ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICACTRP 2017

International Conference on Advances in Computational Techniques and Research Practices

Noida Institute of Engineering & Technology, Greater Noida

Vol. 6, Special Issue 2, February 2017

Big Data Analysis

Akanksha Paul[1], Harsh Raj Vardhan[2], Rishabh Kumar[3], Pragya[4]

Student, Department of Computer Science & Engineering, Buddha Institute of Technology[1,2,3]

Assistant Professor, Department of Computer Science & Engineering, Buddha Institute of Technology[4]

Abstract: The society is moving towards the use of instruments like computer and mobileand as aresult, organizations are producing and storing vast amounts of data. The data called as big data is captured, stored and processed for doing analytical analysis. These data is used for the future prediction by an organization. The big data is now not only confined to a single industry. Almost all organizations are involved in big data analysis. Big data is of vast amount and is so complex that it can't be processed using traditional datamanagement tools or processing applications. Data mining is performed to gain information from the unstructured voluminous data. This paper discuss about the formats of the big data that is available in the stock. It reveals about the stocked big data sources which can be the web, the social media data and the black box data and also that how complicated it is for the management to capture such a high speed voluminous data.Managing and gaining insights from the produced data is achallenge and key to competitive advantage. Analytics solutionsthat mine structured and unstructured data are important as theycan help organizations gain insights not only from their privatelyacquired data, but also from large amounts of data publicly availableon the Web.

Keywords: Analytical analysis, big data, structured data, unstructured data,big data analysis, black box, data mining.

I. INTRODUCTION

People are becoming social and trending towards more use of online networking. A large amount of bag data has burst in the last decade of 21st century. People are trending towards use of social media like Facebook, LinkedIn, Twitter, etc. Online retail sites like eBay, Flipkart, Amazon, etc are coming in use which are constantly generating a large amount of data [1]. People are habitual about using these sites which generates huge amount of data.

According to research, the size of big data in 2011 is 1.8 Zettabytes (1.8 trillion gigabytes) approximately. Till 2020 this data will increase 50 times more [2]. Managing and gaining insights from these produced data is challenging. The nature of big data is unstructured and are generated by modern technologies such as from that from web logs, radio frequency ID (RFID), sensors embedded in devices, machinery, vehicles, Internet searches, social networks such as Facebook, portable computers, smart phones and other cell phones, GPS devices, and call center records**[**3]. To use the big data efficiently, it is combined with structured data such as relational database.

Big data is not confined to a single industry. It is a combination of data management technologies that has been evolved from the last decades. Big data has enabled many of the companies to store, manage and manipulate vast amount of data and use it for the organizations improvement. The storage of bag data enables the organizations to understand the need of the customer [4]. For example, in manufacturing companies the big data helps a lot. It determines how the customer is reacting towards their products. It is also estimated that after some years what the customer will need. The big data is used for future prediction.

Many people think that big data is generally confined to social media data. However, big data not only include social media data. The big data includes black box data, social media data, stock exchange data, power grid data, transport data and search engine data[6].

A. Black Box Data

Black box is used in aircraft, helicopter, etc. The black box is a general term used for Flight Data Recorder and the Cockpit Voice Recorder of an aircraft. It includes recording of microphone for each pilot, the radio, attitude of the craft, engine performance, autopilot setting, etc.

B. Social Media Data

Social media like Facebook, Twitter, etc holds information about the views, news, likes, user information etc. YouTube stores big audio and video files.

C. Stock Exchange Data

Stock exchange data is market data which includes information about securities and stock trades from stock exchanges to stock brokers and stock traders.

D. Power Grid Data

The power grid data holds information consumed by a particular node with respect to a base station.

E. Transport Data

The transport data includes the details of vehicles, its capacity, distance travelled, availability of the vehicle and detail of the person who has hired the vehicle

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICACTRP 2017

International Conference on Advances in Computational Techniques and Research Practices

Noida Institute of Engineering & Technology, Greater Noida

Vol. 6, Special Issue 2, February 2017

F. Search Engine Data

Online search engines stores images, link data and metadata for the document.

Fig. 1Types of big data

II. LITERATURE SURVEY

Data has been around and there has always been a need of storage, processing and management of data. However, the amount and type of data captured, stored, processed and managed depends upon various factors including the necessity of humans, available tools and technologies for storage, processing, management, effort and cost, making decisions, and so on.

