-serviceable areas, thereby boosting growth.

TABLE 1: E-Business Opportunities, Challenges, Drivers and Barriers

e-Business related opportunities	e-Business related challenges
1.Cost savings	1. Lack of awareness
2. Efficiency and productivity gains	2. Increasing market competition, where retailers are highly concentrated
3. Increasing information about the market and the customers	3. Economic return of e-business
4. Online selling and multi-channel approach	
e-Business drivers	**e-Business barriers**
1. Trends in demography and lifestyle	1.Lack of interest in internet based applications and sales systems among a large number of retailers
2. Improved systems for increasing the efficiency of supply chain relationships	2.ITC skill gap
3 new private tables	3 trust and security issue

II.TOP 10 COUNTRIES IN TERMS OF POPULATION AND CORRESPONDING E-HOUSEHOLDS

India"s internet penetration with total e-households at 46 million against China"s 207 million is one of the reasons behind India"s poor B2C sales growth. Only 16% of India"s total population was online in 2013 and of the online users only 14% or 28 million were online buyers. India, therefore, was still in a nascent or immature stage of evolution of online retail spending. China was in ascending stage at 50%, whereas Japan (69%), Australia (57%) and South Korea (70%) were in mature stage.
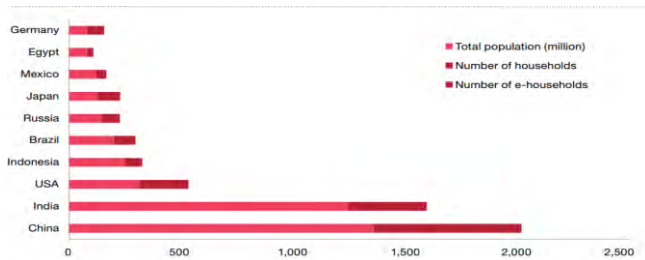
Fig 1 Figure showing statistics of e-households

India's growth potential Since the e-business industry is fast rising, changes can be seen over a year. The sector in India has grown by 34% (CAGR) since 2009 to touch 16.4 billion USD in 20142. The sector is expected to be in the range of 22 billion USD in 2015.

Fig 2. E-commerce ecosystem

Fig 3. India's e-business and retail growth

Currently, eTravel comprises 70% of the total e-business market. eTailing, which comprises of online retail and online marketplaces, has become the fastest-growing segment in the larger market having grown at a CAGR of around 56% over 2009-2014. The size of the eTail market is pegged at 6 billion USD in 2015. Books, apparel and accessories and electronics are the largest selling products through eTailing, constituting around 80% of product distribution. The increasing use of smartphones, tablets and internet broadband and 3G has led to developing a strong consumer base likely to increase further. This, combined with a larger number of homegrown eTail companies with

their innovative business models has led to a robust eTail market in India rearing to expand at high speed.

III. KEY MARKET FACTORS TO BE EVALUATED BEFORE ENTERING A NEW E BUSINESS

To achieve their vision, e-business companies will need to understand the intricate landscape of new markets in addition to their own internal capabilities and limitations. The following factors must be considered:

• Market size: Before moving too aggressively into a new market, it is important to consider how sizable the overall opportunity is.
• e-business readiness: It is essential to fully understanding the payment and logistical infrastructure, consumer behavior, retail opportunity and technological developments.
• Scope of growth: It is also important to look at the internet penetration, demographics of the online buying population and understand which phase of development each market is in.
• Barriers to entry: Players should understand the regulatory environment and connect with solution providers, content distribution networks, and digital agencies.
• Competition: There is also a need to do an in-depth assessment of what competitors are doing, their online strategy and the nature of each offering.

IV. EXPANDING THE PRODUCT BASKET

There is a recent trend of relatively newer products such as grocery, hygiene, and healthcare products being purchased online. Similarly, lingerie and Indian jewellery has also been in great demand among customers outside India. Export comprises 95% of cross-border e-business, with the US, UK, Australia, Canada and Germany being the major markets.

Industry Speak: Some of the key concerns are listed below:

• Generation and sustenance of traffic: Competition from established e-business players is making it difficult for private label brands to generate traffic on their white-label websites.
• High customer acquisition cost: The customer acquisition costs have been rising due to intense competition by the relatively better off companies with more funds.
• Last-mile delivery: Poor last-mile connectivity, especially in remote areas with larger population, is another problem faced by Indian e-Tailers.

• High payment cost: CoD services impose substantial financial cost. In India, unlike in developed markets, CoD continues to be a preferred route of payment.
• Low profitability: Profitability is negatively impacted by high customer acquisition costs, free shipping and high rejection rate of CoD orders.
• Regulatory barriers: Regulatory barriers in the Indian e-business market are higher as compared to more mature markets.
• Skilled manpower: Lack of talent availability and high attrition are causing manpower crunch, which is fast becoming a hurdle.

V. TOP 10 THINGS THE E-BUSINESS COMPANIES NEED TO DO TO ACCELERATE GROWTH

Customer experience: As the customers progress from research to purchase to fulfillment stages, their expectations change fast. e-business companies need to understand these change drivers and adapt their proposition accordingly. Easy transitions between ordering on tablets, mobile phones or PCs will have to be facilitated. Besides, convenient multichannel returns and delivery options need to be developed along with the provisions of touch and feel the product before buying. They should also ensure sufficient after sales service and support. Online product reviews and ratings, videos, more advanced sizing and fitting tools should be provided.

Technological advancements: e-business companies constantly have to upgrade their offerings with changing technology. For instance, shopping through mobiles have truly arrived, they need to devise easy to use mobile apps for their websites. They need to ensure that their websites have the required speed to do fast business, especially during sale, deals and discounts. Solutions enabling seamless integration of back-end and front-end infrastructure, customer experience enhancement initiatives, integrated inventory management and analytics would be crucial for the e-business firms.

Convergence of online and off line channels: As the customers progress from research to purchase to fulfillment stages, their expectations change fast. e-business companies need to understand these change drivers and adapt their proposition accordingly. Easy transitions between ordering on tablets, mobile phones or PCs will have to be facilitated. Besides, convenient multichannel returns and delivery options need to be developed along with the provisions of touch and feel the product before buying. They should also ensure sufficient after sales service and support. Online product reviews and ratings, videos, more advanced sizing and fitting tools should be provided.

Delivery experience: With lack of integrated end to end logistics platform, the e-business industry is facing issues related to procurement operations and transportation. Online purchases from Tier-2 and Tier-3 cities are expected to significantly increase, thanks to the emergence of low cost smartphones, however, poor lastmile connectivity could act as a deterrent. Keeping control on logistics and on ground fleet management, especially courier companies, is essential for growth.

Payments and transactions: India continues to be a cash-based society due to limited banking and credit card penetration. This, combined with a lack of consumer trust in online merchants, has forced companies to offer CoD services, which imposes significant financial cost for firms in the form of labour, cash handling and higher returns of purchased items. Data protection and the integrity of the system that handles the data and transactions are serious concerns. Companies should take necessary action for management even if this imposes a cost on them.

Tax and regulatory environment: Laws regulating e-business in India are still evolving and lack clarity. Favorable regulatory environment would be key towards unleashing the potential of e-business and help in efficiency in operations, creation of jobs, growth of the industry, and investments in back-end infrastructure. Furthermore, the interpretation of intricate tax norms and complex inter-state taxation rules make e-business operations difficult to manage and to stay compliant to the laws. With the wide variety of audience the e-business companies cater to, compliance becomes a serious concern. Companies will need to have strong anti-corruption programs for sourcing and vendor management, as well as robust compliance frameworks. It is important for the e-business companies to keep a check at every stage and adhere to the relevant laws, so as to avoid fines.

Operational framework: Business models have been evolving rapidly in the e-business sector largely due to heightened competition and the inability of players to sustain high costs. Companies in e-business will need to adapt and innovate constantly to sustain their businesses. Furthermore, several of these companies entered into the e-business industry as startups and have grown to a huge size aided by the continuous growth in the market but lack well defined capabilities and organizational structure. System building, financial and talent management become a key focus.

VI. CONCLUSION

According to the PwC report Future of India - The Winning Leap, emergence of new technologies, especially mobile, in India has sparked a social change that"s difficult to quantify. While mobile, internet, and social media penetration and growth can be quantified; describing the changes in social values and lifestyles that have accompanied those trends is far more challenging. New technologies such as virtual walls and virtual mirrors will further help improve the retail customer experience, thereby encouraging greater consumption. Virtual mirrors let shoppers „try on" clothes and accessories virtually before making buying decisions. Virtual walls help customers scan barcodes for items on an electronic wall using their mobile phones and place orders with retailers. Tesco in South Korea was an early adopter of this technology. In India, HomeShop18 has launched India"s first virtual-shopping wall. Scan N Shop at New Delhi"s international airport uses a similar technological interface. A key outcome of the technology revolution in India has been connectivity, which has fuelled unprecedented access to information. Millions of people who had little means to join the national discourse can now gain new insights into the world around them. Farmers know crop prices. Consumers understand global standards of product and service quality. Rural Indians recognize the differences between the opportunities available to them and those available to their urban counterparts. And citizens have a mass forum for expressing their political opinions. The upshot of this connectivity revolution has been empowerment of Indians. The number of mobile subscribers in India jumped from 261 million in 2007-2008 to 910 million in 2013-2014. Along with telephony, internet penetration is soaring in rural and urban India. Moreover, the number of rural internet users is growing by 58% annually. Increases in the number of smartphones and 3G subscriptions are further driving this growth. Indeed, the number of smartphone users is expected to grow at a CAGR 91% from 2012 through 2016, jumping from 29 million to 382 million. Similarly, the number of 3G subscribers could expand at a CAGR of 84%—from 23 million to 266 million—during the same period. Thanks to rising internet penetration, the gross number of online users in India now exceeds the number of people who have completed primary education. This shift emphasizes the increasing relevance of India"s digital economy. The number of internet users soared from approximately 20 million in 2004 to nearly 250 million in 2014. By contrast, the number of people who have studied beyond the eighth standard is about 200 million, indicating that even uneducated people Conclusion are accessing the internet. While increases in the use of traditional options for gaining knowledge, such as education, may be linear, the proliferation of knowledge through the use of new digital technologies appears exponential. The e-business industry in India may currently be behind its counterparts in a number of developed countries and even some emerging markets. However, with India"s GDP growth pegged at 6.4% by the International Monetary Fund and the World Bank, it is expected to grow rapidly. Moreover, the Indian e-business industry has access to funds from within the country and international investors. Overall, the e-business sector is maturing and a number of serious players are entering the market. What differentiates the Indian e-business market from that of a country like China is that while market concentration in China is largely on account of Alibaba-owned Taobao and Tmall (with these players holding a higher percentage of market share than the top players in most of the other major markets), in India the market share is divided amongst several e-business companies, each coming up with its own business model.

As a result, customers have a wide range of products and services to choose from. In our view, there is humongous potential for e-business companies owing to the growing internet user base and advancements in technology. However, this will not be without its share of challenges, be it operational, regulatory, or digital. How a company prepares itself to meet these challenges will decide whether or not it succeeds.

REFERENCES

[1] "E-Commerce and E-Business/Introduction" http://en.wikibooks.org/wiki/E-Commerce_and_E-Business/Introduction
[2] Lee In (2008),"E-Business models, services & communications" Idea group (IGI).
[3] Zappala S. Gray C. (2006), " Impact of e-commerce on consumers & small firms", Ashgate Hampshire U.K.
 [4] "Amazon.com" http://en.wikipedia.org/wiki/Amazon.com .
[5] "EBay.com", http://en.wikipedia.org/wiki/EBay.
[6] "The Impact of E-Business on the Retail Market" - http://voices.yahoo.com/the-impact-e-business-retail-market- 548868.html
[7] "ICT and e-Business Impact in the Retail Industry" A Sectoral e-business Watch study report by Empirica GmbH, Version 4.0 September 2008

Logistics in E-Business

Pankaj Mudholkar
Assistant Professor, TIMSCDR

Kirti Singh
Student, TIMSCDR

Rajni Yadav
Student, TIMSCDR

Abstract - **In this paper we first summarize the categorization and prioritization of the most crucial R&D topics that should be dealt with in order to help enterprises to face the new logistics challenges of e-business. To achieve efficiency and operational quality in supply networks we then suggest a research agenda with three interrelated R&D themes. Fist theme relates to integrated supply network structure with suitable visibility and collaboration. The second theme handles research and development of new logistics service concepts and the third theme covers topics related to using new identification methods in supply networks. According to the findings of the study the R&D initiatives in these fields ought to be interdisciplinary and cover a wide range of business related issues.**

I. INTRODUCTION

Many companies are struggling with the question of how e-business impacts the supply network. It can eliminate some intermediaries (such as the wholesalers or retailers), but it also fosters the emergence of new players like logisticians, whose role is to adapt traditional logistics chains to take into account the requirements of e-business. Managing the flow of information has become as vital as managing the flow of material (Demkes, 2000) and the Internet has made it possible to connect all of those working within a single supply network. There are new technologies available for optimizing the supply network and most concepts for improving supply network efficiency are based on sharing information throughout the supply network in order to improve transparency and agility.

The questions regarding the e-business and supply network management cover many overlapping themes. It is not always easy to see what the interrelation of these different topics is and therefore the overall picture of e-business logistics needs categorizing into different research and development themes. Guidelines are needed so that the resources can be allocated to the areas in the field of logistics and supply network management, which mostly support development of profitable business practices.

We use the term "e-business logistics" to describe the phenomenon we are focusing in. It is difficult to define e-business logistics comprehensively because the potential impact of e-business on logistics and supply network management is not yet fully understood. One possible definition is that e-business logistics simply means processes necessary to transfer the goods sold over the Internet to the customers. On the other hand it is also possible to define e-business logistics as a wide-ranging topic related to supply network integration with improved operational quality

and cost efficiency, which is the viewpoint taken in this paper. In this paper we first summarize the research methods and results presented in Auramo et al. (2002). We then present a more detailed analysis of the research agenda from supply network point of view using the research program started by National Technology Agency of Finland as a case to illustrate the interrelationship between the identified research themes.

II. KEY R&D TOPICS IN E-BUSINESS LOGISTICS

The method used by Auramo et al. (2002) included aspects of the Delphi method (Linstone, 1975), which employs an iterative process of summarizing and evaluating the respondents' views on a consensus view (McKinnon, Forster, 2000). The method consists of focus interviews, to be followed by a workshop. Ideally, the people interviewed should also attend the workshop. This kind of two -phased process enables the evaluation and prioritization of the preliminary findings. The study (Auramo et al., 2002) consisted of 5 stages as illustrated in the Figure 1. First, the preliminary E-Logistics Vision was created based on the literature study to be used as a discussion guideline during the interviews (Stage 1).

Fifty focus interview sessions were organized, with 65 people interviewed in total (Stage 2). The interviews can be divided into four categories: the leading edge companies (trade and manufacturing companies that have a reputation of being early adapters of new business trends), logistics service companies, consultants and IT companies and researchers at universities and research centers.

The objective of the interviews was to determine the key research and development topics in the field of e-business logistics. The preliminary vision was used as a discussion guideline, which became more defined during the interview process. However, the interview sessions were kept unstructured and open to promote the free flow of ideas. The interviewees from industry, trade and service sectors were responsible for the development of e-business in their companies. Researchers and other professionals in the fields of logistics, supply chain management, e-business and future technologies were also interviewed.

In Stage 3, the data from the interviews was analyzed. The key R&D topics identified during the interview phase were listed. Similar topics were divided into eight preliminary categories, which were formulated during the process. Preliminary categories and respective R&D topics were incorporated into three final categories.

In Stage 4, a workshop panel was organized and the findings from the interviews were prioritized, further evaluated and specified. The workshop was attended by a total of 45 people of which 30 were the same previously interviewed. The question: "What should be the key research and development topics in the field of e-business logistics?" was thoroughly discussed in five parallel groups. The group discussions were semi-structured based on categorized and listed topics. Firstly, the categorization was presented to the workshop and participant's approval was sought. Secondly, the participants were encouraged to identify the most important research and development topics and prioritize them within each category in a consensus view. During the interview and workshop process a total of 80 professional opinions were analyzed. The objective of Stage 5 was to formulate a research and development agenda to guide future research work in the field of e-business logistics. The presented themes of the research agenda are combinations of prioritized R&D topics.

