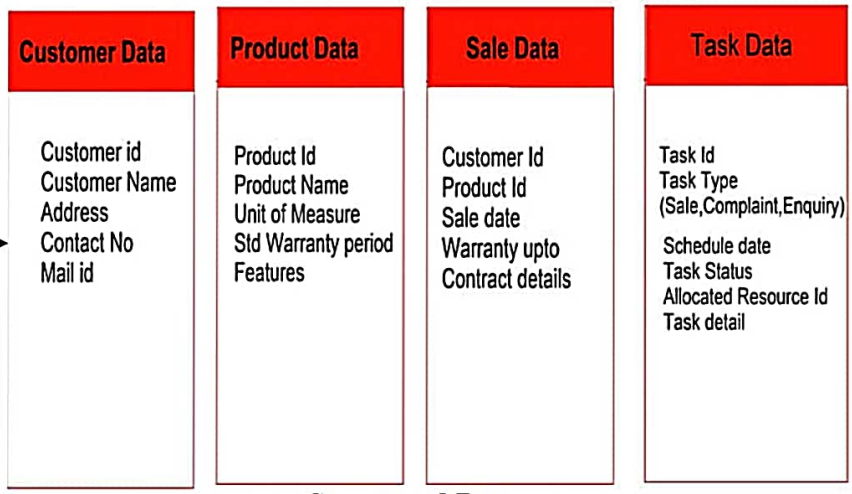
What is Data Modeling?

- A data model is an abstract view of the data referred to in the product being developed.
- Data Modeling explaining & visualizing how the data be used by the software and defining data objects that will be stored in a database. **Example:**

Unstructured Data

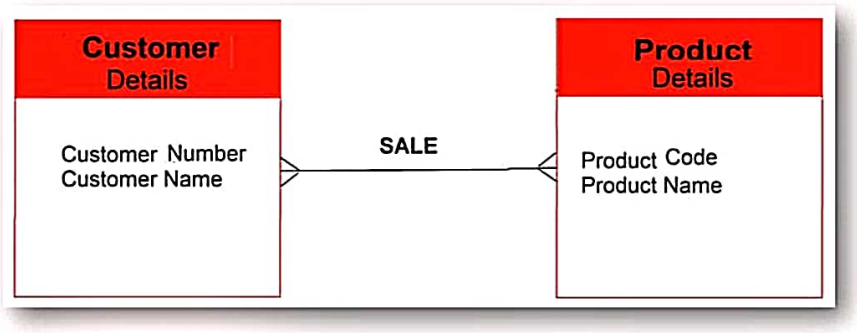


Structured Data

Types of Data Models

Type 1: Conceptual Data Model

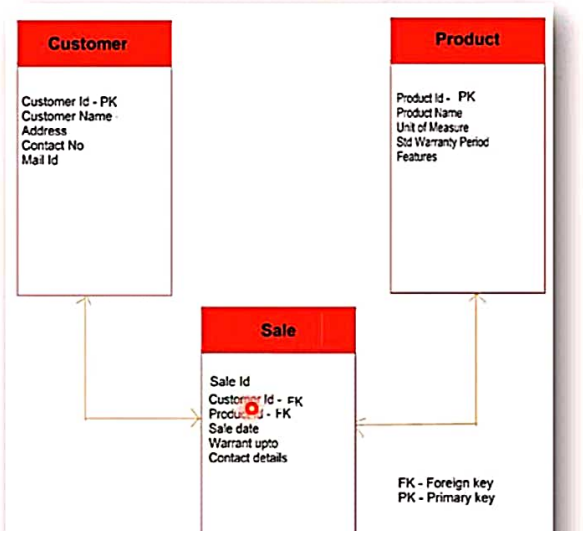
- It gives a front view of the of each data entity and not a technical detail of data.
- It presents all the **data entities** referred to by the business and their characteristics called **attributes** and the dependency or the connection between entities called **Relationships.**
- It makes communication easy and clear as all stakeholders & reducing communication gaps.



Types of Data Models

Type 2: Logical Data Model

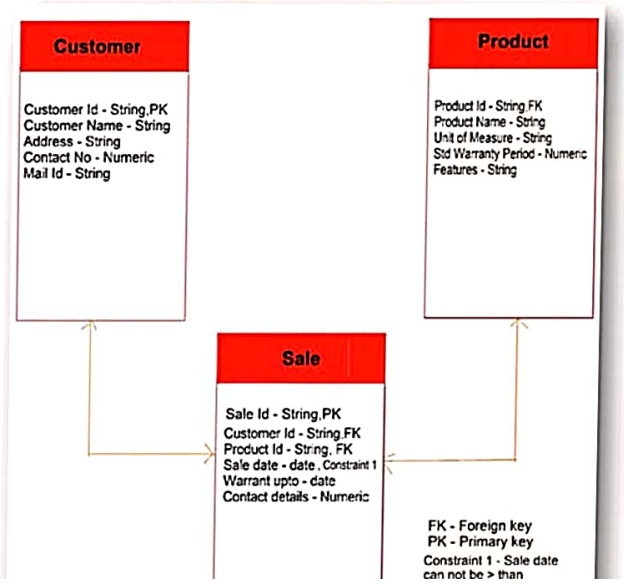
- It gives a detailed description of each data entity their attributes & relationship between two entities giving business purpose to each data.
- The relationship is presented more explicitly with details like Primary key, Foreign key & Parent-child dependent entity type.



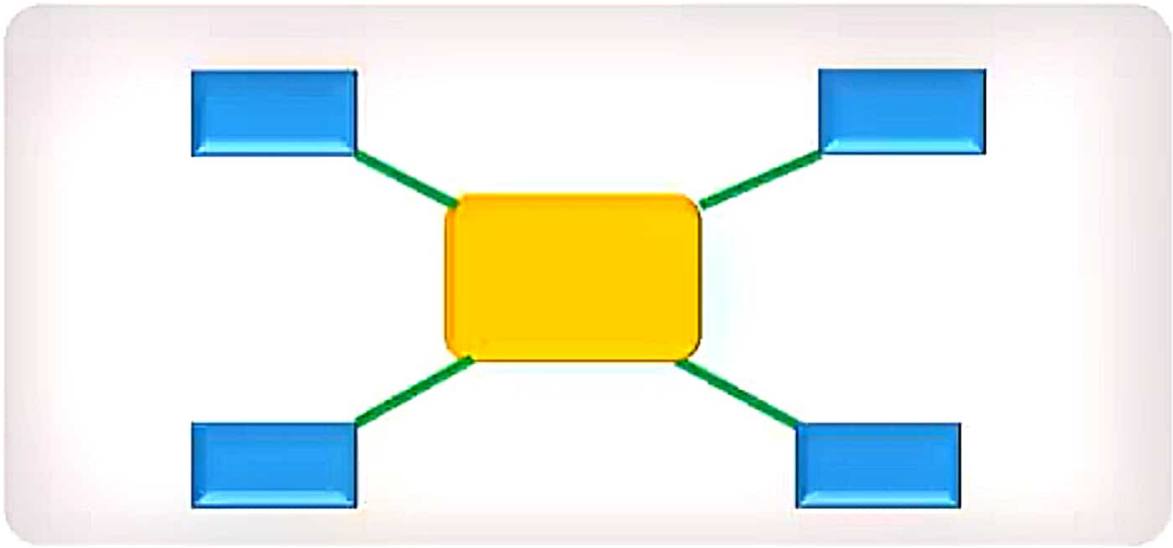
Types of Data Models

Type 3: Physical Data Model

- A physical data model is the layout of the actual database with all its components.
- It gives a technical view of the data i.e., the table name, column name, data type, constraints, indexes, primary key, triggers, stored procedures, etc.
- It is developed by using databases SQL, ORACLE etc



