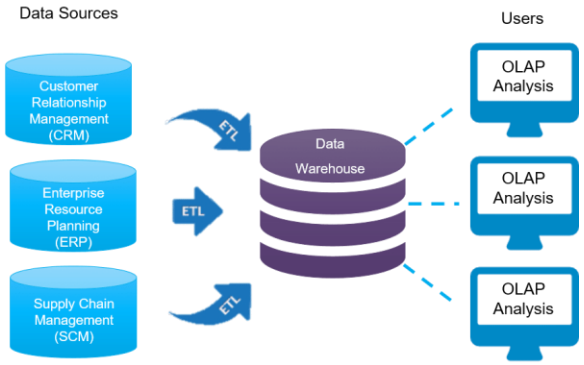


Data Warehouse

Data Warehouse?

menggabungkan data

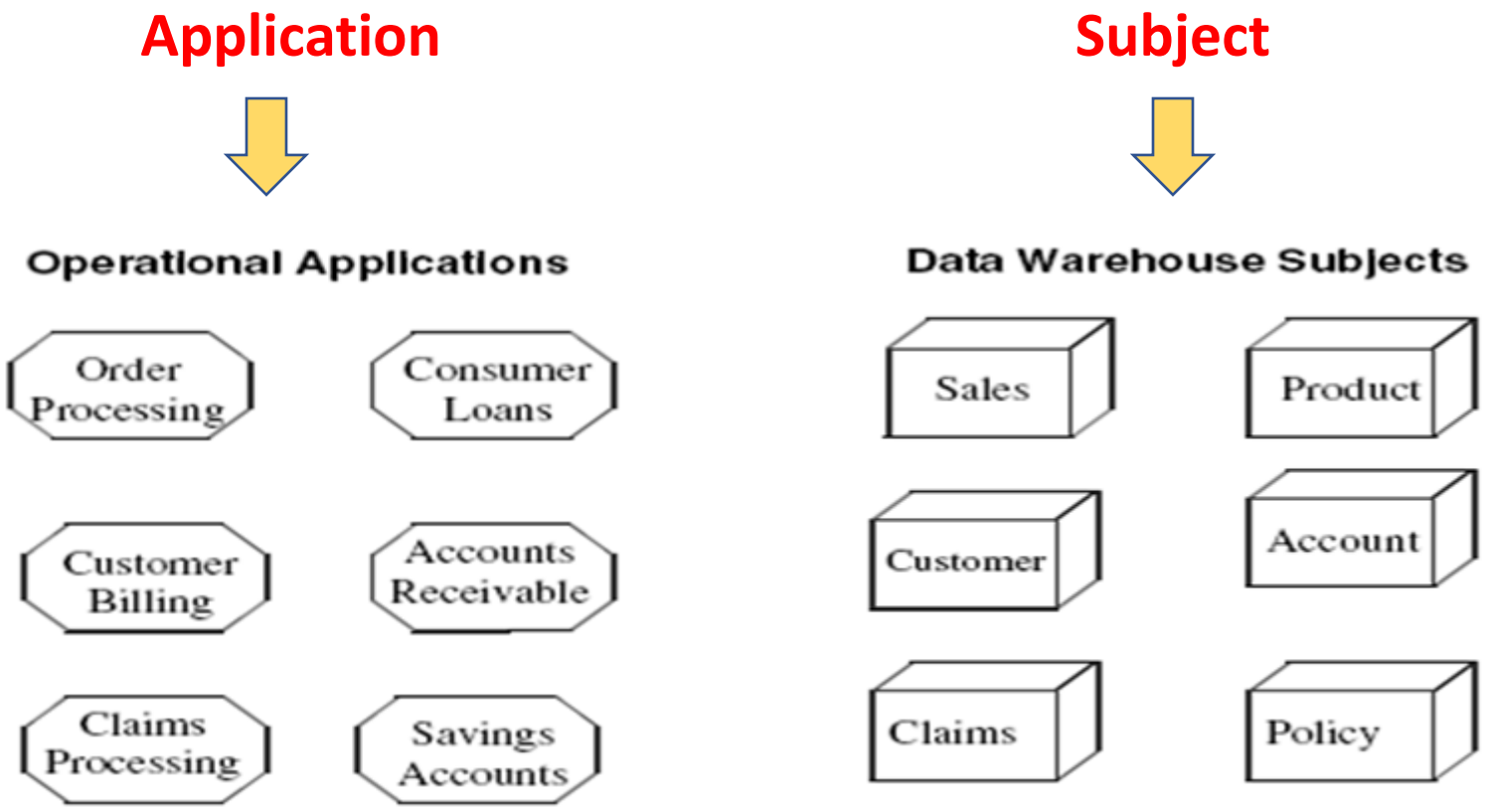
A data warehouse is a system that *retrieves* and *consolidates* data *periodically* from the source systems into a *dimensional* or *normalized data store*.

