Contents

[Snowflake Overview and Architecture. 2](#_Toc131756324)

[Serverless features. 2](#_Toc131756325)

[Spooling 2](#_Toc131756326)

[SnowSQL 2](#_Toc131756327)

[Data storage cost calculation 3](#_Toc131756328)

[Compute cost: 3](#_Toc131756329)

[Snowflake construct 3](#_Toc131756330)

[Which are valid types of windows in Snowflake? 3](#_Toc131756331)

[External functions 4](#_Toc131756332)

[Warehouse 4](#_Toc131756333)

[**Failsafe** 5](#_Toc131756334)

[Bill and chargeability 5](#_Toc131756335)

[**Edition question** 5](#_Toc131756336)

[Share 6](#_Toc131756337)

[Data replication 7](#_Toc131756338)

[Clone. 7](#_Toc131756339)

[Limitation while cloning. 8](#_Toc131756340)

[Clustering. 8](#_Toc131756341)

[how to check clustering history. 8](#_Toc131756342)

[Enable clustering 8](#_Toc131756343)

[Disable clustering key 8](#_Toc131756344)

[Search optimization: 9](#_Toc131756345)

[Enable search optimization. 9](#_Toc131756346)

[Check search optimization cost. 9](#_Toc131756347)

[To check the credit of search optimization 9](#_Toc131756348)

[Masking policy 9](#_Toc131756349)

[Dynamic data masking 9](#_Toc131756350)

[External tokenization 9](#_Toc131756351)

[External stages. 10](#_Toc131756352)

[How can you integrate Snowflake with AWS S3? 10](#_Toc131756353)

[Loading data 10](#_Toc131756354)

[Unloading 10](#_Toc131756355)

[Which options can you specify when unloading data to an external stage? 11](#_Toc131756356)

[MFA 11](#_Toc131756357)

[Privileges 11](#_Toc131756358)

[SAML: Saml\_identify\_provider 12](#_Toc131756359)

[Resource monitor 13](#_Toc131756360)

[Consideration in resource monitor 13](#_Toc131756361)

[Credit calculation of resource monitor 14](#_Toc131756362)

# Snowflake Overview and Architecture.

Active data stores in table.

File store in snowflake location

Historical data maintain for time travel and fail safe.

## Serverless features.

Snow pipe

Data replication

M.V

Automatic clustering

Search optimization service

Spooling

writes out query result to file. In db terms spooling is equal to export of data.

SnowSQL

allows you to execute a set of queries from a script file. Which command can you use for this? ! source.

Connect and abort are two commands that cannot be used by snow sight.

User credential can be specified in the config file, in the shell when running snowsql with -u flag.

# Data storage cost calculation

Amount Stored - Daily Average

# Compute cost:

The amount of time warehouses have run.

The sizes of running warehouses.

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

Data Integration

Business Intelligence

SQL Editors

Security and Governance

Advanced Analytics

Programmatic Interfaces

# Snowflake construct

Which Snowflake construct can you use to group data by date values and perform aggregations based on that grouping? - datebucket

## Which are valid types of windows in Snowflake?

Unbounded-preceding.

Unbounded following

Query result cache can be used across different user and different worksheet.

# External functions

An *external function* calls code that is executed outside Snowflake.

The remotely executed code is known as a *remote service*

# Warehouse

1. Snowflake compute costs depend on which of the following?

The amount of time warehouses have run.

The size of running warehouses.

Virtual warehouse credits are charged per second of use.

1. **Scale up**--> resizing warehouse

**Scale out**--> adding additional cluster into multi warehouse

1. What are the resources that a virtual warehouse provides? Select all that apply.

Virtual warehouses provide CPU, Memory & temporary storage to process queries

Mutli-cluster warehouse concepts

1. When you set minimum and maximum cluster as same it is called maximized.

When you set minimum and maximum cluster differently, it’s called auto scale.

1. When a warehouse is resized, which queries make use of the new size?--> subsequent easy
2. compute layer, virtual warehouse layer, Query processing layer ---> All are same
3. What happens when a virtual warehouse is resized to a LARGER size?

The existing queries are not impacted and will continue executing to completion Queries in queue & future queries are executed on the new servers

1. Warehouse can be re-size at any time, even when it is running queries and running query will not stop.
2. Query complexity can be addressed by increasing the size of a virtual warehouse. (This is also referred to as Scaling Up).  
     
