GJCRT : Global Journal of Computer Science Research & Technology
ISSN (online): 2349- 0608

Big Data: Challenges and Analytics

Meenakshi Garg
(Research Scholar)
Assistant Prof.,Govt. Bikram College of Commerce, Patiala.
meenagarg2000@yahoo.com

ABSTRACT
"Big Data has become a highlighted buzzword from last few years. Data is called "Big Data" because of its large Volume (in Exabyte and beyond), Veracity, Velocity and Variety. Most of this data is unstructured or semi structured and divergent in nature. The volume and the divergent of data with the speed it generated, makes it difficult to manage with current technologies. Usage of InfoTech has created huge challenges as well as opportunities for data storage and management to make data as a resource which is available and actionable. The volume of data exceeds the capacity of current on-line storage system and processing systems. So, Storage and data transport are technological issues. This paper presents a broad overview of big data, its current status, challenges, and forecast to the future.

***KEYWORDS*:** Big Data, Big Data Processing, Challenges and analytics

INTRODUCTION

"Big Data" refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage process and access it. From the last few years there is a huge increase in our ability to collect data from independent or connected applications. This data flood has outrun our capability to process, analyze, store and interpret these datasets. Consider the Internet data. This data comes in the form of digital pictures, posts to social media sites, intelligent sensors, cell phones etc. In Every Minute:-

- **Email** users post more than 204 million messages
- **Mobile Web** get 217 new users
- **Google** get over 2 million search queries
- **YouTube** users transfer 72 hours of video content
- **Facebook** users share 684,000 bits of content
- **Twitter** users send beyond 100,000 tweets
- **Consumers** spend $272,000 on Web shopping
- **Apple** receives around 50,000 application downloads
- **Brands** receive more than 34,000 Facebook 'likes'
- **Instagram** users share 220,000 new photos
- **Flickr** users add 3,125 new photos
- **WordPress** users publicize close to 350 new blog posts

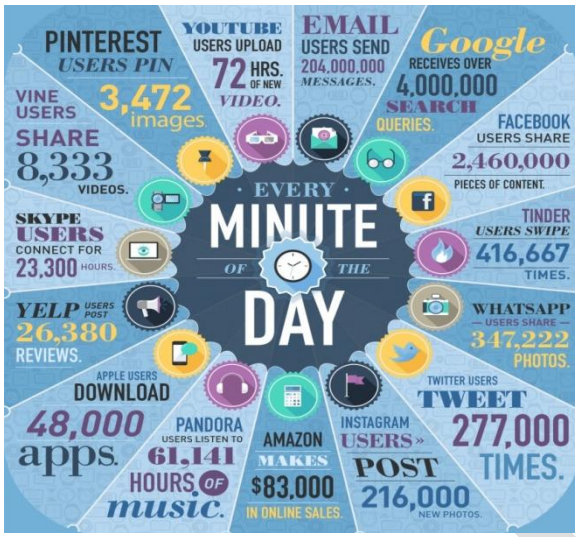

Figure1:-Data Produced in Every Minute by Different sources [2]

Google plays a big role in helping people find online content and data and its growth was visualized in the Amazing facts Google inforgraphic from Promodo last year. Google processes 20 petabytes of information per day. The WordPress a Global Phenomenon inforgraphic from Yoast reported that over 72 million websites were built with WordPress in 2012. By January 2014 that number climbed over 86 million. This type of data requires a different processing approach called big data.

STEP OF BIG DATA PROCESS

- **Collect**: Data is gathered from the data sources and distributed across multiple nodes – often a grid – each of which processes a subset of data in parallel.
- **Process**: The system then uses that same high-powered parallelism to perform fast computations against the data on every node. Next, the nodes shorten the resulting data findings into more consumable data sets to be used by either a human being (in the case of analytics) or machine (in the case of large-scale interpretation of results).
- **Manage**: Often the big data being processed is heterogeneous, originating from distinct systems. Nearly all of that data needs to be implied, defined, cleansed and audited for security purposes.
- **Measure**: Companies will often measure the rate at which data can be integrated with records, and whether the rate of integration or correction is increasing over time. Business requirements should actuate the type of measurement and the ongoing tracking.
- **Consume**: The resulting use of the data should fit in with the original requirement for the processing. For instance, if carrying a few hundred terabytes of social media interactions demonstrates whether and how social media data delivers additional product taken, then there should be rules for how social media data is accessed and amend. This is equally important for machine-to-machine data transfer.
- **Store**: As the "data-as-a-service" trend gets shape, increasingly the data stays in a single location, while the programs that access it rotate around. Whether the data is stored for short-term batch processing or longer-term, storage solutions should be deliberately addressed.
- **Govern**: Data governance enclose the policies and oversight of data from a business angle. As defined, data governance applies to each of the six preceding stages of big data. By establishing processes and guiding rules, governance sanctions behaviors around data. And big data needs to be commanded according to its expected consumption.