In the ancient periods, humans used primitive way to capture and store data like carving on stones, metal sheets, wood etc. As the centuries passed and due to inventions and advancements, human started storing of data and information on paper, cloth, etc. As time progressed, the medium of capturing, storing and management became punching cards followed by magnetic drums, laser disks, floppy disks, magnetic tapes and finally today we are storing data on various devices like USB drives, compact disks, hard drives, etc [8].

In fact, the capturing, storing and processing of data done by the human in past centuries has enabled human beings to pass the knowledge and research from one generation to another so that the next generation need not bother and try to search the same thing that has been invented in the earlier centuries. The next generation does not have to re-invent the wheel again. They can use the wheel invented in the earlier centuries.

We can clearly see this the amount of data storage has been increasing exponentially, and today with the help of cloud infrastructure one can store unlimited amount of data. Today, Terabytes and Petabytes of data is being generated, captured, processed, stored, and managed.

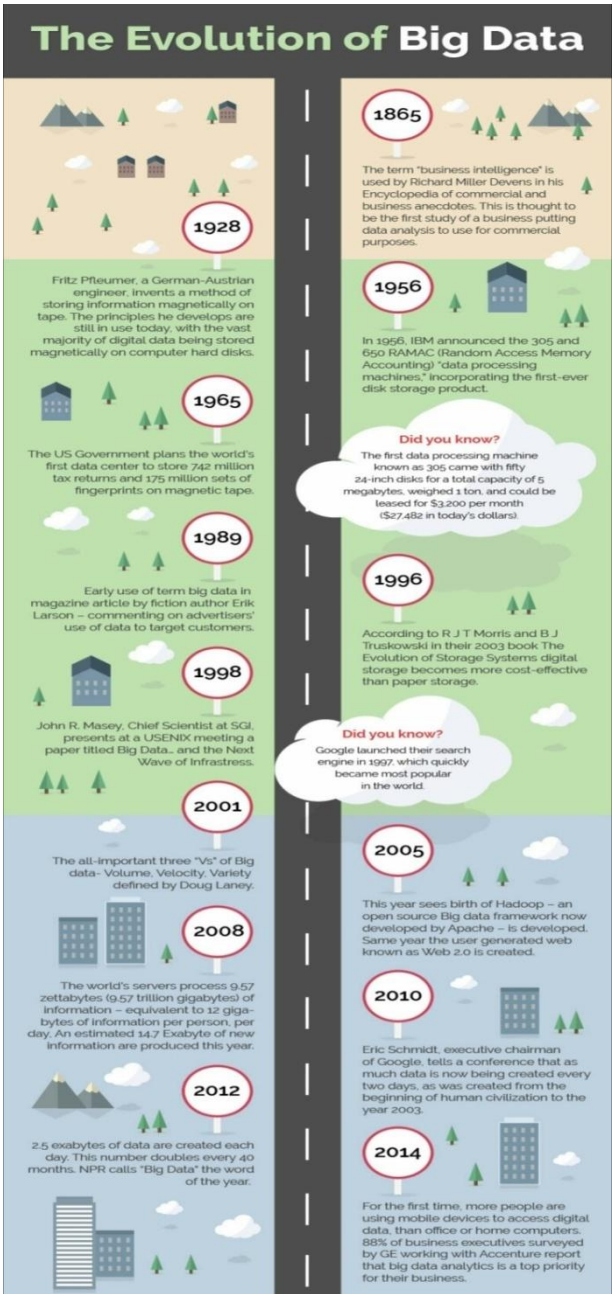

Fig. 2 The evolution big data

III.CHARACTERISTICS OF BIG DATA

As organizations are growing the data related to them also increases exponentially and today there is lots of complexity to their data. Most of the big organizations have data in multiple applications and in different formats. The data is spread so much that it is hard to categorize these data using single algorithm or mechanism. Big organizations are facing challenges to keep the data on a single platform.

Big data relates to data creation, storage, retrieval and analysis that are in terms of volume, velocity and variety.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICACTRP 2017

International Conference on Advances in Computational Techniques and Research Practices

Noida Institute of Engineering & Technology, Greater Noida

Vol. 6, Special Issue 2, February 2017

The 3Vs that define big data are Variety, Velocity and Volume [7].

