Name: Asmi Jain

Div: TY-3

Roll No: 25

Batch: B

Subject: DWM

Experiment No 01

AIM : For the objective given below, build Data warehouse/ DataMart.Write detailed Problem statement and design dimensional modeling (Creation of star and snowflake schema) Implementation of all dimension tables and fact table.

Problem statement :

Design a data warehouse for a regional weather bureau. The weather bureau has about 100 probs, which are scattered throughout various land and ocean locations in the region to collect basic weather data, including air pressure, temperature and precipitation at each hour. All data are sent to the central station, which has collected such data for more than 10 years. Design Star schema and Snowflake schema such that it should facilitate efficient querying and online analytical processing and derive general weather patterns in multidimensional space. Explain all aspects of the diagram. Design Star and Snowflake schema for above case.

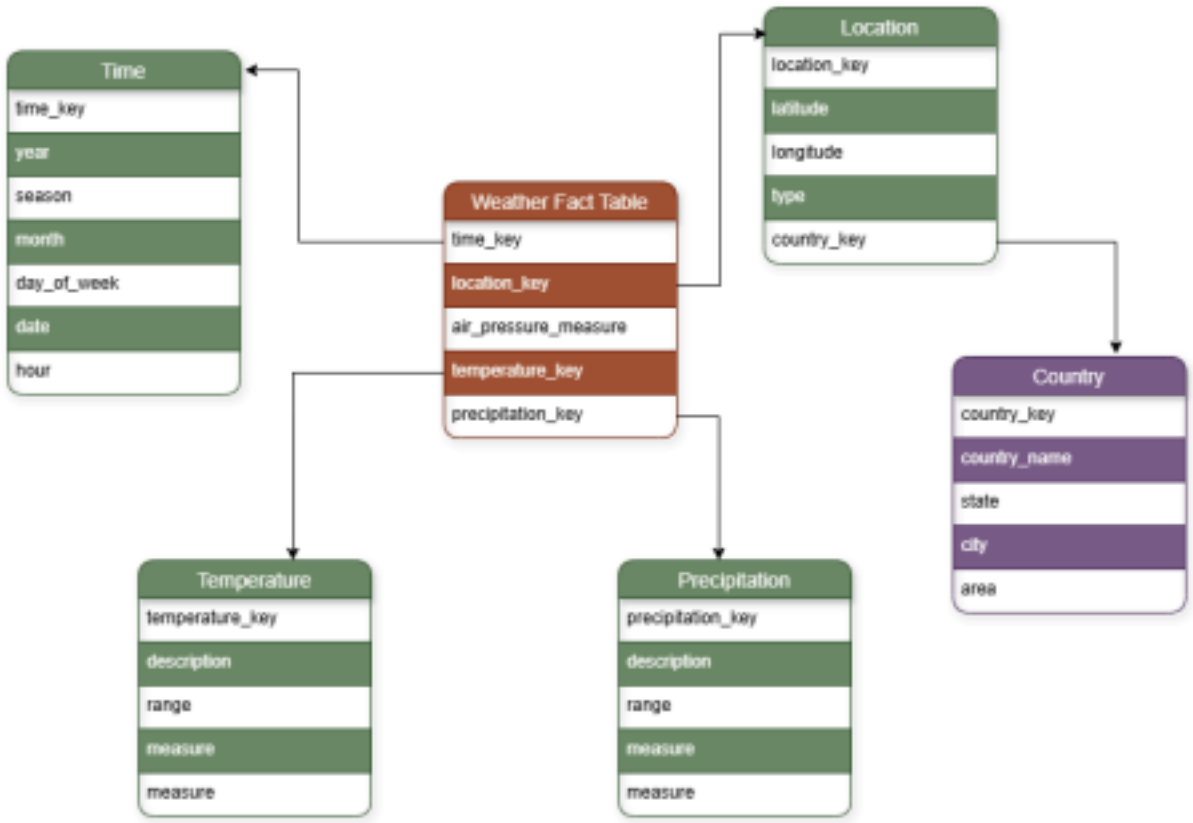
THEORY :

- ➢Star Schema: Star schema is the simplest method for arranging data in a data warehouse. It contains a fact table at the center connected to dimension tables around it. Star schema is most effective for quick and simple data query execution.

➢Snowflake Schema: Snowflake schema is a more complex method of storing data in which fact tables, dimension tables and sub-dimension tables are connected through foreign keys. Snowflake is most effective for in-depth data query analyses.

