---- about SF----

snowflake (SaaS) -- no hardware to select, no software to install. run in a public cloud infrastructure.

Snowflake cannot be run on private cloud infrastructures (on-premises or hosted)

All components of snowflake's service run in a public cloud infrastructure

ongoing management,maintenance is handled by SF itself

- analytic db developed for the cloud ; it is not built on an existing database/ big data or hadoop thing

- faster, easier to use and flexible than traditional warehouse

- handles availability, optimization, authentication, configuration, resource management and data protection.

----- SF ARCHITECTURE-----

-- shared disk architectures use multiple nodes to access data stored on a single storage system (Symmetric Multiprocessing SMP)

-- shared nothing architectures store a portion of data in each node and in each cluster in the data warehouse (Massive parallel processing - MPP)

snowflake combines shared disk and shared nothing architectures and innovates new design and takes full advantage of

snowflake multi cluster shared data architectures consist of 3 separate layers:

-compute (Query Processing) : where queries are executed. Each virtual warehouse is an MPP compute cluster composed of multiple compute nodes

-storage (Database Storage): data is stored in the database in columnar format

- All data stored as an internal, optimized, compressed columnar format (micro-partitions - represents logical structure of table) - Can't access the data directly, only through Snowflake (SQL etc)

- services (Query Processing) (that manages it all)

----- Warehouse---

A warehouse provide the required resources: CPU, MEMORY, TEMPORARY STORAGE

TO perform select statements and DML operations, a warehouse must be running and in use session

When create a warehouse using create warehouse statement: X- small is the default size

When create a warehouse using web interface: X- large is the default size

Auto suspend and auto resumes apply only to the entire warehouse not to individual clusters in the warehouses.

Multi cluster warehouses enable you to scale compute resources to manage your user and query concurrency as they change

Scaling policy can be set when the warehouse is created or at any time afterwards ((through web or using SQL)

You can increase or decrease the number of clusters at any time even while the warehouse is running and executing statements (through web or using SQL)

You can monitor multi cluster warehouses through history or worksheet

Multi cluster datawarehouse enables numbers of users to connect to the same size warehouse

Multi warehouses are best utilized for scaling resources to improve concurrency for users/queries. They are not beneficial for improve performance of slow running queries or data loading.

Sf charges you for the number of clusters in Multi cluster WH

-------- SNOWFLAKE PRICING & REGIONS   -----

Snowflake pricing is based on usage , you pay for only storage and resources you use

storage costs are based on amount of compressed data stored in database

compute costs are based on warehouse size and how long the warehouse runs

Compute usage is billed on a per-second basis, with a minimum of 60 seconds.

In other words, Snowflake offers almost unlimited computing

and storage capabilities by utilizing cloud storage and computing.

snowflake edition:standard, premier, enterprise,business critical, virtual private

Note: Snowflake accounts do not support more than one region. You

will need to create a Snowflake account for each region

The storage and compute

resource costs are based on the amount of compressed data stored in the

database tables and whatever is needed for data recovery

Note: Credits are billed on a per-second basis only when virtual

warehouses are running. If a virtual warehouse is suspended, it will

not accrue charges.

Remember: Snowflake does not charge to load data into your

Snowflake environment from any external stage; however, your

cloud storage provider (Amazon S3 or Microsoft Azure) might charge

a separate egress fee if your data storage is located in a region or

network different from your Snowflake account.

Once a new release has been deployed, Snowflake does not move all accounts to the release at the same time

Day 1

Stage 1 (*early access*) for designated Enterprise accounts.

Day 1 or 2

Stage 2 (*regular access*) for all Standard Edition failsafe accounts.

Day 2

Stage 3 (*final*) for all Enterprise Edition and VPS accounts.

This staged approach only applies to new releases. For patch releases, all accounts are moved to the patch release on the same day.

