Dimensional Modeling

Dimensional Modeling

- **Dimensional modeling** (DM) names a set of techniques and concepts used in data warehouse design.
- Dimensional modeling is one of the methods of data modeling, that help us store the data in such a way that it is relatively easy to retrieve the data from the database.
- Dimensional modeling always uses the concepts of facts (measures), and dimensions (context).

Dimensional Models

- A denormalized relational model
  - Made up of tables with attributes
  - Relationships defined by keys and foreign keys
- Organized for understandability and ease of reporting rather than update
- Queried and maintained by SQL or special purpose management tools.

Benefits of dimensional modeling

Understand ability –

- Compared to the normalized model, the dimensional model is easier to understand and more intuitive.
- In dimensional models, information is grouped into coherent business categories or dimensions, making it easier to read and interpret.
- Simplicity also allows software to navigate databases efficiently.

Benefits of dimensional modeling

Query performance

- Dimensional models are more denormalized and optimized for data querying, while normalized models seek to eliminate data redundancies and are optimized for transaction loading and updating.
- The predictable framework of a dimensional model allows the database to make strong assumptions about the data which may have a positive impact on performance.
- Each dimension is an equivalent entry point into the fact table, and this symmetrical structure allows effective handling of complex queries.
- Query optimization is simple, predictable, and controllable.

Benefits of dimensional modeling

Extensibility

- Dimensional models are scalable and easily accommodate unexpected new data.
- Existing tables can be changed in place either by simply adding new data rows into the table or executing SQL alter table commands.
- No queries or applications that sit on top of the data warehouse need to be reprogrammed to accommodate changes

Entity-Relationship vs. Dimensional Models

Relational Modeling	Dimensional Modeling
Data is stored in RDBMS	Data is stored in RDBMS or Multidimensional databases
Tables are units of storage	Cubes are units of storage
Data is normalized and used for OLTP. Optimized for OLTP processing	Data is de normalized and used in data warehouse and data mart. Optimized for OLAP
Several tables and chains of relationships among them	Few tables and fact tables are connected to dimensional tables
Volatile (several updates) and time variant	Non volatile and time invariant
Detailed level of transactional data	Summary of bulky transactional data (Aggregates and Measures) used in business decisions
SQL is used to manipulate data	MDX is used to manipulate data
Normal Reports	User friendly, interactive, drag and drop multidimensional OLAP Reports

Entity-Relationship vs. Dimensional Models

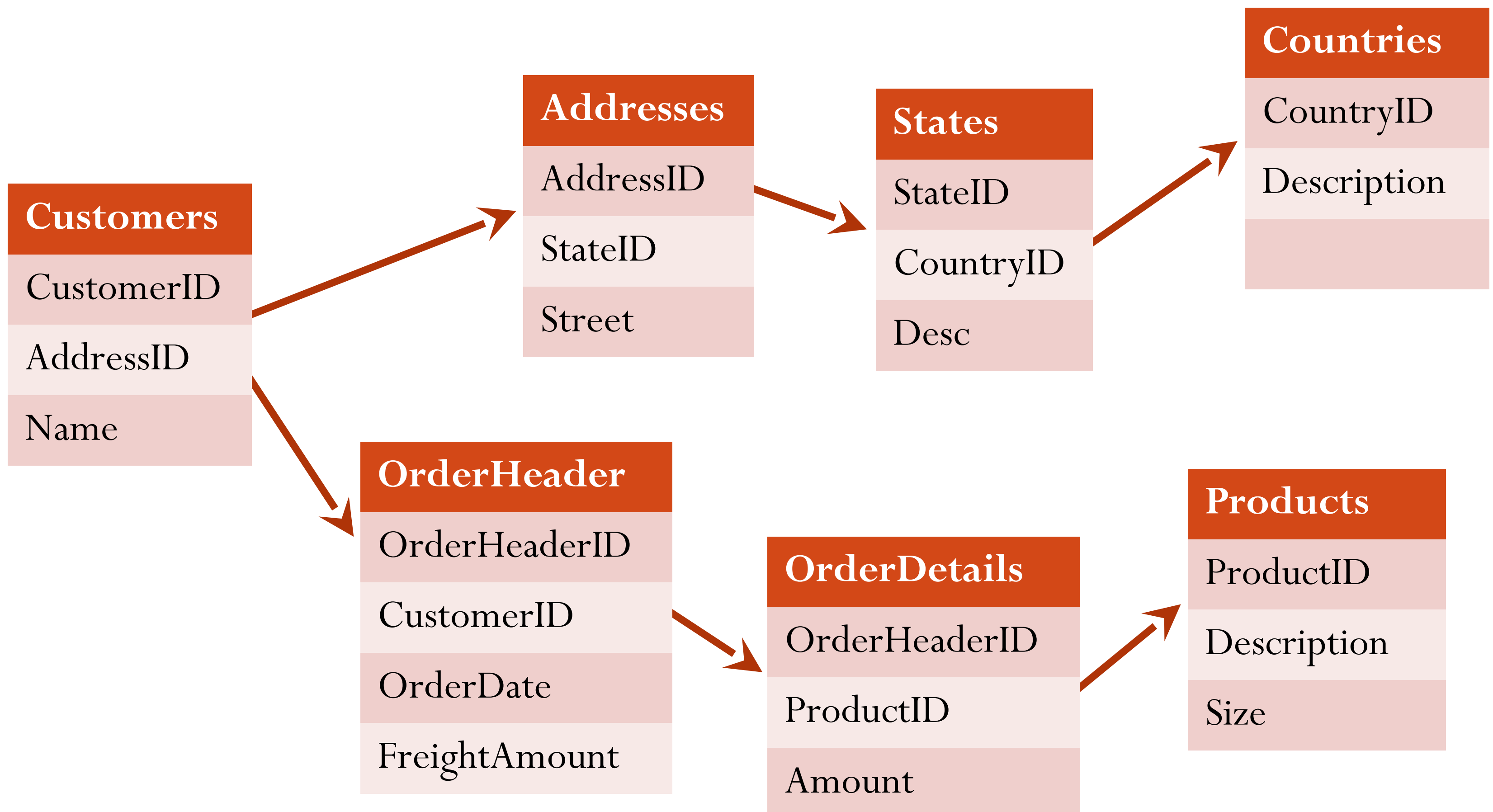
- One table per entity
- Minimize data redundancy
- Optimize update
- The Transaction Processing Model

- One fact table for data organization
- Maximize understandability
- Optimized for retrieval
- The data warehousing model