Dimensional Modelling :

(creation of star and snowflake schema)



Code :

```
CREATE TABLE Time (
    time_key INT PRIMARY KEY,
    year INT,
    season VARCHAR(20),
    month INT,
```

```
    day_of_week VARCHAR(20),
    dates DATE,
    hour INT
)

CREATE TABLE Country (
    country_key INT PRIMARY KEY,
    country_name VARCHAR(50),
    state VARCHAR(50),
    city VARCHAR(50),
    area VARCHAR(100)
)

CREATE TABLE Location (
    location_key INT PRIMARY KEY,
    latitude DECIMAL(10, 6),
    longitude DECIMAL(10, 6),
    type VARCHAR(50),
    country_key INT,
    FOREIGN KEY (country_key) REFERENCES Country(country_key)
)

CREATE TABLE Temperature (
    temperature_key INT PRIMARY KEY,
    description VARCHAR(100),
    range VARCHAR(50),
    measure DECIMAL(5,2)
)
```

```
CREATE TABLE Precipitation (
    precipitation_key INT PRIMARY KEY,
    description VARCHAR(100),
    range VARCHAR(50),
    measure DECIMAL(5,2)
)

CREATE TABLE WeatherFactTable (
    time_key INT,
    location_key INT,
    air_pressure_measure DECIMAL(5,2),
    temperature_key INT,
    precipitation_key INT,
    PRIMARY KEY (time_key, location_key),
    FOREIGN KEY (time_key) REFERENCES Time(time_key),
    FOREIGN KEY (location_key) REFERENCES Location(location_key),
    FOREIGN KEY (temperature_key) REFERENCES Temperature(temperature_key),
FOREIGN KEY (precipitation_key) REFERENCES Precipitation(precipitation_key) )

INSERT INTO Time VALUES (1, 2025, 'Winter', 1, 'Monday', TO_DATE('2025/01/28', 'yyyy/mm/dd'),
10)

INSERT INTO Time VALUES (2, 2025, 'Winter', 1, 'Tuesday', TO_DATE('2025/01/28', 'yyyy/mm/dd'),
12)

INSERT INTO Country VALUES (1, 'USA', 'California', 'Los Angeles', 'Downtown')
```

```
INSERT INTO Country VALUES (2, 'Canada', 'Ontario', 'Toronto', 'North York')

INSERT INTO Location VALUES (1, 34.0522, -118.2437, 'Urban', 1)

INSERT INTO Location VALUES (2, 43.7001, -79.4163, 'Urban', 2)

INSERT INTO Temperature VALUES (1, 'Cold', '-5 to 5°C', 2.5)

INSERT INTO Temperature VALUES (2, 'Mild', '10 to 20°C', 15.0)

INSERT INTO Precipitation VALUES (1, 'Light Rain', '0 to 5 mm', 3.2)

INSERT INTO Precipitation VALUES (2, 'Heavy Rain', '10 to 50 mm', 25.4)

INSERT INTO WeatherFactTable VALUES (1, 1, 20.25, 1, 1)

INSERT INTO WeatherFactTable VALUES (2, 2, 10.80, 2, 2)
SELECT * FROM WeatherFactTable



SELECT W.*, T.year, T.month, T.dates, L.latitude, L.longitude, C.country_name, Temp.description AS
Temperature, Prec.description AS Precipitation

FROM WeatherFactTable W

JOIN Time T ON W.time_key = T.time_key

JOIN Location L ON W.location_key = L.location_key

JOIN Country C ON L.country_key = C.country_key

JOIN Temperature Temp ON W.temperature_key = Temp.temperature_key

JOIN Precipitation Prec ON W.precipitation_key = Prec.precipitation_key

WHERE T.dates = TO_DATE('2025/01/28', 'yyyy/mm/dd')


SELECT L.location_key, C.city, AVG(W.air_pressure_measure) AS Avg_Pressure

FROM WeatherFactTable W

JOIN Location L ON W.location_key = L.location_key

JOIN Country C ON L.country_key = C.country_key

GROUP BY L.location_key, C.city


SELECT L.location_key, C.city, P.description, W.air_pressure_measure
```

