PUNE INSTITUTE OF COMPUTER TECHNOLOGY

Subject: ADBMS (LP LAB)

Name: Aditya Kangune **Roll No. : 33323**

Batch: K11 **Academic Year: 2021-22**

Case Study
Star and Snowflake schema

Aim:

Design conceptual model using Star and Snowflake schema for any one database.

Theory:

Multidimensional Schema:

The relations in a data warehouse schema can be classified as fact tables and dimension tables.

Fact table:

1. Fact tables record information about individual events, such as sales in this case, and are usually very large.
2. The attributes in the fact table can be classified as either dimension attributes or measure attributes.

Measure Attributes:

The measure attributes store quantitative information, which can be aggregated upon.

Dimension Attributes:

1. Dimension attributes are dimensions upon which measure attributes, and summaries of measure attributes, are grouped and viewed.
2. To minimize storage requirements, dimension attributes are usually short identifiers that are foreign keys into other tables called dimension tables.

Data that can be modeled using dimension attributes and measure attributes is called multidimensional data.

Following is the Star and snowflake schema for a sales database:

1. Star Schema:

1. Star schema includes a fact table at the center with multiple dimension tables.
2. It consists of having foreign keys from the fact table to the dimension tables.
3. Star schema is the simplest data warehouse schema and is optimized for querying large databases.

Model of Star schema:

1. In Star Schema, Business process data, which holds the quantitative data about a business is distributed in fact tables, and dimensions which are descriptive characteristics related to fact data.

2. Often, A Star Schema having multiple dimensions is termed a Centipede Schema. It is easy to handle a star schema that has dimensions of a few attributes.

Advantages of star schema:

Simpler Queries:

Join logic of star schema is quite cinch in comparison to other join logic which is needed to fetch data from a transactional schema that is highly normalized.

Simplified Business Reporting Logic:

In comparison to a transactional schema that is highly normalized, the star schema makes simpler common business reporting logic, such as as-of reporting and period-over-period.

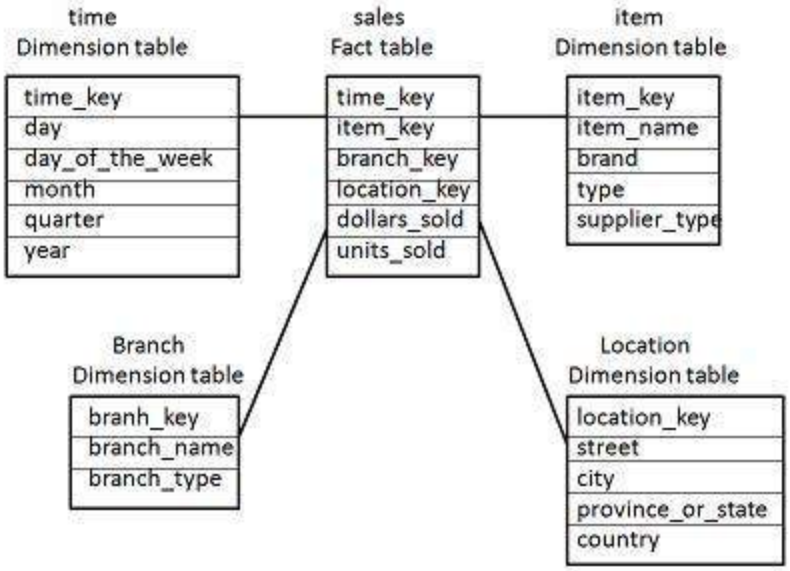
Feeding Cubes:

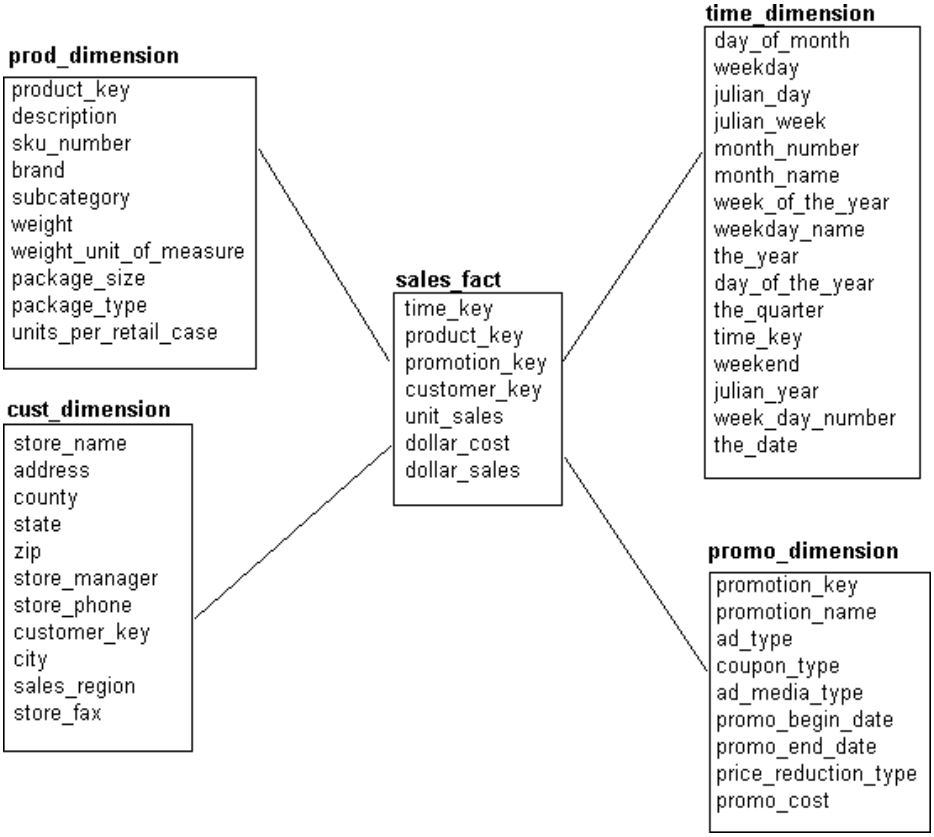
Star schema is widely used by all OLAP systems to design OLAP cubes efficiently. In fact, major OLAP systems deliver a ROLAP mode of operation which can use a star schema as a source without designing a cube structure.

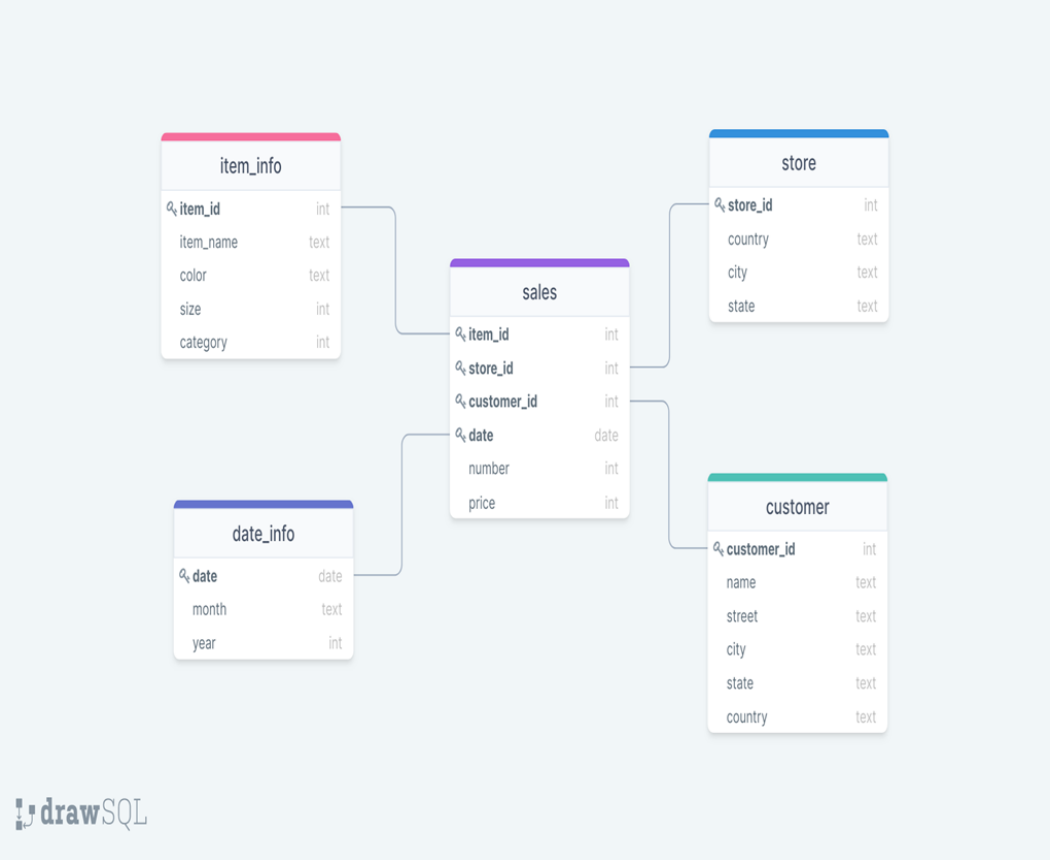
Disadvantages of star schema:

1. **Data integrity** is not enforced well since in a highly de-normalized schema state.
2. **Not flexible** in terms if analytical needs as a normalized data model.
3. Star schemas don't **reinforce many-to-many relationships** within business entities – at least not frequently.

Star schema examples:


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


prod_dimension
product_key
description
sku_number
brand
subcategory
weight
weight_unit_of_measure
package_size
package_type
units_per_retail_case
cust_dimension
store_name
address
county
state
zip
store_manager
store_phone
customer_key
city
sales_region
store_fax
sales_fact
time_key
product_key
promotion_key
customer_key
unit_sales
dollar_cost
dollar_sales
time_dimension
day_of_month
weekday
julian_day
julian_week
month_number
month_name
week_of_the_year
weekday_name
the_year
day_of_the_year
the_quarter
time_key
weekend
julian_year
week_day_number
the_date
promo_dimension
promotion_key
promotion_name
ad_type
coupon_type
ad_media_type
promo_begin_date
promo_end_date
price_reduction_type
promo_cost



2 . Snowflake Schema:

1. Snowflake Schema is an extension of the star schema.
2. A snowflake schema is used to model complex data warehouse designs that have multiple levels of dimensions.
3. Snowflake schema uses smaller disk space, but multiple dimension tables reduce the query performance.
4. The snowflake schema is a variant of the star schema.
5. Here, the centralized fact table is connected to multiple dimensions. In the snowflake schema, dimensions are present in a normalized form in multiple related tables.
6. The snowflake structure materialized when the dimensions of a star schema are detailed and highly structured, having several levels of relationship, and the child tables have multiple parent tables.
7. The snowflake effect affects only the dimension tables and does not affect the fact tables.

What is snowflaking?

1. The snowflake design is the result of further expansion and normalized of the dimension table. In other words, a dimension table is said to be snowflaked if the low-cardinality attribute of the dimensions has been divided into separate normalized tables.
2. These tables are then joined to the original dimension table with referential constraints (foreign key constrain).
3. Generally, snowflaking is not recommended in the dimension table, as it hampers the understandability and performance of the dimension model as more tables would be required to be joined to satisfy the queries.

Characteristics of snowflake schema:

The dimension model of a snowflake under the following conditions:

1. The snowflake schema uses small disk space.
2. It is easy to implement the dimension that is added to the schema.
3. There are multiple tables, so performance is reduced.
4. The dimension table consists of two or more sets of attributes that define information at different grains.
5. The sets of attributes of the same dimension table are being populated by different source systems.