transformasi mengubah data sesuai dengan formatnya

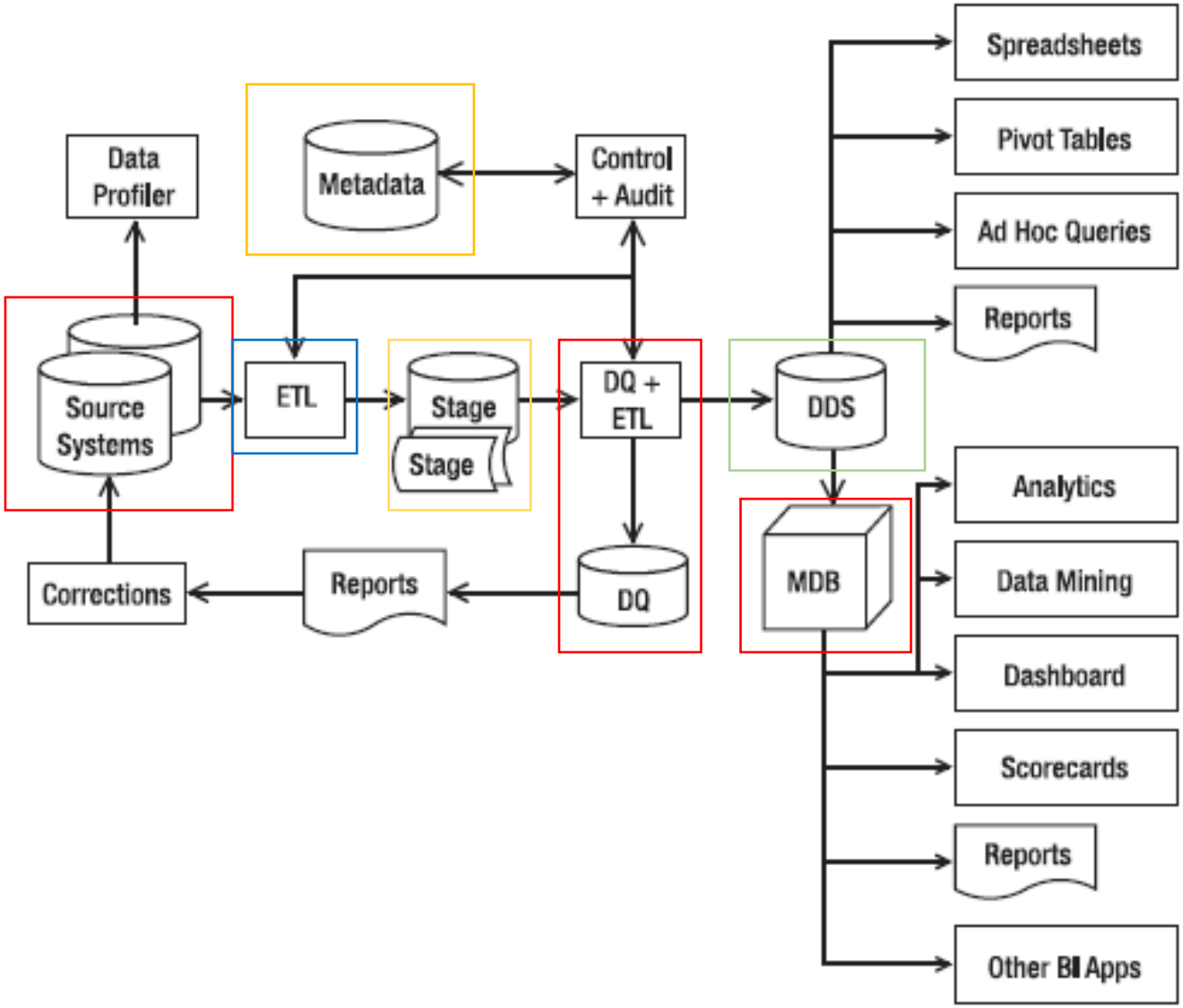
It usually keeps years of *history* and is *queried* for *business intelligence* or other *analytical activities*.

