

A Survey Paper of Study about Big Data

N Pooja Choudhry¹, , Shruti Upadhyay² Rakesh Patel³ Prateek Kumar Singh⁴

Pooja Choudhry

B.E. Student

Department of CSE -Kirodimal Institute of Technology, Raigarh (C.G.)

Email-poojachoudhry333.pc@gmail.com

Shruti Upadhyay

B.E. Student

Department of CSE -Kirodimal Institute of Technology, Raigarh (C.G.)

Email-u.shruti.su@gmail.com

Rakesh patel

Assistant Professor

HOD, IT Department, CSVTU University, Chhattisgarh

Department of IT-Kirodimal Institute of Technology, Raigarh (C.G.)

Email-rakeshpatel.kit@gmail.com

Prateek Kumar Singh

Lecturer

Department of IT-Kirodimal Institute of Technology, Raigarh (C.G.)

Email-prateek.kitiraigarh@gmail.com

Abstract: Big data is the term for data sets so large and complicated that it becomes difficult to process using traditional data management tools or processing applications. In this paper there is a information about Big Data , characteristics , structure , Big Data Analytics , application of analytics and benefits of big data. This paper also revels the recent progress on big data.

Keywords: Big Data, Characteristics, Structure, Big Data Analytics, Applications, Benefits.

1. Introduction

Big data is the term for data sets so large and complicated that it becomes difficult to process using traditional data management tools or processing applications. 'Big Data' is similar to 'small data', but bigger in size. But having data bigger it requires different approaches:- Techniques, tools and architecture. Big Data generates values from the storage and processing of very large quantities of digital information that cannot be analyzed with traditional computing techniques. Big Data may well be the Next Big thing in the IT world. Big data burst upon the scene in the first decade of the 21st century. The first organization to embrace it was online and starts up firms like Google, eBay, Linked, and face book were built around big data from the beginning. Like many new information technologies, big data can bring about dramatic cost reductions, substantial improvements in the time required to perform a computing task or new product and product and service offerings.



Fig 1:- Big Data

Walmart handles more than 1 million customer transaction every hour. Facebook handles 40 billion photos from its user base. Decoding the human genome originally took 10 years to process; now it can be achieved in one week.



Fig 2:- customer transaction

The Three Characteristics of Big Data V3s:-

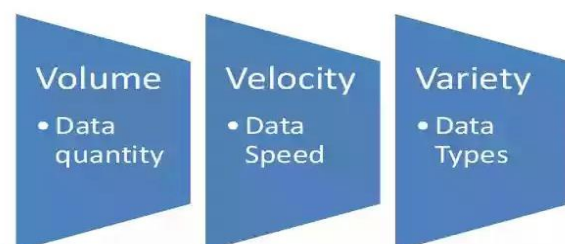


Fig 3:- Characteristics of Big Data

1st character of Big Data Volume:-

- A typical PC might have had 10 gigabytes of storage on 2000.
- Today, face book ingests 500 terabytes of new data every day.
- Boeing 737 will generate 240 terabytes of flight data during a single flight across the US.
- The smart phones, the data they creates and consumes; sensors embedded into everyday objects will soon result in billions of new, constantly-updated data feeds containing environmental, location, and other information, including video.

2nd character of Big Data Velocity:-

- Click streams and impressions capture user behaviour at millions of events per second.
- High-frequency stock trading algorithm reflects market changes within microseconds.
- Machine to machine processes exchange data between billions of devices.
- Infrastructure and sensors generates massive log data in real-time.
- On-line gaming system support millions of concurrent users, each producing multiple inputs per seconds.

3rd Character of Big Data Variety:-

- Big Data isn't just numbers, dates and strings. Big Data is also geospatial data, 3D data, audio and video, and unstructured text, including log files and social media.
- Traditional database systems were designed to address smaller volumes of structured data, fewer updates or a predictable, consistent data structure.
- Big Data analysis includes different types of data.

Why Big Data: - Growth of Big Data in needed

- Increase of storage capacities
- Increase of processing power
- Availability of data(different data types)
- Every day we create 2.5quintillion bytes of data; 90% of the data in the world today has been created in the last two years alone
- FB generates 10 TB daily
- Twitter generates 7TB of data daily
- IBM claim 90% of today's stored data was generated in just the last two years.

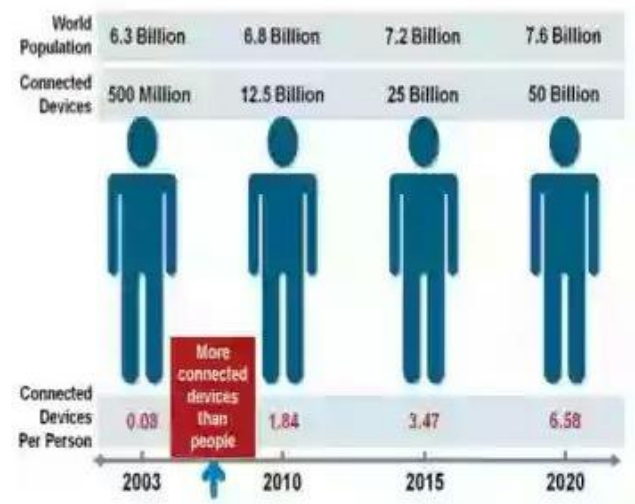


Fig 4:- Growth in Big Data

The Structure of Big Data:-

➤ Structured:-

For geeks and developers (not the same things ^^) Structured data is very banal. It concerns all data which can be stored in database SQL in table with rows and columns. They have relational key and can be easily mapped into pre-designed fields. Today, those data's are the most processed in development and the simplest way to manage information's. But structured data's represent only 5 to 10% of all informatics data's. So let's introduce semi structured data.