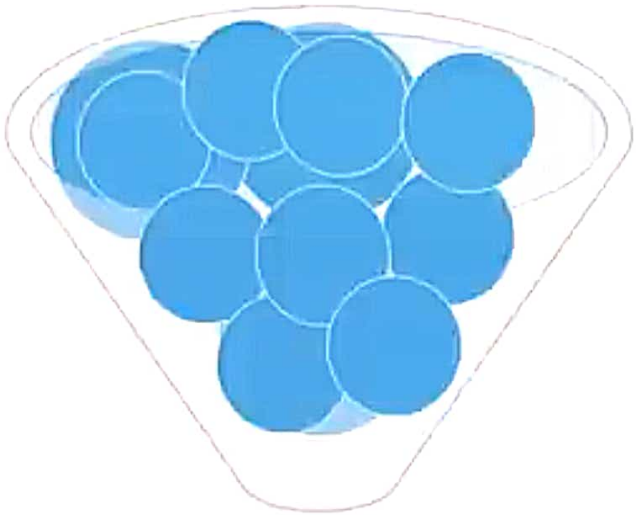
Dimension & Fact

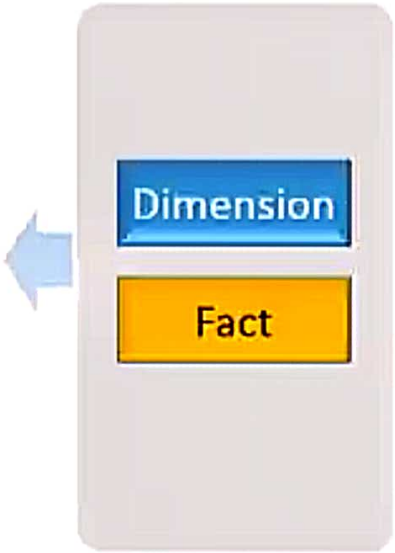


Dimension

Fact

Dimension & Fact

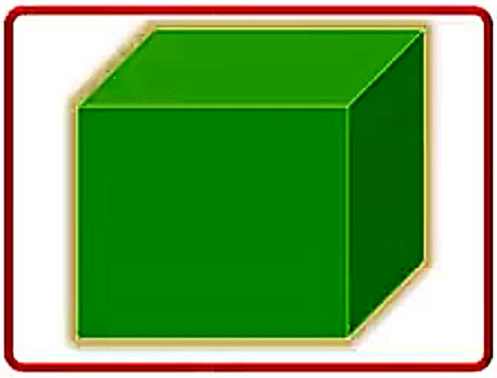




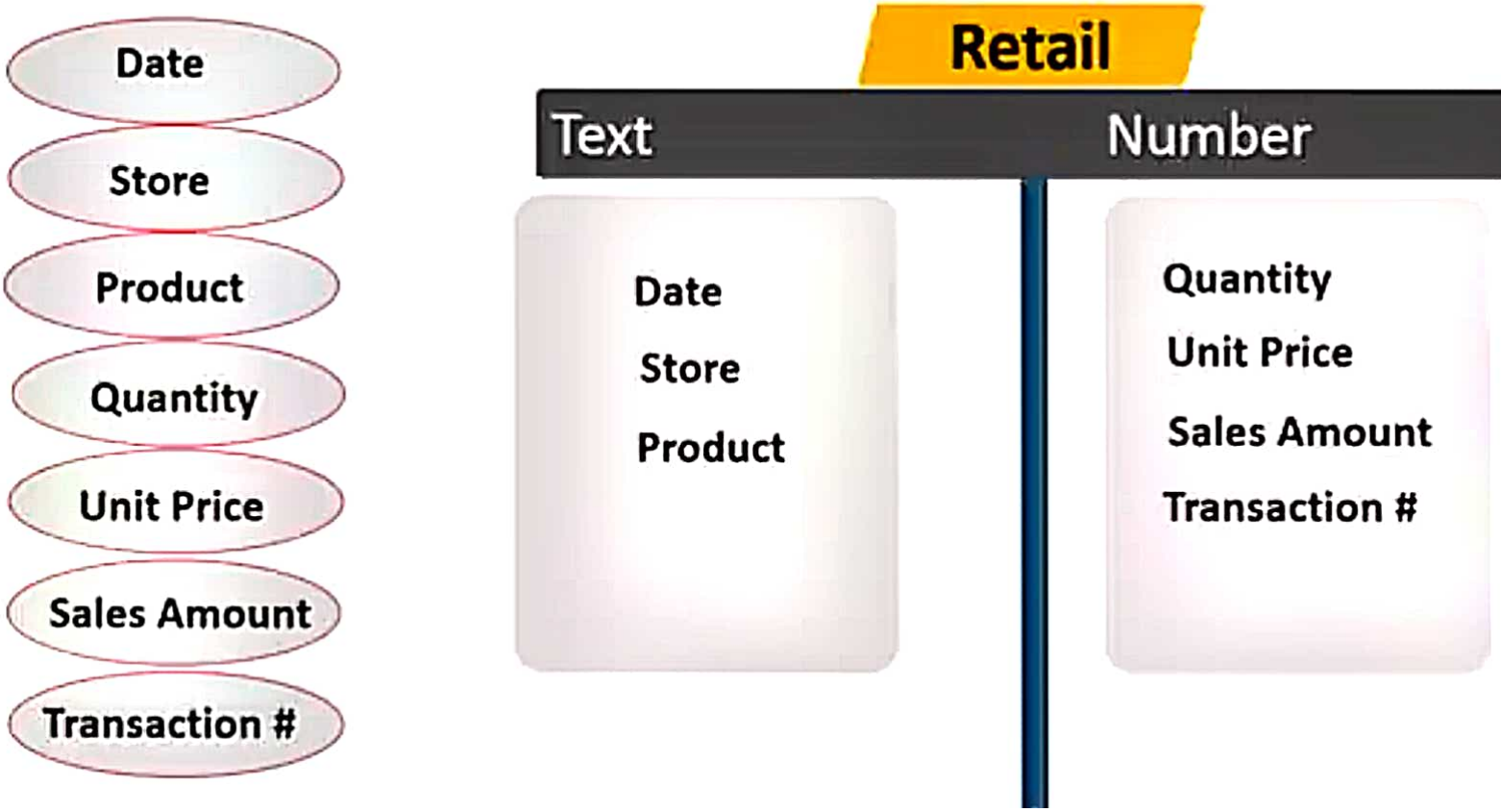
Dimension & Fact

- Date
- Store
- Product
- Quantity
- Unit Price
- Sales Amount
- Transaction #