FROM WeatherFactTable W

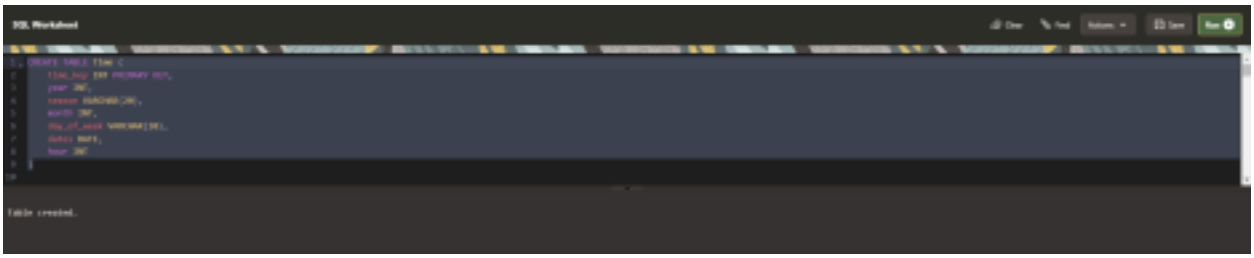
JOIN Location L ON W.location_key = L.location_key

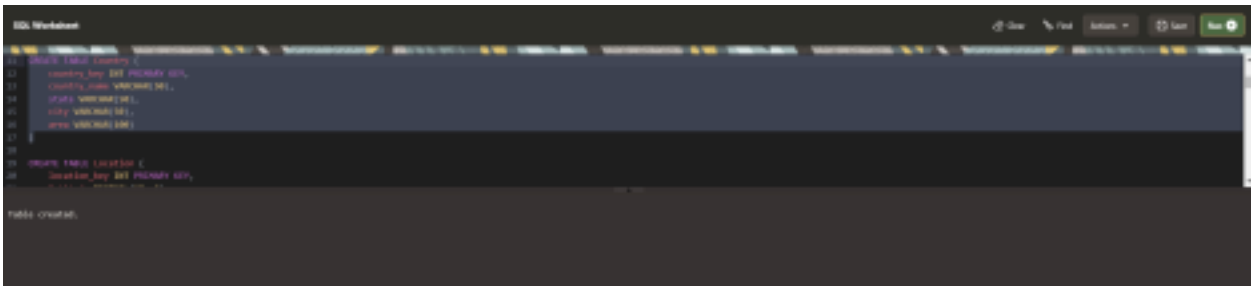
JOIN Country C ON L.country_key = C.country_key

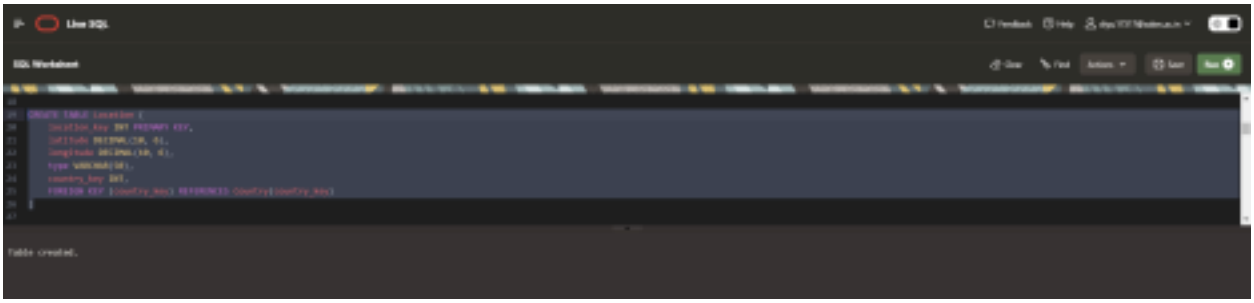
JOIN Precipitation P ON W.precipitation_key = P.precipitation_key

