BIG DATA

- 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.

DATA LAKE

- A data lake is a centralized repository designed to store, process, and secure large amounts of data.
- Data can be in structured, semi-structured, and unstructured format.
- Data lake can store data in its native format and process any variety of it, ignoring size limits.
- SQL, Python & R technologies are commonly used in data lake.
- Used for ML &AI in its current state or for analytics with processing.
- In data lake, data can be organized and put into database or data-warehouse.

DATABASE

- Database is designed to capture and record data.
- In database, data has the ability to be live and real time data.
- In database, data stored in tables with rows and columns.
- In database, data is highly organized (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.