Decision Support Systems
aka
Analytical Systems

Decision Support Systems

- Systems that are used to transform data into information, to manage the organization: OLAP vs OLTP

OLTP vs OLAP

	Transactions	**Analysis**
Objectives	Operations	Decisions
Users	Operatives	Managers and analysists
Detail	Analytic	May be aggregated
Data origin	Internal	Internal and external
Applications	Known a priori	Ad hoc
Data items Per operation	Few (tens)	Many (millions)
Updates	Frequent, Small	Rare, Mostly massive
State of data	Current data	Historic
Stored in	Data Base	Data Warehouse / DSS

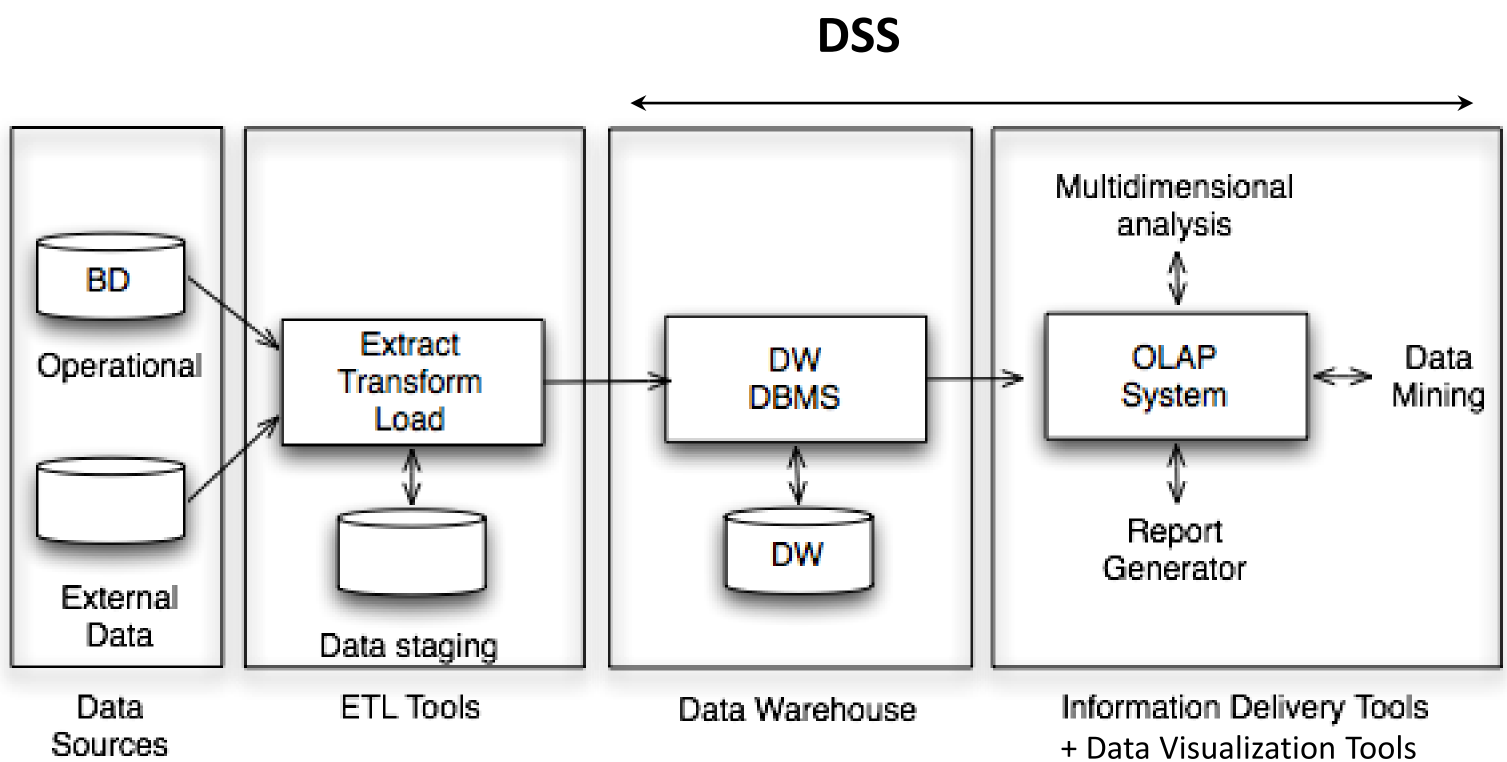
Operational systems vs DSS

- Operational systems:
  - Data are organized in a DB.
  - Data are managed by a traditional DBMS.
  - Applications used to perform structured operational activities
- Decision Support Systems:
  - Data are organized in a separate specialized DB (Data Warehouse (DW)).
  - Data are managed by a specialized DBMS.
  - Business Intelligence applications used to analyze data

Data Warehouse

- The first definition of data warehouse was provided by William Inmon in 1990.
- A DW is a specialized database
  - static (non volatile),
  - with integrated data from different data sources
  - organized to analyze subjects of interest,
  - with historical data,
  - used to produce summarized data to support decision-making processes

Building the data warehouse



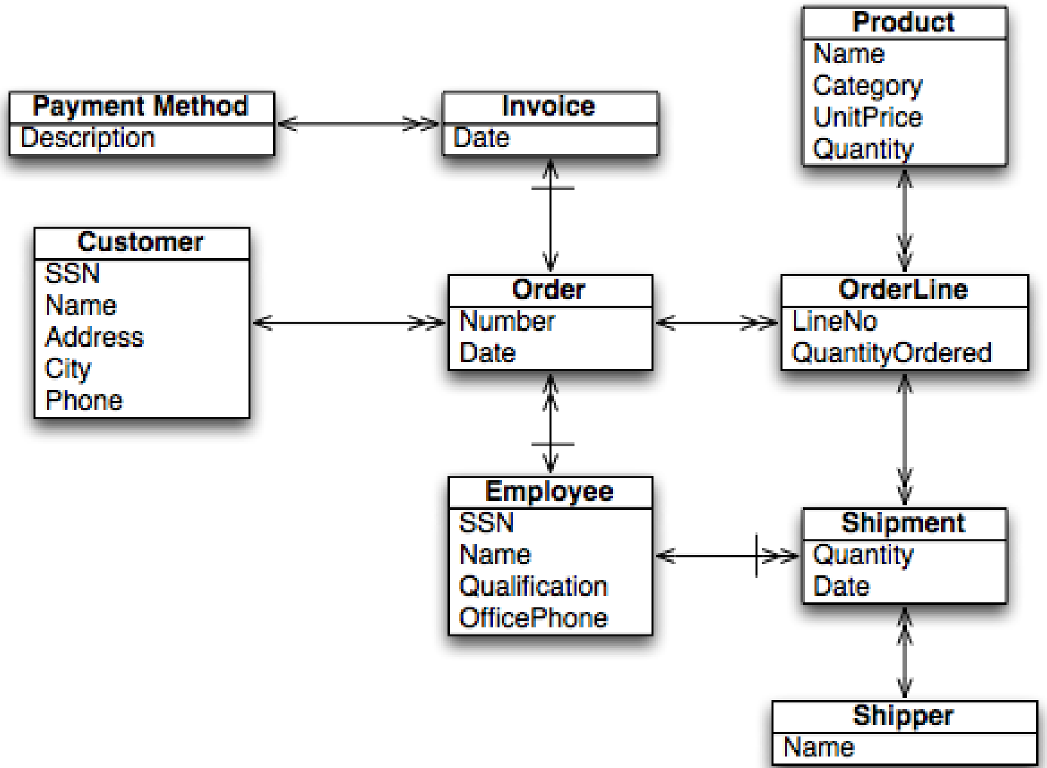
DW design

- The creation of a data warehouse takes place gradually, at different levels of abstraction: a **conceptual model**, a **logical model** and **physical model**.