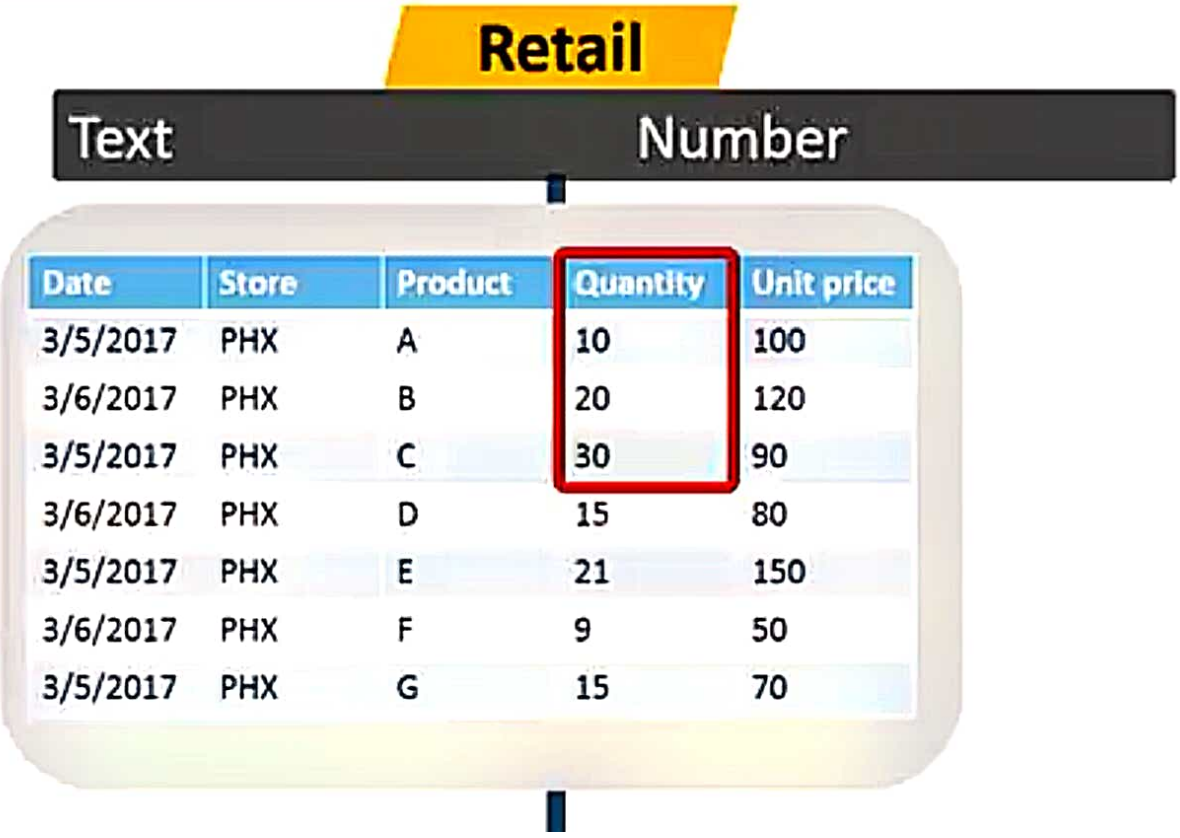
Retail



Dimension & Fact

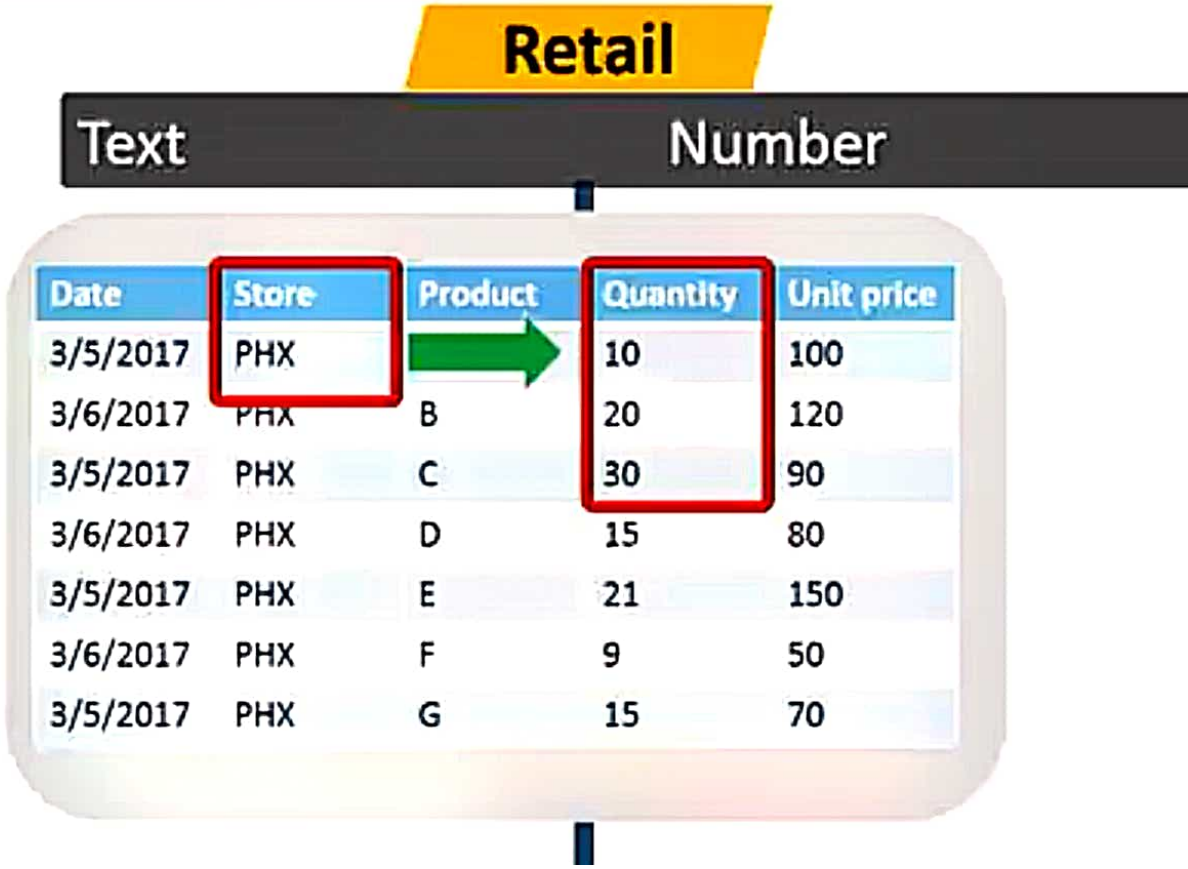


Dimension & Fact

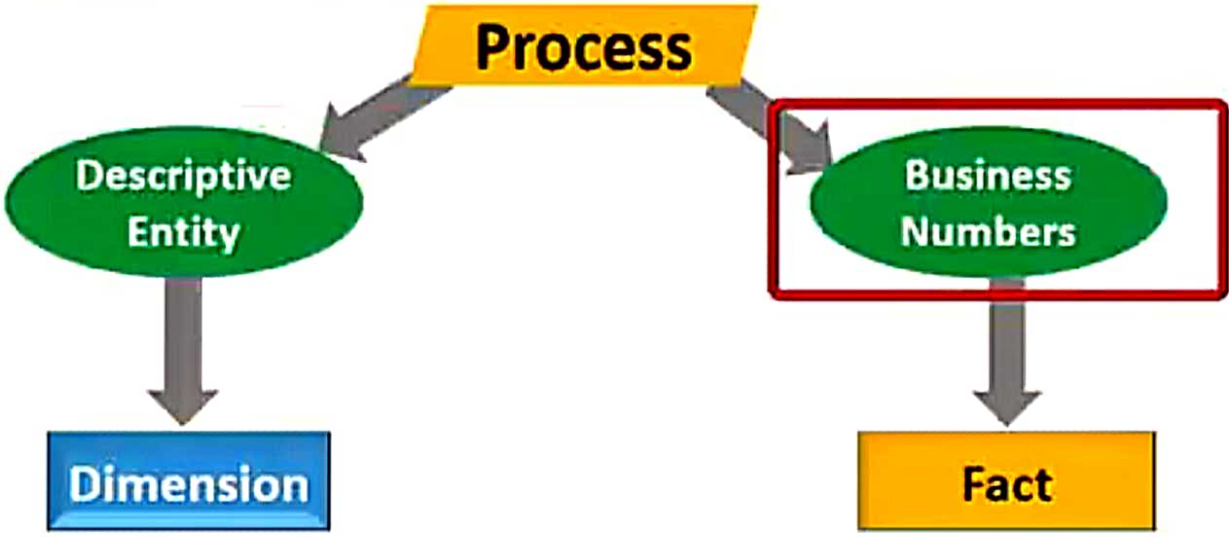


Date	Store	Product	Quantity	Unit price
3/5/2017	PHX	A	10	100
3/6/2017	PHX	B	20	120
3/5/2017	PHX	C	30	90
3/6/2017	PHX	D	15	80
3/5/2017	PHX	E	21	150
3/6/2017	PHX	F	9	50
3/5/2017	PHX	G	15	70