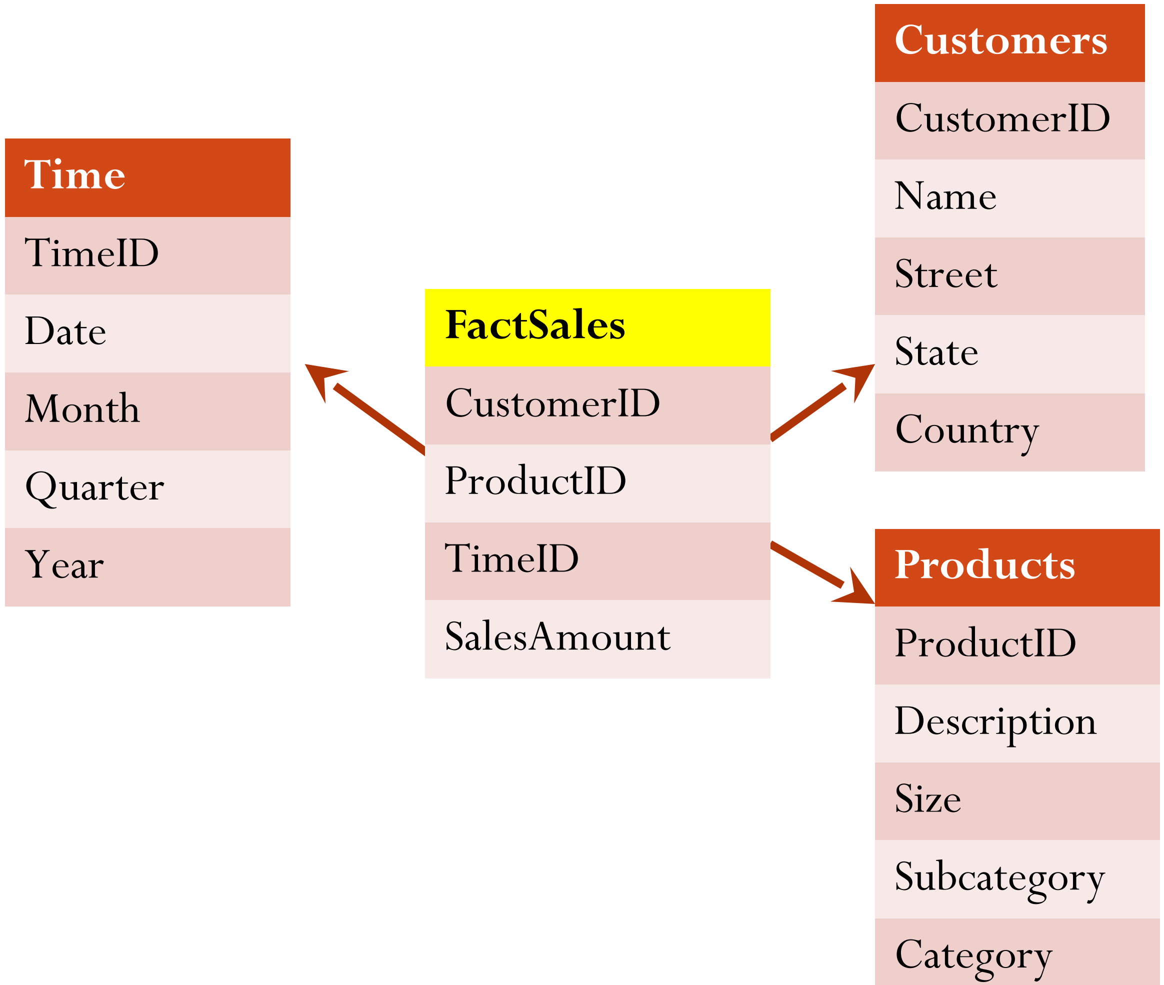
Facts & Dimensions

- There are two main types of objects in a dimensional model
  - **Facts** are quantitative measures that we wish to analyse and report on.
  - **Dimensions** contain textual descriptors of the business. They provide *context* for the facts.

A Transactional Database



A Dimensional Model



Fact Tables

- A fact table stores quantitative information for analysis and is often denormalized.
- Contains two or more foreign keys.
- Tend to have huge numbers of records.
- Useful facts tend to be numeric and additive.
- A fact table holds the data to be analyzed, and a dimension table stores data about the ways in which the data in the fact table can be analyzed. Thus, the fact table consists of two types of columns. The foreign keys column allows joins with dimension tables, and the measures columns contain the data that is being analyzed.

Example :Fact Table

- Suppose that a company sells products to customers. Every sale is a fact that happens, and the fact table is used to record these facts. For example:
-

Time ID	Product ID	Customer ID	Unit Sold
4	17	2	1
8	21	3	2
8	4	1	1
5	20	2	5
3	4	4	7

Dimension Table

- A dimension table stores attributes, or dimensions, that describe the objects in a fact table.
- A data warehouse organizes descriptive attributes as columns in dimension tables.
- For Example: A customer dimension's attributes could include first and last name, birth date, gender, etc., or a website dimension would include site name and URL attributes.
- A dimension table has a primary key column that uniquely identifies each dimension record (row).

Example:Dimension Table

Customer ID	Name	Gender	Income	Education	Region
1	Brian Edge	M	2	3	4
2	Fred Smith	M	3	5	1
3	Sally Jones	F	1	7	3

- The dimension table is associated with a fact table using this PRIMARY key.
- It is not uncommon for a dimension table to have 50 to 100 attributes;
- Dimension tables tend to have fewer rows than fact tables

- Dimension tables are referenced by fact tables using keys.
- When creating a dimension table in a data warehouse, a system-generated key is used to uniquely identify a row in the dimension. This key is also known as a surrogate key.
- The surrogate key is used as the primary key in the dimension table.
- The surrogate key is placed in the fact table and a foreign key is defined between the two tables. When the data is joined, it does so just as any other join within the database.

- describe the “who, what, where, when, how, and why” associated with the event.

Product Dimension
Product Key (PK)
SKU Number (Natural Key)
Product Description
Brand Name
Category Name
Department Name
Package Type
Package Size
Abrasive Indicator
Weight
Weight Unit of Measure
Storage Type
Shelf Life Type
Shelf Width
Shelf Height
Shelf Depth
...

Facts and Dimensions

Criteria	Fact Attributes	Dimension Attributes
Purpose	Measurements for reporting or analysis	Constraints or qualifiers for the measurements
Data type	Additive or semi-additive quantitative data	Textual, descriptive
Size	Larger number of records	Smaller number of records
Reporting use	Main report contents	Row or report headers
Examples	Measurements for sales	About time, people, departments, objects, geographic units

Strengths of the Dimensional Model

- Predictable, standard framework
- Respond well to changes in user reporting needs
- Relatively easy to add data without reloading tables
- Standard design approaches have been developed
- There exist a number of products supporting the dimensional model