Data models for DW

- The Dimensional Fact Model (DFM) is a graphical conceptual model.
- The Relational Data Model, as a logical model
- The Multidimensional Model (called Cube), useful to understand OLAP operations

Design of a Data Mart



Number of items ordered,
by product, **by** customer, **by** month

Total revenue **by** product category,
by customer, **by** year

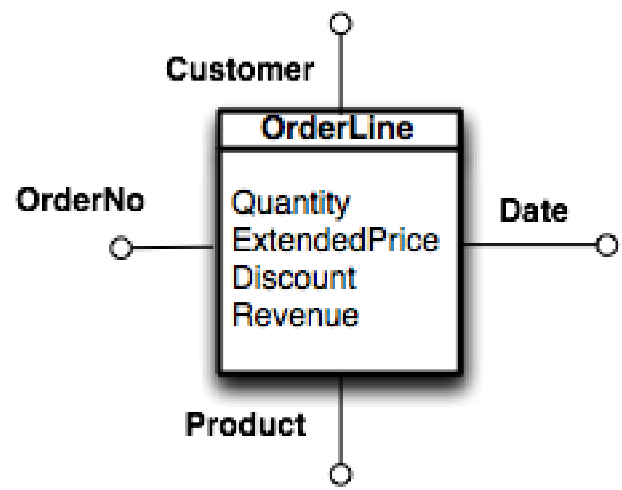
Total revenue from Italian customers **by** customer city,
by year, **by** quarter

DATA BASE

BUSINESS QUESTIONS

A data model for conceptual design

- Object data model:
  - Classes with
    - Attributes
- Data warehouse data model
  - Facts tables with
    - Measures
    - Dimensions

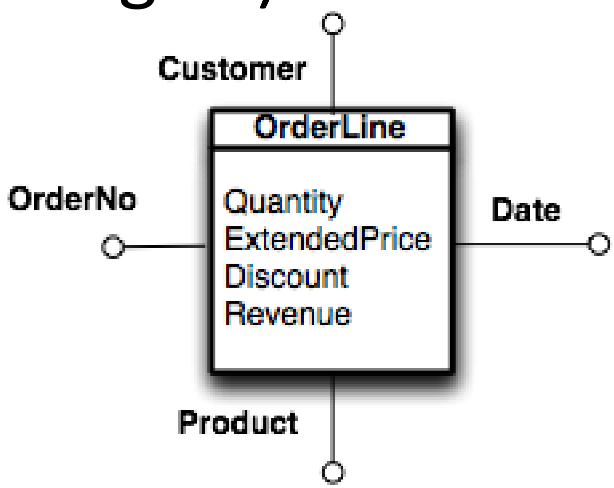


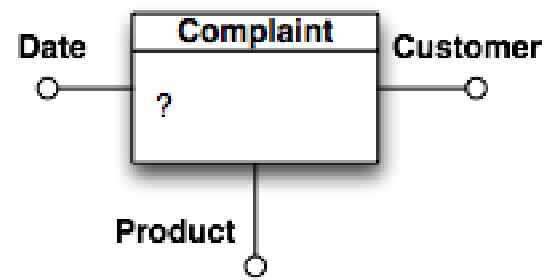
The facts

- Facts are typically transactions / events
  - Sales, clicks, complains, visits
- Periodic facts
  - One fact for a group of transactions made over a period of time.
  - Example: the monthly balance of all monthly transactions
- Accumulating fact
  - One fact for the entire lifetime of an evolving event that has a duration and change over time
  - Example: the life cycle of a mortgage application

The measures

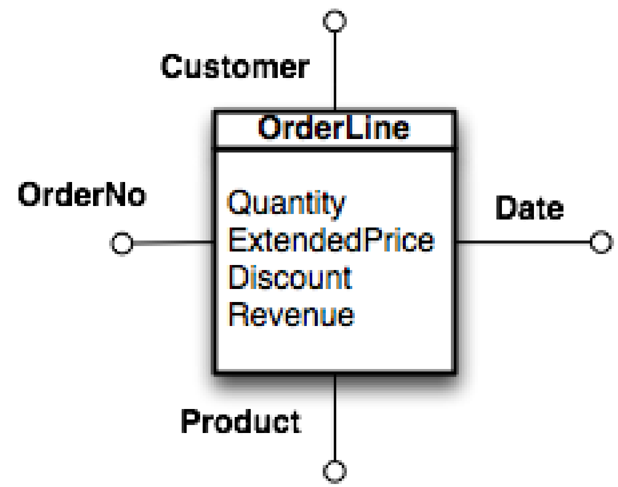
- Measures: quantative information whose aggregation is of interest
- Aggregation is often Sum, but not always (count, average…)



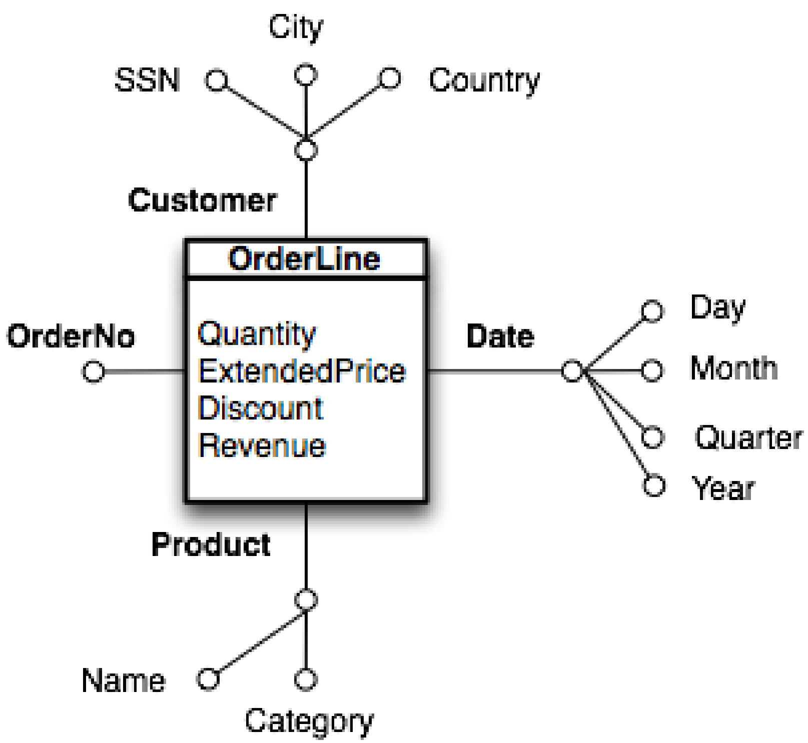


Dimensions

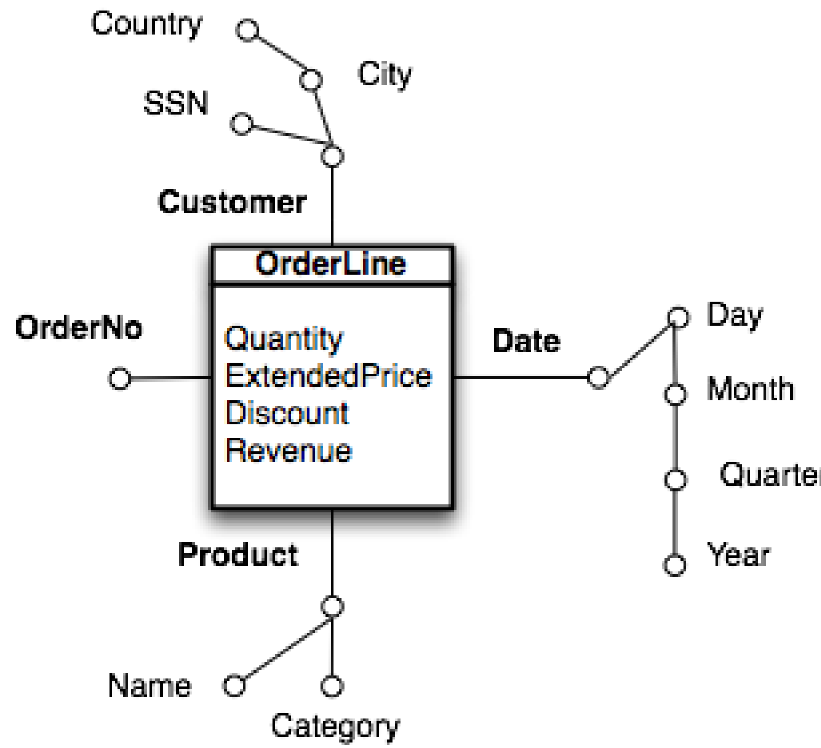
- Dimensions relate to who, what, why, when, and where:
  - Who is involved? What is about? When happens? Where occurs?
- Dimensions: the variables that influence the measures and indicate possible intervention levers
- How does a measure depend on a dimension?