Government Regions are only supported for Snowflake accounts on Business Critical Edition (or higher)

In addition, Snowflake does not move data between accounts, so any data in an account in a region remains in the region unless users explicitly choose to copy, move, or replicate the data

- On Demand: Usage-based pricing with no long-term licensing requirements.
- Capacity: Discounted pricing based on an up-front Capacity commitment.

Caches - Snowflake Caches different data to improve query performance and assist in reducing cost >Metadata Cache - Cloud Services layer - Improves compile time for queries against commonly used tables

>Result Cache - Cloud Services layer - Holds the query results - If Customers run the exact same query within 24 hours, result cache is used, and no warehouse is required to be active

>Local Disk Cache or Warehouse Cache - Storage Layer - Caches the data used by the SQL query in its local SSD and memory - This improves query performance if the same data was used (less time to fetch remotely) - Cache is deleted when the Warehouse is suspended

Which cache type gets purged regularly? ->warehouse cache

User A and User B can access one another's result sets from the Results Cache, as long as which of the following are true? (Choose two) ->same role, run same text or query within 24hrs

You are viewing the Query History table in the History area of the WebUI. You want to see if a query pulled data from long-term centralized storage. Where will you look and what will you look for?

→ Look in the Bytes Scanned column for a green bar.

N.B: each layer scales independently and includes built in redundancy

Snowflake encryption is AES 256

you only need to create databases, table, warehouses, load data and execute queries then snowflake handles the rest

Which of the following terms are associated with the Cloud Services Layer?

->query planning, query optimization and query compiling

Which of the following statements are true about Snowflake?

-> Storage can increase or decrease without any effect on virtual warehouse sizes.

-> Compute can be scaled up, down, out, or in and there is no effect on storage used.

-> Two Virtual Warehouses can access the same data at the same time without causing contention issues.

What attributes make Snowflake a true SaaS solution?

-> No hardware to purchase or configure.

-> No maintenance upgrades or patches to install.

-> Transparent releases don't require user intervention.

Which of the following terms describes Snowflake's Architecture?

-> Shared Data

Which statements are true about storage relationships?

-> Snowflake Tables are stored within Schemas.

-> Snowflake Schemas are stored within Databases

Snowflake data storage costs are calculated based on:

-> Compressed Size

-> Amount Stored - Daily Average

Snowflake data storage costs include which types of data?

-> Persistent data stored in permanent tables

-> Data retained to enable data recovery (time travel and fail-safe)

Snowflake compute costs depend on which of the following?

-> The amount of time warehouses have run.

-> The sizes of running warehouses.

What duties can a Snowflake professionals avoid that are common tasks for traditional on-premise database and IT staff?

->Maintaining metadata

->Maintaining statistics

->Maintaining the physical security of a server room (key cards, door locks, etc)

Which of the following are true of Multi-Cluster warehouses? Select all that apply.

Adds clusters automatically based on query activity

Scales down when query activity slows

The following factors affect data load rates:

Physical location of the Stage

Gzip compression efficiency

What is the recommended approached for making a variant column accessible in a BI tool?

A view

Which of the following are true about the variant data type in Snowflake? Select all that apply.

Optimized storage based on repeated elements
Can be queried using JSON path notation

True or false: Snowflake offers tools to extract data from source systems.
FALSE

Which of the following statements are true about Snowflake Data sharing? Select all that apply.

Consumers can query shared tables in the same query as their own tables
Data Sharing is integrated with role-based access controls

Select all of the answers that describe Snowflake micro-partitions. Micro-partitions:

Are the physical data files that comprise Snowflake's logical tables
→Enable horizontal and vertical query pruning

Table definitions
References to all of the micropartition files for that table
Tracking of all versions of the table data within the data retention window

-----catalog &objects---

you can only load the tables in SF -- not views or schemas

you can't modify a table, view or schema

you can create or drop a warehouse

- suspend or resume , configure or transfer ownership of a warehouse to a different role

- You can open/ close and save sheets

history page allows you to view and drill into details in last 14 days.

auto refresh refreshes every 10sec

VM is independent and it doesn't share resources with others

data is not directly accessible to customers; data is accessible via SQL

Note: Account activation must occur within 72 hours or you will

need to create another trial account.

