**Name:** Ahsan Bilal

# Big Data

- Large Dataset collection which is not processed by our traditional machines
  *Example*: Assume a smartphone is generating 4 Exabytes of data per month in an unstructured manner or a variety of data like images, text, voice, or videos and we have 5 billion mobile users it will be a lot of data that our machines will not handle

- Classification of Big data:
  - ➔ Volume: Data is in a large amount
  - ➔ Velocity: Data increases exponentially
  - ➔ Variety: Data may be structured, Semi-Structured ,or even in unstructured form
  - ➔ Varacity: How accurate and trustworthiness of this dataset is
  - ➔ Value: By analysing this data we can get value insights which will help our business to grow

- Tech &Tools:
  Hadoop, Spark, etc

# Database

- A database is an organized collection of structured data that can be accessed, managed, and updated. It can be thought of as a digital filing system where information is stored in an organized and structured manner, making it easy to retrieve and manipulate the data.
- Databases are commonly used to store a wide range of information, including customer information, inventory data, financial records, and more. They can be managed using database management systems (DBMS), which allow users to create, update, and query the data stored in the database.
- Overall, a database helps to ensure that data is stored securely, is easily accessible, and can be efficiently managed and analyzed.
- Tech & Tools:
  SQL Server, MySQL, Oracle, etc

**Name:** Ahsan Bilal

# Data Warehouse

- Big Collection of data from different databases which will use to make stratergies But it demand structured data
  *Example :* Walmart have many branches and their every brach data is stored in a database but if we want to get some insights from our all branches data it will be difficult for us to gather data from all databases here data warehouse concept making things easy for us and it will allow us to store data from different databases or places and get insights from them and make strategies to grow your business according to that analysis

- Subject oriented: Data is stored in different marts according to theri subjects
- Integrated: We can integrate data from different sources in our data warehouse
- Non Volatile: One data is stored in data warehouse it will become non editable you can delete it but cant update it
- Time Varient: Data warehouse can store historical data means you can store old data in data warehouse

# Data Lake

- A Data Lake is a large and centralized repository that stores raw, unprocessed data in its native format, allowing for flexible and scalable data storage and analysis.

- Unlike traditional data storage methods, a Data Lake does not require data to be pre-processed or structured before storage. Instead, it accepts and stores all data types, including structured, semi-structured, and unstructured data, in their original format. This makes it possible to store large amounts of data without the need for time-consuming data preparation or transformation.

- A Data Lake can be used for a wide range of purposes, such as data exploration, machine learning, and big data analytics. Data scientists and analysts can access and analyze the data stored in the Data Lake, and can use advanced analytics tools to derive insights from the data.

- Overall, a Data Lake is a powerful tool for organizations that need to store and analyze large amounts of data from a variety of sources, without having to worry about data transformation and structure.

**Date:** 23/3/2023