PATNA WOMEN'S COLLEGE 1940

Computer Science (BCA)

Explore–Journal of Research for UG and PG Students
ISSN 2278 – 0297 **(Print)**
ISSN 2278 – 6414 **(Online)**

http://www.patnawomenscollege.in/journal

Big Data: The Next Frontier for Innovation, Competition & Productivity

• **Ankita Pandey** • **Ankita** • **Anuradha**
• **Amrita Prakash**

Received : November 2014
Accepted : March 2015
Corresponding Author : Amrita Prakash

Abstract : *This research work is on study of Big Data and how it is transforming our lives and the business we do. It also includes the study of how to use the ever increasing volumes of data in an appropriate manner. It deals with the factors which fuels Big Data and the methods by which all the data generated can be converted into meaningful outputs. The study also deals with practical examples of how Big Data is used to deliver the real values.*

Keywords *: Big Data, Volumes of Data.*

Ankita Pandey
BCA III year, Session: 2012-2015,
Patna Women's College, Patna University, Patna,
Bihar, India

Ankita
BCA III year, Session: 2012-2015,
Patna Women's College, Patna University, Patna,
Bihar, India

Anuradha
BCA III year, Session: 2012-2015,
Patna Women's College, Patna University, Patna,
Bihar, India

Amrita Prakash
Assistant Professor, Department of BCA,
Patna Women's College, Bailey Road,
Patna – 800 001, Bihar, India
E-mail : bitzbinni@gmail.com

Introduction :

What is a Big Data ?

Every day, we create 2.5 quintillion bytes of data — so much that 90% of the data in the world today has been created in the last two years alone. This data comes from everywhere: sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction records, and cell phone GPS signals to name a few.

This data is "BIG DATA."

Big Data is fundamentally networked. Its value comes from the patterns that can be derived by making connections between pieces of data, about an individual, about individuals in relation to others, about groups of people, or simply about the structure of information itself. But as data Big Data is more than *big* data—it doesn't have to be big to be different—it's a qualitative shift mostly. Big Data is not (primarily) about size/volume; it is about very many very small data produced by / about connected individuals (big data is small data—it can also be slow data).

It means collecting a wider range of data than the organisation's current or traditional analysis requires. So instead of running the same monthly

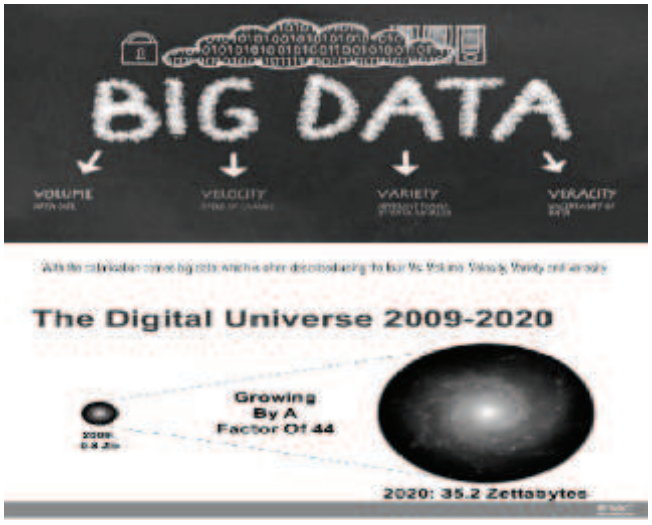
reports on the financial system, for example, or on website statistics, the business incorporate data that may have no currently known relationship with how it analyses its data today.

Second, it means deriving new insights by combining this disparate data (that may also be external, such as geospatial or social) in new ways. This is more about 'data discovery' than about 'weekly reporting packages'. The third part is about more effective visualisation helping stakeholders absorb, share and exploit insights from new data analyses.