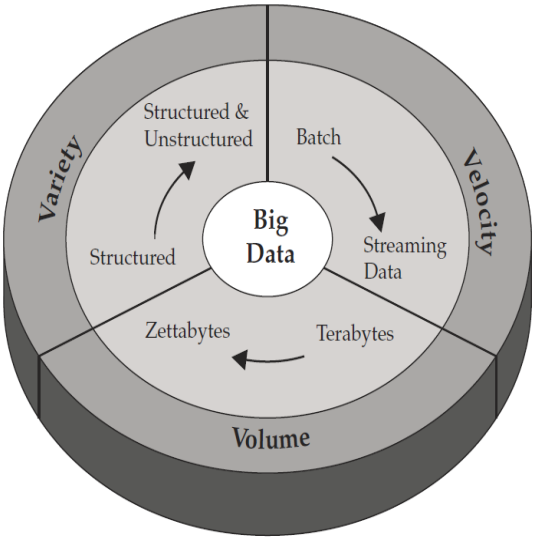

Fig. 3 Characteristics of big data

1) Variety: Variety refers to the different formats in which data is being generated and stored. Different applications generate and store data in different formats. For example, access file, text file, PDF, video file etc.
In today's world, there are large volumes of unstructured data is generated apart from structured data in big organizations.
Before the advancements in big data technologies, the industry didn't have any powerful and reliable tools and technologies which can work with huge volume of unstructured data that can be seen today. In today's world, organization needs not to rely on the structured data from enterprise databases.

The organizations need to store both inside and outside of an enterprise generated data. Apart from the traditional flat files, spreadsheets, relational database, etc., there are a lot of unstructured data such as images, audio files, video files, web logs, sensor data, and many more are stored [4].
The aspect of varied data formats is referred to as variety in the big data.

2) Velocity: Velocity refers to the speed at which the data is being generated. Different applications have different latency requirements and in today's competitive worlds, decision makers want the necessary and important data and information in the least amount of time as possible.
In different areas of technology, data is generated constantly at different speeds. For example, the likes, shares or post on Facebook or tweets on Twitter are generated with a high speed.
This speed aspect of data generation is referred to as velocity in the big data.

3) Volume:Volume refers to the size of data. With the advancements of technology and social media, the amount of data is growing with a very high rate.
This data is spread across different places, in different formats, in large volume i.e., in Gigabytes, Terabytes, Petabytes, and even more. Today the data is not only generated by humans but large amount of data is generated by machines.
The size aspect of data is referred to as volume of the big data.

IV. BIG DATA TECHNOLOGIES

For organizations of all sizes, data management has shifted from an important competency to a critical differentiator that can determine market winners. Organizations are defining new initiatives and re-evaluating existing strategies to examine how they can transform their business using big data.
Big data refers to technologies and initiatives and involve data that is too diverse, fast changing for conventional technologies.
But today, new technologies make it possible to realize value from big data. For example, retailers can track user web clicks to identify behavioural trends that improve campaigns, pricing and stock. Government and even Google can detect and track the emergence of disease outbreaks via social media signals. Oil and gas companies can take the output of sensors in their drilling equipments to make more efficient and safer drilling decisions.

For doing accurate analysis, big data technologies are important. These technologies provide more efficient data sets that can be used in decision making which will result in greater operational efficiencies, cost reductions, and reduced risks for the business.
There are various technologies are now in the market and many new are emerging. Some of these technology vendors are Amazon, Microsoft, IBM, etc.
The big data is dominated by two classes of technology systems that provide operational capabilities for real-time interactive workloads where data is primarily captured and stored and systems that provide analytical capabilities for respective, complex analysis that may touch most or all of the data. The classes of technologies are complementary and frequently deployed together [6].

A. Operational Big Data

For operational big data workloads, NoSQL big data system such as document databases have emerged to address a broad set of applications and other architecture such as graph databases are optimized for more specific applications. NoSQL technologies that are developed to overcome the shortcomings of relational databases in modern computing are faster and scale much more quickly and inexpensive than the relational databases.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICACTRP 2017

International Conference on Advances in Computational Techniques and Research Practices

Noida Institute of Engineering & Technology, Greater Noida

Vol. 6, Special Issue 2, February 2017

NoSQL big data systems are designed to take the advantage of new cloud computing architectures that have emerged over the past decade to allow vast and heavy computations to be run inexpensively and efficiently. This makes operational big data workloads much easier to manage, and cheaper and faster to implement.

