Data Warehousing and Data Mining Experiment 1

Experiment 1 - Program on uninformed search method

Name: Khan Mohammed Ashaaf

Class and Batch: TE3 Batch-B

Roll no: 32

Aim: Prepare the objective of given Mini- Project with respect to building Data warehouse/Data Mart

1. Write Detailed Problem statement and design dimensional modelling (creation of star and snowflake schema)
2. Implementation of all dimension tables and fact table. Develop an application that uses GUI components.

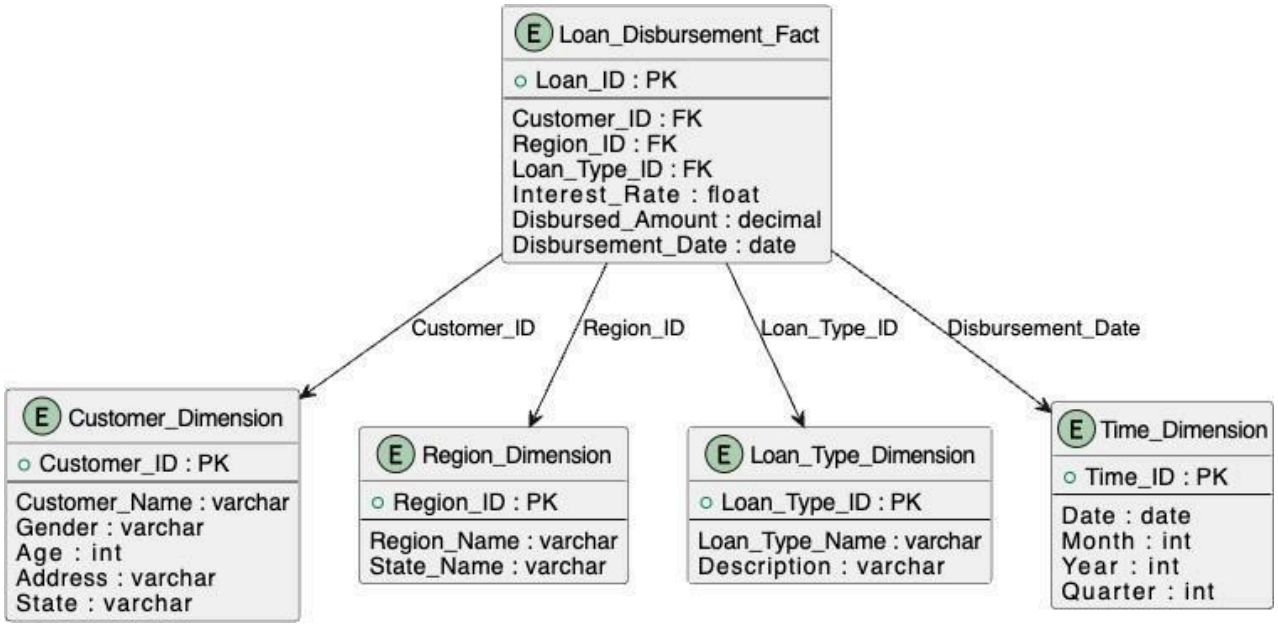
Problem Statement:

A bank wants to develop a data warehouse for effective decision-making regarding its loan schema. The bank provides loans to customers for various purposes, such as house building, car purchase, education, personal expenses, etc. The entire country is categorized into multiple regions: North, South, East, and West. Each region consists of a set of states. Loans are distributed to customers at interest rates that change from time to time. Additionally, at a given point in time, different types of loans have different interest rates. The data warehouse should record an entity for each loan disbursement to a customer based on the above business scenario.

Design an IPD (Information Package Diagram) clearly, explaining all aspects of the diagram. Also, design Star and Snowflake schemas for the given case.

