2015-2016 ODD SEM

V SEM

BATCH(2013-2017)

DATAWAREHOUSE AND DATA MINING

Components of a Data Warehouse

INTERNAL DATA SOURCES
- Operational Data
- Customer Data
- Manufacturing Data
- Historical Data

EXTERNAL DATA SOURCES
- External Data

Extract and Transform → Data Warehouse ↔ Data Access and Analysis
- Queries and reports
- OLAP
- Data mining

Information Directory
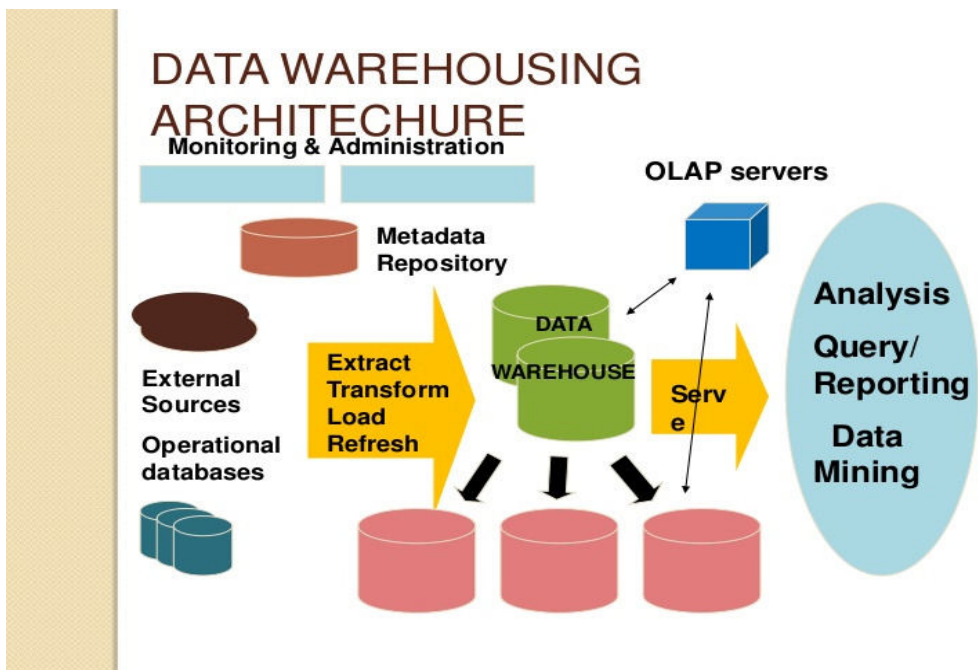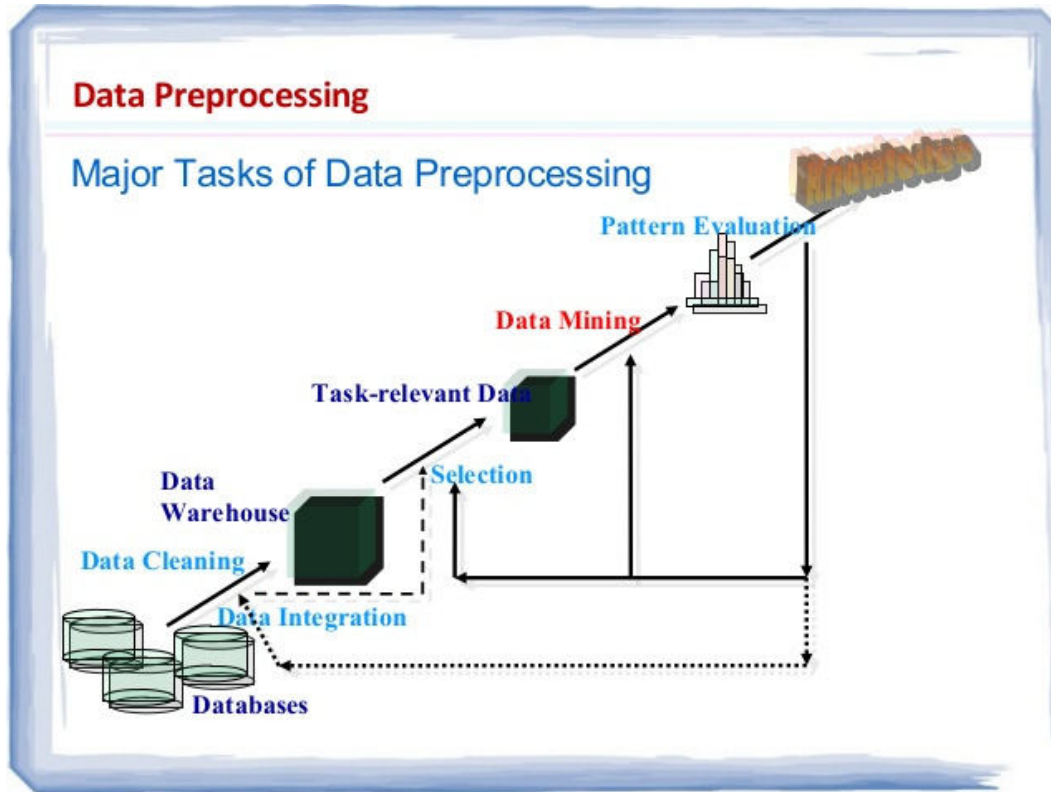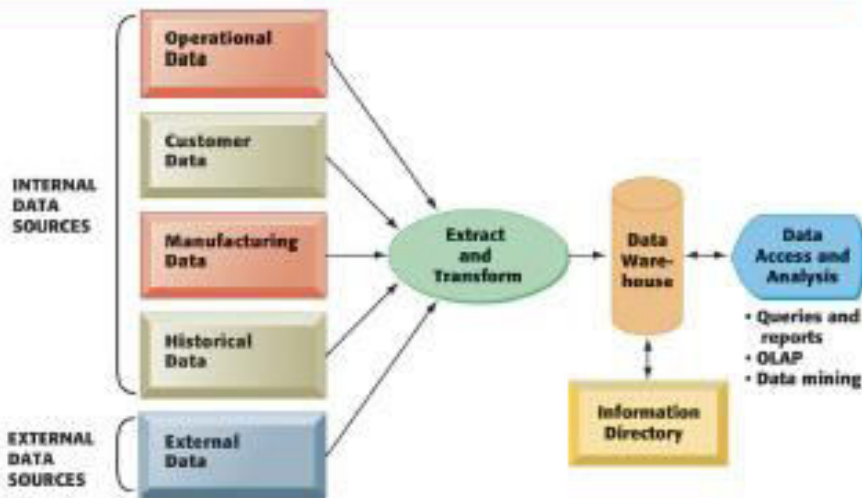
The data warehouse extracts current and historical data from multiple operational systems inside the organization. These data are combined with data from external sources and reorganized into a central database designed for management reporting and analysis. The information directory provides users with information about the data available in the warehouse.