➤ Semi-Structured:-

Semi-structured data is information that doesn't reside in a relational database but that does have some organizational properties that make it easier to analyze. With some process you can store them in relation database (it could be very hard for some kind of semi structured data), but the semi structure exists to ease space, clarity or compute Examples of semi-structured: CSV but XML and JSON documents are semi structured documents, No SQL databases are considered as semi structured. But as Structured data, semi structured data represents a few parts of data (5 to 10%) so the last data type is the strong one: unstructured data.

➤ Unstructured:-

Unstructured data represent around 80% of data. It often includes text and multimedia content. Examples include e-mail messages, word processing documents, videos, photos, audio files, presentations, web pages and many other kinds of business documents. Note that while these sorts of files may have an internal structure, they are still considered « unstructured » because the data they contain doesn't fit neatly in a database.

Unstructured data is everywhere. In fact, most individuals and organizations conduct their lives around unstructured

data. Just as with structured data, unstructured data is either machine generated or human generated.



Fig 5:- Structure of Big Data

Big Data Analytics:-

Big data analytics is the process of examining large data sets containing a variety of data types -- i.e., big data -- to uncover hidden patterns, unknown correlations, market trends, customer preferences and other useful business information. The analytical findings can lead to more effective marketing, new revenue opportunities, better customer service, improved operational efficiency, competitive advantages over rival organizations and other business benefits. The primary goal of big data analytics is to help companies make more informed business decisions by enabling data scientists, predictive modelers and other analytics professionals to analyze large volumes of transaction data, as well as other forms of data that may be untapped by conventional business intelligence (BI) programs. That could include Web server logs and Internet click stream data, social media content and social network activity reports, text from customer emails and survey responses, mobile-phone call detail records and machine data captured by sensors connected to the Internet of Things. Some people exclusively associate big data with semi-structured and unstructured data of that sort, but consulting firms like Gartner Inc. and Forrester Research Inc. also consider transactions and other structured data to be valid components of big data analytics applications Application of big data analysis.



Fig 6:- Application of Big Data

Future of Big Data:-

- \$15 billion on software firms only specializing in data management and analytics.
- This industry on its own is worth more than \$100 billion and growing at almost 10% a year which is roughly twice as fast as the software business as a whole.
- In February 2012, the open source analyst firm Wiki on released the first market forecast for Big Data, listing \$5.1B revenue in 2012 with growth to \$53.4B in 2017.
- The McKinsey Global Institute estimates that data volume is growing 40% per year, and will grow 44X between 2009 and 2020.

How Big Data impact on IT:-

- Big data is troublesome force presenting opportunities with challenges to IT organizations.
- By 2015 4.4 million IT jobs in Big Data; 1.9 million is in US itself.
- India will require a minimum of 1 lakh data scientists in the next couple of year in addition to data analysts and data managers to support the Big Data space.

Benefits of Big Data:-

- I. Real Time big data isn't just a process for storing peta bytes or exabytes of data in a data warehouse, it's about the ability to make better decisions and take meaningful action at the right time.
- II. Fast forward to the present and technologies like Hadoop give you the scale and flexibility to store data before you know how you are going to process it.

- III. Technologies such as Map Reduce, Hive and Impala enable you to run queries without changing the data structure underneath.
- IV. Our newest research finds that organization is using big data to target customer-centric outcomes, top into internal data and build a better information ecosystem.
- V. Big Data is already an important part of the \$64 billion database and data analytics market.
- VI. It offers commercial opportunities of a comparable scale to enterprise software in the late 1980s.
- VII. And the internet boom of the 1990s, and the social media explosion of today.

Prateek Kumar Singh Working as a Lecturer
Department of IT-Kirodimal Institute of Technology,
Raigarh (C.G.)
Email-prateek.kitraigarh@gmail.com

Conclusion:-

In this research paper I have gained information about the Big Data, different characteristics of data which are also known as 3vs and different structures of data. There is also information about Big Data Analytics and application of analytics, benefits of big data. Big data has a growing relevance in the modern information society and specially IT society. It represents the key technology as part of the industry for example Sensor embedded in an engine and so on. Uses of Big Data analytics in Smart Healthcare, Manufacturing, Homeland Security, and Traffic control, Search Quality etc. Simply I can say that Big Data play an important role in human life.

References

- [1] www.Slideshare.com
- [2] www.wikipedia.com
- [3] www.wikipedia.com
- [4] http://en.wikipedia.org/wiki/Big_data
- [5] www.slideshare.net/rahulaga/big_data
- [6] <http://jeremyronk.wordpress.com>
- [7] www.cse.wustl.edu/~jain/cse570-13/ftp/bigdata2/index.html
- [8] http://Searchbusinessanalytics.techtarget.com/definition/big_data_analytics

Author Profile

Pooja Choudhry B.E Student
Department of CSE -Kirodimal Institute of Technology,
Raigarh (C.G.)
Email-poojachoudhry333.pc@gmail.com

Shruti Upadhyay B.E Student
Department of CSE -Kirodimal Institute of Technology,
Raigarh (C.G.)
Email-u.shruti.su@gmail.com

Rakesh Patel Working as an HOD & Assistant Professor,
Department of IT-Kirodimal Institute of Technology,
Raigarh (C.G.)
Email-rakeshpatel.kit@gmail.com