BIBA CLASS 1

Data Engineering

This is the process of designing, building and scaling systems that organize data for the analytics.

To analyse we have, ETL(Extract, Transform, Load) model.

IOT Devices

Data API

Frontend

SQL Database

Analytics

The Data API’s carry data to the SQL Databases from IOT Devices. The SQL Database analyse the data and provides data to the Frontend i.e, Mobiles, Computers,etc.

An Introduction to Data Warehousing

Data Warehousing is the process of storing large amount of raw data. It is

* Subject Oriented : Data is organized based on subject but not application. It provides topic wise information rather than overall process. Eg: Product will be the subject name and product id, product name, product brand will be the attributes.
* Integrated : Data Warehouse is constructed by integrating various data sources like relational DB, flat files, on-line transaction records.
* Time Varient : Data Warehouse can store data that can lasts upto 5-10 years
* Non-volatile: Data can’t be removed from Data Warehouse.

Data Classification:

There are 3 types of data. They are Raw data, Processed data and Cooked data.

* Raw Data: Non refined data will be present here. No schemas are applied. Data is in the JSON format.
* Processed Data: Here schemas are applied on data
* Cooked Data: Only important data is absorbed from the processed data.

A large amount of data will be produced from different appliactions. So all this data is known as Big Data. To manage all these, Data Engineering is the solution.

Big Data Properties:

* Volume: The size of data.
* Veracity: States about the reliability(consistency) of data.
* Velocity: The speed of data.
* Variety: Different data.

Batch Processing:

Data 🡪 Storage 🡪 Analytics 🡪 Insight

The data will be stored in batch wise. Systems use to periodically complete high-volume, repetitive data jobs. At last, we are going to check whether data is stored or not.

Steam Processing:

Data came through the live streaming in YouTube, Instagram, etc. It is known as the continuing flow of data and storing it.

🡪Big data works on Map Reduce: This is key value pairing. Organizing data in key value pairs.

Examples for Data Engineering are Hadoop, Apache Spark, etc.

Data storage

Data can be stored in Relational database and Document store.

Operational Data Store

Operational data store is a process where large amount of data is gathered from different locations and stored in a single place for the business operations.

OLTP(Online Transaction Processing)

It is a type of database system in which the systems are expected to respond to user requests and process them in real time.

Example: Online payment, Sending messages, etc.

Difference between Data Warehouse and OLTP

Datawarehouse is a collection of large amount of data from different sources. OLTP is the day to day transaction data that keeps changing.

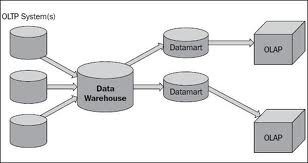
Datawarehouse has past data. OLTP has current data.

Datawarehouse has more than 1000TB data stored in it. OLTP has database less than 100MB.

Data Mart

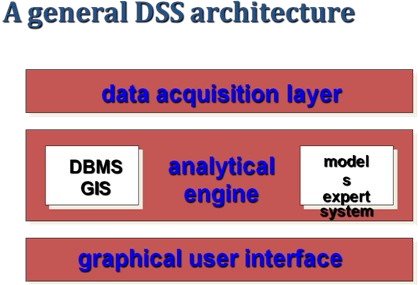
Data Mart is a part of Data Warehouse. It is used to analyse specific information regarding a business more efficiently.

Datawarehouse Architecture



Decision support system(DSS)

A company needs decision support system for quick decision making.



Chakradhar Bhogapurapu, BIBA