   Increased workload i.e more queries & more users accessing the system concurrently is addressed by adding more clusters, normally to a multi-cluster virtual warehouse. (This is also referred to as Scaling Out).
3. Virtual warehouses are independent of each other and don't share the CPU or memory or cache. They do operate on the shared data though.
4. If a virtual warehouse is suspended it is likely that the results stored in its cache will be cleared.
5. When a virtual warehouse is resized to a smaller size, the cache associated with the servers is dropped.
6. Compared to a single cluster virtual warehouse, a multi-cluster virtual warehouse supports which of the following 🡪 Resize,auto-resume,stop,auto-suspend.
7. What are the virtual warehouse sizing approach recommended by Snowflake?

Snowflake recommended approach is to experiment to arrive at the best size

1. Which minimum Snowflake license allows Mulitcluster data virtual warehouse capability?-enterpise
2. **If you have multiple virtual warehouses in your Snowflake system, they will each be accessing the same shared data.--yes**
3. **When a virtual warehouse is resized to a SMALLER size, the required servers are only removed when ?** The servers are removed only when they are not being used by a statement

# **Failsafe**

1. Failsafe retention for permanent is 7 days
2. For a transient and Temporary table Snowflake provides how many days of failsafe retention ? -0 days
3. What is failsafe?

After time travel expiration on table, data can be still retrieved up to max 7 day.

1. Data protected by Failsafe can be recovered by? snowflake support
2. Failsafe cannot be disable.
3. safe can be used only in failure scenarios for data protection and can be accessed only by the Snowflake support
4. All editions of snowflake provide the failsafe feature

# Bill and chargeability

Snow pipe uses a server less compute due to which the billing is based on the actual data processed, COPY command uses virtual warehouse resources therefore the billing is based on how long the virtual warehouse was active

# **Edition question**

Federated Authentication & MFA are available in all editions; therefore the minimum edition is Standard where these functionalities are available.

The staged release process through which Snowflake releases updates is first applied to early access account who have opted for early access, then standard & premier accounts and finally Enterprise & upwards

# Share

1. share can be done on below

Database, Schemas, Tables all types of tables except temporary , Secure views & Secure UDFs can be shared, external table can also share.

All editions of Snowflake provide customer dedicated virtual warehouses, therefore the lowest edition providing such functionality is Standard

1. Sharing with a non-snowflake user requires creation of a reader account as Snowflake allows sharing only between Snowflake account
2. For a consumer a shared database is read-only so they cannot create new objects or modify/append data but A user in consumer account can create their own database form the share if they have the ACCOUNTADMIN role OR the IMPORT SHARE privileges.
3. Only one database can be added to a share. However multiple schemas, tables and views can be added to a share.
4. A snowflake share can have multiple consumers added to it.
5. Ownership of share cannot granted to other.
6. A snowflake share can be defined without a consumer added to it. One or more consumers can be added to the share after wards.
7. Provider can share to many consumer and a consumer can take data from many providers(many to many)
8. Consumer then can also be a provider for another consumer.
9. Consumer are not billed for storage, only for compute they billed , actual data resides in provider account so provider billed for data storage.
10. A read only database created on consumer side from share.
11. Provider can create multiple data share out of one database or from more than one database.
12. Cross cloud and cross data sharing can be implemented through data replication.

# Data replication

An account on aws and east-us region want to share data to azure cloud on west region then

First a replication of database will be made, then secondary db can share data to azure cloud.

Replication is permitted on same organization.

To enable replication use below command on both provider and consumer account.

System$global\_Account\_Set\_parameter(account locator,’enable\_Account\_data\_base\_replication, true)

setup a replica of a database in a target Snowflake account?

.

Alter database provider\_database enable replication to accounts consumer\_Account\_region.locator.

On consumer account do below

Create database db\_name

As replica of region.locator.providerdabase;

Alter database db\_name refresh.

# Clone.

Files that are already copied from the stage to the source table can be loaded again into a table cloned from the source table. 🡪 YES

A clone can also be again cloned multiple time with no limitation.

A cloned database doesn't inherit the source permissions, but schemas, tables & views contained inside the cloned database will inherit the source permissions.