Stage 1	**Preliminary E-Logistics Vision**		
Stage 2	**Interviews** The leading edge companies	Logistics service companies IT-companies and consultants	Universities and research centres
Stage 3	**R&D Topics**		
Stage 4	**Prioritisation of topics; Workshop**		
Stage 5	**Research Agenda**		
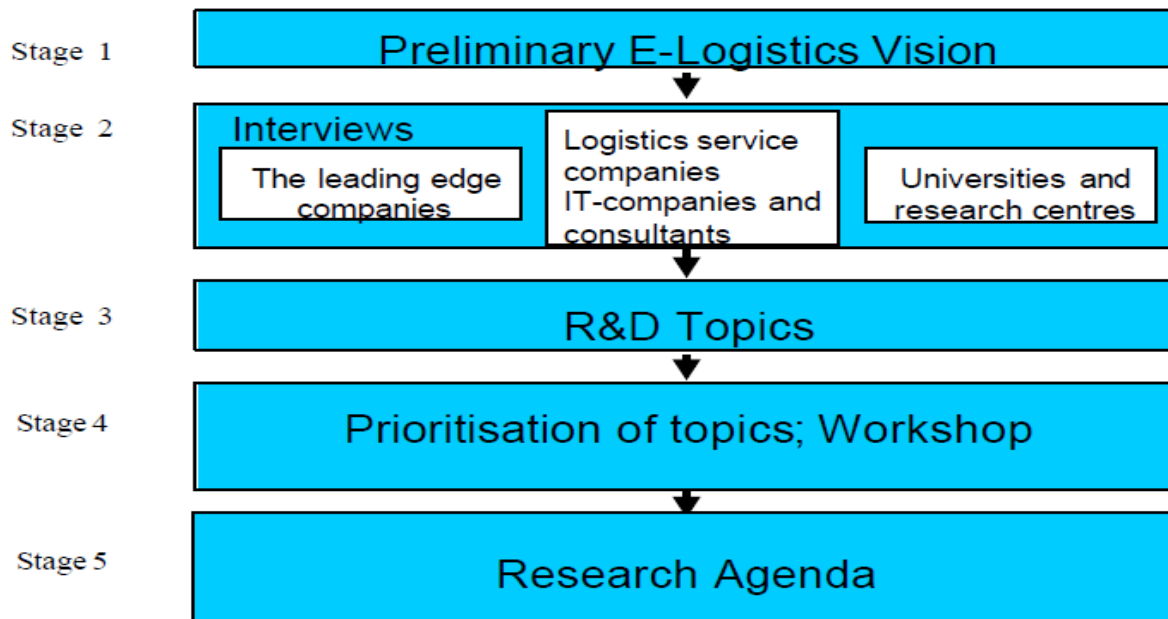
Figure 1: The stages of the study.

III. CATEGORIZATION AND PRIORITIZATION OF THE KEY R&D TOPICS

This section of the paper presents the categorization and prioritization of the key R&D topics found relevant to the development of e- Business logistics (Auramo et al, 2002). The key R&D topics identified during the interviews of logistics professionals were divided into three categories. The categories were:

1. Information flow: the research and development issues dealing with the transparency of the information and its real-time distribution to the necessary parties in the supply network.

2. Value chains: research and development topics related to potential changes in value chains and new collaboration possibilities within networked companies.

3. Physical material flow: utilization of new technologies and their impact on distribution and logistical infrastructure and consolidation possibilities.

planning (Demkes, 1999). Increased visibility and the transfer of real-time information bring new effectiveness to supply network management. To fully benefit from emerging e-business solutions, trading partners must be able to more easily exchange data between their back-

The R&D topics	Importance	The R&D topics	Importance
Development of standards and interpreters.	5	Development of POS (point of sales)- and forecasting data management	2
Integrated ERP (enterprise resource planning)-systems in the supply networks	4	Information management in the supply network: responsibilities, correctness and rights	2
PDM (Product Data Management)	2	Centralized vs. decentralized data management	1

Table I: R&D topics related to the information flow and their importance according to the workshop.

The R&D topics were prioritized within each of the three categories in the workshop. The issues were discussed in five parallel sessions. The participants of each session were encouraged to identify the most important topics and to achieve a consensus view. Three categories and the prioritization within the categories will be covered in more detail. The prioritization of topics is presented in The Tables I, II and III. The numbers 0-5 in the columns headed "Importance" describe how many of the groups identified the topic as "very important".

Information flow
Information and communication technology have played a key role in transforming logistics and supply chain.

Value Chains
Internet technology has dramatically altered the trade-off between vertical and virtual integration (Shah, 2001). The question, "Do the advantages of focusing on core competencies and outsourcing the rest outweigh the cost of managing added complexity", is very relevant. One of the major challenges of e-business is the potential change in the value chains of different Industries. The Internet overturns the old rules about competition and strategy (Werbach, 2000). The changes will vary in Workshop identified the topic as "very important".

end databases and core business applications (Webster, 2001). This must be accomplished within the enterprise as well. Within the supply network the enterprise operates in. The most important R&D topics regarding the information flow within the supply network according to the workshop are presented in the Table I. The results of the prioritization work done during the workshop focuses more on technological issues related to information transfer across the different interfaces within the supply network. However, the overall findings of the study suggest that it is at least as important to foster research and development initiatives that help to understand what information should be shared and how to use the shared information to add value to the supply network.

However, the value chain will only be changed if there is added value for the supply net. The identified key questions were: how to identify and measure the added value for the whole supply net or how to optimize the entire value chain? The R&D topics that help to understand the integration processes and their importance according to the workshop are presented in theTableII.

The R&D topics	Importance	The R&D topics	Importance
Changes in value chain	4	Logistics networks, 4 pl ™	2
Evaluation and development of different e business models	3	Management of channel conflict	2
Modeling and optimizing of the supply network	3	Responsibilities and roles in the e-market places	2
Management of outsourced resources	3		

Table II: R&D topics related to the value chains and their importance according to the workshop.

The main conclusion was that development of IC technology enables business process reengineering and fosters potential changes in the value chains. This creates requirements for new types of logistics services that are supporting new value chains. Research and development initiatives in this area should be focused so that they support the development of integrated and transparent supply network structures.

Physical Material Flow
The development of e-business will and has already set increased quality requirements on logistics services and on the logistics infrastructure as a whole. Customers are demanding products and deliveries customized to their specific needs and flexible outsourcing of supply chain operations is a necessity in volatile business environments (Kärkkäinen & Holmström, 2001). It will be important to be able to provide a different level of service to different customer segments or to individual clients at home. These services must create added value for the clients and at the same time optimize the overall cost structure of the supply network. Table III

Itsummarizes the key research and development topics and their importance according to the workshop. The numbers 0-5 describe how many of the groups in the workshop identified the topic as "very important".

As a conclusion, R&D topics related to physical material flow should be divided into two major themes: technology related R&D to support the development of new logistics services to enable the necessary changes in the value chains and analysis of potential changes in the national logistics infrastructure to support e-business development.

The R&D topics	Importance	The R&D topics	Importance
National logistics infrastructure and requirements of e-business to it.	4	Improvement of order fulfillment accuracy	4
Home delivery	3	Service level variation	1
Management of regional differences	3	Physical logistics of e-market places.	1
Identification technologies	4	Material handling technologies	1
Mobile technologies	3	Positioning technologies	1
Consolidation of material flows, cross-docking	2	Reverse logistics, direct deliveries	0

Table III R&D topics related to physical material flow and their importance according to the workshop.

IV. RESEARCH AGENDA TOWARDS SUPPLY NETWORKS

In the second part of this paper we will present a proposal for the research agenda to support the development of e-business logistics i.e. to support the development of supply network integration with improved operational quality and cost efficiency. The research and development themes we want to focus on are combinations of different R&D topics presented in the previous chapter supported with the findings from the international body of knowledge in the field. The proposed themes are:

Supply Integrated network structure with suitable Visibility:

Creating integrated supply networks with suitable visibility is a very complicated exercise. There are at least three major elements that need to be solved almost simultaneously. Solutions are needed to enable information flow across company interfaces. There must also be know-how of what information should be shared and how to utilize the shared information. And last but not least there are potential changes in the structures of the value chains. During the iterative process of interviews and workshops conducted by Auramo at al (2002) the information transfer within the supply network was seen as one of the key research and development areas. But information transfer is not enough. There needs to be solutions how real-time information and data can be shared among the relevant players within the supply network. E-collaboration is According to e-collaboration experts it is important to understand what information should be shared and how to add value to the information shared (Cameron&Gromley, 1998; CPFR, 2001; Holmström et al. 2000). When transferring and sharing information within the supply network the questions of who manages and owns the information and who guarantees the correctness of the information become vital issues. This can be related to the potential changes in the structures of the value chains. Finally, it seems that proper tools and technologies are currently available. The question is how to change the current operations models and how to

Research on identification methods in the supply network:

Technological innovations and huge increases in efficiency are often connected together (Brynjolfsson & Hitt, 1998). Newtechnologies, like product identification technologies, mobile technologies and applications that utilize satellite location technologies (GPS) (Jedd, 2000; Shulman, 1999, Radding, 1994) are seen as enablers for

i. integrated supply network structure with suitable visibility and collaborative usage of real-time data;

ii. Research and development of new logistics service concepts and their effect on the whole supply network; and

iii. research on the effects and possibilities of using new identification methods such as RFID (e.g. product, parcel or batch identity) in the supply network.

It manages the changes when implementing new business rules.

R&D of new logistics services in the supply network:

More sophisticated customer demand chains and electronic business poses new challenges to supply chain management. According to the interviews the development of e-business will increase the importance of, especially, delivery accuracy; delivery frequency and delivery lead times based on to differentiated customer demands. Customers are demanding products and deliveries customized to their specific needs A major challenge is to identify and develop new logistics service concepts, which could more efficiently manage, consolidate and optimize both information and material flows in networked environment. These could be service providers like 4PLs ☐ (Bauknight, 2001; Bade et al., 1999), information intermediaries (Timmers, 1999) or specialized logistics companies. New evolving operating models, like cross-docking (Daugherty, 1994), consolidation (Hall, 1987, Gooley, 2000) and direct deliveries together with material handling technologies and automation, enable efficient supply networks. New logistics services would enable companies to focus on their core competencies (Prahalad & Hamel, 1990), however the problem is how to manage the networked resources. Another problem is related to the selection of service providers.

efficient supply networks. According to the workshop opinion especially mobile data transfer together with RFID (radio frequency identification) may lead to great efficiency improvements. Use of wireless identification technology could be one of the keys to significantly speed up and increase accuracy in sorting and distribution (Jones, 1999, Boxall, 2000 and Lindström, 2000). For example, wireless product identification is already used with great benefits in the functional areas of innovative companies and there is potential for its use in

supply chain wide solutions, i.e. item level supply chain management (Kärkkäinen & Holmström, 2001). But the current knowledge is still insufficient and further research should be supported in this field However, research and development related to this theme should be approached not only from the technology-perspective but also from business need-

Interrelationships between the R&D themes:
National Technology Agency of Finland has started a four-year (2002 -2005) research program for developing e-business logistics. The budget of the program is 25 million euros and it focuses on logistical challenges in e-business environment. The program is targeted for research institutes as well as for company driven projects. Currently preliminary R&D proposals from research institutes have been received by the National Technology Agency. The division
of the proposals was as follows:
i. 86% of these R&D proposals were related to the problem area of integrated supply network. Research institutes were planning to, for example, model and pilot new operations models and develop demand forecasting and information sharing.
ii. 50% of the R&D proposals were aimed to developing new logistics services necessary in networked.

perspective. There are some relevant questions related to these perspectives:
Technology-perspective: What are the possibilities that new technologies bring?
Business need-perspective: Are there new technology demands when developing supply network management? Are the emerging technologies fully applied when developing the networks?

The objective seems to be to identify emerging service needs and develop suitable service concepts.
iii. 14% of the R&D proposals focused on applying emerging identification technologies in developing supply chain management.
As percentage numbers show, most of the R&D initiatives are overlapping across at least two of the three research themes Described above. The division of the R&D proposals from the research institutes supports the understanding that even though the presented R&D themes seem to be independent entities, there is a very strong interrelationship between them in the context of e-business logistics.

Technology innovations are enablers to business process re -engineering aiming towards supply network integration.

V. CONCLUSIONS

This paper was divided into two parts. The first part of the paper summarized the categorization and prioritization of the most crucial R&D topics that should be dealt with in order to help enterprises to face the new logistics challenges of e-business. On the second part of the paper a research agenda towards supply networks was presented with three themes to support the development of e-business logistics. In order to achieve efficiency and operational quality in supply networks we suggest three interrelated R&D themes: The integrated supply network structure with suitable visibility and collaborative usage of real-time data;□ □ research and development of new logistics service concepts and their effect on the whole supply network; and research on the effects and possibilities of using new identification methods in the supply network. The R&D initiatives in these fields ought to be interdisciplinary and cover a wide range of business related issues. There ought to be space for development initiatives with both shorter and longer time span. In addition to the R&D initiatives aiming to develop new technologies it is as important to

The opportunities to utilize currently available and emerging technologies and innovations in supply networks. However, new technologies are enablers to network integration, successful interaction requires confidence and excellent professional skills from the various supply network partners.

REFERENCES

[1] Auramo J, Aminoff A, Punakivi M, (2002) *Research agenda for e-business logistics based on professional opinions,* Forthcoming in the International Journal of Physical Distribution and Logistics Management
[2] Bade, D., Mueller, J. (1999), *New for the millennium: 4PL*, Transportation & Distribution, Vol. 40 Iss.2, Feb 1999 , pp.78-80.
[3] Bauknight, D. (2001), "Fourth Party Logistics -- Breakthrough Performance in Supply Chain Outsourcing", http://www.accenture.com
[4] Boxall, G., (2000), "The Use of RFID for Retail Supply Chain Logistics", *Tag 2000*, 24.5.2000,
[5] Brynjolfsson, Erik & Hitt, Lorin M., (1998), "Beyond Productivity Paradox: Computers are the Catalyst for Bigger Changes",
Communications of the ACM, Vol.41, No. 8, August 1998.

Page Ranking Algorithm Analysis

Rashmi Vipat
Assistant Professor, TIMSCDR

Sowrabh Kukalkar
Student, TIMSCDR

Rahul Singh Karki
Student, TIMSCDR

***Abstract** -* **In our day-to-today life, we use Search Engines to search for various information across and over the Internet. Internet is a huge ocean of data and increasing every seconds and has a great importance. Such diverse information result in various tool to cut down the time for search. Page Rank is an algorithm used to measure the importance of webpages using hyperlinks between pages. The main disadvantage is that older pages has a good page rank, because a new page, even a very good one, will not have many links unless it is part of an existing site. We review two approaches for ranking: HITS concept and Page Rank method. Both approaches focus on the link structure of the Web to find the importance of the Web pages. The Page Rank algorithm calculates the rank of individual web page and Hypertext Induced Topic Search (HITS) depends upon the hubs and authority framework. A fast and efficient page ranking mechanism for web retrieval remains as a challenge.**

Index Terms - Ranking, Page Rank, HITS, Hyperlink

I. INTRODUCTION

The web as we all know is the largest source of data. During the past few years the World Wide Web has become the foremost and most popular way of communication and information dissemination. It serves as a platform for exchanging various kinds of information, ranging from research papers, and educational content, to multimedia content, software and personal logs. Every day, the web grows by roughly a million electronic pages, adding to the hundreds of millions pages already on-line. So with the rapid growth of information sources available on the World Wide Web, it has become increasingly necessary for users to use automated tools to find the desired information resources, and to track and analyze their usage patterns. The web creates new challenges for information retrieval. The amount of information on the web is growing rapidly, as well as the number of new users inexperienced in the art of web research. People are likely to surf the web using its link graph, often starting with high quality human maintained indices such as Yahoo! or with search engines.

Web mining can be broadly defined as the extraction and mining of useful information from the World Wide Web. Web Structure Mining is the process of discovering information from the Web, finding information about the web pages and inference on hyperlink, finding authoritative web pages, retrieving information about the relevance and the quality of the web page. Thus Web structure mining focuses on the hyperlink structure of the web. We review two approaches: HITS concept and Page Rank method. Both approaches focus on the link structure of the Web to find the importance of the Web pages. Mainly In links to the pages and out links from the page can give idea about the context of the page. Page Rank does not rank web sites as a whole, but it calculates the rank of

individual web page and Hypertext Induced Topic Search (HITS) depends upon the hubs and authority framework.
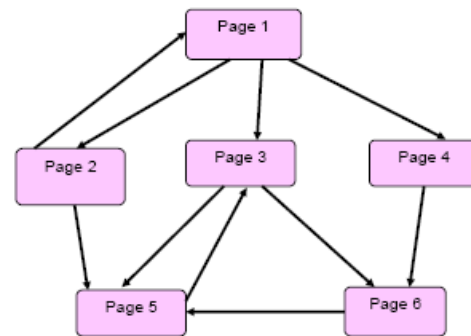

Fig 1: Web Graph

We provide here an overview of Recursive Data Mining. The rest of this paper is organized as follows: Section II introduces Background and Related Work; followed by conclusion, acknowledgement and references

II. BACKGROUND AND RELATED WORK

A web search engine typically consists of:

1. Crawler: used for retrieving the web pages and web contents
2. Indexer: stores and indexes information on the retrieved pages
3. Ranker: Measure the importance of Web Pages returned
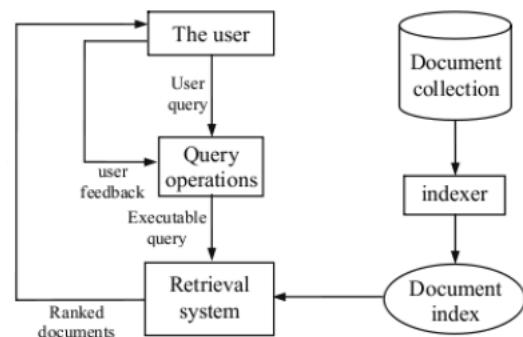4. Retrieval Engine: performs lookups on index tables against query

Fig 2: A search engine system
during a search operation

A user issues a query which is first checked before it is forwarded and compared to documents indexes.