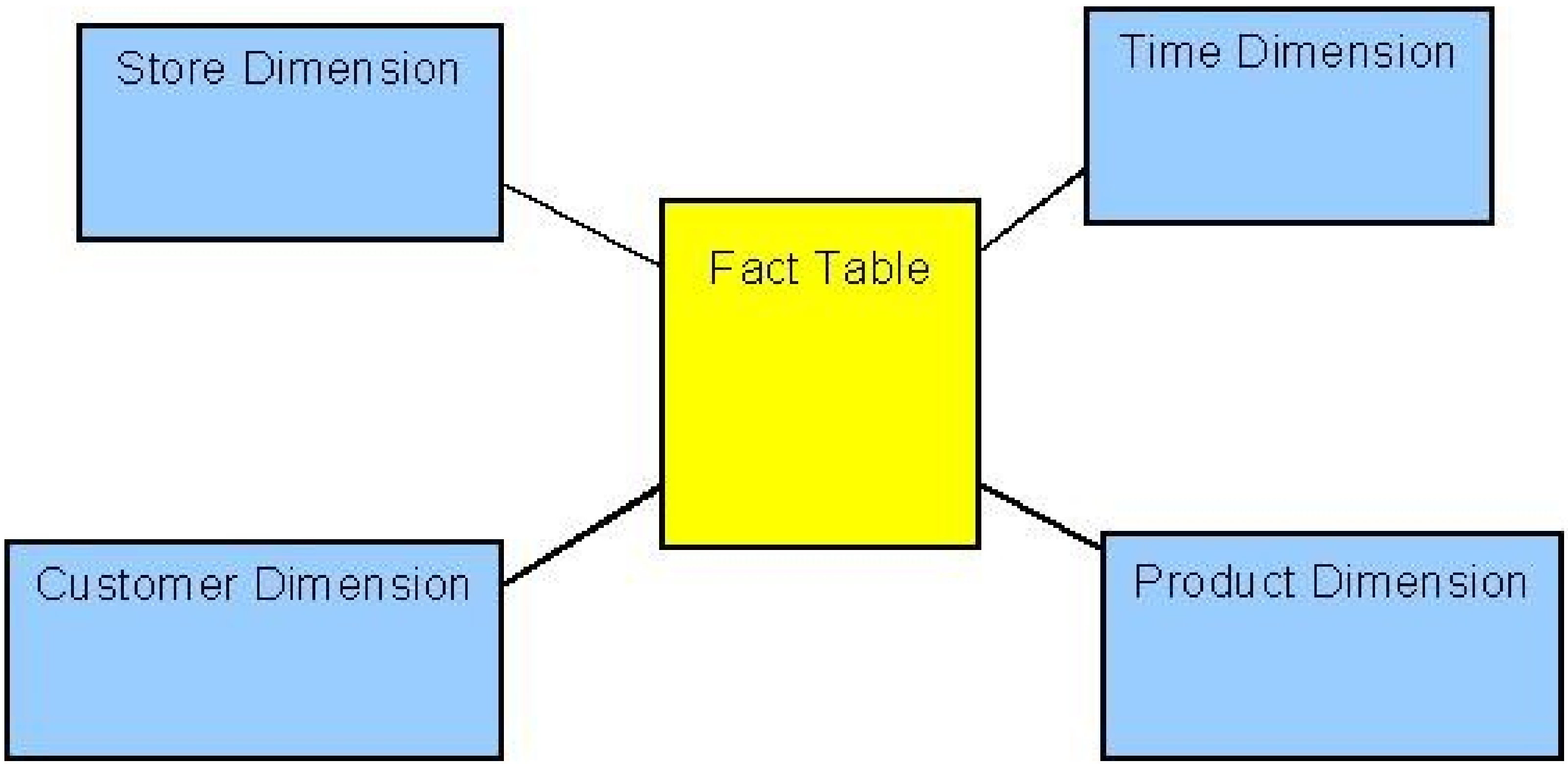
Conceptual Modeling of Data Warehouses

- Star schema
- Snowflake schema
- Fact constellations

Star schema

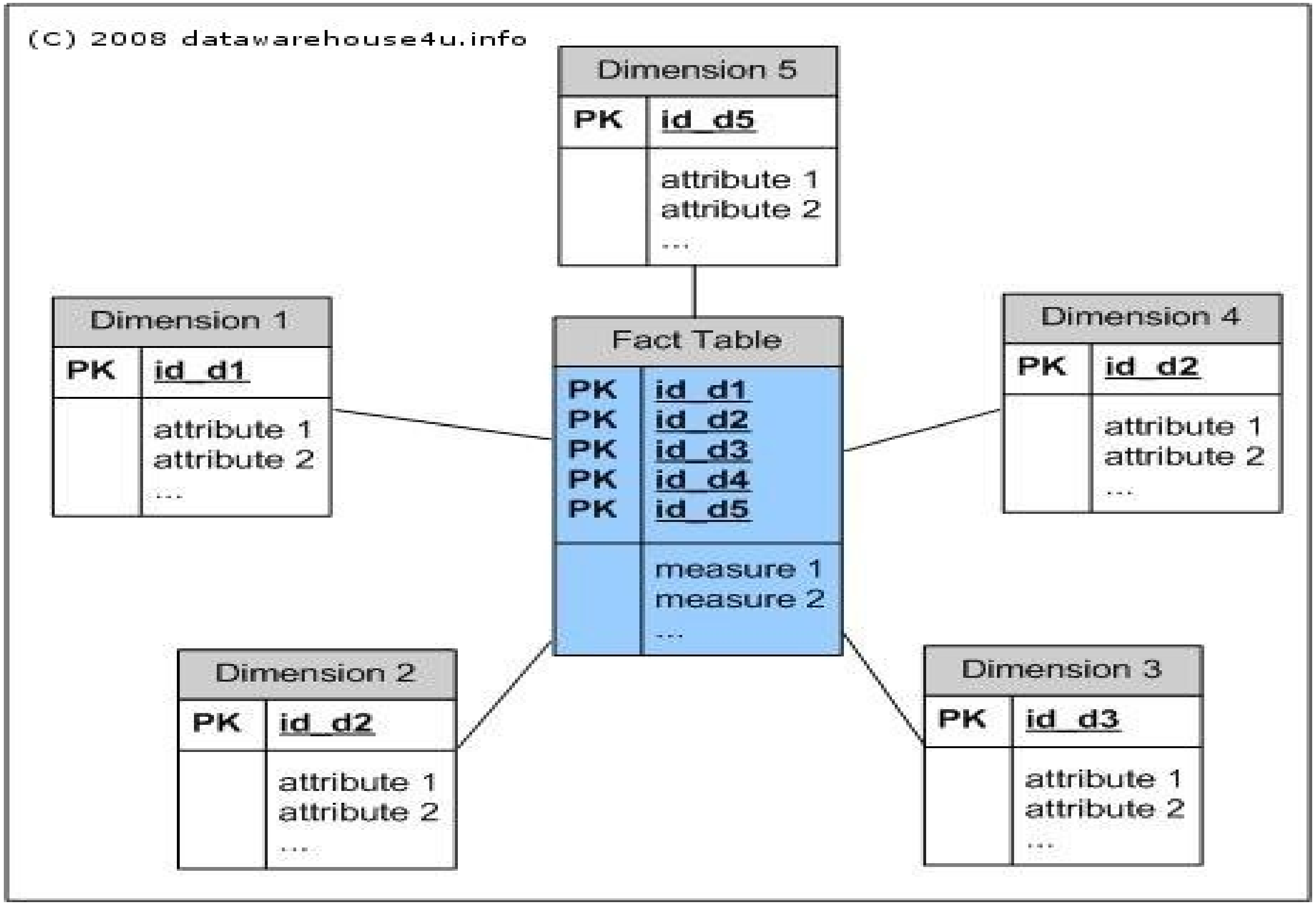
- The star schema architecture is the simplest data warehouse schema.
- It is called a star schema because the diagram resembles a star, with points radiating from a center.
- The center of the star consists of fact table and the points of the star are the dimension tables.
- Usually the fact tables in a star schema are in third normal form(3NF) whereas dimensional tables are de-normalized.

