

Overview of Big Data

Tushar B. Kute,
<http://tusharkute.com>

Data and Big Data

- “90% of the world’s data was generated in the last few years.”
- Due to the advent of new technologies, devices, and communication means like social networking sites, the amount of data produced by mankind is growing rapidly every year.
- The amount of data produced by us from the beginning of time till 2003 was 5 billion gigabytes. If you pile up the data in the form of disks it may fill an entire football field.
- The same amount was created in every two days in 2011, and in every eight minutes in 2015. This rate is still growing enormously.

What is Big Data

- Big Data is a collection of large datasets that cannot be processed using traditional computing techniques.
- It is not a single technique or a tool, rather it involves many areas of business and technology.

What Comes Under Big Data?

- **Black Box Data:** It is a component of helicopter, airplanes, and jets, etc. It captures voices of the flight crew, recordings of microphones and earphones, and the performance information of the aircraft.
- **Social Media Data:** Social media such as Facebook and Twitter hold information and the views posted by millions of people across the globe.
- **Stock Exchange Data:** The stock exchange data holds information about the 'buy' and 'sell' decisions made on a share of different companies made by the customers.
- **Power Grid Data:** The power grid data holds information consumed by a particular node with respect to a base station.

What Comes Under Big Data?

- **Transport Data:** Transport data includes model, capacity, distance and availability of a vehicle.
- **Search Engine Data:** Search engines retrieve lots of data from different databases.
- **Structured data:** Relational data.
- **Semi Structured data:** XML data.
- **Unstructured data:** Word, PDF, Text, Media Logs.

Benefits of Big Data

- Using the information kept in the social network like Facebook, the marketing agencies are learning about the response for their campaigns, promotions, and other advertising mediums.
- Using the information in the social media like preferences and product perception of their consumers, product companies and retail organizations are planning their production.
- Using the data regarding the previous medical history of patients, hospitals are providing better and quick service.

Big Data Technologies

- Operational Big data
- Analytical Big data

Operational Big Data

- These include systems like MongoDB that provide operational capabilities for real-time, interactive workloads where data is primarily captured and stored.
- NoSQL Big Data systems are designed to take advantage of new cloud computing architectures that have emerged over the past decade to allow massive computations to be run inexpensively and efficiently.

Analytical Big Data

- These includes systems like Massively Parallel Processing (MPP) database systems and MapReduce that provide analytical capabilities for retrospective and complex analysis that may touch most or all of the data.
- MapReduce provides a new method of analyzing data that is complementary to the capabilities provided by SQL, and a system based on MapReduce that can be scaled up from single servers to thousands of high and low end machines.

Challenges in Big Data

- The major challenges associated with big data are as follows:
 - Capturing data
 - Curation
 - Storage
 - Searching
 - Sharing
 - Transfer
 - Analysis
 - Presentation

Thank you

This presentation is created using LibreOffice Impress 4.2.7.2, can be used freely as per GNU General Public License

Web Resources

<http://mitu.co.in>
<http://tusharkute.com>

Blogs

<http://digitallocha.blogspot.in>
<http://kyamputar.blogspot.in>

tushar@tusharkute.com