III. HITS ALGORITHM

The HITS algorithm stands for "Hypertext Induced Topic Selection" and is used for rating and ranking websites based on the link information when identifying topic areas. Kleinberg's hypertext-induced topic selection (HITS) algorithm is a very popular and effective algorithm to rank documents based on the link information among a set of documents. The algorithm presumes that a good hub is a document that points to many others, and a good authority is a document that many documents point to. Hubs and authorities exhibit a *mutually reinforcing relationship*: a better hub points to many good authorities, and a better authority is pointed to by many good hubs. To run the algorithm, we need to collect a base set, including a root set and its neighborhood, the in-and out-links of a document in the root set. HITS calculate hub and authority scores per query for the focused subgraph of the web. A good authority must be pointed to by several good hubs while a good hub must point to several goods authorities. User queries are generally divided into two types. The specific query where the user requires exact matches and narrow information, and broad-topic query for user who look for narrow answers *and* information relation to the broad topic. HITS concentrates on the latter type and aims to find the most authoritative and informative pages for the topic of the query.

HITS algorithm, can be stated as follows:
1. Using existing system, get the root set for the given query.
2. Add all the pages linking to and linked from pages in the root set, giving an extended root set or base set.
3. Run iterative eigenvector based computation over a matrix derived from the adjacency matrix of the base set.
4. Report the top establishment and hubs.

IV. PAGE RANK

Page Rank is the numerical value, weighting 1-10 for any single website or web page. Page Rank represents the PR (Page Rank) strength of any given website. However, Page Rank should not be confused with the actual ranking strength of any website (i.e. your actual SERP's). Page rank is generally calculated by search engines by analyzing and calculating various back links pointing towards a particular website. Google Page rank algorithm uses its own complex algorithmic formulas to calculate Page Rank of any given website or web page. So far so good, but the real question comes up "Whether Page Rank has any relationship with the actual rankings of your website?" The answer is a big NO. There are several hundred webmasters discussing increase/decrease in Page Rank on various forums, but none of them realize the importance of their actual SERPs as compared to Page Rank. There are several hundred websites with low PR (1 or 2), and several of these websites genuinely overshadow other high-ranking competitor websites with

higher Page Ranks (generally 4-5 or even higher than that). There is a wrong misconception among several webmasters that higher the page rank, higher will be their rankings on search engines. The truth is that your increase or decrease in page rank has no relationship with the actual rankings of your website.

There are several webmasters who will argue that if any website sees a steep decline in their Page Rank, then it could be an indication of drop in their actual SERPs. If your website is having some major SEO issue related to content or links on your website, then your actual rankings (SERPs) will drop first. Once you observe a drop in SERPs, your Page Rank will eventually drop after a period of 2-3 months. Till the next update in Page Rank, the Google Toolbar will keep pulling PR Data from its stored database. Your SERPs relies on the current status of your website whereas PR relies on the stored database which gets updated only once in every 2-3 or 3-4 months.

Features of Page Rank Algorithm are:
- It is the query independent algorithm that assigns a value to every document independent of query.
- It is Content independent Algorithm.
- It concerns with static quality of a web page.
- Page Rank value can be computed offline using only web graph.
- Page Rank is based upon the linking structure of the whole webpage.
- Rank does not rank website as a whole but it is determined for each page individually.
- Page Rank of pages Ti which link to page A does not influence the rank of page A uniformly.
- More the outbound links on a page T, less will page A benefit from a link to it.
- Page Rank is a model of user's behavior

V. COMPARISON BETWEEN PAGE RANK AND HITS ALGORITHM

HITS, like Page and Brin's PageRank, is an iterative algorithm based on the linkage of the documents on the web. However it does have some major differences:
- It is executed at query time, not at indexing time, with the associated hit on performance that accompanies query-time processing. Thus, the *hub* and *authority* scores assigned to a page are query-specific.
- It is not commonly used by search engines.
- It computes two scores per document, hub and authority, as opposed to a single score.
- It is processed on a small subset of „relevant" documents, not all documents as was the case with PageRank.

Link graph features such as in-degree and PageRank have been shown to significantly improve the performance of text retrieval algorithms on the web. The HITS algorithm is also believed to be of interest for web search; to some degree, one

may expect HITS to be more informative that other link-based features because it is query-dependent: it tries to measure the interest of pages with respect to a given query. However, it remains unclear today whether there are practical benefits of HITS over other link graph measures. This is even truer when we consider that modern retrieval algorithms used on the web use a document representation that incorporates the document's anchor text, *i.e.* the text of incoming links. This, at least to some degree, takes the link graph into account, in a query-dependent manner. Comparing HITS to PageRank or in-degree empirically is no easy task.

There are two main difficulties: scale and relevance. Scale is important because link-based features are known to improve in quality as the document graph grows. If we carry out a small experiment, our conclusions won't carry over to large graphs such as the web. However, computing HITS efficiently on a graph the size of a realistic web crawl is extraordinarily difficult. Relevance is also crucial because we cannot measure the performance of a feature in the absence of human judgments: what is crucial is ranking at the top of the ten or so documents that a user will peruse. To our knowledge, this paper is the first attempt to evaluate HITS at a large scale and compare it to other link-based features with respect to human evaluated judgment. Our results confirm many of the intuitions we have about link-based features and their relationship to text retrieval methods exploiting anchor text. This is reassuring: in the absence of a theoretical model capable of tying these measures with relevance, the only way to validate our intuitions is to carry out realistic experiments. However, we were quite surprised to find that HITS, a query-dependent feature, is about as effective as web page in-degree, the most simpleminded query-independent link-based feature. This continues to be true when thelink-based features are combined with a text retrieval algorithm exploiting anchor text.

The ranking function has many parameters like the type-weights and the type-prox-weights. Figuring out the right values for these parameters is something of a black art. In order to do this, we have a user feedback mechanism in the search engine. A trusted user may optionally evaluate all of the results that are returned. This feedback is saved. Then when we modify the ranking function, we can see the impact of this change on all previous searches which were ranked. Although far from perfect, this gives us some idea of how a change in the ranking function affects the search results.

For search-engine optimization purposes, some companies offer to sell high Page Rank links to webmasters. As links from higher-PR pages are believed to be more valuable, they tend to be more expensive. It can be an effective and viable marketing strategy to buy link advertisements on content pages of quality and relevant sites to drive traffic and increase a Webmaster's link popularity. However, Google has publicly warned webmasters that if they are or were discovered to be selling links for the purpose of conferring Page Rank and reputation, their links will be devalued (ignored in the calculation of other pages' Page Ranks). The practice of buying and selling links is intensely debated across the Webmaster's community. Google advises webmasters to use the no follow HTML attribute value on sponsored links.

According to Matt Cutts, Google is concerned about webmasters who try to game the system, and thereby reduce the quality and relevancy of Google search results. Trust Rank is one of the popular method for improving rankings is to increase artificially the perceived importance of a document through complex linking schemes. Google's PageRank and similar methods for determining the relative importance of Web documents have been subjected to manipulation. Trust Rank can be bolted onto Page Rank to significantly improve search relevancy.

Additionally, in computing the level of relevance, we require a match between the query and the text on the expert page which qualifies the hyperlink being considered. This ensures that hyperlinks being considered are on the query topic. For further accuracy, we require that at least 2 non-affiliated experts point to the returned page with relevant qualifying text describing their linkage. The result of the steps described above is to generate a listing of pages that are highly relevant to the user's query and of high quality.

ACKNOWLEDGEMENT

REFERENCES

[1] HAVELIWALA, T. H. Topic-Sensitive PageRank: A Context-SensitiveRanking Algorithm for Web Search. In IEEE Transactions on Knowledge and Data Engineering (July 2003).

[2] B. Liu. Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data. Springer, 2006.

[3] Robots Exclusion Protocol:
http://info.webcrawler.com/mak/projects/robots/exclusion.html

[4] Google Search Engine http://google.stanford.edu

[5] [Abiteboul 97] Serge Abiteboul and Victor Vianu, *Queries and Computation on the Web*. Proceedings of the International Conference on Database Theory. Delphi, Greece 1997.

[6] [Bagdikian 97] Ben H. Bagdikian. *The Media Monopoly*. 5th Edition. Publisher: Beacon, ISBN: 0807061557

[7]. Web Growth Summary: http://www.mit.edu/people/mkgray/net/web-growth-summary.html

Research on Barcode Positioning System

Rashmi Vipat
Assistant Professor, TIMSCDR

Sundeep Sharmadhikari
Student, TIMSCDR

Abstract – **Aiming at the disadvantages of the traditional positioning technology, barcode positioning system is introduced in this paper. Based on Otsu method, a novel barcode image binarization is put forward by comparing varieties of image binarization methods domestically and abroad. Moreover, we have a systematic research on histogram and binarization mechanism, and also give the calculation of histogram and derive a formula of Otsu method. Finally, the histogram and binarization of one-dimensional barcode image are realized with the specific examples. After experiments for scanned barcode image, the result has demonstrated effectiveness of the method. Keywords: binarization, histogram, Otsu method, barcode positioning.**

I. INTRODUCTION

At present, the common positioning methods of the transportation system are in two ways: One is positioning with laser or ultrasonic [1], whose advantage is that positioning accuracy and speed, but it only works on a straight line; and the other is positioning with the rotary encoder, whose advantage is can be positioned in the curve direction, but its accuracy is poor, and has the cumulative error. In view of these phenomena, the introduction of barcode positioning system can perfectly resolve the above problems. The barcode positioning system is a new measurement and positioning system, which is evolved from the large-scale logistics transportation system, is a breakthrough of the traditional positioning technology, and represents the development direction of positioning technology in the modern largescale transportation system. In this paper, the improved binarization based on Otsu method is put forward and applied to barcode positioning system according to its own characteristic of laser scanned barcode image. Finally, the histogram and binarization of one-dimensional barcode image are realized with the specific examples in Matlab programming environment, which has an excellent practical value.

II. BASIC PRINCIPLES OF BARCODE POSITIONING

The barcode positioning system is composed of the barcode reader and the barcode tape, etc. It works with the barcode reader installed on the robot, and the barcode tape installed on a walking track. When the robot is walking on the track, the barcode reader scans the current barcode constantly, and outputs the robot's current location information through the built-in decoder. The schematic diagram of the barcode positioning system is as shown in Fig. 1. This system is mainly composed of the barcode tape, the barcode reader and the controller. The barcode tape has printed a number of fixed-length barcodes with certain rules. The barcode reader consists of scanning system, signal shaping system and decoder. Laser emitted from the light source scans the barcode, diffuse-reflected light from the barcode is absorbed by the photoelectric converter, and the reflected light signal is converted into the corresponding electrical signal. The electrical signal is converted into digital signal after signal shaping. The decoder distinguishes the number of bars and spaces by measuring the number of digital signal 0/1, and distinguishes the width of bars and spaces by measuring the duration of digital signal 0/1. According to the encoding rules, the signal combination of the bars and spaces is converted into location data. Finally, the decoded location data are sent to the controller through the interface.
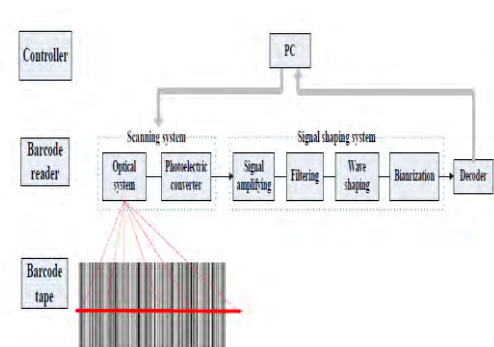

Fig 1: The schematic diagram

III. THE BINARIZATION OF BARCODE IMAGE SYSTEM

In the process of barcode scanning, the light and distance would have great changes every time, so the barcode image (also referred to as digital signal) after signal shaping is binarization processed with a dynamic threshold method.

Histogram:

In the view of probability, if the pixel intensity (gray level) is regarded as a random variable, then the percentage that the pixel number of a certain gray level accounts the total pixel number reflects the statistical characteristics of one image, which can be described as Probability Density Function (PDF) and be defined as gray histogram. Gray histogram is a function of gray level, which indicates the pixel number of a certain gray level and reflects the frequency of every gray level. Let r represent the pixel gray level. Normalizing the pixel gray level, then the value of r will be restricted to the range $[0,1]$, where r = 0 represents black and r = 1 represents white. For a given image, it is random to take every gray level in $[0,1]$. In other words, the gray level is a random variable. In the discrete form, r_i denotes the discrete gray level and Pr (r_i) denotes the Probability Density Function. In order to simplify the discussion, the gray level histogram is normalized and regarded as a probability distribution:

Where r_i denotes the pixel number at r_i , n denotes the total pixel number, n_i n denotes the frequency of the gray level, l denotes the total number of gray level.

$$P_r(\eta)=\frac{n_i}{n} \quad (0 \le \eta \le 1 \quad i=0,1,2,\text{L},l-1) \qquad (1)$$

After that, make a diagram in the rectangular coordinate system for the relationship of ri and P(ri) , you get the histogram. Suppose that the gray level of the continuous image can be changed smoothly from the center of higher gray level to the edge of lower gray level. Now select a gray level D1, and define a contour that connects all the points of the gray level D1 . The contour forms a closed curve which surrounded by the gray level that greater than or equal to D1. And then define another contour with greater gray level D2 .As shown in Fig.2, A1 and A2 are the areas surrounded by the contour of gray level D1 and D2 , respectively. A(D) is the threshold area function, which represents the area surrounded by the gray level D . Then the image histogram can be defined as

$$H(D) = \lim_{\Delta D \to 0} \frac{A(D)-A(D+\Delta D)}{D-(D+\Delta D)} = \lim_{\Delta D \to 0} \frac{A(D)-A(D+\Delta D)}{-\Delta D} = -\frac{d}{d/D}A(D) \quad (2)$$

Therefore, the histogram of a continuous image is the negative of its threshold area function derivative. And the symbol of negative is due to that the threshold area decreases with D increases.

For the discrete function, set .D be 1, then

$$H(D)=A(D)-A(D+1) \qquad (3)$$

For the digital image, the area of gray level D is the pixel number of the gray level that greater than or equal to D .
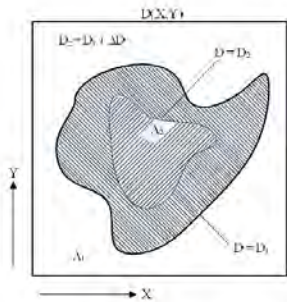

Fig 2: The characteristics of histogram

The characteristics of histogram are as follows:

 (1) The histogram is the statistic of every gray level, which reflects the occurrence number or frequency rather than the location of a certain gray level. In other words, it contains only the probability of a gray level but loses its location information in the image.

 (2) Any image can identify its corresponding histogram uniquely. However, different images may have the same histogram. That is, the relationship between one image and its histogram is many-to-one mapping relationship.

(3) The histogram is obtained with the pixel statistic of the same gray level, so the histogram of every sub-area is equivalent to the histogram of the whole image.

The calculation of histogram: Let the pixels of a given image be represented in L gray levels (L.256 , that is 8-bit gray level), then the histogram pBuffer[0,L ,L-1] of the gray image (M x N) can be available with the following algorithm [9]:

Step1. Initialization: pBuffer[i]=0 (i=0,L ,L-1) ;
Step2. Statistic: pBuffer[f (x,y)]++ (x=0,L ,M-1; y=0,L ,N-1);
Step3. Normalization: pBuffer[f (x,y)]/.M x N .

Among them, the normalization of histogram is optional, you could neglect this operation if special process is needless.

OTSU: After the foregoing process of barcode image, we may select a simply global threshold method for binarization. Because of simple computation and well adaptation, the Otsu method becomes one of the most popular methods of threshold selection. This paper chooses the Otsu method for image binarization. The Otsu method is an adaptive threshold method, which is OTSU for short, is also known as the maximum between-class variance method. It dichotomizes the image into two parts of the background and objective according to the gray level of the image. The greater the betweenclass variance between the two parts, the greater the difference of the two parts. When parts of the objective are wrongly divided into the background or parts of the background are wrongly divided into the objective, it will lead to the difference smaller [4]. Therefore, the segmentation that makes the between-class variance maximum means that the error dividing probability is the minimum.