Dimensions have attributes and hierarchies



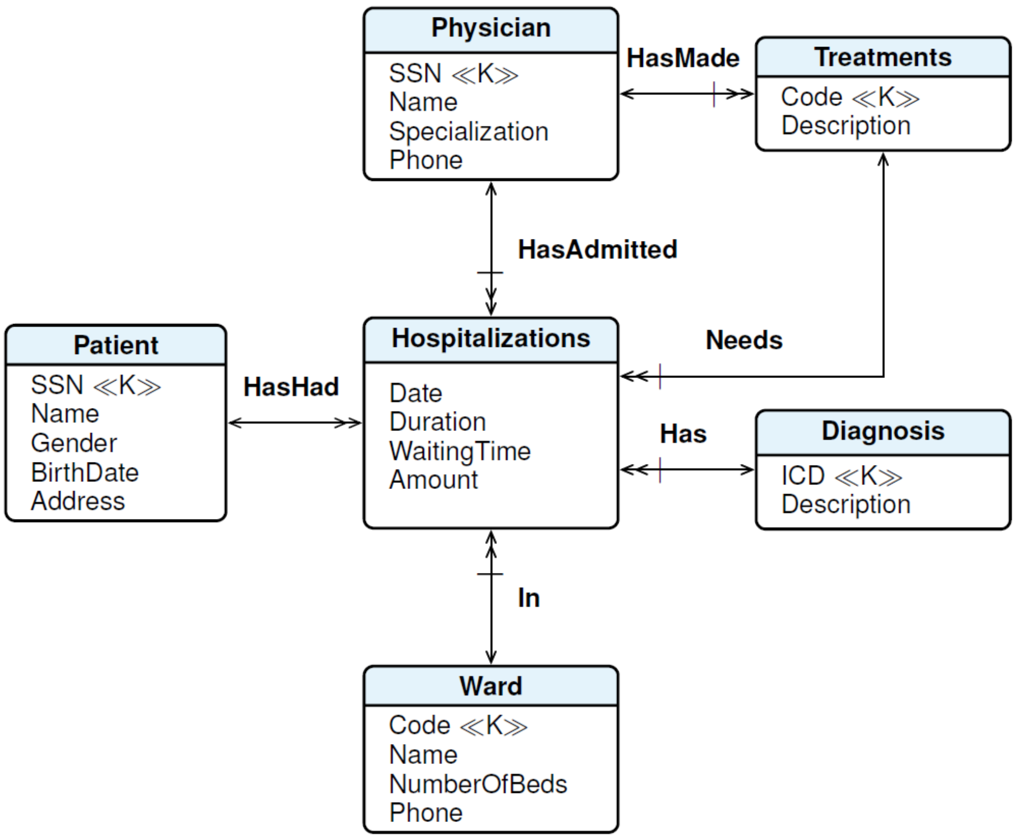
Without hierarchies



With hierarchies

Case study

An hospital needs a DM to extract information from their operational database with information about inpatients treatments.



1. Total billed amount for hospitalizations, **by** diagnosis code and description, **by** month (year).
2. Total number of hospitalizations and billed amount, **by** ward, **by** patient gender (age at date of admission, city, region).
3. Total billed amount, average length of stay and average waiting time, **by** diagnosis code and description, **by** name (specialization) of the physician who has admitted the patient.
4. Total billed amount, and average waiting time of admission, **by** patient age (region), **by** treatment code (description).

Requirements specification

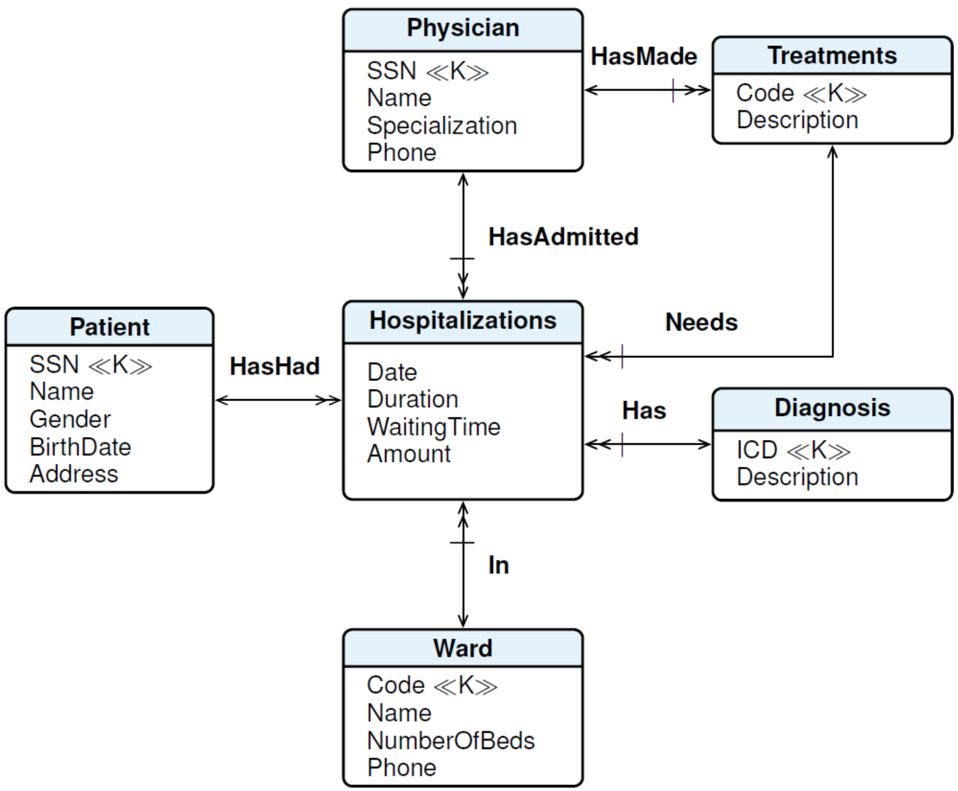
Requirements analysis	Dimensions	Measures	Metrics
Total billed amount for hospitalizations, by diagnosis code and description, by month (year).	Diagnosis (ICD, Description), Date (Month, Year)	Amount	Total Amount
Total number of hospitalizations and billed amount, by ward, by patient gender (age at date of admission, city, region).	Ward, Patient (Gender, Age, City, Region)	Amount	Total number Total Amount
Total billed amount, average length of stay and average waiting time by diagnosis code and description, by name (specialization) of the physician who admitted the patient.	Diagnosis (ICD code, Description), Physician (Name, Specialization)	Amount, Duration, WaitingTime	Total Amount Average Duration Average WaitingTime
Total billed amount, and average waiting time for admission by patient age (region), by treatment code (description).	Patient (Age, Region), Treatment (Code, Description)	Amount, Duration, WaitingTime	Total Amount Average WaitingTime

Requirements specification

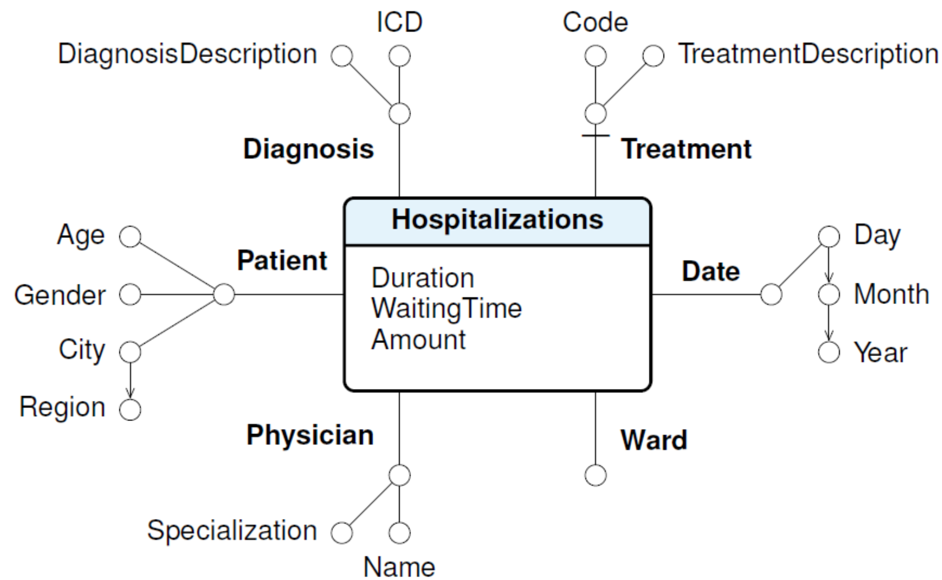
Fact granularity

Description	A fact is a hospitalization of a patient, assuming that they may require one treatment only
Preliminary dimensions	Patient, Date, Ward, Diagnosis, Treatment, Physician
Preliminary measures	Duration, WaitingTime, Amount