Factors that have lead the world to make use of big data are:

- Understanding and navigating federated big data sources
- Managing and storing huge volume of any data
- Structuring and control of data
- Managing streaming data
- Analyzing unstructured data
- Integrating and governing all data sources
- Big Data Exploration: to find, to visualize, to understand all big data to improve decision making
- Security/Intelligence Extension: Lower risk, detecting fraud and monitoring cyber security in real-time

Big data spans four dimensions: Volume, Velocity Variety and Veracity



Volume: Refers to the vast amount of data generated every second. We are not talking Terabytes but Zettabytes or BronotoBytes. If we take all the data generated in the world between the beginning of time and 2008, the same amount of data will soon be generated every minute. New big data tools use distributed system so that we can store and analyze data across database that are dotted anywhere in the world

Velocity: Refers to the speed at which new data is generated and the speed at which data moves around. Just think of social media message going viral in seconds. Technology allows us now to analyze the data while it is being generated (sometimes referred to as in-memory analytics), without ever putting it into database.

Variety: Refers to the different types of data we can now use. In the past we only focused on structured data that neatly fitted into tables or relational database, such as financial data. In fact 80% of the world's data is unstructured (text, image, video, voice, etc.) with big data technology we can now analyze and bring together data of different types such as messages, social media conversations, photos, sensor, data, video or voice recordings etc.

Veracity: Refers to the messiness or trustworthiness of the data. Many forms of big data quality and accuracy are less controllable(just think of twitter posts with hash tags ,abbreviations, types and colloquial speech as well as the reliability and accuracy of content) technology now allows us to work with this type of data.

Big Data Framework

Here is a framework to understand big data strategies and the techniques used with each strategy. The first dimension is labeled business objective. When developing big data capabilities, companies try to measure or experiment.

The second dimension is labeled data type. In their normal course of functioning, companies collect data on their operations (e.g., sales) and capture it in their database that has a structure or schema. We call this transactional data. In other instances, companies deal with data that come from sources other than transactions and are typically unstructured (e.g., social media data).

Four Big Data strategies

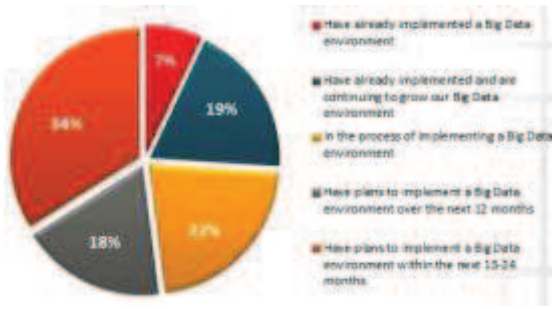
1. Performance Management
   - Performance management involves understanding the meaning of big data in company databases using pre-determined queries and multidimensional analysis.
   - The data used for this analysis are transactional, for example, years of customer purchasing activity, and inventory levels and turnover.
   - A big benefit for report developers is that they can interact with different aspects of business data including HR, marketing, sales, customer service, and manufacturing data, and get multiple perspectives of how the business is doing.
2. Data Exploration
   - Data exploration makes heavy use of statistics to experiment and get answers to questions that managers might not have thought of previously.
   - This approach leverages predictive modeling techniques to predict user behavior based on their previous business transactions and preferences.
   - Cluster analysis can be used to segment customers into groups based on similar attributes that may not have been on analysts' radar screens. Once these groups are discovered, managers can perform targeted actions such as customizing marketing messages, upgrading service, and cross/up-selling to each unique group.
   - The results of these experiments can help predict which combination of these variables will lead to the highest conversion rate of site visitors to qualified leads, and qualified leads to customers.
   - The rise in robust statistical/ analytical techniques can lead to fast, direct results for data exploring organizations. The big challenge is the lack of qualified statisticians with expertise in the latest business analytical techniques.
3. Social Analytics
   - Social analytics measure the vast amount of non-transactional data that exists today. Much of this data exist on social media platforms, such as conversations and reviews on Facebook, Twitter, and Yelp.

- Social analytics measure three broad categories: awareness, engagement, and word-of-mouth or reach.
- Awareness looks at the exposure or mentions of social content and often involves metrics such as the number of video views and the number of followers or community members.
- Engagement measures the level of activity and interaction among platform members, such as the frequency of user-generated content.
- Reach measures the extent to which content is disseminated to other users across social platforms. Reach can be measured with variables such as the number of retweets on Twitter and shared likes on Facebook.

