BIG DATA

What is Big Data?

There is no place where Big Data does not exist. The curiosity about what is Big Data has been soaring in the past few years. Forbes reports that every minute, users watch *4.15 million YouTube videos*, send *456,000 tweets* on Twitter, post *46,740 photos* on Instagram and there are *510,000 comments* posted and *293,000 statuses* updated on Facebook!

Just imagine the huge chunk of data that is produced with such activities. This constant creation of data using social media, business applications, telecom and various other domains is leading to the formation of Big Data.

Big Data refers to the large amounts of data which is pouring in from various data sources and has different formats. Even previously there was huge data which were being stored in databases, but because of the varied nature of this Data, the traditional relational database systems are incapable of handling this Data. Big Data is much more than a collection of datasets with different formats; it is an important asset which can be used to obtain enumerable benefits.

The three different formats of big data are:

1. *Structured:* Organised data format with a fixed schema. Ex: RDBMS
2. *Semi-Structured:* Partially organised data which does not have a fixed format. Ex: XML, JSON
3. *Unstructured:* Unorganised data with an unknown schema. Ex: Audio, video files etc.

Following are the characteristics associated with Big Data:

I. Validity: correctness of data
II. Variability: dynamic behaviour
III. Volatility: tendency to change in time
IV. Vulnerability: vulnerable to breach or attacks
V. Visualization: visualizing meaningful usage of data

Basically, Big Data Analytics is largely used by companies to facilitate their growth and development. This majorly involves applying various data mining algorithms on the given set of data, which will then aid them in better decision making.

There are multiple tools for processing Big Data such as *Hadoop*, *Pig*, *Hive*, *Cassandra*, *Spark*, *Kafka*, etc. depending upon the requirement of the organisation.

Scope of Big Data

- *Numerous Job opportunities***:** The career opportunities pertaining to the field of Big data include, Big Data Analyst, Big Data Engineer, Big Data solution architect etc. According

to IBM, 59% of all Data Science and Analytics (DSA) job demand is in Finance and Insurance, Professional Services, and IT.

- **Rising demand for Analytics Professional:** An article by Forbes reveals that "IBM predicts demand for Data Scientists will soar by 28%". By 2020, the number of jobs for all US data professionals will increase by 364,000 openings to 2,720,000 according to IBM.
- **Salary Aspects:** Forbes reported that employers are willing to pay a premium of $8,736 above median bachelor's and graduate-level salaries, with successful applicants earning a starting salary of $80,265
- **Adoption of Big Data analytics:** Immense growth in the usage of big data analysis across the world.

Topics Covered in Big Data:-

- Introduction of Big Data
- Case for Hadoop Ecosystem
- Hadoop Installation
- Planning Your Hadoop Cluster:-The Hadoop Distributed File System (Hdfs)
- MapReduce
- MapReduce with Example
- Developing a MapReduce Application
- MapReduce Types & Formats
- Apache Pig
- Hive
- Introuction to Hbase
- Apache Spark
- PySpark Basics
- Aggregating Data with Pair Rdds

Job Opportunities & Meeting the Skill Gap:

The demand for Analytics skill is going up steadily but there is a huge deficit on the supply side. This is happening globally and is not restricted to any part of geography. In spite of Big Data Analytics being a 'Hot' job, there is still a large number of unfilled jobs across the globe due to shortage of required skill. A McKinsey Global Institute study states that the US will face a shortage of about 190,000 data scientists and 1.5 million managers and analysts who can understand and make decisions using Big Data by 2018.

India currently has the highest concentration of analytics globally. According to Srikanth Velamakanni, co-founder and CEO of Fractal Analytics, there are two types of talent deficits: Data Scientists, who can perform analytics and Analytics Consultant, who can understand and use data. The talent supply for this job title, especially Data Scientists is extremely scarce and the demand is huge.

Job Titles and Type of Analytics:

From a career point of view, there are so many options available, in terms of domain as well as nature of job. Since Analytics is utilized in varied fields, there are numerous job titles for one to choose from.

- Big Data Analytics Business Consultant
- Big Data Analytics Architect
- Big Data Engineer
- Big Data Solution Architect
- Big Data Analyst
- Analytics Associate
- Business Intelligence and Analytics Consultant
- Metrics and Analytics Specialist