Data mart conceptual schema



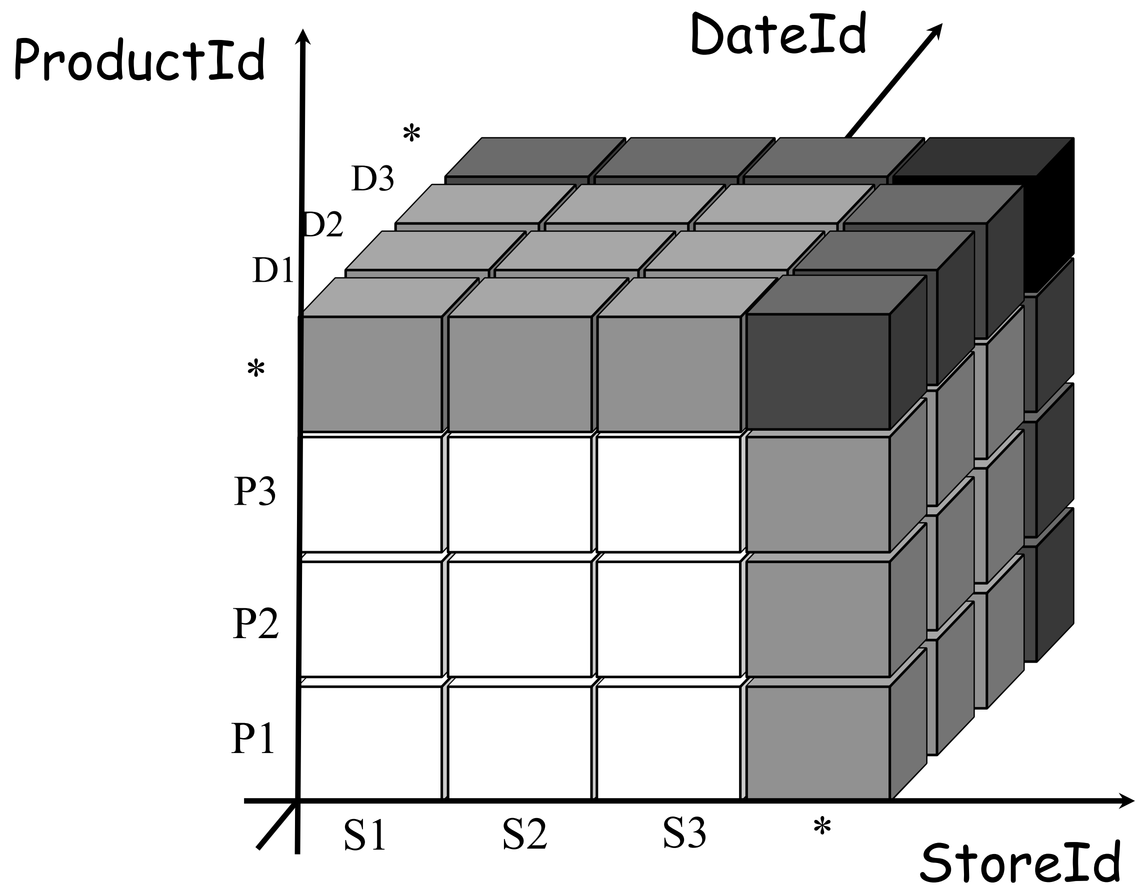
DATA BASE



DATA MART

Multidimensional model (cube)

- The cube model helps understanding interactive data analysis

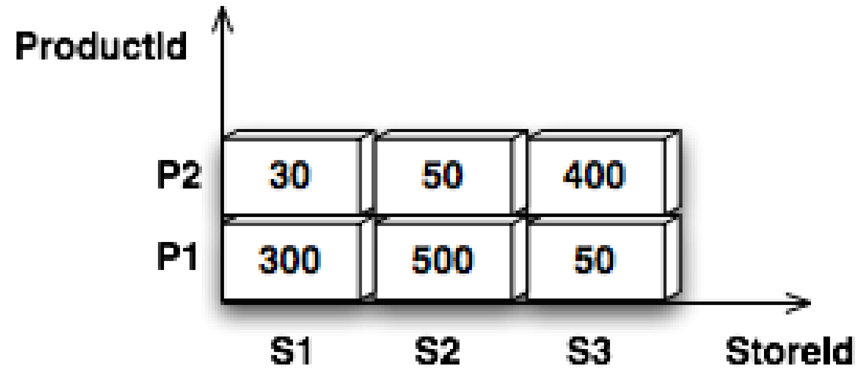


2-D Cube

Sales

StoreId	ProductId	Qty
S1	P1	300
S2	P1	500
S3	P1	50
S1	P2	30
S2	P2	50
S3	P2	400

Fact Table



2-D Cube

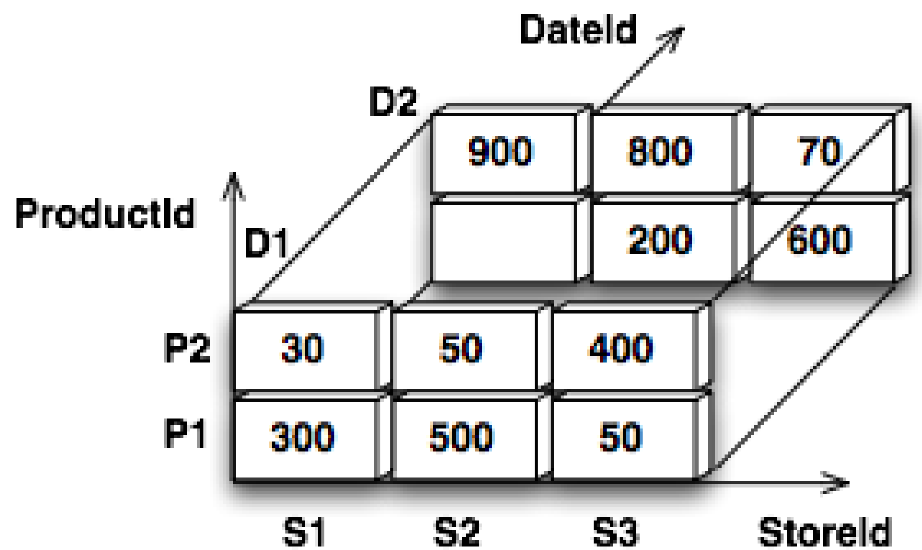
	StoreId		
ProductId	S1	S2	S3
P1	300	500	50
P2	30	50	400

Cross tabulation

3-D Cube

Sales

StoreId	ProductId	DateId	Qty
S1	P1	D1	300
S2	P1	D1	500
S3	P1	D1	50
S1	P2	D1	30
S2	P2	D1	50
S3	P2	D1	400
S2	P1	D2	200
S3	P1	D2	600
S1	P2	D2	900
S2	P2	D2	800
S3	P2	D2	70

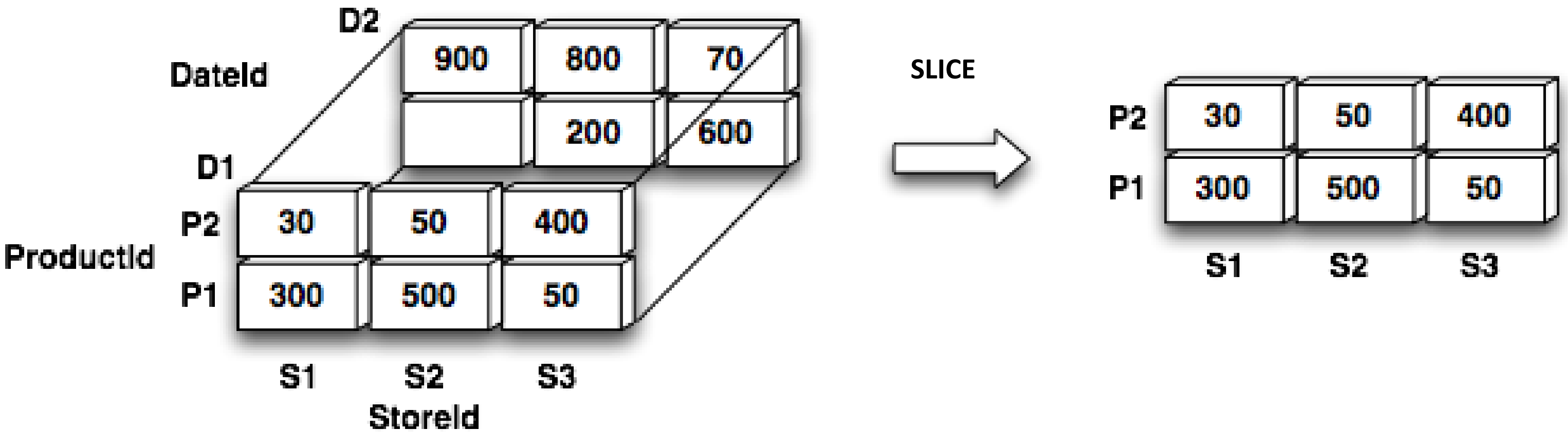


Fact Table

3-D Cube

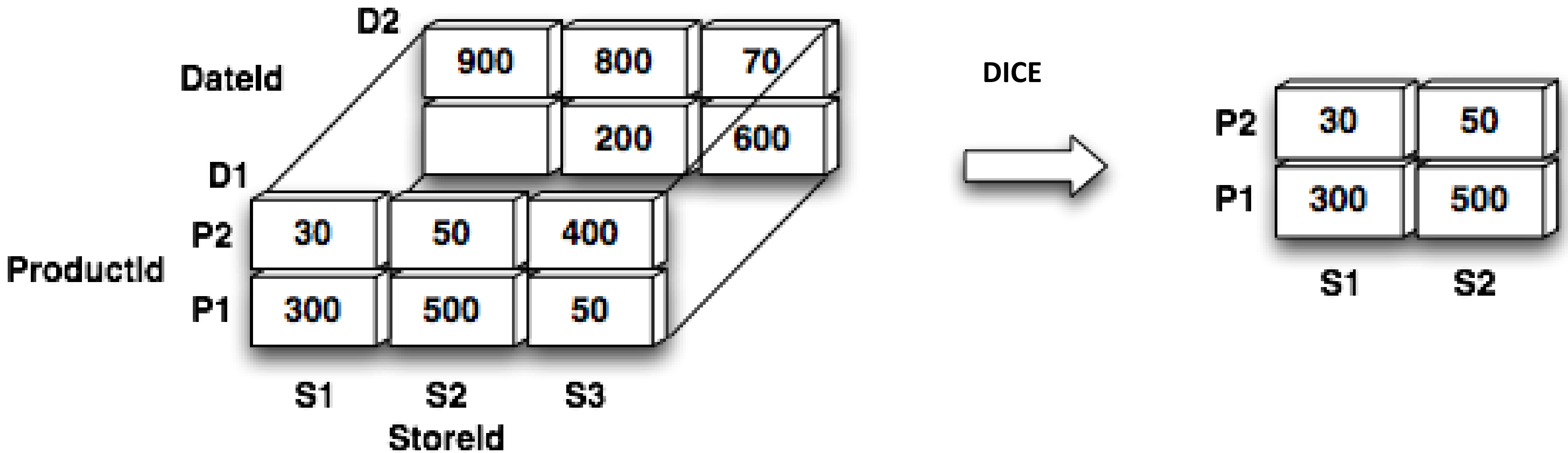
Cube operator: Slice

- Sales **SLICE FOR** DateId = ‘D1’;