4. Decision Science

- Decision science involves experiments and analysis of non-transactional data.
- Unlike social analyzers who focus on social analytics to measure known objectives, decision scientists explore social big data as a way to conduct "field research" and to test hypotheses.
- Techniques used by decision scientists involve listening tools that perform text and sentiment analysis.

Implementation of Big Data Environment :



Provocations of Big Data

1. Automating Research Changes the Definition of Knowledge.

Big Data not only refers to very large data sets and the tools and procedures used to manipulate and analyze them, but also to a *computational turn* in thought and research (Burkholder 1992). Just as Ford changed the way we made cars – and then transformed work itself – Big Data has emerged a system of knowledge that is already changing the objects of knowledge, while also having the power to inform how we understand human networks and community.

We would argue that Big Data creates a radical shift in how we think about research. If we are observing the automation of particular kinds of research functions, then we must consider the inbuilt flaws of the machine tools. It is not enough to simply ask, as Anderson suggests 'what can science learn from Google?', but to ask how Google and the other harvesters of Big Data might change the *meaning* of learning, and what new possibilities and new limitations may come with these systems of knowing.

2. Claims to Objectivity and Accuracy are Misleading

'Sociology has been obsessed by the goal of becoming a quantitative science.' Yet sociology has never reached this goal, because of where it draws the line between what is and is not quantifiable knowledge in the social domain. Big Data offers the humanistic disciplines a new way to claim the status of quantitative science and objective method. It makes many more social spaces quantifiable. In reality, working with Big Data is still subjective, and what it quantifies does not necessarily have a closer claim on objective truth – particularly when considering messages from social media sites. But there remains a mistaken belief that qualitative researchers are in the business of interpreting stories and quantitative researchers are in the business of producing facts. In this way, Big Data risks reinscribing established

divisions in the long running debates about scientific method.

All researchers are interpreters of data. As computational scientists have started engaging in acts of social science, there is a tendency to claim their work as the business of facts and not interpretation. A model may be mathematically sound, an experiment may seem valid, but as soon as a researcher seeks to understand what it means, the process of interpretation has begun. The design decisions that determine what will be measured also stem from interpretations.

3. Bigger Data are Not Always Better Data

Social scientists have long argued that what makes their work rigorous is rooted in their systematic approach to data collection and analysis. Ethnographers focus on reflexively accounting for bias in their interpretations. Experimentalists control and standardize the design of their experiment. Survey researchers drill down on sampling mechanisms and question bias. Quantitative researchers weigh up statistical significance. These are but a few of the ways in which social scientists try to assess the validity of each other's work. Unfortunately, some who are embracing Big Data presume the core methodological issues in the social sciences are no longer relevant. There is a problematic underlying ethos that bigger is better, that quantity necessarily means quality.

Twitter provides an example in the context of a statistical analysis. First, Twitter does not represent 'all people', although many journalists and researchers refer to 'people' and 'Twitter users' as synonymous. Neither is the population using Twitter representative of the global population. Nor can we assume that accounts and users are equivalent. Some users have multiple accounts. Some accounts are used by multiple people. Some people never establish an account, and simply access Twitter via the web. Some accounts are 'bots' that produce automated content without involving a person. Furthermore, the notion of an 'active' account is problematic. While some users post content frequently through Twitter, others participate as 'listeners' (Crawford 2009, p. 532). Twitter Inc. has revealed that 40 percent of active users sign in just to listen (Twitter, 2011). The very meanings of 'user' and 'participation' and 'active' need to be critically examined.

Twitter has become a popular source for mining Big Data, but working with Twitter data has serious methodological challenges that are rarely addressed by those who embrace it. When researchers approach a dataset, they need to understand – and publicly account for– not only the limits of the dataset, but also the limits of which questions they can ask of a dataset and what interpretations are appropriate.

4. Not All Data Are Equivalent

Some researchers assume that analyses done with small data can be done better with Big Data. This argument also presumes that data is interchangeable. Yet, taken out of context, data lose meaning and value. Context matters. When two datasets can be modeled in a similar way, this does not mean that they are equivalent or can be analyzed in the same way. Consider, for example, the rise of interest in social network analysis that has emerged alongside the rise of social network sites and the industry-driven obsession with the 'social graph'. Countless researchers have flocked to Twitter and Facebook and other social media to analyze the resultant social graphs, making claims about social networks.

