Data Warehouse
Technologies

Data Warehouse Technologies

Basic Functions of Databases

1. *Data Manipulation and Management*
2. *Reading Data with the Purpose of Displaying and Reporting*
3. *Data Analysis*

Data Warehouse Technology

Data Warehouse Technologies

Data Warehouse Characteristics

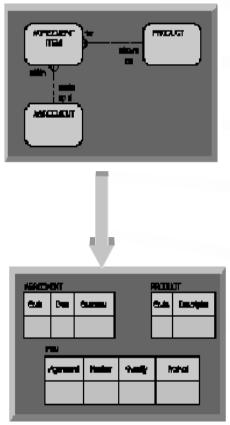
1. *Organizations need more about trends of data for Strategic Analysis rather than individual records*
2. *Data in warehouses are used to display rather than manipulation*
3. *Last minute Change in data does not affect analysis results*
4. *Data modeling and design in data warehouses are different than relational data modeling*

Data Warehouse Technologies

For extracting information from data warehouses

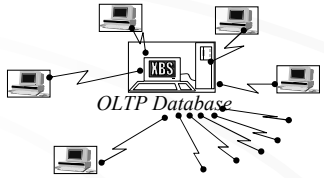
1. *OLAP (On-Line Analytical Processing)*
2. *Data Mining*

Data Warehouse Technologies



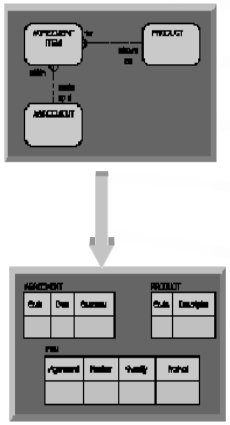
OLTP (On-Line Transaction Processing) Databases

They are used in daily data processing and operations. They are modeled by Confirming the first 3 normalization rule.



OLTP Database

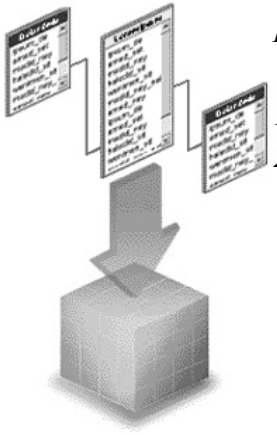
Data Warehouse Technologies



OLTP (On-Line Transaction Processing) Databases

1. *Supports relational rules (conforms to 3 normal forms)*
2. *Ideal for daily data processing*
3. *Not suitable for historical queries covering wide range of data (OLAP)*
4. *Not suitable for Automatic Information Extraction (Data Mining)*

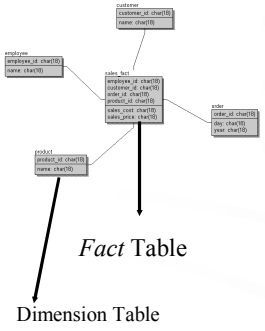
Data Warehouse Technologies



Data Modeling in Data Warehouses

1. *Multidimensional usage*
2. *Data Modeling Methods*
   a. *Star schema*
   b. *Snow flake schema*
   c. *A single Denormalized table*

Data Warehouse Technologies

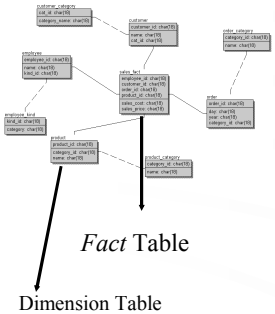


Data Modeling in Data Warehouses

Star Schema

There is only one fact table. It is associated with many dimension tables. Fact table contains values related to cube dimensions. Columns in Dimension tables contain cube measures.