IV. EXPERIMENTAL RESULTS

To verify the rationality of the method, we compare the performance of the proposed method with traditional method, such as [3] (referred to as Otsu N) and [5] (referred to as L. L. Li). As shown in Fig.3 and Table 1, the histogram and binarization of one-dimensional barcode image are realized with the specific examples in Matlab programming environment, and the result with threshold of Otsu is near to the result with manual threshold (the optimal threshold). a) is the original barcode image, b) is the histogram of image, c) is the result with manual threshold, d) is the result with default threshold, e) is the result with threshold of Otsu method.

Therefore, the barcode image that binarization processed with the dynamic threshold method has satisfactory performance. However, the optimal threshold may be available after several attempts if without Otsu method. Because of the relatively low measurement accuracy, laser scanning usually cannot satisfy the requirements for the measurement of the barcode positioning system. Through analysis in principle and lots of experiments, a BP neural network model can be established to realize soft-ware compensation for the measurement error.

Fig.3 Experimental comparison on barcode images

Table 1: Comparison segmentation processes

Image	Manual threshold	Default threshold	Otsu
Image1	159	123	155
Image2	146	215	140

V. CONCLUSION

Regard the real-time of barcode positioning system and the characteristic of laser scanned barcode image as the starting point, the simulation experiments are realized with barcode image binarization base on Otsu method. Besides, we give the calculation of histogram and derive a formula of Otsu method. The method selects the optimal threshold according to its own information of barcode image. After binarization processing, the edge of barcode image is clear and the background noise is little, so that it provides a favorable foundation for the barcode positioning system.

REFERENCES

[1] G. Hueber, T. Ostermannn, T. Bauernfeind, R. Raschhofer, R Hanelauer, "New Approach of Ultrasonic Distance Measurement Technique in Robot Applications ", 5th International Conference on Signal Processing Proceedings, 2000, pp. 2066-2069.
[2] L. Jiang, "The Barcode Location and Recognition Base on Run-length Coding", Electronics World, Vol. 3, 2012, pp. 16-19.
[3] Otsu N, "A Threshold Selection Method From Gray-level Histograms", IEEE Transaction on Systems, Man and Cybernetics, Vol. 9, No. 1, 1979, pp. 62-66.
[4] J. P. Wu, Z. X. Yang, D. Han and Z. F. Bai, "2D Barcode Image Binarization Based on Wavelet and Otsu Method", Computer Engineer, Vol. 36, No. 10, 2010, pp. 190-192.
[5] L. L. Li, S. X. Deng and X. H. Ding, "Binarization Algorithm Based on Image Partition Derive From Da-Jing Method", Control & Automation, Vol. 21, No. 8, 2005, pp. 76-77.

[6] X. Y. Xu, E.M. Song and L. H. Jin, "Characteristic Analysis of Threshold Based on Otsu Criterion", Acta Electronic Sinica, Vol. 12, 2005, pp. 2716-2719.
[7] Y. S. Wu, J. Deng, "Barcode Location by Extracting Edge Lines", Journal of Computer Applications, Vol. 30, No. 5, 2010, pp. 1246-1250. [8] Q. Du, J. Y. Xiang and S. C. Yuang, "An Improved Method of Maximum Classes Square Error", Infrared Technology, Vol. 25, No. 5, 2003, pp. 33-36.
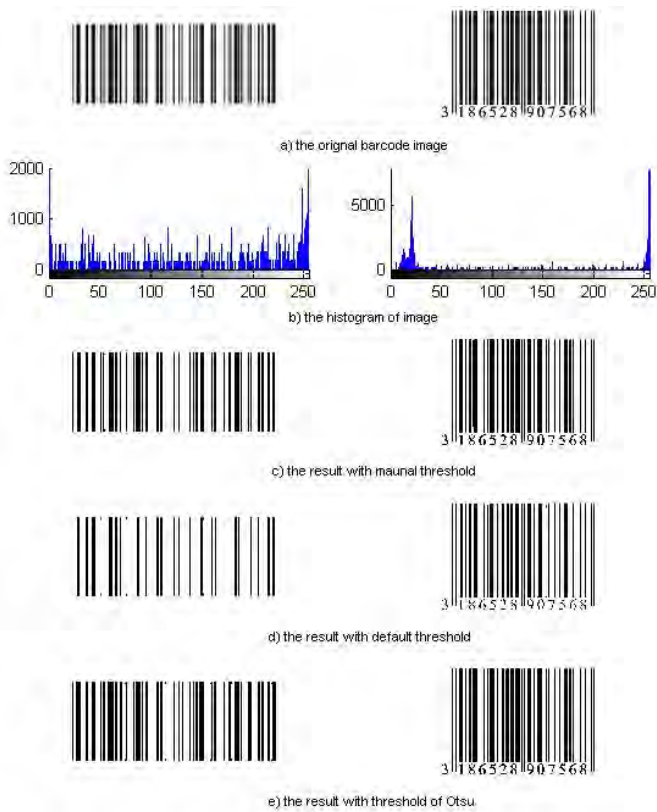
Security of E-Commerce

Pankaj Mudholkar
Assistant Professor, TIMSCDR

Kirti Giri
Student, TIMSCDR

Neeraj Singh
Student, TIMSCDR

Maroti Pawar
Student, TIMSCDR

Mustafa Mh. Shaikh
Student, TIMSCDR

Abstract—**E-commerce Security is a part of the Information Security framework and is specifically applied to the components that affect e-commerce that include Computer Security, Data security and other wider realms of the Information Security framework. E-commerce security has its own particular nuances and is one of the highest visible security components that affect the end user through their daily payment interaction with business.**

I. INTRODUCTION

Electronic commerce, or e-commerce, refers to economic activity that occurs online. E-commerce includes all types of business activity, such as retail shopping, banking, investing and rentals. Even small businesses that provide personal services, such as hair and nail salons, can benefit from e-commerce by providing a website for the sale of related health and beauty products that normally are available only to their local customers. Electronic commerce draws on technologies such as mobile commerce, electronic funds transfer, supply chain management, Internet marketing, online transaction processing, electronic data interchange (EDI), inventory management systems, and automated data collection systems. Modern electronic commerce typically uses the World Wide Web at least at one point in the transaction's life-cycle, although it may encompass a wider range of technologies such as e-mail, mobile devices social media, and telephones as well. E-commerce differs from e-business in that no commercial transaction, an exchange of value across organizational or individual boundaries, takes place in e-business. The buying and selling of products, services by business and consumers through an electronic medium, without using any paper documentsM-commerce shares security concerns with other technologies in the field. Privacy concerns have been found, revealing a lack of trust in a variety of contexts, including commerce, electronic health records, e-recruitment technology and social networking, and this has directly influenced users. Security is one of the principal and continuing concerns that restrict customers and organizations engaging with ecommerce. Web e-commerce applications that handle payments (online banking, electronic transactions or using debit cards, credit cards, PayPal or other tokens) have more compliance issues, are at increased risk from being targeted than other websites and there are greater consequences if there is data loss or alteration. Online shopping through shopping websites having certain steps to buy a product with safe and secure. The e-commerce industry is slowly addressing security issues on their internal networks. There are guidelines for securing systems and networks available for the ecommerce systems personnel to read and implement. Educating the consumer on security issues is still in the infancy stage but will prove to be the most critical element of the e-commerce security architecture.

II. RELATED WORKS

Security is one of the principal and continuing concerns that restrict customers and organizations engaging with ecommerce. The aim of this paper is to explore the perception of security in e-commerce B2C and C2C perspectives. E-commerce tends to be at a higher echelon for risk and attacks. This is so because according to our definition, E-Commerce is the transaction of goods and services; and the payment for those goods and services over the Internet. Therefore, the physical place where all of these transactions occur is at the Server level. The server can be viewed as the central repository for your "E-Commerce Place of Business" Threats to E-Commerce servers fall into two general categories: (1) Threats from an actual attacker(s); and (2) Technological failure. In terms of the former, the motivation is primarily psychological. The intent is to garner personal information from people for the sheer purposes of exploitation (such obtaining Credit Card and Bank Account information; Phishing schemes, obtaining usernames and passwords, etc.) Web applications increasingly integrate third-party services. The integration introduces new security challenges due to the complexity for an application to coordinate its internal states with those of the component services and the web client across the Internet. Each phase of E-commerce transaction has a security measures.

E-commerce Transaction Phases			
Information Phase	Negotiation Phase	Payment Phase	Delivery Phase
Security Measures			
Confidentiality Access Control Integrity Checks	Secure Contract Identification Digital Signatures	Encryption	Secure Delivery Integrity Checks

Threats to E-Commerce servers fall into two general categories: (1) Threats from an actual attacker(s); and (2) Technological failure. In terms of the former, the motivation

is primarily psychological. The intent is to garner personal information from people for the sheer purposes of exploitation (such obtaining Credit Card and Bank Account information; Phishing schemes, obtaining usernames and passwords, etc.). With the latter, anything related to the Internet can cause problems. This can be anything from a network not configured properly to data packets being lost, especially in a wireless access environment. Even poorly written programming code upon which your E-Commerce site was developed can be very susceptible to threats. Most E-Commerce Servers utilize a Windows Operating System (such as Windows 2000 and 2003 Server), a Web Server Software to host the E-Commerce Site (such as Internet Information Services, or IIS), and a database (such as Access 2000 or SQL Server 2000) which contains your customer information and transaction history.

These platforms have had various security flaws associated with them, which has made them wide open to threats and attacks. As a result, there has been a move in the business community to adopt more robust and secure platforms. A prime example of this is the use of Linux as the operating system, Apache as the Web Server Software, and either PostGRESql or My SQL as the database (these are database languages created from the Structured Query Language, or SQL). These latter platforms will be explored in much more detail in subsequent articles.

A. *Units*
Computerised technology may include:
- personal computers or terminals (stand-alone or networked)
- scanning equipment
- bar coding equipment
- point-of-sale terminals
- hand-held equipment
- software, including:
- word processing
- databases
- spreadsheets
- financial
- inventory

- electronic data interchange (EDI).

Systems may include:
- electronic
- manual.

e -commerce may include:
- business to business
- business to end-consumer (direct)
- consumer to consumer.

Maintenance requirements may include:
- cleaning
- hardware or software upgrades
- preventative maintenance
- OHS requirements.

Reporting may be:
- formal and informal
- to individuals or groups
- written, faxed, emailed or spoken.

Sources of information may include:
- legislation and regulations
- work procedures and internal manuals
- company operating procedures and instructions
- manufacturer specifications
- supplier and customer instructions
- manifests
- codes of practice
- quality assurance systems, procedures and policies
- awards, workplace agreements, and other industrial arrangements
- verbal communication, including face-to-face, telephone, internet and radio
- written instructions and communication such as data exchange, letters and emails
- safety instructions and workplace signage
- electronic or hard copy
- policy and protocols.

Basic transactions may include:
- sales
- quotations
- returns
- administration of accounts
- electronic transactions, including:
- electronic data interchange (EDI)
- MIME
- value added networks
- payments received by:
- cash
- cheque
- credit card
- EFTPOS
- cash on delivery (COD)
- direct credit.

Transaction security may include:
- principles of digital cryptography, steganography and public-key cryptosystems

- cryptographic standards (e.g. government security policy, international standards)
- digital signatures, digital escrow, certification
- virus protection
- secure communications
- intrusion detection and countermeasures
- copy and counterfeit detection
- privacy and anonymity protocols.

Credit checks may be:
- automated or manual
- completed internally or by external agent.

Business policy and procedures in regard to:
- acquisition and sale of products and services
- reporting mechanisms
- interaction with customers
- information technology systems
- processing e-commerce transactions.

B. *Equations*

Understanding the "state of": First thing is first, you should know every number in this formula for your site, and in fact, you should know it from memory. If you don't have these numbers, or if your analytics doesn't provide it, the first step is getting it in place. There really is no excuse to not having this data. By understanding where you are today, you'll be more informed to make decisions on how to move forward. And we suggest tracking this data over time, if you even do this monthly you'll have a better understanding of some of the fluctuations in sales and how they are impacted by each of these factors.

Estimating Revenue Impact:It's like a fun math game, plug in some hypotheticals and see what comes out. Many retailers fall into the trap of thinking if they want more sales they need to increase traffic. That's not necessarily so, and furthermore, you might be better off investing in the other two components. Many retailers don't have the resources required to focus on all three areas at once, so how can you tell where to focus first? Plug in some numbers and see how it affects revenue. But be realistic with it, you can't expect to easily double conversion rates or traffic, but look at it in terms of small gains. If you increase conversion rate by 10%, will that work out better than increasing traffic by 10%? And how about that often-overlooked average order amount? Certainly each factor requires different expertise, and might require different levels of time invested on acheiving those gains. But by having a better understanding the "state of", coupled with some experience with the underlying tactics to improving those factors, you can make educated decisions about where to focus your energy on.

C. *Some Common Mistakes*

Below are some of the most common mistakes found on e-commerce websites:

1) *Load Time*
Many websites take a long time before users are even able to see them, never mind start browsing. This is usually due to bad design or because the owner uses a slow server. It is also a crucial factor for the search engines, so make your shop fast.

1) *Browser Compatibility*
Many people who are not experienced at designing or setting up out-of-the-box websites make the mistake of not checking what their creation looks like on various browsers. With several options available, it is essential that you test what your store looks like on each of them. Your website may display fine on some, while on others it could appear broken. It is fundamental that you test your website on as many browsers as possible and make any necessary adjustments to give all your visitors the same quality experience.

2) *Style Over Substance*
The number of websites designed only with style and looks in mind is staggering. This would not be bad, if they were actually easy and pleasant to use, but often the final result is the complete opposite. Easy navigation and general clarity should be a priority way ahead of the visual style. Simplicity is usually the best way forward.

3) *Contact Details*
Make these easy to find. Customers will not feel confident about buying from your website, unless they are sure that they can contact someone with questions or in the case something goes wrong. Do not hide your office number or e-mail, or you'll risk looking untrustworthy.

4) *Product Detail*
The lack of detailed product description is a very common "crime" among online stores. It is as if some retailers have forgotten that the internet is a virtual tool and consumers are not actually able to touch or try on any of the products. That is why they rely on the information you can provide them on the website. Make it comprehensive; put yourself in the position of the shopper and try to be as specific as possible when creating descriptions.

5) *SEO*
Lack of even basic search engine optimisation is very common among online entrepreneurs. Many of them simply don't realise that it is a difficult task to appear high in organic search results. There is a lot of information about SEO available online if you wish to implement it yourself or you may opt to hire an expert to do this work for you. Just remember that good SEO is essential for your store to rank well in the search engines results.

III. SECURITY TOOLS

1) Visible Contact Information: As simple as it sounds, customers feel better being able to put a "face" to a business, rather than just an e-mail. Displaying a phone number at the top of the page or in the "about" section of your site – along with an address linked to Google Maps – helps boost the visitor's confidence and makes your online business feel more personable and real to the shopper instead of a possible scam.

By listing the company's contact information, your customer's will know they have an immediate form of communication should questions or concerns arise about the order and transaction. Even if the phone number you list is your mobile, people will subconsciously feel more secure with an older and familiar form of communication.

2) Customer's Testimonials and Reviews : Showing that other customers have had good experiences with your site and your products not only makes your site more reputable but also increases sales. Services like PowerReviews will generate "tag-based" reviews that will connect with your target market. Though it is one thing to connect with customers as a company, user reviews are an element on your store that connects customer to customer – reviews prove that online shoppers have had an experience with your company and product. This community-like structure is known to be fairly honest and affective throughout social media and e-commerce due to the customer's ability to relate to the reviewer.
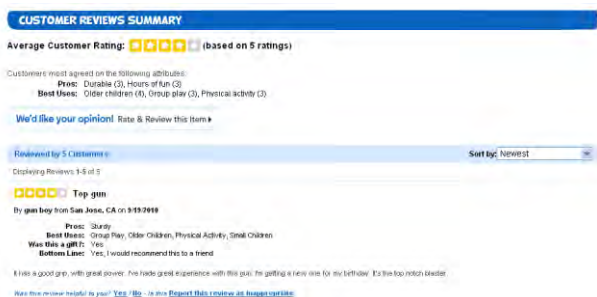

3) Dedicated SSL Certificate : When placing an online transaction securely, a browser will access your site through a Secure Socket Layer connection (SSL). In order to ensure the connection is secure, the web browser will access the site's SSL certificate. When using 3dcart's services, you are given access to a free shared SSL certificate that uses your base *.3dcartstores.com URL for its authentication. Please note however, that the use of a *shared* SSL certificate changes the domain name displayed during checkout. Obtaining a Dedicated SSL Certificate from security companies such as Geotrust, Verisign or Comodo will further bring your customer some peace of mind while checking-out. If possible, get an EVSSL (Extended Validation SSL) since it is the maximum level of security. The Dedicated SSL Certificate not only verifies your identity as a company and proves that the customer's information is encrypted during the process of transmitting sensitive information to complete an online transaction, but it will *also* be assigned specifically to your

domain name. Thereby keeping your base URL the same throughout the whole transaction process.