To improve the user interactions with data, most operational systems need to provide insights into patterns and some amount of real-time intelligence about the active data in the system.

4) Advantages of operational data analysis:

- Understand how your products are used in the fields.
- Reduce service cost and chum.
- Enable value-added product innovation.

5) Benefits:

- Can enhance revenue and cut cost.
- Reduce cost of customer support and increase customer satisfaction.
- Optimize service offering according to consumption patterns.
- Ability to retain customers through understanding their experience.

B. Analytical Big Data

Analytical big data workloads include MRP (Massively Parallel Processing) database systems and MapReduce. They provide analytical capability for complex analysis that may touch most or all of the data.

These technologies are also a reaction to the limitations of traditional relational databases MapReduce provides a new method of analyzing data that is complementary to the capabilities provided by SQL.

6) Advantages of operational data analysis:

- Level-2 Heading:A level-2 heading must be in Italic, left-justified and numbered using an uppercase alphabetic letter followed by a period. For example, see heading “C. Section Headings” above.

C. Operational vs Analytical Systems

TABLE I OPERATIONAL VS ANALYTICAL SYSTEMS

Features	Operational	Analytical
Latency	1ms-100ms	1min-100min
Concurrency	1000-100,000	1-10
Access Pattern	Writes and reads	Reads
Queries	Selective	Unselective
Data Scope	Operational	Retrospective
End User	Customer	Data Scientist
Technology	NoSQL	MapReduce, MRP Database

V. SOURCES OF BIG DATA

As the varieties of formats of data are evolving, the sources of data are also evolving. With the advancement of the technology, the amount of the data generated from different sources is also increasing with a high rate [7].

Sources of big data can broadly be classified into six different categories.

Fig. 4 Sources of big data

7) Enterprise Data: There is vast amount of data that is generated by the enterprises. These data are in different formats including flat files, emails, word documents, spreadsheets, PDF documents, HTML pages or documents, etc.

These data are spread across an organisation. The different enterprise data are captured and valuable information is extracted from these data. Since, these are in high amount capturing, storing and managing of these data is tough enough.

Thus, the data which is spread across an organisation in different formats is referred as enterprise data.

8) Transactional Data: Every enterprise has some kind of applications which include different forms of transactions. These applications can be web applications, mobile applications, CRM systems, etc.

To support the transactions, a backend infrastructure which includes a relational database is required. The transactional data are structured data as these are stored in relational database. This is referred as transactional data.

9) Social Media Data: There is a large amount of data getting generated on social networks like user profiles, likes, shares, etc on Facebook, tweets on Twitter, etc.

The social media usually includes unstructured data formats which include text, images, audio files, video files, etc.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICACTRP 2017

International Conference on Advances in Computational Techniques and Research Practices

Noida Institute of Engineering & Technology, Greater Noida

Vol. 6, Special Issue 2, February 2017

This category of data source is referred to as social media data.

10) Activity Generated Data:There is a large amount data that is generated by the machines in comparison to the data generated by humans.

Activity generated data includes the data from medical devices, censor data, surveillance video, satellites data, cell phone towers data, industrial machinery data, and other data that is generated by the machines.

The type of data that is automatically generated by the machines is referred as activity generated data.

11) Public Data:Data which is publically available on web are public data. These data includes the data published by the government, research data published by research institutes, data from weather and meteorological department, census data, Wikipedia, open source data feeds, and other data which are freely available to the public.

12) Archive Data: Organisations have a lot of data as archive which are either not required anymore or very rarely required. Today, no organisation want to discard any data, they want to capture and store as much data as it is possible to do.

The archived data includes the scanned copy of document and agreements, records of ex-employees, completed projects, banking transactions, etc.

These data is less frequently accessed hence they are termed as archive data.

VI.FORMATS OF DATA

Data captured in an organisation are present in different formats. These captured data needs to be classified such that it become easy to understand the formats of data.

Thus the data is broadly classified into two categories: Structured Data and Unstructured Data [7].

A. Structured Data

Structured data refers to the data which has a pre-defined data model or schema. The structured data is stored in relational database.

The term structured data refers to big data of defined length and defined format. Structured data includes data in the relational databases, data from CRM systems, XML files, etc. According to the experts 20 percent of the total data is structured data. The data which user uses and generally deals with comes under structured data. The structured data is generally stored in database, commonly relational database [4].