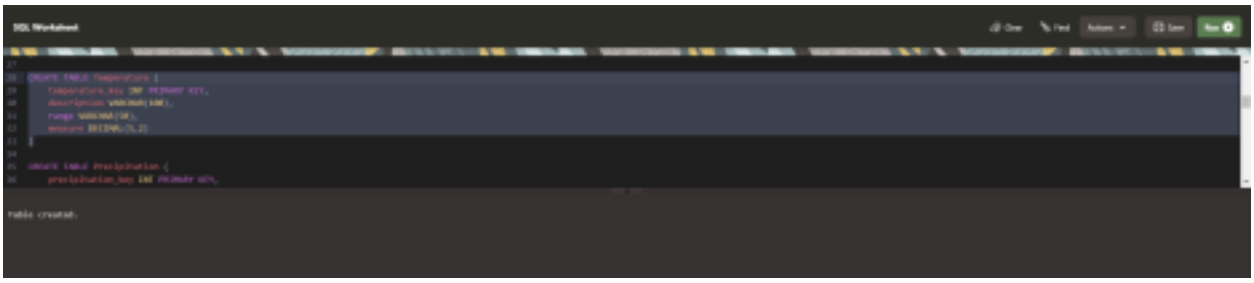
ORDER BY P.measure DESC

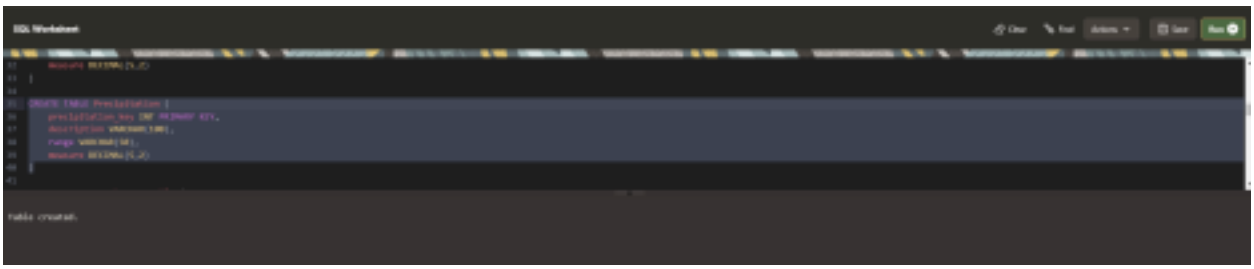
Output :

