Snowflake Interview Questions Part I

By
Janardhan Bandi

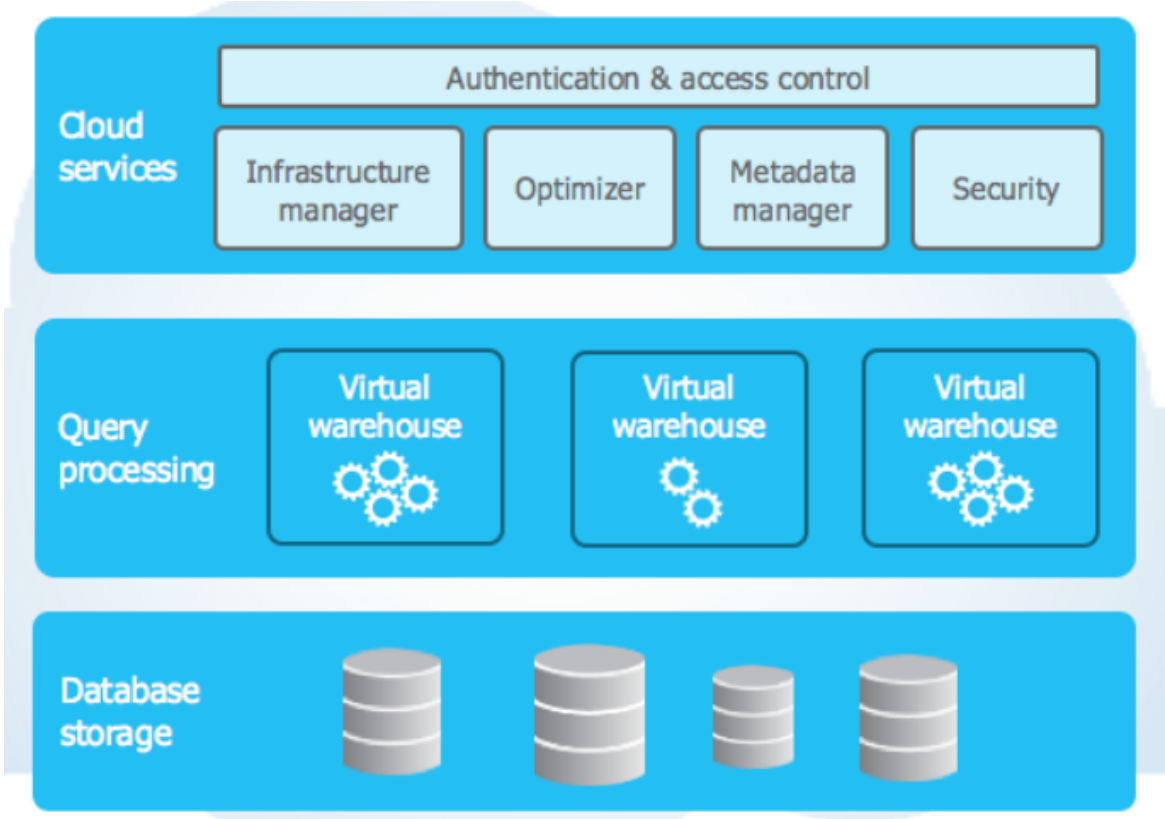
1. What is Snowflake and what kind of Database it is?

Ans: Snowflake is a cloud based Data Warehouse available as SaaS(Software as a Service), Snowflake enables Data storage And Data analytic solutions.

Snowflake doesn’t have their own infrastructure and currently it can be set up on Azure, AWS and GCP.

Snowflake is pure SQL database, it organizes the data into multiple micro partitions that are internally optimized and compressed. It uses a columnar format to store.

2. Explain the architecture of Snowflake.



→ The cloud services layer is a collection of services that coordinate activities across Snowflake.

→ Actual processing unit of snowflake, it processes the queries using Virtual Warehouses.

→ Data is stored in Columnar format in Micro Partitions.

3. What are the Advantages of Snowflake over traditional Databases?
(or) What are the new features available in Snowflake?

Ans: Lot of new features and Advantages.

Pay as you use

No infrastructure maintenance

Easy Data loading

Time Travel and Fail Safe

Zero copy Cloning

Easy Data Sharing

Tasks and Streams

Explained clearly in below video, link given in the Description.

https://youtu.be/LRdArUYm7jE

4. What are stages in Snowflake and write a query to create a Stage.

Ans: Snowflake Stages are the locations where data files are stored

There are 2 types of stages in Snowflake.

External Stages: If the data that needs to be loaded into Snowflake is stored in other cloud regions like AWS S3 or Azure or GCP then then we can use External Stages.