Star schema



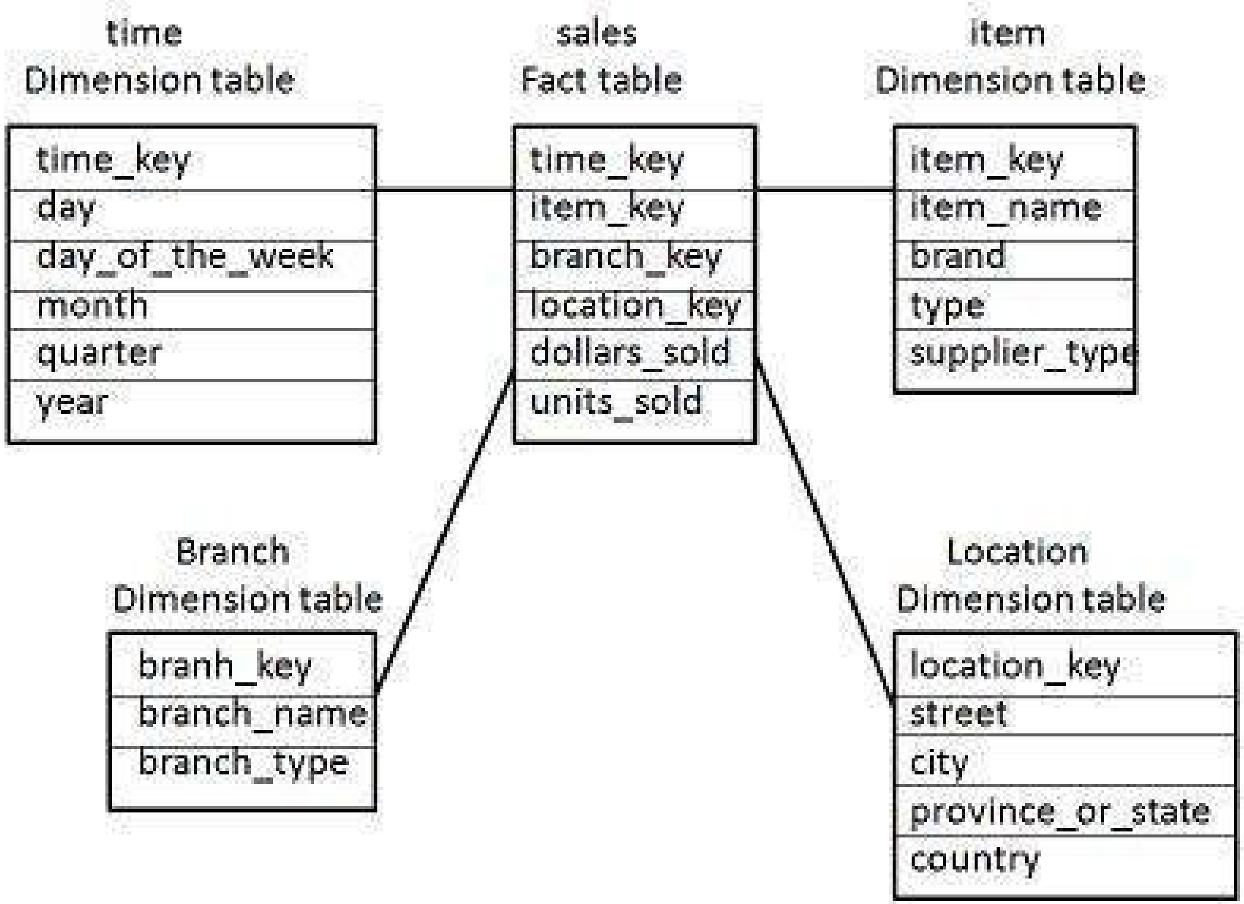
http://www.1keydata.com/datawarehousing/star-schema.html

Star schema



(C) 2008 datawarehouse4u.info
Dimension 5
PK
id_d5
attribute 1
attribute 2
...
Dimension 1
PK
id_d1
attribute 1
attribute 2
...
Fact Table
PK
PK
PK
PK
PK
id_d1
id_d2
id_d3
id_d4
id_d5
measure 1
measure 2
...
Dimension 4
PK
id_d2
attribute 1
attribute 2
...
Dimension 2
PK
id_d2
attribute 1
attribute 2
...
Dimension 3
PK
id_d3
attribute 1
attribute 2
...

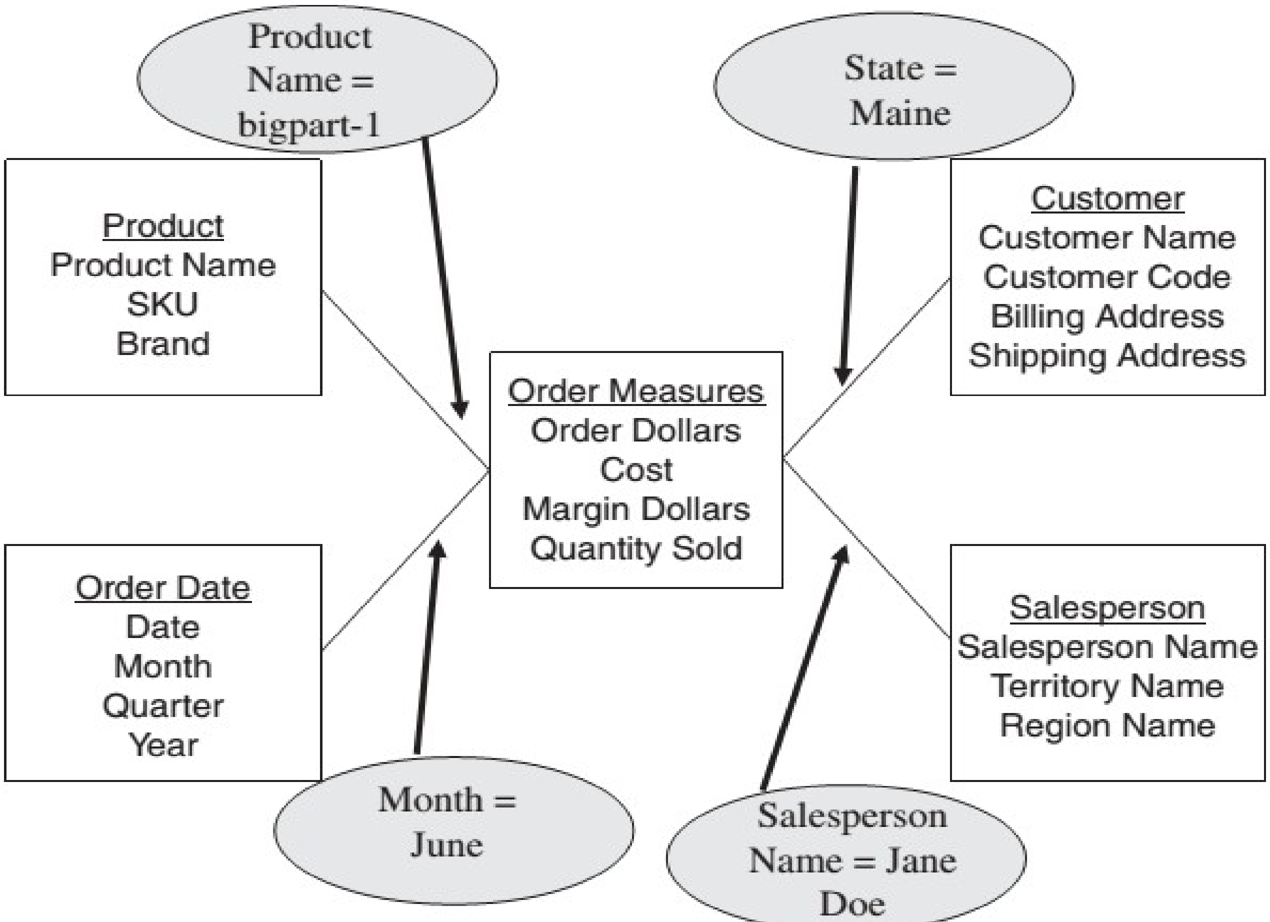
Star schema



time
Dimension table
time_key
day
day_of_the_week
month
quarter
year
sales
Fact table
time_key
item_key
branch_key
location_key
dollars_sold
units_sold
item
Dimension table
item_key
item_name
brand
type
supplier_type
Branch
Dimension table
branh_key
branch_name
branch_type
Location
Dimension table
location_key
street
city
province_or_state
country