It is typically updated in *batches*, not every time a transaction happens in the source system.

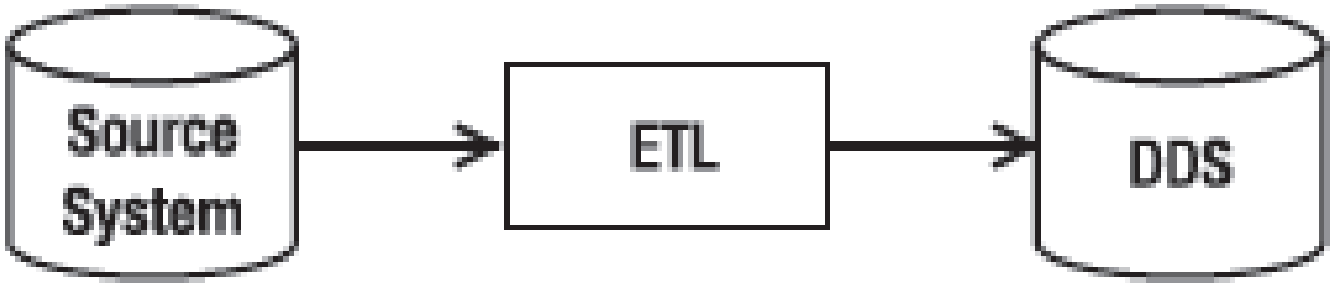
Subject Oriented Data



A diagram of a data warehouse system



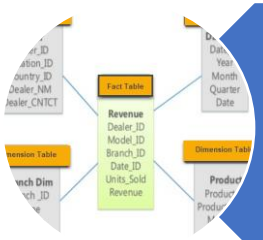
Simplest form of a data warehouse system



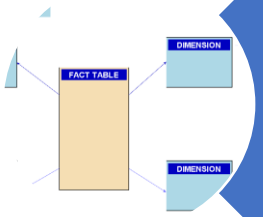
Dimensional Data Store

- A data warehouse is a system that retrieves data from source systems and puts it into a dimensional data store or a normalized data store.
- A DDS is one or several databases containing a collection of dimensional data marts.
- A dimensional data mart is a group of related fact tables and their corresponding dimension tables containing the measurements of business events categorized by their dimensions.