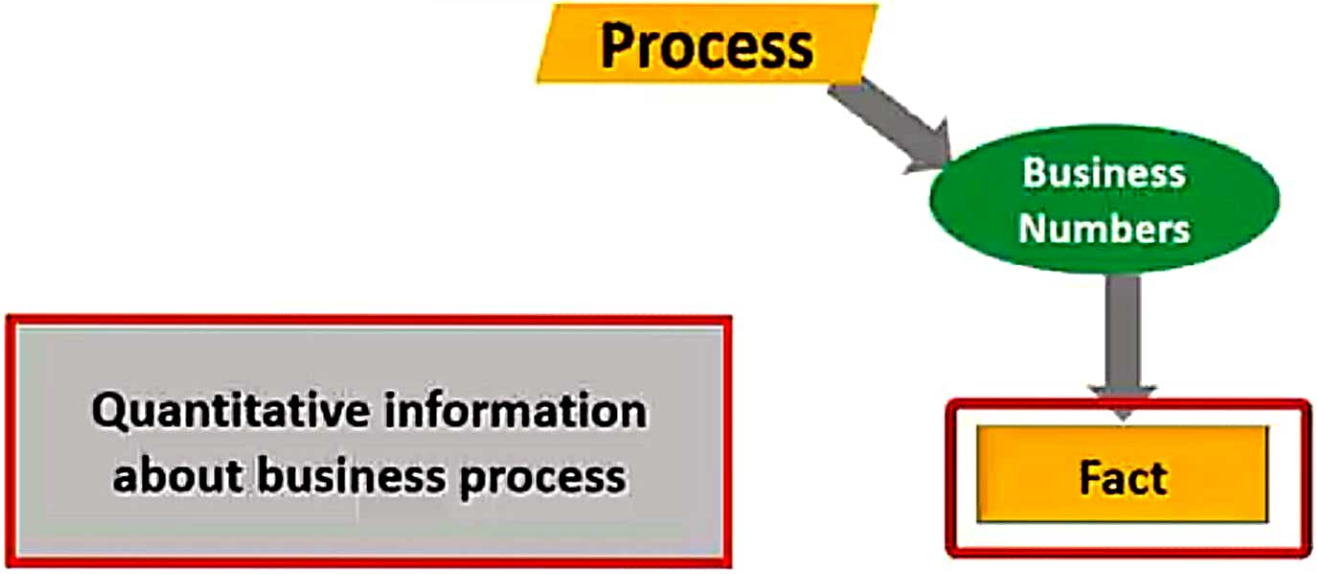
Dimension & Fact



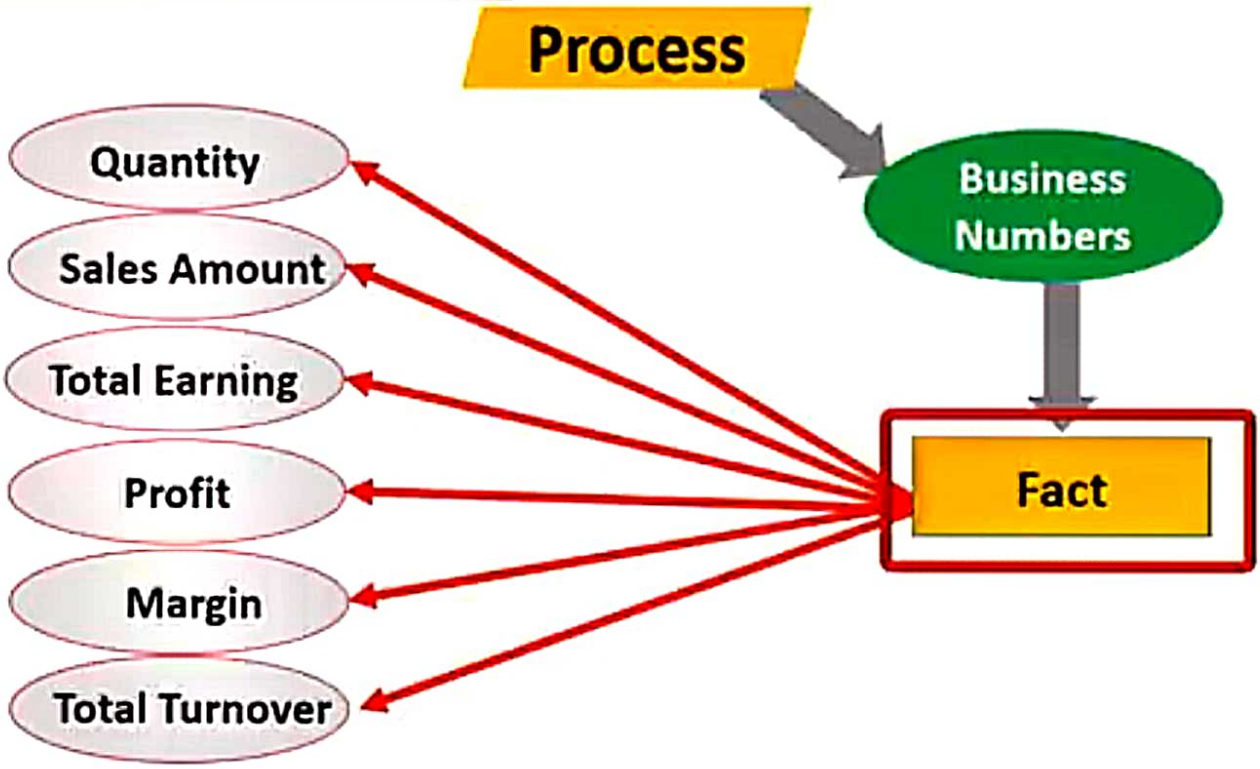
Dimension & Fact



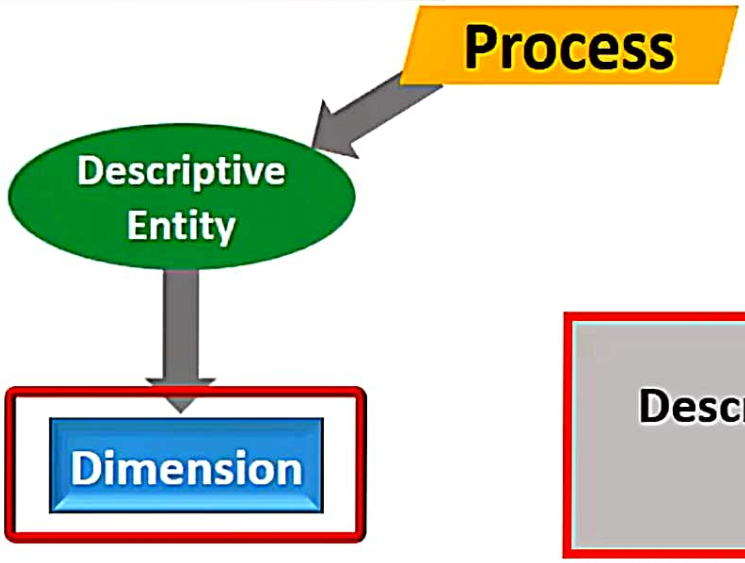
Dimension & Fact



Dimension & Fact

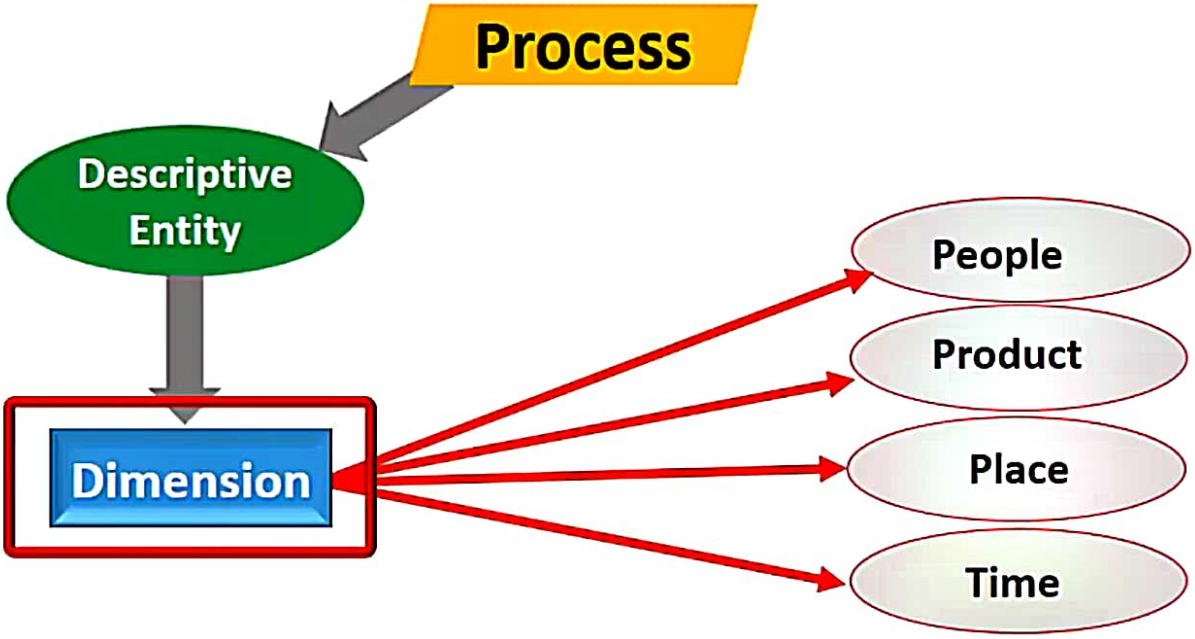


Dimension & Fact



Describe facts or business numbers

Dimension & Fact



Customer_id (PK)	Customer_name	Phone_number	State
Q7A55Q8A	Leonard F.	855-966-566	California
A8DG8B2H	Sherlock H	967-968-563	NewYork
Q8W2C26F	Monica Bing	852-563-988	New York
D2B2GNHY	Rachel	859-563-556	New York
C56T89THE	Sheldon Cooper	859-898-456	California

Customers
(dim table)

Product_id (PK)	Product Name	Product Description
ABHEUMSHR	ABC Laptop	ABC Laptop, 8BG, 256GB
OLRGIDTBRM	ABC Mobile	ABC Mobile, 6.1, 4GB, 64 GB
DHBVDVADV	ABC Laptop Charger	ABC laptop charger with adaptor
LKMTNYKOEF	Headphones	Wireless headphones
ZDVBFSGBGD	Tony TV	56' Smart TV

Products
(dim table)

Customer_id

Product_id

Transaction_id	Product_id (FK)	Customer_id (FK)	Transaction_date	Quantity	Amount
123456789	ABHEUMSHR	Q7A55Q8A	24th Mar 2021	1	30000
547896124	OLRGIDTBRM	A8DG8B2H	01st Nov 2021	1	10000
475896125	DHBVDVADV	Q8W2C26F	11th Feb 2021	5	500
859641237	LKMTNYKOEF	D2B2GNHY	31st Dec 2021	1	1500
178594025	ZDVBFSGBGD	A8DG8B2H	14th Jan 2021	12	1200
685247830	OLRGIDTBRM	C56T89THE	06th April 2021	10	10000
965214738	DHBVDVADV	D2B2GNHY	25th May 2021	15	15000
605853459	DHBVDVADV	Q7A55Q8A	8th Oct 2021	6	600
259035721	ABHEUMSHR	Q7A55Q8A	19th Jun 2021	2	56000
741256780	OLRGIDTBRM	A8DG8B2H	01ST Dec 2021	2	10000

Transactions
(fact table)