However, the networks produced through social media and resulting from communication traces are not necessarily interchangeable with other social network data .Just because two people are physically co-present – which may be made visible to cell towers or captured through photographs – does not mean that they know one another .Furthermore, rather than indicating the presence of predictable objective patterns, social

network sites facilitate connectedness across structural boundaries and act as a dynamic source of change: taking a snapshot, or even witnessing a set of traces over time does not capture the complexity of all social relations.

5. Just Because it is Accessible Doesn't Make it Ethical

In 2006, a Harvard-based research project started gathering the profiles of 1,700 college based Facebook users to study how their interests and friendships changed over time. This supposedly anonymous data was released to the world, allowing other researchers to explore and analyze it. What other researchers quickly discovered was that it was possible to de-anonymize parts of the dataset: compromising the privacy of students, none of whom were aware their data was being collected.

The case made headlines, and raised a difficult issue for scholars: what is the status of so called 'public' data on social media sites? Can it simply be used, without requesting permission? What constitutes best ethical practice for researchers? Privacy campaigners already see this as a key battleground where better privacy protections are needed. The difficulty is that privacy breaches are hard to make specific – is there damage done at the time? What about twenty years hence? 'Any data on human subjects inevitably raise privacy issues, and the real risks of abuse of such data are difficult to quantify '.Even when researchers try to be cautious about their procedures, they are not always aware of the harm they might be causing in their research.

6. Limited Access to Big Data Creates New Digital Divides

Much of the enthusiasm surrounding Big Data stems from the perception that it offers easy access to massive amounts of data. But who gets access? For what purposes? In what contexts? And with what constraints? While the explosion of research using data sets from social media sources would suggest that access is straightforward, it is anything but. Thay only social media companies have access to really large social data – especially transactional data. An anthropologist working for Facebook or a sociologist working for google will have access to data that the rest of the scholarly community will not. Some companies restrict access to their data entirely; other sell the privilege of access for a high fee; and others offer small data sets to university-based researchers. This produces considerable unevenness in the system: those with money – or those inside the company– can produce a different type of research than those outside. Those without access can neither reproduce nor evaluate the methodological claims of those who have privileged access.

It is also important to recognize that the class of the Big Data rich is reinforced through the university system: top-tier, well-resourced universities will be able to buy access to data, and students from the top universities are the ones most likely to be invited to work within large social media companies. Those from the periphery are less likely to get those invitations and develop their skills. The result is that the divisions between those who went to the top universities and the rest will widen significantly.

In addition to questions of access, there are questions of skills. Wrangling APIs, scraping and analyzing big swathes of data is a skill set generally restricted to those with a computational background. When computational skills are positioned as the most valuable, questions emerge over who is advantaged and who is disadvantaged in such a context. This, in its own way, sets up new hierarchies around 'who can read the numbers', rather than recognizing that computer scientists and social scientists both have valuable perspectives to offer. Significantly, this is also a gendered division. There are complex questions about what kinds of research skills are valued in the future and how those skills are taught. How can students be educated so that they are equally

comfortable with algorithms and data analysis as well as with social analysis and theory?

Why should entrepreneurs adopt Big Data ?

Today enterprises are in the midst of a transformation that compels them to be analytics-driven to remain competitive in the coming years. Analyzing their data will not only enable them to get a thorough insight about their business, but also give significant competitive edge in the market place. Example of analytics-driven activities includes service innovation, customer experience improvements, detection and remediation of anomalies, and reduction in time to market for products and services. To meet the need for analytics-driven activities, businesses need to collect, analyze, and store more and more data, often from their diverse sources. They should also adopt such tools and workflows that allow them to quickly and continuously analyze results from stored and transient, high velocity and voluminous data and change their business functions accordingly. This means the systems will be more agile than their predecessors.

Here are some facts: -

- The New York Stock Exchange generates about one terabyte of new trade data per day
- Facebook hosts approximately 10 billion photos, taking up one petabyte of storage
- Ancestry.com, the genealogy site, stores around 2.5 petabytes of data
- The internet archive stores around 2 petabytes of data, and is growing at a rate of 20 terabytes per month
- The Large Hadron Collider near Geneva, Switzerland, will produce about 15 petabytes of data per year

According to Gartner report, while big data has certainly got people's attention, 2014 will be the year that is catapulted to the forefront. Most of the enterprises will start to focus on processes and hit their stride for investment in big data.

