Question 1: **Correct**
Snowflake is available on premise

- True
- False **(Correct)**

Explanation
Snowflake's data warehouse is a true SaaS offering. More specifically:

There is no hardware (virtual or physical) for you to select, install, configure, or manage.

There is no software for you to install, configure, or manage.

Ongoing maintenance, management, and tuning is handled by Snowflake.

Snowflake runs completely on cloud infrastructure. All components of Snowflake's service (other than an optional command line client), run in a public cloud infrastructure.

Snowflake uses virtual compute instances for its compute needs and a storage service for persistent storage of data. Snowflake cannot be run on private cloud infrastructures (on-premises or hosted).

Snowflake is not a packaged software offering that can be installed by a user. Snowflake manages all aspects of software installation and updates.

https://docs.snowflake.com/en/user-guide/intro-key-concepts.html

Question 2: **Correct**
Pick the true statement for snowflake architecture

- Shared nothing architecture
- Shared disk architecture
- Multi-Cluster Shared Data architecture **(Correct)**

Explanation
Snowflake's architecture is a hybrid of traditional shared-disk database architectures and shared-nothing database architectures. Similar to shared-disk architectures, Snowflake uses a central data repository for persisted data that is accessible from all

compute nodes in the data warehouse. But similar to shared-nothing architectures, Snowflake processes queries using MPP (massively parallel processing) compute clusters where each node in the cluster stores a portion of the entire data set locally. This approach offers the data management simplicity of a shared-disk architecture, but with the performance and scale-out benefits of a shared-nothing architecture.

https://docs.snowflake.com/en/user-guide/intro-key-concepts.html

Question 3: **Correct**
The three key layers of snowflake are

- Extraction, Ingestion, Load
- Database storage, Query Processing, Cloud Services **(Correct)**
- Database, Virtual Warehouse, Data Experience

Explanation
Snowflake's unique architecture consists of three key layers:

Database Storage

Query Processing

Cloud Services

https://docs.snowflake.com/en/user-guide/intro-key-concepts.html

Question 4: **Correct**
How is data loaded into snowflake?

- Snowflake loads the data in parquet format on the underlying cloud storage
- Snowflake loads the data in JSON format on the underlying cloud storage
- Snowflake reorganizes the data into its internal optimized, compressed, columnar format at stores on the underlying cloud storage **(Correct)**

Explanation
When data is loaded into Snowflake, Snowflake reorganizes that data into its internal optimized, compressed, columnar format. Snowflake stores this optimized data in cloud storage.

Snowflake manages all aspects of how this data is stored — the organization, file size, structure, compression, metadata, statistics, and other aspects of data storage are handled by Snowflake. The data objects stored by Snowflake are not directly visible nor accessible by customers; they are only accessible through SQL query operations run using Snowflake.

https://docs.snowflake.com/en/user-guide/intro-key-concepts.html

Question 5: **Correct**
Query processing in snowflake is done by...

- Snowflake processes queries using AWS EMR running spark
- Snowflake processes queries running spark on a EC2 instance
- Snowflake process queries using 'Virtual Warehouses'
  (Correct)

Explanation
Query execution is performed in the processing layer. Snowflake processes queries using “virtual warehouses”. Each virtual warehouse is an MPP compute cluster composed of multiple compute nodes allocated by Snowflake from a cloud provider.

Each virtual warehouse is an independent compute cluster that does not share compute resources with other virtual warehouses. As a result, each virtual warehouse has no impact on the performance of other virtual warehouses.

https://docs.snowflake.com/en/user-guide/intro-key-concepts.html

Question 6: **Correct**
Which of the below services are provided by cloud services

- [x] Metadata management
  (Correct)
- [x] Authentication
  (Correct)
- [x] Infrastructure management
  (Correct)
- [ ] Query execution

Explanation

The cloud services layer is a collection of services that coordinate activities across Snowflake. These services tie together all of the different components of Snowflake in order to process user requests, from login to query dispatch. The cloud services layer also runs on compute instances provisioned by Snowflake from the cloud provider.

Among the services in this layer:

Authentication

Infrastructure management

Metadata management

Query parsing and optimization

Access control

Question 7: **Correct**
How will you store JSON data in snowflake

- Using a column with datatype as JSON
- Using a column with datatype as VARCHAR
- Using a column with datatype as VARIANT **(Correct)**

Explanation
Snowflakes variant datatype allows you to store semi-structured non relational data like JSON, AVRO, XML

Question 8: **Correct**
You have two virtual warehouses in your snowflake instance. You have updated the data in the storage layer using one of the warehouses. When will the other warehouse be able to see the data

- Immediately **(Correct)**
- After 30 minutes once snowflake completes data synchronization
- You will need to trigger the data synchronization process for the other warehouse to see the data

Explanation

Virtual warehouses uses the same data storage layer. So If one of the warehouses updates the data, it is immediately available to all the warehouses

Question 9: **Correct**

Zero-copy cloning in snowflake is powered by which service?

- (•) Metadata store of the service layer **(Correct)**
- ( ) SSD cache of Virtual warehouse
- ( ) Query result cache

Explanation

Metadata store is a key component of the services layer. It supports zero copy cloning, data sharing and time travel

Question 10: **Correct**

What influences snowflake pricing?

- ( ) Amount of data queried from snowflake
- ( ) Amount of data scanned during querying snowflake
- (•) Snowflake pricing is based on usage and it charges only for storage and compute **(Correct)**

Explanation

There are only two const components for which Snowflake charges its customers. They are storage and compute. This is very important to remember when you optimize snowflake for cost.

Question 11: **Correct**

Compute cost in snowflake depends on

- ( ) The actual query execution time
- ( ) The query execution time and the time the query waits for the resource
- (•) The warehouse size and how long the warehouse runs **(Correct)**

Explanation

This is the beauty of snowflake. The cost does not depend on how many queries you run on the warehouse. It depends on which warehouse size you have chose and how long the warehouse was alive.

Question 12: **Correct**

You are an account administrator and you want to use a third party product in snowflake. Where will you go to enable the third party product?

- The third party product's webpage and contact them through contact us
- Call snowflake support to enable the product
- Enable the product through Partner Connect in Snowflake web console **(Correct)**

Explanation
You can enable 3rd party products by navigating to partner connect on Snowflake web console

Question 13: **Correct**
Compute in snowflake is provided by

- The cloud providers VM instances
- Only EC2 instances on AWS
- Virtual warehouses **(Correct)**

Explanation
Snowflake is a SaaS and runs on all the three major cloud providers AWS, Azure and GCP. It abstracts the computing resources through its virtual warehouse concept. Virtual warehouses are one or more cluster of servers that provide compute resources

Question 14: **Correct**
Once you have selected a warehouse size, you cannot resize the size of the warehouse

- True
- False **(Correct)**

Explanation
You can always resize a warehouse to make it bigger or smaller. However the queries that are already in process will not be able to use the resized warehouse. The resized warehouse will be used by all future queries

Question 15: **Correct**
What is the easiest way to monitor the queries that are run on snowflake?

- Create a tableau dashboard and connect to snowflake
-

All queries go to cloudwatch and use cloudwatch to monitor

- Click on the the History tab and monitor all queries that are executed in the last 14 days
  (Correct)

Explanation
The history page displays all queries that are executed in last 14 days.

Question 16: **Correct**
You are a snowflake architect hired by an organization. They want you to design their warehouse strategy. Which one of the strategy will you pick up.

- Both loading and analysis of data will be done by a single warehouse to reduce cost
- You do not need to use warehouse for loading data, only query analysis will require a warehouse
- You will recommend to use a multi-warehouse strategy. The load workload will be done by one ware house and the query analysis workload will be done by another warehouse
  (Correct)

Explanation
Multi-warehouse strategy allows you to isolate workloads by the workload type. These also helps manage access, cost accounting in a better way. Hence you should recommend multi-warehouse strategy. Please note that cost does not depend on how many warehouses you have. Compute cost depends on the size of the warehouse and the amount of time they are running.

Question 17: **Correct**
In your snowflake environment, you have a medium warehouse. The medium warehouse is used by business to run adhoc queries. The warehouse has auto suspend set at 15 minutes. You have noticed that all of the queries that run on this warehouse finishes within a minute. What will you do to optimize cost of compute in this case?

- Since all the queries are completing in a minute, you should delete this warehouse and tell the users to use another existing warehouse
- You will tell the users that they do not need snowflake to run their queries, they should load their data in another on premise database and run query from there
- You will recommend to reduce the auto suspend time to 1 minute
  (Correct)

Explanation

All virtual warehouses can be setup with 'Auto Suspend' setting. This setting suspends the warehouse if it is idle for the time that has been setup for the configuration. In this case, all the queries are completing in 1 minute but since the auto suspend is turned on, the warehouse will keep on running for 15 minutes. Compute cost depends on how long the warehouse is running. So, you should reduce the auto suspend time setting

Question 18: **Correct**

When a warehouse does not have enough resources available to process the queries, what happens to the incoming queries?

- Queries are aborted immediately
- Snowflake automatically resizes the warehouse
- The queries are queued and then executed when the resources are available again **(Correct)**

Explanation

If the warehouse is already overloaded with queries, the new queries are queued for execution. As and when the resources get available, the warehouse executes the query. You can monitor if your warehouse is overloaded by tracking the query overload parameter or going to the warehouse view in the snowflake web UI.

Question 19: **Incorrect**

In you organization, you have snowflake Enterprise addition. You notice that consistently, your queries are getting queued on the warehouses and delaying your ETL process. What are the possible solution options that you can think of?

- Resize the warehouse **(Correct)**
- Use multi-cluster warehouse **(Correct)**
- Set the auto-resize parameter of the warehouse
- Contact snowflake support to increase the number of servers for the warehouse

Explanation

When you see that you are experiencing query overload on a consistent basis, you know that the warehouse is getting overworked. In this case you can either resize your warehouse to make it bigger or use a multi-cluster warehouse. In a multi-cluster ware house you can mention the minimum and maximum servers to instantiate. Multi-cluster warehouse will then spin up additional servers based on the query load. Multi-cluster warehouses are an enterprise edition feature

Question 20: **Correct**

You can load data using PUT command through worksheets in snowflake web ui

- True
- False **(Correct)**

Explanation
You cannot use worksheets to load data into snowflake using PUT. PUT only works from SNOWSQL

Question 21: **Incorrect**
What are the valid data loading options in snowflake?

- Using snowsql and sql **(Correct)**
- Using snowpipe **(Correct)**
- Using 3rd part ETL tool **(Correct)**
- Using the cloud providers data upload tools **(Incorrect)**

Explanation
You can bulk load data into snowflake using SNOWSQL, Snowpipe or 3rd party ETL tools like Informatica, Talend, Pentaho etc.