4) Privacy Policy: Adding a Privacy Policy and associated trust seal to your store and checkout pages is essential to gaining customer confidence and converting visitors into customers. The privacy policy reinforces the fact that you are a dedicated retailer who will keep customer's information secure and private, without sharing or reselling the information to third parties. TRUSTe is one of the most recognized privacy organizations, used by top retailers like eBay, Microsoft and Apple. The presence of their seal has been proven to increase order value by at least $10 and sales by over 29%!
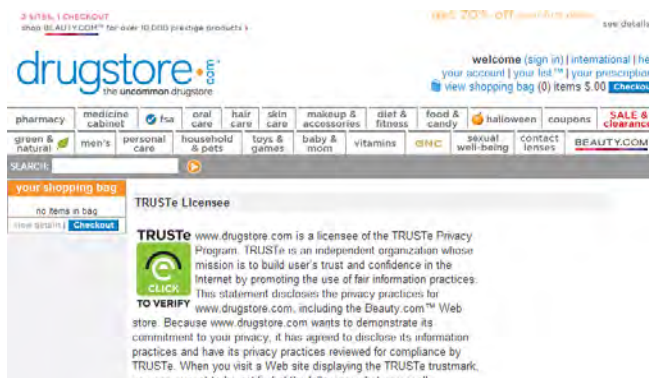
5) Security Seals from Reputable Companies: Displaying a reputable security seal like McAfee Secure or VeriSign TrustSeal conveys the message that your business has gone through a validation process and establishes an association with the entity providing the seal. Some seals illustrate that your company has passed an identity verification process while other may show that a specific, trusted company is handling the payments for your business, In a way, these seals help provide some assurance to first time visitors who are eager to place an order, but may still have doubts regarding the background of the business.

D. Authors and Affiliations

The template is designed so that author affiliations are not repeated each time for multiple authors of the same affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization). This template was designed for two affiliations.

1) For author/s of only one affiliation (Heading 3): To change the default, adjust the template as follows.

a) Selection (Heading 4): Highlight all author and affiliation lines.

b) Change number of columns: Select the Columns icon from the MS Word Standard toolbar and then select "1 Column" from the selection palette.

c) Deletion: Delete the author and affiliation lines for the second affiliation.

2) For author/s of more than two affiliations: To change the default, adjust the template as follows.

a) Selection: Highlight all author and affiliation lines.

b) Change number of columns: Select the "Columns" icon from the MS Word Standard toolbar and then select "1 Column" from the selection palette.

c) Highlight author and affiliation lines of affiliation 1 and copy this selection.

d) Formatting: Insert one hard return immediately after the last character of the last affiliation line. Then paste down the copy of affiliation 1. Repeat as necessary for each additional affiliation.

e) Reassign number of columns: Place your cursor to the right of the last character of the last affiliation line of an even numbered affiliation (e.g., if there are five affiliations, place your cursor at end of fourth affiliation). Drag the cursor up to highlight all of the above author and affiliation lines. Go to Column icon and select "2 Columns". If you have an odd number of affiliations, the final affiliation will be centered on the page; all previous will be in two columns.

IV. INTEGRAL REQUIREMENT

1.Integral requirements
a) privacy – information exchanged must be kept from unauthorized parties
b) integrity – the exchanged information must not be altered or tampered with
c) authentication – both sender and recipient must prove their identities to each other and
d) non-repudiation – proof is required that the exchanged information was indeed received (Holcombe, 2007).
These basic maxims of eCommerce are fundamental to the conduct of secure business online. Further to the fundamental maxims of eCommerce above, eCommerce providers must also protect against a number of different external security threats, most notably Denial of Service (DoS). These are where an attempt is made to make a computer resource unavailable to its intended users though a variety of mechanisms discussed below. The financial services sector still bears the brunt of e-crime, accounting for 72% of all attacks. But the sector that experienced the greatest increase in the number of attacks was eCommerce. Attacks in this sector have risen by 15% from 2006 to 2007 (Symantec, 2007).

2. Privacy

Privacy has become a major concern for consumers with the rise of identity theft and impersonation, and any concern for consumers must be treated as a major concern for eCommerce providers. According to Consumer Reports Money Adviser (Perrotta, 2008), the US Attorney General has announced multiple indictments relating to a massive international security breach involving nine major retailers and more than 40 million credit- and debit-card numbers. US attorneys think that this may be the largest hacking and identity-theft case ever prosecuted by the justice department. Both EU and US legislation at both the federal and state levels mandates certain organizations to inform customers about information uses and disclosures. Such disclosures are typically accomplished through privacy policies, both online and offline (Vail et al., 2008).

In a study by Lauer and Deng (2008), a model is presented linking privacy policy, through trustworthiness, to online trust, and then to customers' loyalty and their willingness to provide truthful information. The model was tested using a sample of 269 responses. The findings suggested that consumers' trust in a company is closely linked with the perception of the company's respect for customer privacy (Lauer and Deng, 2007). Trust in turn is linked to increased customer loyalty that can be manifested through increased purchases, openness to trying new products, and willingness to participate in programs that use additional personal information. Privacy now forms an integral part of any e-commerce strategy and investment in privacy protection has been shown to increase consumer's spend, trustworthiness and loyalty.

The converse of this can be shown to be true when things go wrong. In March 2008, the Irish online jobs board, jobs.ie, was compromised by criminals and users' personal data (in the form of CV's) were taken (Ryan, 2008). Looking at the real-time responses of users to this event on the popular Irish forum, Boards.ie, we can see that privacy is of major concern to users and in the event of their privacy being compromised users become very agitated and there is an overall negative effect on trust in e-commerce. User comments in the forum included: "I'm well p*ssed off about them keeping my CV on the sly"; "I am just angry that this could have happened and to so many people"; "Mine was taken too. How do I terminate my acc with jobs.ie"; "Grr, so annoyed, feel I should report it to the Gardai now" (Boards.ie, 2008).

3. Integrity, Authentication & Non-Repudiation

In any e-commence system the factors of data integrity, customer & client authentication and non-repudiation are critical to the success of any online business. Data integrity is the assurance that data transmitted is consistent and correct, that is, it has not been tampered or altered in any way during transmission. Authentication is a means by which both parties in an online transaction can be confident that they are who

they say they are and non-repudiation is the idea that no party can dispute that an actual event online took place. Proof of data integrity is typically the easiest of these factors to successfully accomplish. A data hash or checksum, such as MD5 or CRC, is usually sufficient to establish that the likelihood of data being undetectably changed is extremely low (Schlaeger and Pernul, 2005). Notwithstanding these security measures, it is still possible to compromise data in transit through techniques such as phishing or man-in- the-middle attacks (Desmedt, 2005). These flaws have led to the need for the development of strong verification and security measurements such as digital signatures and public key infrastructures (PKI).

REFERENCES

[1] Internet Commerce: Digital Models for Business, Lawrence et al, Wiley
[2] Electronic Commerce: A Manager's Guide, Kalakota et al, Addison-Wesley
[3] Frontiers of Electronic Commerce, Kalakota et al, Addison-Wesley
[4] http://www.cs.usyd.edu.au/~bob/Icomm

Smartwatches Innovation or Idiocy

Dr. Vinita Gaikwad
Associate Professor
TIMSCDR

Sowrabh Kukalkar
Student, TIMSCDR

Kanka Hattiangadi
Student, TIMSCDR

Mohamed Aejaz Tazak
Student, TIMSCDR

Abstract **- They say the smartwatch is the obvious progression in the mobile sector. Some market analysts remain unconvinced. Is the smart watch a replacement for your Smartphone or an addition to it? Should there be some crossover functionality, or is the watch best marketed as an addition? Is it a more convenient accessory or an entirely new tool? Does it make anything that much easier? Now with a smart watch, things are kind of going in a different direction. You'll again need multiple chargers and you'll have to monitor multiple batteries. It definitely doesn't replace your Smartphone since most of these modern smart watches need a connection to your Smartphone in order to function properly.**

Index Terms - Typing, One handed usability and safety, Convenience, budget, battery life

I. Introduction

Technology is invading our lives. Every day we perform a multitude of operations using computing devices; we check out bank account balance, we pay parking and we exchange documents with our colleagues and friends. Mobile, social networks and cloud computing are the paradigms that have changed the online user experience; these platforms manage today almost all of the information in the internet, an impressive and priceless amount of data.

So, it seems smartwatches will be swooping in from all directions and while a brand new category is always welcome, this one treads a thin red line. If smartwatches remain just a glorified remote control to your phone, then this whole business model will collapse. A smartwatch will need to carve its own identity, have built-in GPS for true navigation and location services, multiple sensors that track your health above and beyond any fitness tracker, visually log your entire day, look better than any traditional watch to get over the 'nerd alert' problem, have a battery life of about 15 days and make sure that the entire world exists on your wrist. Else it will become a geek-and-nerd domain and die a quick death, just like the first digital watches that came in being heralded as the Holy Grail and were soon being sold by street side vendors by the kilo.
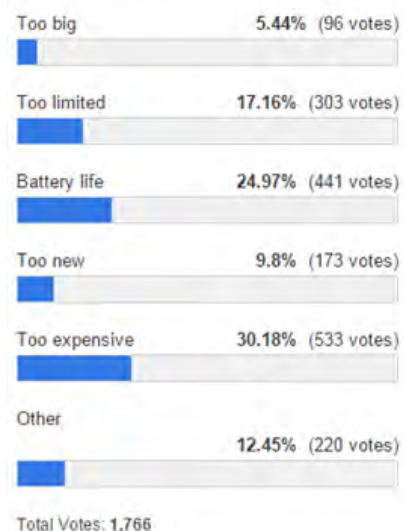
Smartwatches are going to be the next big thing. They'll keep us connected, keep us informed, and make it feel like we're living in the future. Virtually every mobile phone maker has dipped its toes into smartwatch waters, including Apple, which is going to release its much-hyped smartwatch next year. There are also a number of companies whose sole job is to make smart watches, and they've been doing some really interesting things. The reality is that most people probably don't have a need for a smartwatch, at least not yet. But there are some uses for a smartwatch, and if you're already comfortable wearing a traditional watch on your wrist, there are some good reasons to consider upgrading to a smartwatch. A good smartwatch does three things: it makes it easy to see the time, it makes it easy to see notifications that show up on your phone, and it looks good, so you don't feel like you're wearing a computer on your wrist everywhere you go. It should allow you to keep your phone in your bag or pocket and only pull it out for important notifications.

Many smartwatches do extra things too, such as counting your steps, performing web searches, and running a variety of apps. Some even let you do your best Dick Tracy impression, using your voice to order a pizza or send a quick message. But the best smartwatches focus on the core features and put the extra features on the sidelines.

II. Data collected

Why don't you have a smartwatch yet?

Too big	5.44%	(96 votes)
Too limited	17.16%	(303 votes)
Battery life	24.97%	(441 votes)
Too new	9.8%	(173 votes)
Too expensive	30.18%	(533 votes)
Other	12.45%	(220 votes)

Total Votes: 1,766

Low Battery life and its high price contributes the main reason for its low business.

A. Battery life

Some have a legitimate hangup in terms of battery life. This is the hardest one to refuse, because they are not entirely wrong. Smartwatches today will last a day, or a week depending of your flavor. Personally, I plug my phone in

every night when I go to sleep and I can just as easily do the same with my watch. But there are several good counter points, such as forgetting to plug it in, having it get knocked off the charger, or going out of town for a weekend. In any of these cases, your watch is dead by morning, or at least shortly thereafter. If you ask me it's a very minor inconvenience to trade for the great convenience that smartwatches bring, but in this article, I'm asking you, not the other way around.

B. Too expensive

Smartwatches aren't cheap. Some older smartwatches – like the original Pebble for example – can be had for less than $100. But most are priced starting at $200 and up from there. Way, WAY up in some cases. This is a pretty big barrier of entry for the average consumer. Maybe if prices could come down (some already are) to a more comfortable $100 – $150, adoption would be more prevalent.

C. Too big

I know they're not for everyone. Those with slender wrists for example don't have a heck of a lot of options. So maybe you don't want this huge eyesore hanging off your hand. There's a lot of technology that needs to get packed into this frame so it's no surprise that smartwatches tend to be bigger, and therefore more bulky than some would like.

D. Too limited

Others are looking for a device that eliminates the need for a smartphone. They want a watch that will make phone calls, receive texts, and heck why not even watch Netflix (spoiler alert – because it's ridiculous)? These people are looking for the wrong experience. Or at least they're not looking for the experience that today's smartwatches can deliver. Of course in my opinion, they shouldn't deliver that experience, but that's neither here nor there.

III. CONCLUSION

The wrist is valuable accessory space. Watches aren't just watches anymore, if they ever were. They're fashion statements in a way that today's gadgets-ooo, a black rectangle!-aren't. One-look-fits-all isn't going to cut it, and color variations don't count (looking your way Apple).

Too many sizes. In fact, one size doesn't fit all, either! Men and women wear different-sized watches. Beyond that, watch size is another aesthetic choice that people who wear things on their wrists care very much about. How many SKUs are these companies prepared to manufacture?

Which makes for terrible UI problems. Adding a micro-sized display to your platform's lineup is problematic enough on its own for apps. Allowing for displays with multiple degrees of tiny is guaranteed chaos. Scaling horrors, ahoy!

But mostly one giant size. The obvious answer to that would be to stick with just one display size, which iWatch and Samsung and Microsoft rumors currently peg at 1.5-inches. That is a very big thing to lug around all day.

Battery life will be horrid. The sad thing is, it still won't be big enough to fit in all of the internals it needs alongside a battery that's worth a damn. While there's no way a smartwatch will be as battery-intensive as your phone, you're still going to have to plug the damn thing in every few days.

Imagine input on that tiny display. Making your smartwatch do what you want is going to be a royal pain, unless you've got needlepoint fingers, or until voice command technology advances so far beyond where it's at today we might as well be talking about getting a few more feet of lift out of our hoverboards.

How much will these things cost? Oh man, are you really ready to drop $150 on yet another gadget that already does things your preexisting gadgets can? That's how much the Sony SmartWatch runs you. Even if an Apple iWatch manages to match that, it's a whole lot of change.

It's another data eater. Not to mention that a smartwatch that only runs on Wi-Fi would be effectively useless; when you're in your house or a coffee shop or another Wi-Fi accessible location, you're using other devices. So get ready to tack yet another gadget onto your data plan, unless it's pure Bluetooth (there's that battery life again) or some sort of Airplay-like ad hoc wireless hookup with your phone.

REFERENCES

[1] http://www.tomsguide.com/us/best-smartwatches,review-2156.html
[2] http://pocketcalculatorshow.com/nerdwatch/seiko-computer-watch-fun/
[3] http://appleinsider.com/articles/13/03/21/new-samsung-smart-watch-will-be-companys-third-stab-at-wrist-accessory
[4] http://techcrunch.com/2013/09/15/pebbles-eric-migicovsky-is-uninterested-in-a-potential-acquisition/?ncid=tcdaily
[5] http://fortune.com/2014/04/17/why-investors-are-pouring-millions-into-crowdfunding/
[6] http://www.pocket-lint.com/news/121756-acer-we-are-looking-at-wearable-coming-in-2014
[7] http://www.connectedly.com/best-smartwatches-ces-2014

Strategies for Effective Mobile Computing

Dr. Vinita Gaikwad
*Associate Professor,
TIMSCDR*

Shahbaz N. Shaikh
Student, TIMSCDR

Rahul D. Sharma
Student, TIMSCDR

Kishan Mallah
Student, TIMSCDR

***Abstract* -This research paper provides the Strategies of accessibility of windows phones over the android mobile devices. Basically android mobile devices are the largest selling product. Over the period of time the android devices are sharing 40 to 45 percent of market share. The Primary goal of this Study was to inculcate additional features which would increase its usage capability and productivity. The Secondary goal is to evaluate an explanation on how the application works and also describe how the available features and functions can be further improved. As windows phone has the capability to have their brand worldwide because of some drawback . With windows phone 8.1 Microsoft's OS finally became as full featured as its rivals, which along with its stylish, unique design means it's very tempting alternative to iOS or Android. But with Android 4.4.4 KITKAT Google has improved on its OS in numerous ways too and while it's since gone even further with Android 5.0.It's KITKAT which is still running on most android phones. Both Operating systems, have been thoughtfully designed and undergone years of polish to reach the point they're at now, but there are numerous differences between them.**

Index Terms - Application Ecosystem, Customizability, Attractiveness, Limited Multitasking, Limited Accessibility, XBox games, Windows Phone, Consistency, Devices, Offline Maps, Expandable Storage

I. INTRODUCTION

Mobile operating system have came a long way since the first generation of apple's iOS, back in 2007. With major advancements in hardware, and reduced cost of processors and sensors, the possibilities for mobile software have really opened up, and software developers have been quick to take advantage of these new-found powers. So what does that spell for shoppers in the market for a new phone?