Advantages:

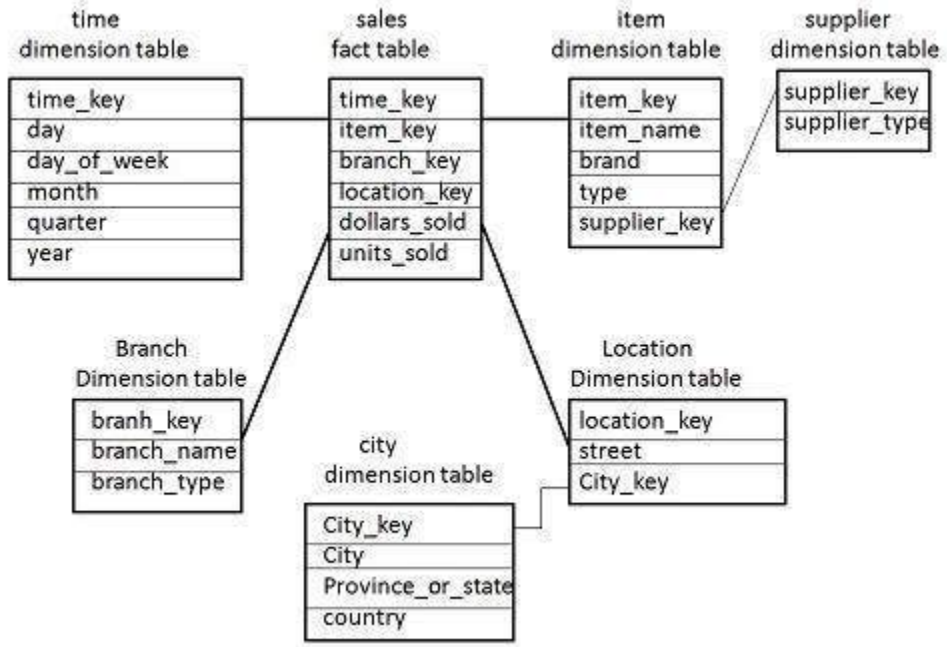
There are two main advantages of snowflake schema given below:

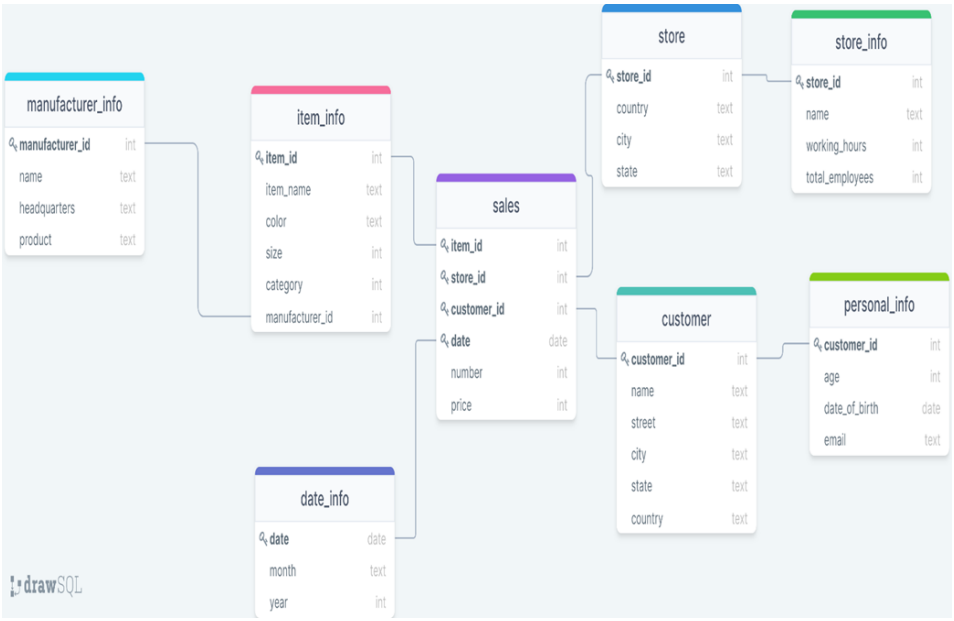
1. It provides structured data which reduces the problem of data integrity.
2. It uses small disk space because data are highly structured.