Caution: If you log out of Snowflake, all running queries in the

worksheet will cease.

The history goes back 14 days -- history page two weeks

that are in a worksheet

Note: Larger virtual warehouses may not result in better

performance for data loading or query processing.

Note: If queries are queuing more than desired, another virtual

warehouse can be created, and queries can be manually redirected to

the new virtual warehouse. In addition, resizing a virtual warehouse

can enable limited scaling for query concurrency and queuing;

however, virtual warehouse resizing is primarily for improving query

performance.

Snowflake allows for custom column separators (a.k.a. column delimiters). In order to use a caret symbol as a custom column separator, what option must you first choose?

--> other

Snowflake does NOT allow for custom field enclosures like tildes. What three options are offered for field enclosures?

--> single quote, double quote, none

Why did we recommend making the description field 62 characters wide instead of 60?

--> Because Snowflake does not provide an "Other" option for field enclosures.

You can't change the name of a Datawarehouse after it is created

Warehouses are owned by roles not individual users

To change the warehouse that will be used to run a SQL command within a specific worksheet (for example, changing the worksheet so that it uses SMALL_WH), what two options are available?

-->Run a SQL command like "USE WAREHOUSE SMALL_WH;"

-->Update the Warehouse field in the Context Menu located above the worksheet.

snowflake has three panes: navigation tree, sql entry pane, and results/preview pane

Not all Snowflake Editions have Elastic Data Warehousing. Check all Snowflake editions that have Elastic Data Warehousing enabled.

--> enterprise edition, business critical, virtual private snowflake

Scaling a warehouse DOWN will decrease the number of servers. (e.g. Medium to Small)

Scaling a warehouse IN will decrease the number of clusters. (e.g. Max to Min)

Scaling a warehouse OUT will increase the number of servers. (e.g. Min to Max)

Scaling a warehouse OUT will increase the number of clusters. (e.g. Min to Max)

Data Marts can improve efficiency but can lead to inaccuracies due to replication.

Snowflake Warehouses have access to all of a company's data at all times.

Both Data Marts and Snowflake Warehouses can be assigned to a business' departments for their specific needs.

True

False

Scaling up is a manual process. (e.g. Small to Medium)   TRUE

Scaling down is an automated process. False

Snapping back is a manual process (e.g. Max clusters to min clusters) false

Scaling out is an automated process (e.g. Min clusters to max clusters) true

When configuring a Warehouse using a Snowflake edition that has Elastic Data Warehousing enabled, what facets or components will you need to configure that are not needed in accounts where Elastic Data Warehousing is not enabled. (Choose two).

--> minimum and maximum clusters, scaling policy

Snowflake Stages can be defined as either External or Internal.

Note: The date slider has a minimum/maximum range of 8 hours to

14 days. Load monitoring data is not available previous to 14 days.

There are four types of query statuses: Running, Queued, Queued

Provisioning, and Queued Repairing.

• Running: Queries that were actively running during

the interval. These queries may have started before ordering the interval.

• Queued: Queries that are in wait status. The wait could be because of the warehouse load being maxed out and therefore, would need to wait for running queries to complete processing.

• Queued Provisioning: Queries that are in wait status because the warehouse is provisioning, usually after a warehouse resumes.

• Queued Repairing: Queries that are in wait status while the warehouse is repaired. While rare, this occursonly for a few minutes.

multicluster virtual warehouse :

supported files are csv, json/xml, parquet, Avro, ORC

Note: Splitting larger data files allows the load to scale linearly. Using a larger warehouse (X-Large, 2X-Large, etc.) will consume more credits and may not result in any performance increase.

