Snowflake Interview Questions Part II

By
Janardhan Bandi

26. What is snowpipe and write syntax for creating snowpipe.

Ans: Snowpipe is Snowflake's continuous data ingestion service. Snowpipe loads data within minutes after files are added to a stage and submitted for ingestion.

→ Snowpipe is serverless compute model.

→ Snowpipe provides a "pipeline" for loading fresh data in micro-batches as soon as it is available.

```
CREATE OR REPLACE PIPE PIPE_NAME
AUTO_INGEST = TRUE
AS
COPY INTO DBNAME.SCHEMANAME.TABLENAME  FROM @EXTERNAL_STAGE_NAME;

DESC PIPE PIPE_NAME;
```

27. What are the roles available in Snowflake?

Ans: Roles are the entities to which privileges on snowflake objects can be granted and revoked.

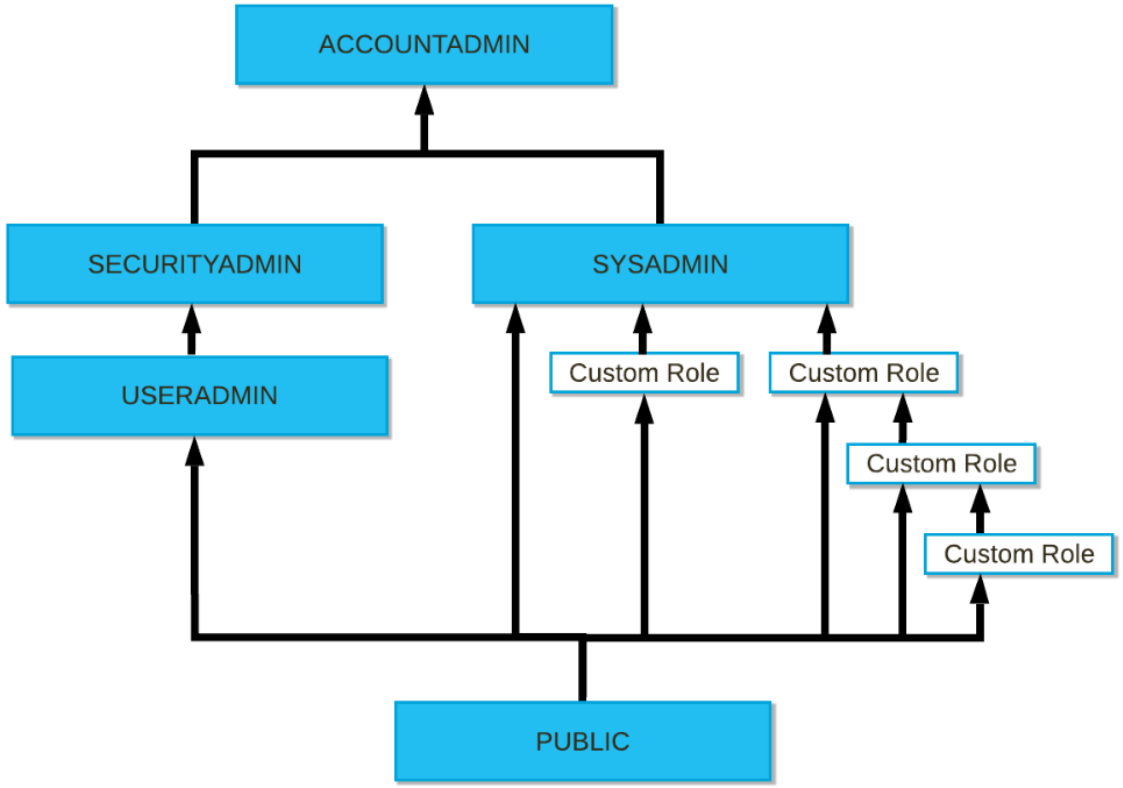
→ Roles are assigned to users to allow them to perform actions required for business functions in their organization.
→ A user can be assigned multiple roles.

Two types of roles in Snowflake.

1. System defined roles
2. Custom roles

27. What are the roles available in Snowflake? -- Continued

Ans:



https://docs.snowflake.com/en/user-guide/security-access-control-overview.html

28. What are the editions of Snowflake and which one you are using in your project?

Ans: There are 4 editions of Snowflake.

1. Standard
2. Enterprise
3. Business Critical
4. Virtual Private

→ These editions differ in terms of multi clusters, time travel, security and many other features.
→ Cost depends on Snowflake edition we choose

Standard - $2.7/Credit
Enterprise - $4/Credit
Business Critical - $5.4/Credit

Most of the organizations use either Enterprise or Business Critical

29. What is the virtual warehouse size you are using in your project? and how many clusters?

Ans:

→ We have to choose size based on the data size you are dealing with and the queries complexity. Size can be anything from XS to 6XL

→ Number of clusters depends on the number of concurrent jobs you are running. You can start with single cluster and you can scale out when your jobs increases.