Data Warehouse Technologies



Data Modeling in Data Warehouses

Snow Flake

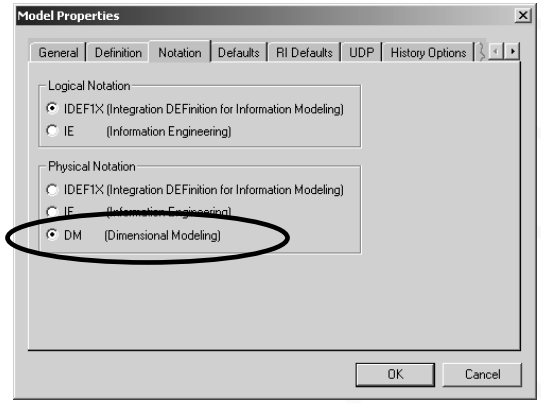
Dimension tables are associated with other tables.

Fact Table

Dimension Table

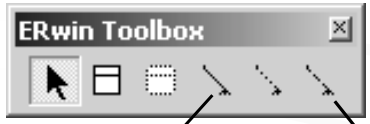
Data Warehouse Technologies

Multi Dimensional Modeling using ERwin



Data Warehouse Technologies

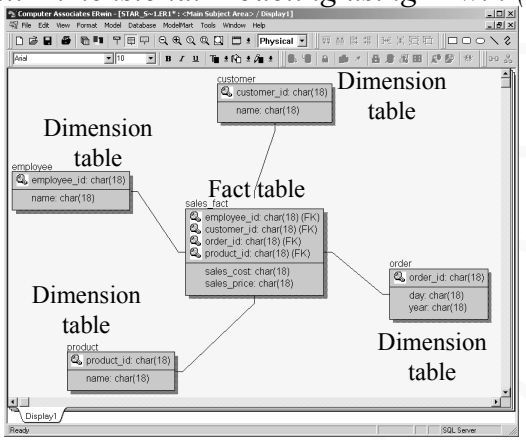
Multi Dimensional Modeling using ERwin



Identifying Relationship Non-identifying relationship

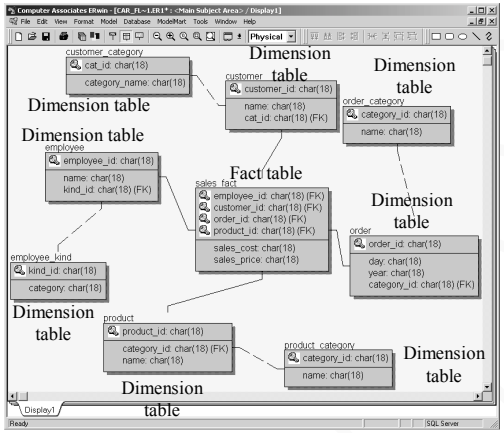
Data Warehouse Technologies

Multi Dimensional Modeling using Erwin (Star)



Data Warehouse Technologies

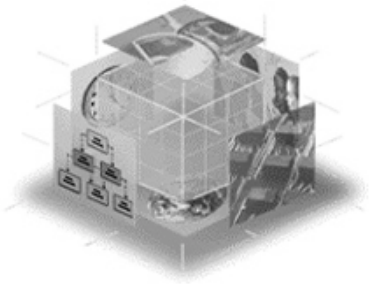
Multi Dimensional Modeling using Erwin (Snow-flake)



Data Warehouse Technologies

Cubes

Multi-dimensional modeling is first issue in OLAP analysis and query performances of Data Warehouses.



Data Warehouse Technologies

Cubes

OLAP manager prepares the infrastructure to transfer the data in relational databases into organizational information.

He uses data cubes for this purpose.

Data Warehouse Technologies

Dimensions of Cubes

Dimensions of cube are various categories obtained from data.

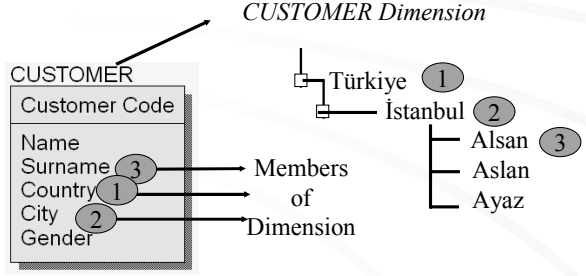
Typical examples to these dimensions are time, geographical place, or product.