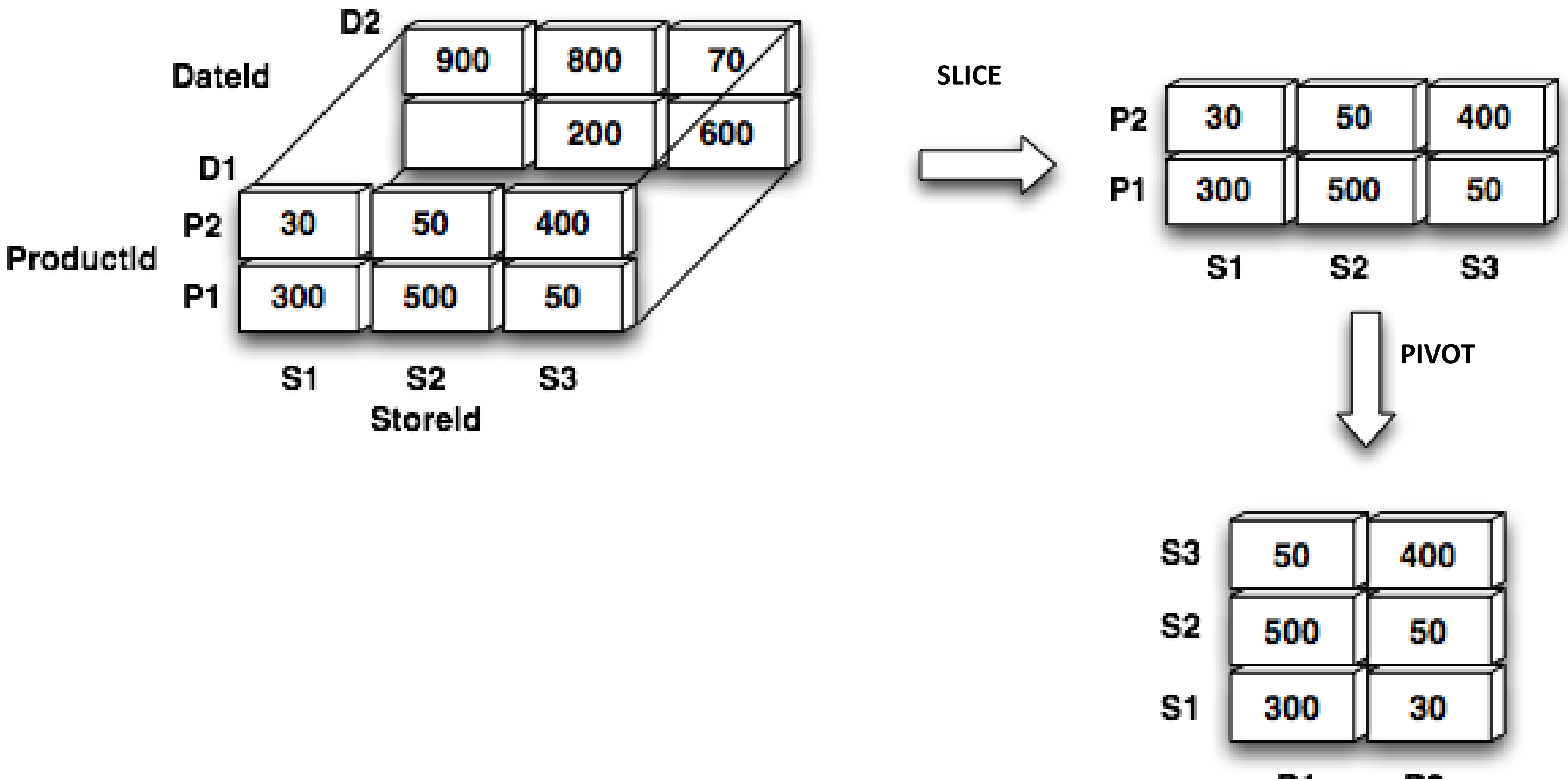
Cube operator: Dice

- Sales DICE FOR DateId = ‘D1’
  StoreId IN (‘S1’, ‘S2’);

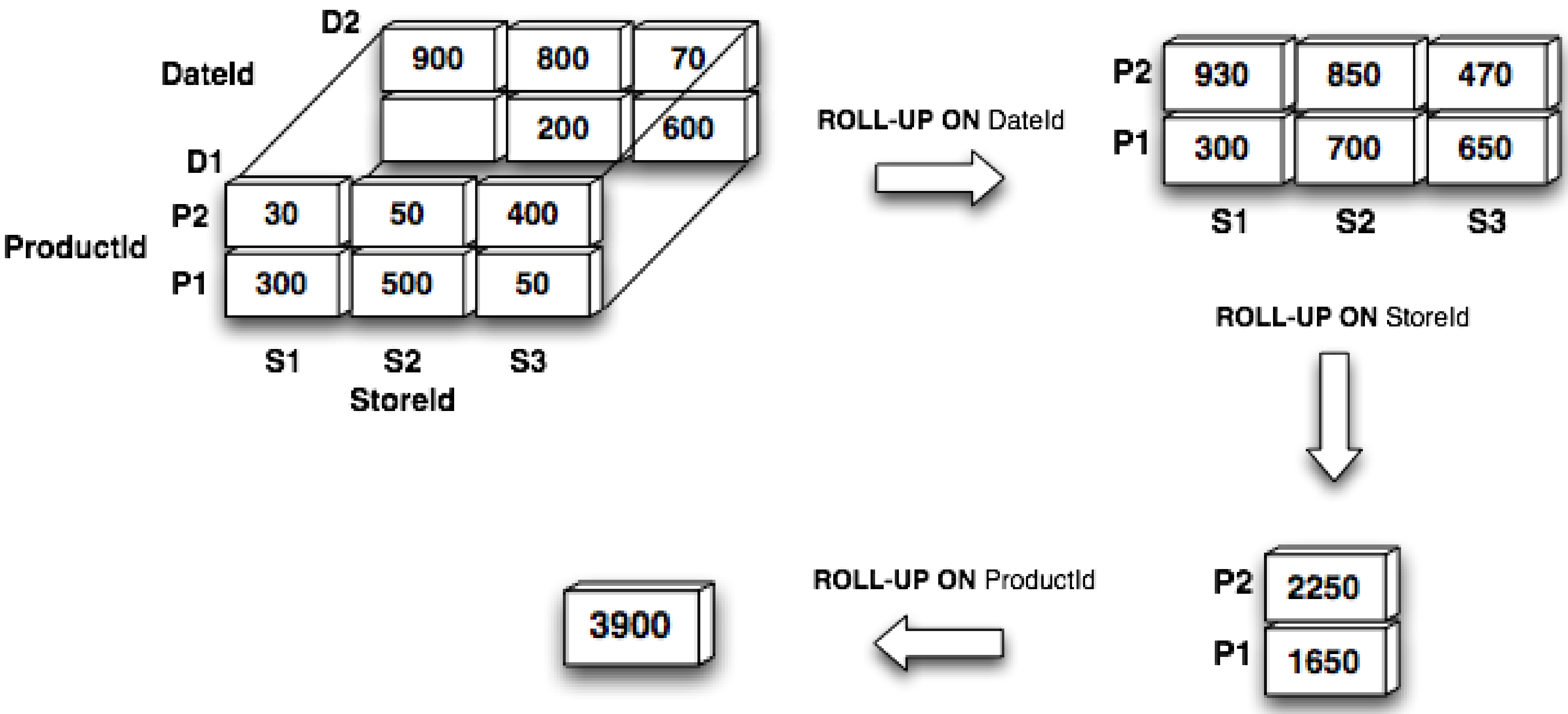


Cube operator: Pivot

- **PIVOT** (Sales **SLICE FOR** DateId = ‘D1’);