Querying Star Schema

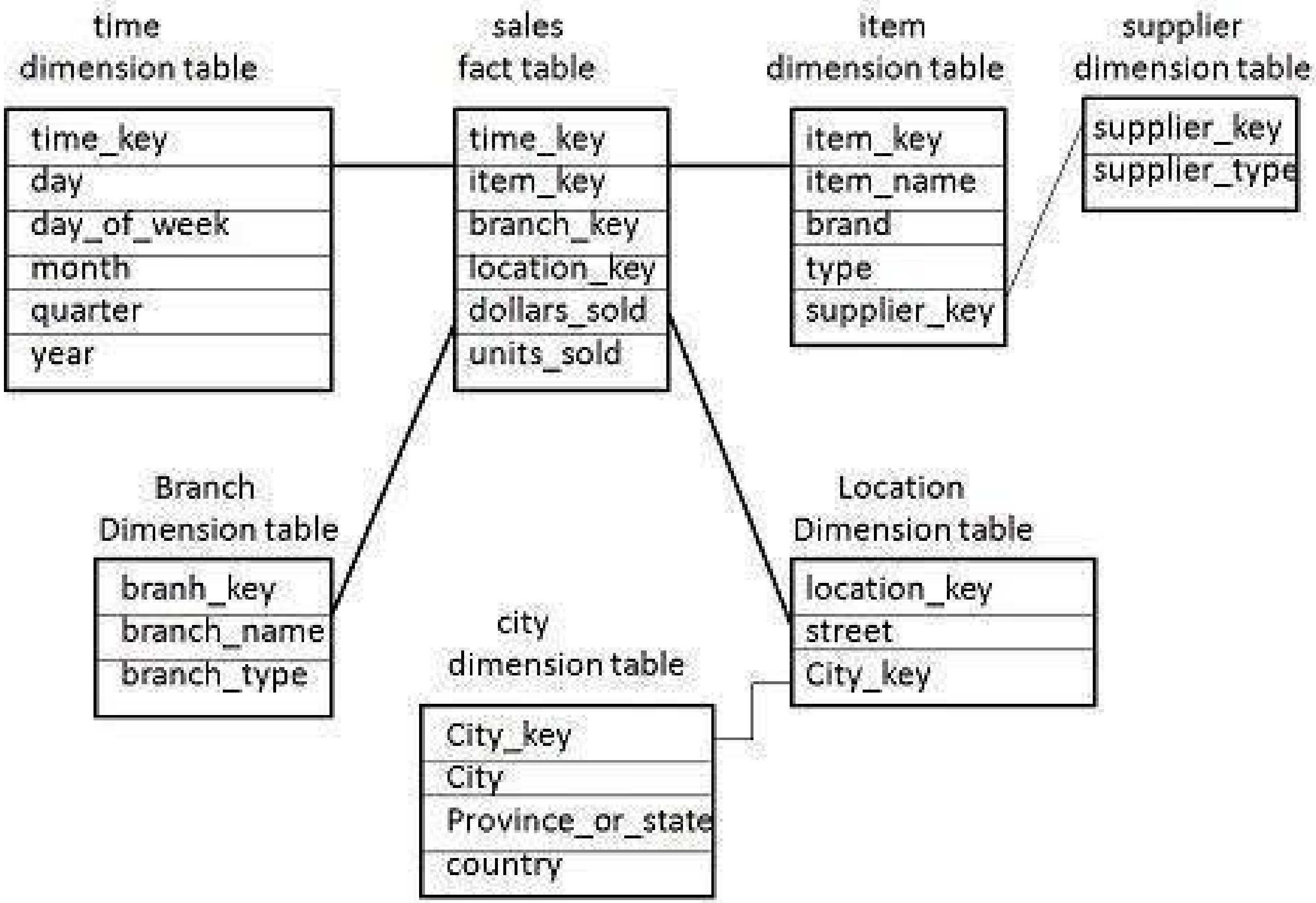
If marketing department wants the quantity sold and order dollars for productbigpart-1, relating to customers in the state ofMaine, obtained by salesperson Jane Doe, during the month of June.



Snowflake Schema

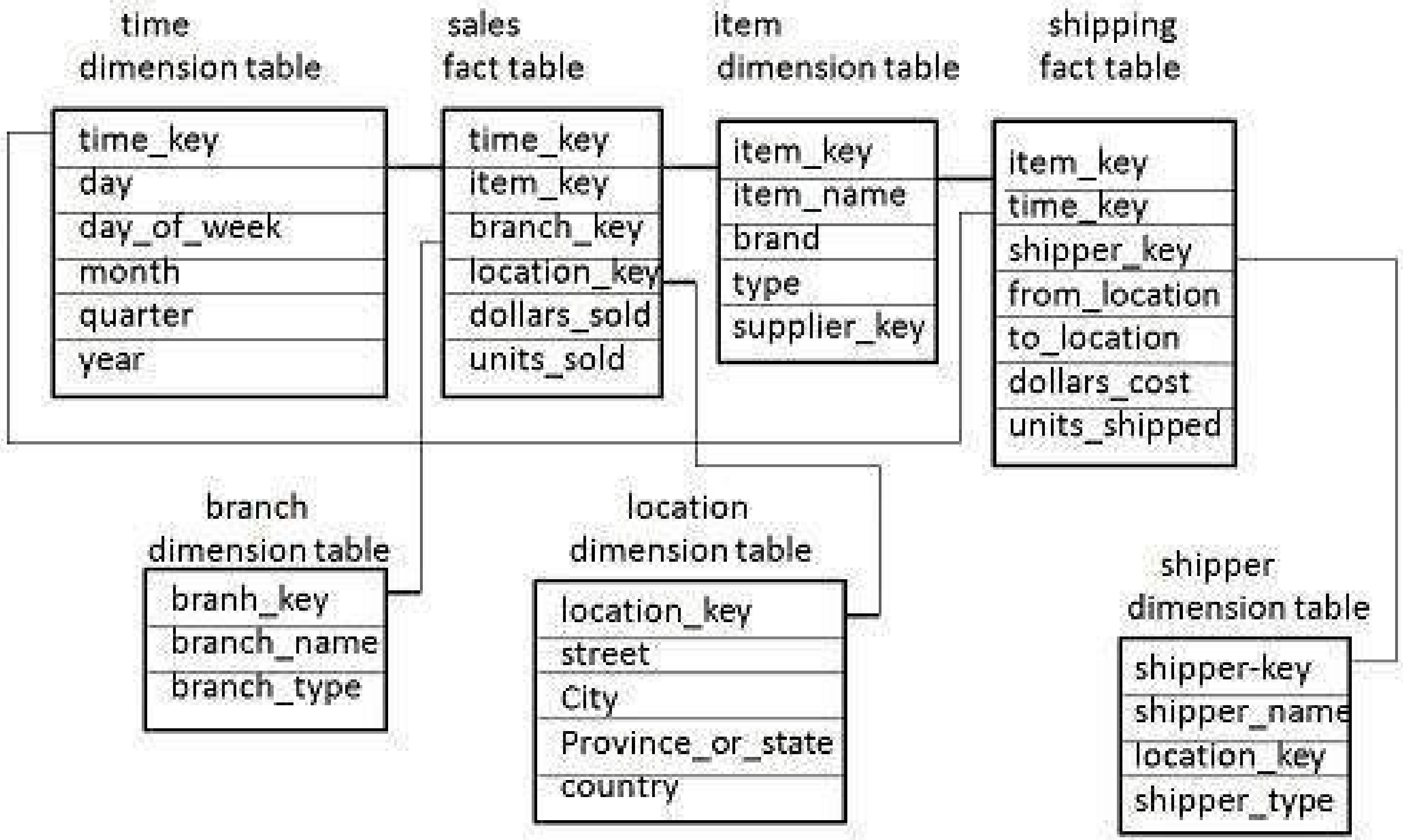
- Some dimension tables in the Snowflake schema are normalized.
- The normalization splits up the data into additional tables.
- Unlike Star schema, the dimensions table in a snowflake schema are normalized. For example, the item dimension table in star schema is normalized and split into two dimension tables, namely item and supplier table.

Snowflake Schema



Fact Constellation Schema

- A fact constellation has multiple fact tables. It is also known as galaxy schema.
- The following diagram shows two fact tables, namely sales and shipping.
- It is also possible to share dimension tables between fact tables. For example, time, item, and location dimension tables are shared between the sales and shipping fact table.


time
dimension table
time_key
day
day_of_week
month
quarter
year
sales
fact table
time_key
item_key
branch_key
location_key
dollars_sold
units_sold
item
dimension table
item_key
item_name
brand
type
supplier_key
shipping
fact table
item_key
time_key
shipper_key
from_location
to_location
dollars_cost
units_shipped
branch
dimension table
branh_key
branch_name
branch_type
location
dimension table
location_key
street
City
Province_or_state
country
shipper
dimension table
shipper-key
shipper_name
location_key
shipper_type

Business Model

As always in life, there are some disadvantages to 3NF:

- Performance can be truly awful. Most of the work that is performed on denormalizing a data model is an attempt to reach performance objectives.
- The structure can be overwhelmingly complex. We may wind up creating many small relations which the user might think of as a single relation or group of data.