Question 22: **Correct**
Snowflake allows only loading of structured data

- True
- False **(Correct)**

Explanation
In snowflake, you can load and store both structured and semi-structured data like XML, JSON, AVRO, Parquet and ORC. Variant datatype is usually used store semi-structured data

Question 23: **Correct**
What are the usual data loading steps in snowflake?

- Source -> Snowflake Stage -> Snowflake table **(Correct)**

- Source -> Cloud storage -> Snowflake stage -> Snowflake table
- Source -> Snowflake temp table -> Snowflake transient table-> Snowflake permanent table

Explanation
Data from source is usually batch loaded to Snowflake stage which is a snowflake object that loads the data as raw files in the underlying cloud storage object. From the snowflake stage, the data is copied into snowflake tables for further processing.

Question 24: **Correct**
You have several CSV files loaded into your named snowflake internal stage. You want to load files from the stage into a table using pattern matching to only load uncompressed CSV files whose names include the string `sales`. Which is the command that you will use to do the same?

- ```
  copy into mytable
    from @my_int_stage
    file_format = (format_name = myformat)
    pattern='.*sales.*[.]csv';
  ```

  (Correct)
- ```
  copy into mytable
    from @my_int_stage
    regex='.*sales.*[.]csv';
  ```
- ```
  copy into mytable
    from @my_int_stage
    match_pattern='.*sales.*[.]csv';
  ```

Explanation
To load data from staged files to an existing table, the files must already be staged in one of the following locations:

1. Named internal stage (or table/user stage). Files can be staged using the PUT command.

2. Named external stage that references an external location (Amazon S3, Google Cloud Storage, or Microsoft Azure).

3. External location (Amazon S3, Google Cloud Storage, or Microsoft Azure).

In this case we have the file staged in a named internal stage. To load files matching a pattern, we will need to use the PATTERN keyword.
  ```

```
[ PATTERN = '<regex_pattern>' ]
```

https://docs.snowflake.com/en/sql-reference/sql/copy-into-table.html

Question 25: **Incorrect**

Snowflake is compliant with which certifications?

- HIPAA, PCI DSS, HIPAA, PCI DSS and FedRAMP
  (Correct)
- HIPAA, PCI DSS
  (Incorrect)
- Only HIPAA
- Only FedRAMP

Question 26: **Correct**

When you sign up for a snowflake trial account, what are three snowflake editions that are offered

- [ ] Free-Tier
- [x] Standard
  (Correct)
- [ ] Enterprise Data Security
- [x] Enterprise
  (Correct)
- [x] Business critical
  (Correct)

Question 27: **Correct**

What are the snowflake editions available as of today?

- Standard Edition,Enterprise Edition,Business Critical Edition,Virtual Private Snowflake
  (Correct)
- Standard Edition,Enterprise Edition,ESD,Business Critical Edition,Virtual Private Snowflake

- Standard Edition,Enterprise Edition

Explanation
Snowflake offers multiple editions to chose from.

The Snowflake Edition that your organization chooses determines the unit costs for the credits and the data storage you use. Other factors that impact unit costs are the regions where your Snowflake account is located and whether it is an *On Demand* or *Capacity* account:

On Demand: Usage-based pricing with no long-term licensing requirements.

Capacity: Discounted pricing based on an up-front Capacity commitment.

https://docs.snowflake.com/en/user-guide/intro-editions.html#

Question 28: **Correct**
Which snowflake edition supports private communication between Snowflake and your other VPCs through AWS PrivateLink

- Standard
- Premier
- Enterprise
- Business critical **(Correct)**

Explanation
Only business critical and above supports private communication between Snowflake and your other VPCs through AWS PrivateLink or your other VNets through Azure PrivateLink

Question 29: **Correct**
Which cloud providers are supported by snowflake

- Google Cloud Platform **(Correct)**
- Azure **(Correct)**
- AWS

(Correct)

Explanation
A Snowflake account can be hosted on any of the following cloud platforms:

Amazon Web Services (AWS)

Google Cloud Platform (GCP)

Microsoft Azure (Azure)

On each platform, Snowflake provides one or more regions where the account is provisioned.

Note : GCP support has started from 2020
- https://www.snowflake.com/news/snowflake-announces-general-availability-on-google-cloud/

https://docs.snowflake.com/en/user-guide/intro-cloud-platforms.html

Question 30: **Correct**
Snowflake automatically stores data in encrypted form in all editions

- True
  (Correct)
- False

Explanation
Protecting customer data is one of Snowflake's highest priorities. Snowflake encrypts all customer data by default, using the latest security standards, at no additional cost. Snowflake provides best-in-class key management, which is entirely transparent to customers. This makes Snowflake one of the easiest to use and most secure data warehouses

https://docs.snowflake.com/en/user-guide/security-encryption.html

Question 31: **Incorrect**
You have created a lot of shares with role as ACCOUNTADMIN. You want to create the shares with SYSADMIN as role, so you have granted 'CREATE SHARE' privilege to SYSADMIN. How do you change the ownership of the existing shares?

- Execute 'GRANT OWNERSHIP ON <SHARE NAME> TO ROLE SYSADMIN'

- Execute 'GRANT OWNERSHIP TO ROLE SYSADMIN ON <SHARE NAME>
  (Incorrect)

- The only option is to drop and recreate the share. Ownership of a share cannot be granted to another role
  (Correct)

Explanation

```
The ACCOUNTADMIN role has the privileges to create a share.

You can also grant CREATE SHARE and IMPORT SHARE to other roles, enabling the tas
ks to be delegated to other users in the account. For more information, see Enabl
ing non-ACCOUNTADMIN Roles to Perform Data Sharing Tasks.
```

Ownership of a share cannot be transferred to another role. It needs to be deleted and recreated. This is by design and is done for security reasons. Users with this role can expose any object they own.

```
Ownership of a share, as well as the objects in the share, may be either through
a direct grant to the role or inherited from a lower-level role in the role hiera
rchy. For more details, see Role Hierarchy and Privilege Inheritance.