Star Schema:

Diagram:



Source Code:

```
-- Customer Dimension
CREATE TABLE Customer_Dimension (
    Customer_ID INT PRIMARY KEY,
    Customer_Name VARCHAR(100) NOT NULL,
    Gender VARCHAR(10),
    Age INT,
    Address VARCHAR(255),
    State VARCHAR(50)
);
-- Region Dimension
CREATE TABLE Region_Dimension (
    Region_ID INT PRIMARY KEY,
    Region_Name VARCHAR(50) NOT NULL,
    State_Name VARCHAR(50) NOT NULL
);
-- Loan Type Dimension
CREATE TABLE Loan_Type_Dimension (
    Loan_Type_ID INT PRIMARY KEY,
    Loan_Type_Name VARCHAR(50) NOT NULL,
    Description VARCHAR(255)
);
-- Time Dimension
CREATE TABLE Time_Dimension (
    Time_ID INT PRIMARY KEY,
    Time_Date DATE NOT NULL,
    Month INT NOT NULL,
    Year INT NOT NULL,
    Quarter INT NOT NULL
);
-- Loan Disbursement Fact Table
CREATE TABLE Loan_Disbursement_Fact (
    Loan_ID INT PRIMARY KEY,
    Customer_ID INT,
    Region_ID INT,
    Loan_Type_ID INT,
```

```
    Interest_Rate FLOAT,
    Disbursed_Amount DECIMAL(15,2),
    Disbursement_Date DATE,
    Time_ID INT,
    FOREIGN KEY (Customer_ID) REFERENCES Customer_Dimension(Customer_ID),
    FOREIGN KEY (Region_ID) REFERENCES Region_Dimension(Region_ID),
    FOREIGN KEY (Loan_Type_ID) REFERENCES Loan_Type_Dimension(Loan_Type_ID),
    FOREIGN KEY (Time_ID) REFERENCES Time_Dimension(Time_ID)
);
-- Show Tables
SELECT table_name FROM user_tables;
-- Show Columns for Each Table
SELECT column_name, data_type FROM user_tab_columns WHERE table_name = 'CUSTOMER_DIMENSION';
SELECT column_name, data_type FROM user_tab_columns WHERE table_name = 'REGION_DIMENSION';
SELECT column_name, data_type FROM user_tab_columns WHERE table_name = 'LOAN_TYPE_DIMENSION';
SELECT column_name, data_type FROM user_tab_columns WHERE table_name = 'TIME_DIMENSION';
SELECT column_name, data_type FROM user_tab_columns WHERE table_name = 'LOAN_DISBURSEMENT_FACT';
```

Output Screenshot:

Table created.

Table created.

Table created.

Table created.

Table created.

TABLE_NAME
CUSTOMER_DIMENSION
LOAN_DISBURSEMENT_FACT
LOAN_TYPE_DIMENSION
REGION_DIMENSION
TIME_DIMENSION

Download CSV

5 rows selected.

COLUMN_NAME	DATA_TYPE
CUSTOMER_ID	NUMBER
CUSTOMER_NAME	VARCHAR2
GENDER	VARCHAR2
AGE	NUMBER
ADDRESS	VARCHAR2
STATE	VARCHAR2

Download CSV

6 rows selected.

COLUMN_NAME	DATA_TYPE
REGION_ID	NUMBER
REGION_NAME	VARCHAR2
STATE_NAME	VARCHAR2

Download CSV

3 rows selected.

COLUMN_NAME	DATA_TYPE
LOAN_TYPE_ID	NUMBER
LOAN_TYPE_NAME	VARCHAR2
DESCRIPTION	VARCHAR2

Download CSV

3 rows selected.

