Young Scientist- Tomorrow's Science Begins Today
Vol. 1, Issue – 2016
ISSN: 2581-4737

INNOVATION, OPPORTUNITIES AND CHALLENGES IN BIG DATA

ABHILASH SHUKLA*
*Assistant Professor, P.G and Research Department of Computer, Sri Paramakalyani College.
Correspondence E-mail Id: editor@eurekajournals.com

HOW IS BIG DATA DIFFERENT FROM "LITTLE DATA"?

Let's take a hypothetical scenario in which you suddenly find a leakage in a water pipe in your garden. You take a bucket and some fixing material to settle the issue. After some time, you realize that the leak is much bigger that you cannot fix it by yourself and you need a call a plumber and tell him to bring bigger tools. While waiting for plumber to come, you are still using the bucket to drain the water. Sooner or later, you notice that a monstrous underground stream has opened and you have to handle millions of litres of water each second.

Now to handle this situation, you don't just need new buckets, yet a totally new way to deal with the problem just because the volume and velocity of water has grown. To prevent the town from flooding, perhaps you require your government to build a massive dam that requires a tremendous civil engineering expertise and an elaborate control system. To make things worse, everywhere water is gushing out from nowhere and everyone is scared with the variety.

WELCOME TO WORLD OF BIG DATA!!

WHAT IS BIG DATA

Based on my so far understanding about big data, the simplest way to define it as "*Big data is considered as so huge and complex data sets, that it becomes difficult or impossible to process those using traditional database management applications*".

Young Scientist- Tomorrow's Science Begins Today
Vol. 1, Issue – 2016
ISSN: 2581-4737

BIG DATA ORIGINS

Some time ago in 2002, Google was wondering how to handle extensive stacks of piled up data. They initially attempted to save files manually inside a customized distributed file system architecture, but after sometime they realize that its very careful, tedious and at the same time very time-consuming process. Google then sought an automated approach to solve this problem. Subsequently, to solve this problem they released a white paper titled "Google Distributed File System".

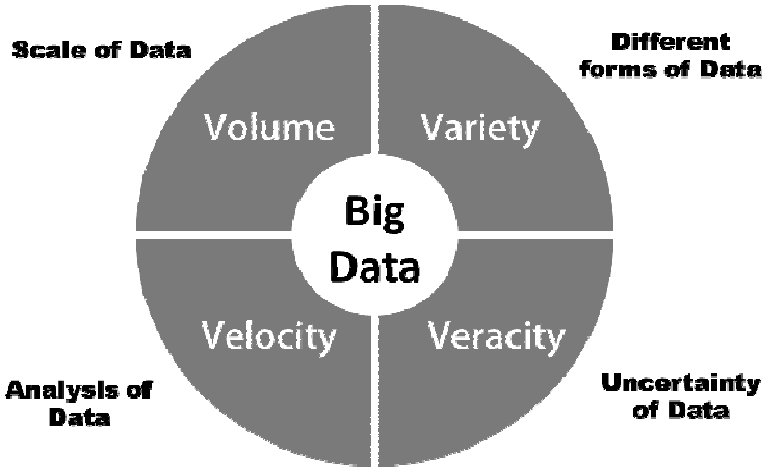
In the year 2002-2003, a search framework development specialist named Doug Cutting began practically working on the white paper's content. Word went around and Yahoo employed Cutting. In 2005, Hadoop was consequently implemented by Yahoo, donated to Apache, as a full-fledged project.

Google now went a step ahead and searched for a comparative achievement in analytics. Another white paper 'Google MapReduce Algorithm' was thus created. Cutting and Yahoo took help of the paper to implement an analytics solution. The project was fully implemented and passed on to Apache again. This is how Hadoop evolved into a complete framework.

