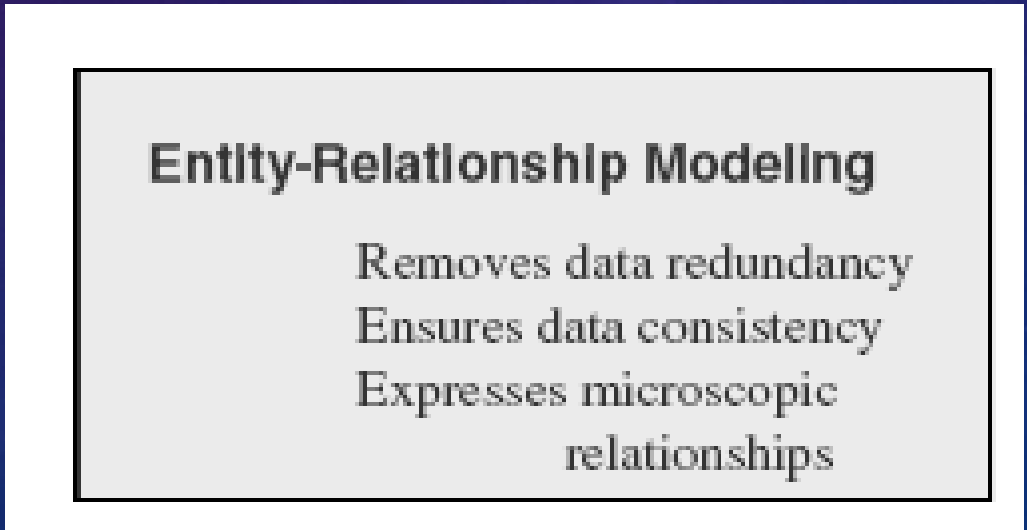
DATA WAREHOUSING
LECTURE 6

ENGR. MADEHA MUSHTAQ
DEPARTMENT OF COMPUTER SCIENCE
IQRA NATIONAL UNIVERSITY

ER MODELING

- We are familiar with data modeling for operational or OLTP systems.
- We adopt the Entity-Relationship (E-R) modeling technique to create the data models for these systems.
- Figure below lists the characteristics of OLTP systems and shows why E-R modeling is suitable for OLTP systems.



ER MODELING

- Entity relationship diagrams provide a visual starting point for database design.
- The diagrams are used to design or analyze relational databases used in business processes.
- Coupled with normalization drives out all the redundancy out of the database.
- Change (add or delete) the data at just one point.
- Resulted in success of OLTP systems.