- Changes over time
- 3 types of SCDs
  - i) Does not track changes, and it will overwrites previous dimension information with current dimension information.
  - ii) Track changes; includes 4 supplementary attributes.
  - iii) Do not track the entire history, only current and original state of a dimension.

A small, single-subject data warehouse subset that provides decision support to a small group of people.

- **Top-down approach**
  Enterprise data warehouse is created first, then data is extracted to create one or more data marts.
- **Bottom-up approach**
  Build independent data marts as needed, enterprise data warehouse will be built later from existing data marts.

OLAP

Slowly changing dimension

Data Mart

Characteristics of data warehouse
- Historical
- Time Variant
- Non-volatile
- Denormalized
- Subject-oriented
- Integrated
- Aggregated

- It is an advanced data analysis environment that supports decision making, business modeling and operations research.
- 3 kind of OLAP cubes
  - i) **Relational OLAP**: Greater scalability
  - ii) **Multidimensional OLAP**: Rapid Response
  - iii) **Hybrid OLAP**: Combination of ROLAP and MOLAP

Data Warehouse and OLAP

Schemas

Data warehouse

Constellation Schema
- Important for bottom-up approach
- Consists of 2 or more fact table

Snowflake Schema
- Multiple dimensions
- Quite similar to Star Schema

Star Schema
- Most popular schema
- Consists only of two levels (fact & dimension)

- Designed to facilitate reporting and analysis.
- Retrieve and analyse data, to extract, transform and load data, and to manage the data dictionary are also considered essential components of a data warehousing system.