

Name : Muhammad Ahad khan

BIG DATA

Ans: Big data is a general term for representing modern data.

- Big data is mostly present in unstructured format.
- Big data mostly stored in Petabyte or Exabyte as its name (big data) clear that vast amount of data is present.
- Earlier few years, data is increasing gradually but now data is increasing exponentially, so it is called big data.
- Earlier few years, normally data can be stored on any local server but now big data is distributed globally and stored on a large amount of servers (data centers).
- Hadoop & spark (technologies) are mostly used for accessing big data from different servers. 2

DATA LAKE

A data lake is a place where big amounts of data can be stored, processed, and kept safe.

Data can be organized, semi-organized, or not organized at all.

Data lake can keep data in its original format and process it in any way, regardless of its size.

SQL, Python, and R are all popular technologies used in data lake.

Used for ML and AI as they are now, or for analytics that involve processing.

Data can be sorted and put into a database or data warehouse in a data lake.

DATABASE

The data collection and storage capabilities of a database are its primary strengths.

Data stored in a database has the capacity to be active data that is updated in real time.

Tables consisting of rows and columns are used to contain the data in a database.

The data in a database are extremely well structured. (structured data).

DATA WAREHOUSE

Data warehouse is designed for analytical process.

- Data will be transfer from database to data warehouse by ELT process.
- Data that is fetched from database to data warehouse, is not in fresh format, however, data is in summarized form.
- Data warehouses are used for analysis and processes on very large amount of data with very high speed.