→ Also VW size and number of clusters depends on the Environment, Dev, Test, Prod etc. based on the data and jobs you are dealing within that environment.

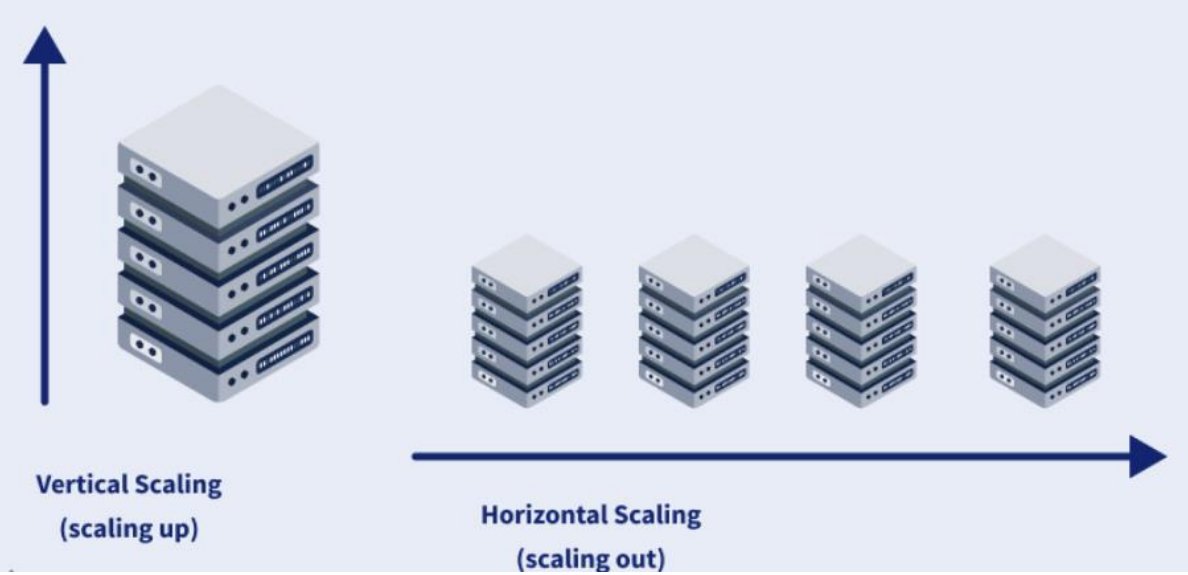
You can say we are using ‘M’ size with 4 clusters

30. What is Vertical scaling and Horizontal scaling?

Ans:

Vertical Scaling (Scale up) : Increasing the Virtual Warehouse Size to reduce the processing time and make you queries running faster.

Horizontal Scaling (Scale out): Increasing the number of clusters to avoid queries going into queues, you need to scale out when your customer base grows or when your parallel queries/jobs increases.



31. What are the diff table types in Snowflake?

Ans: Snowflake supports 3 types of tables.

1. Permanent Tables: For permanent tables, time travel is 0-90 days of retention period and 7 days fail safe period.

2. Transient Tables: Similar to permanent table but use where data protection and data recovery is not required, 0-1 day retention period and does not support fail safe.

3. Temporary Tables: Only active in that session and gets dropped once we close the session, 0-1 day retention period and does not support fail safe. Can be used in stored procedure for intermediate data storage.

Note: If you create any Database/Schema as Transient then all the tables under that Database/Schema are Transient by default.

32. What is the use of transient tables and temporary tables?

Ans:

Transient tables are specifically designed for transitory data that needs to be maintained beyond each session, but does not need the same level of data protection and recovery provided by permanent tables. You can create the stage tables and intermediate work tables as Transient.

Temporary tables are designed for storing non-permanent and transitory data, you can use them in stored procedures for intermediate data processing, these will get dropped when the execution of stored proc completed.

33. What is the retention period of Permanent, Transient and Temp tables?

Ans:
Permanent tables: 0-90 days retention period and 7 days fail safe period after retention period is completed. We can adjust retention period.

alter table tablename set data_retention_time_in_days=30;

Transient tables: 0-1 day retention period and does not support fail safe.

Transient tables: 0-1 day retention period and does not support fail safe.

Default retention period for all types of tables is 1.

34. Can you create a Temp table with the same name as a Permanent table?

Ans: Yes we can create, and all the queries will be fetching the data from Temporary table in that session.

35. How can you convert a Temp/Transient table into Permanent table?

Ans:

After creation, transient tables cannot be converted to any other table type.