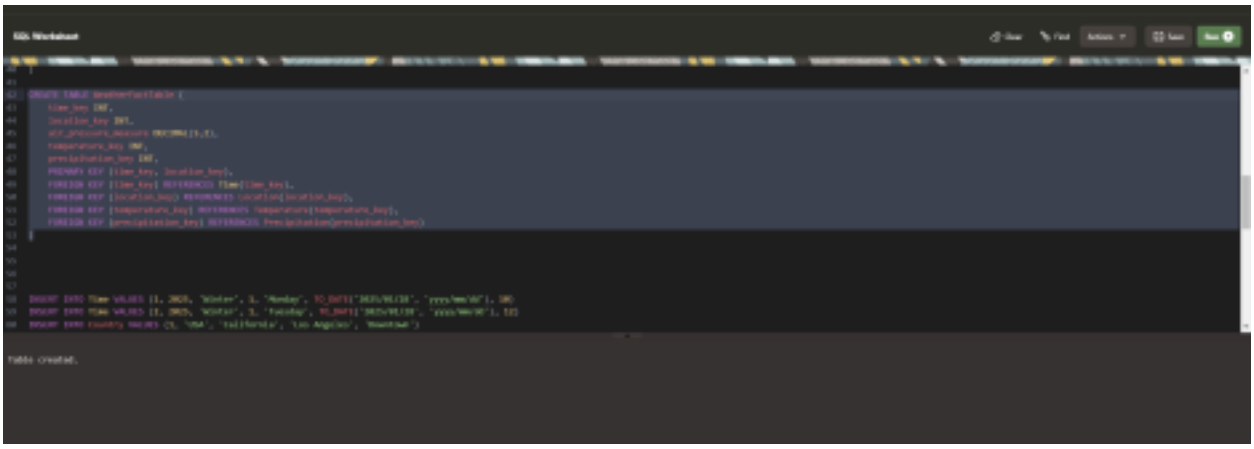






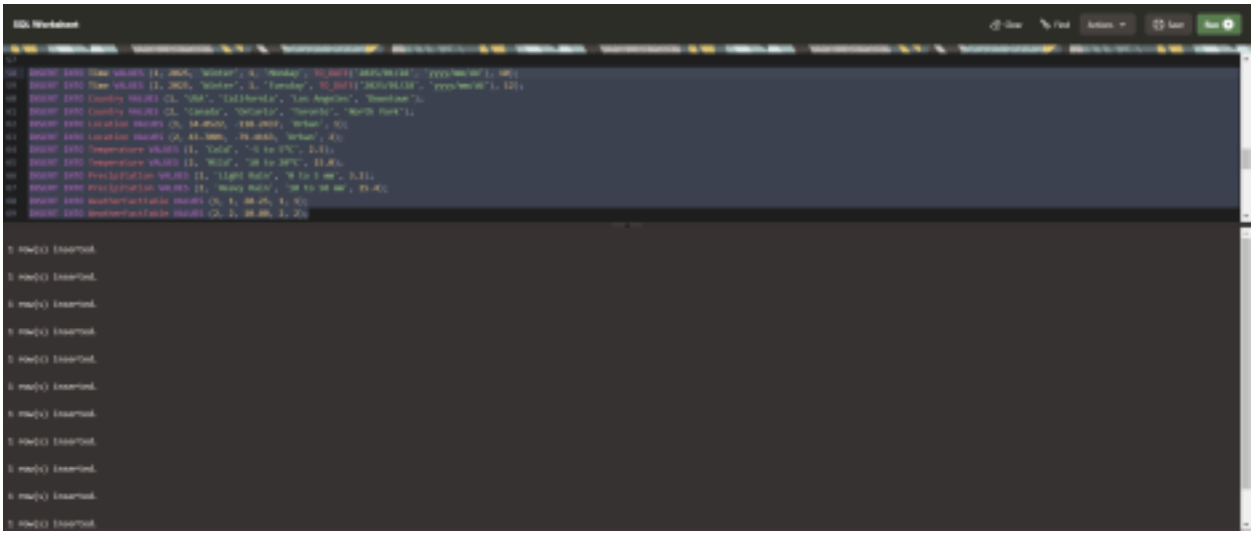






```
SQL Worksheet

CREATE TABLE weathervariable (
    time_key INT,
    location_key INT,
    air_pressure_measure DECIMAL(5,2),
    temperature_key INT,
    precipitation_key INT,
    PRIMARY KEY (time_key, location_key),
    FOREIGN KEY (time_key) REFERENCES Time(time_key),
    FOREIGN KEY (location_key) REFERENCES Location(location_key),
    FOREIGN KEY (temperature_key) REFERENCES Temperature(temperature_key),
    FOREIGN KEY (precipitation_key) REFERENCES Precipitation(precipitation_key)
);

INSERT INTO Time VALUES (1, 2025, 'Winter', 3, 'Monday', TO_DATE('2025/01/13', 'yyyy/mm/dd'), 20);
INSERT INTO Time VALUES (2, 2025, 'Winter', 3, 'Tuesday', TO_DATE('2025/01/14', 'yyyy/mm/dd'), 21);
INSERT INTO Country VALUES (1, 'USA', 'California', 'Los Angeles', 'Downtown');
```

Table created.



```
SQL Worksheet

INSERT INTO Time VALUES (1, 2025, 'Winter', 3, 'Monday', TO_DATE('2025/01/13', 'yyyy/mm/dd'), 20);
INSERT INTO Time VALUES (2, 2025, 'Winter', 3, 'Tuesday', TO_DATE('2025/01/14', 'yyyy/mm/dd'), 21);
INSERT INTO Country VALUES (1, 'USA', 'California', 'Los Angeles', 'Downtown');
INSERT INTO Country VALUES (2, 'Canada', 'Ontario', 'Toronto', 'North Park');
INSERT INTO Location VALUES (1, 34.0522, -118.2437, 'Urban', 1);
INSERT INTO Location VALUES (2, 43.6532, -79.3832, 'Urban', 2);
INSERT INTO Temperature VALUES (1, 'Cold', '-5 to 5°C', 2.5);
INSERT INTO Temperature VALUES (2, 'Mild', '10 to 20°C', 15.0);
INSERT INTO Precipitation VALUES (1, 'Light Rain', '0 to 1 mm', 0.5);
INSERT INTO Precipitation VALUES (2, 'Heavy Rain', '10 to 30 mm', 20.0);
INSERT INTO weathervariable VALUES (1, 1, 30.25, 1, 1);
INSERT INTO weathervariable VALUES (2, 2, 29.80, 2, 2);
```

1 row(s) inserted.

1 row(s) inserted.

1 row(s) inserted.

1 row(s) inserted.

1 row(s) inserted.