Cloning objects

DATABASE | SCHEMA | TABLE|STAGE | FILE FORMAT | SEQUENCE | STREAM | TASK

## Limitation while cloning.

Tables are cloned, which means the internal stage associated with each table is also cloned. ..

External tables cannot be cloned.

Stages 🡪 External named stages that were present in the source when the cloning operation started are cloned. Internal named stages are not cloned.

To clone a table, you need a SELECT privileges on the source table and for all other objects that can be cloned you need the USAGE privilege.

# User and role important concept.

1. First create user.
2. Create a role and assign to user.
3. Then whatever the access you want to give to that role give it through role.

Create user deb

Password=’abcd1234’

create role deb\_role

grant role check\_vw to user FROM\_SYS\_ADMIN;

grant usage on database pat\_mdm\_dev to role deb\_role

grant usage on schema pat\_mdm\_dev .schema1, to role deb\_role

grant select on all future views in schema PAT\_MDM\_DEV.CDW\_REPORTING to role

deb\_role

note: if you create role with any user apart from account admin it will be by default inherit by security admin role.

# Clustering.

A clustering key is a subset of columns in a table (or expressions on a table) that are explicitly designated to co-locate the data in the table in the same [micro-partitions](https://docs.snowflake.com/en/user-guide/tables-clustering-micropartitions). This is useful for very large tables where the ordering was not ideal (at the time the data was inserted/loaded) or extensive DML has caused the table’s natural clustering to degrade.

clustering depth of a table measure? The average depth of the overlapping micro-partitions

Helps with range lookup queries 🡪 clustering.

## how to check clustering history.

SELECT \* FROM TABLE (information\_schema.AUTOMATIC\_CLUSTERING\_HISTORY());

## Enable clustering

ALTER TABLE CLUSTERED\_TABLE CLUSTER BY(CLUSTERING\_KEY);

## Disable clustering key

ALTER TABLE t1 SUSPEND RECLUSTER;

Which query can you use to view the number of partitions scanned in a query clustering a table upon a column would save in Snowflake?

SELECT SYSTEM$CLUSTERING\_INFORMATION('ALL\_REVIEWS', '(PRODUCT\_CATEGORY)');

Disable caching:

ALTER SESSION SET USE\_CACHED\_RESULT = FALSE;

Which type of constraint will Snowflake enforce? -> Not null constraint.

# Search optimization:

It’s like creating hash index on each column in table.

Help with point lookup but not with range queries.

Search optimization does not work on range of data.

Search optimization requires an additional data structure.

## Enable search optimization.

ALTER TABLE table\_name ADD SEARCH OPTIMIZATION.

To see search optimization works on not, table should have large numbers of data and large set of distinct values.

Compute and storage cost will be applied.

## Check search optimization cost.

System$estimate\_search\_optimization\_costs(‘ALL reviews’);

## To check the credit of search optimization

Select \* from table (information schema. search\_optimization\_history(date\_range\_start=>dateadd(h,-7,current\_timestamp)

# Masking policy

Masking data at column level

# Dynamic data masking

# External tokenization

Row level policy :

Which performance optimizations can you perform on external tables in Snowflake? 🡪 External partitioning on a column

Stages.

Snowflake compress and encrypted file within stages.

In which objects does Snowflake store your configurations for loading data from files? - file format.

# External stages.

GCS copying data 🡪 It is possible to bypass external stages while copying data from GCS.

Azure 🡪 There are two types to connect azure

1. Storage account integration
2. SAS Token

Update and delete operation are not allowed on external table.

Which value will you take from your AWS account and copy into the specification of your Snowflake storage integration object, 🡪 The IAM User ARN

## How can you integrate Snowflake with AWS S3?

Using a Snowflake storage integration object

Using an Access Key ID and an Access Key Secret

## Loading data

Put command is used to move data into stage.

Copy command is used to put data in table.

## Unloading

Copy into command is used to unload the data from a table into stage.