I grew up tinkering with PCs and love having control over every aspect of my device, and Android is flexible enough to accommodate my affinity or customization option and would rather have an easy-to-use touch screen device that lets you chat and take great photos – Why not consider a Windows Phone device?[6]

A. iOS

iOS (originally iPhone OS) is a mobile operating system developed by Apple Inc. and distributed exclusively for Apple hardware. It is the operating system that presently powers all of the company's iDevices.

Originally unveiled in 2007 for the iPhone, it has been extended to support other Apple devices such as the iPod Touch , iPad , iPad Mini and second-generation Apple TV onward. As of January 2015, Apple's App Store contained more than 1.4 million iOS applications, 725,000 of which are native for iPad. These mobile app shave collectively been downloaded more than 75 billion

times. It had a 21% share of the Smartphone mobile operating system units. The user interface of iOS is based on the concept of direct manipulation, using multi-touch gestures. Interface control elements consist of sliders, switches, and buttons. Interaction with the OS includes gestures such as swipe, tap, pinch, and reverse pinch, all of which have specific definitions within the context of the iOS operating system and its multi-touch interface. Internal accelerometers are used by some applications to respond to shaking the device or rotating it in three dimensions.

iOS shares with OS X some frameworks such as Core Foundation and Foundation; however, its UI toolkit is Cocoa Touch rather than OS X's Cocoa, so that it provides the UIKit framework rather than the AppKit framework. It is therefore not compatible with OS X for applications. Also while iOS also shares the Darwin foundation with OS X, Unix-like shell access is not available for users and restricted for apps, making iOS not fully Unix-compatible either.

B. Android

Android owns the lion's share when it comes to Smartphone in India because of its availability across several brands and because it meets the needs of a wide range of users. From gaming to multimedia entertainment to photography to everyday calling messaging, Android does it all, while offering users a tremendous control and messaging, Android does it all, while offering users a tremendous control over their experience. Want a simple phone-and-SMS only experience? Prefer a custom lock screen and custom lock screen and launcher with information widgets? Need an enhanced camera app? Android's got you covered, with apps and setting for all of the above.[1]

C. Windows Phone

Currently at version 8.1, Windows Phone includes a novel interface that first-timers can get used to in a pinch, and allow quick access to message, contacts and content you care about right +on the home screen. It cleverly combines widgets with icons in its information live Tiles, and most apps generally follow a unified interface so there are no surprises when you're sending a message, uploading a file or browsing the web.

II. WINDOWS PHONE USABILITY

✓ Customizable home screen, and real-time, in-app updates (live Tiles)
✓ Integration with Microsoft's core-business apps
✓ Consistent GUI and APIs across all platforms

✓ Power efficiency through cross-platform OS

A. *New start screen*

Microsoft's redesigned start screen (pictured above) spans the width of the screen. You'll be able to resize each individual live tile on the start screen to one of three sizes. There's also support for new color themes, which differ slightly by carrier and manufacturer.

There's also support for in-app purchases, and there's a "wallet" for linking credit cards and storing loyalty cards. Phones get over-the-air OS updates, instead of needing tethering to the desktop to update.

In architecture Windows Phone 8 heavily overlaps with Windows 8 for desktops and tablets, so there will be a great deal of similarity in the way that the two operating systems handle security, gaming, networking, media playback, and so forth. Being able to sync apps like Office 2013 documents and notes, music, and photos is easy and useful.

B. *User interface*

Windows Phone features a user interface based on Microsoft's "Metro" design language, and was inspired by the user interface in the Zune HD. The home screen, called the "Start screen", is made up of "Live Tiles", which have been the inspiration for the Windows 8 live tiles. Tiles are links to applications, features, functions and individual items (such as contacts, web pages, applications or media items). Users can add, rearrange, or remove tiles. Tiles are dynamic and update in real time – for example, the tile for an email account would display the number of unread messages or a tile could display a live update of the weather. Since Windows Phone 8, live tiles can also be resized to either a small, medium, or large appearance.

Several features of Windows Phone are organized into "**hubs**", which combine local and online content via Windows Phone's integration with popular social networks such as Facebook, Windows Live, and Twitter. For example, the Pictures hub shows photos captured with the device's camera and the user's Facebook photo albums, and the People hub shows contacts aggregated from multiple sources including Windows Live, Facebook, and Gmail. From the hub, users can directly comment and 'like' on social network updates. The other built-in hubs are Xbox Music and Video, Xbox Live Games, Windows Phone Store, and Microsoft Office.

Windows Phone uses multi-touch technology. The default Windows Phone user interface has a dark theme that prolongs battery life on OLED screens as fully black pixels do not emit light. Alternatively, users may choose a light theme in their phone's settings menu. Light theme has white background which is more energy efficient on LCD as white pixels are actually transparent and do not required energy. The user may also choose from several accent colors. User interface elements such as links, buttons and tiles are shown in the user's chosen accent color. Third-party applications can be automatically themed with these colors. Windows Phone 8.1 introduces transparent tiles and a customizable background image for the Start screen. The image is visible trough the transparent area of the tiles and features a parallax effect when scrolling which gives an illusion of depth. If the user does not pick a background image the tiles render with the accent color of the theme.[1]

C. *Text input*

Users input text by using an on-screen virtual keyboard, which has a dedicated key for inserting emoticons, and features spell checking and word prediction. App developers (both in house and ISV) may specify different versions of the virtual keyboard in order to limit users to certain character sets, such as numeric characters alone. Users may change a word after it has been typed by tapping the word, which will invoke a list of similar words. Pressing and holding certain keys will reveal similar characters. The keys are somewhat larger and spaced farther apart when in landscape mode. Phones may also be made with a hardware keyboard for text input. Users can also add accents to letters by holding on an individual letter. Windows Phone 8.1 introduces a new method of typing by swiping trough the keyboard without lifting the finger, in a manner similar toSwype and SwiftKey.

D. *Web browser*

Internet Explorer on Windows Phone allows the user to maintain a list of favorite web pages and tiles linking to web pages on the Start screen. The browser supports up to 6 tabs, which can all load in parallel. Other features include multi-touch gestures, smooth zoom in/out animations, the ability to save pictures that are on web pages, share web pages via email, and support for inline search which allows the user to search for a word or phrase in a web page by typing it. Tabs are synced with Windows 8.1 devices using Internet Explorer 11.

E. *Contacts*

Contacts are organized via the "People hub", and can be manually entered into contacts or imported from Facebook, Windows Live Contacts, Twitter, LinkedIn, Google, and Outlook. A "What's New" section shows a user's Face book news feed and a "Pictures" section show pictures from those social networks, while a "Me" section within the "People" hub shows a user's own social network status and wall and allows them to view social network updates. Contacts can also be pinned to the Start Screen. The contact's "Live Tile" displays their social network status and profile picture on the home screen. Clicking on a contact's tile or accessing their card within the "People" hub will reveal their recent social network activity as well as the rest of their contact information.

If a contact has information stored on multiple networks, users can link the two separate contact accounts, allowing the information to be viewed and accessed from a single card. As of Windows Phone 7.5, contacts can also be sorted into "Groups". Here, information from each of the contacts is combined into a single page which can be accessed directly from the Hub or pinned to the Start screen.

F. *Email*

Windows Phone supports Outlook.com, Exchange, Yahoo! Mail and Gmail natively and supports many other services via the POP and IMAP protocols. Updates added support

for more services such as iCloud and IBM Notes Traveler. Contacts and calendars may be synced from these services as well. Users can also search through their email by searching in the subject, body, senders, and receivers. Emails are shown with threads, and multiple email inboxes can be combined into a single view (a feature commonly referred to as "combined inbox") or can viewed separately.

G. Multimedia

Xbox Music and Xbox Video are built-in multimedia hubs providing entertainment and synchronization capabilities between PC, Windows Phone, and other Microsoft products. The two hubs were previously combined until standalone apps were released in late 2013, shortly before Windows Phone 8.1 debuted. The hubs allow users to access music, videos, and podcasts stored on the device, and links directly to the "Xbox Music Store" to buy or rent music and the "Xbox Video Store" to purchase movies and TV episodes. Xbox Music also allows the user to stream music with an Xbox Music Pass. When browsing the music by a particular artist, users are able to view artist biographies and photos. The Xbox Music hub also integrates with many other apps that provide video and music services, including, but not limited to, iHeartRadio, YouTube, and Vevo. This hub also includes Smart DJ which compiles a playlist of songs stored on the phone similar to the song or artist selected.

The Pictures hub displays the user's Face book and OneDrive photo albums, as well as photos taken with the phone's built-in camera. Users can also upload photos to social networks, comment on photos uploaded by other people, and tag photos posted to social networks. Multi-touch gestures permit zooming in and out of photos.

An official file manager app called Files, which is available for download from the Windows Phone Store, enables users to move and rearrange documents, videos, music and other files within their device's hard drive or to an external SD card.

H. Media support

Windows Phone supports WAV, MP3, WMA, AMR, AAC/MP4/M4A/M4B and 3GP/3G2 standards. The video file formats supported on WP include WMV, AVI, MP4/M4V, 3GP/3G2 and MOV
(QuickTime) standards. These supported audio and video formats would be dependent on the codes contained inside them. It has also been previously reported that the DivX and Xvid codes within the AVI file format are also playable on WP devices.[2]

Windows Phone does not support DRM protected media files that are obtained from services other than Xbox Music Pass.

The image file formats that are supported include JPG/JPEG, PNG, GIF, TIF and Bitmap (BMP).

Users can also add custom ringtones which are less than 1MB in size and less than 40 seconds long. DLNA streaming and stereoscopic 3D are also supported.

I. Games

The "Games hub" provides access to games on a phone along with Xbox Live functionality, including the ability for a user to interact with their avatar, view and edit their profile, see their achievements and view leader boards, and send messages to friends on Xbox Live. The hub also features an area for managing invitations and turn notifications in turn-based multiplayer games. Games are downloaded from Windows Phone Store.

J. Search

Bing is the default search engine on Windows Phone handsets because its functions are deeply integrated in the OS (which also include the utilization of its map service for location-based searches and queries). However, Microsoft has stated that other search engine applications can be used.

In the area of location-based searches, Bing Maps (which is powered by Nokia's location services) provides turn-by-turn navigation service to Windows Phone users, and Local Scout shows interest points such as attractions and restaurants in the nearby area. On Nokia devices, Nokia's Here Maps is preinstalled in place of Bing Maps. Furthermore, Bing Audio allows the user to match a song with its name, and Bing Vision allows the user to scan barcodes, QR codes, and other types of tags.

K. Cortana

Every Windows Phone has either a dedicated physical Search button or an on-screen Search button, which was previously reserved for a Bing Search app, but has been replaced on Windows Phone 8.1 devices in the United Kingdom and United States by Cortana, a digital personal assistant which can also double as an app for basic searches. Cortana allows users to do tasks such as set calendar reminders and alarms, and recognizes a user's natural voice, and can be used to answer questions (like current weather conditions, sports scores, and biographies). The app also keeps a "Notebook" to learn a user's behavior over time and tailor reminders for them. Users can edit the "Notebook" to keep information from Cortana or reveal more about themselves.

L. Office suite

All Windows Phones come preinstalled with Microsoft Office Mobile, which provides interoperability between Windows Phone and the desktop version of Microsoft Office. Word Mobile, Excel Mobile, PowerPoint Mobile, and SharePoint Workspace Mobile apps are accessible through a single "Office Hub," and allow most Microsoft Office file formats to be viewed and edited directly on a Windows Phone device. The "Office Hub" can access files from OneDrive and Office 365, as well as files which are stored locally on the device's hard drive. Although they are not preinstalled in Windows Phone's "Office Hub," OneNote Mobile, Lync Mobile, and OneDrive for Business can be downloaded separately as standalone applications from the Windows Phone Store.[2]

M. Multitasking

Multitasking in Windows Phone is invoked through long pressing the "back" arrow, which is present on all Windows Phones. Windows Phone 7 uses a card-based task switcher, whereas later versions of Windows Phone utilize true background multitasking.[2]

IV. LIMITATIONS OF WINDOWS PHONE

Windows phones beautiful start screen with tiles, needs apps to fill that space as a background . The biggest downfall currently is application ecosystem of Microsoft, even after four years its initial release Many popular applications including are still missing. In recent months, We have seen the popular arrival of application such as tumblr, Pintrest,. Games are a similar story, While we have a collection of great games, many of top quality games we see on other platforms take months to get to Windows Phone.

A. Lack of customizability

You can rearrange your live Tiles, select your new favorite color, and maybe even add a background – that is where customizing windows phones start screen and operating system stops. If you don't like an aspect of the phone, say the keyboard, there are no options to swap it out for another. Microsoft has made sure that core system component can't be changed, and while this may not be an issue for some, many of us have become accustomed to swapping in our favorite keyboard, dialer, or SMS app.

B. Lack of attractiveness

There at least 20 colors you can choose from, including primary standards like Red, cobalt(blue),and yellow, as well as more unique shades like Emerald, Magenta, and Steel. Unlike the android devices there is no such concept of wallpaper. So, the limited colors are available for limited numbers of Apps. You can only choose between black and white for your phone's background. You can only have two views of apps one is tiles view and second one is List view. Limitations in Tiles availability some of the apps only available in the list view. Search of apps in vast list so complicated. Unlike the android devices the Theme concept are not available.

C. Limited multitasking

Microsoft introduced multitasking with its Windows Phone 7 'Mango' launch. If you are like me, you have been happily using the feature and enjoying it, but if you are like me, you look outside the Microsoft ecosystem, you find other systems are doing the same job better.

To start, Windows Phone place a limit on the number of application you can run in the background. Other operating system, such as Android and IOS, don't have such a swallow limit. You might find that while can be an issue if it was something you sincerely needed.

With the rise of phablets, we have also seen manufactures, such as Sony and LG, adding multi-app viewing to their device. A Samsung Galaxy user can easily 'snap' two app side by side, and I find that it does come in useful while answering emails and trying to jot down notes at the same time.

D. Limited accessibility

Windows phones 8.1 dramatically improve how Microsoft's smart phone platform deals with notification from the OS and from apps. Now, in addition to live title update and banner notifications, the system includes a new notification manager called Action Center.

First of all the people for share purpose most probably use Bluetooth. The Bluetooth notifications are not removed after the Bluetooth gets off. The major issue for a user is they have to accept all the files request one by one.

FUTURE SCOPE / RECOMMENDATIONS

A. Notifications and settings

One swipe down from the top of your screen opens action centre, where you can see notifications and quickly turn on WiFi, Bluetooth and more. The new volume control gives you separate sliders for your ringer and apps. You can back up more stuff to the cloud. And Storage Sense keeps you in control of your storage space – see what's taking up space, set your SD card as the default location for installing apps and more.

B. Better browsing

Internet Explorer includes inPrivate browsing, reading mode and new settings to help conserve mobile data as you browse. Plus, your tabs can now sync to your PC and Windows tablet – just sign in everywhere with your Microsoft account and you're good to go!
But we also suggest that's the Microsoft add some feature of offline reading and after closing of the browser the history may be deleted.

C. Naming of the tiles

The feature called Tiles in Windows Phone Home Screen shows the Apps in the tile format. The resizing of the Tiles are possible in Windows Phone 8.1 But the Renaming of tiles are not possible there as well.
Suggestion is that Microsoft should put Rename feature and as well the folder feature for grouping up the apps.

VI. CONCLUSION

While android offers greater app flexibility, Windows Phone offers great potential, Better integration over more platforms and fluidity. With the changing smart phone market, the world needs a solid platform that has a quality and the potential to develop into a full range solution, that spans across mobile, desktop and the cloud. The strategies are already in place, and if Nokia and Windows phone can weather the storm, they may one day be able to unset Android form its iron grip of the smart phone OS market.

REFERENCES

[1] http://www.pcadvisor.co.uk/buying-advice/mobile-phone/3414135/which-is-better-android-or-windows-phone/
[2] http://www.usability247.com/blog/mobile-usability-android-vs-iphone-vs-windows-infographic/
[3] http://www.hongkiat.com/blog/windows-phone-better-than-android/
[4] https://www.google.co.in/webhp?sourceid=chrome-instant&ion=1&espv=2&ie=UTF-8#q=windows+phone+uses
[5] http://www.winbeta.org/news/top-five-reasons-avoid-windows-phone-and-keep-your-androidios-device
[6]http://www.timesofindia.indiatimes.com

Study of Concurrency Control in Distributed Database System

Shirshendu Maitra
Assistant Professor
TIMSCDR

Amar Barge
Student,
TIMSCDR

Sonal Gupta
Student,
TIMSCDR

Mayuri Natekar
Student,
TIMSCDR

Pooja Jadhav
Student,
TIMSCDR

Mayank Jain
Student,
TIMSCDR

Abstract—**This electronic document explains about how concurrency control is managed in Distributed Database Management System and how it can be achieved**

Index Terms- concurrency, fragmentation, transaction, security.