After creation, temporary tables cannot be converted to any other table type.

36. What are different caches available in Snowflake?

Ans: Two types of caches available in Snowflake.

- **Result Cache:** It holds the results of every query executed in the past 24 hours. These are available across virtual warehouses, so query results returned to one user is available to any other user on the system who executes the same query, provided the underlying data has not changed.
- **Local Disk Cache:** It is used to cache data used by SQL queries. Whenever data is needed for a given query it's retrieved from the *Remote Disk* storage, and cached in SSD and memory.

37. What are streams in Snowflake? Write a query to create a Stream.

Ans: A Stream object records the delta of change data capture (CDC) information for a table such as a staging table including inserts and other data manipulation language (DML) changes like Update and Delete.

```
CREATE OR REPLACE STREAM STREAM_NAME ON TABLE TABLE_NAME;
```

This stream will record all the changes occurring to the table over time.

38. How the Stream tracks the changes occurring a table?

Ans: Every stream contains 2 fields, based on the values of these 2 fields we can identify the record is a Insert record or Update record or Delete record.

METADATA$ACTION – Insert/Delete
METADATA$ISUPDATE – True/False

METADATA$ACTION=INSERT and METADATA$ISUPDATE=FALSE ➔ Insert Record
METADATA$ACTION=DELETE and METADATA$ISUPDATE= FALSE ➔ Delete Record
METADATA$ACTION= INSERT and METADATA$ISUPDATE= TRUE ➔ Update Record

Note: A Streams records Update operation as a set of Delete(delete old record) and Insert(insert updated record).

Given an example merge query to process all changes Insert/Delete/Update in the description of this video.

39. What are the types of streams available in Snowflake?

Ans: Three types of stream.

Standard Streams: Supported for streams on tables, directory tables(external stages), or views. A standard stream tracks all DML changes to the source object, including inserts, updates, and deletes.

Append-Only Streams: Supported for streams on tables, directory tables, or views. An append-only stream tracks row inserts only. Update and delete operations are not recorded.

Insert-Only Streams: Supported for streams on external tables only. An insert-only stream tracks row inserts only; they do not record delete operations.

40. What is a Task and write syntax to create a Task?

Ans: A Task object is used schedule the execution of a SQL statement, including statements that call stored procedures.

→ We have to provide the Virtual Warehouse to execute the sql statement.
→ We can also use Cron scheduling in Tasks.

```
CREATE OR REPLACE TASK TASK_NAME
    WAREHOUSE = COMPUTE_WH
    SCHEDULE = '1 MINUTE'
    AS
    'SQL Statement';

CREATE OR REPLACE TASK TASK_NAME
    WAREHOUSE = COMPUTE_WH
    SCHEDULE = 'USING CRON 0 7 * * * UTC'
    AS
   'SQL Statement';
```

41. How can you implement column level security in Snowflake?

Ans: We can do this by using Dynamic data masking feature available in Snowflake. Snowflake supports using Dynamic Data Masking on columns of tables and views. We can partially mask the data or fully mask the data from un-authorized users.

First we have to create a masking policy and then we can apply this policy on the columns we wanted to implement security.

```
-- Creating masking policy
create or replace masking policy policy_name
    as (val varchar) returns varchar ->
            case
            when current_role() in ('SYSADMIN', 'ACCOUNTADMIN') then val
            else '#######'
            end;

-- Applying policy on a column
ALTER TABLE IF EXISTS table_name MODIFY COLUMN column_name
SET MASKING POLICY policy_name;
```

42. Write syntax to create a stored proc and an UDF.

Ans:

```
create or replace procedure proc_name()
        returns datatype
        language sql /* can write in sql, java, jave script, scala */
        AS
        $$
                sql statements;
        $$;



create function area_of_circle(radius float)
  returns float   /* can write in sql, java, jave script */
  as
  $$
      pi() * radius * radius
  $$
  ;
```

43. Best practices you followed in Snowflake?

Ans:

1. Choose appropriate table type
2. Define cluster keys on large table only and choose proper cluster keys
3. Reduce default retention period
4. Enable auto suspend and auto resume
5. Choose appropriate warehouse size, use scale-up and scale-out effectively
6. Understand storage and compute costs

44. What is the difference between Where and Having?

Ans:

Where clause is used to filter data while Having is used to filter the summary data after applying the grouping functions like count, sum, avg etc.

Can you use both where and having in single sql statement?
Yes, we can, below is one example.

Query to fetch list of depts that contains more than 10 Active employees.

```
SELECT DEPT_NO, COUNT(1) FROM EMP
WHERE EMP_STS='A'
GROUP BY DEPT_NO HAVING COUNT(1) > 10;
```

45. Query to delete duplicate records from a table?

Ans:

```
DELETE FROM TABLE_NAME
WHERE ROWID  NOT IN
      ( SELECT MAX(ROWID) FROM TABLE_NAME GROUP BY Key1, Key2) ;
```

If your database doesn’t support ROWID then use Rank approach or below temporary table approach.