Also, it’s possible for the same role to own a share and the objects in the share
.
```

https://docs.snowflake.com/en/user-guide/security-access-privileges-shares.html#enabling-non-accountadmin-roles-to-perform-data-sharing-tasks

Question 32: **Incorrect**
Select all options that are true for ORDER BY in snowflake

- ☑ All data is sorted according to the numeric byte value of each character in the ASCII table. UTF-8 encoding is supported
  (Correct)

- ☐ For numeric values, leading zeros before the decimal point and trailing zeros (`0`) after the decimal point have no effect on sort order.
  (Correct)

- ☑ Unless specified otherwise, NULL values are considered to be higher than any non-NULL values. As a result, the ordering for NULLS depends on the sort order:
  1. If the sort order is `ASC`, NULLS are returned last; to force NULLS to be first, use `NULLS FIRST`.

2. If the sort order is `DESC`, NULLS are returned first; to force NULLS to be last, use `NULLS LAST`.
(Correct)

Explanation
You should at least remember the below two points. Even if this does not appear in certification, you need to know this when you will work with snowflake

1. If the sort order is `ASC`, NULLS are returned last; to force NULLS to be first, use `NULLS FIRST`.

2. If the sort order is `DESC`, NULLS are returned first; to force NULLS to be last, use `NULLS LAST`.

https://docs.snowflake.com/en/sql-reference/constructs/order-by.html#usage-notes

Question 33: **Correct**
Column level security in Snowflake allows the application of a masking policy to a column within a table or view. Which two features are related to column level security

- ☑ Dynamic data masking
(Correct)
- ☑ External tokenization
(Correct)
- ☐ Data loss prevention(DLP)

Explanation
Column level security is still in preview feature and is available in enterprise edition or above. This is a very useful feature and you will need this while designing your solution on snowflake

https://docs.snowflake.com/en/user-guide/security-column.html#column-level-security

Question 34: **Correct**
In snowflake, what are the two major cost categories

- ☑ Storage
(Correct)
- ☑ Compute
(Correct)

- ☐ Cloud services
- ☐ Network Traffic

Explanation
This is a very important thing to remember. In snowflake you pay only for storage and compute. So, everything that you do in snowflake, you must keep this in perspective. For example while designing queries, check if your query can be satisfied by the metadata store. Metadata store in snowflake is in the services layer and does not charge you for compute or storage. Running MIN() and MAX() functions on your field doesn't require any warehouse compute as it is stored in the metadata store, hence it does not cost you anything. Also queries like 'SHOW COLUMNS' does not require a running warehouse to execute as the data comes from the metadata store

Question 35: **Incorrect**
What are the installment options available for snowflake?

- ☑ Snowflake hosted accounts on AWS **(Correct)**
- ☑ Hybrid on-premise and cloud **(Incorrect)**
- ☐ On-premise
- ☑ Snowflake hosted accounts on Azure **(Correct)**

Explanation
This is very important to know. Snowflake does not have an on-premise option. Since last year it was available on AWS and Azure. From this year onwards, it is also available on GCP

Snowflake's data warehouse is a true SaaS offering. More specifically:

1. There is no hardware (virtual or physical) for you to select, install, configure, or manage.

2. There is no software for you to install, configure, or manage.

3. Ongoing maintenance, management, and tuning is handled by Snowflake.

Snowflake **runs completely on cloud infrastructure**. All components of Snowflake's service (other than an optional command line client), run in a public cloud infrastructure.

https://docs.snowflake.com/en/user-guide/intro-key-concepts.html#data-warehouse-as-a-cloud-service

Question 36: **Incorrect**
How is the data storage cost computed for snowflake?

- ☐ Based on amount of compressed data **(Correct)**
- ☐ Based on amount of un-compressed data
- ☑ Based on daily average of data stored **(Correct)**
- ☐ Amount stored on first day of month
- ☐ Amount stored on last day of month

Explanation
Usage for **data storage** is calculated on the daily average amount of data (in bytes) stored in the system for:

1. Files staged for bulk data loading/unloading (can be stored compressed or uncompressed).

2. Database tables, including historical data for Time Travel (always compressed by Snowflake).

3. Fail-safe for database tables (always compressed by Snowflake).

4. Clones of database tables that reference data deleted in the table that owns the clones.

Please look at point #3 above. Snowflake has a storage cost for fail-safe. Now by this time you may already know that Transient tables do not have fail-safe. So for use cases where it makes sense to use transient tables, you can choose transient table to save

some costs. Note that storage costs are not that much so weigh in on what you are gaining by going with transient table.

Question 37: **Incorrect**
Which type of data incur snowflake storage cost?

- [ ] Data stored in permanent tables
  (Correct)
- [x] Data retained to enable data recovery(fail-safe and time travel)
  (Correct)
- [ ] Cached results
- [ ] Semi structured data stored in the cloud storage(like AWS S3, GCS)

Explanation
Storage cost is charged for below type of data

1. Files stored in Snowflake locations (i.e. user and table stages or internal named stages) for bulk data loading/unloading. The user who stages a file can choose whether or not to compress the file to reduce storage.

2. Data stored in database tables, including historical data maintained for Time Travel. Snowflake automatically compresses all data stored in tables and uses the compressed file size to calculate the total storage used for an account.

3. Historical data maintained for Fail-safe.

Question 38: **Correct**
How often does snowflake release new feature

- ( ) Daily
- (•) Weekly
  (Correct)
- ( ) Monthly
- ( ) Annually

Explanation
Snowflake is committed to providing a seamless, always up-to-date experience for our users while also delivering ever-increasing value through rapid development and continual innovation.

To meet this commitment, we deploy new releases and patch releases each week. This allows us to regularly deliver service improvements in the form of new features, enhancements, and fixes. The deployments happen transparently in the background; users experience no downtime or disruption of service, and are always assured of running on the most-recent release with access to the latest features.

https://docs.snowflake.com/en/user-guide/intro-releases.html#snowflake-releases

Question 39: **Correct**
Snowflake can carry out transformations after loading files staged by partner software (ELT)

- TRUE
  (Correct)
- FALSE

Explanation
Snowflake, the cloud data platform, offers secure data sharing that eliminates the need for data extraction or transformation between departments, geographies, or partners. For primary data source loading, Snowflake works with a range of data integration partners and allows users to choose either ETL or transform data after loading (ELT). Snowflake removes the worry from data integration and allows you to focus on results.

Question 40: **Correct**
From a snowflake perspective what is common about Fivetran, Informatica, Matillion, Segment, Stitch and Talend

- They are all snowflake data integration partner
  (Correct)
- They are all snowflake competitors
- They are all programming interface patterns

Explanation
These products are data integration partner of snowflake and can be used to load data into snowflake

Question 41: **Incorrect**
Looker is a snowflake business intelligence partner

-

TRUE
(Correct)

- FALSE
(Incorrect)

Question 42: **Correct**
What is the name of the Snowflake's Command Line Interface tool?

- SnowCLI
- SnowSQL
(Correct)
- SnowCMD
- SnowSpark

Explanation
SnowSQL is the next-generation command line client for connecting to Snowflake to execute SQL queries and perform all DDL and DML operations, including loading data into and unloading data out of database tables.

SnowSQL (`snowsql` executable) can be run as an interactive shell or in batch mode through `stdin` or using the `-f` option.

https://docs.snowflake.com/en/user-guide/snowsql.html#snowsql-cli-client

Question 43: **Correct**
What is the best way to get the latest ODBC connector for use with Snowflake?

- Download it from snowflake web UI
(Correct)
- Search google and download any version of ODBC
- Compile in .NET

Question 44: **Correct**
Snowflake has scalar and tabular user-defined functions (UDFs), with support for both SQL and JavaScript.

- TRUE
(Correct)

- FALSE

Explanation
Snowflake currently supports two types of UDFs, *SQL* and *JavaScript*:

A SQL UDF evaluates an arbitrary SQL expression and returns either scalar or tabular results.

A JavaScript UDF lets you use the JavaScript programming language to manipulate data and return either scalar or tabular results.

https://docs.snowflake.com/en/sql-reference/user-defined-functions.html#udfs-user-defined-functions

Question 45: **Incorrect**
This object in snowflake returns a set of rows instead of a single, scalar value, and can be accessed in the FROM clause of a query

- UDTF
  (Correct)
- UDF
  (Incorrect)
- Stored procedure

Explanation
A UDTF returns a set of rows instead of a single, scalar value, and can be accessed in the FROM clause of a query. Snowflake supports both SQL and JavaScript UDTFs. This topic covers SQL UDTFs.

https://docs.snowflake.com/en/sql-reference/udf-table-functions.html#sql-udtfs-user-defined-table-functions

Question 46: **Incorrect**
By default, result reuse is enabled in snowflake. if you want to disable it what will you do?

- Execute `ALTER SESSION SET USE_CACHED_RESULT=FALSE`
  (Correct)
- Execute `ALTER SESSION SET QUERY_CACHE=FALSE`
-

Execute `ALTER SESSION SET QUERY_RESULT_CACHE =FALSE`
(Incorrect)

Explanation
You will need to change it with the alter session query. This may not come in certification. But you will probably using it a lot when you actually work on a snowflake project. You will use it to test the performance once you have optimized your query. For example if a table is not partitioned well and you want to check the query performance after properly partitioning the table(either through clustering or reloading the table using natural order)

Question 47: **Correct**
Each time a persisted result for a query is reused, Snowflake resets the 24-hour retention period for the result, up to a maximum of 31 days from the date and time that the query was first executed. After 31 days, the result is purged and the next time the query is submitted, a new result is generated and persisted.

- TRUE
  (Correct)
- FALSE

Explanation
This is a very important point to remember. The query result of an already run query is persisted for 24 hours and every time the query is run the 24 hour window is reset for another 24 hours till 31 days

Question 48: **Correct**
Snowpipe guarantees that files are loaded in the same order they are staged

- TRUE
- FALSE
  (Correct)

Explanation
For each pipe object, Snowflake establishes a single queue to sequence data files awaiting loading. As new data files are discovered in a stage, Snowpipe appends them to the queue. However, multiple processes pull files from the queue; and so, while Snowpipe generally loads older files first, there is no guarantee that files are loaded in the same order they are staged.

Question 49: **Correct**
Snowpipe is recommended to load data in which of the below scenario?

- You have small volume of frequent data
  (Correct)
-

You have a huge volume of data generated as part of a batch schedule

Explanation
Please not the key word 'SMALL VOLUME' and 'FREQUENT'. This question can be framed in multiple ways.

Question 50: **Correct**
What should be your minimum snowflake edition for data sharing capability

- (•) Standard **(Correct)**
- ( ) Enterprise
- ( ) Business critical

Explanation

Data Sharing

Feature/Service	Standard	Enterprise	Business Critical	VPS
As a data provider, securely share data with other accounts.	✔	✔	✔	✔
As a data consumer, query data shared with your account by data providers.	✔	✔	✔	✔
Secure data sharing across regions and cloud platforms (through data replication)	✔	✔	✔	✔
Snowflake Data Marketplace, where providers and consumers meet to securely sharing data.	✔	✔	✔	✔
Data Exchange, a private hub of administrators, providers, and consumers that you invite to securely collaborate around data.	✔	✔	✔	✔

Question 51: **Correct**
Select the snowflake edition that allow only a maximum of 1 day of time travel

- (•) Standard **(Correct)**
- ( ) Enterprise
- ( ) VPS
- ( ) Business Critical

Explanation
Standard Time Travel (up to 1 day) for accessing/restoring modified and deleted data is available for all editions. Extended Time Travel (up to 90 days) is available for all editions except Standard

Question 52: **Correct**

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

- TRUE
  (Correct)
- FALSE

Explanation
You do not have to mug up the answer.

Just remember what prevents the loaded files to be reloaded again. It is the load metadata. Snowflake maintains detailed metadata for each table into which data is loaded. This data expires after 64 days. Through this data, Snowflake knows that the file has been already loaded to the table. When you clone the table, this metadata is not cloned and hence you will be able to load the data again to the cloned table.

Refresh your memory on load metadata

Load Metadata

Snowflake maintains detailed metadata for each table into which data is loaded, including:

Name of each file from which data was loaded

1. File size

2. ETag for the file

3. Number of rows parsed in the file

4. Timestamp of the last load for the file

5. Information about any errors encountered in the file during loading

This load metadata expires after 64 days. If the LAST_MODIFIED date for a staged data file is less than or equal to 64 days, the COPY command can determine its load status for a given table and prevent reloading (and data duplication). The LAST_MODIFIED date is the timestamp when the file was initially staged or when it was last modified, whichever is later.

If the LAST_MODIFIED date is older than 64 days, the load status is still known if ***either*** of the following events occurred less than or equal to 64 days prior to the current date:

The file was loaded successfully.

The initial set of data for the table (i.e. the first batch after the table was created) was loaded.

However, the COPY command cannot definitively determine whether a file has been loaded already if the LAST_MODIFIED date is older than 64 days ***and*** the initial set of data was loaded into the table more than 64 days earlier (***and*** if the file was loaded into the table, that also occurred more than 64 days earlier). In this case, to prevent accidental reload, the command skips the file by default.

Question 53: **Correct**
What is the technique called which snowflake uses to limit the number of micro-partitions retrieved as part of a query?

- Pruning
  (Correct)
- Selective Filter
- Indexing
- Clustering

Explanation
Query Pruning

The micro-partition metadata maintained by Snowflake enables precise pruning of columns in micro-partitions at query run-time, including columns containing semi-structured data. In other words, a query that specifies a filter predicate on a range of values that accesses 10% of the values in the range should ideally only scan 10% of the micro-partitions.

For example, assume a large table contains one year of historical data with date and hour columns. Assuming uniform distribution of the data, a query targeting a particular hour would ideally scan 1/8760th of the micro-partitions in the table and then only scan the portion of the micro-partitions that contain the data for the hour column; Snowflake uses columnar scanning of partitions so that an entire partition is not scanned if a query only filters by one column.

In other words, the closer the ratio of scanned micro-partitions and columnar data is to the ratio of actual data selected, the more efficient is the pruning performed on the table.

For time-series data, this level of pruning enables potentially sub-second response times for queries within ranges (i.e. “slices”) as fine-grained as one hour or even less.

Not all predicate expressions can be used to prune. For example, Snowflake does not prune micro-partitions based on a predicate with a subquery, even if the subquery results in a constant.

Question 54: **Correct**
Select two choices that are true about snowflake roles

- ☐ Snowflake users has a limit on the number of roles that they can assume
- ☑ Snowflake user can have one or more roles
  (Correct)
- ☐ Privileges can be directly assigned to a user
- ☑ For a particular session, only one role can be active at a given time
  (Correct)

Question 55: **Correct**
Snowflake waits till all the servers are provisioned for a new virtual warehouse, before it executes query on that warehouse

- ◉ TRUE
  (Correct)
- ○ FALSE

Explanation
Snowflake does not begin executing SQL statements submitted to a warehouse until all of the servers for the warehouse are successfully provisioned, unless any of the servers fail to provision:

1. If any of the servers for the warehouse fail to provision during start-up, Snowflake attempts to repair the failed server(s).

2. During the repair process, the warehouse starts processing SQL statements once 50% or more of the requested servers are successfully provisioned.

Question 56: **Incorrect**
If you want customer dedicated virtual warehouse, which is the lowest snowflake edition that you should opt for

- Business Critical
- Enterprise **(Incorrect)**
- Standard **(Correct)**

Explanation
Customer dedicated virtual warehouse is available in all editions
Question 57: **Correct**
You need to contact Snowflake Support team if retrieval of data is required from fail safe

- TRUE **(Correct)**
- FALSE

Explanation
There is nothing to explain here, you need to remember this.
Question 58: **Correct**
After how many days do the load metadata of a table expire?

- 64 **(Correct)**
- 14
- 7
- 365

Explanation
This load metadata `expires after 64 days.` If the LAST_MODIFIED date for a staged data file is less than or equal to 64 days, the COPY command can determine its load status for a given table and prevent reloading (and data duplication). The LAST_MODIFIED date is the timestamp when the file was initially staged or when it was last modified, whichever is later.
Question 59: **Correct**

Let us imagine you have a permanent table named EMPLOYEE, you have dropped the table. Then you created another table with the same name. If you execute the UNDROP command now, what will happen?

- (•) UNDROP command will fail
  (Correct)
- ( ) It will work similar to insert overwrite
- ( ) The new table will be renamed with a suffix of v1.

Explanation
Let us do a hand's on and learn it ourselves.