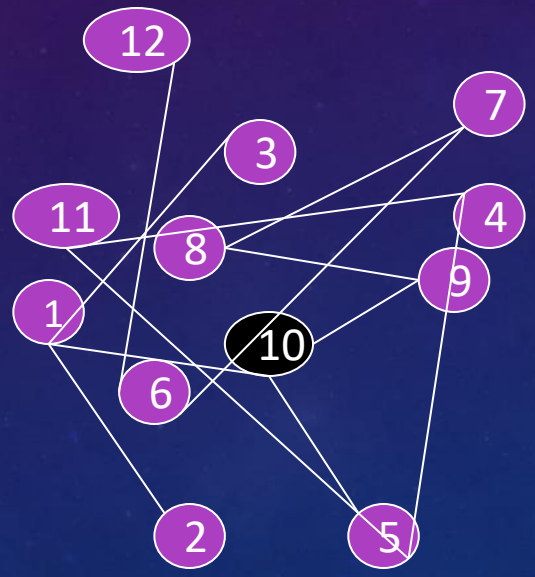
NEED FOR DIMENSIONAL MODELING

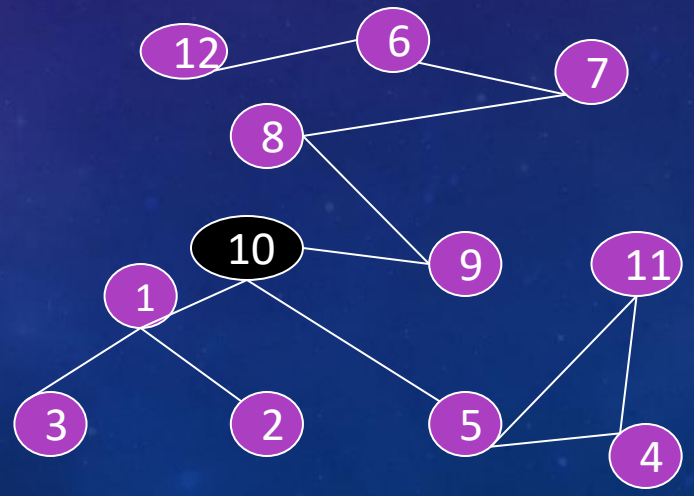
- Lets have a look at a typical ER data model first.

- Some Observations:
- All tables look-alike, as a consequence it is difficult to identify:
  - Which table is more important ?
  - Which is the largest?
  - Which tables contain numerical measurements of the business?
  - Which table contain nearly static descriptive attributes?

NEED FOR DIMENSIONAL MODELING

- Many topologies for the same ER diagram, all appearing different.
  - Very hard to visualize and remember.
  - A large number of possible connections to any two (or more) tables.





NEED FOR DIMENSIONAL MODELING

- ER and Normalization result in large number of tables which are:
  - Hard to understand by the users (DB programmers)
  - Hard to navigate optimally by DBMS software
- Real value of ER is in using tables individually or in pairs.
- Too complex for queries that span multiple tables with a large number of records.

ER VS. DM

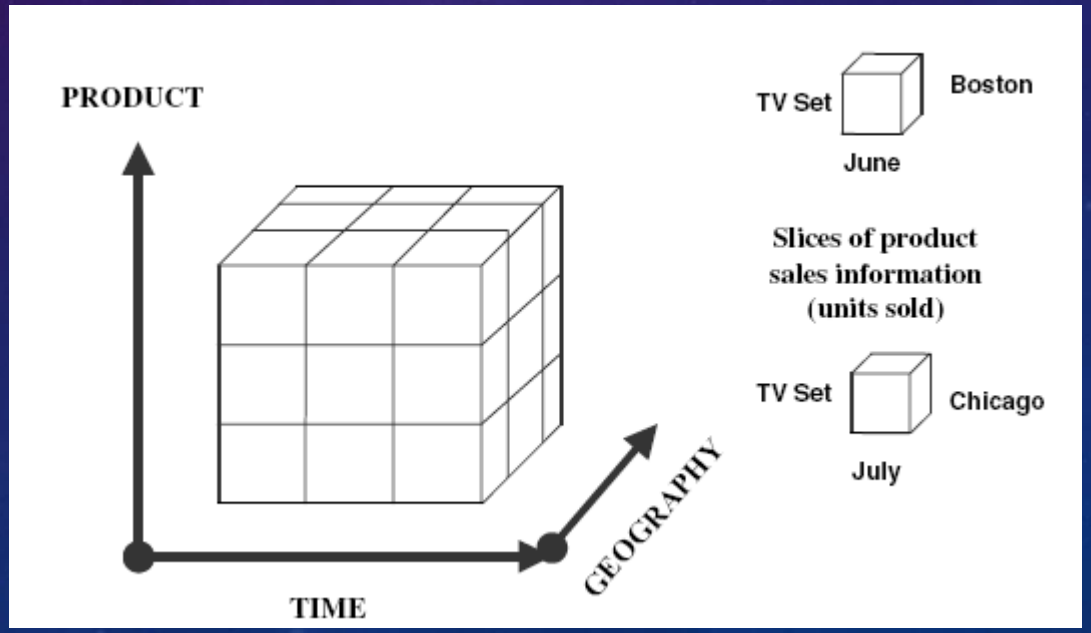
ER	DM
Constituted to optimize OLTP performance.	Constituted to optimize DSS query performance.
Models the micro relationships among data elements.	Models the macro relationships among data elements with an overall deterministic strategy.
A wild variability of the structure of ER models.	All dimensions serve as equal entry points to the fact table.
Very vulnerable to changes in the user's querying habits, because such schemas are asymmetrical.	Changes in users' querying habits can be accommodated by automatic SQL generators.