FOUR DIMENSIONS OF BIG DATA

Gartner Group defines big data as information that spans four dimensions:

A. Volume
B. Velocity
C. Variety
D. Veracity



VOLUME

Enterprises are immersed with consistently growing data of assorted types, effortlessly storing up terabytes or even petabytes of information. Some of the examples are:

ISSN: 2581-4737

- Around 6 million people are using the digital media and it is estimated that about 2.5 quintillion bytes of data is being generated every day
- Turn more than ten terabytes of tweets created every day into product sentiment analysis.
- Analyze shipping patterns of more than twelve million packages every day as a method for enhancing flight paths.
- Analyze real-time financial market information as a method for detecting when trading opportunities exist.
- **Greater than 3,500 amount of big data stored across the world (in petabytes)**

VELOCITY

Velocity manages to what extent it takes to process approaching data. For example, an organization may need to completely process multi day's transactions before the following day begins. Other examples:

- Near about **Three Million Emails** sent **every second**.
- More than **Fifty Million Tweets** taken place **per day**.
- More than **Twenty Hours of Video** uploaded **every minute**
- Analyze more than five hundred million daily call centre interactions in real-time in order to predict customer churn faster

VERACITY

Veracity alludes to the steady flow of data and the speed at which it moves. Sometimes data travels through applications too quick to be processed by traditional RDBS. Sometimes seconds matter or insights have a very short shelf life. Some of the other examples are:

- 27% of businesses are not sure if the data they are working on is accurate.
- Logistics companies like railroad operators use thermometers, microphones, and ultrasound to capture data about their engines as a means of identifying equipment at risk for failure.
- Cable companies track media consumption and engagement, advertising, and customer retention, as well as the status of operations and infrastructure.

VARIETY

This is a big one, especially when it comes to understanding the broad depth that encompasses big data. Big data is any type of data. This includes structured data, as well as unstructured data like text, sensor data, audio, video, log files, and more. Some of the other examples are:

- People to People: through virtual communities, social networks, web logs
- People to Machine: through medical devices, ecommerce, bank cards, computers
- Machine to Machine: through sensors, GPS devices, bar code scanners, scientific research

ISSN: 2581-4737

- Algorithmic monitoring of surveillance cameras to identify target points of interest
- Attempt to enrich images with metadata using signals from a variety of different sources

Big Data if often hailed as a critical tool that provides competitive advantage, but make effective use of Big Data tools is real life business scenarios offers plenty of challenges.

BIG DATA CATEGORIES

Big data can be categories in three ways:

A. Structured

B. Unstructured

C. Semi-structured

STRUCTURED DATA

Data which is stored in a fixed format is called as Structured Data. In structured data the data is formatted so that it is easily accessible and can be used for analysis.

UNSTRUCTURED DATA

Any data whose structure is not classified is known as unstructured data. Unstructured data is very huge in size. Unstructured data usually consists of data that contains combination of text, images, files, etc. They do not use conventional database models.

SEMI-STRUCTURED DATA

It contains both structured as well as unstructured data. The data is not organized in a repository but has associated information which makes it accessible.

Young Scientist- Tomorrow's Science Begins Today
Vol. 1, Issue – 2016
ISSN: 2581-4737

WHY TO USE BIG DATA?

Big Data is important not in terms of volume but in terms of what you do with the data and how you utilize it to make analysis in order to benefit your business and organisation.

The reasons why every company is inclined towards adopting big data:

- **Timely**-Gain instant insights from diverse data sources
- **Better analytics**-Improvement of business performance through real-time analytics
- **Vast amount of data**-Big data technologies manage huge amounts of data
- **Insights**-Can provide better insights with the help of unstructured and semi-structured data
- **Decision-making**-Helps mitigate risk and make smart decision by proper risk analysis
- **Cost**
- **Product development**

Big data when teamed up with Analytics help you determine root causes of failure in businesses, analyse sales trends based on analysing the customer buying history. Also help determine fraudulent behaviour and reduce risks that might affect the organisation.