COLUMN_NAME	DATA_TYPE
TIME_ID	NUMBER
TIME_DATE	DATE
MONTH	NUMBER
YEAR	NUMBER
QUARTER	NUMBER

Download CSV

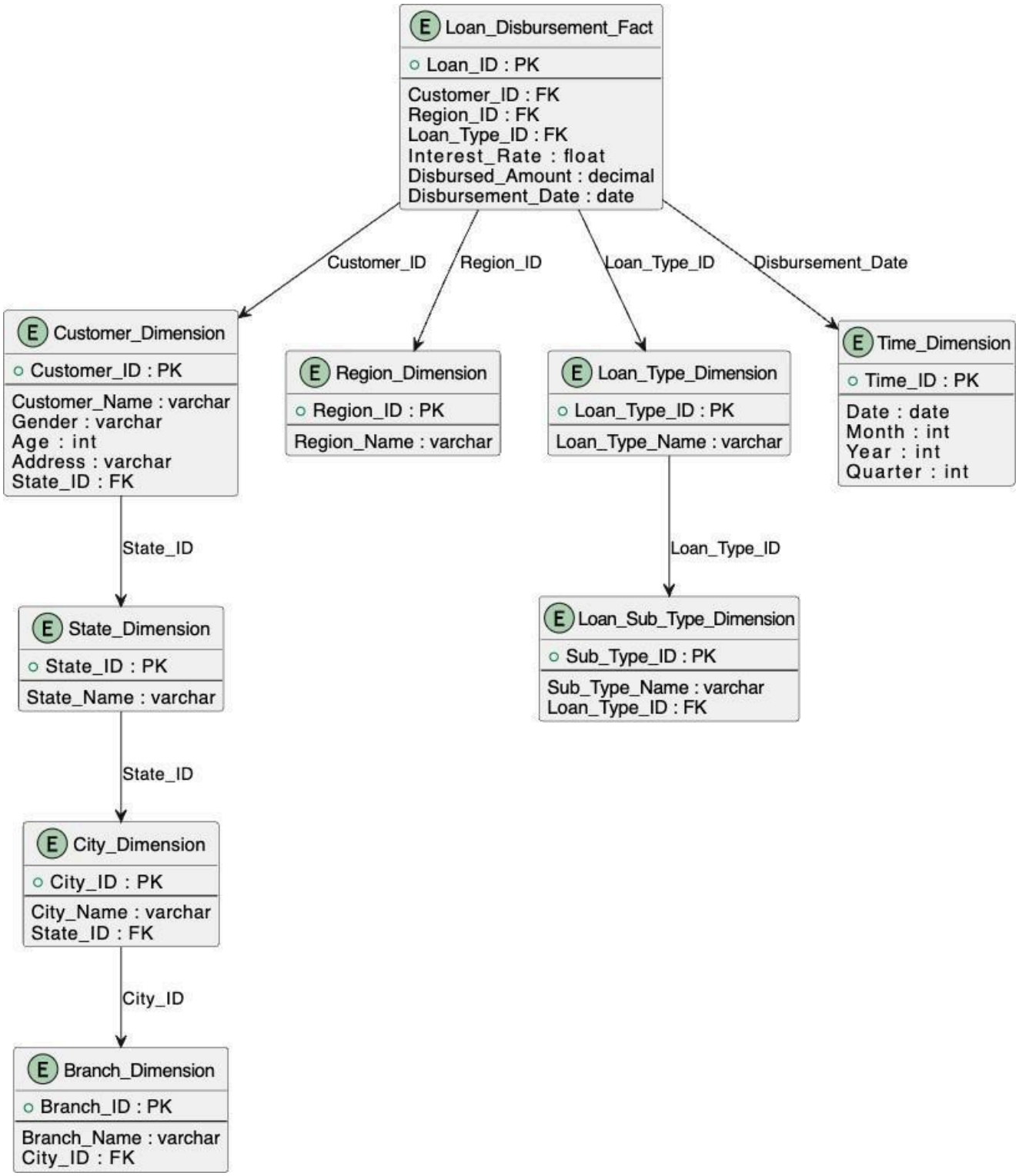
5 rows selected.