DIMENSIONAL MODELING

- A simpler logical model optimized for decision support.
- Inherently dimensional in nature, with a single central fact table and a set of smaller dimensional tables.
- Multi-part key for the fact table.
- Dimensional tables with a single-part PK.
- Keys are usually system generated.

DIMENSIONAL NATURE OF BUSINESS DATA

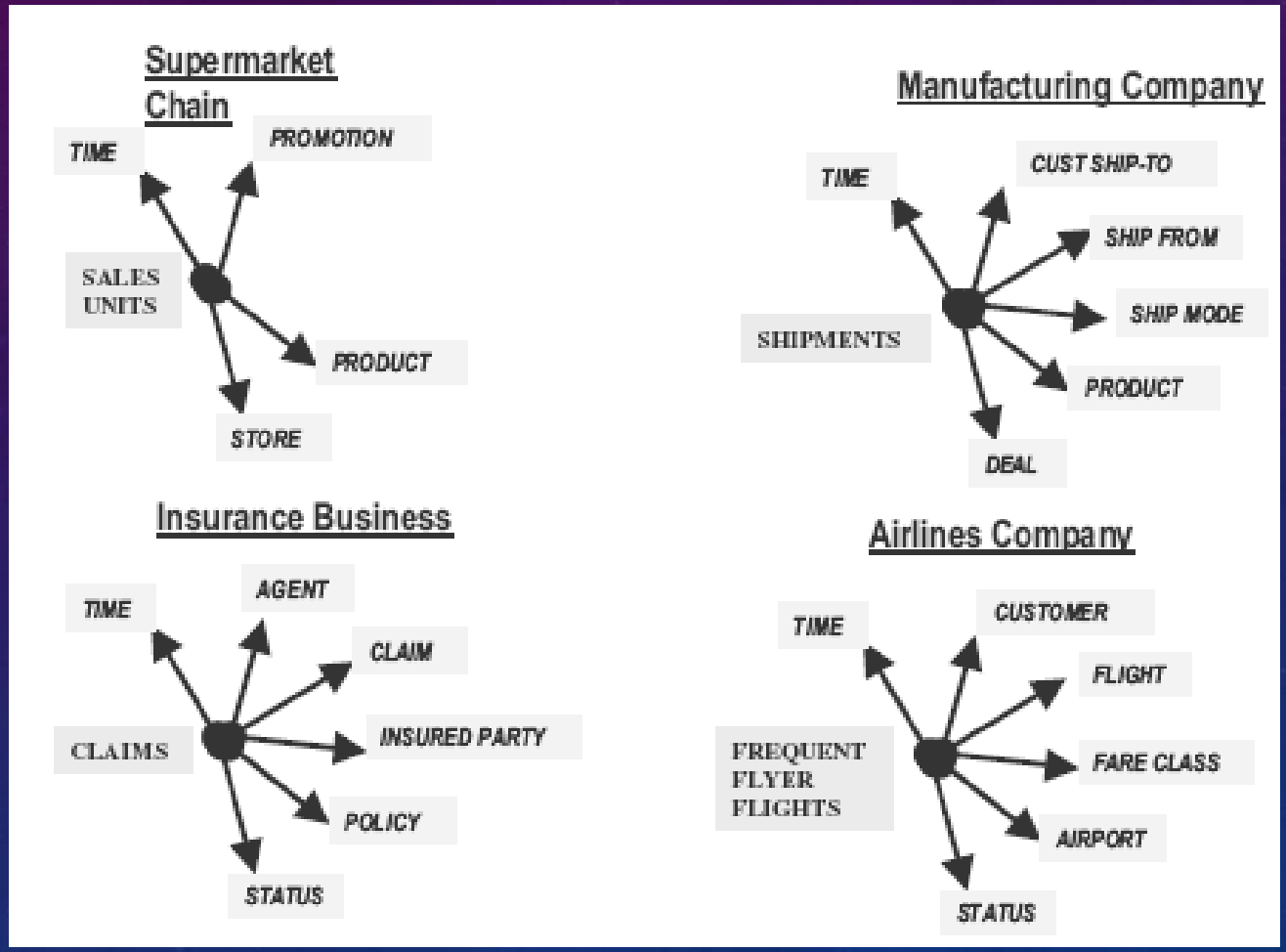
- To get a good grasp of the dimensional nature of business data, let us look at an example:
- Figure below shows the analysis of sales units along the three business dimensions of product, time, and geography.