1. Create a temp table with unique records of the actual table
2. Delete the data from actual table and insert data from temp table to actual table
3. Drop temp table

46. What is the diff between Union and Union All?

Ans:

UNION eliminates duplicate records after clubbing all records

UNION ALL will not eliminate the duplicate records.

When it comes to performance UNION ALL gives better performance because it doesn’t need to perform duplicate elimination task.

Note: To perform UNION and UNION ALL, list of columns and order of columns should be same in all statements or tables.

47. Below are two tables with one column in each
What is the number of records after each type of join?

Table A
1
2
1
3
null

Table B
1
2
2
4
null

Ans:

A inner join B - 4

A left outer join B - 6

A right outer join B - 6

A full outer join B - 8

48. Query to fetch Dept wise 3rd highest salary?

Ans:

```
SELECT DEPT_ID, SALARY FROM EMP
QUALIFY ROW_NUMBER() OVER(PARTITION BY DEPT_ID ORDER BY SALARY DESC) = 3;
```

If your database does not support Qualify, use below query.

```
SELECT DEPT_ID, SALARY FROM
(       SELECT DEPT_ID, SALARY,
        DENSE_RANK() OVER(PARTITION BY DEPT_ID ORDER BY SALARY DESC) RNK
        FROM EMP
) ABC
WHERE RNK=3;
```

49. What is the difference between Coalesce and Decode?

Ans:

COALESCE() : The COALESCE() function examines the first expression, if the first expression is not null, it returns that expression Otherwise, it does a COALESCE of the remaining expressions. In simple words COALESCE() function returns the first not-null expression in the list.

Syntax – COALESCE (expr_1, expr_2, ... expr_n)

DECODE(): The DECODE function decodes an expression in a way similar to the IF-THEN-ELSE logic used in various languages. The DECODE function decodes expression after comparing it to each search value. If the expression is the same as search, result is returned.

Syntax –

DECODE(col|expression, search1, result1 [, search2, result2,...,][, default])

50. What is diff between Primary Key, Unique Key and Surrogate Key?

Ans:

Unique key contains unique values in that field, it does not allow duplicates but allows nulls.

Primary key helps identifying the record uniquely in the table, basically it is Unique+Not Null, means it won’t allow duplicates and nulls.

Surragte key is a key with no business meaning and it is just a number (mostly we define Surragate Id as Numeric) assigned to each row in the table.

Note 1: A table can contain any number of unique keys but contains only one primary key.

Note 2: A primary key can be combination of multiple columns that define a unique record in the table, but surrogate key is a single column.

51. How to convert a timestamp to date in Snowflake?

Ans:

To extract Date from Timestamp:

select to_date('2022-05-22 00:00:00'); -- 2022-05-22

To extract Year, Month, Day from Timestamp:

select year(to_date('2022-05-22 00:00:00')); -- 2022

select month(to_date('2022-05-22 00:00:00')); -- 05

select day(to_date('2022-05-22 00:00:00')); -- 22

52. How to fetch only numeric characters from a string in Snowflake?

Ans:

```
SELECT  TRIM(REGEXP_REPLACE(string, '[^[:digit:]]', '')) AS Numeric_value

FROM

( SELECT ' Area code for employee ID 12345 is 6789.' AS string ) a;
```

53. Tell me some performance tuning techniques of SQL queries.

Ans:

1. Define proper Indexes
2. Define proper partitioning keys. (Cluster keys in Snowflake).
3. Select only required fields, don’t put SELECT * blindly.
4. Replace OR with UNION if possible.
5. Use UNION instead of UNION ALL if you are sure there are no duplicates.
6. Use CTEs instead of subqueries if you want to use same result set at multiple places in the query.
7. Use Inner Join instead of putting Cross Join + Where clause.
8. Collecting missing stats on table will help in Teradata.
9. Use materialized views for faster data retrieval and continuous data availability.
10. Look at query execution plan or Query profile to understand where is the bottleneck.

Some other frequently asked sql questions.

1. What is the diff between NVL and NVL2.
2. What are the DML commands in sql?
3. What is diff between varchar and nvarchar?
4. What is an index in database?
5. What is the diff between case and decode?
6. What are the types of slowly changing dimensions?
7. How a data warehouse is different from a database?
8. What is the diff between rank() and dense_rank()?
9. Write a query to find nth highest salary.
10. Write a query to fetch all the employee details with even number emp_id.

Thank You!

janardhan520@gmail.com

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