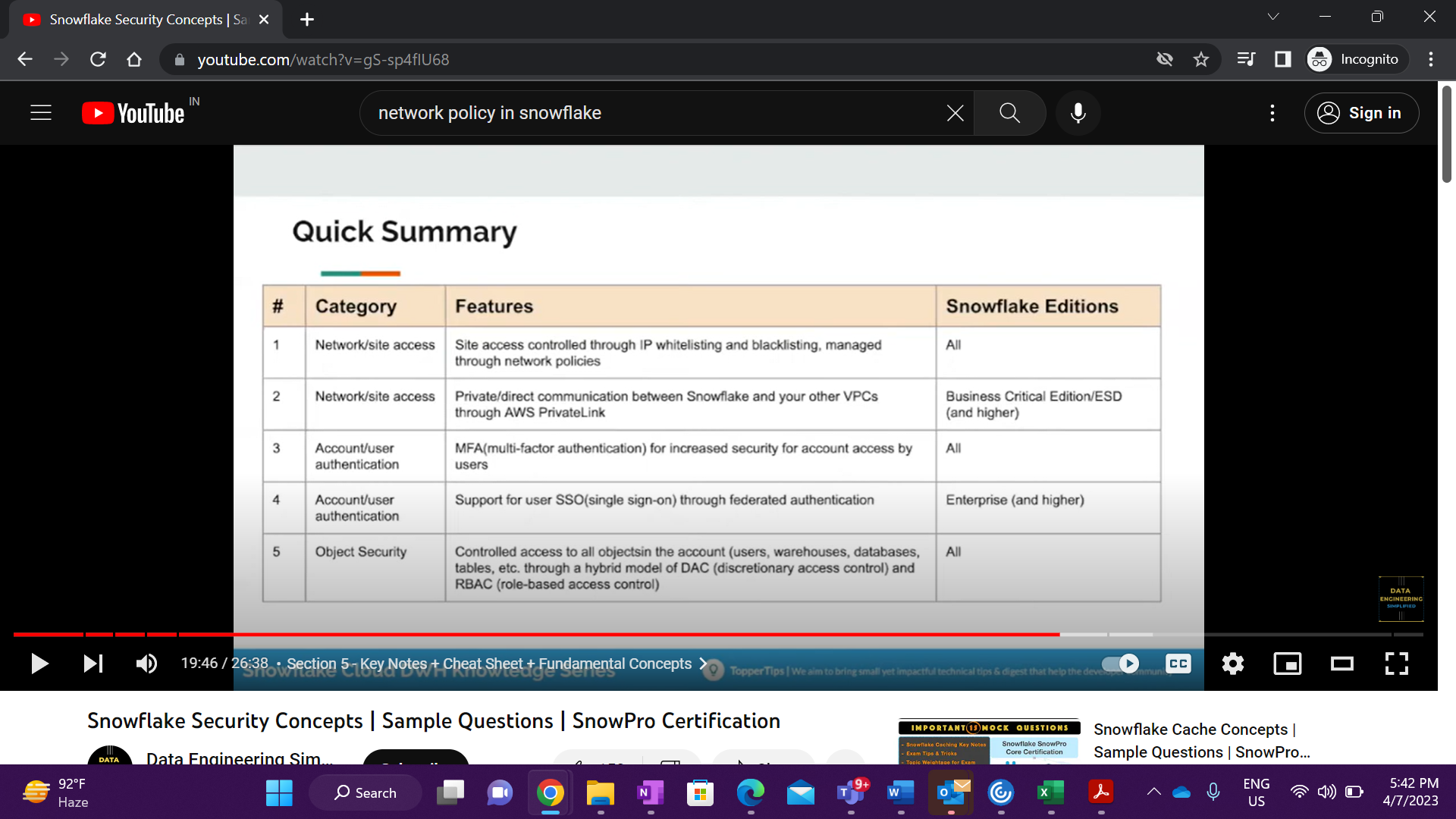
Get command is used to export table data into file system.