I. INTRODUCTION

With the rising need for secure, reliable and accessible information in today's technology environment, the need for distributed databases and client/server applications is also increasing. A distributed database is a single logical database that is spread physically across computers in multiple locations that are connected by data network links. Distributed database is a kind of virtual database whose component parts are physically stored in a number of distinct real databases at a number of distinct locations. The users at any location can access data at anywhere in the network as if the data were all stored at the user's own location. The objective of a distributed database management system (DDBMS) is to control the management of a distributed database (DDB) in such a way that it appears to the user as a centralized database.

II. MOTIVATION

There are various conditions that encourage the use of distributed databases:

1) Data network costs and reliability: If the data is distributed and the applications are related to these data, it may be much more inefficient, in terms of network costs, to partition the data and do the processing at each site. On a different hand, the cost of having smaller computing powers at each site is much more less than the cost of having an equivalent power of a single processor.

2) Database recovery: Replicating data on separate computers is one master plan for ensuring that a damaged database can be quickly recovered and users can have access to data while the primary site is being rebuild. Replicating data across various computer sites is one natural form of a distributed database.

3) Data sharing: Even to a certain extent complex business decisions require sharing data across business units, so it must be convenient to consolidate data across local databases on demand.

4) Data retrieved by a transaction may be stored at a various sites, making it possible to execute the transaction in parallel.

5) Increased reliability and availability: When a centralized system fails, the database is inaccessible to all users. In contrast to centralized systems, distributed system will carry on to function at some degrade level, however, even when a component fails.

6) Faster response: This speeds up query processing since communication and central computer delays are minimized.

III. DISTRIBUTED DATABASE DESIGN

Distributed Database Systems are needed for the applications where data and its accesses are inherently distributed and to increase the availability during failures. However, some additional factors have been considered for a Distributed Database:

A. Data Fragmentation:

In Distributed Databases, we need to specify and understand the logical unit of Database Distribution and allocation. The database may be broken up into logical units called fragments which will be stored at various sites. The easiest logical units are the tables themselves. Two Types of Data Fragmentation are as follows.

1) Horizontal fragmentation: A horizontal fragment of a table is a subset of rows in it. So horizontal fragmentation spilts a table 'horizontally' by selecting the relevant rows and these fragments can be assigned to different sides in the distributed system.

2) Vertical fragmentation: *A vertical fragment of a table keeps only some attributes of it. It spilts a table vertically by columns. It is essential to include the primary key of the table in each vertical fragment so that the full table can bere constructed if needed.*

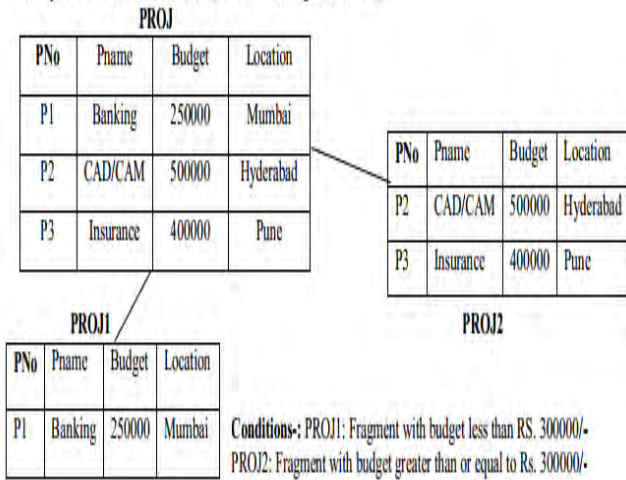**Example:** Consider a relation PROJ (PNo, Pname, Bugdet, Location)

PROJ

PNo	Pname	Budget	Location
P1	Banking	250000	Mumbai
P2	CAD/CAM	500000	Hyderabad
P3	Insurance	400000	Pune

PNo	Pname	Budget	Location
P2	CAD/CAM	500000	Hyderabad
P3	Insurance	400000	Pune

PROJ1

PNo	Pname	Budget	Location
P1	Banking	250000	Mumbai

PROJ2

Conditions:- PROJ1: Fragment with budget less than RS. 300000/-
PROJ2: Fragment with budget greater than or equal to Rs. 300000/-

Fig. 1 Horizontal Fragmentation

Example: Consider a Relation PROJ (PNo, Pname, Budget, Location)

PROJ

PNo	Pname	Budget	Location
P1	Banking	250000	Mumbai
P2	CAD/CAM	500000	Hyderabad
P3	Insurance	400000	Pune

PROJ1

PNo	Budget
P1	250000
P2	500000
P3	400000

PROJ2

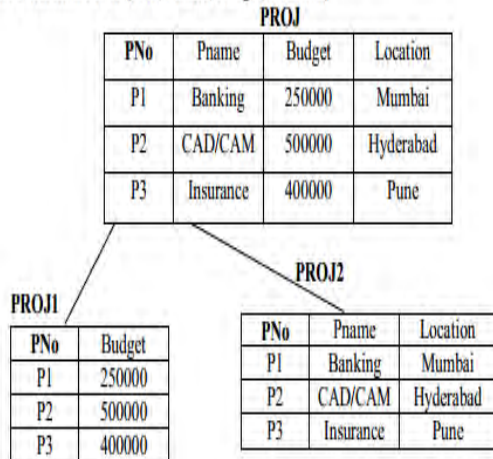| PNo	Pname	Location
P1	Banking	Mumbai
P2	CAD/CAM	Hyderabad
P3	Insurance	Pune

Fig. 2 Vertical Fragmentation

B. Data Replication

A popular option for data distribution as well as for fault tolerance of a database is to store a separate copy of the database at each of two or more sites.A replicate of each fragment can be maintained at several sites. Data replication is the process of deciding which fragments will be replicated.

IV. FUNDAMENTALS OF TRANSACTION MANAGEMENT AND CONCURRENCY CONTROL

Transaction: A transaction consists of a series of operations performed on a database. The major issue in transaction management is that if a database was in a consistent state prior to the initiation of a transaction, then the database should return to a consistent state after the transaction is completed.

Properties of Transaction: A Transaction has four properties that lead to the consistency and reliability of a distributed database. These are Atomicity, Consistency, Isolation, and Durability.

1) *Atomicity:* This refers to the fact that a transaction is treated as a unit of operation. It dictates that either all the actions related to a transaction are completed or none of them is carried out.
2) *Consistency:* The consistency of a transaction is its correctness. In different words, a transaction is a correct program that maps one consistent database state into another.
3) *Isolation:* According to Isolation, each transaction should see a consistent database at all times. Consequently, no other transaction can read or modify data that is being modified by another transaction.
4) *Durability:* This property ensures that as a transaction commits, its results are immutable and cannot be taken away from the database. This means that whatever occurs after the COMMIT of a transaction, whether it is a processor crash or aborts of other transactions, the results already committed are not modified or undone.

Concurrency Control: In distributed database systems, database is typically used by various users. These systems usually allow multiple transactions to run concurrently. Concurrency control is the activity of coordinating concurrent accesses to a database in a multiuser database management system (DBMS). Concurrency control permits users to access a database in a multi-programmed fashion while preserving the illusion that each user is executing alone on a dedicated system. The major technical difficulty in attaining this goal is to prevent database updates performed by one user from interfering with database retrievals and updates performed by another. When the transactions are updating data concurrently, it may lead to various problems with the consistency of the data.

V. DISTRIBUTED CONCURRENCY CONTROL ALGORITHMS

In this paper, we consider some of the distributed concurrency control algorithms. We outline the salient aspects of these four algorithms in this section. In order to do so, we must first explain the structure that we have assumed for distributed transactions. Before understanding the algorithms, we need to know about the distributed transactions. *Distributed Transaction:* A distributed transaction is a transaction that runs in various processes, usually on different machines. Each process works for the transaction. A distributed transaction processing system follows the ACID properties:

1) Recoverable processes. Recoverable processes log their actions and therefore can restore earlier states if a failure occurs.

2) A commit protocol. A commit protocol allows multiple processes to coordinate the committing or aborting of a transaction. The most important commit protocol is the two-phase commit protocol.

A. Distributed Two-Phase Locking (2PL):

In order to ensure serializability of parallel executed transactions elaborated different methods of concurrency control. One of these methods is locking method. There are various forms of locking method. Two phase locking protocol is the basic concurrency control protocols in distributed database systems. The important approach of this protocol is "read any, write all". Transactions set read locks on items that they read, and they change their read locks to write locks on items that need to be updated. To read an item, it meets the need to set a read lock on any replicate of the item, so the local copy is locked; to update an item, write locks are required on all replicates. Write locks are obtained as the transaction runs, with the transaction blocking on a write request until all of the copies of the item to be updated have been successfully locked. All locks are intact until the transaction has successfully committed or aborted. The 2PL Protocol oversees locks by determining when transactions can acquire and release locks. The 2PL protocol forces each transaction to make a lock or unlock request in two steps:

1) Growing Phase: A transaction may obtain locks but may not release any locks.

2) Shrinking Phase: A transaction may release locks but not obtain any new lock.

The transaction first enters into the Growing Phase, makes appeal for required locks, then goes into the Shrinking phase where it releases all locks and cannot make any more appeals. Transactions in 2PL Protocol should get all needed locks before getting into the unlock phase. Although the 2PL protocol guarantees serializability, it does not assure that deadlocks do not happen.

B. Wound-Wait (WW):

The second algorithm is the distributed wound-wait locking algorithm. It follows the similar approach as the 2 PL protocol. The difference lies in the fact that it differs from 2PL in its handling of the deadlock problem: unlike 2PL protocol, rather than maintaining waits-for information and then checking for local and global deadlocks, deadlocks are prevented through the use of timestamps in this algorithm. Each transaction is enumerated according to its initial startup time, and newer transactions are controlled from making older ones wait. If an older transaction appeals a lock, and if the appeal would lead to the older transaction waiting for a younger transaction, the younger transaction is "wounded" – it is restarted unless it is already in the second phase of its commit protocol. Younger transactions can hold on for older transactions so that the possibility of deadlocks is eliminated.

C. Basic Timestamp Ordering (BTO):

A timestamp is a unique identifier created by the DBMS to identify a transaction. Typically, timestamp values are assigned in the order in which the transactions are submitted to the system, so a timestamp can be thought of as the transaction start time. The third algorithm is the basic timestamp ordering algorithm. The idea for this scheme is to order the transactions based on their timestamps. A schedule in which the transactions participate is then serializable, and the equivalent serial schedule has the transactions in order of their timestamp values. This is called timestamp ordering (TO).

D. Distributed Optimistic(OPT):

The fourth algorithm is the distributed, timestamp-based, optimistic concurrency control algorithm. which operates by exchanging certification information during the commit protocol. For each data item, a read timestamp and a write timestamp are maintained. Transactions may read and update data items freely, storing any updates into a local workspace until commit time. For each read, the transaction must remember the version identifier (i.e., write timestamp) associated with the item when it was read. Then, when all of the transaction's cohorts have completed their work, and have reported back to the master, the transaction is assigned a globally unique timestamp.

VI. CONCLUSION

In this paper we have discuss about the distributed database system that is considered to be more reliable than centralized database system. We also explain the concurrency control algorithms-: distributed 2PL and wound-wait algorithm. It is really essential for database to have the ACID properties to perform.

REFERENCES

[1] Ambler, Scott. Introduction to Concurrency Control, 2006
[2] Ambler, Scott. Introduction to Concurrency Control, 2006
[3] Arun Kumar Yadav& Ajay Agarwal, An Approach for Concurrency Control in Distributed Database System, Vol. 1, No. 1, pp. 137-141, January-June (2010)
[4] Coronel, Carlos, Peter Rob. Database Systems, sixth ed. Thomson Course Technology, 2004.
[5] "Database System Concepts" by Abraham Silberschatz, Henry Korth, and S. Sudarshan
[6] "Database Management Systems" by Raghu Ramakrishnan
[7] Fundamentals of DBMS, Lakhanpal Publisher, III edition (2008)
[8] Gupta V.K., Sheetlani Jitendra, Gupta Dhiraj and Shukla
[9] Brahma Datta, Concurrency control and Security issues in Distributed database system, Vol. 1(2),70-73, August (2012)
[10] Kumar, V. Transaction Management Concurrency Control Mechanisms, 2012
[11] Navathe Elmasri, Database Concepts, Pearson Education, V edition (2008)
[12] Swati Gupta, Kuntal Saroha, Bhawna, Fundamental Research of Distributed Database, IJCSMS International Journal of Computer Science and Management Studies, Vol. 11, Issue 02, Aug 2011.

Working of GOOGLE Search Engine

Rashmi Vipat
Assistant Professor
TIMSCDR

Ami Shah
Student, TIMSCDR

Palak Shah
Student, TIMSCDR

Priya Sharma
Student, TIMSCDR

Abstract - **Google's search engine is a powerful tool. Without search engines like Google, it would be practically impossible to find the information you need when you browse the Web. Like all search engines, Google uses a special algorithms to generate search results. Google uses automated programs called spiders or crawlers, just like most search engines. Also like other search engines, Google has a large index of keywords and where those words can be found. What sets Google apart is how it ranks search results, which in turn determines the order Google displays results on its search engine results page (SERP). The prototype with a full text and hyperlink database of at least 24 million pages is available at http://google.stanford.edu/ . In this research paper, we are going to present how Google search engine works. Apart from the problems of scaling traditional search techniques to data, there are new challenges involved with using the additional information present in hypertext to produce better search results.**

Indexed terms - World Wide Web, Crawlers, Spiders, Information Retrieval, Page Rank, Google

I. INTRODUCTION

Google Search, commonly referred to as Google Web Search or just Google, is a web search engine owned by Google Inc. It is the most-used search engine on the World Wide Web, handling more than three billion searches each day. As of February 2015 it is the most used search engine in the US with 64.5% market share.

The main purpose of Google Search is to hunt for text in publicly accessible documents offered by web servers, as opposed to other data, such as images or data contained in databases. It was originally developed by Larry Page and Sergey Brin in 1997. Google Search provides several features beyond searching for words. These include synonyms, weather forecasts, time zones, stock quotes, maps, earthquake data, movie show times, airports, home listings, and sports scores. There are special features for numbers, dates, and some specific forms, including ranges, prices, temperatures, money and measurement unit conversions, calculations, package tracking, patents, area codes, and language translation. In June 2011 Google introduced "Google Voice Search" to search for spoken, rather than typed, words. In May 2012 Google introduced a Knowledge Graph semantic search feature in the U.S.

Google Architecture Overview: In this section, we will give a high level overview of how the whole system works as pictured in Figure 1. Further sections will discuss the applications and data structures not mentioned in this section. Most of Google is implemented in C or C++ for efficiency and can run in either Solaris or Linux. In Google,

the web crawling (downloading of web pages) is done by several distributed crawlers. There is a URLserver that sends lists of URLs to be fetched to the crawlers. The web pages that are fetched are then sent to the storeserver. The storeserver then compresses and stores the web pages into a repository. Every web page has an associated ID number called a docID which is assigned whenever a new URL is parsed out of a web page. The indexing function is performed by the indexer and the sorter. The indexer performs a number of functions. It reads the repository, uncompresses the documents, and parses them. Each document is converted into a set of word occurrences called hits. The hits record the word, position in document, an approximation of font size, and capitalization. The indexer distributes these hits into a set of "barrels", creating a partially sorted forward index. The indexer performs another important function. It parses out all the links in every web page and stores important information about them in an anchors file. This file contains enough information to determine where each link points from and to, and the text of the link. The URLresolver reads the anchors file and converts relative URLs into absolute URLs and in turn into docIDs. It puts the anchor text into the forward index, associated with the docID that the anchor points to. It also generates a database of links which are pairs of docIDs. The links database is used to compute PageRanks for all the documents. The sorter takes the barrels, which are sorted by docID (this is a simplification, see Section 4.2.5), and resorts them by wordID to generate the inverted index. This is done in place so that little temporary space is needed for this operation. The sorter also produces a list of wordIDs and offsets into the inverted index. A program called DumpLexicon takes this list together with the lexicon produced by the indexer and generates a new lexicon to be used by the searcher. The searcher is run by a web server and uses the lexicon built by DumpLexicon together with the inverted index and the PageRanks to answer queries. It happens billions of times a day in the blink of an eye.