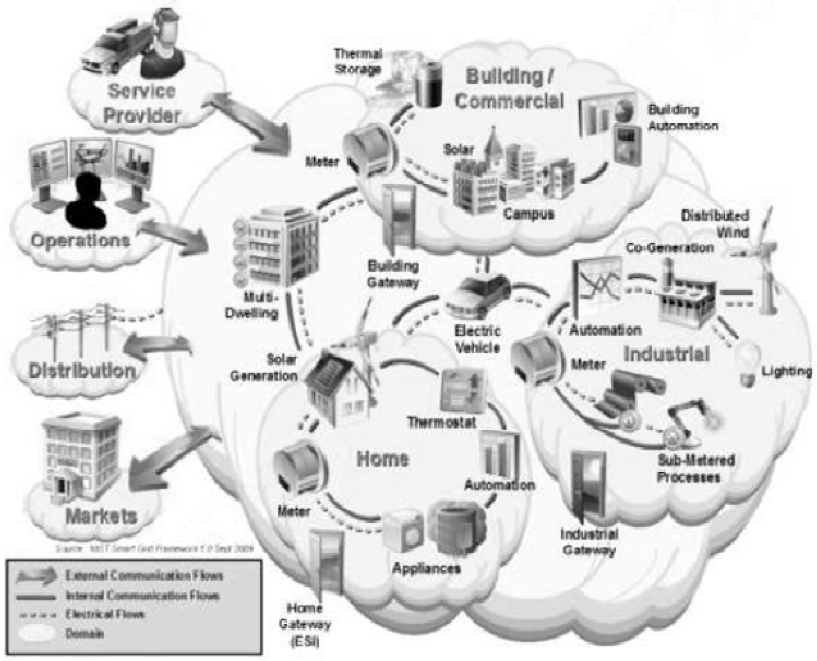
BIG DATA OPPORTUNITIES

Now let's discuss how big data is changing business –

IMPROVING BUSINESS EFFICIENCY

Big data is providing rich data about each product and process.

ISSN: 2581-4737



These data's are helping smart businesses to conquer the marketplace. Engineers are analyzing big data and searching for approaches to make forms run all the more effectively. When constraints are discovered and removed, the business performance increases.

PERSONALIZED MARKETING

Big data analysis can enable business to foresee what products customers may require in the future. Imagine how your business would benefit from being able to market the products that you knew your customers needed and knew enough information about them to customize for specific needs

ADVANCED BUSINESS INTELLIGENCE

Prior the ascent of big data, business intelligence was somewhat restricted. Incorporating advanced analytics for big data with business intelligence frameworks is an important step toward picking up a full return on investment.

Presently numerous organizations are adapting by employing business intelligence specialists since they help take an organization to the next level.

PROACTIVE CUSTOMER SERVICE

With big data, organizations can know precisely what their clients require before the client raises the query. This sort of proactive client service will revolutionize business that desires to differentiate themselves based on their customer service. Big data analysis can also enable client to help team to proactively contact clients on accounts where predictive analysis determines the customers might have made a purchase decision but in future, they can be a potential customer.

ISSN: 2581-4737

COST REDUCTION

Organizations are currently using big data to precisely discover trends and predict future events inside their respective ventures. Realizing when something might happen can enhance forecasts and planning. Planners can determine when to produce and how much to produce. They can even determine how much inventory to keep in hand.