TEST_DB.PUBLIC

```
1  CREATE OR REPLACE EXT_STG_AZURE
2      STORAGE_INTEGRATION = azsf_jana_apr22
3      URL = 'azure://optumstagejana22.blob.core.windows.net/datalakejana'
4      FILE_FORMAT = FILE_FORMAT_AZURE;
```

Internal Stages: Stores the Data files internally, we can copy files to Internal stages by using PUT command from Snowsql.

5. Write the syntax of a Copy command to load a file into snowflake table.

Ans:

TEST_DB.PUBLIC

```
COPY INTO JANA_DB.STAGE_TBLS.STG_EMPL_DTLS
    FROM @EXT_STG_AZURE
    file_format= (type = csv field_delimiter=',' skip_header=1);
    FILES=('empl_dtls1.txt', 'empl_dtls2.txt');
    (or) PATTERN='.*empl_dtls.*';
```

6. Can you use Where clause in Copy command?

Ans: **No,** we can't use where clause in Copy command.

But We can do some transformations while loading the data by using copy.

- Select only required fields
- Can use functions like substring, cast etc.
- Can use Case statement

TEST_DB.PUBLIC

```
COPY INTO JANA_DB.STAGE_TBLS.STG_EMPL_DTLS
    FROM (select
            s.$1,
            s.$3,
            substring(s.$4,1,5)
            CASE WHEN CAST(s.$5 as int) < 0 THEN 'ABC' ELSE 'XYZ' END
          FROM @EXT_STG_AZURE s)
    file_format= (type = csv field_delimiter=',' skip_header=1);
    FILES=('empl_dtls1.txt');
```

7. How can you load a json file to Snowflake?
(or) how can you process and load semi-structured data?

Ans: We can store this semi-structure data into a table by using a data type called ‘Variant’. Then we can read this data from Variant, we can process it into rows and columns and load it into another table.

```
CREATE OR REPLACE TABLE JANA_DB.STAGE_TBLS.PETS_DATA_JSON_RAW
(raw_file variant);
```

Note: I have clearly explained how to process this data in below video and given all queries in the comments section of that video.

https://youtu.be/RSam2qmelQ8

8. Tell me some performance tuning techniques in Snowflake.

1. Use Cluster keys effectively
   - → Don’t define on small tables
   - → Define on filter columns
   - → Define on join keys
   - → Define on function based columns
2. Make use of Results cache for faster retrieval of Data.
3. Use materialized views Wisely
   - → on more frequently accessed tables
   - → on tables with less frequent data changes
4. And other common sql tuning techniques like
   - → Select only required columns
   - → Replace ‘OR’ with Union
   - → Union All is always better if we are sure there are no duplicates
   - → Try to avoid inequality with ‘OR’ condition
   - → Avoid unnecessary joins
   - → Avoid using ‘distinct’

9. How can you handle if the data coming from file is exceeds the length of a filed in the table

Ans: We can handle this by using (TRUNCATECOLUMNS = TRUE) in Copy command. If we don’t specify this property, Copy command will fail. By default this property is set to FALSE.

TEST_DB.PUBLIC

```
1  COPY INTO JANA_DB.STAGE_TBLS.STG_EMPL_DTLS
2      FROM @EXT_STG_AZURE
3      file_format= (type = csv field_delimiter=',' skip_header=1);
4      FILES=('empl_dtls1.txt')
5      TRUNCATECOLUMNS = TRUE;
```

10. How the Cost is calculated in Snowflake?

Ans: There are two types of costs in Snowflake.

→ Storage Cost
→ Compute Cost

Note: I have clearly explained how both of these costs calculated in below video.

https://youtu.be/0uJ2s0JS7SI

11. What is a clustering in Snowflake?

Ans: Clustering is basically grouping a bunch of values together so that it improves your query performance.

We define cluster keys on big tables, below are the best practices to define cluster keys.

- → Don’t define on small tables
- → Define on filter columns
- → Define on join keys
- → Define on function based columns

We can define cluster keys at the time creating tables, also we can add or modify cluster keys by using Alter Statement.

12. How many cluster keys is advised on single table?

Ans: Snowflake recommends a maximum of **3 or 4 columns** (or expressions) for clustering keys on tables. Adding more than 3-4 columns tends to increase costs more than benefits.

13. Write a query to retrieve data that was deleted from a table.

Ans:

TEST_DB.PUBLIC

```
SELECT * FROM TABLE_NAME
    before (timestamp => '2022-07-03 00:00:00.000'::timestamp);

SELECT * FROM TABLE_NAME
    at (offset => -60*60*24*2);

SELECT * FROM TABLE_NAME
    before (statement => '019b9ee5-0500-8473-0043-4d8300073062');
```

14. What are all the objects we can restore after delete or drop?

Ans:

We can restore the deleted data from any table based on the Time Travel Retention Period defined on the table. Based on the edition of snowflake, retention period can be 1 to 90 days.