```
CREATE OR REPLACE TABLE EMPLOYEE(EMPLOYEE_NAME VARCHAR);


INSERT INTO EMPLOYEE VALUES ('AKSHAY'),('RUPAK'),('JEETU'); -- Please note the sy
ntax, this is bulk insert.


DROP TABLE EMPLOYEE;


CREATE OR REPLACE TABLE EMPLOYEE(EMPLOYEE_NAME VARCHAR);


UNDROP TABLE EMPLOYEE;
```

You will get the below error

```
SQL compilation error: Object 'EMPLOYEE' already exists.
```

Question 60: **Incorrect**
Which of the below are true with respect to snowflake web ui?

- [x] Each worksheet can have different role, warehouse, schema and database
  (Correct)
- [ ] Each worksheet is a separate session
  (Correct)

- You will be able to run a 'PUT' command from worksheet
- You must use the same role and warehouse for all worksheets

Explanation
These are the easy set of questions in the exam. You should not miss it. Please log on to the WEB UI and validate each statement. Try to run the below. Can you run it

```
CREATE STAGE CSV_STAGE;

PUT 'D:/MYFILE.CVS' @CSV_STAG;
```

This will not run and you will get below error

```
SQL compilation error: The command is not supported from the UI: PUT
```

Question 61: **Correct**
When a network policy includes values in both the allowed and blocked IP address lists, Snowflake applies the ***blocked*** IP address list first.

- TRUE
  (Correct)
- FALSE

Explanation
When a network policy includes values in both the allowed and blocked IP address lists, Snowflake applies the ***blocked*** IP address list first.

Do ***not*** add `0.0.0.0/0` to the blocked IP address list. `0.0.0.0/0` is interpreted to be “all IPv4 addresses on the local machine”. Because Snowflake resolves this list first, this would block your own access. Also, note that it is not necessary to include this IP address in the allowed IP address list.

Question 62: **Incorrect**
Lets say you executed a transaction in a snowflake session. Due to some reason the session disconnects and the transaction now is in a detached state. The transaction cannot be committed or rolled back. The object on which the transaction was applied is also now locked. if you do not do anything and let snowflake eventually abort the transaction, how long will you need to wait?

- 60 minutes
-

15 minutes
(Incorrect)

- 4 hours
(Correct)

Explanation
If a transaction is running in a session and the session disconnects abruptly, preventing the transaction from committing or rolling back, the transaction is left in a detached state, including any locks that the transaction is holding on resources. If this happens, you might need to abort the transaction.

To abort a running transaction, the user who started the transaction or an account administrator can call the system function, SYSTEM$ABORT_TRANSACTION.

```
If the transaction is left open, Snowflake typically rolls back the
transaction after it has been idle for four hours.
```

Question 63: **Incorrect**
Snowflake does not support nested transactions

- TRUE
(Correct)

- FALSE
(Incorrect)

Explanation
Overlapping Transactions

This section describes overlapping transactions.

A stored procedure that contains a transaction can be called from within another transaction. The outer transaction can be in an outer stored procedure or can be outside any stored procedure.

The inner transaction is ***not*** treated as nested; instead, the inner transaction is ***a separate transaction***. Snowflake calls these “autonomous scoped transactions” (or simply “scoped transactions”), because each transaction executes in a conceptually ***independent scope***.

Note

Terminology note:

The terms “inner” and “outer” are commonly used when describing nested operations, such as nested stored procedure calls. Although Snowflake supports nested procedure calls, `Snowflake does not support nested transactions`;

https://docs.snowflake.com/en/sql-reference/transactions.html#overlapping-transactions

Question 64: **Correct**
You have cloned a table. Which of the below queries will work on the cloned table?

- DROP TABLE <TABLE_NAME>
- SELECT * FROM <TABLE_NAME>
- SHOW TABLES LIKE '<TABLE_NAME>'
- ALL OF THE ABOVE
  (Correct)

Explanation
Cloning creates a copy of an existing object in the system. This command is primarily used for creating zero-copy clones of databases, schemas, and non-temporary tables; however, it can also be used to quickly/easily create clones of other schema objects (i.e. external stages, file formats, and sequences).

The command is a variation of the object-specific CREATE <object> commands with the addition of the `CLONE` keyword.

When you clone a table, it is actually a regular table which shares the micro-partitions with the table from which it has been cloned

Question 65: **Correct**
Select the term that is associated with compute layer?

- Query optimization
- Query planning
- Query processing
  (Correct)

Explanation
Compute layer has virtual warehouses which are responsible for query processing

Question 66: **Correct**
You can use the query profiler view only for completed queries

- TRUE
- FALSE **(Correct)**

Question 67: **Incorrect**
You need snowflake instances in different regions. You will need to have separate account for each region

- TRUE **(Correct)**
- FALSE **(Incorrect)**

Question 68: **Correct**
In a multi-cluster warehouse, you will need to manually scale the warehouses(minimum cluster to maximum cluster)

- TRUE
- FALSE **(Correct)**

Explanation
Multi-cluster warehouses enable you to scale compute resources to manage your user and query concurrency needs as they change, such as during peak and off hours. And it is 100% automated.

Question 69: **Correct**
Which of the below statement will you use to recreate a specified object?

- GET_DDL **(Correct)**
- GET_SCHEMA
- SHOW SCHEMA

Explanation
GET_DDL

Returns a DDL statement that can be used to recreate the specified object. For databases and schemas, GET_DDL is recursive, i.e. it returns the DDL statements for recreating all supported objects within the specified database/schema.

GET_DDL currently supports the following object types:

Databases (see CREATE DATABASE)

Schemas (see CREATE SCHEMA)

Tables (see CREATE TABLE) (not external tables)

Views (see CREATE VIEW)

Streams (see CREATE STREAM)

Tasks (see CREATE TASK)

Sequences (see CREATE SEQUENCE)

File formats (see CREATE FILE FORMAT)

Pipes (see CREATE PIPE)

UDFs (see CREATE FUNCTION)

Stored Procedures (see CREATE PROCEDURE)

Try it out

```
CREATE OR REPLACE TABLE EMPLOYEE(EMPLOYEE_NAME VARCHAR);

select get_ddl('TABLE', 'EMPLOYEE', true);
```

What do you get?

Question 70: **Correct**

This command can be used to list streams for the current/specified database or schema, or across your entire account.

- (●) SHOW STREAMS **(Correct)**
- ( ) DISPLAY STREAMS
- ( ) LIST STREAMS

Explanation
SHOW STREAMS

Lists the streams for which you have access privileges. The command can be used to list streams for the current/specified database or schema, or across your entire account.

The output returns stream metadata and properties, ordered lexicographically by database, schema, and stream name (see Output in this topic for descriptions of the output columns). This is important to note if you wish to filter the results using the provided filters.

https://docs.snowflake.com/en/sql-reference/sql/show-streams.html#show-streams

Question 71: **Incorrect**
When you use this parameter with SHOW STREAMS, The output also includes an additional `dropped_on` column, which displays

1. Date and timestamp (for dropped streams).

2. `NULL` (for active streams).

- HISTORY **(Correct)**
- TERSE
- DROPPED=TRUE **(Incorrect)**

Explanation
Parameters

`TERSE`

Returns only a subset of the output columns:

`created_on`

`name`

`kind` (rename of `type` column in full set of columns)

`database_name`

`schema_name`

`tableOn` (rename of `table_name` column in full set of columns)

`HISTORY`

Optionally includes dropped streams that have not yet been purged (i.e. they are still within their respective Time Travel retention periods). If multiple versions of a dropped stream exist, the output displays a row for each version. The output also includes an additional `dropped_on` column, which displays:

Date and timestamp (for dropped streams).

`NULL` (for active streams).

Default: No value (dropped streams are ***not*** included in the output)

https://docs.snowflake.com/en/sql-reference/sql/show-streams.html#parameters

Question 72: **Correct**
The table functions in INFORMATION_SCHEMA can be used to return account-level usage and historical information for storage, warehouses, user logins, and queries

- TRUE **(Correct)**
- FALSE

Explanation
List of Table Functions

The table functions in INFORMATION_SCHEMA can be used to return account-level usage and historical information for storage, warehouses, user logins, and queries:

https://docs.snowflake.com/en/sql-reference/info-schema.html#list-of-table-functions

Question 73: **Correct**

What can you expect if the filters specified in an INFORMATION_SCHEMA query are not sufficiently selective?

- AN ERROR
  (Correct)
- WILL SHOW RESULTS WITH AN WARNING
- WILL SWITCH TO A LARGE WAREHOUSE

Explanation
To prevent performance issues, the following error is returned if the filters specified in an INFORMATION_SCHEMA query are not sufficiently selective:

```
Information schema query returned too much data. Please repeat query with more selective predicates.
```

https://docs.snowflake.com/en/sql-reference/info-schema.html#general-usage-notes

Question 74: **Incorrect**
Information Schema Views will require a warehouse to execute whereas SHOW command does not require one.

- TRUE
  (Correct)
- FALSE
  (Incorrect)

Explanation
The INFORMATION_SCHEMA views provides a SQL interface to the same information provided by the SHOW <objects> commands. You can use the views to replace these commands; however, there are some key differences to consider before switching:

Considerations	SHOW Commands	Information Schema Views
Warehouses	Not required to execute.	Warehouse must be running and currently in use to query the views.
Pattern matching/filtering	Case-insensitive (when filtering using LIKE).	Standard (case-sensitive) SQL semantics. Snowflake automatically converts unquoted, case-insensitive identifiers to uppercase internally, so unquoted object names must be queried in uppercase in the Information Schema views.
Query results	Most SHOW commands limit results to the current schema by default.	Views display all objects in the current/specified database. To query against a particular schema, you must use a filter predicate (e.g. `... WHERE table_schema = CURRENT_SCHEMA()...`). Note that Information Schema queries lacking sufficiently selective filters return an error and do not execute (see General Usage Notes in this topic).

Question 75: **Incorrect**

When unloading data into multiple files, you will use this copy option to specify the maximum size of each file created

- MAX_FILE_SIZE
  (Correct)
- MAX_SIZE
- MAX_FILE_BYTES
  (Incorrect)

Explanation
Bulk Unloading into Single or Multiple Files

The COPY INTO <location> command provides a copy option (SINGLE) for unloading data into a single file or multiple files. The default is SINGLE = FALSE (i.e. unload into multiple files).

Snowflake assigns each file a unique name. The location path specified for the command can contain a filename prefix that is assigned to all the data files generated. If a prefix is not specified, Snowflake prefixes the generated filenames with `data_`.

Snowflake appends a suffix that ensures each file name is unique across parallel execution threads; e.g. `data_stats_0_1_0`.

When unloading data into multiple files, use the `MAX_FILE_SIZE` copy option to specify the maximum size of each file created.

Question 76: **Correct**
How do you truncate a date (from a timestamp) down to the year, month, and day

- Use concatenation of the date part
- Split the date into year, month and day
- Use date_trunc and execute a query as below

```
1.  select to_date('2015-05-08T23:39:20.123-07:00') as "DATE1",
2.         date_trunc('YEAR', "DATE1") as "TRUNCATED TO YEAR",
3.         date_trunc('MONTH', "DATE1") as "TRUNCATED TO MONTH",
4.         date_trunc('DAY', "DATE1") as "TRUNCATED TO DAY";
```

(Correct)

Explanation

You do not need to remember the query(just remember the function **date_trunc**) and it will most probably not come in your certification, but this will be an useful function when you actually start working on snowflake

https://docs.snowflake.com/en/sql-reference/functions/date_trunc.html#examples

Question 77: **Correct**
You are trying to set a variable by suing the set variable function `SET MY_VARIABLE ='XXXX'` . You got an error as below

```
Assignment to 'MY_VARIABLE'  not done because value exceeds size limit of variables. Its size is 312; the limit is 256
```

What is the reason for this error?

- (•) The size of string or binary variables is limited to 256 bytes
  (Correct)
- ( ) This is a temporary error due to insufficient memory in VM instance
- ( ) Your warehouse is not big enough to accomodate this

Explanation
Variables can be set by executing the SQL statement `SET` or by setting the variables in the connection string when you connect to Snowflake.

The size of string or binary variables is limited to 256 bytes.

https://docs.snowflake.com/en/sql-reference/session-variables.html#initializing-variables

Question 78: **Incorrect**
Snowflake supports specifying a SELECT statement instead of a table in the COPY INTO <location> command

- ( ) TRUE
  (Correct)
- (•) FALSE
  (Incorrect)

Explanation
Bulk Unloading Using Queries

`Snowflake supports specifying a SELECT statement instead of a table in the COPY INTO <location> command`. The results of the query are written to one or more files as specified in the command and the file(s) are stored in the specified location (internal or external).

SELECT queries in COPY statements support the full syntax and semantics of Snowflake SQL queries, including JOIN clauses, which enables downloading data from multiple tables.

https://docs.snowflake.com/en/user-guide/data-unload-overview.html#bulk-unloading-using-queries

Question 79: **Incorrect**
Using **COPY INTO *<location>*** command, you can unload data from a table (or query) into the below locations.

- ☑ Named internal stage (or table/user stage)
  (Correct)
- ☐ Named external stage
  (Correct)
- ☐ External location (Amazon S3, Google Cloud Storage, or Microsoft Azure)
  (Correct)
- ☐ Local drive

Explanation
COPY INTO *<location>*

Unloads data from a table (or query) into one or more files in one of the following locations:

1. Named internal stage (or table/user stage). The files can then be downloaded from the stage/location using the **GET command**.

2. Named external stage that references an external location (Amazon S3, Google Cloud Storage, or Microsoft Azure).

3. External location (Amazon S3, Google Cloud Storage, or Microsoft Azure).

Question 80: **Correct**
To download files from the stage/location loaded through COPY INTO <LOCATION> command, you will use

- GET
  (Correct)
- UNLOAD
- COPY INTO

Explanation
GET

Downloads data files from one of the following Snowflake stages to a local directory/folder on a client machine:

Named internal stage.

Internal stage for a specified table.

Internal stage for the current user.

Typically, this command is executed after using the COPY INTO <location> command to unload data from a table into a Snowflake stage.

https://docs.snowflake.com/en/sql-reference/sql/get.html#get

Question 81: **Incorrect**
GET does ***not*** support downloading files from external stages

- FALSE
  (Incorrect)
- TRUE
  (Correct)

Explanation
`GET does not support downloading files from external stages.` To download files from external stages, use the utilities provided by the cloud service.

The following Snowflake clients do not support GET:

Go Snowflake Driver

.NET Driver

Node.js Driver

The ODBC driver supports GET with Snowflake accounts hosted on the following platforms:

Amazon Web Services (using ODBC Driver Version 2.17.5 and higher).

Google Cloud Platform (using ODBC Driver Version 2.21.5 and higher).

Microsoft Azure (using ODBC Driver Version 2.20.2 and higher).

Question 82: Incorrect
You will use this parameter to specify the number of threads to use for downloading the files using GET command

- PARALLEL = <INTEGER>
  (Correct)
- DOWNLOAD_PARALLEL=<INTEGER>
  (Incorrect)
- DUMP=<INTEGER>

Explanation
Optional Parameters

`PARALLEL = integer`

Specifies the number of threads to use for downloading the files. The granularity unit for downloading is one file.

Increasing the number of threads can improve performance when downloading large files.

Supported values: Any integer value from `1` (no parallelism) to `99` (use 99 threads for downloading files).

Default: `10`

`PATTERN = 'regex_pattern'`

Specifies a regular expression pattern for filtering files to download.

Default: No value (all files in the specified stage are downloaded)

https://docs.snowflake.com/en/sql-reference/sql/get.html#optional-parameters

Question 83: **Correct**
Following commands cannot be executed from worksheets

- [x] PUT
  (Correct)
- [x] GET
  (Correct)
- [ ] SHOW
- [ ] LIST <STAGE>

Question 84: **Correct**
If file format options are specified in multiple locations, the load operation applies the options in the following order of precedence.

- (•) 1. COPY INTO TABLE statement.
  2. Stage definition.
  3. Table definition.
  (Correct)
- ( ) 1. Stage definition.
  2. COPY INTO TABLE statement.
  3. Table definition.
- ( ) 1. COPY INTO TABLE statement.
  2. Table definition.
  3. Stage definition.

Explanation
If file format options are specified in multiple locations, the load operation applies the options in the following order of precedence:

1. COPY INTO TABLE statement.

2. Stage definition.

3. Table definition.

Question 85: **Correct**
Which are the two metadata columns for staged files

- ☑ **METADATA$FILENAME**
  (Correct)
- ☐ **METADATA$FILEFORMAT**
- ☑ **METADATA$FILE_ROW_NUMBER**
  (Correct)

Explanation
Metadata Columns

Currently, the following metadata columns can be queried or copied into tables:

METADATA$FILENAME

Name of the staged data file the current row belongs to. Includes the path to the data file in the stage.

METADATA$FILE_ROW_NUMBER

Row number for each record in the container staged data file.

https://docs.snowflake.com/en/user-guide/querying-metadata.html#metadata-columns

Question 86: **Incorrect**
Metadata columns for staged files can only be queried by name

- ○ TRUE
  (Correct)
- ◉ FALSE
  (Incorrect)

Explanation
Query Limitations

Metadata cannot be inserted into existing table rows.

Metadata columns can only be queried by name; as such, they are not included in the output of any of the following statements:

SELECT *

SHOW <objects>

DESCRIBE <object>

Queries on INFORMATION_SCHEMA views

https://docs.snowflake.com/en/user-guide/querying-metadata.html#query-limitations

Question 87: **Incorrect**
Only named stages (internal or external) and user stages are supported for COPY transformations

- TRUE
  (Correct)
- FALSE
  (Incorrect)

Explanation
Usage Notes

This section provides usage information for transforming staged data files during a load.

Supported Stages