Note: VARIANT "null" values (not to be confused with SQL NULL values) are not loaded to the table. To avoid this, extract

semi structured data elements containing "null" values into relational columns before loading them, Alternatively, if the "null" values in

your files indicate missing values and have no other special meaning, Snowflake recommends setting the file format option STRIP_NULL_

VALUES to TRUE when loading the semi structured data files.

Note: S3 transmits a directory list with each COPY statement used by Snowflake, so reducing the number of files in each directory

improves the performance of your COPY statements.

-Listing specific files to load from a stage is generally the fastest.

i.e:copy into sample_table from @%sample_data/data1/ files=

('sample_file1.csv', 'sample_file2.csv', 'sample_file3.csv')

ie with pattern: copy into sample_table from @%sample_data/data1/

pattern='.*sample_file[^0-9{1,3}$$].csv';

Note: Querying is primarily for performing simple queries during

the data loading only and is not intended to replace querying already

loaded tables.

If a file is large, then it should be loaded using SnowSQL or

Snowpipe (if < 50MB then using the web user is fine)

Snowpipe is an autoscaling Snowflake cloud service that provides continuously loaded data into the Snowflake data warehouse from internal and external stages

Caution: The option auto_ingest is not available unless it is explicitly enabled on your Snowflake account. Please contact

Snowflake Supports for the enable options in your Snowflake account.

Credits are billed per second, with a 60-second minimum.

Note: Make sure that data is in a compressed format in the Snowflake staging area. Another consideration is to use external

storage options like Amazon S3 where you can set the data lifecycle policy and archive cold data.

Snowflake charges a fee for unloading data into S3 or Blog Storage within the same region or across regions

Note :Snowflake won't charge you for loading data from external storage

Note: A table with clustering keys defined is considered to be clustered. Clustering keys aren't important for all tables. Whether

to use clustering depends on the size of a table and the query performance, and it is most suitable for multiterabyte tables.

Note: Materialized views are designed to improve query performance for workloads composed of common, repeated query patterns. However,

materializing intermediate results incurs additional costs. As such, before creating any materialized views, you should consider whether the

costs are offset by the savings from reusing these results frequently.

-- Security---

Every securable object is owned by a single role, which is typically the role used to create the object. ownership can be transferred from one role to another

A user can be assigned multiple roles

Roles can be also granted to other roles, creating a hierarchy of roles

Sysadmin – create warehouse and databases
securityadmin- manages monitor resources can create roles and users and custom roles

not even the ACCOUNTADMIN role can modify or drop objects created by a custom role.

Privileges granted to roles at a lower level are inherited by roles at a higher level

If the stage is an internal (i.e. Snowflake) stage (option B), data files are automatically encrypted by the client on the local machine prior to being transmitted to the internal stage

Account and table master keys are automatically rotated by Snowflake when they are more than 30 days old

While key rotation ensures that a key is transferred from its active state (originator usage) to a retired state (recipient usage), rekeying ensures that a key is transferred from its retired state to being destroyed.

Time Travel and Fail-safe retention periods are not affected by rekeying

Only account administrators and security administrators (i.e. users with the ACCOUNTADMIN or SECURITYADMIN role) can create, alter, or drop network policies.

When creating or editing network policies using the web interface, enclosing Network Policy Properties in single quotes is not required; however, when using SQL, network policy properties must be enclosed in single quotes

Snowflake automatically blocks all IP addresses not included in the allowed list.

Snowflake encrypts all the data automatically, including data at rest and in transit.

Snowflake provides multifactor authentication and performs federated authentication

Note Discretionary access control (DAC) is when each object has

an owner, who can in turn grant access to that object. Role-based

access control (RBAC) is when access privileges are assigned to

roles, which are in turn assigned to users.