We can un-drop the tables, schemas and databases that were dropped by mistake or wantedly.

15. Write a query to create a table with previous version of another table.

(or) Write a query for Clone with Time Travel.

Ans:

TEST_DB.PUBLIC

```
1  CREATE OR REPLACE TABLE DATABASE.SCHEMA.TABLE_PREV_VER
2      CLONE DATABASE.SCHEMA.TABLE_NAME at (OFFSET => -60*1.5);
```

16. What are all the objects that can be cloned in Snowflake?

Ans: Below objects can be cloned in Snowflake.

Data Containment Objects

- Databases
- Schemas
- Tables
- Streams

Data Configuration and Transformation Objects

- Stages
- File Formats
- Sequences
- Tasks

17. What are Secure Views in Snowflake?

Ans:

If we define the views with secure keyword then unauthorized users can’t see the definition of views using GET_DDL, SHOW VIEWS, DESC commands.

Normal views allows anyone to see the view definition.

TEST_DB.PUBLIC

```
1  CREATE OR REPLACE SECURE VIEW SALES_DATA
2  AS
3      SELECT ID, AMOUNT, TRAN_DATE FROM SALES WHERE STATE = 'TX';
```

18. What are all the objects that can be shared in Snowflake?

Ans: The following Snowflake database objects can be shared.

- Tables.
- External tables.
- Secure views.
- Secure materialized views.
- Secure UDFs.

19. What is a materialized view in Snowflake?

Ans: A materialized view is a pre-computed data set derived from a query specification and stored for later use.

Because the data is pre-computed, querying a materialized view is faster than executing a query against the base table of the view.

TEST_DB.PUBLIC

```
1  CREATE OR REPLACE MATERIALIZED VIEW ORDERS_MV
2  AS
3  SELECT
4  YEAR(O_ORDERDATE) AS YEAR,
5  MAX(O_COMMENT) AS MAX_COMMENT,
6  MIN(O_COMMENT) AS MIN_COMMENT,
7  MAX(O_CLERK) AS MAX_CLERK,
8  MIN(O_CLERK) AS MIN_CLERK
9  FROM ORDERS.TPCH_SF100.ORDERS
10 GROUP BY YEAR(O_ORDERDATE);
```

20. How to refresh the data in materialized views?

Ans: No need to manually refresh material views. After you create a materialized view, a background process automatically maintains the data in the materialized view.

To see the last time that a materialized view was refreshed, check the REFRESHED_ON and BEHIND_BY columns in the output of the command SHOW MATERIALIZED VIEWS.

To see the refresh history of any Materialized View

MATERIALIZED_VIEW_REFRESH_HISTORY(
[DATE_RANGE_START => <constant_expr>]
[, DATE_RANGE_END => <constant_expr>]
[, MATERIALIZED_VIEW_NAME => '<string>'])

21. What is difference btn Star schema and Snowflake schema

Ans: Star schema contains a fact table and the dimension tables those are denormalized but the snowflake schema contains a fact table and may be normalized dimension tables.

When it comes to performance in snowflake schema there will be more number of joins when compared to star schema because of its normalized dimension tables. So Start schema gives better performance.

22. What is the retention period in Business critical edition and how can you increase or reduce it?

Ans: The default retention period in Business critical edition is 90 days.

We can’t increase beyond 90 days but we can change it by using below Alter command.

ALTER TABLE TABLENAME SET DATA_RETENTION_TIME_IN_DAYS=30;

23. Snowflake is an ETL or ELT?

Ans: Snowflake supports both ETL and ELT.

We can transform and load the data at the same time we can load the data to snowflake and transform it.

24. Does snowflake supports indexes?

Ans: No we can’t define indexes on Snowflake tables instead we can use cluster keys on large tables for better performance.

25. How to grant select on all future tables in a schema and database level?

Ans:

1. Schema level:

use role accountadmin;

grant usage on database dbname to role role_name;

grant usage on schema dbname.schemaname to role role_name;

grant select on future tables in schema dbname.schemaname to role role_name;

2. Database Level:

use role accountadmin;

grant usage on database dbname to role role_name;

grant usage on future schemas in database dbname to role role_name;

grant select on future tables in database dbname to role role_name;

Thank You!

janardhan520@gmail.com

https://www.youtube.com/channel/UCNTGAQaxJMxZLS0GR1VHOKg