```
Only named stages (internal or external) and user stages are supported for COPY t
ransformations.
```

Supported File Formats

The following file format types are supported for COPY transformations:

CSV

JSON

Avro

ORC

Parquet

XML

Question 88: **Correct**
The VALIDATION_MODE parameter does not support COPY statements that transform data during a load.

- FALSE
- TRUE
  (Correct)

Explanation
VALIDATION_MODE Parameter

The VALIDATION_MODE parameter does not support COPY statements that transform data during a load.

https://docs.snowflake.com/en/user-guide/data-load-transform.html#validation-mode-parameter

Question 89: **Incorrect**
Following transformations are not supported in a COPY command. Select three.

- FLATTEN
  (Correct)
- JOIN
  (Correct)
- GROUP BY
  (Correct)
- NVL

Explanation
Note that COPY transformations do ***not*** support the FLATTEN function,
or JOIN or GROUP BY (aggregate) syntax.

This is a very important topic, please read it

https://docs.snowflake.com/en/user-guide/data-load-transform.html#transforming-data-during-a-load

Question 90: **Correct**

Snowflake recommends to compress your data files when you are loading large data sets.

- TRUE
  (Correct)
- FALSE

Explanation

`We recommend that you compress your data files when you are loading large data sets.` See CREATE FILE FORMAT for the compression algorithms supported for each data type.

When loading compressed data, specify the compression method for your data files. The COMPRESSION file format option describes how your data files are ***already*** compressed in the stage. Set the COMPRESSION option in one of the following ways:

1. As a file format option specified directly in the COPY INTO <table> statement.

2. As a file format option specified for a named file format or stage object. The named file format/stage object can then be referenced in the COPY INTO *<table>* statement.

Question 91: **Correct**
Snowpipe is a serverless function

- TRUE
  (Correct)
- FALSE

Explanation

Automated data loads leverage event notifications for cloud storage to inform Snowpipe of the arrival of new data files to load. Snowpipe copies the files into a queue, from which they are loaded into the `target table in a continuous, serverless fashion` based on parameters defined in a specified pipe object.

Question 92: **Incorrect**
When calling the rest endpoints in snowpipe, below authentication methods are supported

- key pair authentication with JSON Web Token
  (Correct)
- user id and password

- Both
  (Incorrect)

Explanation
Authentication

Bulk data load

Relies on the security options supported by the client for authenticating and initiating a user session.

Snowpipe

When calling the REST endpoints: Requires key pair authentication with JSON Web Token (JWT). JWTs are signed using a public/private key pair with RSA encryption.

Question 93: **Correct**
Load history of Snowpipe expires after how many days.

- 30
- 14
  (Correct)
- 64

Explanation
Load History

Bulk data load

Stored in the metadata of the target table for 64 days. Available upon completion of the COPY statement as the statement output.

Snowpipe

Stored in the metadata of the pipe for 14 days. Must be requested from Snowflake via a REST endpoint, SQL table function, or ACCOUNT_USAGE view.

Question 94: **Correct**
SnowPipe can load a file with same name if it has been modified later.

- TRUE
-

FALSE
(Correct)

Explanation
Data Duplication

Snowpipe uses file loading metadata associated with each ***pipe object*** to prevent reloading the same files (and duplicating data) in a table. This metadata stores the path (i.e. prefix) and name of each loaded file, and prevents loading files with the same name even if they were later modified (i.e. have a different eTag).

Question 95: **Incorrect**
SNOWPIPE REST APIs support both internal and external stage

- TRUE
(Correct)

- FALSE
(Incorrect)

Explanation
Snowpipe supports both internal (Snowflake) stages and external stages, i.e. S3 buckets. Please note that this is true for REST APIs. SNOWPIPE Auto ingest supports only external stage

Question 96: **Incorrect**
Which of the below are SNOWPIPE REST APIs. Select three.

- `insertFiles`
(Correct)

- `insertReport`
(Correct)

- `loadHistoryScan`
(Correct)

- `loadHistoryProgress`
(Incorrect)

Explanation
SNOWPIPE has three APIs

Data File Ingestion

1. Endpoint: `insertFiles`

Load History Reports

1. Endpoint: `insertReport`

2. Endpoint: `loadHistoryScan`

https://docs.snowflake.com/en/user-guide/data-load-snowpipe-rest-apis.html#snowpipe-rest-api

Question 97: **Incorrect**
A successful response from the SNOWPIPE insertFiles API means that the files are ingested.

- ◉ TRUE
  (Incorrect)
- ○ FALSE
  (Correct)

Explanation
Endpoint: `insertFiles`

Informs Snowflake about the files to be ingested into a table. A successful response from this endpoint means that Snowflake has recorded the list of files to add to the table. `It does not necessarily mean the files have been ingested`

Question 98: **Incorrect**
The insertReport SNOWPIPE API can retrieve file ingestion events and report it. The events are retained for infinite time until deleted manually.

- ◉ TRUE
  (Incorrect)
- ○ FALSE
  (Correct)

Explanation
The SNOWPIPE REST API topic is relevant for both SNOWPRO CORE and SNOWPRO ARCHITECT certifications. Please read this chapter thoroughly

https://docs.snowflake.com/en/user-guide/data-load-snowpipe-rest-apis.html#snowpipe-rest-api

Endpoint: `insertReport`

Retrieves a report of files submitted via `insertFiles` whose contents were recently ingested into a table. Note that for large files, this may only be part of the file.

Note the following limitations for this endpoint:

1. The 10,000 most recent events are retained.

2. Events are retained for a maximum of 10 minutes.

An event occurs when data from a file submitted via `insertFiles` has been committed to the table and is available to queries. The `insertReport` endpoint can be thought of like the UNIX command tail. By calling this command repeatedly, it is possible to see the full history of events on a pipe over time. Note that the command must be called often enough to not miss events. How often depends on the rate files are sent to `insertFiles`.

Question 99: **Correct**
To help avoid exceeding the rate limit (error code 429), snowflake recommends relying more heavily on `insertReport` than `loadHistoryScan`

- (•) TRUE **(Correct)**
- ( ) FALSE

Explanation
Endpoint: `loadHistoryScan`

Fetches a report about ingested files whose contents have been added to table. Note that for large files, this may only be part of the file. This endpoint differs from `insertReport` in that it views the history between two points in time. There is a maximum of 10,000 items returned, but multiple calls can be issued to cover the desired time range.

Important

This endpoint is rate limited to avoid excessive calls. To help avoid exceeding the rate limit (error code 429), we recommend relying more heavily on `insertReport` than `loadHistoryScan`. When calling `loadHistoryScan`, specify the most narrow time range that includes a set of data loads. For example, reading the last

10 minutes of history every 8 minutes would work well. Trying to read the last 24 hours of history every minute will result in 429 errors indicating a rate limit has been reached. The rate limits are designed to allow each history record to be read a handful of times.

Question 100: **Incorrect**
You have a warehouse. It ran for 62 seconds and then was suspended. After that it resumed and ran for 20 seconds. For how many seconds will you be billed

- 122 **(Correct)**
- 20
- 92 **(Incorrect)**

Explanation
When a virtual warehouse starts for the first time, the first one minute is anyway charged after that the billing is per second

Warehouses are only billed for credit usage when they are running. When a warehouse is suspended, it does not accrue any credit usage.

The credit numbers shown here are for a full hour of usage; however, credits are billed per-second, with a 60-second (i.e. 1-minute) minimum:

Each time a warehouse is started or resized to a larger size, the warehouse is billed for 1 minute's worth of usage based on the hourly rate shown above.

After 1 minute, all subsequent billing is per-second.

Stopping and restarting a warehouse within the first minute does not change the amount billed; the minimum billing charge is 1 minute.

When a warehouse is increased in size, credits are billed only for the ***additional*** servers that are provisioned. For example, changing from Small (2) to Medium (4) results in billing charges for 1 minute's worth of 2 credits.

https://docs.snowflake.com/en/user-guide/credits.html#virtual-warehouse-credit-usage

Question 1: **Incorrect**

Which of the following best describes snowflakes processing engine?

- Leverages Apache Spark
- Leverages Map Reduce
  (Incorrect)
- A derivative of presto
- Native SQL
  (Correct)

Question 2: **Incorrect**

Which factors affect your snowflake data loading rate?

- Physical location of the stage
  (Correct)
- RAM on the virtual warehouse
  (Incorrect)
- GZIP compression efficiency
  (Correct)
- Thread size

Explanation
https://www.snowflake.com/blog/how-to-load-terabytes-into-snowflake-speeds-feeds-and-techniques/

Load rates for your own data files may differ based on a number of factors:

Location of your S3 buckets – For our test, both our Snowflake deployment and S3 buckets were located in us-west-2

Number and types of columns – A larger number of columns may require more time relative to number of bytes in the files.

Gzip Compression efficiency – More data read from S3 per uncompressed byte may lead to longer load times.

Question 3: **Correct**

You ran a query in snowflake and it took 2 minute 3 seconds to run, you ran the query again and it returned the results in less than a second. What might have happened?

- When the query ran for the second time, it used the persisted query results from the query result cache
(Correct)
- The query used the meta data cache
- All the partitions were scanned from the SSD cache

Explanation
When a query is executed, the result is persisted (i.e. cached) for a period of time (currently 24 hours). At the end of the time period, the result is purged from the system.

Snowflake uses persisted query results to avoid re-generating results when nothing has changed (i.e. “retrieval optimization”).

In addition, you can use persisted query results to post-process the results (e.g. layering a new query on top of the results already calculated).

https://docs.snowflake.com/en/user-guide/querying-persisted-results.html#using-persisted-query-results

Question 4: **Incorrect**

You are loading data to a snowflake internal stage area using the PUT Command, the data will be encrypted on the client's machine and you will not be able to turn off the encryption.

- TRUE
(Correct)
- FALSE
(Incorrect)

Explanation
One of the biggest worries people have about moving to the cloud is security. One key piece of providing enterprise class security is the ability to encrypt the data in your data warehouse environment. With Snowflake, your data is automatically encrypted by default.

https://www.snowflake.com/blog/automatic-encryption-data/

Question 5: **Incorrect**

Please select the ones that are true for a snowflake task

- [x] A task can execute a single SQL Statement
(Correct)
- [ ] A task can execute a call to a store procedure
(Correct)
- [ ] A task can execute multiple SQL statements
- [ ] A task can execute a function

Explanation
Currently, a task can execute a single SQL statement, including a call to a stored procedure.

https://docs.snowflake.com/en/user-guide/tasks-intro.html#introduction-to-tasks

Question 6: **Correct**

Which object in snowflake records data manipulation language changes made to tables, including inserts, updates, and deletes, as well as metadata change about each change?

- ( ) Tasks
- (•) Table stream
(Correct)
- ( ) Snowpipe
- ( ) Change Data Capture

Explanation
A stream object records data manipulation language (DML) changes made to tables, including inserts, updates, and deletes, as well as metadata about each change, so that actions can be taken using the changed data. This process is referred to as change data capture (CDC). An individual table stream tracks the changes made to rows in a *source table*. A table stream (also referred to as simply a “stream”) makes a “change table” available of what changed, at the row level, between two transactional points of time in a table. This allows querying and consuming a sequence of change records in a transactional fashion.

Question 7: **Incorrect**

A table stream(also known as stream) itself does ***not*** contain any table data

- ( )

TRUE
(Correct)

- FALSE
(Incorrect)

Explanation
This is very important to know.

Note that a stream itself does ***not*** contain any table data. A stream only stores the **offset** for the source table and returns CDC records by leveraging the versioning history for the source table. When the first stream for a table is created, a pair of hidden columns are added to the source table and begin storing change tracking metadata. These columns consume a small amount of storage. The CDC records returned when querying a stream rely on a combination of the *offset* stored in the stream and the *change tracking metadata* stored in the table.

https://docs.snowflake.com/en/user-guide/streams.html#overview-of-table-streams

Question 8: Skipped

The current offset for a stream can be determined by querying which function?

- SYSTEM$STREAM_GET_TABLE_TIMESTAMP
(Correct)

- SYSTEM$DISABLE_BEHAVIOR_CHANGE_BUNDLE
- SYSTEM$TASK_DEPENDENTS_ENABLE
- SYSTEM$USER_TASK_CANCEL_ONGOING_EXECUTIONS

Explanation
SYSTEM$STREAM_GET_TABLE_TIMESTAMP returns a timestamp indicating the transactional point when the stream contents were last consumed using a DML statement.

https://docs.snowflake.com/en/sql-reference/functions/system_stream_get_table_timestamp.html#system-stream-get-table-timestamp

Question 9: **Incorrect**

Streams support repeatable read isolation

- TRUE
(Correct)

- FALSE
  (Incorrect)

Explanation
Streams support repeatable read isolation. In repeatable read mode, multiple SQL statements within a transaction see the same set of records in a stream. This differs from the read committed mode supported for tables, in which statements see any changes made by previous statements executed within the same transaction, even though those changes are not yet committed.

https://docs.snowflake.com/en/user-guide/streams.html#repeatable-read-isolation

Question 10: **Incorrect**

A stream stores data in the same shape as the source table with some additional columns. Which are those additional columns?

- ☐ METADATA$ACTION
  (Correct)
- ☐ METADATA$ISUPDATE
  (Correct)
- ☑ METADATA$ROW_ID
  (Correct)
- ☐ METADATA$COLUMN_ID

Explanation
A stream stores data in the same shape as the source table (i.e. the same column names and ordering) with the following additional columns:

METADATA$ACTION

Indicates the DML operation (INSERT, DELETE) recorded.

METADATA$ISUPDATE

Indicates whether the operation was part of an UPDATE statement. Updates to rows in the source table are represented as a pair of DELETE and INSERT records in the stream with a metadata column METADATA$ISUPDATE values set to TRUE.

Note that streams record the differences between two offsets. If a row is added and then updated in the current offset, the delta change is a new row. The METADATA$ISUPDATE row records a FALSE value.

METADATA$ROW_ID

Specifies the unique and immutable ID for the row, which can be used to track changes to specific rows over time.

Question 11: Skipped

What are the various types of table streams?

- Standard
  (Correct)
- Append-only
  (Correct)
- Insert-only
  (Correct)
- Upsert

Explanation
https://docs.snowflake.com/en/user-guide/streams.html#types-of-streams

Please note insert-only stream is in preview mode

The following stream types are available based on the metadata recorded by each:

Standard

A standard (i.e. delta) table stream tracks all DML changes to the source table, including inserts, updates, and deletes (including table truncates). This stream type performs a join on inserted and deleted rows in the change set to provide the row level delta. As a net effect, for example, a row that is inserted and then deleted between two transactional points of time in a table is removed in the delta (i.e. is not returned when the stream is queried).

Append-only

An append-only table stream tracks row inserts only. Update and delete operations (including table truncates) are not recorded. For example, if 10 rows are inserted into a table and then 5 of those rows are deleted before the offset for an append-only stream is advanced, the stream records 10 rows.

An append-only stream returns the appended rows only and therefore can be much more performant than a standard stream for extract, load, transform (ELT) and similar scenarios that depend exclusively on row inserts. For example, the source table can be truncated immediately after the rows in an append-only stream are consumed, and the record deletions do not contribute to the overhead the next time the stream is queried or consumed.

Insert-only

Supported on external tables only. An insert-only stream tracks row inserts only; they do not record delete operations that remove rows from an inserted set (i.e. no-ops). For example, in-between any two offsets, if File1 is removed from the cloud storage location referenced by the external table, and File2 is added, the stream returns records for the rows in File2 only. Unlike when tracking CDC data for standard tables, Snowflake cannot access the historical records for files in cloud storage.

Question 12: Skipped

How will you know if a stream has become stale?

- Execute `DESCRIBE STREAM` or `SHOW STREAMS`
  (Correct)
- Execute `SHOW TABLE STREAMS`
- Execute `SHOW PIPES`

Explanation
To determine whether a stream has become stale, execute the DESCRIBE STREAM or SHOW STREAMS command. In the command output, when the STALE column value for the stream is TRUE, the stream is stale.

Question 13: Skipped

What are the two ways available in snowflake before change tracking metadata is recorded for a table?

- Change tracking is enabled on the table (using ALTER TABLE … CHANGE_TRACKING = TRUE)
  (Correct)
- A stream is created for the table (using CREATE STREAM)
  (Correct)
- Enable CDC on the table

Explanation

Currently, ***either*** of the following must be true before change tracking metadata is recorded for a table:

1. Change tracking is enabled on the table (using ALTER TABLE … CHANGE_TRACKING = TRUE).

2. A stream is created for the table (using CREATE STREAM).

Either option adds a pair of hidden columns to the table and begins storing change tracking metadata. The columns consume a small amount of storage.

No change tracking metadata for the table is available for the period before one of these conditions is satisfied.

Question 14: Skipped

Which role permissions are required to create and manage streams ? Please specify the OBJECT and the PERMISSION.

- ☐ DATABASE - USAGE
  (Correct)
- ☐ SCHEMA - USAGE, CREATE STREAM
  (Correct)
- ☐ TABLE - SELECT
  (Correct)
- ☐ TABLE - DELETE

Explanation
https://docs.snowflake.com/en/user-guide/streams.html#required-access-privileges

Question 15: Skipped

Streams ensure **exactly once semantics** for new or changed data in a table

- ○ TRUE
  (Correct)
- ○ FALSE

Explanation
Tasks can be combined with table streams for continuous ELT workflows to process recently changed table rows. **Streams ensure exactly once semantics for new or changed data in a table.**

https://docs.snowflake.com/en/user-guide/tasks-intro.html#introduction-to-tasks

Question 16: Skipped

Lets say that you have scheduled a task. At time T0, the task has started executing. If it is still executing when the next scheduling execution time occurs, what happens to that task

- It will be skipped
  (Correct)
- It will wait for the previous task to complete
- It will be aborted

Explanation
Snowflake ensures only one instance of a task with a schedule (i.e. a standalone task or the root task in a tree of tasks) is executed at a given time. If a task is still running when the next scheduled execution time occurs, then that scheduled time is skipped.

Question 17: Skipped

Which role or privileges are required to view TASK_HISTORY?

- Account administrator (i.e. users with the ACCOUNTADMIN role)
  (Correct)
- Task owner (i.e. role that has the OWNERSHIP privilege on a task)
  (Correct)
- Any role that has the global MONITOR EXECUTION privilege
  (Correct)
- Any one with SYSADMIN role

Explanation
The following roles (or roles with the specified privileges) can use SQL to view the task history within a specified date range:

Account administrator (i.e. users with the ACCOUNTADMIN role).

Task owner (i.e. role that has the OWNERSHIP privilege on a task).

Any role that has the global MONITOR EXECUTION privilege.

Question 18: Skipped

You want to query the history of task usage within a specified date range. What will you do?

- You will query the TASK_HISTORY table function **(Correct)**
- You will run 'SHOW TASK HISTORY'
- You will run 'DESCRIBE TASKS'

Explanation
TASK_HISTORY

This table function can be used to query the history of task usage within a specified date range. The function returns the history of task usage for your entire Snowflake account or a specified task.