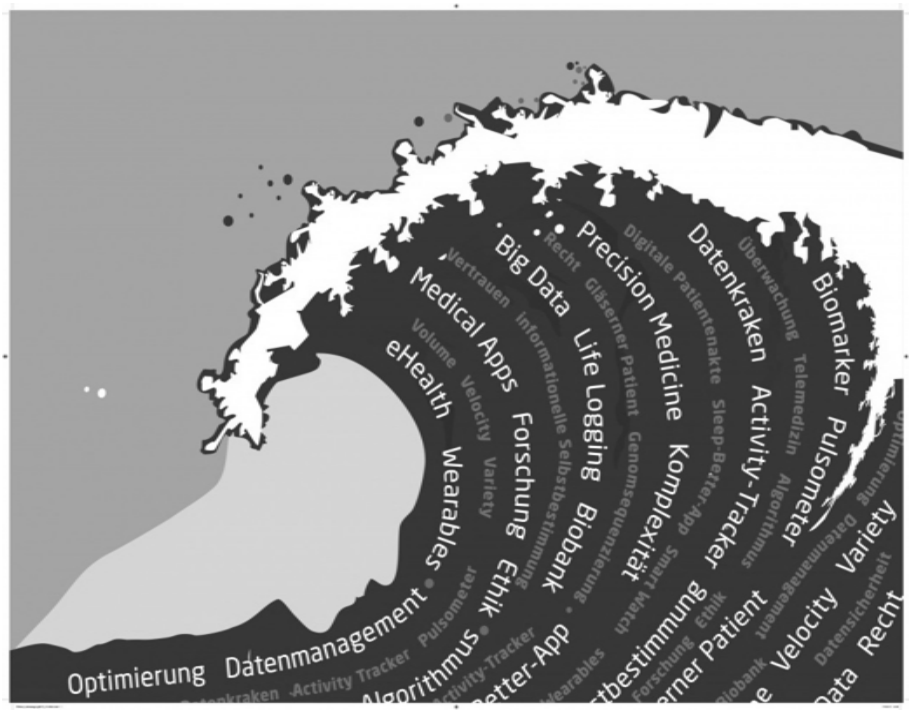
CUSTOMER RESPONSIVE PRODUCTS

Big data guarantees to enhance customer service as well as enables organizations to make customer responsive products. Product design can help in satisfying the necessities of customers in ways that have never been possible. Rather than clients disclosing to you what they are looking for in a product, you can use data analysis to predict what they are looking for in a product. Even this can help in creating a better picture of what a future product should look like.

According to estimates, it is revealed that the global market share of big data and its allied technology is expected to reach approximately $57 billion by the year 2020, generating a need for over 50,000 skilled Big Data experts and data scientists. The US is the forerunner in the generation, consumption and deployment of data and this technology. Thus the big data is connecting and building the digital future.

BIG DATA INNOVATIONS

Big Data technologies are very beneficial to the businesses in order to boost the efficiency and develop new data driven services.



ISSN: 2581-4737

There are a number of uses of big data. For example: in analysing a set of data containing weather report to predict the next week's weather. Some of big data innovations discussed below:

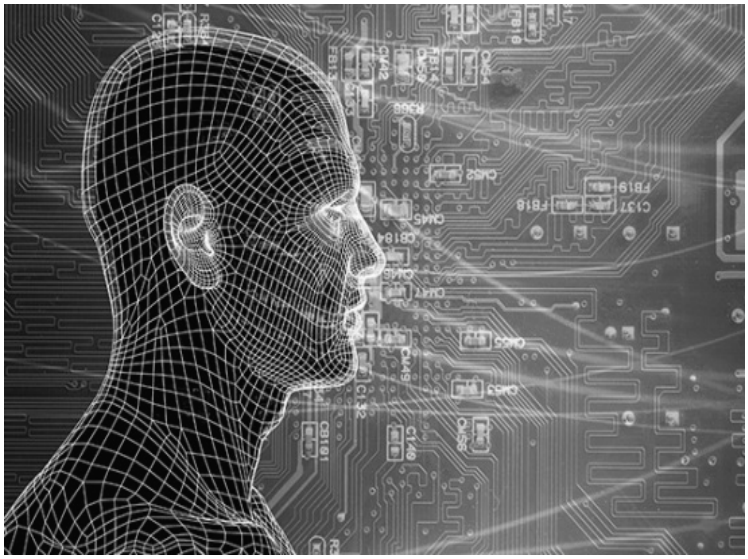
SOCIAL MEDIA ANALYSIS AND RESPONSE

The flood of updates from users that flow through social networking sites like Facebook, Twitter, Instagram and others are some of the most obvious examples of big data. Today, organizations are relied upon to monitor what people are saying about them in social media and respond appropriately and if they don't, they rapidly lose clients.

The data is collected and observed in the form of comments, images, social statuses, etc. This helps companies to analyse and come up strategies that will be beneficial for the company's growth

CONTRIBUTION OF BIG DATA IN HEALTH CARE

The contribution of Big Data in Healthcare domain has grown largely. With medical advances there was need to store large amount of data of the patients. Big data is used extensively to store the patients' health history.