COLUMN_NAME	DATA_TYPE
LOAN_ID	NUMBER
CUSTOMER_ID	NUMBER
REGION_ID	NUMBER
LOAN_TYPE_ID	NUMBER
INTEREST_RATE	FLOAT
DISBURSED_AMOUNT	NUMBER
DISBURSEMENT_DATE	DATE
TIME_ID	NUMBER

Download CSV

Snowflake Schema:

Diagram:



Source Code:

```
-- State Dimension
CREATE TABLE State_Dimension (
    State_ID INT PRIMARY KEY,
    State_Name VARCHAR(50) NOT NULL
);
-- City Dimension
CREATE TABLE City_Dimension (
    City_ID INT PRIMARY KEY,
    City_Name VARCHAR(50) NOT NULL,
    State_ID INT,
    FOREIGN KEY (State_ID) REFERENCES State_Dimension(State_ID)
);
-- Branch Dimension
CREATE TABLE Branch_Dimension (
    Branch_ID INT PRIMARY KEY,
    Branch_Name VARCHAR(100) NOT NULL,
    City_ID INT,
    FOREIGN KEY (City_ID) REFERENCES City_Dimension(City_ID)
);
-- Customer Dimension
CREATE TABLE Customer_Dimension (
    Customer_ID INT PRIMARY KEY,
    Customer_Name VARCHAR(100) NOT NULL,
    Gender VARCHAR(10),
    Age INT,
    Address VARCHAR(255),
    State_ID INT,
    FOREIGN KEY (State_ID) REFERENCES State_Dimension(State_ID)
);
-- Region Dimension
CREATE TABLE Region_Dimension (
    Region_ID INT PRIMARY KEY,
    Region_Name VARCHAR(50) NOT NULL
);
```

```
-- Loan Type Dimension
CREATE TABLE Loan_Type_Dimension (
    Loan_Type_ID INT PRIMARY KEY,
    Loan_Type_Name VARCHAR(50) NOT
    NULL
);
-- Loan Sub-Type Dimension
CREATE TABLE Loan_Sub_Type_Dimension (
    Sub_Type_ID INT PRIMARY KEY,
    Sub_Type_Name VARCHAR(50) NOT NULL,
    Loan_Type_ID INT,
    FOREIGN KEY (Loan_Type_ID) REFERENCES Loan_Type_Dimension(Loan_Type_ID)
);
-- Time Dimension (Updated to avoid reserved keyword issue)
CREATE TABLE Time_Dimension (
    Time_ID INT PRIMARY KEY,
    Time_Date DATE NOT NULL,
    Month INT NOT NULL,
    Year INT NOT NULL,
    Quarter INT NOT NULL
);
-- Fact Table (Updated Foreign Key Reference)
CREATE TABLE Loan_Disbursement_Fact (
    Loan_ID INT PRIMARY KEY,
    Customer_ID INT,
    Region_ID INT,
    Loan_Type_ID INT,
    Interest_Rate FLOAT,
    Disbursed_Amount DECIMAL(15,2),
    Disbursement_Date DATE,
    Time_ID INT, -- Reference to the primary key of Time_Dimension
    FOREIGN KEY (Customer_ID) REFERENCES Customer_Dimension(Customer_ID),
    FOREIGN KEY (Region_ID) REFERENCES Region_Dimension(Region_ID),
    FOREIGN KEY (Loan_Type_ID) REFERENCES Loan_Type_Dimension(Loan_Type_ID),
    FOREIGN KEY (Time_ID) REFERENCES Time_Dimension(Time_ID) -- Updated reference to Time_ID
);
```

```
-- Show Tables
SELECT table_name FROM user_tables;

SELECT column_name, data_type
FROM user_tab_columns
WHERE table_name = 'STATE_DIMENSION';

SELECT column_name, data_type
FROM user_tab_columns
WHERE table_name = 'CITY_DIMENSION';

SELECT column_name, data_type
FROM user_tab_columns
WHERE table_name = 'BRANCH_DIMENSION';

SELECT column_name, data_type
FROM user_tab_columns
WHERE table_name = 'CUSTOMER_DIMENSION';

SELECT column_name, data_type
FROM user_tab_columns
WHERE table_name = 'REGION_DIMENSION';

SELECT column_name, data_type
FROM user_tab_columns
WHERE table_name = 'LOAN_TYPE_DIMENSION';

SELECT column_name, data_type
FROM user_tab_columns
WHERE table_name = 'LOAN_SUB_TYPE_DIMENSION';

SELECT column_name, data_type
FROM user_tab_columns
WHERE table_name = 'TIME_DIMENSION';

SELECT column_name, data_type
```

```
FROM user_tab_columns
WHERE table_name = 'LOAN_DISBURSEMENT_FACT';
```