The 4 Step Design Process

- Choose the Data Mart
- Declare the Grain
- Choose the Dimensions
- Choose the Facts

Building a Data Warehouse from a Normalized Database

The steps

- Develop a normalized entity-relationship business model of the data warehouse.
- Translate this into a dimensional model. This step reflects the information and analytical characteristics of the data warehouse.
- Translate this into the physical model. This reflects the changes necessary to reach the stated performance objectives.

Structural Dimensions

- The first step is the development of the structural dimensions. This step corresponds very closely to what we normally do in a relational database.
- The star architecture that we will develop here depends upon taking the central intersection entities as the fact tables and building the foreign key => primary key relations as dimensions.

Steps in dimensional modeling

- Select an associative entity for a fact table
- Determine granularity
- Replace operational keys with surrogate keys
- Promote the keys from all hierarchies to the fact table
- Add date dimension
- Split all compound attributes
- Add necessary categorical dimensions
- Fact (varies with time) / Attribute (constant)

Converting an E-R Diagram

- Determine the purpose of the mart
- Identify an association table as the central fact table
- Determine facts to be included
- Replace all keys with surrogate keys
- Promote foreign keys in related tables to the fact table
- Add time dimension
- Refine the dimension tables

Choosing the Mart

- A set of related fact and dimension tables
- Single source or multiple source
- Conformed dimensions
- Typically have a fact table for each process

Fact Tables

Represent a process or reporting environment that is of value to the organization

- It is important to determine the identity of the fact table and specify exactly what it represents.
- Typically correspond to an associative entity in the E-R model

Grain (unit of analysis)

The grain determines what each fact record represents: the level of detail.

- For example
  - Individual transactions
  - Snapshots (points in time)
  - Line items on a document
- Generally better to focus on the smallest grain

Facts

Measurements associated with fact table records at fact table granularity

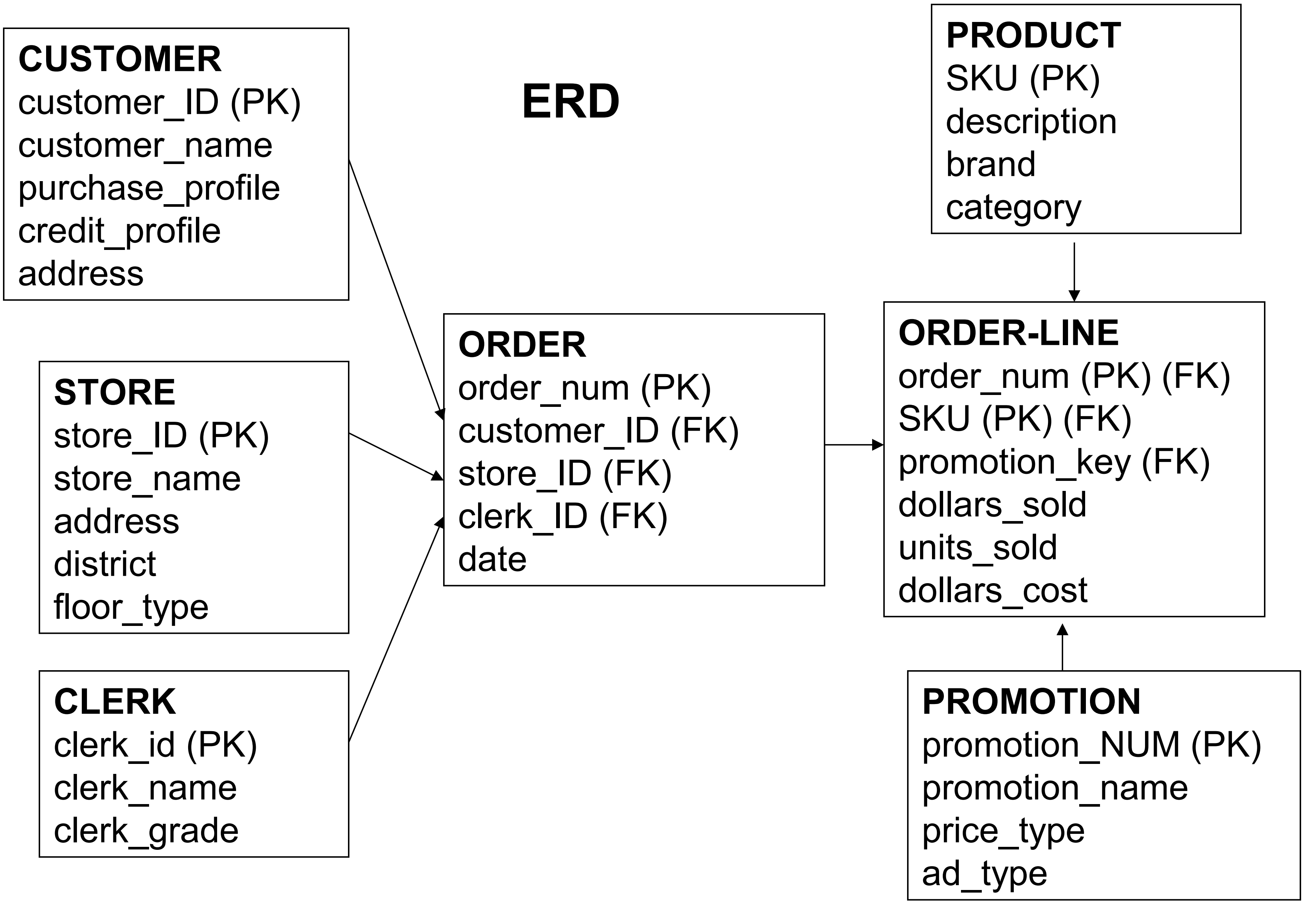
- Normally numeric and additive
- Non-key attributes in the fact table
    - Attributes in dimension tables are constants. Facts vary with the granularity of the fact table