Fact tables and Dimension tables

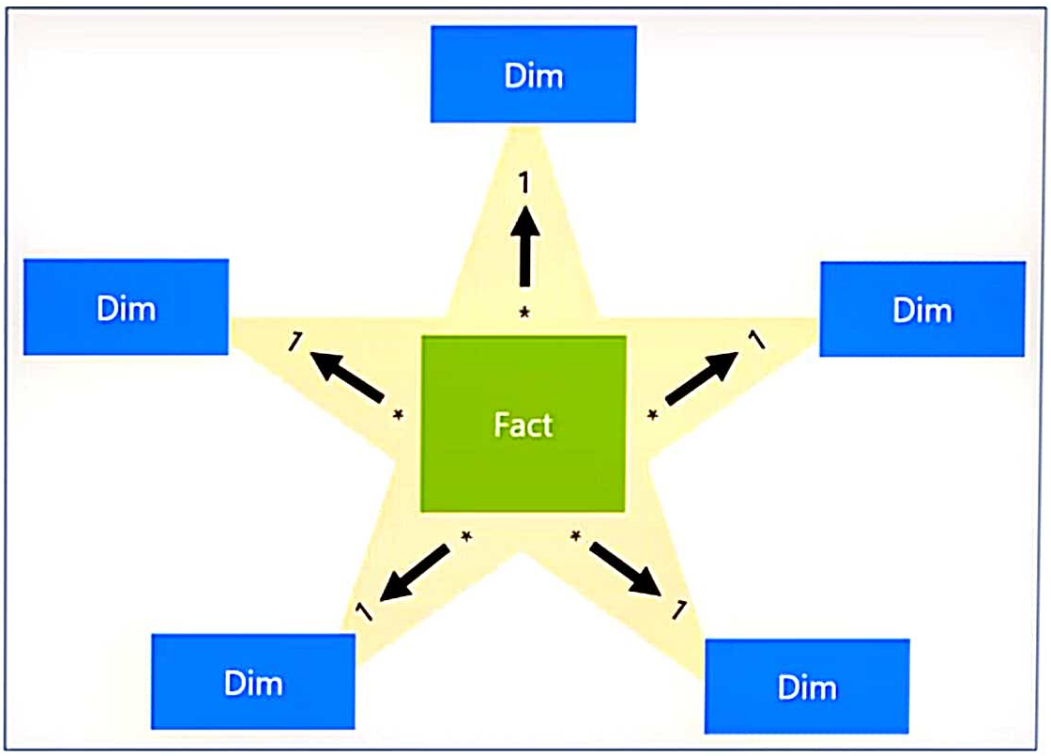
- Fact tables:
  - It contains measurements, facts or metrics of the attributes
  - Hold no meaning in itself
  - Numeric and quantifiable
  - Created or loaded after dimensions are loaded
  - Primary Key is a new column identifying the unique row, references dim tables with FK
- Dimension tables:
  - Gives context to facts, holds attributes for the facts
  - Created or loaded before facts are loaded
  - Primary key is referenced by fact tables

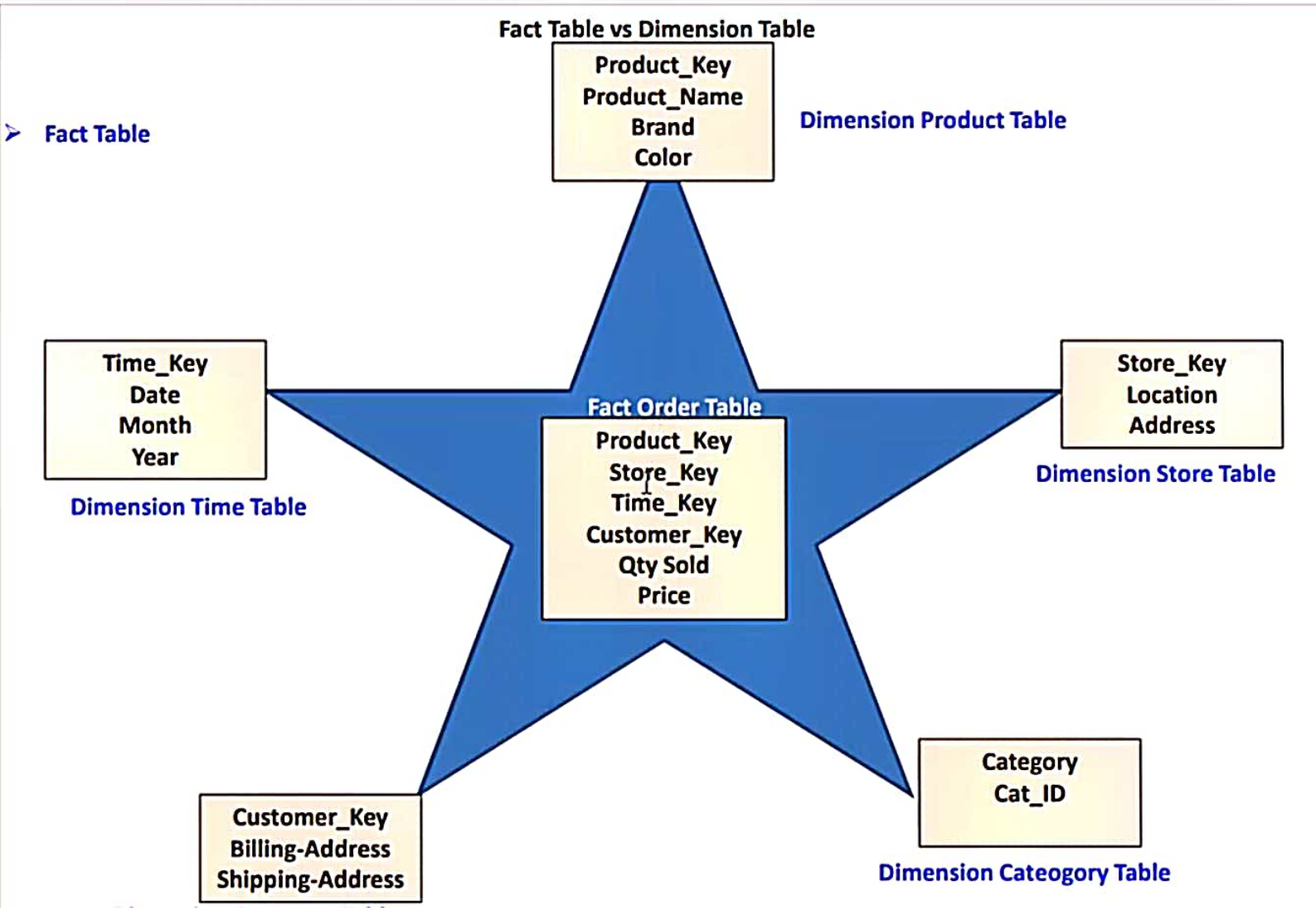
Star schema overview

- **Star schema** is a mature modeling approach widely adopted by relational data warehouses.
- It requires **modelers** to classify their model tables as either **dimension or fact.**
  - **Dimension tables** describe business entities—the *things* you model. Entities can include products, people, places, and concepts including time itself. The most consistent table you'll find in a star schema is a date dimension table. A dimension table contains a key column (or columns) that acts as a unique identifier, and descriptive columns.
  - **Fact tables** store observations or events, and can be sales orders, stock balances, exchange rates, temperatures, etc. A fact table contains dimension key columns that relate to dimension tables, and numeric measure columns.

Star schema overview

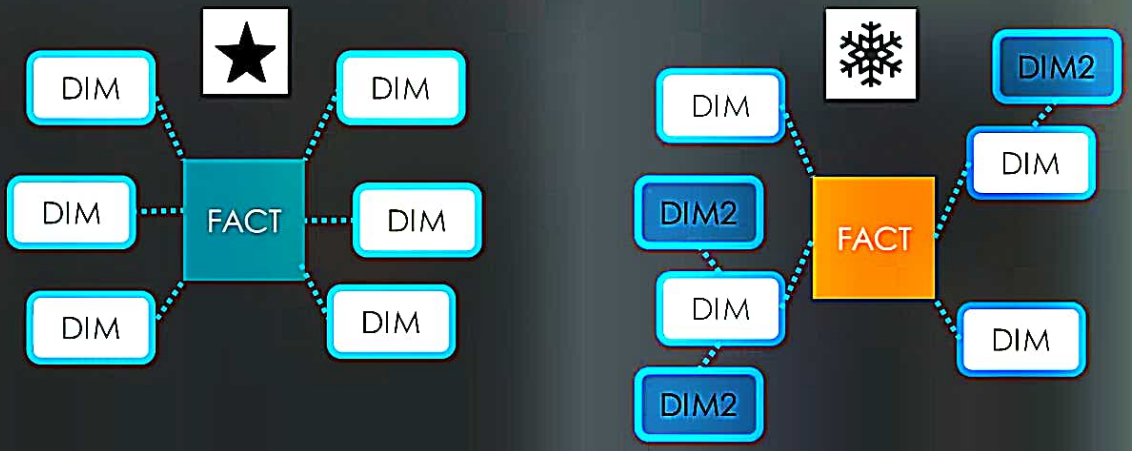
- **Star schema** is a mature modeling approach widely adopted by relational data warehouses.
- It requires **modelers** to classify their model tables as either **dimension or fact.**




Fact Table vs Dimension Table
Fact Table
Product_Key
Product_Name
Brand
Color
Dimension Product Table
Time_Key
Date
Month
Year
Dimension Time Table
Fact Order Table
Product_Key
Store_Key
Time_Key
Customer_Key
Qty Sold
Price
Store_Key
Location
Address
Dimension Store Table
Customer_Key
Billing-Address
Shipping-Address
Category
Cat_ID
Dimension Cateogory Table