Output Screenshot:

Table created.

TABLE_NAME
BRANCH_DIMENSION
CITY_DIMENSION
CUSTOMER_DIMENSION
LOAN_DISBURSEMENT_FACT
LOAN_SUB_TYPE_DIMENSION
LOAN_TYPE_DIMENSION
REGION_DIMENSION
STATE_DIMENSION
TIME_DIMENSION

Download CSV

9 rows selected.

COLUMN_NAME	DATA_TYPE
STATE_ID	NUMBER
STATE_NAME	VARCHAR2

Download CSV

2 rows selected.

COLUMN_NAME	DATA_TYPE
CITY_ID	NUMBER
CITY_NAME	VARCHAR2
STATE_ID	NUMBER

Download CSV

3 rows selected.

COLUMN_NAME	DATA_TYPE
BRANCH_ID	NUMBER
BRANCH_NAME	VARCHAR2
CITY_ID	NUMBER

Download CSV

3 rows selected.

COLUMN_NAME	DATA_TYPE
CUSTOMER_ID	NUMBER
CUSTOMER_NAME	VARCHAR2
GENDER	VARCHAR2
AGE	NUMBER
ADDRESS	VARCHAR2
STATE_ID	NUMBER

Download CSV

6 rows selected.

COLUMN_NAME	DATA_TYPE
REGION_ID	NUMBER
REGION_NAME	VARCHAR2

Download CSV

2 rows selected.

COLUMN_NAME	DATA_TYPE
LOAN_TYPE_ID	NUMBER
LOAN_TYPE_NAME	VARCHAR2

Download CSV

2 rows selected.

COLUMN_NAME	DATA_TYPE
SUB_TYPE_ID	NUMBER
SUB_TYPE_NAME	VARCHAR2
LOAN_TYPE_ID	NUMBER

Download CSV

3 rows selected.

COLUMN_NAME	DATA_TYPE
TIME_ID	NUMBER
TIME_DATE	DATE
MONTH	NUMBER
YEAR	NUMBER
QUARTER	NUMBER

Download CSV

5 rows selected.

COLUMN_NAME	DATA_TYPE
LOAN_ID	NUMBER
CUSTOMER_ID	NUMBER
REGION_ID	NUMBER
LOAN_TYPE_ID	NUMBER
INTEREST_RATE	FLOAT
DISBURSED_AMOUNT	NUMBER
DISBURSEMENT_DATE	DATE
TIME_ID	NUMBER

Download CSV

8 rows selected.

Review Questions:

1. **In a star schema, how is the fact table typically related to the dimension tables?**

Ans) In a star schema, the fact table is typically related to the dimension tables through foreign key relationships. The fact table contains measurable data (quantitative values), while each dimension table provides descriptive attributes related to the facts. The dimension tables are directly connected to the fact table without further normalization, forming a star-like structure.

2. **What is the main difference between a star schema and a snowflake schema?**

Ans) The main difference between a star schema and a snowflake schema is normalization:

- Star Schema: Dimension tables are denormalized, meaning they contain redundant data to improve query performance and simplify joins.
- Snowflake Schema: Dimension tables are normalized by splitting them into multiple related tables to reduce redundancy and improve storage efficiency, but at the cost of more complex queries and additional joins.