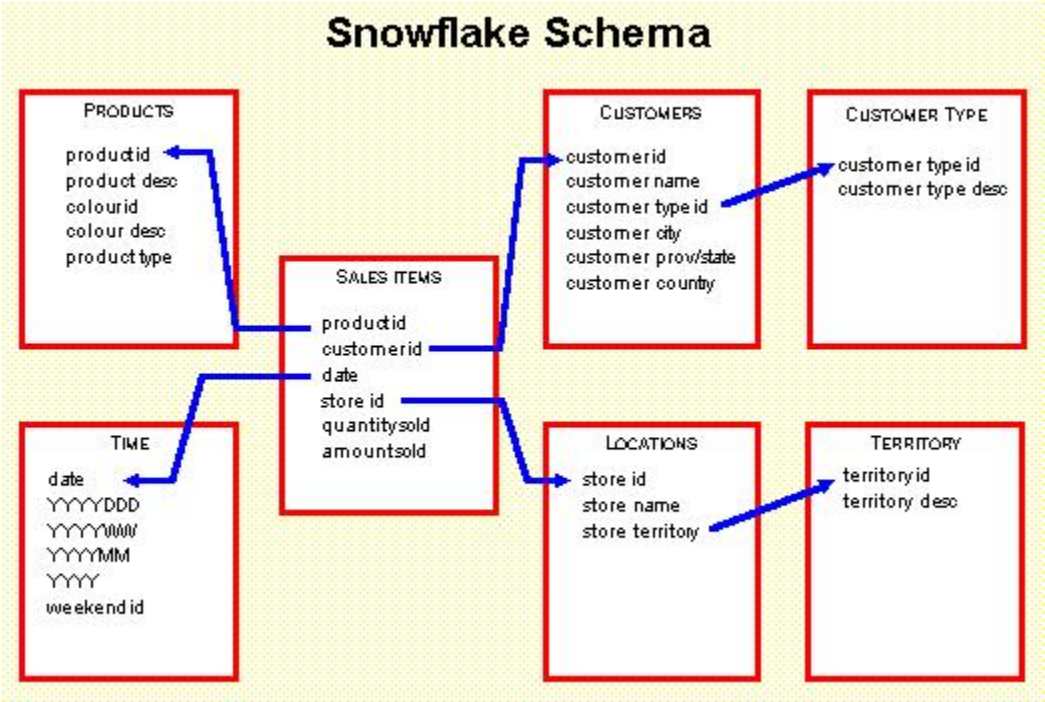
Disadvantages:

1. Snowflaking reduces space consumed by dimension tables but compared with the entire data warehouse the saving is usually insignificant.
2. Avoid snowflaking or normalization of a dimension table, unless required and appropriate.
3. Do not snowflake hierarchies of one dimension table into separate tables. Hierarchies should belong to the dimension table only and should never be snowflakes.
4. Multiple hierarchies that can belong to the same dimension have been designed at the lowest possible detail.

Snowflake schema examples:




manufacturer_info
manufacturer_id int
name text
headquarters text
product text
item_info
item_id int
item_name text
color text
size int
category int
manufacturer_id int
sales
item_id int
store_id int
customer_id int
date date
number int
price int
store
store_id int
country text
city text
state text
store_info
store_id int
name text
working_hours int
total_employees int
customer
customer_id int
name text
street text
city text
state text
country text
personal_info
customer_id int
age int
date_of_birth date
email text
date_info
date date
month text
year int
drawSQL


Snowflake Schema
PRODUCTS
productid
product desc
colourid
colour desc
product type
SALES ITEMS
productid
customerid
date
store id
quantitysold
amountsold
CUSTOMERS
customerid
customer name
customer type id
customer city
customer prov/state
customer country
CUSTOMER TYPE
customer type id
customer type desc
TIME
date
YYYYDDD
YYYYWW
YYYYMM
YYYY
weekend id
LOCATIONS
store id
store name
store territory
TERRITORY
territoryid
territory desc

Conclusion:

Conceptual models using Star and Snowflake schema were designed after the completion of this case study.