CHALLENGES OF BIG DATA

- **Storage**
  Clearly not enough hard disk/ devices. Distributed storage is still not enough, Manufactures cannot make enough storage devices in time. Speed in writing to devices, bigger data paths/ data-bus.

- **Processing**
  - Integrated data using Filters
  - "What" Data and "How"?
  - Useful Data processing system Design
  - Power?
  - Latency and Bandwidth
  - Taxonomy and Ontology
  - How to analyze big data-No standard way of doing that yet
- **Security/Privacy**
  What is to be assure the Data sources? As IT logs are also now a source of big data.
- **Data Policies**
- e.g. storage, computing, analytical software
  - E.g. new types of analyses
- **Technology and techniques**
  E.g. Privacy, security, intellectual property
- **Access to Data**
  E.g. integrate multiple data sources

BIG DATA ANALYTICS

Big data analytics refers to the process of gather, organizing and analyzing huge sets of data i.e. big data to discover patterns and other advantageous information. Big data analytics will help organizations to better understand the information contained within the data and will also help identify the data that is most important to the business and for future decisions in business. Big data analysts basically want the knowledge that comes from analyzing the data. For most organizations, big data analysis is a challenge. Because considering the very large volume of data ,different formats of data (structured and unstructured) data collected across the entire organization with many different ways. So different types of data can be combined, contrasted and analyzed to find patterns and other useful information.

- The first challenge is in breaking down data silos to access all data an organization stores in different places and often in different systems.
- A second big data challenge is in creating platforms that can pull in unstructured data as easily as structured data. This massive volume of data is typically so large that it's difficult to process using traditional database and software methods.[4]

The following are the six big trends in Big Data Analytics:-[3]

- **Big data analytics in the cloud**

Hadoop, a framework and set of tools for processing very large data sets, was by origin designed to work on clusters of physical machines."Now lots of technologies available for data processing in the cloud," For Example Amazon's hosted BI data warehouse, Google's Big Query data analytics service, IBM's Bluemix platform and Amazon's Kinesis data processing. "The coming state of big data will be a hybrid of on-premises and cloud".

- **Hadoop: The new data operating system**

Distributed analytic frameworks like as MapReduce, are evolving into distributed resources that are gradually turning Hadoop into a general-purpose data operating system. With these systems we can perform many different data manipulations and analytics operations by plugging them into Hadoop as the distributed file storage system.

- **Big data lakes**

A data lake, also called an enterprise data lake or enterprise data hub, turns that model on its head it provides tools for people to evaluate the data, along with a high-level definition of what data exists in the lake. It's a very incremental model for building a large-scale database.

- **More predictive analytics**

With big data, analysts have not only more data to work with, but also the power to handle large numbers of records with many attributes. Traditional machine learning uses statistical analysis based on a sample of a total data set. Now we have the ability to do very large numbers of records and very large numbers of attributes per record and that increases predictability.

- **Deep learning**

Deep learning, a set of machine-learning techniques based on neural networking, is still growing but shows great potential for solving business issues. "Deep learning, empower computers to recognize items of interest in large quantities of unstructured and binary data and to deduce relationships without needing specific models or programming instructions".

- **In-memory analytics**

The use of in-memory databases to speed up analytic processing is increasingly popular and highly beneficial in the right setting. Hybrid transaction/analytical processing (HTAP) — allowing transactions and analytic processing to reside in the same in-memory database.

CONCLUSION

This paper presents the fundamental view of Big Data. This view include the increase in data. The amount of information and knowledge that can be extracted from the digital universe is continuing to expand as users come up with new ways to massage and process data. As Facebook and Twitter are producing, collectively, around 50 gigabytes of data per day, and three times every year, within a few years. We are indeed facing the challenge of "big data becoming really very big data". And big data analytics that is the process examining big data to uncover hidden patterns and important information that can be used to make better decisions. And also presents some trends in big data analytics.

REFRENCES

[1] www.sas.com/en_us/news/sascom/2012q4/big-data-delivery.html

[2] http://aci.info/2014/07/12/the-data-explosion-in-2014-minute-by-minute-infographic/

[3] www.computerworld.com/article/2690856/8-big-trends-in-big-data-analytics.html

[4]www.webopedia.com/TERM/B/big_data_analytics.html

[5] Wei Fan, Albert Bifet " Mining Big Data: Current Status, and Forecast to the Future" Volume 14, Issue 2

[6] B R Prakash, Dr. M. Hanumanthappa "Issues and Challenges in the Era of Big Data Mining" Volume 3, Issue 4 July-August 2014

[7] M.H.Padgavankar, Dr.S.R.Gupta "Big Data Storage and Challenges" (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 2218-2223

[8] "Big Data: Challenges and Opportunities" International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-4 Issue-2, July 2014