FACULTY OF EGINEERING

DATA MINING & WAREHOUSEING
LECTURE-06

MR. DHIRENDRA
ASSISTANT PROFESSOR
RAMA UNIVERSITY

OUTLINE

- ❖ **DATA WAREHOUSING – SCHEMAS**

- ❖ **STAR SCHEMA**

- ❖ **SNOWFLAKE SCHEMA**

- ❖ **FACT CONSTELLATION SCHEMA**
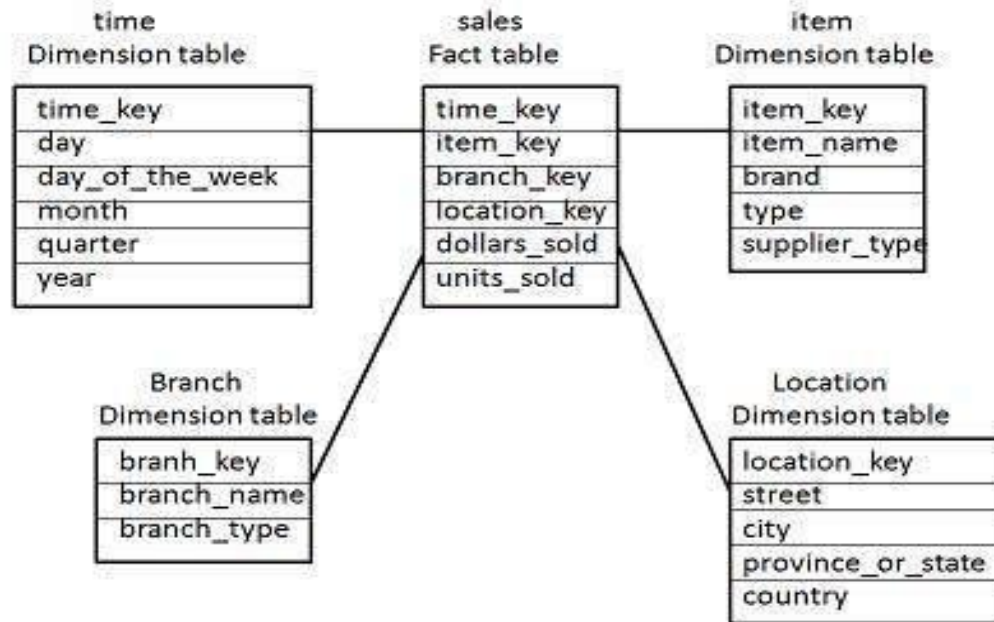
- ❖ **MCQ**

- ❖ **REFERENCES**

DATA WAREHOUSING - 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.

- Much like a database, a data warehouse also requires to maintain a schema.

- A database uses relational model, while a data warehouse uses Star, Snowflake, and Fact Constellation schema.
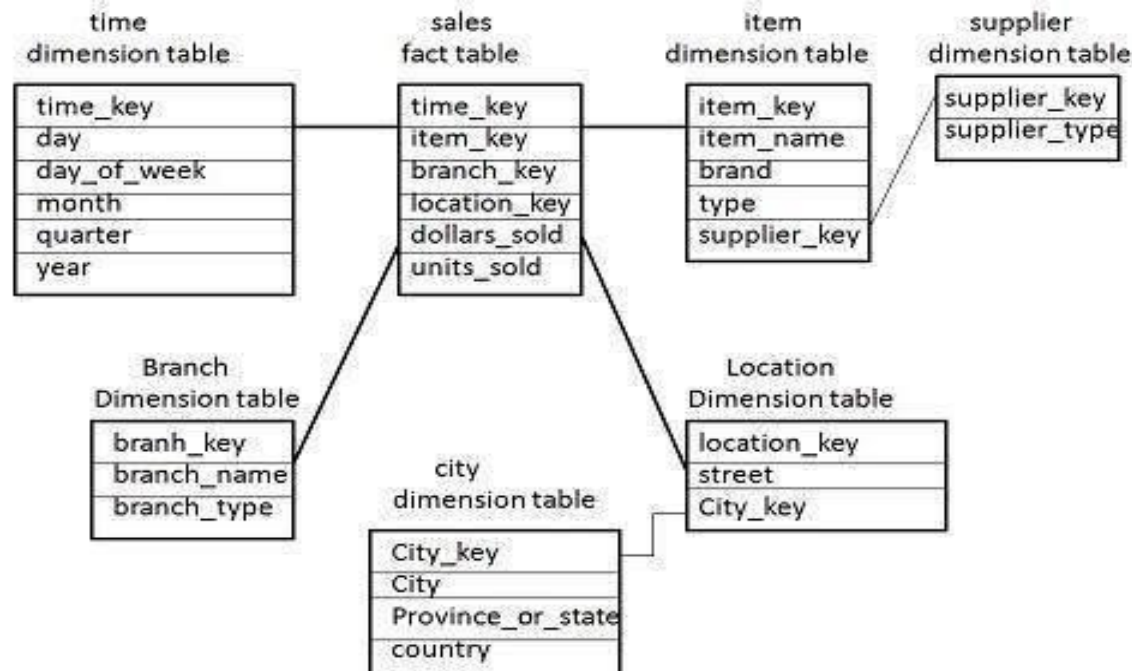
STAR SCHEMA

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

- This dimension table contains the set of attributes.