Mega trends using Big Data

Here are some mega trends in different sector of enterprises which indicates a growing need of big data in enterprises: -

Mega Trend 1: In healthcare sector

Traditionally, the healthcare industry has trailed behind other industries in the use of big data. Part of the problem stems from resistance to change. Providers are accustomed to making treatment decisions independently, using their own clinical judgment, rather than relying on protocols based on big data. Other obstacles are more structural in nature. Many healthcare stakeholders have underinvested in information technology because of uncertain returns. Although their older systems are functional, they have limited ability to standardize and consolidate data. The nature of the healthcare industry itself also creates challanges: while there are many players, there is no possibility to share data easily among different

providers or facilities, partly because of privacy concerns. And even within a single hospital, payer or pharmaceutical company, important information often remains soloed within one group or department because organizations lack procedures for integrating data and communicating findings.

But as the market is becoming competitive and a core demand for more is everywhere, stakeholders have now started to access promising new threads of knowledge. Pharmaceutical-industry experts, payers and providers are now beginning to analyze big data to obtain insights. Although these efforts still remain in their early stages, they could collectively help the industry address problems related to variability in healthcare quality and escalating healthcare spend. For instance, researchers can mine the data to see what treatments are most effective for particular conditions, identify patterns related to drug side effects or hospital re-admissions, and gain other important information that can help patients and reduce costs. Many innovative companies in the private sector, both established players and new entrants, are now focusing and building applications and analytical tools that help patients, physicians and other healthcare stakeholders identify value and opportunities. Fortunately, as their technological capabilities and understanding advance, we expect that innovators will develop solutions which will transform healthcare in a big way, some of which could help substantially reduce the soaring cost of healthcare by delivering consistent, quick and accurate diagnosis or prognosis.

Mega Trend 2: In educational sector

New technologies allow schools, colleges and universities to analyze absolutely everything that happens - from student behavior, testing results, careers developments of students and educational needs based on changing societies. A lot of these data have already been stored and are used for statistical analysis by government agencies. With the increasing value from high quality online education and development, all data get a completely new meaning. Big data allows changes in the educational field that will revolutionize the way students learn and teachers teach. Some are mentioned below :

To improve student results: The overall goal of big data within the educational system should be to improve student results. Better students are good for society, organizations and educational institutions. Currently, the answers to assignments and exams are the only measures to assess the performance of students. However, during his/her student life, every student generates a unique data trail. This data trail can be analyzed in real-time to deliver an optimal learning environment for the student and to gain a better understanding in his/her individual behavior. It is possible to monitor every action of the students - how long they take to answer a question, which sources they use, which questions they skipped, how much research was done, what is the relation to other questions answered, which tips works best for which student etc. Answers to questions can be checked instantly and can automatically (except for essays) give instant feedback to students.

To improve the learning experience in real-time: When students start working on their own, in their customized learning program, the vast amount of teaching, which most of the time is covered by general topics that have to appeal to all students from different levels, can be done online and by themselves. The professor can monitor all students in real-time and start a much more interesting and deeper conversation on the topic of choice. This will give students the possibility to gain a better understanding of the topics.

To create mass-customized programs: All this data will help to create a customized program for each individual student. Big data allows customization at colleges and universities, even if they have 10.000s of students. This will be created

with blended learning; a combination of online and offline learning. It will give students the opportunity to develop their own personalized program, following those classes that they are interested in, working at their own pace, while having the possibility for (offline) guidance by professors. Providing mass customization in education is a challenge, but algorithms make it possible to track and assess each individual student.

To reduce dropouts, increase results: All these analyses will improve the students' results and perhaps also reduce dropout rates at universities or colleges. Dropouts are expensive for educational institutes and for the society. When students are closely monitored, receive instant feedback and coached based on their personal needs, it can help reduce dropout rates, as mentioned in post one of the technical site.