Dimensions of these cubes are not limited to 3.

Data Warehouse Technologies

Dimensions and Hierarchies

Data in dimensions are built by classifying the data in columns of database tables in a specific hierarchy.

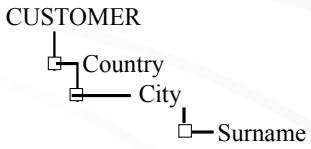


Data Warehouse Technologies

Dimensions and Hierarchies

Dimension hierarchies are classified into levels which are called dimension members.

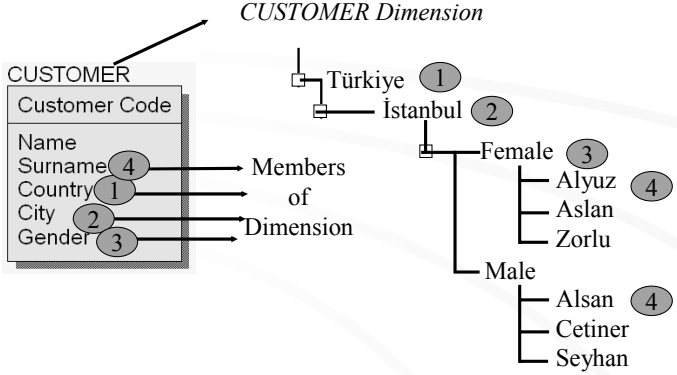
Each level in a dimension comes together to construct another level.



Data Warehouse Technologies

Dimensions and Hierarchies

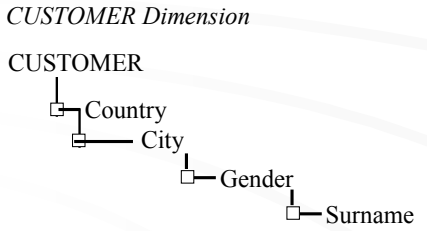
Suppose we define now the members as follows:



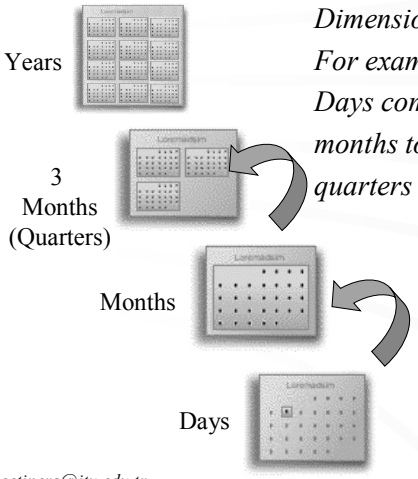
Data Warehouse Technologies

Dimensions and Hierarchies

Suppose we define now the members as follows:



Data Warehouse Technologies



Dimensions and Hierarchies

For example; in a time dimension, Days come together to make months months to make quarters, and quarters to make years.

Data Warehouse Technologies

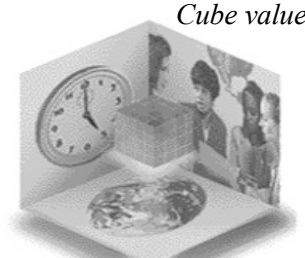
Cube Values, or measures

They are the values in database table to be analyzed quantitatively.

For example; sale price, cost, budget, working hour, maintenance duration, profit etc.

Data Warehouse Technologies

Cube values, or measures



These measures can be analyzed against different dimension categories in cube.

Sale price, cost, budget, working time, maintenance duration, profit etc.

Data Warehouse Technologies

For Example; You can analyze costs, sale prices, and profits (measures) for a certain product (product dimension) in different countries (geographical location dimension) in the last two years (time dimension).