- There is a fact table at the center. It contains the keys to each of four dimensions.

- The fact table also contains the attributes, namely dollars sold and units sold.

- The following diagram shows the sales data of a company with respect to the four dimensions, namely time, item, branch, and location.

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

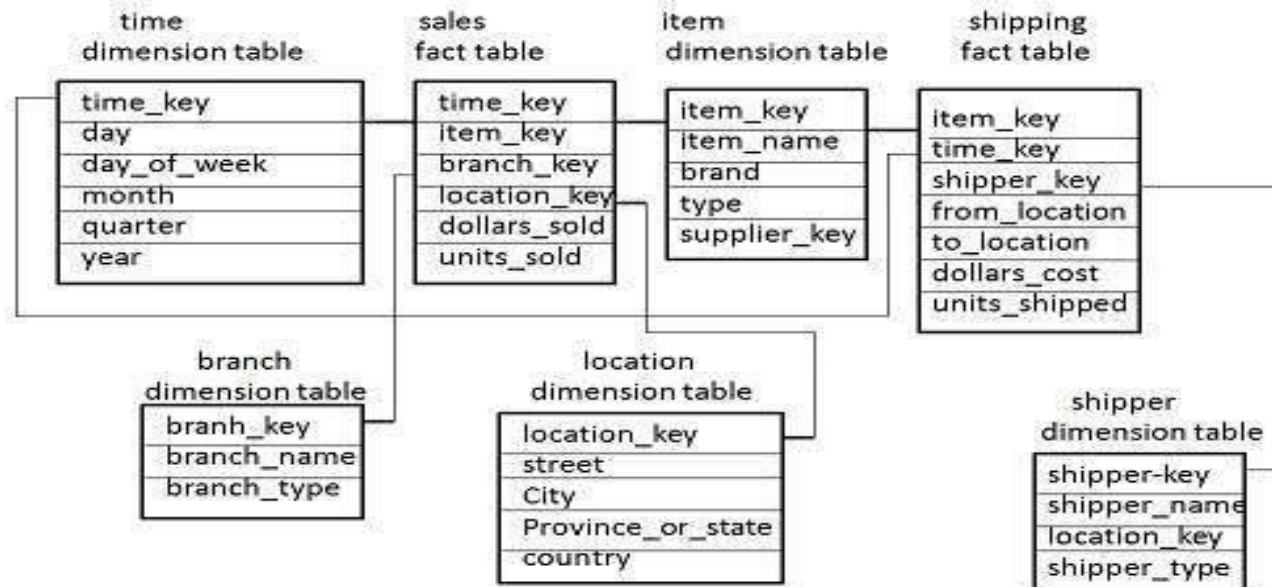| location_key |
| street |
| city |
| province_or_state |
| country |

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.

- Now the item dimension table contains the attributes item_key, item_name, type, brand, and supplier-key.

- The supplier key is linked to the supplier dimension table. The supplier dimension table contains the attributes supplier key and supplier type.
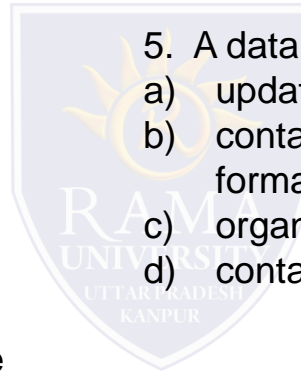
FACT CONSTELLATION SCHEMA

- 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.

- The sales fact table is same as that in the star schema.

- The shipping fact table has the five dimensions, namely item_key, time_key, shipper_key, from_location, to_location.

- The shipping fact table also contains two measures, namely dollars sold and units sold.

- 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.

Multiple Choice Question

1. Detail data in single fact table is otherwise known as_____.
a) monoatomic data.
b) diatomic data.
c) atomic data.
d) multiatomic data.

2.. _____test is used in an online transactional processing environment.
a) MEGA.
b) MICRO
c) MACRO.
d) ACID.

3. _____ is a good alternative to the star schema.
a) Star schema.
b) Snowflake schema.
c) Fact constellation.
d) Star-snowflake schema.

4. The biggest drawback of the level indicator in the classic star-schema is that it limits_____.
a) quantify
b) qualify
c) flexibility
d) ability

5. A data warehouse is _____.
a) updated by end users.
b) contains numerous naming conventions and formats
c) organized around important subject areas.
d) contains only current data.

REFERENCES

- https://www.tutorialspoint.com/dwh/dwh_overview.htm

- http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf
  DATA MINING BOOK WRITTEN BY Micheline Kamber
- https://www.javatpoint.com/three-tier-data-warehouse-architecture