Mega Trend 3: In social sector

We all know that social media generates huge amounts of data. The explosive growth of social media is one of the reasons that 90% of all the data in the world has been generated in the last two years alone. There are over a billion active users of social media network worldwide, many of whom are frequently active and can be connected by means of their smart phones and tablets. Social media indeed has become a main communication network in the daily lives of people around the world. Did you know that any activity in social media is generating insurmountable data? These big data are waiting to be explored. In fact, social media now embodies the leading and biggest source of consumer data.

Mega Trend 4: In government sector

Although the business sector is leading big data application development, the public/government sector has begun to derive insights to help support decision making in real time from fast-growing in-motion data from multiple sources, including web, biological and industrial sensors, video, email, and social communications. Many white papers, journal articles and business reports have proposed ways governments can use big data to help them serve their citizens and overcome national challenges (such as rising healthcare costs, job creation, natural disasters and terrorism).There is also some skepticism as to whether it can actually improve government operations, as governments must develop new capabilities and adopt new technologies to transform it into information through data organization and analytics.

Mega Trend 5: In Insurance sector

The insurance industry is one of the most data driven industries on the planet, so it is amazing how few of them use crowd sourcing solutions to enhance their business. Joe Martinez, an Insurance Adjuster from Austin, TX outlines the problem. "Most insurance companies are content to stand on the shoulders of experts, adjusters like me, accountants and actuaries. They miss out on a big opportunity every year. Big data would allow the insurance companies to determine variables that they have not been looking at, like the link between mobile phone usage and accidents. "There are data points other than demographics that can more accurately assess risk than demographic data points like age or sex.

Mega Trend 6: In sales sector

Sales department has a good service that can only be acquired through financial transaction, and most buyers are already motivated. It becomes a process of facilitating the transaction and this is where data comes in. According to a top automobile salesman from Redmond, WA says that analytics play a huge part in how he pursues his leads. "I was able to use data on how long it has been since a customer bought a vehicle and how often they had to get it fixed, to calculate a percentage of them that were likely to be unhappy with their vehicle. By focusing direct marketing on this segment I actually saw a huge jump in sales from those who were in the market for a new car without even realizing it. This data gave me jump on the competition."

Big Data can also help sales teams determine what to concentrate on. The COO of a company that uses data to help companies engage, retain and develop the workforce, talks about a study that asked employees to rank the importance of sales skills. After comparing employee data with the real data, the company found that focus on 'value creation' was the most important aspect to the sales process and not 'sales presentation.' "In this case, the analysis showed that investing time and money on emphasizing value creation would likely to have the biggest impact on sales revenue," explains the COO.

Apart from the above explained sectors there are many more other sectors e.g. agriculture, non-profits, securities, life sciences, surveillance, defense, intelligence, cyber etc., which are also focusing on adopting big data as per trend shows.

Changing Business Models as the Mega Trends Gain Ground

All the major industries (healthcare, government, education, social, insurance and sales) tracked as part of this research have given clear indication of shifting business models to respond better to their customers journey and be at par with the competitors.

Our researchers worked to identify how mega trends in enterprises are impacting the traditional business models:

1. How to improve efficiency and reduced costs of all machines operated while farming

2. How to improve productivity and efficiency of crops as well as animals

3. How to mitigate weather conditions and optimize pricing for products

On the other hand, for any non-profitable/profitable enterprise, it is really important for a nonprofit enterprise to turn profitable while any profitable enterprise to continue as a profitable organization.

Staying Relevant with Big Data

IT providers will need to play an industry-defining role by transforming into a 'master of big data game' to stay relevant in the emerging industry context. Providers are not dealing with just discrete technologies or one monolithic industry.

Enterprises today are seeking business transformation capabilities from IT providers by adopting big data for:

Risk Management - Inaccurate risk assessment can lead to poor decision making, high costs and scrutiny from regulators. With so much data to analyze, companies need a systemic approach to effectively identify and assess all of the risk exposures, known and unknown, that their business faces. Organizations must identify every risk, from threats to the company's reputation to data breaches and risk of non-compliance with regulations, and weigh them against business opportunities. By analyzing with big data, they can strike a balance between risk and opportunity.