```
TASK_HISTORY(
      [ SCHEDULED_TIME_RANGE_START => <constant_expr> ]
      [, SCHEDULED_TIME_RANGE_END => <constant_expr> ]
      [, RESULT_LIMIT => <integer> ]
      [, TASK_NAME => '<string>' ] )
```

Retrieve the execution history for tasks in the account within a specified 30 minute block of time within the past 7 days:

```
select *
  from table(information_schema.task_history(
    scheduled_time_range_start=>to_timestamp_ltz('2018-11-9 12:00:00.000 -0700'),
    scheduled_time_range_end=>to_timestamp_ltz('2018-11-9 12:30:00.000 -0700')));
```

Question 19: Skipped

Which SNOWFLAKE object stores a generated identity and access management (IAM) entity for your external cloud storage, along with an optional set of allowed or blocked storage locations (Amazon S3, Google Cloud Storage, or Microsoft Azure)?

STORAGE INTEGRATION
(Correct)

- STORAGE GROUP
- STORAGE CLUSTERS

Explanation
A storage integration is a Snowflake object that stores a generated identity and access management (IAM) entity for your external cloud storage, along with an optional set of allowed or blocked storage locations (Amazon S3, Google Cloud Storage, or Microsoft Azure). Cloud provider administrators in your organization grant permissions on the storage locations to the generated entity. This option allows users to avoid supplying credentials when creating stages or when loading or unloading data

Question 20: Skipped

A single storage integration can support multiple external stages

- TRUE
(Correct)

- FALSE

Explanation
A single storage integration can support multiple external stages. The URL in the stage definition must align with the storage location specified for the STORAGE_ALLOWED_LOCATIONS parameter.

Please see example below

```
create storage integration s3_int
  type = external_stage
  storage_provider = s3
  storage_aws_role_arn = 'arn:aws:iam::001234567890:role/myrole'
  enabled = true
  storage_allowed_locations = ('s3://mybucket1/path1/', 's3://mybucket2/path2/');
create storage integration gcs_int
  type = external_stage
  storage_provider = gcs
  enabled = true
  storage_allowed_locations = ('gcs://mybucket1/path1/', 'gcs://mybucket2/path2/'
);
```

```
create storage integration azure_int

  type = external_stage

  storage_provider = azure

  enabled = true

  azure_tenant_id = '<tenant_id>'