Data Warehouse Technologies

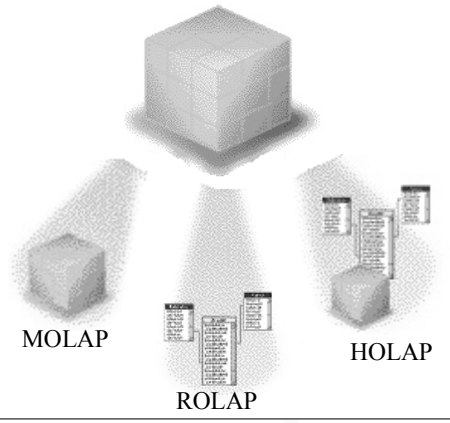
For example; How many computers did we sell in İstanbul in the first quarter of this year and what is the difference if we compare with the sales in the first quarter of last year?

Question: Can any manager write an SQL statement to bring this report?

Answer: They don't need to write SQL statements in OLAP.

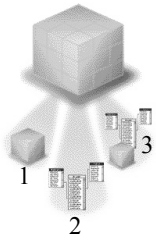
Data Warehouse Technologies

Storing Data in Cubes



Data Warehouse Technologies

Storing Data in Cubes

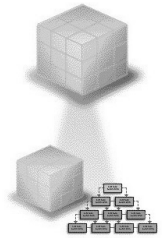


1. MOLAP-Multidimesional OLAP
2. ROLAP-Relational OLAP
3. HOLAP-Hybrid OLAP

Selection of a method is dependent on database size and the usage of data.

Data Warehouse Technologies

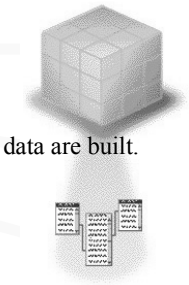
MOLAP- Multi Dimensional OLAP



- MOLAP is high performance multi dimensional data store format.
- In MOLAP, data is stored physically in OLAP server.
- It gives the best performance.
- Suitable for small-to-middle size organizations.

Data Warehouse Technologies

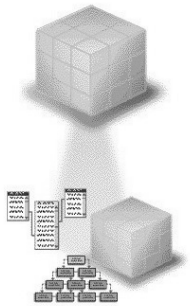
ROLAP- Relational OLAP



- Data is in the actual (OLTP) database.
- Different relational tables which refer to actual data are built.
- Suitable for large databases which are not queried frequently.

Data Warehouse Technologies

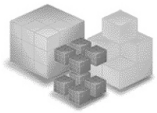
HOLAP- Hybrid OLAP



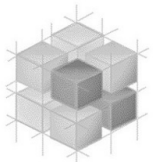
- Denormalized data are stored in tables in MOLAP whereas others are stored in relational (OLTP) database.

Data Warehouse Technologies

Data partioning

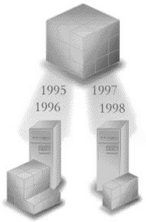


P*artioning*



Partitions may be held differently as MOLAP, ROLAP, or HOLAP.

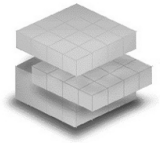
Distributed Data



Partitions may be held on different servers.

Data Warehouse Technologies

Data Slices

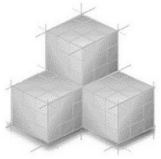


A data set in a partition

Example: Sale measures for a specific product for all years.

Data Warehouse Technologies

Virtual Cubes



Same as views in Relational (OLTP) Databases.

Data Warehouse Technologies

Dr.Codd suggests 12 rules for OLAP systems.

1. *Multidimensional conceptual view*
2. *Transparency (Example, accessing easily from MS-Excel etc)*
3. *Tool should only access to objects which are used for analysis*
4. *Consistent reporting performance (performance should not decrease tremendously with the increase of dimensions)*
5. *Client/Server Architecture*
6. *Each dimension should be equally structural and operational.*
7. *Dynamic Matrix handling*
8. *Multi user support*
9. *Unlimited operations between dimensions and automatic grouping of levels within dimensions*
10. *Operations such as Pivoting, Drill-down and Roll-up should be done easily (e.g. drag and drop processes) by selecting the cells of cube*
11. *Flexible reporting (Visual support specially)*
12. *Unlimited dimensions and support for grouping any number of levels*