ETL

DW implementation includes loading preliminary data, implementing transformation program, design user interface and develop standard query and reports.

Datawareh implementation.

Extract Transform Load (ETL) covers a process of how the data are loaded from the source system to the DW.

The sequence is extract, clean, transform, load

24x7

→ Extract

The objective of extract step is to retrieve all the required data from the source system with as little resources as possible & in such a manner that it doesn't affect the performance, response time of the system.

Ways to perform extract

1) Update Notification: System is able to provide the notification that the record has been changed & describe the change.
2) Incremental Extract: The system is able to identify which record has been modified & provide an extract of such record.
3) Full Extract: It requires keeping a copy of the last extract in the same format in order to be able to identify changes.

→ Clean

Cleaning is the important step as it ensures the quality of the data in the data warehouse

e.g.: 1) Making identifier unique (Male/M, Female/F)
2) Convert + no. zip codes to standard format.

3) Transform

This step applies set of rules to transform the data from the source to the target. It requires joining data, sorting, deriving new calculated values etc.

4) Load

During the load it is necessary to ensure that load is performed directly & with as little resources as possible.

The target of this process is a database.

ETL Tools (Eg)

1) Informatica - Power Center
2) IBM - Cognos Data Manager
3) Microsoft - SQL Server Integration Services
4) AB - initio

Data Warehouse Design / (View) Design of DWH. (w.r.t S.E).

Four different views regarding the design of D/w are

1) Top-down View

It allows the selection of the relevant information necessary for the DW.

2) Data Source View

It exposes the information being captured, stored and managed by operational systems

3) Data warehouse View

It includes fact & dimension tables.

4) Business Query View

It is perspective of data in DW from the viewpoint of the end user.

The warehouse design process consist of the following steps:

1) choose a business process to model
   e.g. sales, inventory, account, administration etc.

2) choose the grain of the business process
   grain is the atomic level of data to be represented.

3) choose the dimension that will apply to each fact table. e.g. Time, item, supplier, etc.

4) choose the ~~measures~~ measures that will popuate each fact table record. Measures are numeric quantities. e.g. Dollars sold, unit sold.