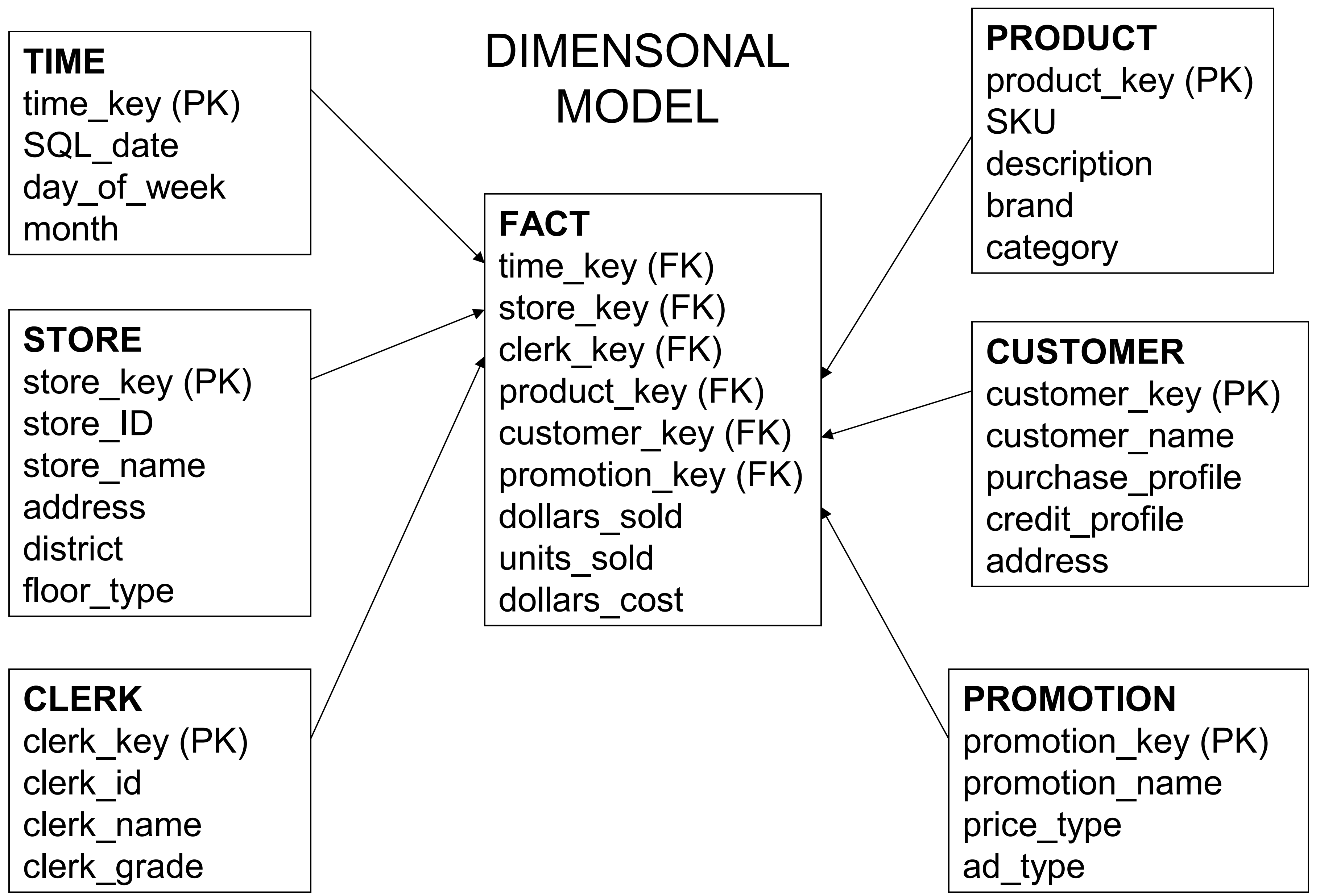
Dimensions

A table (or hierarchy of tables) connected with the fact table with keys and foreign keys

- Preferably single valued for each fact record (1:m)
- Connected with surrogate (generated) keys, not operational keys
- Dimension tables contain text or numeric attributes

ERD

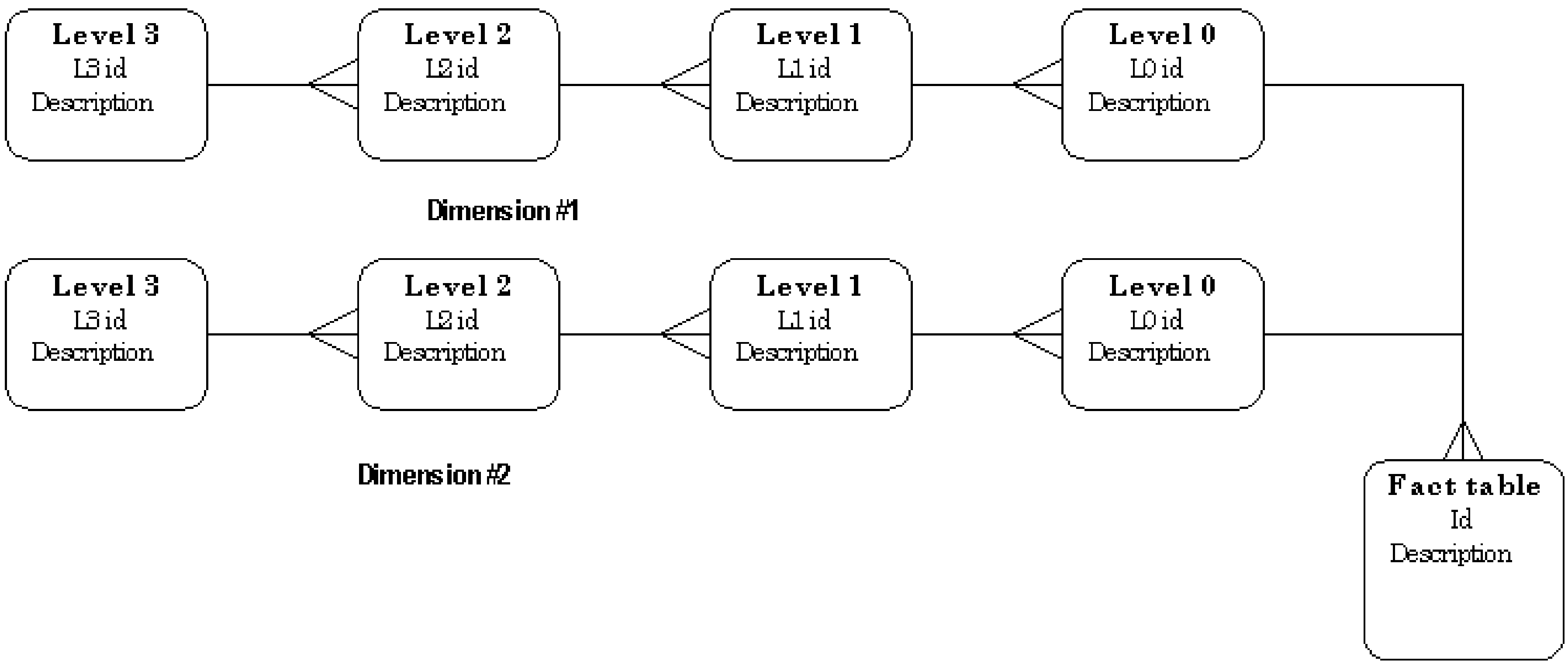

CUSTOMER
customer_ID (PK)
customer_name
purchase_profile
credit_profile
address
STORE
store_ID (PK)
store_name
address
district
floor_type
CLERK
clerk_id (PK)
clerk_name
clerk_grade
ORDER
order_num (PK)
customer_ID (FK)
store_ID (FK)
clerk_ID (FK)
date
PRODUCT
SKU (PK)
description
brand
category
ORDER-LINE
order_num (PK) (FK)
SKU (PK) (FK)
promotion_key (FK)
dollars_sold
units_sold
dollars_cost
PROMOTION
promotion_NUM (PK)
promotion_name
price_type
ad_type


DIMENSONAL
MODEL
TIME
time_key (PK)
SQL_date
day_of_week
month
STORE
store_key (PK)
store_ID
store_name
address
district
floor_type
CLERK
clerk_key (PK)
clerk_id
clerk_name
clerk_grade
FACT
time_key (FK)
store_key (FK)
clerk_key (FK)
product_key (FK)
customer_key (FK)
promotion_key (FK)
dollars_sold
units_sold
dollars_cost
PRODUCT
product_key (PK)
SKU
description
brand
category
CUSTOMER
customer_key (PK)
customer_name
purchase_profile
credit_profile
address
PROMOTION
promotion_key (PK)
promotion_name
price_type
ad_type