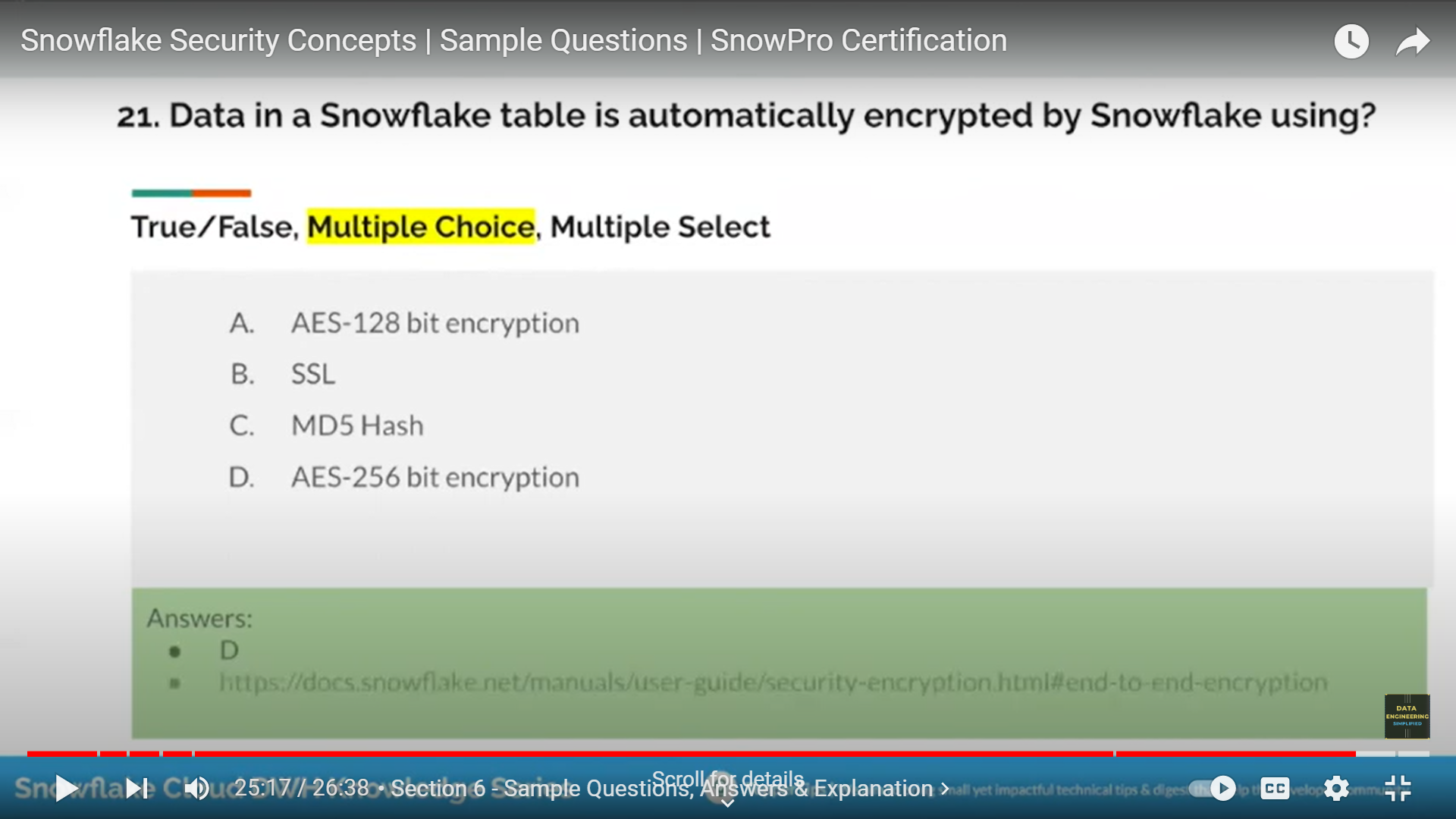
## Which options can you specify when unloading data to an external stage?

That the data should be unloaded into only one file= single=TRUE

That the data should overwrite the contents of the file being unloaded to.

# Network policy





# MFA

Duo application is required for MFA

**MINS\_TO\_BYPASS\_MFA 🡪 set a time by which user can bypass MFA and user can log in without MFA.**

**DISABLE\_MFA 🡪 Permanently disable MFA for user.**

# Privileges

Privileges are granted to roles, and roles are granted to users,

Import share: it is for data consumers.