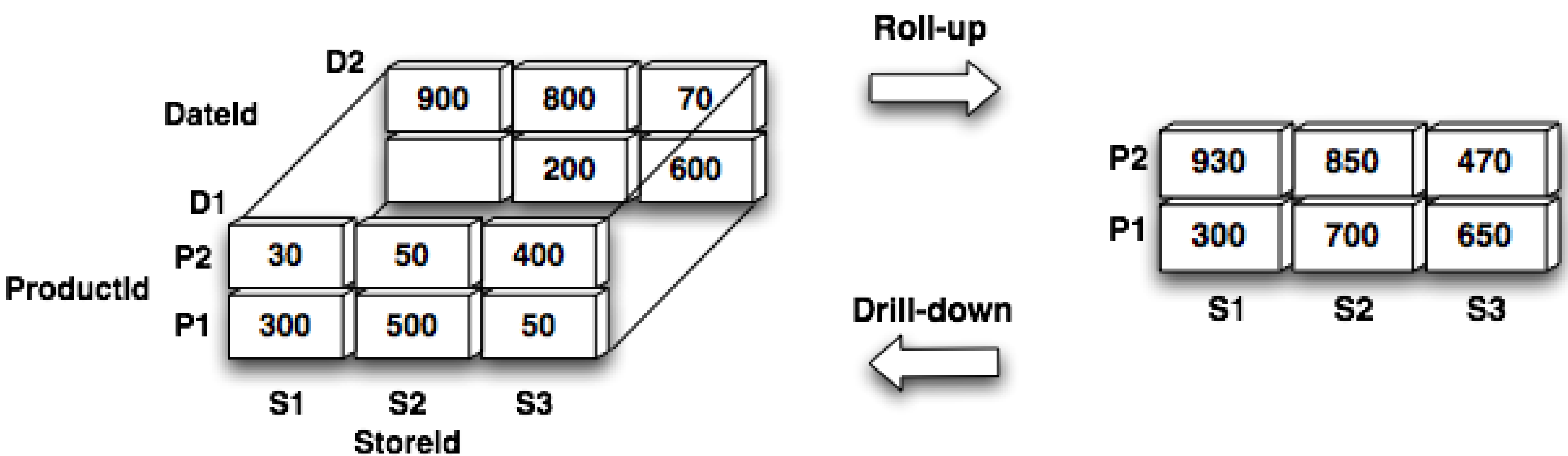


Cube operators: Roll-up



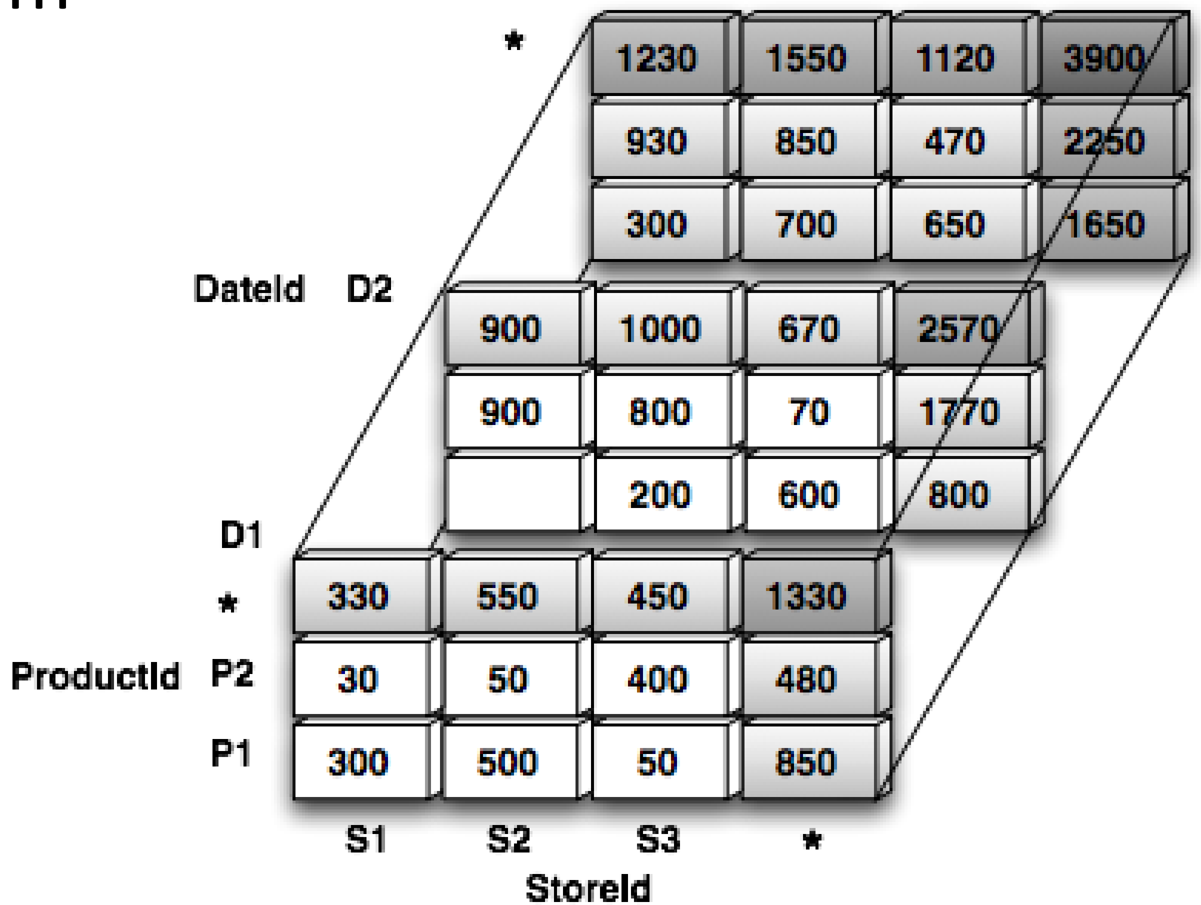
Cube operators: Roll-up and Drill-down

- Roll-up aggregates data by dimension reduction or by navigating dimension hierarchy
- Drill-down: the reverse
- SALES **ROLL-UP** ON DateId



Extended cube

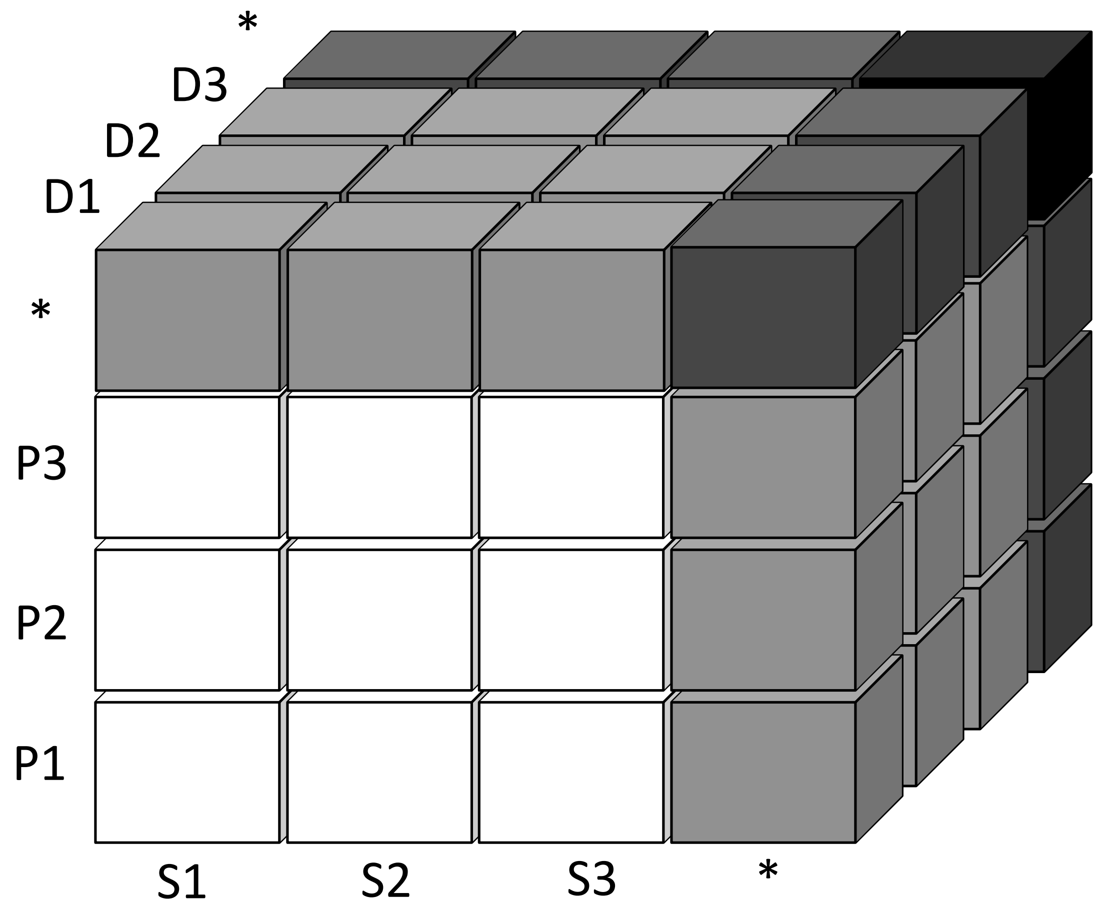
- Extend each dimension with *, mapped to the sum