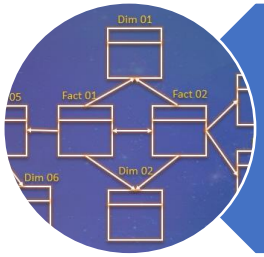
Schema dimensional data store



a star schema

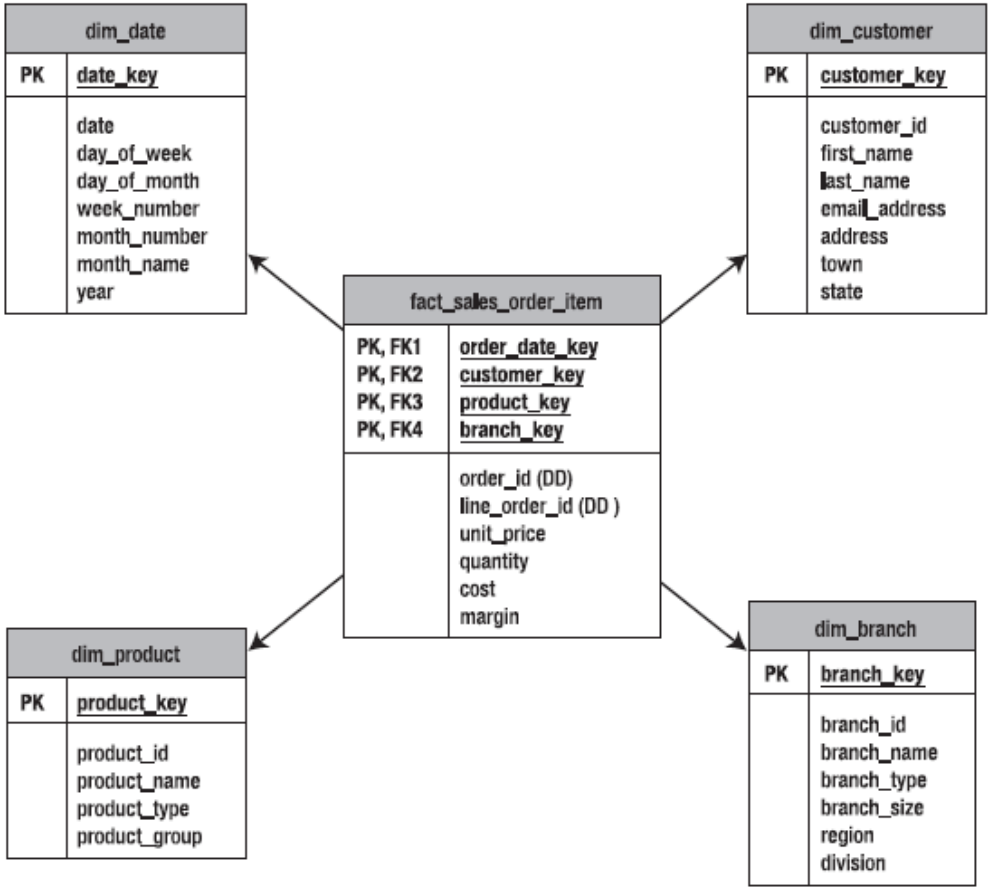


a snowflake schema

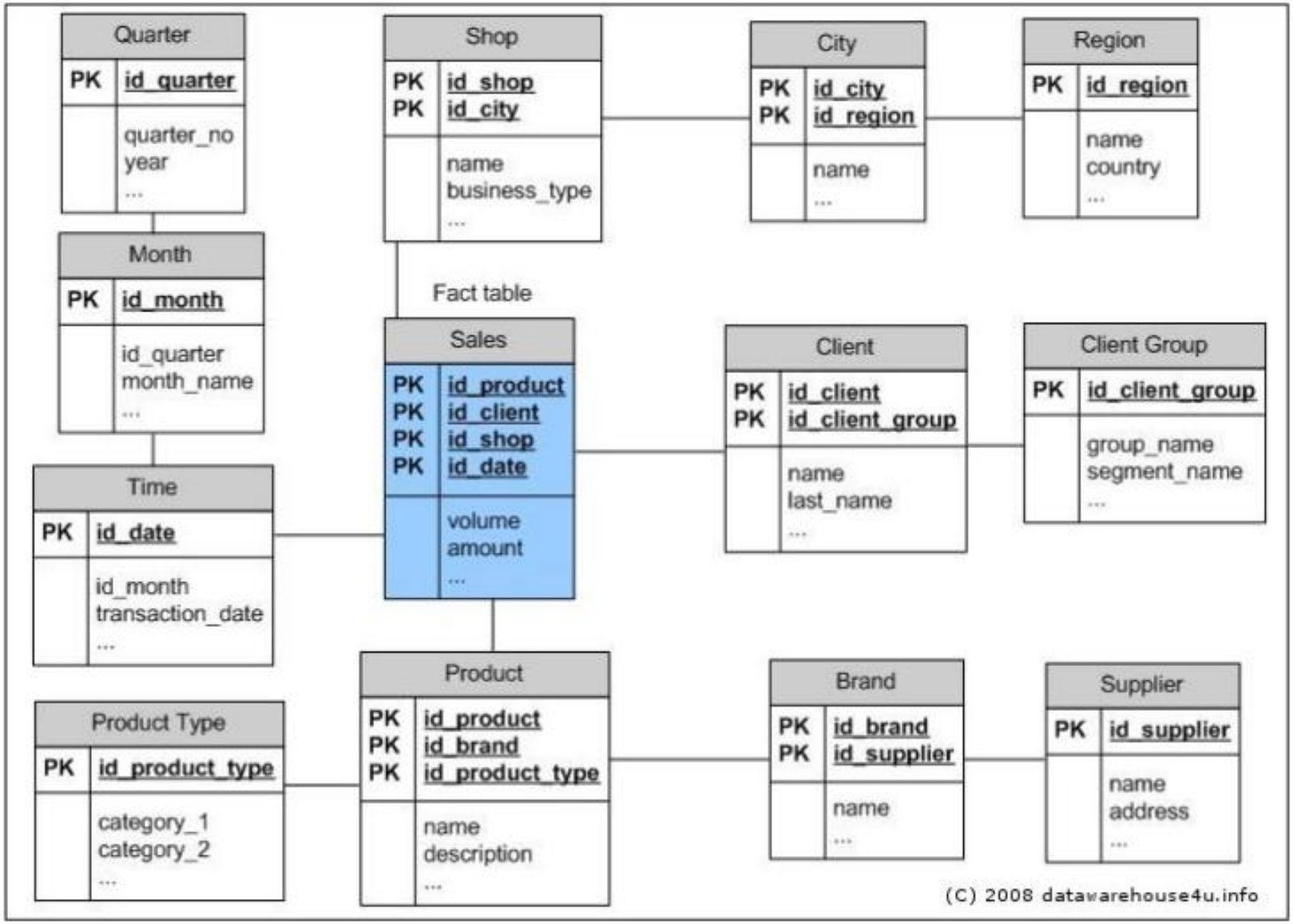


a galaxy schema or a *fact constellation schema :*

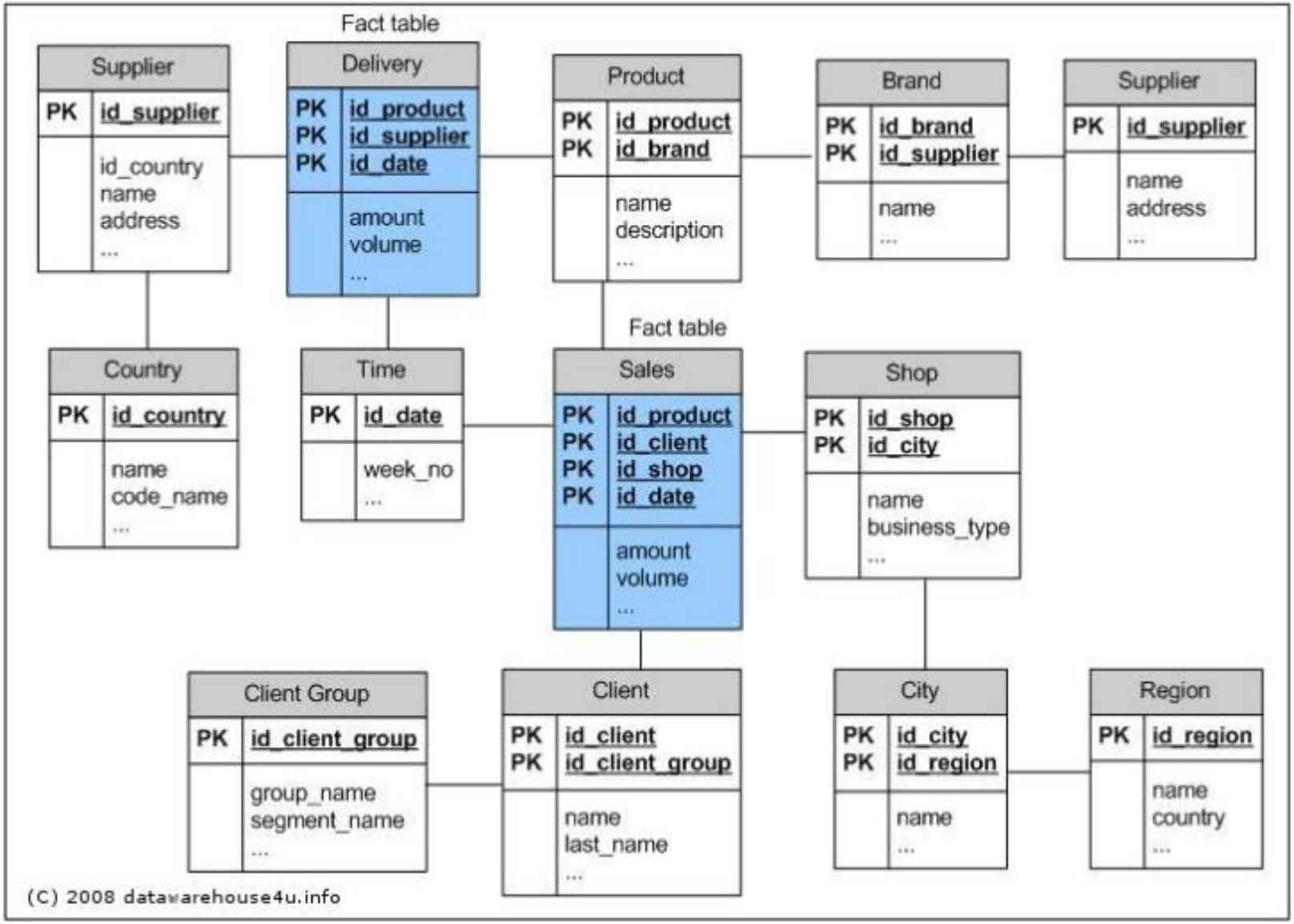
Star Schema dimensional data store