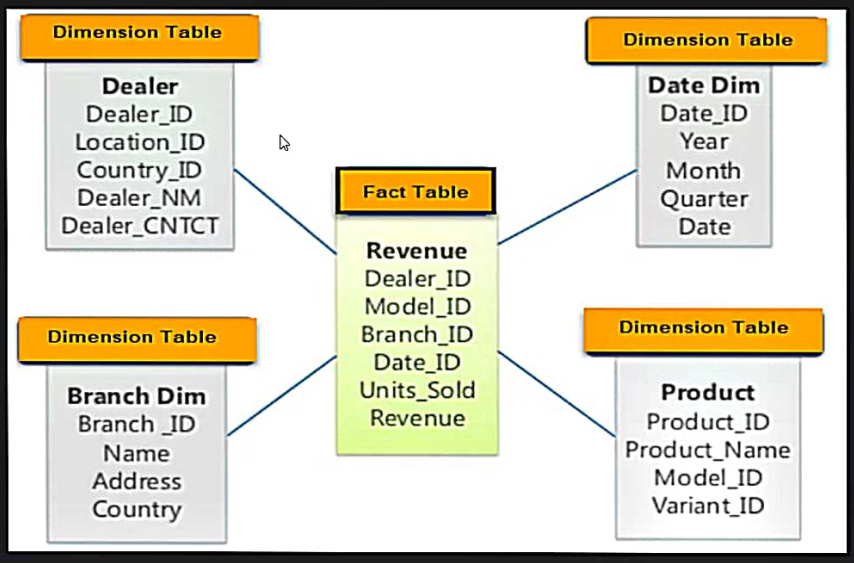
Data Warehouse - Schemas

- Schema is a logical description of the entire database.
- It includes the name and description of records of all record types including all associated data-items and aggregates.



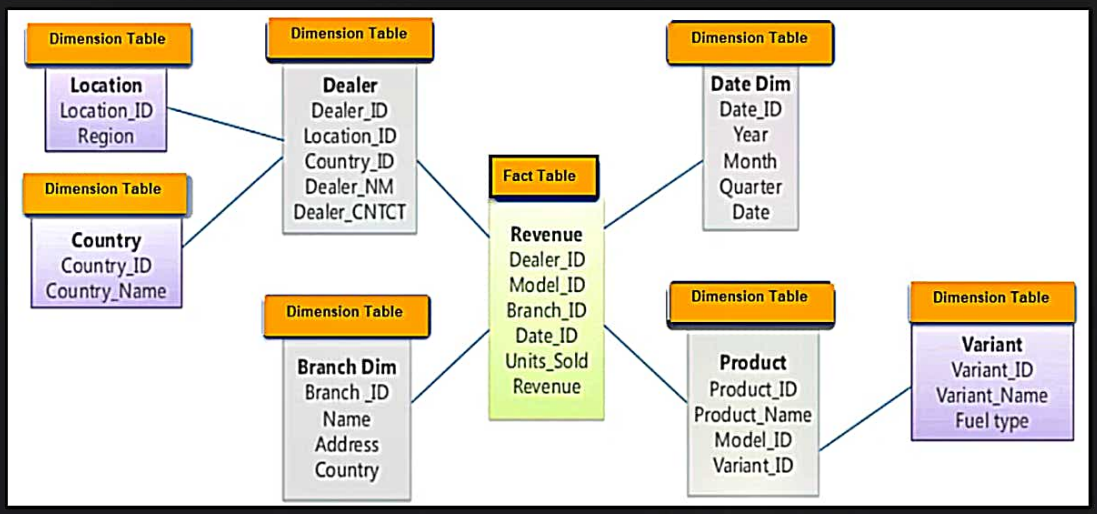
Star Schema-

Each dimension in a star schema is represented with only one-dimension table.



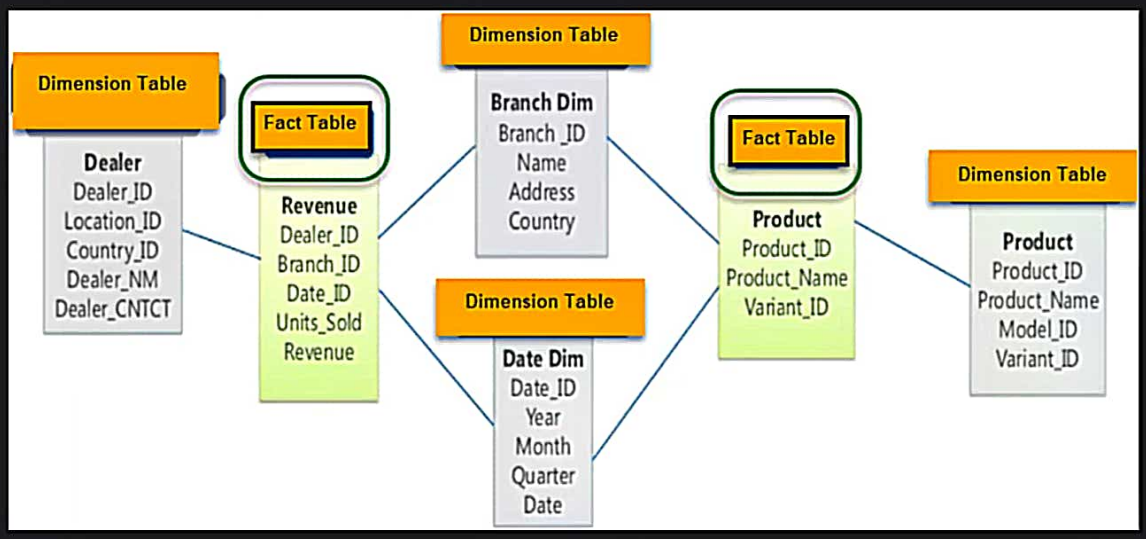
Snowflake Schema

- Some dimension tables in the Snowflake schema are normalized.
- The normalization splits up the data into additional tables.



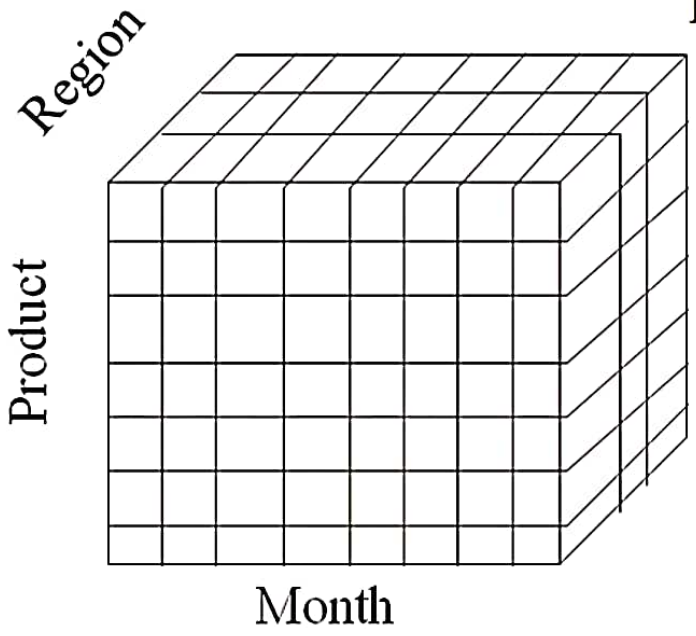
Fact Constellation Schema-

A fact constellation has multiple fact tables. It is also known as galaxy schema.