DIMENSIONAL NATURE OF BUSINESS DATA

- These three dimensions are plotted against three axes of coordinates.
- In this case, the business data of sales units is three dimensional because there are just three dimensions used in this analysis.
- If there are more than three dimensions, we extend the concept to multiple dimensions and visualize multidimensional cubes, also called hyper cubes.

EXAMPLES OF BUSINESS DIMENSIONS



Examples of business dimensions

EXAMPLES OF BUSINESS DIMENSIONS

- For the supermarket chain, the measurements that are analyzed are the sales units.
- These are analyzed along four business dimensions.
- For the insurance company, we would want to analyze the claims data by agent, individual claim, time, insured party, individual policy, and status of the claim.
- For the example of the airlines company the business dimensions are time, customer, specific flight, fare class, airport, and frequent flyer status.

EXAMPLES OF BUSINESS DIMENSIONS

- The example analyzing shipments for a manufacturing company show some other business dimensions.
- Here we see the dimensions of time, ship-to and ship-from locations, shipping mode, product, and any special deals.
- What we find from these examples is that the business dimensions are different and relevant to the industry and to the subject for analysis.
- We also find the time dimension to be a common dimension in all examples.
- Almost all business analyses are performed over time.

PROCESS OF DIMENSIONAL MODELING

- Five Step process of DM:
  - Choosing the Business Process
  - Choosing the Grain
  - Choosing the Facts
  - Choosing the Dimensions
  - Choosing the duration of the database

PROCESS OF DIMENSIONAL MODELING

- Choose the Business Process:
  - A business process is a major operational process in an organization.
  - Selecting the subjects from the information packages for the first set of logical structures to be designed.
  - Typically supported by a legacy system (database) or an OLTP.
  - Examples: Orders, Invoices, Inventory etc.

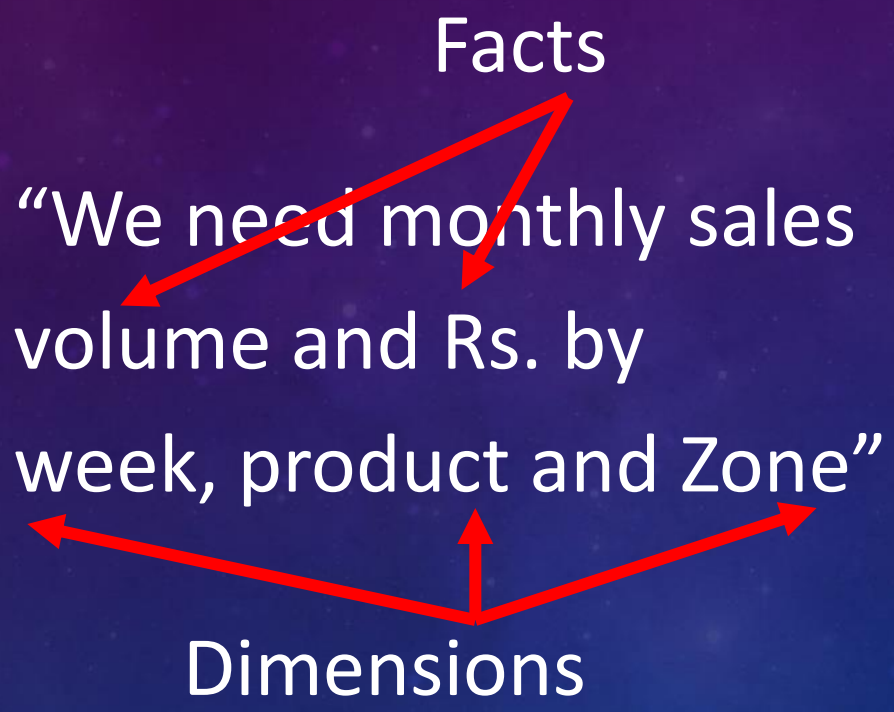
PROCESS OF DIMENSIONAL MODELING

- Choosing the Grain:
  - Determining the level of detail for the data in the data structures.
  - Grain is the fundamental, atomic level of data to be represented.
  - Grain is also termed as the unit of analyses.
  - Example grain statements:
  - Typical grains
    - Individual Transactions
    - Daily aggregates
    - Monthly aggregates

