

Snowflake Full Notes (Data Analyst + Engineer)

1. What is Snowflake?

Snowflake is a cloud-based data warehouse used for storing and analyzing data.

2. Architecture

- Storage Layer
- Compute Layer (Virtual Warehouse)
- Cloud Services

3. Create Database & Table

```
CREATE DATABASE sales_db;  
USE DATABASE sales_db;
```

```
CREATE TABLE orders(  
    order_id INT,  
    customer STRING,  
    amount INT  
) ;
```

4. Insert Data

```
INSERT INTO orders VALUES  
(1,'Aman',500),  
(2,'Ravi',800);
```

5. Select Query

```
SELECT * FROM orders;
```

6. Aggregations

```
SELECT customer, SUM(amount)  
FROM orders  
GROUP BY customer;
```

7. Virtual Warehouse

```
CREATE WAREHOUSE compute_wh  
WITH WAREHOUSE_SIZE = 'XSMALL'  
AUTO_SUSPEND = 60;
```

```
USE WAREHOUSE compute_wh;
```

8. Joins

```
SELECT o.order_id, c.city  
FROM orders o  
JOIN customers c  
ON o.customer = c.customer;
```

9. Window Functions

```
SELECT customer,  
SUM(amount) OVER(PARTITION BY customer)  
FROM orders;
```

10. Stage & File Load

```
CREATE STAGE my_stage;
```

```
PUT file://data.csv @my_stage;
```

```
COPY INTO orders
FROM @my_stage/data.csv
FILE_FORMAT = (TYPE = CSV);

11. Time Travel
SELECT * FROM orders AT(offset => -60*5);

12. Streams & Tasks
CREATE STREAM order_stream ON TABLE orders;

CREATE TASK daily_task
WAREHOUSE = compute_wh
SCHEDULE = '1 DAY'
AS
INSERT INTO archive SELECT * FROM orders;

13. Interview Questions
- Difference Snowflake vs Redshift
- What is Virtual Warehouse
- Micro-partition
- Clustering
- Fail-safe
- Time Travel

14. Performance Tips
- Use clustering keys
- Avoid SELECT *
- Use proper warehouse size
- Cache usage
```