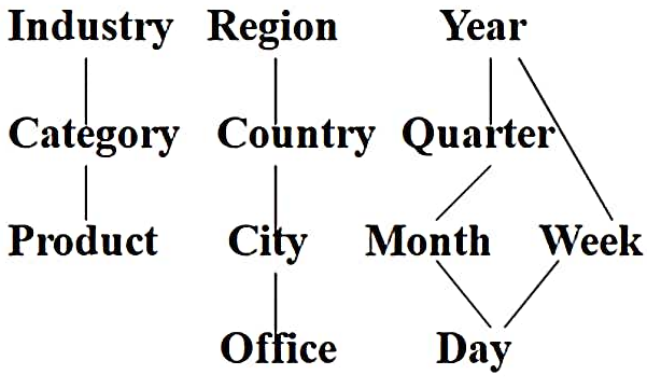


Multidimensional Data

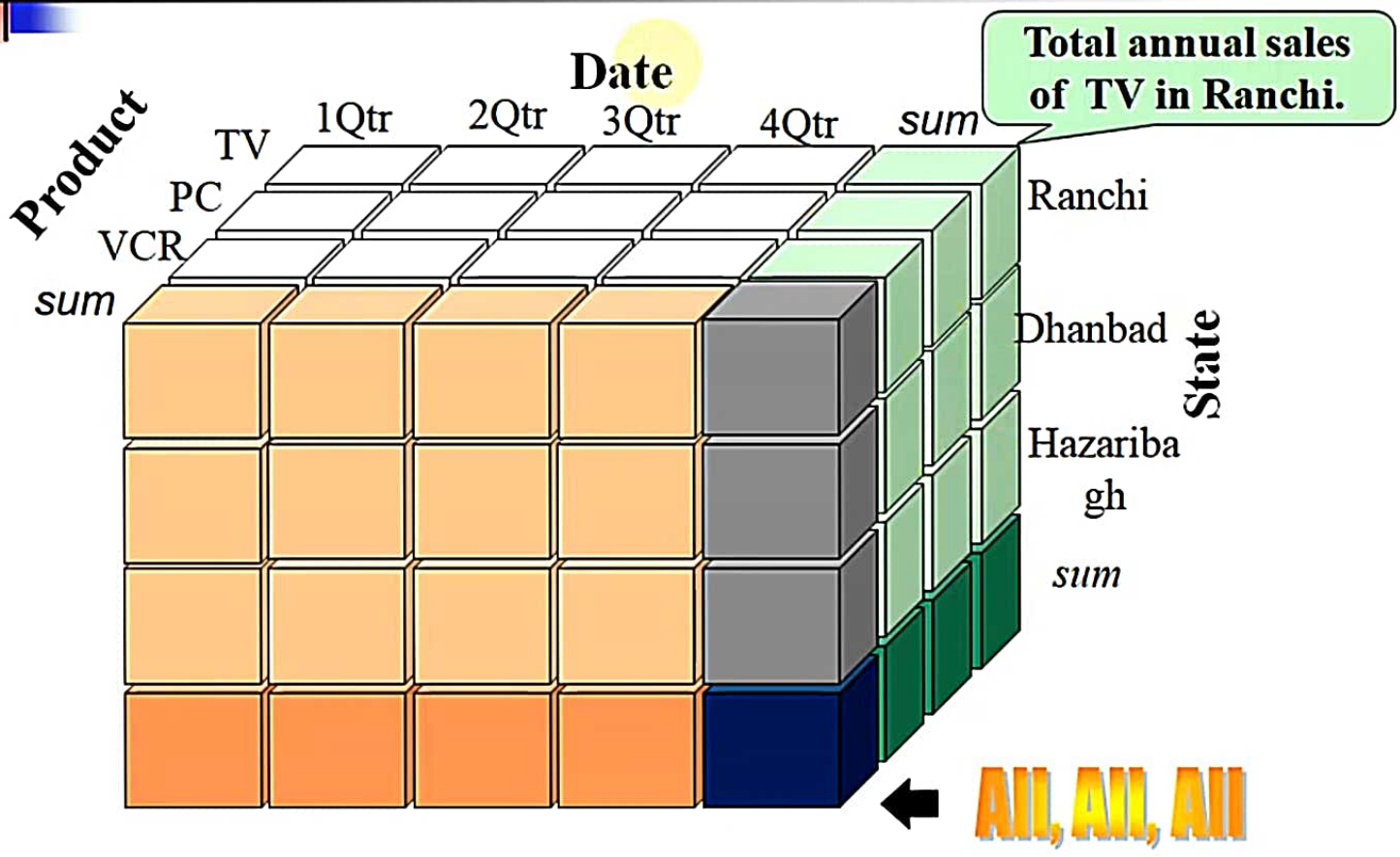
- Sales volume as a function of product, month, and region



Dimensions: Product, Location, Time
Hierarchical summarization paths



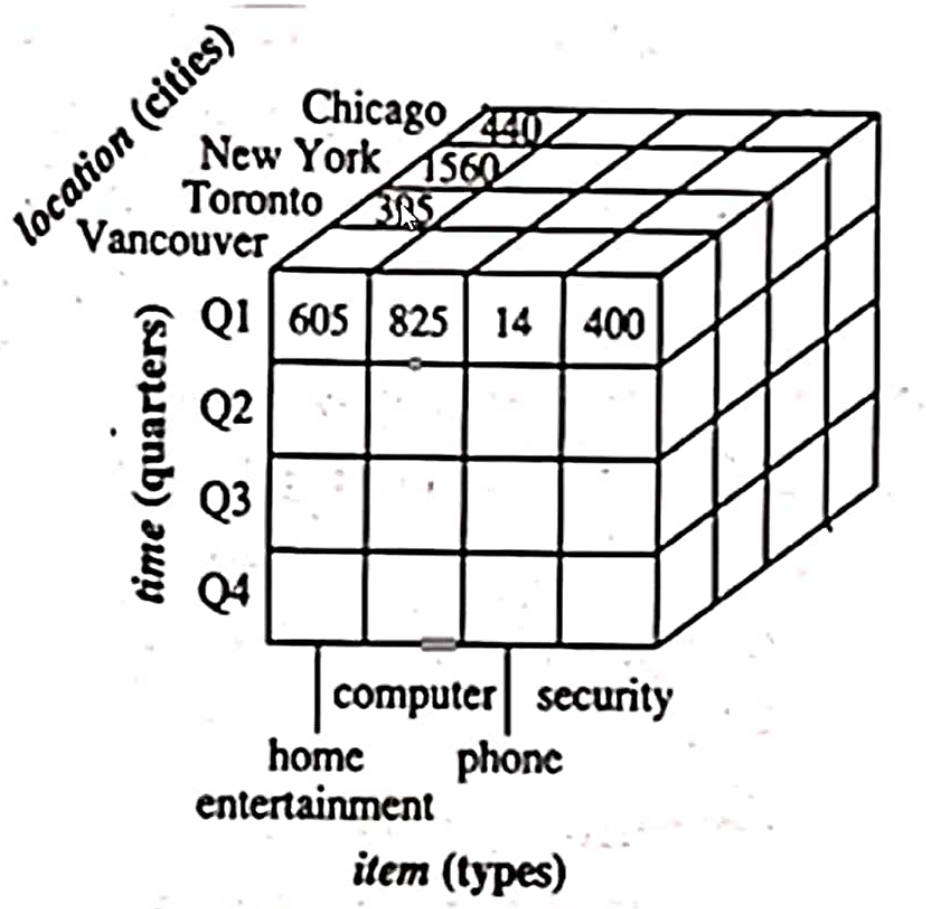
A Sample Data Cube



Typical OLAP Operations

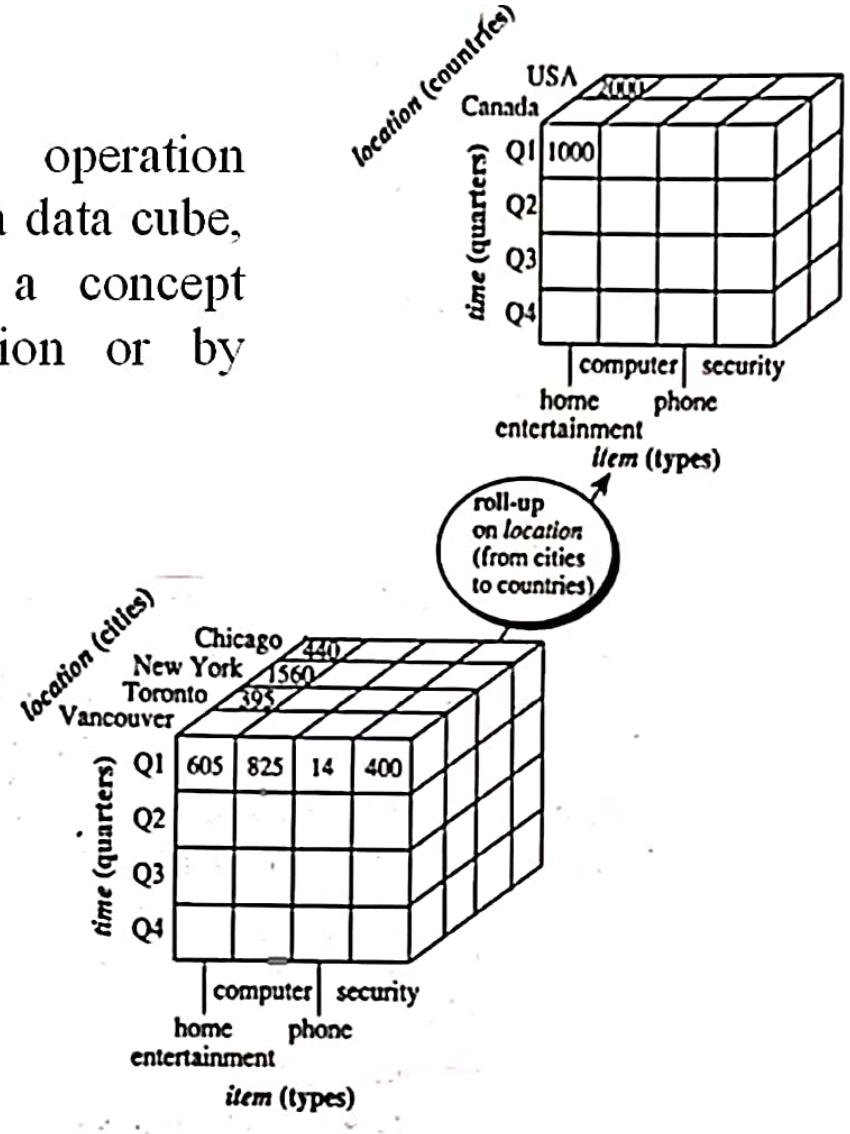
- In the multidimensional model, **data are organized into multiple dimensions**, and each **dimension contains multiple levels of abstraction defined by concept hierarchies**.
- This organization provides **users with the flexibility to view data from different perspectives**.
- A number of **OLAP data cube operations exist to materialize these different views**, allowing **interactive querying and analysis of the data at hand**.
- Hence, OLAP provides a **user-friendly environment for interactive data analysis**.