Driving R&D - Big Data for R&D is less about velocity and more about variety, viability and sometimes volume. The key analytics capability for data is the ability to visualize relationships and patterns. By combining real-world outcomes data with clinical data and through the mining of genetic data and a broader understanding of regional and population data, analytically savvy organizations can begin to recognize research failures faster, design more efficient trials and speed the discovery and approval of new innovation while lowering costs along the way.

Analyzing Consumer Behavior - The emergence of internet, e-commerce and social media has radically altered the landscape of available consumer behavior data. Cash registers and Point-of-Sale (POS) systems are being replaced by e-commerce sites that record every move consumers make, even when they don't buy something. Casual telephone conversations about recent purchases are being replaced by tweets that can be scanned and analyzed by anyone who follows those Twitter feeds. All of this data on actual consumer behavior and experiences is there to be measured and analyzed.

In a nutshell, the ability to capture, translate and leverage increasing amounts of information, from consumer and shopper data, customer or retailer data, social media and the real-time visibility into the demand and supply chains is critical. Investing in the necessary tools and analyzing business-critical information can help develop a deeper level of understanding of customers and products, enhance brand loyalty, increase sales and drive competitive advantage. As such, organizations must equip themselves with the ability to store and quickly process these massive volumes of data with the appropriate technology to develop actionable insights.

A Data Warehouse as a Foundation for Big Data

Even as new big data architectures emerge and mature, business users will continue to analyze data by directly leveraging and accessing data warehouses. The rest of this paper describes how Oracle Database 12c provides a comprehensive platform for data warehousing that combines industry-leading scalability and performance, deeply-integrated analytics, and advanced workload management – all in a single platform running on an optimized hardware configuration.

Exadata : The bedrock of a solid data warehouse solution is a scalable, high-performance hardware infrastructure. One of the long-standing challenges for data warehouses has been to deliver the IO bandwidth necessary for large-scale queries, especially as data volumes and user workloads have continued to increase. While the Oracle Exadata Database Machine is designed to provide the optimal database environment for every enterprise database, the Exadata architecture also provides a uniquely optimized storage solution for data warehousing that delivers order-of magnitude performance gains for large-scale data warehouse queries and along with very efficient data storage. A few of the key features of Exadata that are particularly valuable to data warehousing are:

Exadata Smarts Scans. With traditional storage, all database intelligence resides on the database servers. However, Exadata has database intelligence built into the storage servers. This allows database operations, and specifically SQL processing, to leverage both the storage servers and database servers to vastly improve performance. The key feature is "Smart Scans", the technology of offloading some of the data-intensive SQL processing into the Exadata Storage Server: specifically, row-filtering (the evaluation of where-clause predicates) and column-filtering (the evaluation of the select-list) are executed on Exadata storage server, and a much smaller set of filtered data is returned to the database servers. "Smart scans" can improve the query performance of large queries by an order of magnitude, and in conjunction with the vastly superior IO bandwidth of Exadata's architecture delivers industry-leading performance for large-scale queries.

Exadata Storage Indexes. Completely automatic and transparent, Exadata Storage Indexes maintain each column's minimum and maximum values of tables residing in the storage server. With this information, Exadata can easily filter out unnecessary data to accelerate query performance.

Hybrid Columnar Compression. Data can be compressed within the Exadata Storage Server into a highly efficient columnar format that provides up to a 10 times compression ratio, without any

loss of query performance. And, for pure historical data, a new archival level of hybrid columnar compression can be used that provides up to 40 times compression ratios.

Important case studies on Big Data

These case studies highlight how large companies leverage Big Data for driving productivity.

23andMe : 23andMe is a privately held personal genomics and biotechnology company. The company has developed its whole model around pulling insights from big data to give customers a 360-degree understanding of their genetic history.

CBA : Commonwealth Bank of Australia is using big data to analyze customer risk. Using analytics can get better risk assessment businesses, ongoing cash flow performance and early warning of risk challenges.

Centers for Disease Control : The Centers for Disease Control and Prevention (CDC) is the national public health institute of the United States. Its main aim is to protect people health and safety through the control and prevention of diseases. CDC had to rely on doctor reports of influenza outbreaks. CDC was weeks behind in providing vaccines to the affected patients. Using historical data from the CDC, Google compared search term queries against geographical areas that were known to have had flu outbreaks. Google then found forty five terms correlated with the outbreak of flu. With this data, CDC can act immediately.