Snowflake schema

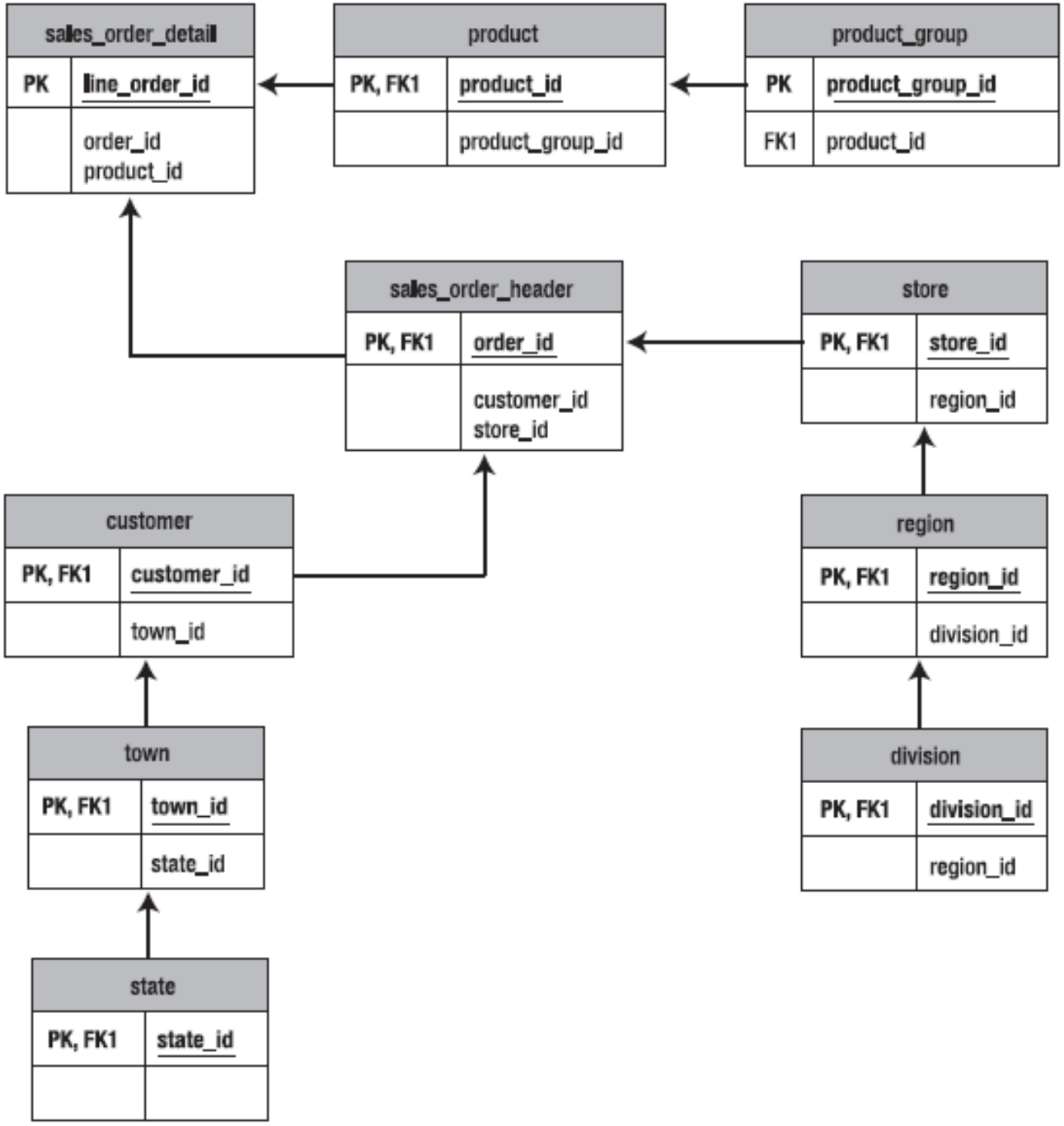


A galaxy schema or a *fact constellation*



Normalized Data Store

- A normalized data store is one or more relational databases with little or no data redundancy.
- A relational database is a database that consists of entity tables with relationships between them.
- Normalization is a process of removing data redundancy by implementing normalization rules.
- There are five degrees of normal forms, from the first normal form to the fifth normal form.
- A normalized data store is usually in third normal form or higher, such as fourth or fifth normal form.