Example of structured data includes numbers, dates and group of words and numbers called strings.

B. Sources of Structured Big Data

Structured data is taking a new role in the world of big data. As the new technologies are evolving new sources of structured data is being generated.

The sources of structured big data are divided into two categories- Machine generated and Human generated.

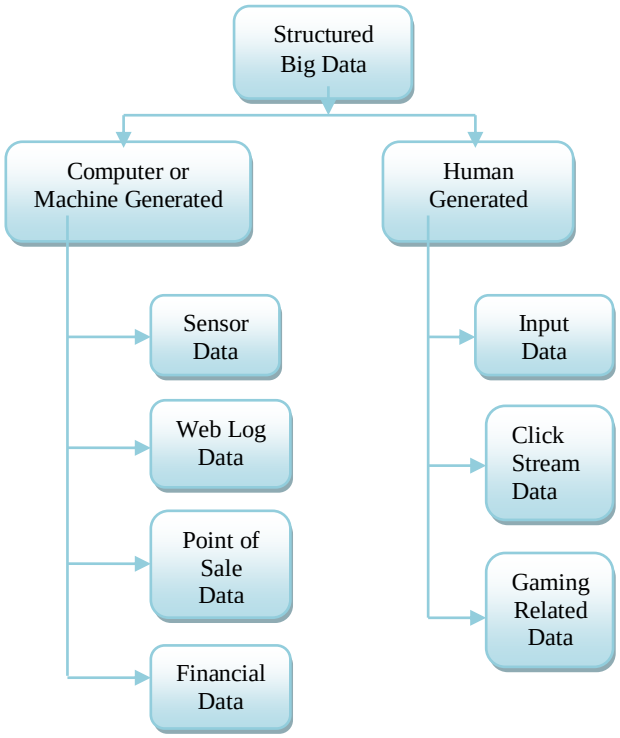

Fig. 5 Sources of structured data

13) Computer or Machine Generated Big Data:Machine generated data are those data which are generally automatically by the machine. Human is not involved in the generation of these data.

Some of the machine generated data are as follows:

a) Sensor Data: Sensor data includes radio frequency

ID tags, smart meters, medical devices and global positioning system data. Because of supply chain management and inventory control companies are interested in using sensor data.

b) Web Log Data:When severs, applications, networks,

and many more operate and perform their functionality the data is captured. The web log data is of huge amount which can be used for many things like to deal with service level agreements or to predict security breaches.

c) Point-of-sale Data: The bar code of the product is when swiped generates all the information related to the product.

d) Financial Data: Many of the financial systems are

programmatic. The financial systems operate on the program rules that run the process automatically. For example, stock trading data. It contains structured data such as the company symbol and dollar value. Some of this data is machine generated.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICACTRP 2017

International Conference on Advances in Computational Techniques and Research Practices

Noida Institute of Engineering & Technology, Greater Noida

Vol. 6, Special Issue 2, February 2017

14) Human Generated Big Data:The data which are generated by human in the interaction with the computers.

a) Input Data: The data which is input by the user into a computer is the input data. For example, name, age, income, and so on. This data can be used to understand the behaviour of the customer.

b) Click-stream Data:Each and every time when user clicks a link on a website, the data is generated. This data can be used by the organisations to determine the behaviour of customer and buying patterns.

c) Gaming-related Data: Whenever a user makes an action or move on a game, it is recorded. This can be useful in understanding how end users move through a gaming portfolio.

C. Unstructured Data

Unstructured data is the information or data that is not organised in a pre-defined data model like relational database. Unstructured information in heavy, but contains the data such as dates, numbers, facts, etc. This makes the unstructured data ambiguous. The ambiguity and irregularity in data make it difficult to understand using traditional approach [10].

Some techniques are used to make the unstructured information understandable. Techniques such as data mining, natural language processing (NLP) and text analytics provide different methods to find pattern and analyze the data to gain information. Unstructured information management architecture (UIMA) is a commonly used framework to process unstructured data to extract valuable information and create structured data about the information.

Example of unstructured data include book, journals, documents, metadata, health records, audio, video, images, files and unstructured text such as body of an email or web page.

D. Problem with Unstructured Data

It is possible to transform unstructured data into structured data. However, structured data is used to machine language which makes it easy to deal with using the computers whereas the unstructured data is usually for humans.