Search engines have two major functions:
  i.   Crawling and Indexing
  ii.   Algorithms
  iii. Fighting Spam

II. CRAWLING AND INDEXING

The journey of a query starts before you ever type a search, with crawling and indexing the web of trillions of documents. Google uses software known as "web crawlers" to discover publicly available web pages.
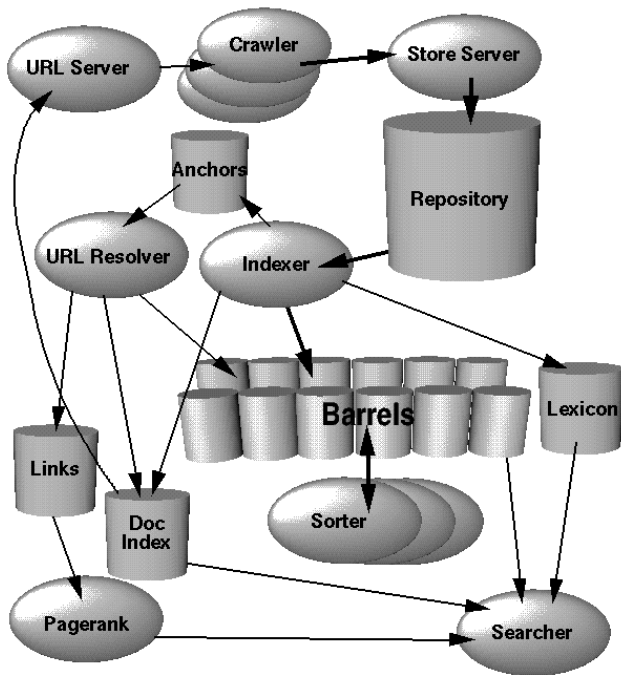
The most well-known crawler is called "Googlebot." Crawlers look at web pages and follow links on those pages, much like you would if you were browsing content on the web. They go from link to link and bring data about those web pages back to Google's servers.
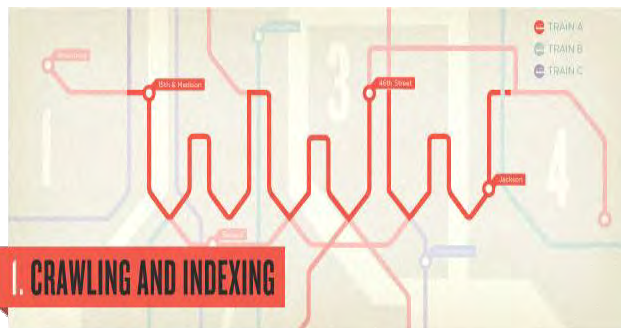
The crawl process begins with a list of web addresses from past crawls and sitemaps provided by website owners. As the crawlers visit these websites, they look for links for other pages to visit. The software pays special attention to new sites, changes to existing sites and dead links. Computer programs determine which sites to crawl, how often, and how many pages to fetch from each site. Google doesn't accept payment to crawl a site more frequently for our web search results. It cares more about having the best possible results because in the long run that's what's best for users and, therefore, our business.

Most websites don't need to set up restrictions for crawling, indexing or serving, so their pages are eligible to appear in search results without having to do any extra work. That said, site owners have many choices about how Google crawls and indexes their sites through Webmaster Tools and a file called "robots.txt". With the robots.txt file, site owners can choose not to be crawled by Googlebot, or

they can provide more specific instructions about how to process pages on their sites.Site owners have granular choices and can choose how content is indexed on a page-by-page basis. For example, they can opt to have their pages appear without a snippet (the summary of the page shown below the title in search results) or a cached version (an alternate version stored on Google's servers in case the live page is unavailable). Webmasters can also choose to integrate search into their own pages with Custom Search.

*Organizing Information by Indexing:*The web is like an ever-growing public library with billions of books and no central filing system. Google essentially gathers the pages during the crawl process and then creates an index, so it knows exactly how to look things up. Much like the index in the back of a book, the Google index includes information about words and their locations. When you search, at the most basic level, algorithms look up your search terms in the index to find the appropriate pages. The search process gets much more complex from there. When you search for "dogs" you don't want a page with the word "dogs" on it hundreds of times. You probably want pictures, videos or a list of breeds. Google's indexing systems note many different aspects of pages, such as when they were published, whether they contain pictures and videos, and much more. With the Knowledge Graph, continuing to go beyond keyword matching to better understand the people, places and things you care about.

III. ALGORITHMS

For a typical query, there are thousands, if not millions, of webpages with helpful information. Algorithms are the computer processes and formulas that take your questions and turn them into answers. Today Google's algorithms rely on more than 200 unique signals or "clues" that make it possible to guess what you might really be looking for. These signals include things like the terms on websites, the freshness of content, your region and Page Rank. When Google was a Stanford research project, it was nicknamed BackRub because the technology checks backlinks to determine a site's importance. Previous keyword-based methods of ranking search results, used by many search engines that were once more popular than Google, would rank pages by how often the search terms occurred in the page, or how strongly associated the search terms were within each resulting page. The PageRank algorithm instead analyzes human-generated links assuming that web pages linked from many important pages are themselves likely to be important. The algorithm computes a recursive score for pages, based on the weighted sum of the PageRanks of the pages linking to them. PageRank is thought to correlate well with human concepts of importance. In addition to PageRank, Google, over the years, has added many other secret criteria for determining the ranking of pages on result lists, reported to be over 250 different indicators, the specifics of which are kept secret to keep

spammers at bay and help Google maintain an edge over its competitors globally.
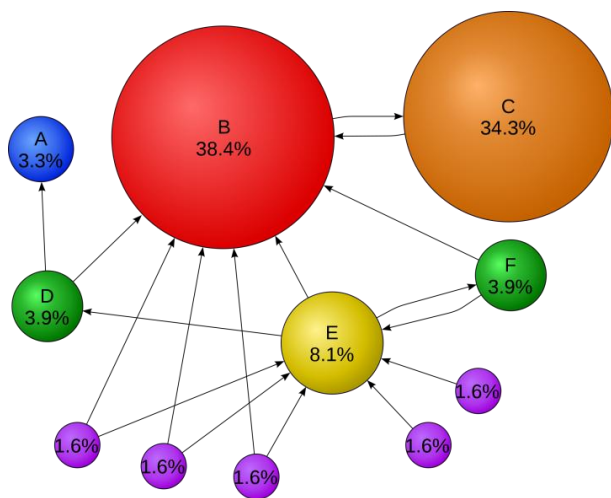
Search Projects:
There are many components to the search process and the results page, and we're constantly updating our technologies and systems to deliver better results. Many of these changes involve exciting new innovations, such as the Knowledge Graph or Google Instant. There are other important systems that we constantly tune and refine. This list of projects provides a glimpse into the many different aspects of search.

Answers: Displays immediate answers and information for things such as the weather, sports scores and quick facts.
Autocomplete: Predicts what you might be searching for. This includes understanding terms with more than one meaning.
Books: Finds results out of millions of books, including previews and text, from libraries and publishers worldwide.
Freshness: Shows the latest news and information. This includes gathering timely results when you're searching specific dates.
Google Instant: Displays immediate results as you type.
Images: Shows you image-based results with thumbnails so you can decide which page to visit from just a glance.
Indexing: Uses systems for collecting and storing documents on the web.
Knowledge Graph: Provides results based on a database of real world people, places, things, and the connections between them.
Mobile: Includes improvements designed specifically for mobile devices, such as tablets and smartphones.
News: Includes results from online newspapers and blogs from around the world.
Query Understanding: Gets to the deeper meaning of the words you type.
Refinements: Provides features like "Advanced Search," related searches, and other search tools, all of which help

you fine-tune your search.
Safe Search: Reduces the amount of adult web pages, images, and videos in your results.
Search Methods: Creates new ways to search, including "search by image" and "voice search."
Site & Page Quality: Uses a set of signals to determine how trustworthy, reputable, or authoritative a source is. (One of these signals is PageRank, one of Google's first algorithms, which looks at links between pages to determine their relevance.)
Snippets: Shows small previews of information, such as a page's title and short descriptive text, about each search result.
Spelling: Identifies and corrects possible spelling errors and provides alternatives.
Synonyms: Recognizes words with similar meanings. Translation and Internationalization Tailors results based on your language and country.
Universal Search: Blends relevant content, such as images, news, maps, videos, and your personal content, into a single unified search results page.
User Context: Provides more relevant results based on geographic region, Web History, and other factors.
Videos: Shows video-based results with thumbnails so you can quickly decide which video to watch.

Non-indexable data:
Despite its immense index, there is also a considerable amount of data available in online databases which are accessible by means of queries but not by links. This so-called invisible or deep Web is minimally covered by Google and other search engines. The deep Web contains library catalogs, official legislative documents of governments, phone books, and other content which is dynamically prepared to respond to a query.

Functionality:
A definition link is provided for many search terms.Google search consists of a series of localized websites. The largest of those, the google.com site, is the top most-visited website in the world. Some of its features include a definition link for most searches including dictionary words, the number of results you got on your search, links to other searches (e.g. for words that Google believes to be misspelled, it provides a link to the search results using its proposed spelling), and many more.

Search syntax:
Google's search engine normally accepts queries as a simple text, and breaks up the user's text into a sequence of search terms, which will usually be words that are to occur in the results, but one can also use Boolean operators, such as: quotations marks (") for a phrase, a prefix such as "+", "-" for qualified terms (no longer valid, the '+' was removed from Google on October 19, 2011), or one of several advanced operators, such as "site:". The web pages of "Google Search Basics" describe each of these additional queries and options (see below: Search options). Google's Advanced Search web

form gives several additional fields which may be used to qualify searches by such criteria as date of first retrieval.

Query expansion:

Google applies query expansion to the submitted search query, transforming it into the query that will actually be used to retrieve results. As with page ranking, the exact details of the algorithm Google uses are deliberately obscure, but certainly the following transformations are among those that occur: Term reordering: in information retrieval this is a standard technique to reduce the work involved in retrieving results. This transformation is invisible to the user, since the results ordering uses the original query order to determine relevance. Stemming is used to increase search quality by keeping small syntactic variants of search terms. There is a limited facility to fix possible misspellings in queries. "I'm Feeling Lucky" "I'm Feeling Lucky" redirects here. For the 2011 book by Douglas Edwards, see I'm Feeling Lucky (book) Google's homepage includes a button labeled "I'm Feeling Lucky". Prior to a change in 2012, when a user typed in a search and clicked on the button the user would be taken directly to the first search result, bypassing the search engine results page. The idea was that if a user is "feeling lucky", the search engine would return the perfect match the first time without having to page through the search results. According to a study by Tom Chavez of "Rapt", this feature cost Google $110 million a year as 1% of all searches use this feature and bypass all advertising. With the introduction of Google Instant, the functionality of the button behaves differently. Currently, the "I'm Feeling Lucky" button changes based on your settings and what webpage you are at. If Google Instant is turned off, the button will direct to the Google Doodles gallery. If Google Instant is turned on and a user hovers over the button, the button text will spin and land on a phrase that starts with "I'm feeling" (e.g. "I'm feeling hungry" or "I'm feeling smart"). Each phrase links to a Google service related to the associated phrase.
Google Chrome and Mozilla Firefox used Lucky Search as the default search string when the user entered a query in the location bar; this functionality was deprecated in later versions.

Rich snippets

See also: Google Searchology § Rich Snippet On May 12, 2009, Google announced that they would be parsing the hCard, hReview, and hProduct microformats and using them to populate search result pages with what they called "Rich Snippets".

V. FIGHTING SPAM

Every day, millions of useless spam pages are created. We fight spam through a combination of computer algorithms and manual review. Spam sites attempt to game their way to the top of search results through techniques like repeating keywords over and over, buying links that pass PageRank or putting invisible text on the screen. This is bad for search because relevant websites get buried, and it's bad for

legitimate website owners because their sites become harder to find. The good news is that Google's algorithms can detect the vast majority of spam and demote it automatically. For the rest, we have teams who manually review sites.

Identifying Spam: Spam sites come in all shapes and sizes. Some sites are automatically-generated gibberish that no human could make sense of. Of course, we also see sites using subtler spam techniques. Check out these examples of "pure spam," which are sites using the most aggressive spam techniques. This is a stream of live spam screenshots that we've manually identified and recently removed from appearing in search results.

Types of Spam: In addition to spam shown above, here are some other types of spam that we detect and take action on.

Cloaking and/or sneaky redirects: Site appears to be cloaking (displaying different content to human users than is shown to search engines) or redirecting users to a different page than Google saw.

Hacked site: Some pages on this site may have been hacked by a third party to display spammy content or links. Website owners should take immediate action to clean their sites and fix any security vulnerabilities.

Hidden text and/or keyword stuffing: Some of the pages may contain hidden text and/or keyword stuffing.

Parked domains: Parked domains are placeholder sites with little unique content, so Google doesn't typically include them in search results.

Pure spam: Site appears to use aggressive spam techniques such as automatically generated gibberish, cloaking, scraping content from other websites, and/or repeated or egregious violations of Google's Webmaster Guidelines.

Spammy free hosts and dynamic DNS providers: Site is hosted by a free hosting service or dynamic DNS provider that has a significant fraction of spammy content.

Thin content with little or no added value: Site appears to consist of low-quality or shallow pages which do not provide users with much added value (such as thin affiliate pages, doorway pages, cookie-cutter sites, automatically generated content, or copied content).

Unnatural links from a site: Google detected a pattern of unnatural, artificial, deceptive or manipulative outbound links on this site. This may be the result of selling links that pass PageRank or participating in link schemes.

Unnatural links to a site: Google has detected a pattern of unnatural artificial, deceptive or manipulative links pointing to the site. These may be the result of buying links that pass PageRank or participating in link schemes.

User-generated spam: Site appears to contain spammy user-generated content. The problematic content may appear on forum pages, guestbook pages, or user profiles.

Taking Action

While algorithms address the vast majority of spam, Google addresses other spam manually to prevent it from affecting the quality of your results. This graph shows the number of domains that have been affected by a manual action over time and is broken down by the different spam types. The numbers may look large out of context, but the web is a really big place. A recent snapshot of index showed that about 0.22% of domains had been manually marked for removal.
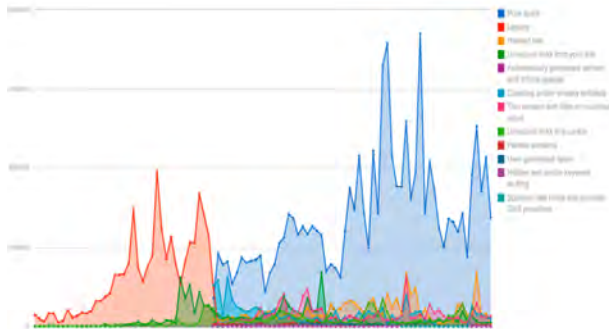
Notifying Website Owners

When Google take manual action on a website, they try to alert the site's owner to help him or her address issues. Google want website owners to have the information they need to get their sites in shape. That's why, over time, Google have invested substantial resources in webmaster communication and outreach.

CONCLUSION

Google is designed to be a scalable search engine. The primary goal is to provide high quality search results over a rapidly growing World Wide Web. Google employs a number of techniques to improve search quality including page rank, anchor text, and proximity information. Furthermore, Google is a complete architecture for gathering web pages, indexing them, and performing search queries over them.

REFERENCES

[1] https://www.google.com/insidesearch/howsearchworks/thestory/index.html

[2] https://en.wikipedia.org/wiki/Google_Search

[3] "The Anatomy of a Large-Scale Hypertextual Web Search Engine". Computer Science Department, Stanford University, Stanford, CA. Retrieved January 27, 2009.

[4] "Google.com Site Info". Alexa Internet.

[5] "Google Help : Cheat Sheet".

About the Institute

Thakur Institute of Management Studies, Career Development and Research (TIMSCDR) was established in the year 2001 with a clear objective of providing quality technical education in tune with international standards and contemporary global requirements. TIMSCDR offers 3 years full time Post Graduate Degree of Master of Computer Applications (MCA). The Institute is approved by All India Council for Technical Education (AICTE), Directorate of Technical Education (DTE), Govt. of Maharashtra and is affiliated to University of Mumbai.

Management's commitment to excellence is reflected in the marvelous infrastructure that is comparable to the finest institution of its type in the country. The sprawling campus with lush green gardens, cafeteria, playgrounds, parking area, hostel accommodation and temple, ensures right academic ambience essential for a center of higher education.

At TIMSCDR, the importance of faculty is well understood which is reflected in qualified and experienced teaching staff. A closely monitored quality assurance mechanism ensures proper coverage of syllabus within right time frame. Additional coaching in comparatively difficult topics in various subjects and individual mentoring are some of the healthy academic practices followed at TIMSCDR.

Application of modern technology in teaching-learning process and day to day governance of the Institute makes TIMSCDR unique. Within 15 years of its existence, TIMSCDR has carved out a niche for itself as one of the leading MCA institutes under University of Mumbai.

The Institute focuses on imparting knowledge to the students that persists even when they pass out and step into the corporate world. Syllabus is given a new dimension through experienced faculty and state of the art infrastructure. Overall personality development through extra curricular activities like Aptitude Tests, Group Discussions, Presentation by students, Mock Interviews, and Seminars etc. have been a hallmark of the Institute.