PROCESS OF DIMENSIONAL MODELING

- Choosing the Facts:
  - Selecting the metrics or units of measurements (such as product sale units, dollar sales, dollar revenue, etc.) to be included in the first set of structures.
  - Choose the facts that will populate each fact table record.
  - Remember that best Facts are Numeric, Continuously Valued and Additive.
  - Examples: Quantity Sold, Amount etc.

PROCESS OF DIMENSIONAL MODELING



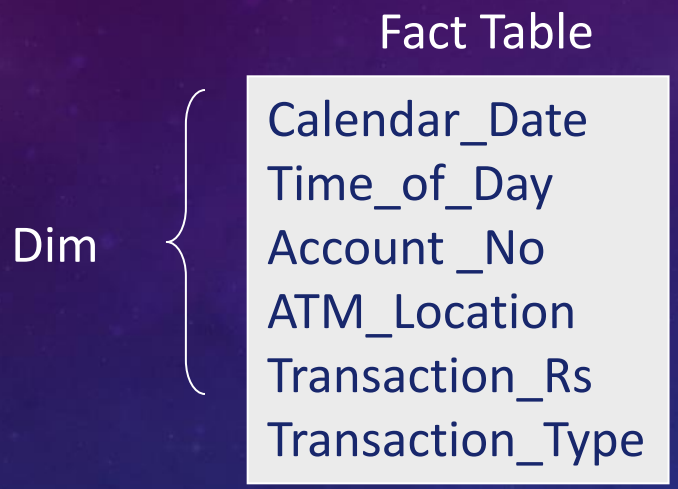
Choosing fact statement

PROCESS OF DIMENSIONAL MODELING

- Choosing the Dimensions:
  - Choosing the business dimensions (such as product, market, time, etc.) to be included in the first set of structures.
  - Choose the dimensions that apply to each fact in the fact table.
  - Typical dimensions: time, product, geography etc.
  - Identify the descriptive attributes that explain each dimension.
  - Determine hierarchies within each dimension.

PROCESS OF DIMENSIONAL MODELING

- The single valued attributes during recording of a transaction are dimensions.



Time_of_day: Morning, Mid Morning, Lunch Break etc.

Transaction_Type: Withdrawal, Deposit, Check balance etc.

How to Identify a Dimension?

PROCESS OF DIMENSIONAL MODELING

- Are dimensions ALWAYS single?
  - Not really
  - What are the problems? And how to handle them

- Calendar_Date (of inspection)
- Reg_No
- Technician
- Workshop
- Maintenance_Operation