Which of these are Snowflake table typesi? temporary,transient,external,permanent

view types: materialized,standard,secure

Materialized views are designed to improve query performance for workloads composed of common, repeated query patterns

Materialized views are particularly useful when:

- Query results contain a small number of rows and/or columns relative to the base table (the table on which the view is defined).
- Query results contain results that require significant processing, including:
  - Analysis of semi-structured data.
  - Aggregates that take a long time to calculate.
- The query is on an external table (i.e. data sets stored in files in an external stage), which might have slower performance compared to querying native database tables.
- The view's base table does not change frequently.

Snowflake does not allow standard DML (e.g. INSERT, UPDATE, DELETE) on materialized views. Snowflake does not allow users to truncate materialized views

As with non-materialized views, a materialized view does not automatically inherit the privileges of its base table. You should explicitly grant privileges on the materialized view to the roles that should use that view

INFORMATION_SCHEMA.VIEWS does not show materialized views. Materialized views are shown by INFORMATION_SCHEMA.TABLES.

If you clone a schema or a database that contains a materialized view, then the materialized view is cloned.

Defining a clustering key on a materialized view is supported and can increase performance in many situations. However, it also adds costs

What are the three Snowflake Stage types? user, table, named

Named stages come in two varieties, what are they? internal or external

Which type of view is most like a table? materialized

Which type of view has an extra layer of protection to hide the SQL code from unauthorized viewing? secure

In a Snowflake account named MX43210, you need to set a user's default namespace to a database called MYDB and the PUBLIC schema. Which of the following

 commands would you use? set default_namespace = mydb.public

Select all statements that are true about the Snowflake container hierarchy:

-Accounts contain databases which contain schemas.

-Schemas contain tables as well as views.

Fail-Safe is a seven-day history of data and is automatically available on which table types?

Permanent

Each Snowflake account comes with two shared databases. One is a set of sample data and the other contains Account Usage information. Check all true statements about these shared databases.

-SNOWFLAKE_SAMPLE_DATA contains several schemas from TPC (tpc.org)

-SNOWFLAKE contains a schema called ACCOUNT_USAGE

-ACCOUNT USAGE is a schema filled with secure views

-The data stored as part of fail-safe is part of storage costs charged to customers.

-Only a Snowflake employee can recover data from fail-safe storage.

Time travel is available for which table types?

-permanent, temporary, transient

Which types of stages are automatically available in Snowflake and do not need to be created or configured?

-user and tables

Which table type disappears after the close of the session and therefore has no fail-safe, and no time travel options after the close of the session?

-temporary

Which of the following object types are stored within schemas?

Stored Procedures, file formats, Stages, Sequences, User Defined Functions

Snowflake Ecosystem Tech Partners are classified into which of the following functional categories?

-Data Integration

-Business Intelligence

-SQL Editors

-Security & Governance

-Advanced Analytics

-Programmatic Interfaces

What functional category does Matillion fall into?

-Data Integration

What functional category does Looker fall into?

-Business Intelligence

Which of the following have drivers/connectors (or information about where to find them) available via Help->Downloads in the Snowflake WebUI?

-GO,Node.js,JDBC,Spark

What makes a Partner Connect Partner different from other partners?

-Can be connected to Snowflake using a streamlined wizard

-Can be connected from within the WebUI

-Includes a streamlined Partner Trial Account Signup

-Includes automated role, user and staging database set up

Is Snowflake HIPAA compliant?

-Yes

Which of the following Snowflake Editions encrypt all data transmitted over the network within a Virtual Private Cloud (VPC)?

-Business Critical

--------Data sharing------

Data providers can manage access to Snowflake data shares using:

Account mapping tables

The web interface does not currently support adding or removing external tables, secure materialized views, or secure UDFs to/from shares. All management of these objects in shares must be performed using SQL.

A new object created in a database in a share is **not** automatically available to consumers.(tables rows are available immediately to the consumers after the change but not objects)

To make the object available to consumers, you must use the GRANT <privilege> … TO SHARE command to explicitly add the object to the share.

the recreated object is treated as a new object and is, therefore, not accessible until the object has been explicitly granted the necessary privileges in the share.

