LAB – 1: DATA PREPARATION FOR STAR SCHEMA

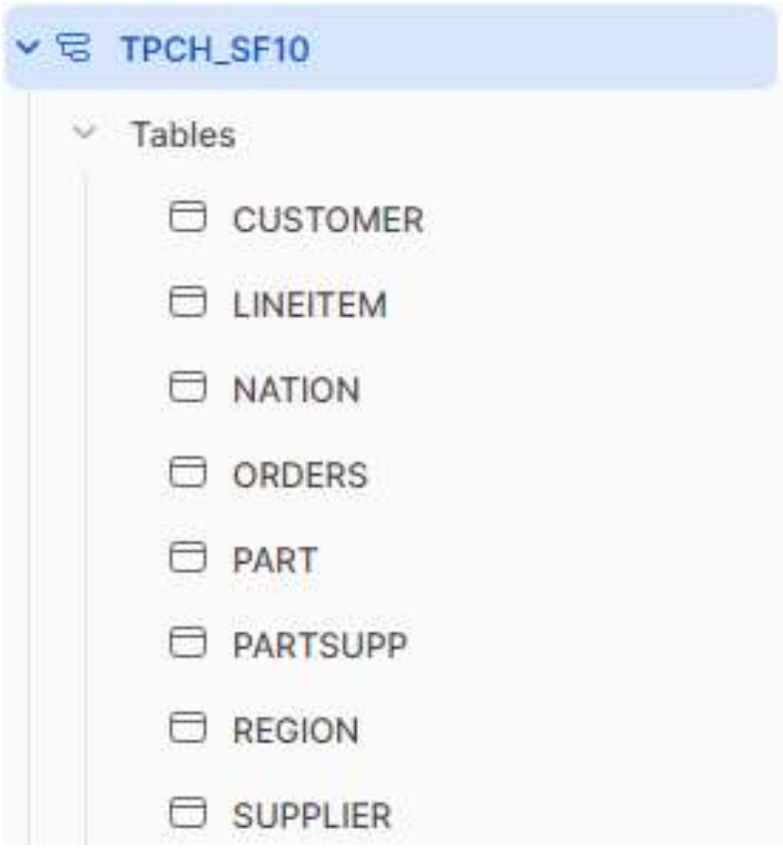
DATA WAREHOUSE:

- A data warehouse is a system used for reporting and data analysis, and is considered a core component of business intelligence.
- It is a central repository of integrated data from one or more disparate sources, and is designed to enable and support business intelligence (BI) activities, especially analytics.
- Key characteristics of Data Warehouse are:
    1. Integration
    2. Subject-oriented
    3. Time-variant
    4. Non-volatile

LAB PROCEDURES:

1. **Create account in Snowflake.**
    a. Snowflake is a data platform that would harness the immense power of the cloud.
    b. Simply, it is a Data Warehouse.
    c. We created our account in this platform.

2. **Understand TPCH_SF10 Schema.**
    a. TPCH_SF10 Schema is a sample data provided by Snowflake.
    b. This schema contains 8 tables.
    c. The table is provided below:



3. **Write SQL statements for the following:**

   a. **What is the total number of orders received in "1995-9-16"?**

   SELECT COUNT(*) AS TOTAL_ORDER FROM ORDERS WHERE O_ORDERDATE = '1995-09-16';

	...	TOTAL_ORDER
1		6223

   b. **What is the total number of orders received in September 1995?**

   SELECT COUNT(*) AS TOTAL_ORDER FROM ORDERS WHERE MONTH(O_ORDERDATE) = '9' AND YEAR(O_ORDERDATE) = '1995';

	...	TOTAL_ORDER
1		187281

c. What is the total number of orders received on each day of September?

SELECT O_ORDERDATE, COUNT(*) FROM ORDERS

WHERE MONTH(O_ORDERDATE) = '9' AND YEAR(O_ORDERDATE) = '1995'

GROUP BY O_ORDERDATE;

	O_ORDERDATE	COUNT(*)
1	1995-09-01	6357
2	1995-09-02	6114
3	1995-09-03	6003
4	1995-09-04	6241
5	1995-09-05	6325
6	1995-09-06	6301
7	1995-09-07	6317
8	1995-09-08	6215
9	1995-09-09	6223
10	1995-09-10	6311
11	1995-09-11	6207
12	1995-09-12	6176
13	1995-09-13	6325
14	1995-09-14	6230
15	1995-09-15	6215
16	1995-09-16	6223
17	1995-09-17	6196
18	1995-09-18	6226
19	1995-09-19	6210
20	1995-09-20	6223
21	1995-09-21	6296
22	1995-09-22	6160
23	1995-09-23	6209
24	1995-09-24	6214
25	1995-09-25	6304
26	1995-09-26	6344
27	1995-09-27	6156
28	1995-09-28	6438
29	1995-09-29	6253
30	1995-09-30	6269

d. What is the total number of orders received in each month of 1995?

```
SELECT MONTH(O_ORDERDATE), COUNT(*) FROM ORDERS
WHERE YEAR(O_ORDERDATE) = '1995'
GROUP BY MONTH(O_ORDERDATE);
```

	MONTH(O_ORDERDATE)	COUNT(*)
1	10	193057
2	1	193902
3	5	193116
4	7	193607
5	4	187016
6	9	187281
7	8	193223
8	6	186946
9	12	193207
10	2	173991
11	3	193719
12	11	186510

e. What is the total number of orders received in all years?

```
SELECT YEAR(O_ORDERDATE), COUNT(*) FROM ORDERS
GROUP BY YEAR(O_ORDERDATE);
```

	YEAR(O_ORDERDATE)	COUNT(*)
1	1992	2281205
2	1993	2276638
3	1994	2275919
4	1995	2275575
5	1996	2281938
6	1997	2275511
7	1998	1333214

f. What is the total number of orders received in all years?

```
SELECT COUNT(*) FROM ORDERS
JOIN CUSTOMER ON ORDERS.O_CUSTKEY =
CUSTOMER.C_CUSTKEY
JOIN NATION ON CUSTOMER.C_NATIONKEY =
NATION.N_NATIONKEY
WHERE NATION.N_NAME = 'BRAZIL';
```

	...	COUNT(*)
1		600955

CONCLUSION:

In this lab, we made a Fact table of ORDERS and for Dimensions we used Time and Location.