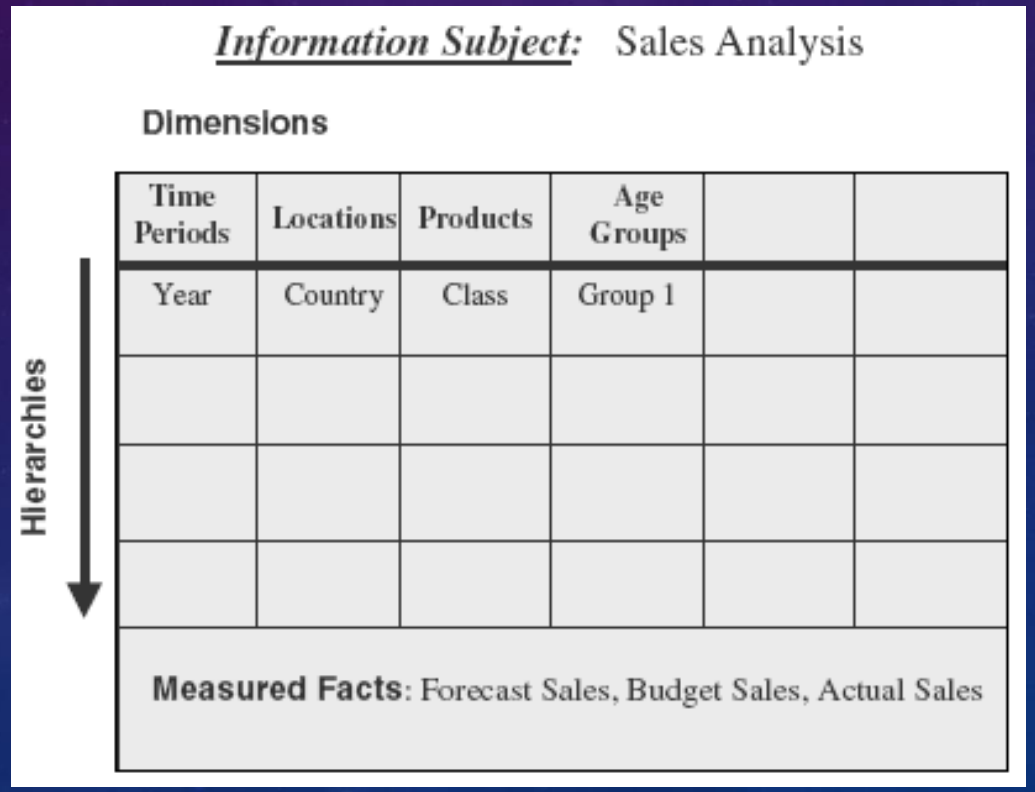
Multivalued Dimensions

PROCESS OF DIMENSIONAL MODELING

- Choosing the duration of the database:
  - Determining how far back in time you should go for historical data.

DIMENSIONAL MODELING

- Let us look at an example for analyzing sales for a certain business as shown in figure:



DIMENSIONAL MODELING

- The subject here is sales.
- The measured facts or the measurements that are of interest for analysis are shown in the bottom section of the figure.
- The business dimensions along which these measurements are to be analyzed are shown at the top of diagram as column headings.
- In our example, these dimensions are time, location, product, and demographic age group.

DIMENSIONAL MODELING

- Each of these business dimensions can contain a hierarchy or levels.
- For example, the time dimension has the hierarchy going from year down to the level of individual day.
- The other intermediary levels in the time dimension could be quarter, month, and week.

