Semester : 7th   Subject : BDA   Academic Year: 2024 - 2025

Module 1 :- Introduction to Big Data & Hadoop

Data :-

The quantities, characters, or symbols on which Operations are performed by a computer. Alpahabets (A-z, a-z) digits (0-9) or special characters (+, -, /, *, <, >, =, etc) which may be stored and transmitted in the form of electrical signals recorded on magnetic, optical, or mechanical recording media.

Information :-

Information is organized or classified data, which has some meaninful values for the receiver. Information is the processed data on which decisions and actions are based.

Big Data :

Big Data is a term used for a collection of data sets that are larged and complex difficult to store and process using available database management tools or traditional data processing applications.

Big data is data whose scale, diversity, and complexity require new architecture, techniques, algorithms and analytics to manage it and extract value and hidden knowledge from it. Normally we work on data of size MB (WordDoc or Excel), or maximum GB (movies) but data in Peta bytes ie 10^{15} byte size is called Big Data. It is stated that almost 90% of today's data has been generated in the past 3 years

Sources of Big Data

These data come from many sources like

- **Social Networking Site:-**
  Facebook, Google, Linkedin all these sites generates huge amount of data on a day to day basis as they have billions of users worldwide.

- **E-Comerce Site**
  Sites like Amazon, flipKart, Alibaba generates huge amount of logs from which users buying trends can be traced.

- **Weather Station:**
  All the weather station and satellite give huge data which are stored and manipulated

Semester : _____   Subject : _____   Academic Year: 20 ___ - 20 ___

to forecast weather.

Telecom Company :

Telecom giants like Airtel, Vodafone, study the user trends and accordingly publish their plans and for this they store the data of its million users.

Share Market

Stock exchange across the world generates huge amount of data through its daily transaction.