|  |  |  |
| --- | --- | --- |
| MODIFY | Resource Monitor, Warehouse, Data Exchange Listing, Database, Schema | Grants the ability to change the settings or properties of an object (e.g. on a virtual warehouse, provides the ability to change the size of a virtual warehouse). |
| MONITOR | User, Resource Monitor, Warehouse, Database, Schema, Task, Failover Group, Replication Group | Grants the ability to see details within an object (e.g. queries and usage within a warehouse). |
| USAGE | Warehouse, Data Exchange Listing, Integration, Database, Schema, Stage (external only), File Format, Sequence, Stored Procedure, User-Defined Function, External Function | Grants the ability to execute a [USE <object>](https://docs.snowflake.com/en/sql-reference/sql/use.html) command on the object. Also grants the ability to execute a [SHOW <objects>](https://docs.snowflake.com/en/sql-reference/sql/show.html) command on the object.  Ex. Use database |

# SAML: Saml\_identify\_provider

it’s an protocol (account parameter) that’s use for communication b/w identity provider and snowflake.

To enable Sso use below

1. Alter account set sso\_login\_page=TRUE;
2. Need to create security integration object.

If you would like to integrate Snowflake with Okta for SSO, which account-level parameter must you configure?

SAML\_IDENTITY\_PROVIDER

What does key pair authentication allow you to do? It allows you to regulate Snowflake access via a combination of a private key and a public key

SELECT \*

FROM SNOWFLAKE.ORGANIZATION\_USAGE.STORAGE\_DAILY\_HISTORY

ORDER BY USAGE\_DATE DESC.—user admin role.

Account parameter setting.

D. ALTER ACCOUNT SET DATE\_OUTPUT\_FORMAT = 'DD/MM/YYYY';

Task history: Account admin role, any role that has global monitor execution privilege, task owner.

# Resource monitor

Send a notification when credit threshold reaches.

Suspend a warehouse at specific date and time, regardless of credit quota reached or not.

Monitor credit usage of warehouses with no interval end date and do a different suspend action at 80% and 90% thresholds.

Resource monitor usage and creation

Users with the ACCOUNTADMIN role only create resource monitors and can monitor it .

Users can be granted privileges to view and modify resource monitors.

Resource monitor properties.

Credit quota

Monitor level.

Schedule

Action

## Consideration in resource monitor

* An account-level resource monitor does not override resource monitor assignment for individual warehouses. If either the account resource monitor or the warehouse resource monitor reaches its defined threshold and a suspend action has been defined, the warehouse is suspended.
* An account-level resource monitor does not control credit usage by the Snowflake-provided Compute resources for serverless features (for example, Snowpipe, automatic reclustering, and materialized views). For a full list of features, see [Serverless Features](https://docs.snowflake.com/en/user-guide/cost-understanding-compute.html#label-serverless-features).
* A warehouse-level resource monitor can monitor, but cannot suspend, credit usage by Cloud Services. After a virtual warehouse is suspended, subsequent queries run against that warehouse can still result in additional cloud services costs. For more details about credit usage for cloud services, see [Cloud Service Credit Usage](https://docs.snowflake.com/en/user-guide/cost-understanding-compute.html#label-cloud-services-credit-usage).

## Credit calculation of resource monitor

The used credits for a resource monitor reflect the sum of credits consumed by all assigned warehouses within the specified interval, along with the cloud services used to support those warehouses during the same interval

By default threshold notification is not enabled.

Types of URL: in unstructured data.

Scoped

File

Presigned.

# SOME USER-Friendly command

## Parameters

Show parameters like ‘%.....%’ for DATABASE|SCHEMA|TABLE name.

# Data warehousing concept

SCD1🡪 keeps only latest and updated version data.

In context of MDM xref table is a table of type1.

SCD2🡪 where history of data get maintained, along with some flag to point current data

In context of MDM history xref table can be considered as SCD2 as if any records gets updated it maintains a history of it.

SCD 3🡪 history is getting maintained in column rather than present in separate row.

We should be knowing which dimension can be changed and according to it we can create column for example previous location and current location.

SCD 4🡪 combination of SCD 1 and 2

History is getting maintained in separate table for any update. If we assume xref and history xref than xref SCD1 data is getting stored and in history xref SCD 2 data is getting stored all together it’s a SCD4

SCD6 🡪 (1+2+3)

Where a history is getting maintained, there will be always a current record along with this dimension changes will get captured in column.

Fact: frequently changeable data like bill\_amount, medicines.

Dimension: Infrequently changed data , name, DOB.etc