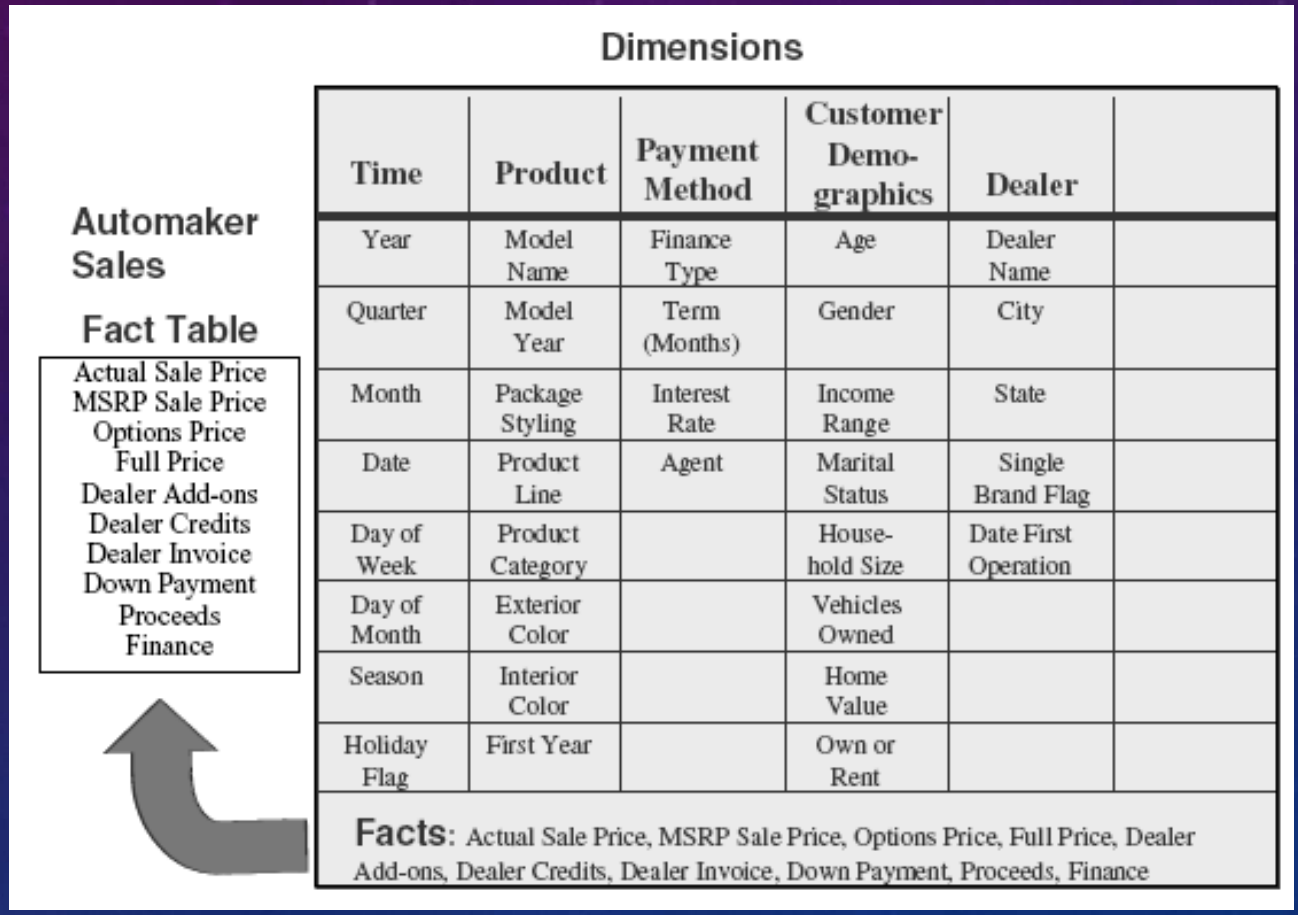
DIMENSIONAL MODELING

- Let us now examine the business dimensions for an automobile manufacturer.
- Let us say that the goal is to analyze sales.
- We want to build a data warehouse that will allow the user to analyze automobile sales in a number of ways.

DIMENSIONAL MODELING

- In the automaker sales figure, we can see the facts at the bottom.
- All these facts will form the fact table as shown in the figure.
- The fact table gets its name from the subject for analysis; in this case, it is automaker sales.
- Each fact item or measurement goes into the fact table as an attribute for automaker sales.

