**BIG DATA**

# Topics

* Introduction to BIG Data
* Importance of BIG Data
* Technologies used in BIG Data Analysis

Introduction to Big Data

Broadly, we can define BIG Data, in terms of Vs, these are **Volume, Variety, Velocity and Veracity** Lets discuss, each of them one by one.

First of all its volume, obviously, we can define this BIG Data with a large sized data and we measure this volume in zeta bytes. Let’s us take an example of an International flight, that contains a lot of sensors, for measuring temperature, carbon di oxide level, speed of plane etc and they do generate more than 10 TB of data for every 30 min, so for the full round of a flight, it will be very much difficult for traditional computers to process the data. Since data is not only large in size but it is complex also. By complexity of data, we mean that data might be consists of Structured or Unstructured sets. These structured set of data are those, which can be easily stored in to rows and column formats and unstructured data, consists of data that do not possess any kind of structure, just like our photographs, videos, presentations etc. Similar kind of data collection can be experienced by various social sites and telecom and IT companies for getting customer behavior. Moreover, this BIG Data is about the data complexity along with the size of the data.

Now next point is about Variety of Data, we know very well that we post a lot of photographs, videos etc. on various social sites. These consist of Unstructured data, and specific techniques are required to process and analyze the same.

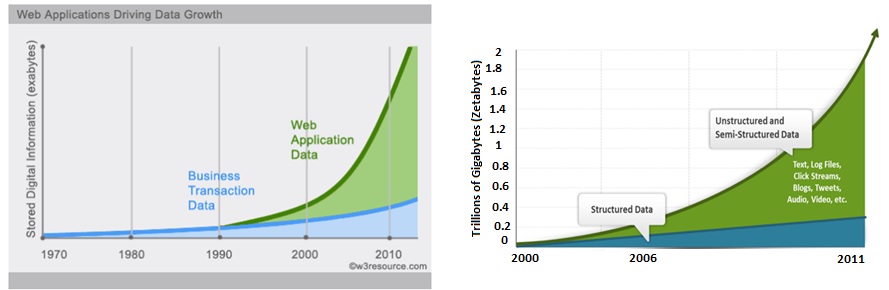
Now if we discuss about Velocity, this is something related to the incoming and outgoing of the data. Let’s take an example of Stock Market or some financial data, there in we have a couple of algorithms, that have to analyze millions of records not in hours, but in seconds, for the smooth running of the business. This is actually a very high rate of data processing, which cannot be done by our traditional databases, more over after data processing, one need to check about the analytics of data as well. And all this is supposed to be done at extremely fast speed.

At last we have something called as Veracity, by this we mean, like how one can trace out the meaningful data from the heap. Let’s consider an example of an e-commerce website or any social website, there in you can find links such as product you may be interested in or people you may know. This is all dependent on you past searches and about our profile.



Now due to advancement in Information Technology and development of a lot of hand held computing devices, we are getting huge amount of data since last couple of years. And this is incrementing at an exponential rate. And it contains very large amount of unstructured data. Due to this, by now out of total data, more that 80 % of the data is Unstructured one and that is still to be analyzed.

Every day social networking sites like facebook are dumping more than 700 TB of data, huge amount of data transaction takes place at stock markets. IT sector that revolutionized almost all domains of our lives including Banking, retails, telecom and so on. Now the great challenge is: - **HOW TO PROCESS THIS HUGE AMOUNT OF DATA**.



**Data warehousing and analytics**



Importance of Big Data

Now why Big Data is so important? The answer is Analytics, there is no use of collecting or getting the huge amount of data, if it is not going to be utilized in commercial / scientific or industrial terms. Analysis of this huge amount of data is required to add value to your business, lets discuss our previous example, lets say if a flight collecting the huge amount of data, we must analyze that data in order to get some meaningful information like, what we can do to save fuel or how one can make some arrangements to prevent accidents before they occur or if we discuss about retail sector, how one can predict the mood and requirements of the customer.

This Big Data Technology can be used for a variety of different purposes, just say for Big Science like the experiment for Large Hadron Collider, Universe simulation, Gene sequencing etc. Under these scenarios, we need to analyze millions of records to be analyzed at a very fast rate.

Another example can be called as for Big Government. Under this case some governments, regularly trace social media accounts, internet activities of suspected people, just in order to prevent terrorist attacks and for the security features as well. Indian Government is also working over a famous Aadhar Card project to provide good governance, insurance, account facilities to Indian citizens.

However, these are just a few examples and based on them, one can easily confirm, that Big data is going to play a vital role in context of upcoming requirements.

Technologies used in BIG Data Analysis

Now from Engineering point of view, the Big Data Analysis, requires a couple of technologies as mentioned below:-

* Hadoop and MapReduce Framesork
* Distributed File system
* Distributed Databases
* NoSQL technologies

We will study all of them one by one during our training course.