Figure 1.12 shows a data cube for AllElectronics sales.



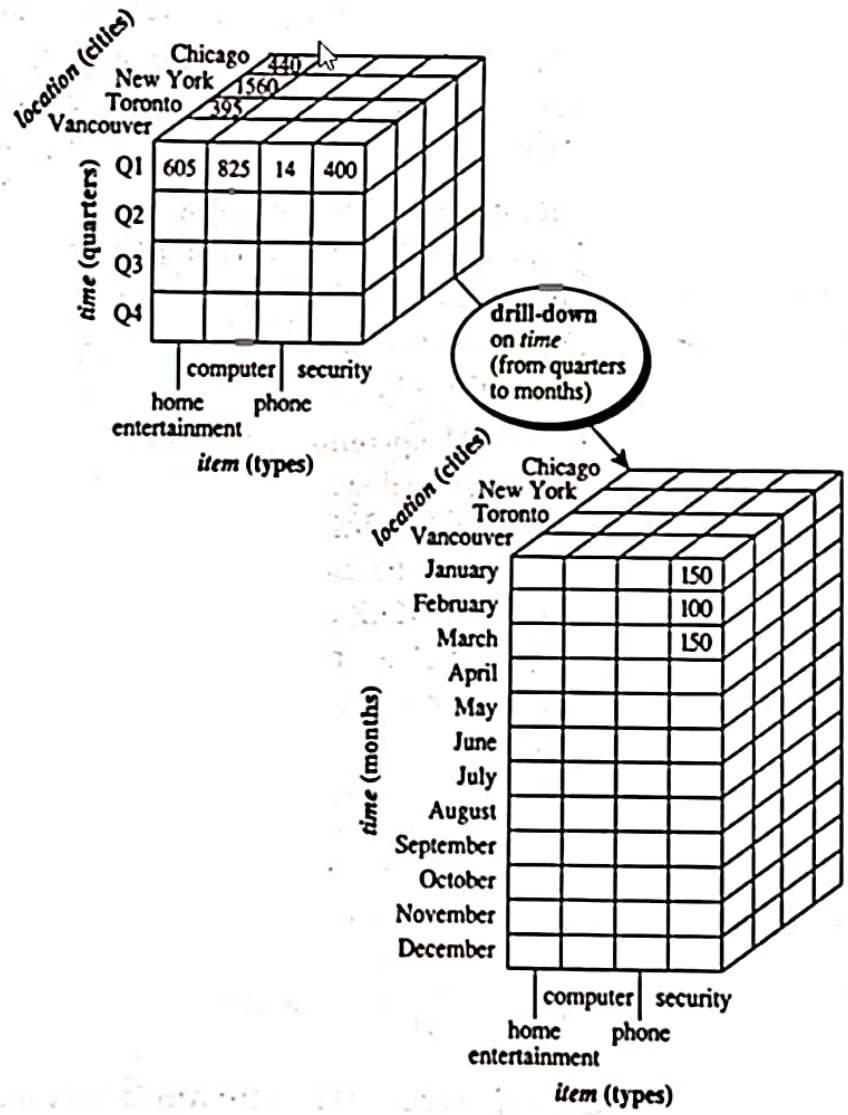
1. Roll-up

The roll-up (drill-up) operation performs aggregation on a data cube, either by climbing up a concept hierarchy for a dimension or by dimension reduction.



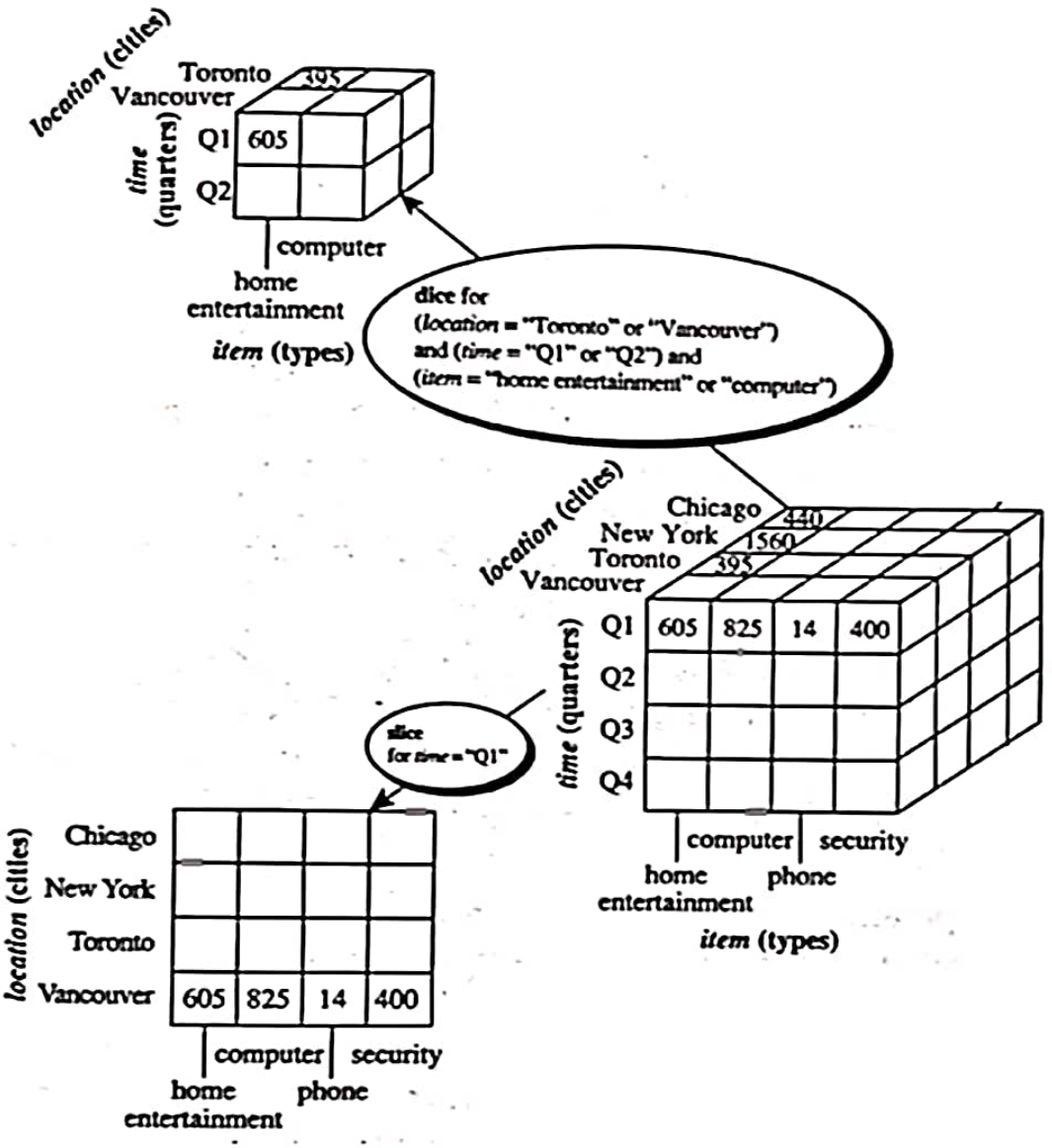
2. Drill-down

Drill-down can be realized by either stepping down a concept hierarchy for a dimension or introducing additional dimensions.



3. Slice and dice

- The slice operation performs a selection on one dimension of the given cube, resulting in a subcube.
- The dice operation defines a subcube by performing a selection on two or more dimensions.


location (cities)
Toronto
Vancouver
395
time (quarters)
Q1
605
Q2
computer
home entertainment
item (types)
dice for
(location = "Toronto" or "Vancouver")
and (time = "Q1" or "Q2") and
(item = "home entertainment" or "computer")
location (cities)
Chicago
440
New York
1560
Toronto
395
Vancouver
Q1
605
825
14
400
Q2
Q3
Q4
time (quarters)
slice
for time = "Q1"
computer
security
home entertainment
phone
item (types)
location (cities)
Chicago
New York
Toronto
Vancouver
605
825
14
400
computer
security
home entertainment
phone

4. Pivot (rotate)

- Pivot (also called rotate) is a visualization operation that rotates the data axes in view to provide an alternative data presentation.