

Laboratory 9

Title of the Laboratory Exercise: Multidimensional data modelling

1. Introduction and Purpose of Experiment

The multidimensional data model is an integral part of On-Line Analytical Processing (OLAP). Dimensional modelling always uses the concepts of facts, measures, and dimensions. Facts are typically (but not always) numeric values that can be aggregated. Dimensions are groups of hierarchies and descriptors that define the facts. By doing this lab, students will be able to design and implement multi-dimensional data model.

2. Aim and Objectives

Aim

- To design and implement concepts of multi-dimensional data modelling

Objectives

At the end of this lab, the student will be able to

- Design star and snowflake schema
- Implement multi-dimensional data modelling

3. Experimental Procedure

- Analyse the problem statement
- Design a data cube which contains a fact table and dimension table
- Document the Results
- Analyse and discuss the outcomes of your experiment

4. Questions

- Design the following multi-dimensional data models for the Banking enterprise
 - Star schema
 - Snowflake schema

5. Calculations/Computations/Algorithms

6. Presentation of Results

Star Schema

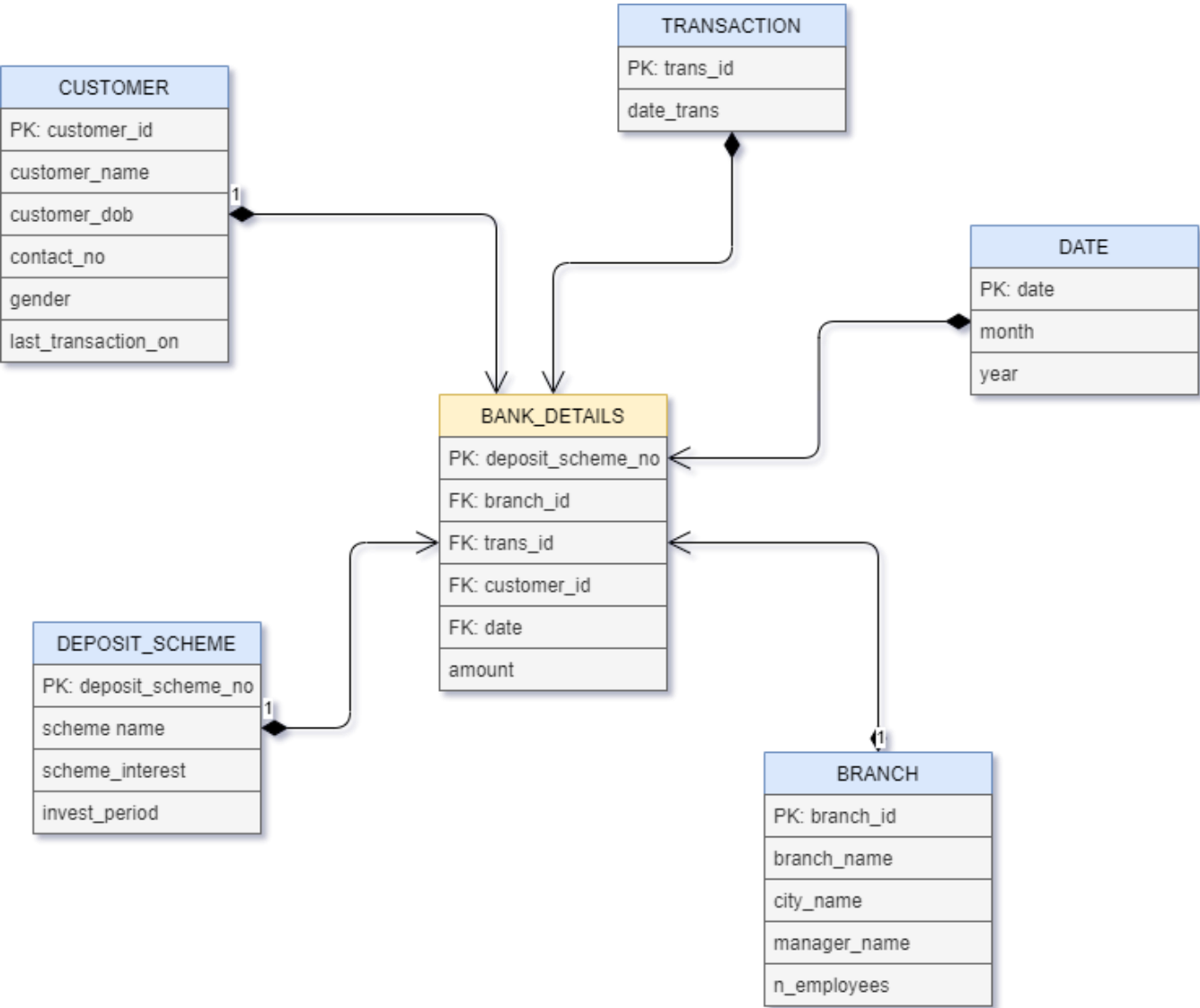


Figure 1 Star Scheme

Snowflake Schema

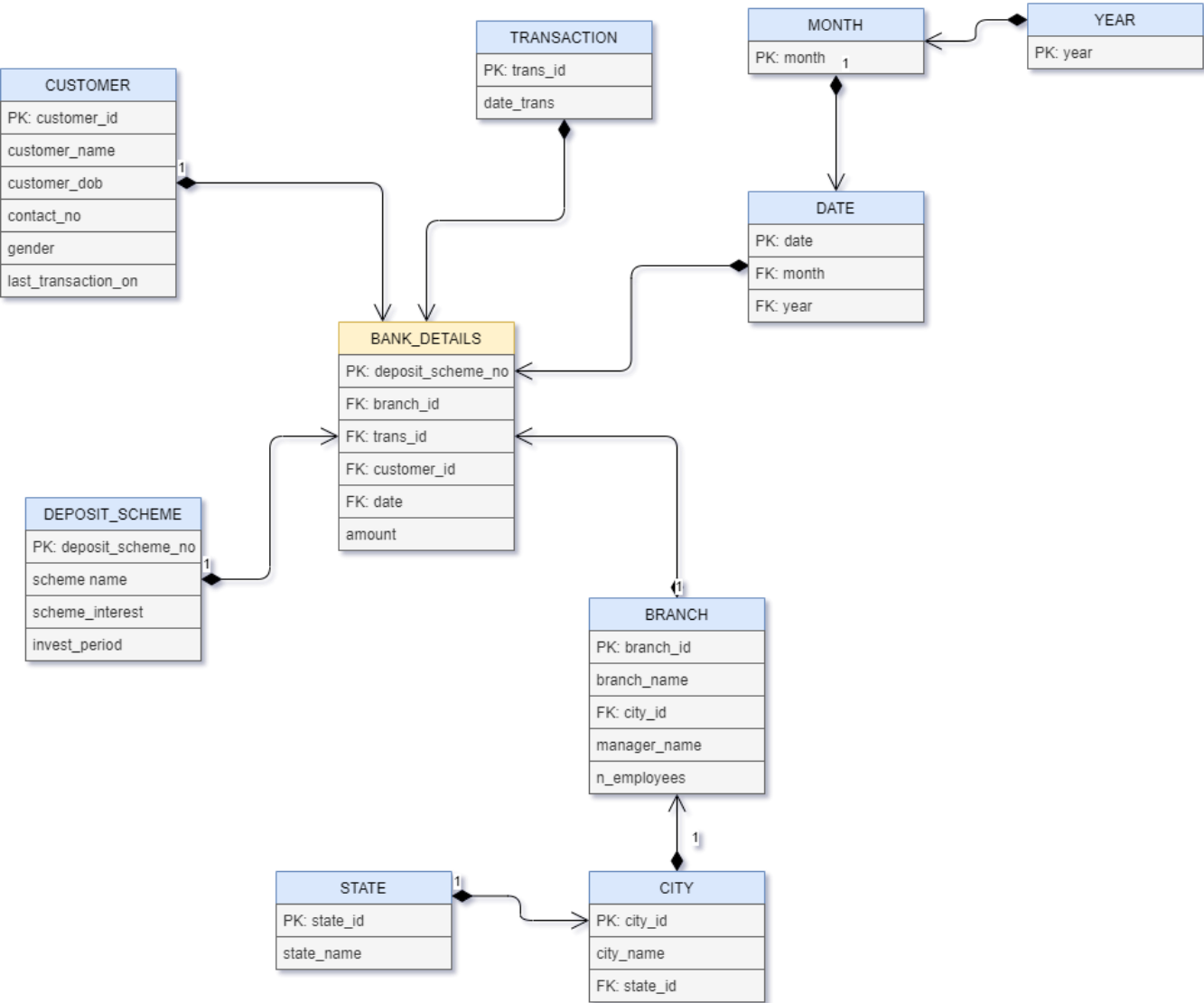


Figure 2 Snowflake Schema

7. Conclusions

Model of Star Schema –

In Star Schema, Business process data, that holds the quantitative data about a business is distributed in fact tables, and dimensions which are descriptive characteristics related to fact data. Sales price, sale quantity, distant, speed, weight, and weight measurements are few examples of fact data in star schema.

Often, A Star Schema having multiple dimensions is termed as Centipede Schema. It is easy to handle a star schema which have dimensions of few attributes.

Advantages of Star Schema –

1. Simpler Queries:

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

2. Simplified Business Reporting Logic:

In compared 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.

3. 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 –

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

Characteristics of snowflake schema: The dimension model of snowflake under the following conditions:

- The snowflake schema uses small disk space.
- It is easy to implement dimension is added to schema.
- There are multiple tables, so performance is reduced.
- The dimension table consist of two or more sets of attributes which define information at different grains.

- 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:

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

Disadvantages:

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

(Source: <https://www.geeksforgeeks.org/snowflake-schema-in-data-warehouse-model/>)

8. Comments

1. Limitations of Experiments

- The primary limitation of the snowflake schema is that the additional levels of attribute normalization adds complexity to source query joins, when compared to the star schema.
- The main limitation of the star schema is that data integrity is not well-enforced due to its denormalized state

2. Limitations of Results

- In snowflake schema you may need complex joins to perform a query due to the number of tables added. Hence query performance will be degraded.
- In star schema, since the Master Data table is acting as the Dimension table which is inside the cube, So the Master data table cannot be reused.

3. Learning happened

- We learnt how to make star and snowflake schema for a given scenario

4. Recommendations

- Design the schema carefully considering all cases, the design is very important before implementation of the physical tables on the db.