Email is an example of unstructured data because the inbox of the email is arranged by date, time or size. If the email would be fully structured, it would also be arranged by exact subject and content, with no deviation which is impractical because people are not focused to subject.

On the other hand, spreadsheets are a structured data. It can be quickly scanned for information because it is properly arranged in a relational database system.

The problem with the unstructured data is its volume. The unstructured data is voluminous, and most of the business interactions require huge amount of data to extract the useful and necessary data as in the web search engine. Because the information is so large, current data mining techniques often miss a piece of useful information that can be utilised efficiently to get the better result. The unstructured data is also ambiguous.

E. Difference between Structured and Unstructured Data

TABLE II STRUCTURED VS UNSTRUCTURED DATA

Features	**Structured Data**	**Unstructured Data**
Representation	Discrete - rows and column	Less defined boundaries and easily addressable
Storage	DBMS or file formats	Unmanaged file structures
Metadata	Syntax	Semantics
Integration Tools	ETL or ELT	Batch processing or manual data entry that involves codes
Standard	SQL, ADO.net, ODBC, etc.	Open XML, SMTP, SMS, CSV, etc.

The term “big data” is associated with unstructured data. Big data refers to large data sets that are difficult to analyze by using traditional tools. Big data includes both structured and unstructured data. However, 90 percent of the big data is unstructured data. Apart of structured and unstructured data, there is one more type of big data, semi-structured data.

F. Semi-Structured Data

Semi-structured data is a form of structured data but does not associated with relational database. It contains tags or markets to separate semantic element. Some types of data appear to be unstructured but are actually semi-structured such as XML.

Semi-structured data are increasingly occurring since the increasing use of internet where full-text documents and databases are not the only forms of data anymore, and different applications need a medium for exchanging information. Semi-structured data is often find in object oriented databases [9].

Semi-structured data include XML, email data interchange messages (EDI), web servers logs and search patterns, sensor data, etc.

G. Pros and Cons of Using Semi-Structured Data Format

1) Advantages: Programming persisting objects from their applications to a data need not to worry about object-relational impedance mismatch. They can often serialize objects via a light-weight library.

It supports nested or hierarchical data which represents complex relationship between entities.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICACTRP 2017

International Conference on Advances in Computational Techniques and Research Practices

Noida Institute of Engineering & Technology, Greater Noida

Vol. 6, Special Issue 2, February 2017

It support lists of objects simplifies data models by avoiding massive translations of lists into relational data model.

2) Disadvantages:The traditional relational data model has a ready-made query language, SQL.

VII. BENEFITS OF BIG DATA

A. Cost Reduction

Big data technologies like Hadoop and cloud-based analytics can provide cost advantages. Big enterprises are using big data techniques.

The big data helps to store vast amount of data at a single platform which is cost effective. The data extracted from different sources is stored in different databases. These captured data id intergraded and useful information is extracted. The big data techniques help in processing and storing vast amount of data in the data ware house. For example, companies are using Hadoop clusters for moving data to enterprise warehouses as needed for production analytical applications.

B. Faster, Better Decision Making

Analytics has improved the process of decision making. Large organizations are needs both faster and better decisions with big data.

For example, Caesars is a leading gaming company is now embracing big data analytics for faster decisions. The company has the data about the users from web click streams and real-time play in slot machines. Using the traditional approach, it is difficult to understand the user need, integrate those data and act on them in real-time.

By acquiring the Hadoop clusters and commercial analytics software it becomes easy to capture each and every tap of the user and use them to perform better decision making. As the amount of captured data increases, the decision making is better performed.

C. New Products and Services

The most interesting use of big data analytics is to create new products and services for customers. Online companies are doing it from many decades but now offline companies have also started doing the same.

The new products launched by the companies are based on the user experience. The user experience is collected from the big data. The big data captured from different devices are analysed and used to enhance user experience by launching new products that the user aspect.

VIII. BIG DATA CHALLENGES

The big data is used for different purpose. However, capturing these big data is not an easy task. The data sets which are captured are occurring with high speed. Capturing such high speed voluminous data and making them useful is a challenge for the companies [3].

A. Understanding and Utilizing Big Data

The main challenge of the companies is to understand the big data. Before dealing with the big data captured from different sources must be understandable by the companies.

Most of the big data captured is unstructured data that is difficult to understand. The data captured is ambiguous. After removing its ambiguity, the data is made understandable.