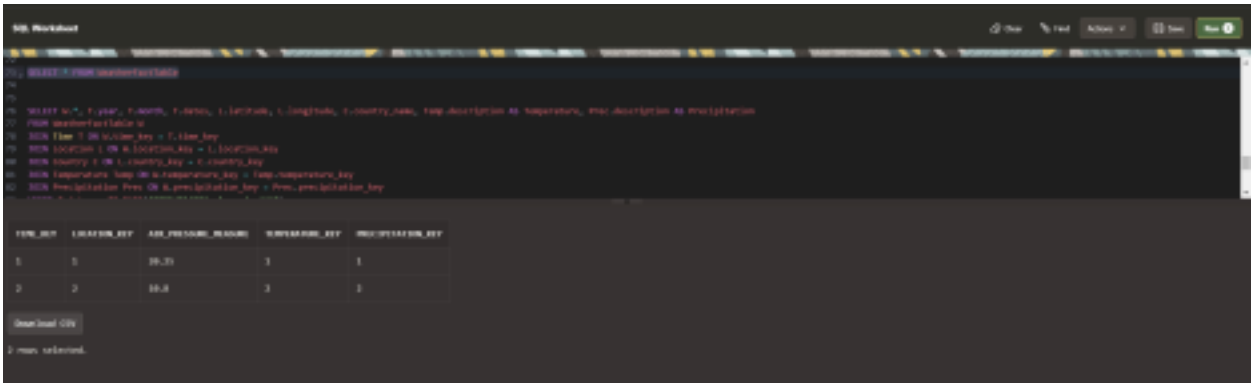
1 row(s) inserted.

1 row(s) inserted.

1 row(s) inserted.

1 row(s) inserted.

1 row(s) inserted.

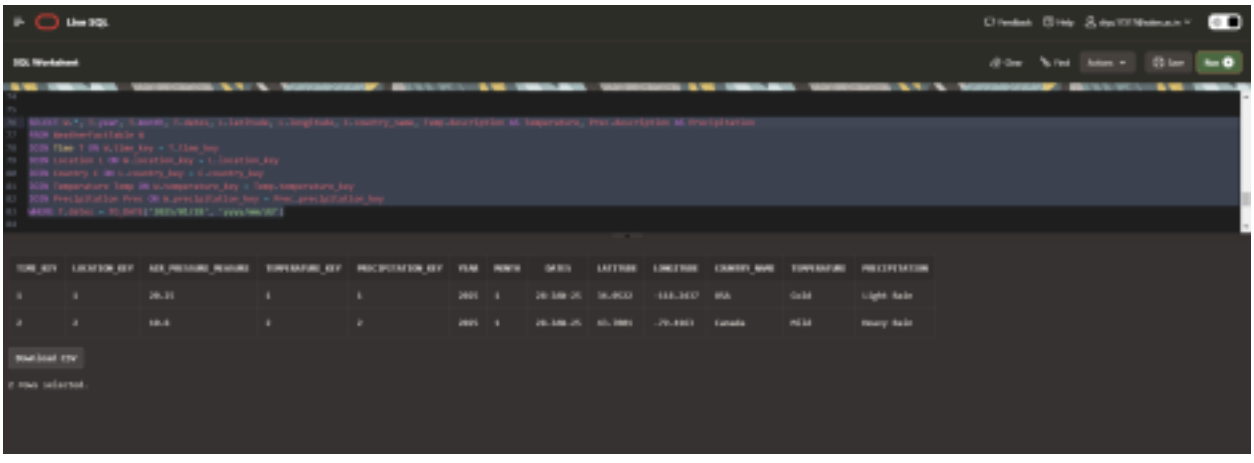


```
SQL Worksheet

1, SELECT * FROM weathervariable
```

TIME_KEY	LOCATION_KEY	AIR_PRESSURE_MEASURE	TEMPERATURE_KEY	PRECIPITATION_KEY
1	1	30.25	1	1
2	2	29.8	2	2

Download CSV

2 rows selected.