Data cube

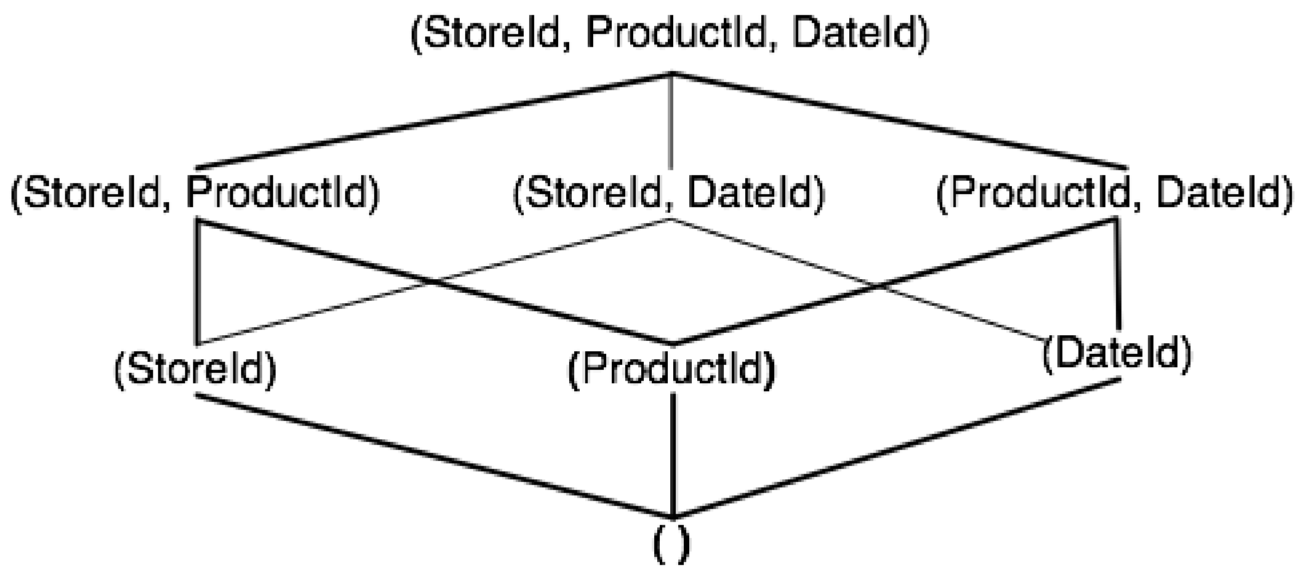
Extended cube

- The cube is now a set of cuboids
- **white** cells are the data cube
- **gray** cells are rolled-up by a dimension
- **dark gray** cells are rolled-up by two dimensions
- **black** cells are rolled-up by all dimensions.



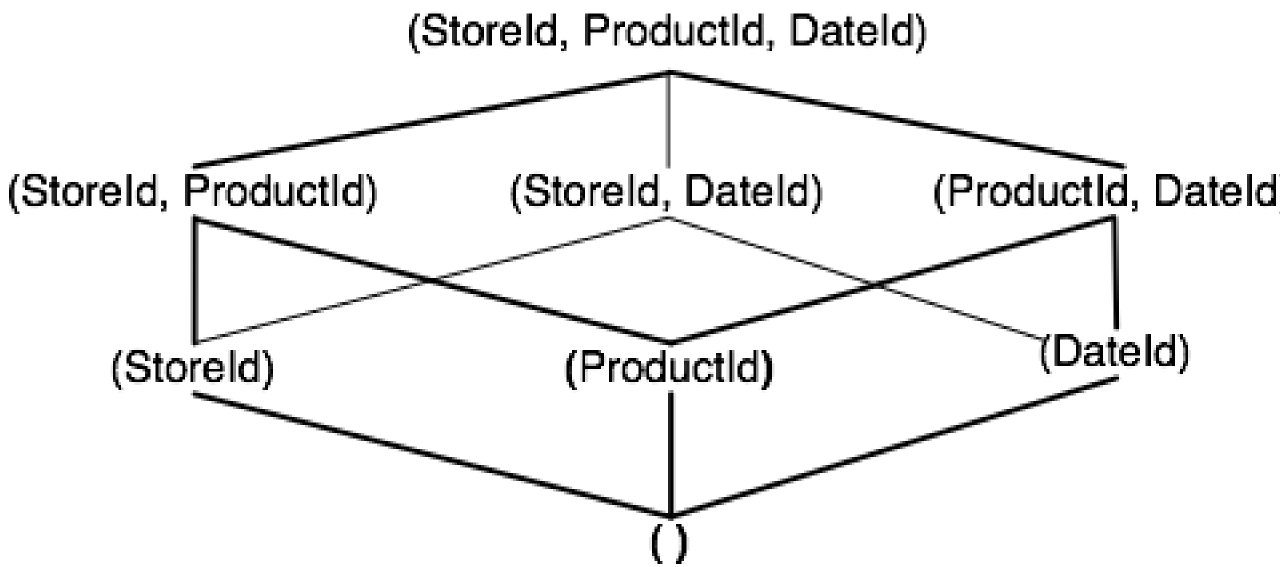
DW Lattice: a lattice of cuboids

- On the set of cuboids is defined the following partial order relation:
  - $C1 \leq C2$ if C1 can be computed by C2



Cuboid materialization

- If the materialization is partial, which cuboids do we store?

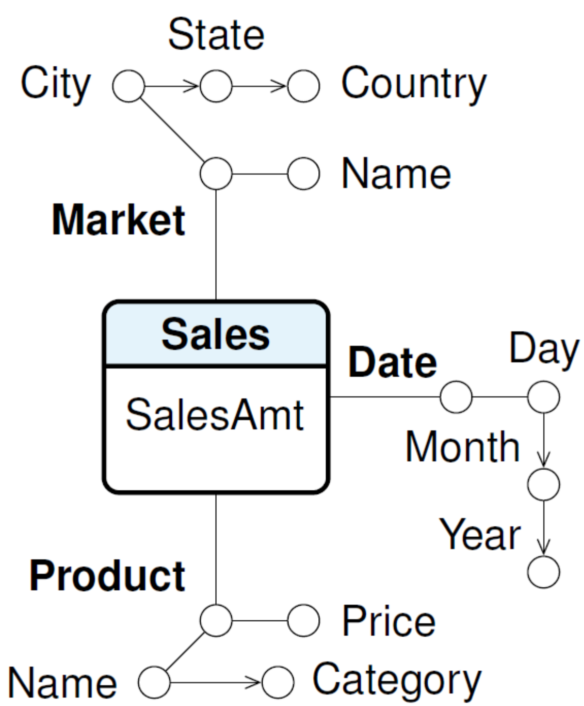


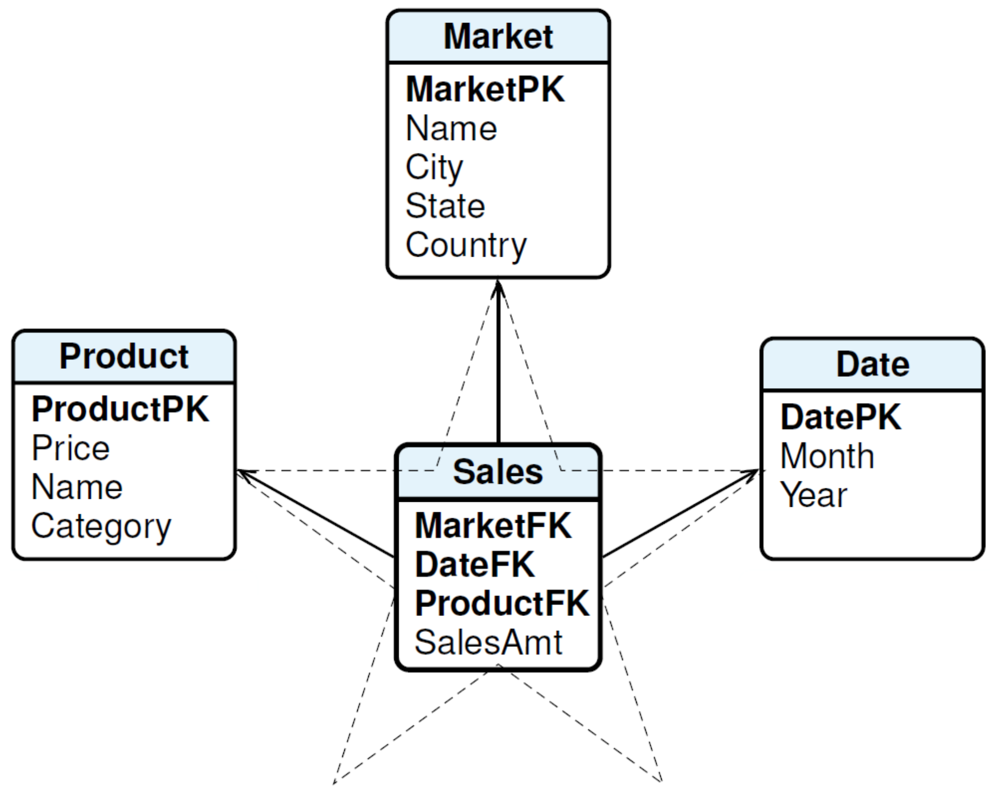
Relational representation

- Relational OLAP systems are relational DBMS extended with specific features to support business intelligence analysis
- A DW is represented with a special kind of relational schema:
  - star schema
  - snowflake schema
  - constellation schema