Delta : Delta Air Lines, Inc. is a major American airline with an extensive domestic and international network. In general the top concern for an airlines would be passenger's lost baggage. Delta looked further into their data and created a solution that would remove the uncertainty of where a passenger's bag might be.

Energy Future Holdings Corporation : Energy Future Holdings Corporation is an electric utility company. The majority of the company's power generation is through coal- and nuclear-power plants. The company used Big data to install smart meters. The smart meters allow the provider to read the meter once every 15 minutes rather than one month.

Google : Google constantly develops new products and services that have big data algorithms. Google uses big data to refine its core search and ad-serving algorithms. Google describes that the self-driving car as a big data application.

Kreditech : Kreditech is a young tech company headquartered in Hafencity, Hamburg. The European company uses Big Data to create a unique credit score for consumers using more than 8000 sources. The analysis also lead to a surprise discovery of correlation between social media behaviour and financial stability.

LinkedIn : LinkedIn is a business-oriented social networking service. Founded in December 2002 and launched in 2003, it is mainly used for professional networking. LinkedIn uses big data to develop product offerings such as people you may know, jobs you may be interested in, who has viewed my profile and more.

McLaren's Formula One racing team : McLaren Racing Limited is a British Formula One team. The racing car team uses real-time car sensor data during car races, identifies issues with its racing cars using predictive analytics and takes corrective actions pro-actively before it's too late.

Singapore healthcare : The healthcare providers in Singapore used analytics to better understand each patient's condition, lifestyle choices, work and home environment. They can create personalized treatment plans tailored to that person's individual behavior.

Sprint : Sprint Corporation, is a United States telecommunications holding company that provides wireless services and is also a major global Internet carrier. It is the third largest U.S. wireless network operator as of 2014.Wireless carrier Sprint, uses smarter computing - primarily big data analytics to put real-time intelligence and control back into the network driving a 90% increase in capacity.

UPS : United Parcel Service of North America, Inc., referred as UPS , is one of the largest shipment and logistics companies in the world. The company tracks data on 16.3 million packages per day for 8.8 million customers, with an average of 39.5 million tracking requests from customers per day. The company stores over 16 petabytes of data.

US Xpress : US Xpress, provider of a wide variety of transportation solutions collects about a thousand data elements ranging from fuel usage to tire condition to truck engine operations to GPS information, and uses this data for optimal fleet management and to drive productivity saving millions of dollars in operating costs.

Verizon : Verizon uses big data to enhance mobile advertising. A unique identifier is created when the user registers in the website. The identifier allows advertiser to use information from the desktop computer. Marketing messages can be delivered to you mobile phone using this information.

Advnatages of Big Data :

- Understanding and Targeting costumers
- Understanding and Optimizing Business Processes
- Improving Science and Research
- Improving Healthcare and Public Health
- Financial Trading
- Improving Sports Performance
- Improving Security and Law Enforcement

Disadvantages of Big Data :

- Big Data violates the privacy principle.
- Data can be used for Manipulating customers.
- Big Data may increase Social Stratification.
- Big data is not useful in Short Run.
- Faces difficulties in parsing and interpreting.

Conclusion :

We believe enterprises will dive into their big data and the approaches would be very different from the kind we see today. You can expect to see a much more real time analytics-focused and agile set of organizations, as they leverage big data in business and information technology. Traditional businesses will not change, but the convergence of innovative technologies to change business process and models will bring in greater agility.

The new business model based on the mega trends will be the key differentiator between competing firms. Therefore, firms must prepare themselves today to remain relevant tomorrow.

References :

Tom Davenport's. *Big data at work : Dispelling the myths, Uncovering the Opportunities.* Harvard Business Press.

www.csi-india.org

www.bigdata-madesimple.com

www.ivey.com/analytics-insights

www.happiestminds.com/sites/default/files/big-data-why-should-enterprises-adopt-to.pdf. (Author : Manish Kumar). Accessed on Mon, Sep. 1, 2014 at 09:47 a.m.