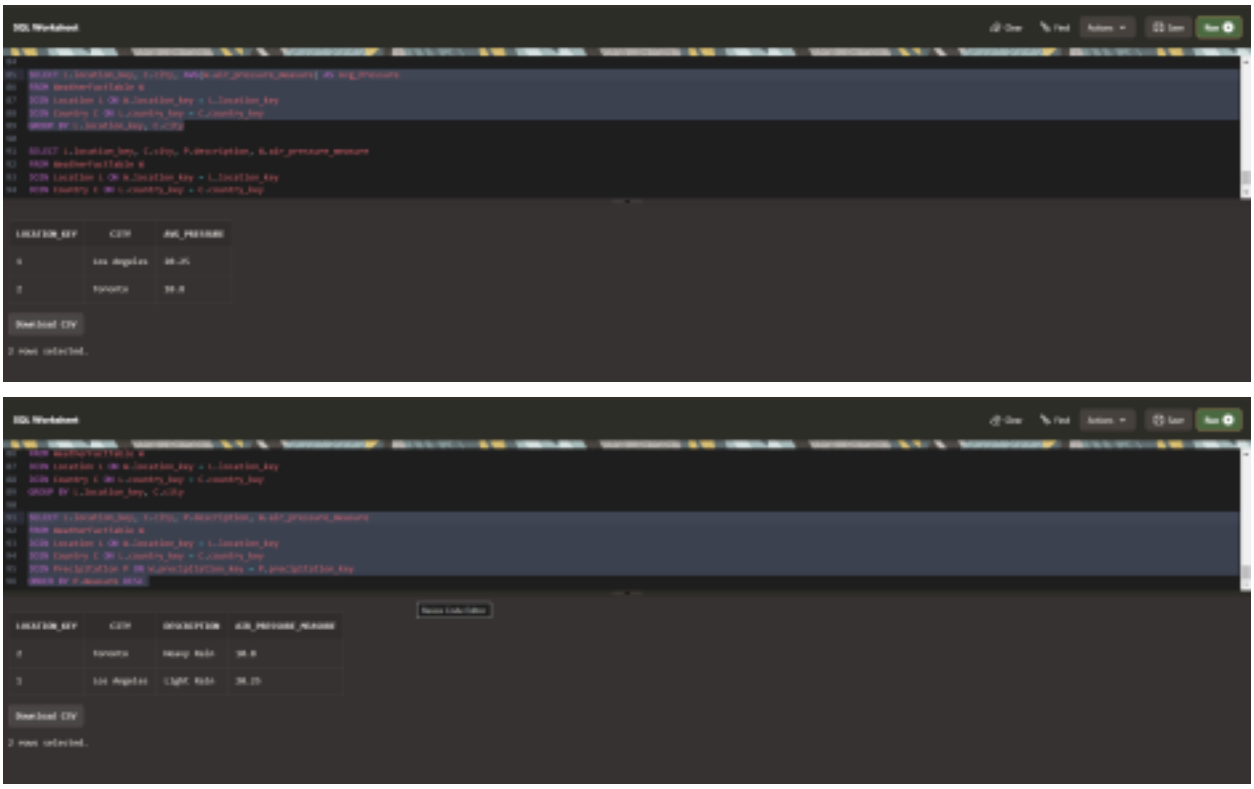


```
Live SQL

SQL Worksheet

SELECT w.*, t.year, t.month, t.dates, l.latitude, l.longitude, c.country_name, temp.description AS temperature, prec.description AS precipitation
FROM weathervariable w
JOIN Time t ON w.time_key = t.time_key
JOIN Location l ON w.location_key = l.location_key
JOIN Country c ON l.country_key = c.country_key
JOIN Temperature temp ON w.temperature_key = temp.temperature_key
JOIN Precipitation prec ON w.precipitation_key = prec.precipitation_key
WHERE t.dates = TO_DATE('2025/01/13', 'yyyy/mm/dd');
```

TIME_KEY	LOCATION_KEY	AIR_PRESSURE_MEASURE	TEMPERATURE_KEY	PRECIPITATION_KEY	YEAR	MONTH	DATES	LATITUDE	LONGITUDE	COUNTRY_NAME	TEMPERATURE	PRECIPITATION
1	1	30.25	1	1	2025	1	2025-01-13	34.0522	-118.2437	USA	Cold	Light Rain
2	2	29.8	2	2	2025	1	2025-01-13	43.6532	-79.3832	Canada	Mild	Heavy Rain

Download CSV

2 rows selected.



Review Question :

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 centrally located and is connected to multiple dimension tables through foreign key relationships, meaning each row in the fact table references one or more corresponding rows in the dimension tables to provide context and descriptive details about the measured data stored in the fact table; essentially, the fact table is the "many" side of a one-to-many relationship with each dimension table.

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

Ans: The primary difference between a star schema and a snowflake schema is that a star schema uses denormalized dimension tables, leading to faster queries but potential data redundancy, while a snowflake schema normalizes dimension tables by breaking them down into multiple related tables, reducing redundancy but potentially increasing query complexity due to more joins required.

GITHUB LINK : https://github.com/asmi-04/DWM-ASMI-25.git