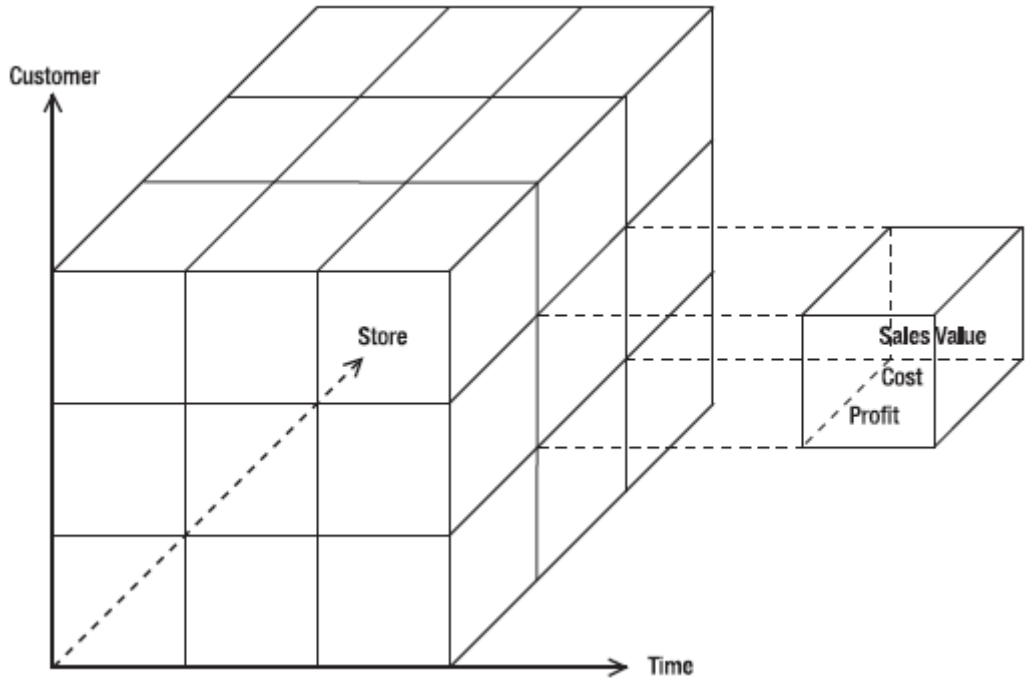


Normalized Data Store

Example of MDBs:

- A cube with three dimensions, or axes: Time, Store, and Customer.
- Assume that each dimension, or axis, has 100 segments, so there are 100 x 100 x 100 = 1 million cells in that cube.
- Each cell represents an event where a customer is buying something from a store at a particular time.
- In each cell there are three numbers:
  - Sales Value (the total value of the products that the customer purchased),
  - Cost (the cost of goods sold + proportioned overheads), and
  - Profit (the difference between the sales value and cost).

A cube with three dimensions



Three categories of business intelegence

- Reporting, such as key performance indicators, global sales figures by business unit and service codes, worldwide customer accounts, consolidated delivery status, and resource utilization rates across different branches in many countries
- OLAP, such as aggregation, drill down, slice and dice, and drill across
- Data mining, such as data characterization, data discrimination, association analysis, classification, clustering, prediction, trend analysis, deviation analysis, and similarity analysis

Examples of business intelegence

- Business performance management,
  - Including producing key performance indicators such as daily sales, resource utilization, and main operational costs for each region, product line, and time period, as well as their aggregates, to enable people to take tactical actions to get operational performance on the desired tracks.
- Customer profitability analysis,
  - to understand which customers are profitable and worth keeping and which are losing money and therefore need to be acted upon.
  - The key to this exercise is allocating the costs as accurately as possible to the smallest unit of business transaction, which is similar to activity-based costing.
- Statistical analysis
  - such as purchase likelihood or basket analysis.
  - Basket analysis is a process of analyzing sales data to determine which products are likely to be purchased or ordered together.
- Predictive analysis
  - such as forecasting the sales, revenue, and cost figures for the purpose of planning for next year's budgets and taking into account other factors such as organic growth, economic situations, and the company's future direction.