Snowflaking & Hierarchies

- Efficiency vs Space
- Understandability
- M:N relationships

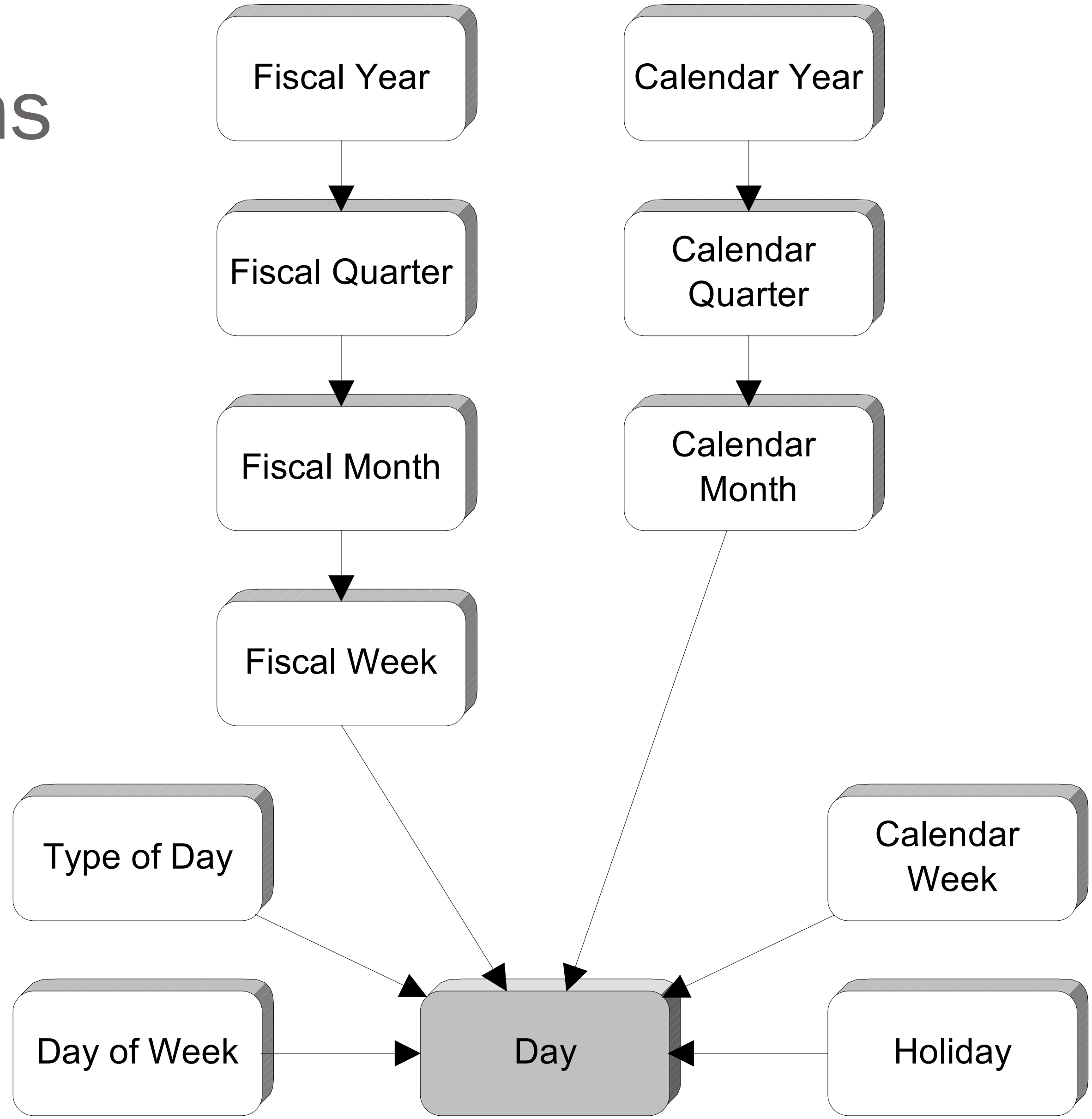
Simple DW hierarchy pattern.



Good Attributes

- Verbose
- Descriptive
- Complete
- Quality assured
- Indexed (b-tree vs bitmap)
- Equally available
- Documented

Date Dimensions



Attribute Name	Attribute Description	Sample Values
Day	The specific day that an activity took place.	06/04/1998; 06/05/1998
Day of Week	The specific name of the day.	Monday; Tuesday
Holiday	Identifies that this day is a holiday.	Easter; Thanksgiving
Type of Day	Indicates whether or not this day is a weekday or a weekend day.	Weekend; Weekday
Calendar Week	The week ending date, always a Saturday. Note that WE denotes	WE 06/06/1998; WE 06/13/1998
Calendar Month	The calendar month.	January,1998; February, 1998
Calendar Quarter	The calendar quarter.	1998Q1; 1998Q4
Calendar Year	The calendar year.	1998
Fiscal Week	The week that represents the corporate calendar. Note that the F	F Week 1 1998; F Week 46 1998
Fiscal Month	The fiscal period comprised of 4 or 5 weeks. Note that the F in the data	F January, 1998; F February, 1998
Fiscal Quarter	The grouping of 3 fiscal months.	F 1998Q1; F1998Q2
Fiscal Year	The grouping of 52 fiscal weeks / 12 fiscal months that comprise the financial year.	F 1998; F 1999

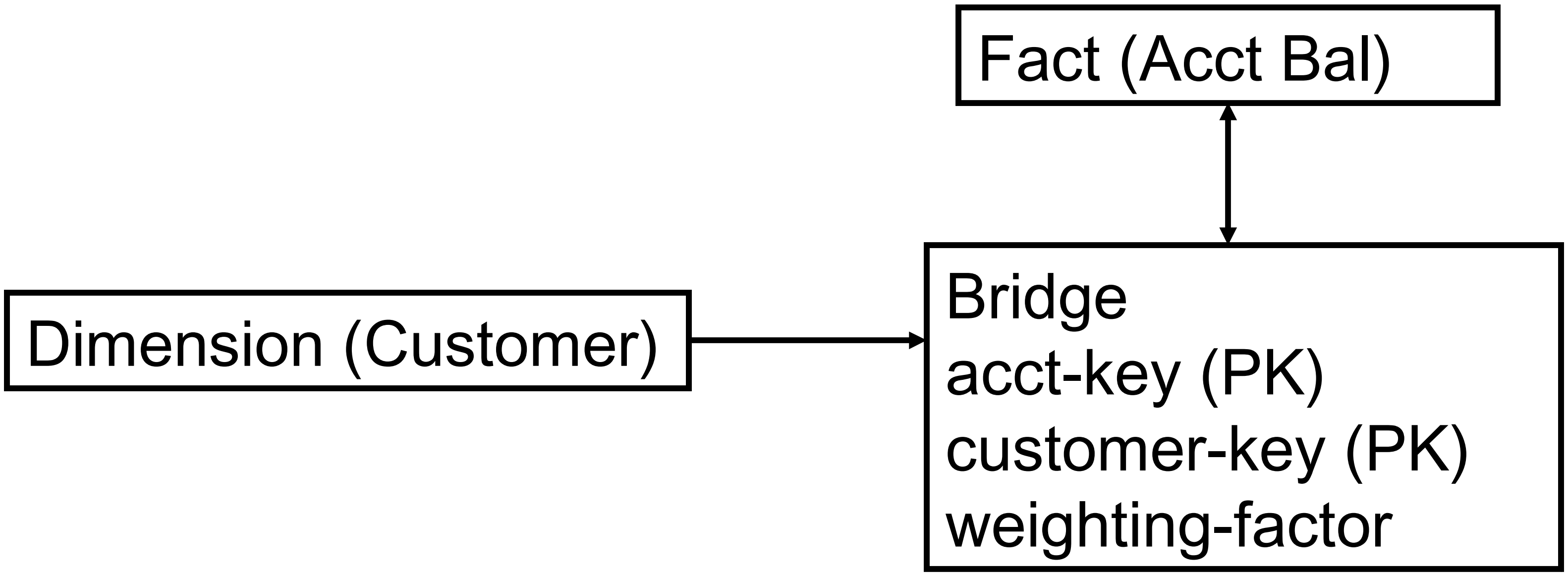
Slowly Changing Dimensions

(Addresses, Managers, etc.)

- Type 1: Store only the current value
- Type 2: Create a dimension record for each value (with or without date stamps)
- Type 3: Create an attribute in the dimension record for previous value

Many to many

- Use a Bridge Table
- Add a weighting factor to correct fact addition



Recursive

- Use a Bridge Table
- Add a level count and bottom flag