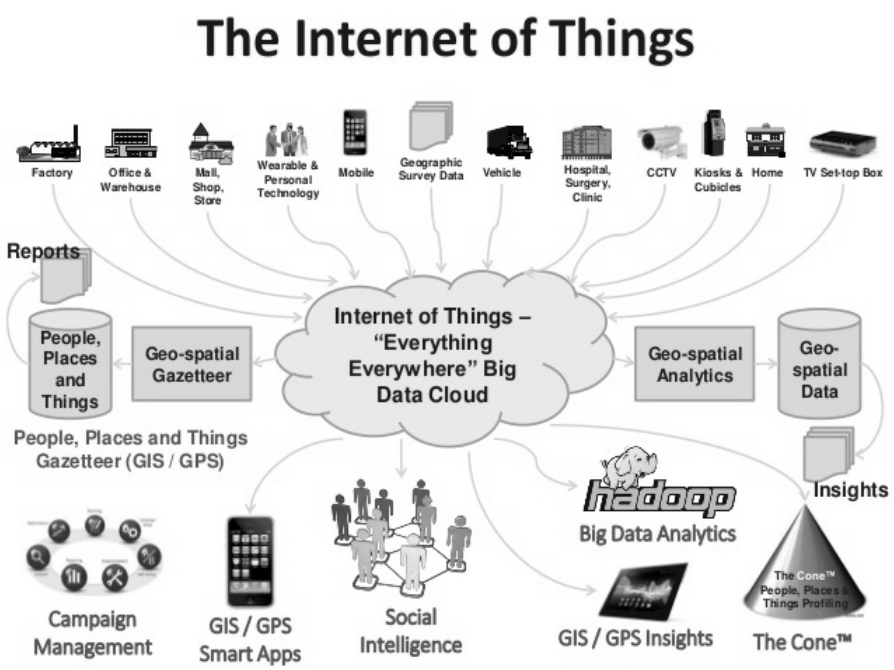


This data can be used to analyse the patients' health condition and to prevent health failures in future.

INTERNET OF THINGS

Enterprises in each industry are starting to explore the potential outcomes of the Internet of Things (IoT).

Young Scientist- Tomorrow's Science Begins Today
Vol. 1, Issue – 2016
ISSN: 2581-4737



As in the preventive maintenance example, they are using sensors to collect data that they can then analyze to achieve actionable insights. They might track customer or product movement, monitor the weather or keep an eye on security camera footage

FRAUD PREVENTION

Fraud detection and prevention is one of the many uses of Big Data today. Credit card companies face a lot of frauds and big data technologies are used to detect and prevent them.

Earlier credit card companies would keep a track on all the transactions and if any suspicious transaction is found they would call the buyer and confirm if that transaction was made. But now the buying patterns are observed and fraud affected areas are analysed using Big Data analytics. This is very useful in preventing and detecting frauds..

360° VIEW OF THE CUSTOMER

Numerous enterprises utilize big data to assemble a dashboard application that gives a complete 360° view of the clients.

ISSN: 2581-4737

These dashboards pull together data from a variety of internal and external sources analyze it and present it to customer service, sales and/or marketing personnel in a way that helps them do their jobs.

WEATHER

Big Data technologies are used to predict the weather forecast. Large amount of data is feeded on the climate and an average is taken to predict the weather This can be useful to predict natural calamities such as floods, etc.

SECURITY INTELLIGENCE

To prevent themselves from criminal action, organizations are also utilizing big data analytics to enable them thwart hackers and cyber-attackers. Working an endeavour IT division creates a huge amount of log data. Furthermore, cyber threat intelligence data is accessible from outside sources, such as law enforcement or security providers. Numerous organizations or now utilizing big data solutions to help them aggregate and analyze all of this internal and external information to enable them prevent, detect and mitigate attacks.

PUBLIC SECTOR

Big Data is used in a lot of government as well as public sectors. Big data provides a lot of facilities such as power investigation, economic promotion, etc.

Big Data is used in many other cases such as Education sector, Insurance services, Transportation, Security Intelligence, etc. Big data has become an important part for analysis and is needed in order to understand the growth of the businesses and build strategies to help it grow further.

ISSN: 2581-4737

DATA WAREHOUSE OFFLOAD

One of the easiest and potentially most cost-effective ways for organizations to start utilizing big data tools is to remove some of the portion from their data warehouses. Indeed, even among the couple of organizations that haven't yet started experimenting with big data analytics, it is common to have a data warehouse that facilitates their business intelligence (BI) efforts.

PRICE OPTIMIZATION

Both business-to-consumer (B2C) and business-to-business (B2B) enterprises are also utilizing big data analytics to optimize the costs that they charge their clients. For any organization, the objective is to set costs so that they expand their income. In case if the cost is too high, they will sell less products which may end up with decreasing their net returns. But if the price is too low, they may leave money on the table.

OPERATIONAL EFFICIENCY

Apart from helping companies optimize their pricing, big data analytics can also help organizations to distinguish other potential opportunities to streamline operations or maximize their profits. Frequently, this specific big data use case is the purview of BI or financial analysts.