Data can be shared across accounts but can't be cloned across accounts

A share can't be cloned in a consumer account but shared data can be copied into a table in the consumer's account

Data sharing from the Enterprise Sensitive Data edition (ESD) of Snowflake to non -ESD editions isn't enabled by default and must be enabled by Snowflake support

----- Query history-----

LEVEL UP: QUERY HISTORY

- Available from the ribbon item "History"
- Allows viewing of 14 days of query history
- Can filter by user, warehouse, session and many other attributes
- Can download previous result sets (within 24 hours)
- Can view other people's queries
- Cannot view other people's results

Auto refresh happens every 10sec

Anybody can view other people's queries. You don't need to have the accountadmin role

If you want to view query history older than 14 days, where can you go to view it? Choose one path and one "term" commonly used.

The Account Usage Share
SNOWFLAKE (Database) -> ACCOUNT_USAGE (Schema) -> QUERY_HISTORY (Secure View)

Which method for viewing query history provides more parameters/fields/attributes?

The QUERY_HISTORY secure view available from the Account Usage Share.

----Fail safe time travel and clone-----

You can't clone a transient table to a permanent table

You can clone entire databases, schemas, and tables.

Time travel cannot Restore tables, schemas, and databases that have been dropped after the retention period lapses

Time Travel cannot be disabled for an account; however, it can be disabled for individual databases, schemas, and tables by specifying DATA_RETENTION_TIME_IN_DAYS with a value of 0 for the object.

Clone

Databases, Schemas, Tables

```
CREATE [ OR REPLACE ] { DATABASE | SCHEMA | TABLE } [ IF NOT EXISTS ] <object_name>
  CLONE <source_object_name>
    [ { AT | BEFORE } ( { TIMESTAMP => <timestamp> | OFFSET => <time_difference> | STATEMENT => <id> } ) ]
  ...
```

Other Schema Objects

```
CREATE [ OR REPLACE ] { STAGE | FILE FORMAT | SEQUENCE } [ IF NOT EXISTS ] <object_name>
  CLONE <source_object_name>
  ...
```

- ▶ OWNERSHIP privilege on the source object
- ▶ Does not contribute to the overall data storage unless modified
- ▶ Does not include the load history of the source table
- ▶ Does not lock the object being cloned
- ▶ Does not automatically transfer the privileges

Clone - Stage

> Only individual external named stages; not internal named stages

> When cloning a database or schema:
> External named stages that were present in the source when the cloning operation started are cloned.
> When tables are cloned, their internal stages are also cloned.
> Internal named stages are **not** cloned.

> Regardless of how a stage was cloned, the clone does **not** include any of the files from the source. i.e. all cloned stages are empty.

When you clone a schema /database you are cloning every objects that belongs to them

Storage Costs for Time Travel and Failsafe

> Calculated By 24 hour period unit
> The individual table rows for DML (Delta)
> Full table for DROP/TRUNCATE

Table Type	Time Travel Retention Period (Days)	Fail-safe Period (Days)	Min , Max Historical Data Maintained (Days)
Permanent	0 or 1 (for Snowflake Standard Edition)	7	7 , 8
	0 to 90 (for Snowflake Enterprise Edition)	7	7 , 97
Transient	0 or 1	0	0 , 1
Temporary	0 or 1	0	0 , 1

----Data loading----

User stages cannot be altered or dropped.

User stage type is designed to store files that are staged and managed by a single user but can be loaded into multiple tables

Table stage type is designed to store files that are staged and managed by one or more users but only loaded into a single table. Table stages cannot be altered or dropped

To stage files to a table stage you must be the table owner (have the role with the OWNERSHIP privilege on the table).

Micro-partitioning is automatically performed on all Snowflake tables. Tables are transparently partitioned using the ordering of the data as it is inserted/loaded.