Star schema

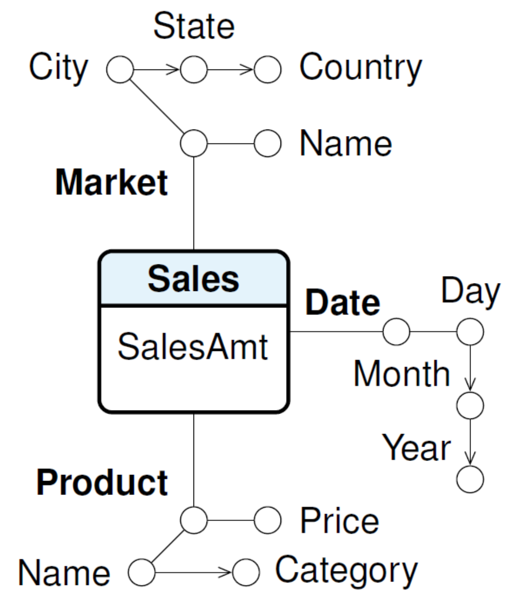
- A fact table with a foreign key for each dimension
- Each dimension has one table with an artificial key

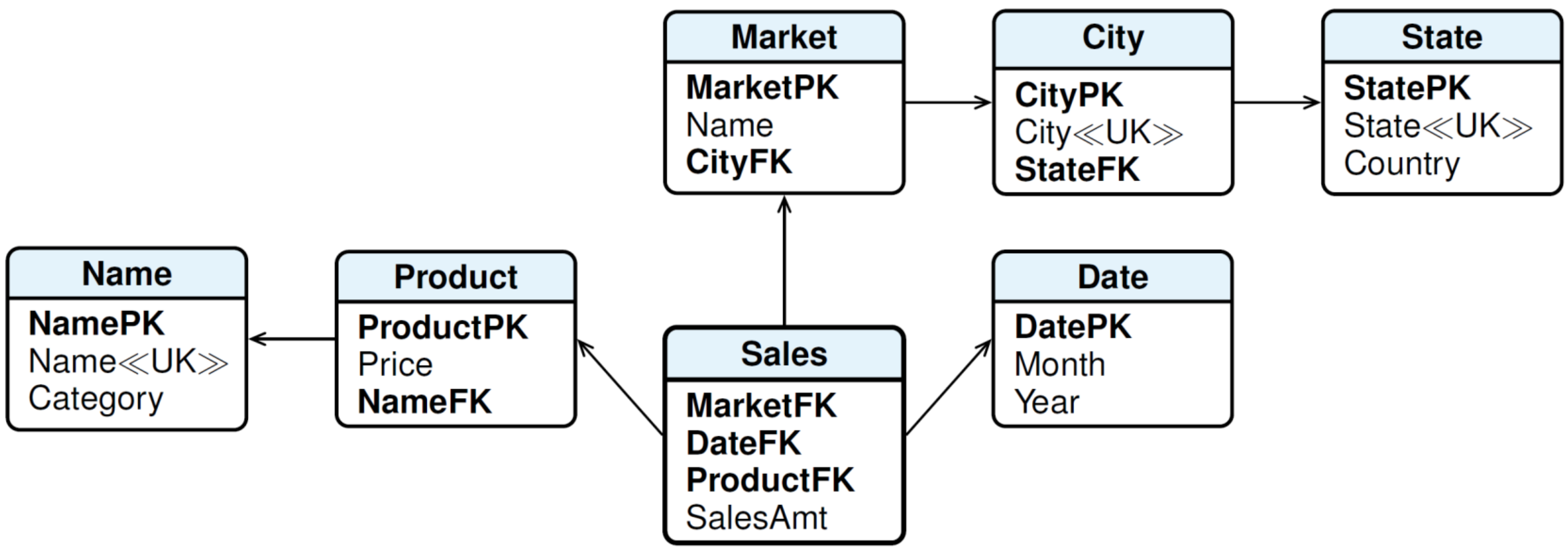




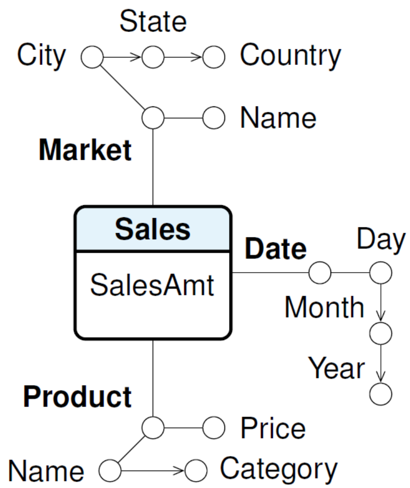
Snowflake schema

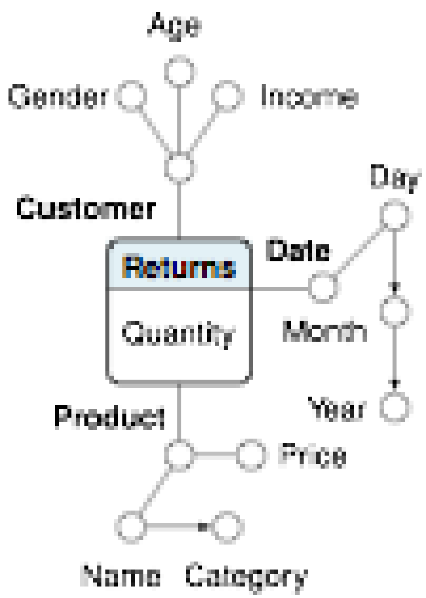
- Dimension tables are 'normalized'
- Not very popular

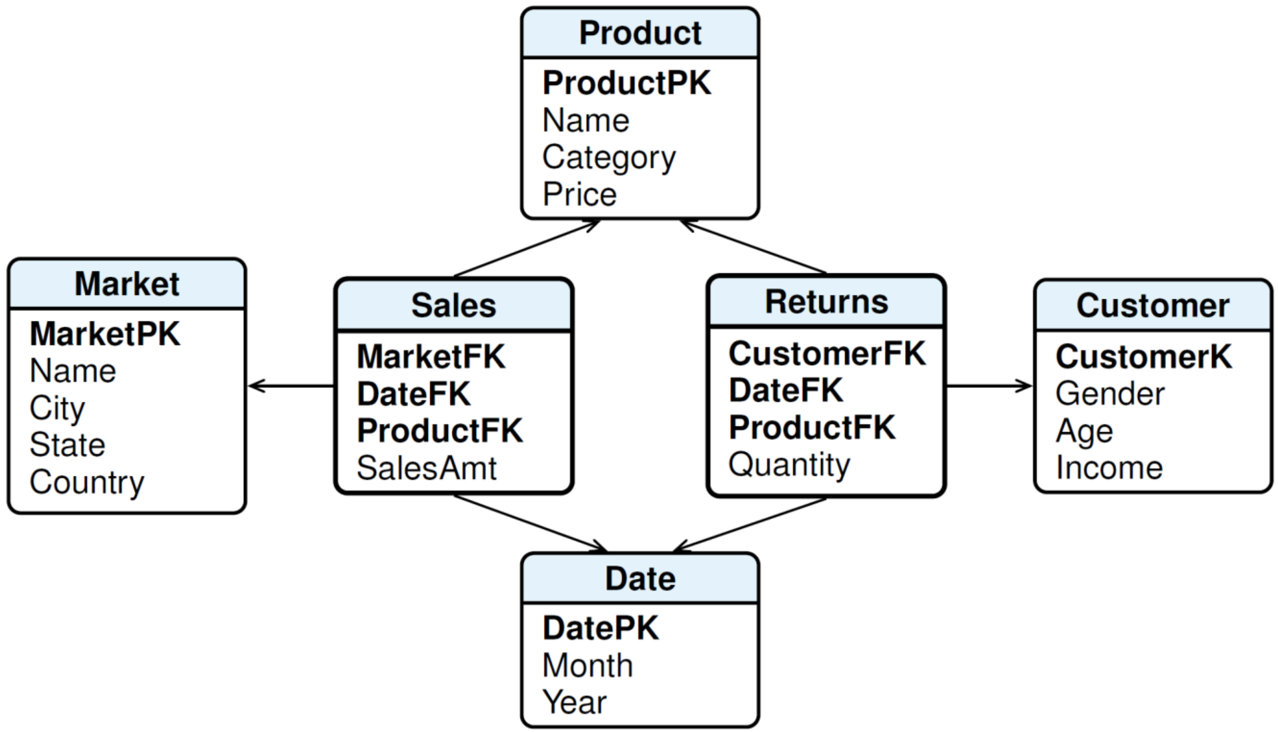




Constellation schema







Data Warehouse Management System

- Interface: what to change?
- Engine: what to change?