RECOMMENDATION ENGINES

As for as popularity is concern, one of the most well-known use cases for big data is the recommendation engine.

When you are watching a movie at Netflix or looking for products from Flipkart, you most likely presently underestimate it that the website will propose similar items that you may appreciate. Obviously, the capacity to offer those suggestions emerges from the utilization of big data analytics to analyze historical data.

ISSN: 2581-4737

PREVENTIVE MAINTENANCE AND SUPPORT

Numerous big data use cases mentioned so far relate to retail or financial organizations, but businesses in manufacturing, energy, construction, agriculture, transportation and similar sectors of the economy can also benefited by big data. In these examples, some of the biggest benefit might come from using big data to improve equipment maintenance.

As with big data itself, the number of ways in which analytics can be applied to IoT solutions seems to be endless.

Other Innovations through Big Data:

- An example of big data might be petabytes (1,024 terabytes) or exabytes (1,024 petabytes) of data consisting of billions to trillions of records of millions of people-all from different sources (e.g. Web, sales, customer contact center, social media, mobile data and so on). The data is typically loosely structured data that is often incomplete and inaccessible.
- Twitter produces over 90 million tweets per day.
- eBay uses two data warehouses at 7.5 petabytes and 40PB as well as a 40PB Hadoop cluster for search, consumer recommendations, and merchandising.
- Walmart handles more than 1 million customer transactions every hour, which are imported into databases estimated to contain more than 2.5 petabytes (2560 terabytes) of data- the equivalent of 167 times the information contained in all the books in the US Library of Congress.

CHALLENGES FOR BIG DATA

- The biggest challenge with Big Data is, it can cause privacy problems, which can be seen, for instance, from the analysis of social networks, it can have our personal details.
  The moment you sign up for any social networking site your privacy is lost. You are and will be watched every time on social network.
  - Your smart phones get them your location
  - Your whatsapp get them chats about current issues.
  - Your blog gets them your views and describes your behaviour.
  - Your mail gets them your contacts and conversation.
  - Your status helps them to plot probability of NAMO becoming PM.
  - Your likes/dislikes make them improve their products.
    And so on….
- More often than not, Data Visualization applications cost a decent some of money, and it may not be possible for especially small companies to spend that much resources upon purchasing them.
- The rate of increase in data is much faster than the existing processing systems. The storage systems are not capable enough to store these data. There is a need to develop a processing system that not only caters to today's needs but also future needs.

ISSN: 2581-4737

- While organizing data there may be many unnecessary data points. The analyst should work hard to separate the wheat from the tares.
- The value of the data decreases over time. Most of the applications like fraud detection in telecom, insurance and banking, require real time or near real time analysis of the transactional data
- If a business isn't used to handling data at such a rapid rate, it could lead to incorrect analysis, which could cause larger problems for the organization.
- With increase in the growth of information technology, huge amount of data is generated. With advent of cloud computing for storage and retrieval of data, there is a need to utilize the big data. Large datasets from internet sources are prone to errors and losses, hence unreliable. The source of the data should be understood to minimize the errors caused while using multiple datasets. The properties and limits of the dataset should be understood before analysis to avoid or explain the bias in the interpretation of data
- With the increase in amount of (structured and unstructured) data generated, there is a need for talent. The demand for people with good analytical skills in big data is increasing. Research says that by end of 2018 itself, as many as 140,000 to 190,000 additional specialists in the area of big data may be required.

CONCLUSION

We have entered an era of Big Data. Through better analysis of the large volumes of data that are becoming available, there is the potential for making faster advances in many scientific disciplines and improving the profitability and success of many enterprises. However, many technical challenges described in this paper must be addressed before this potential can be realized fully. The challenges include not just the obvious issues of scale, but also heterogeneity, lack of structure, error handling, privacy, timeliness, provenance, and visualization, at all stages of the analysis pipeline from data acquisition to result interpretation. These technical challenges are common across a large variety of application domains, and therefore not cost effective to address in the context of one domain alone. Furthermore, these challenges will require transformative solutions, and will not be addressed naturally by the next generation of industrial products. We must support and encourage fundamental research towards addressing these technical challenges if we are to achieve the promised benefits of Big Data.

