ByteWise Fellowship – Week 1

Module: DE Basics

Date :14-March-2023

1. **What is BigData?**

   Big data is a large and complex data which is Hard to manage and process through conventional techniques and tools. It can be structured, Unstructured and semi structured.

   When conventional data is unable to handle the enormous data production then comes Big data and its tools.

   - **5 V's of Big data:**

     Velocity : Data Rapid generation and Storage

     Veracity : Data Accuracy and Quality

     Volume : Size and Quantity Of data

     Value : Insightful data analytics

     Variety : Diversity and heterogeneity of data

2. **What is a Data WareHouse?**

   A data warehouse is a data management system which aggregates data from different resources into one single consistent data storage to support data analytics, Machine learning, AI and data mining on huge amounts of data in ways a typical database cannot.

   It also manages historic as well as current data. The data in DWH is structured, filtered and processed for its own specific purpose.

3. **What is Data Lake?**

   Data Lake is the storage repository that holds the data in its native and raw format (Structured, Unstructured and Semi Structured) at any scale.

   For fast retrieval of the data, data lake associates it with meta tags and identifiers.

4. What is a DataBase ?

The database is a collection of data and information which stores current data to power up an application and enable user and application to interact with the data. There are many databases in the market and all of them have their own characteristics and properties. Following are some examples of databases

- Relational databases: Oracle, MySQL, Microsoft SQL Server, and PostgreSQL
- Document databases: MongoDB and CouchDB
- Key Value databases: Redis and DynamoDB