These types of analyses need to be performed on an ongoing basis as the data landscape changes at an ever-increasing rate.

B. New, Complex and Continuously Emerging Technologies

The technology is evolving day-by-day. As the new technologies are coming in trend the company has to know about these technologies.

In order to utilize big data is new to most organizations, it will be necessary for these organizations to learn about these new technologies at an ever-accelerating pace.

The firms are entering in the world of big data, the firm need to analyze the big data more efficiently which can be done by adopting new technologies.

C. Cloud Based Solutions

With the emergence of the trend of using big data, a new class of business software applications has emerged. By this the company is managed and stores the data at a global centre.

These solutions range from ERP, CRM, Document Management, Data Warehouses and Business Intelligence to many others. These solutions offer companies to use the big data flexibility and cost effectively as compared to the traditional view. It raises a new dimension related to data security.

D. Privacy, Security and Regularity Considerations

The voluminous and complex big data stored in a single platform captured from different sources become tougher to secure and prevent the privacy of the data.

The data as complex and of huge amount, the security of the data become a tough job to do. The firms needs to secure the data so that confidential and/or private business and customer data are not accessed by and/or disclosed to unauthorized parties.

In the regulatory area, the proper storage and transmission of personally identifiable information (PII), including that contained in unstructured data such as emails can be problematic and necessitate new and improved security measures and technologies.

It is very important for most forms to tightly integrate their big data, data security or privacy, and regulatory functions.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICACTRP 2017

International Conference on Advances in Computational Techniques and Research Practices

Noida Institute of Engineering & Technology, Greater Noida

Vol. 6, Special Issue 2, February 2017

E. Archiving and Disposal of Big Data

Overtime big data will lose its value to current decision-making. The big data is voluminous and varied in content and structure. That's why it is necessary to use new tools, technologies, and methods to archive and delete big data, without losingthe effectiveness big data.

F. The Need for IT, Data Analyst, and Management Resources

It is estimated that there is a need for approximately 140,000 to 190,000 more workers with "deep analytical" expertise and 1.5 million more data-literate managers, either retrained or hired.

Therefore, it is likely that any firm that undertakes a big data initiative will need to either retrain existing people, or engage new people in order for their initiative to be successful.

IX. CONCLUSIONS

The big data is generated in huge quantity by the human and the machine. The data is stored in different formats structured and unstructured data which is changed into structured format. These captured data is stored and managed for the future use. The data is used to analyze the user needs for particular applications.

Thus, from the above descriptions we can conclude that the big data helps the big organizations to gain information about the users, their response towards the applications which is analyzed to understand the user and their future need.

The data sets captured from the different sources is of huge amount and need to be protected so that any unauthorized party would not be able to view, read and use that data.

REFERENCES

[1] http://www.sas.com/resources/assest/Big-Data-in-Big-Companies.pdf .
[2] http://www.cse.wustl.edu/~jain/cse570-13/ftp/bigdata2.pdf
[3] http://www.navint.com/images/Big.Data.pdf
[4] http://eecs.wsu.edu/~yinghui/mat/courses/fall%202015/resources/Big%20for%20dummies.pdf
[5] http://www.planetdata.eu/sites/deafult/files/presentations/Big_Data_Tutorial_part4.pdf
[6] http://www.tutorialspoint.com/hadoop
[7] www.zdnet.com/article/top-10-categories-for-big-data-sources-and-mining-technologies/
[8] https://www.mssqltips.com/sqlservertip/3132/big-data-basics--part-1--introduction-to-big-data/
[9] https://en.m.wikipedia.org/wiki/Semi-structured_data
[10] https://en.m.wikipedia.org/wiki/Unstructured_data

BIOGRAPHIES

Rishabh Kumar is pursuing B.Tech from B.I.T GIDA Gorakhpur with Computer Science Stream and currently working as trainee in mobiloitte.

Harsh raj Vardhan is pursuing B.Tech from B.I.T GIDA Gorakhpur with Computer Science Stream and currently working as trainee in mobiloitte.

Akansha paul is pursuing B.Tech from B.I.T GIDA Gorakhpur with Computer Science Stream and currently working as trainee in mobiloitte.

Pragya Kamal completed M.Tech(IT) from M.M.M.U.T.Working in B.I.T,Gorakhpur for 1.5 years as Assistant Professor in department of Computer Science.