DIMENSIONAL MODELING

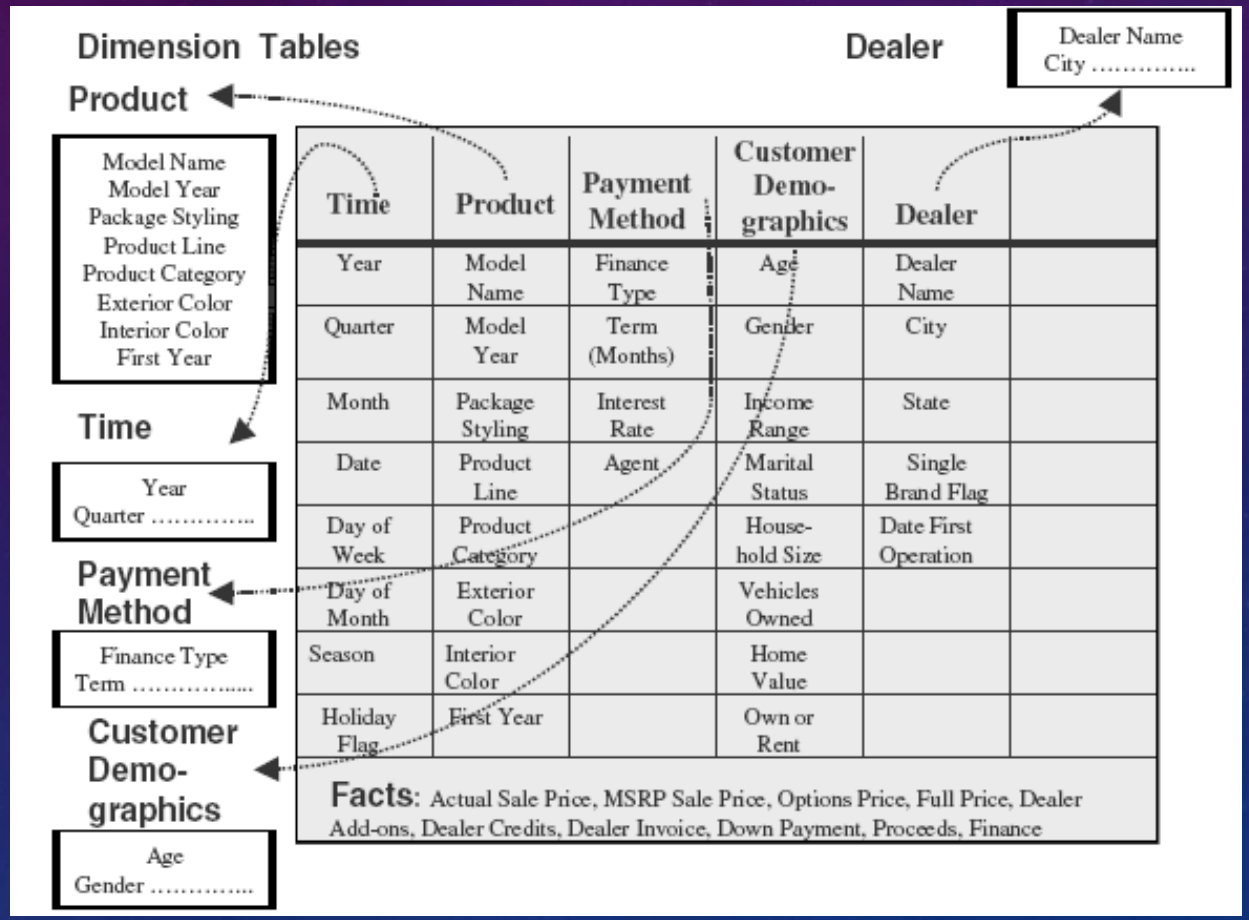


Formation of the automaker sales fact table

DIMENSIONAL MODELING

- Next step is the formation of dimension tables.
- The business dimensions are shown as column heading.
- In the automaker sales example, the business dimensions are product, dealer, customer demographics, payment method, and time.
- We can form dimension tables for each of these dimensions.
- The data items shown within each column would then be the attributes for each corresponding dimension table.

DIMENSIONAL MODELING



Formation of the automaker dimension tables

DIMENSIONAL MODELING

- Let us take one more example. In this case, we want to come up with an information package for a hotel chain.
- The subject in this case is hotel occupancy.
- We want to analyze occupancy of the rooms in the various branches of the hotel chain.
- We want to analyze the occupancy by individual hotels and by room types.
- So hotel and room type are critical business dimensions for the analysis.
- As in the other case, we also need to include the time dimension.

DIMENSIONAL MODELING

Information Subject: Hotel Occupancy

Dimensions

Time	Hotel	Room Type			
Year	Hotel Line	Room Type			
Quarter	Branch Name	Room Size			
Month	Branch Code	Number of Beds			
Date	Region	Type of Bed			
Day of Week	Address	Max. Occupants			
Day of Month	City/State/ Zip	Suite			
Holiday Flag	Construc-tion Year	Refrige-rator			
	Renova-tion Year	Kichen-nette			

Facts: Occupied Rooms, Vacant Rooms, Unavailable Rooms, Number of Occupants, Revenue

Hierarchies / Categories

DIMENSIONAL MODELING

- Here are some of the criteria for combining the tables into a dimensional model:
  - The model should provide the best data access.
  - The whole model must be query-centric.
  - It must be optimized for queries and analyses.
  - The model must show that the dimension tables interact with the fact table.
  - It should also be structured in such a way that every dimension can interact equally with the fact table.

END OF SLIDES