  storage_allowed_locations = ('azure://myaccount.blob.core.windows.net/mycontain
er/path1/', 'azure://myaccount.blob.core.windows.net/mycontainer/path2/');
```

Question 21: Skipped

What will you use to query Snowflake data loading history for the last 365 days (1 year)?

- COPY_HISTORY View
  (Correct)
- LOAD_HISTORY View from INFORMATION SCHEMA
- INSERT_HISTORY View

Explanation
COPY_HISTORY View

This Account Usage view can be used to query Snowflake data loading history for the last 365 days (1 year). The view displays load activity for both COPY INTO <table> statements and continuous data loading using Snowpipe. The view avoids the 10,000 row limitation of the LOAD_HISTORY View.

https://docs.snowflake.com/en/sql-reference/account-usage/copy_history.html#copy-history-view

LOAD_HISTORY View from Information schema

This Information Schema view enables you to retrieve the history of data loaded into tables using the COPY INTO <table> command within the last 14 days. The view displays one row for each file loaded.

https://docs.snowflake.com/en/sql-reference/info-schema/load_history.html#load-history-view

Please also read about the LOAD_HISTORY view from ACCOUNT USAGE schema

https://docs.snowflake.com/en/sql-reference/account-usage/load_history.html#load-history-view

Question 22: Skipped

Which feature in snowflake allows you to load data in micro batches?

- SNOWPIPE
  (Correct)
- COPY INTO COMMAND
- LOAD COMMAND

Explanation
Snowpipe enables loading data from files as soon as they're available in a stage. This means you can load data from files in **micro-batches**, making it available to users within minutes, rather than manually executing COPY statements on a schedule to load larger batches.

Question 23: Skipped

Load history for bulk data load is stored in the metadata of the target table for 64 days

- TRUE
  (Correct)
- FALSE

Explanation
Bulk data load

Stored in the metadata of the target table for 64 days. Available upon completion of the COPY statement as the statement output.

Snowpipe

Stored in the metadata of the pipe for 14 days. Must be requested from Snowflake via a REST endpoint, SQL table function, or ACCOUNT_USAGE view.

https://docs.snowflake.com/en/user-guide/data-load-snowpipe-intro.html#load-history

Question 24: Skipped

Which one of the below requires an user-specified warehouse to execute COPY statements

- Bulk data load
  (Correct)

- Snowpipe
- Both

Explanation
Bulk data load

Requires a user-specified warehouse to execute COPY statements.

Snowpipe

Uses Snowflake-supplied compute resources.

Question 25: Skipped

When you remove the predecessor of a child task, what may happen to the child task. Select two possibilities?

- The child task is removed from the system
- The child task may become the root task
(Correct)
- The child task may become a stand-alone task
(Correct)

Explanation
Tasks is a very important topic for the exam. I have added a lot of questions now on tasks across the question sets. Please ensure that you read the task chapter thoroughly.

Severing the Link Between Predecessor and Child Tasks

If the predecessor for a child task is removed (using ALTER TASK … REMOVE AFTER), then the former child task becomes either a standalone task or a root task, depending on whether other tasks identify this former child task as their predecessor.

If a predecessor task is dropped (using DROP TASK), or if ownership of a predecessor task is transferred to another role (using GRANT OWNERSHIP), then all former child tasks that identified this task as the predecessor become either standalone tasks or root tasks, depending on whether other tasks identify these former child tasks as their predecessor.

Question 26: Skipped

While loading data through the COPY command, you can transform the data. Which of the below transformations are allowed?

- [ ] TRUNCATECOLUMNS
(Correct)
- [ ] REORDER COLUMNS
(Correct)
- [ ] OMIT COLUMNS
(Correct)
- [ ] CAST
(Correct)
- [ ] AGGREGATE

Explanation
Snowflake supports transforming data while loading it into a table using the COPY INTO <table> command, dramatically simplifying your ETL pipeline for basic transformations. This feature helps you avoid the use of temporary tables to store pre-transformed data when reordering columns during a data load.

The COPY command supports:

`Column reordering, column omission, and casts using a SELECT statement.` There is no requirement for your data files to `have the same number and ordering of columns as your target table`.

The `ENFORCE_LENGTH | TRUNCATECOLUMNS` option, which can truncate text strings that exceed the target column length.

`Currently, these copy options support CSV data only.`

Question 27: Skipped

Which of the below is true about snowflake architecture

- ( ) One node shared data
- ( ) Multi Cluster Shared data
(Correct)
- ( )

Multi Node distributed data

Explanation

If you miss this question, it will hurt me a lot:(

You do not need to remember this. Ask yourself the question. How does compute work in Snowflake? Virtual Warehouse, right? What are Virtual Warehouses? They are a cluster of nodes(like EC2 instances if on AWS). Each virtual warehouse is an MPP compute cluster composed of `multiple compute nodes allocated by Snowflake from a cloud provider`. A Large WAREHOUSE has 8 nodes in a cluster. Each node has 8 cores so total 64 cores. All these nodes refer to a single source of data and that data is shared by all the nodes. Snowflake is architected in such a way that, all the nodes have access to the same data and then access it without stepping on each others toes.

Below is a list of all the warehouse sizes and their details.

Warehouse Size	Servers / Cluster	Credits / Hour	Credits / Second	Notes
X-Small	1	1	0.0003	Default size for warehouses created using CREATE WAREHOUSE.
Small	2	2	0.0006	
Medium	4	4	0.0011	
Large	8	8	0.0022	
X-Large	16	16	0.0044	Default for warehouses created in the web interface.
2X-Large	32	32	0.0089	
3X-Large	64	64	0.0178	
4X-Large	128	128	0.0356	

Question 28: Skipped

Snowflake supports multi-factor authentication (MFA) to provide increased login security for users connecting to Snowflake. At this time, Snowflake users are not automatically enrolled in MFA.

- TRUE
  (Correct)
- FALSE

Explanation

MFA is enabled on a per-user basis; however, at this time, users are not automatically enrolled in MFA. To use MFA, users must enroll themselves.

https://docs.snowflake.com/en/user-guide/security-mfa.html#multi-factor-authentication-mfa

Question 29: Skipped

Select two true statements about fail safe from the following

- Fail safe is same as time travel
- Fail-safe provides a (non-configurable) 7-day period during which historical data is recoverable by Snowflake
  (Correct)
- Fail-safe provides a (non-configurable) 90-day period during which historical data is recoverable by Snowflake
- Fail-safe ensures historical data is protected in the event of a system failure or other catastrophic event, e.g. a hardware failure or security breach
  (Correct)

Explanation
Separate and distinct from Time Travel, Fail-safe ensures historical data is protected in the event of a system failure or other catastrophic event, e.g. a hardware failure or security breach.

Fail-safe provides a (non-configurable) 7-day period during which historical data is recoverable by Snowflake. This period starts immediately after the Time Travel retention period ends.

https://docs.snowflake.com/en/user-guide/data-failsafe.html#understanding-viewing-fail-safe

Very important to Note:

```
Fail-safe is not provided as a means for accessing historical data after the Time Travel retention period has ended. It is for use only by Snowflake to recover data that may have been lost or damaged due to extreme operational failures.
```

Question 30: Skipped

When you load a table from a file through COPY command, the file should have the same number and order of columns as the table

- TRUE
- FALSE
  (Correct)

Explanation
Note that the actual field/column order in the data files can be different from the column order in the target table. It is only important that the SELECT list maps fields/columns in the data files to the ***corresponding*** columns in the table.

https://docs.snowflake.com/en/sql-reference/sql/copy-into-table.html#transformation-parameters

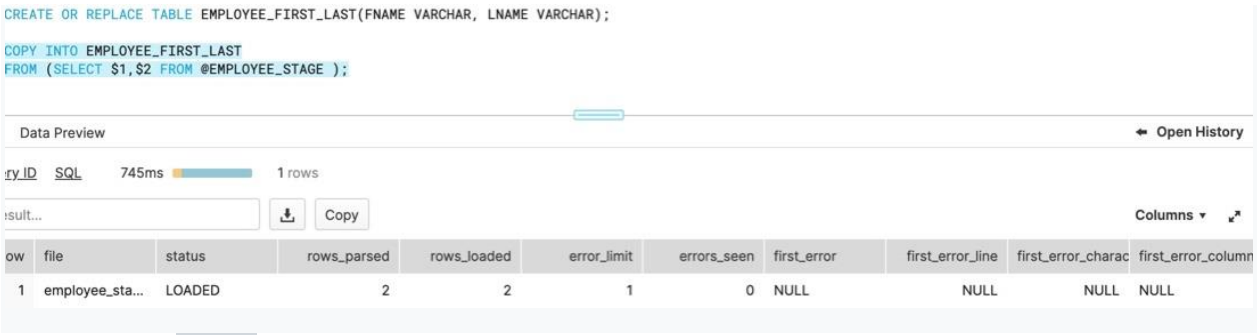
Try it out yourself. You will never forget it.

1. Load a snowflake stage with a file using a PUT command from SNOWSQL

2. In my case I have already loaded an employee file into stage and it looks like as below

3. Do a select on the stage. Please note how I wrote the select. This is how you do select from stage

4. I will now create a table with only first name and last name and load this data into the table



Question 31: Skipped

Failsafe period for temporary and transient table is zero

- TRUE
(Correct)
- FALSE

Explanation

```
This is a question that you need to remember. temporary table lives till the session is alive, so no point in having fail safe, transient tables also do not have fail safe and hence the name:)
```

To help manage the storage costs associated with Time Travel and Fail-safe, Snowflake provides two table types, temporary and transient, which do not incur the same fees as standard (i.e. permanent) tables:

Transient tables can have a Time Travel retention period of either 0 or 1 day.

Temporary tables can also have a Time Travel retention period of 0 or 1 day; however, this retention period ends as soon as the table is dropped or the session in which the table was created ends.

```
Transient and temporary tables have no Fail-safe period.
```

Question 32: Skipped

When the snowflake staged release happens, which accounts get the updates first

- Enterprise or above
- Trial accounts
- Premier accounts
- Accounts who has opted for early access

(Correct)

Explanation
If you have multiple Enterprise Edition (or higher) accounts, you can designate one or more of these accounts to take advantage of the 24-hour period between the early access and final stages for new releases. This can be particularly useful if you maintain separate accounts for development/testing and production.

To designate an account for early access, please contact Snowflake Support.

Once you have designated one or more accounts for early access, you can implement a testing framework similar to the following:

Use CURRENT_VERSION (or a UDF that returns similar results) to verify when your early access account(s) are on the new release.

Use your early access account(s) to test your production workloads against the new release.

If any issues are encountered, notify Snowflake Support, who can work with you to prevent the issues from disrupting your other accounts.

Early access is not required or recommended for all organizations with Enterprise Edition accounts; Snowflake's rigorous release testing and monitoring during deployments is usually sufficient to prevent most issues. Early access is intended primarily for organizations that desire added certainty that their production accounts will not be affected by changes in new releases.

Question 33: Skipped

After you have granted a share to a consumer, you added more objects to the share. You have provided the required access by using the GRANT <privilege> … TO SHARE command.

Please select the true statement

- The added objects will be accessible to the consumers immediately
  (Correct)
- You will need to recreate the share

Explanation

https://docs.snowflake.com/en/user-guide/data-sharing-provider.html#adding-objects-to-a-share

Adding Objects to a Share

You can add objects to an existing share at any time using the GRANT <privilege> … TO SHARE command. Any objects that you add to a share are instantly available to the consumers accounts who have created databases from the share. For example, if you add a table to a share, users in consumer accounts can query the data in the table as soon as the table is added to the share

Question 34: Skipped

For federated authentication and MFA, what is the minimum snowflake edition required?

- Business critical
- Enterprise
- Premier
- Standard
  (Correct)

Question 35: Skipped

A warehouse can be assigned to a single resource monitor only

- TRUE
  (Correct)
- FALSE

Explanation
Assignment of Resource Monitors

A single monitor can be set at the account level to control credit usage for all warehouses in your account.

In addition, a monitor can be assigned to one or more warehouses, thereby controlling the credit usage for each assigned warehouse. Note, however, `that a warehouse can be assigned to only a single resource monitor.`

Question 36: Skipped

Select the objects which do not have a fail safe period

- Transient table
(Correct)
- Temporary table
(Correct)
- Permanent table

Question 37: Skipped

When staging uncompressed files in a Snowflake stage, the files are automatically compressed using gzip, unless compression is explicitly disabled.

- TRUE
(Correct)
- FALSE

Explanation

Feature	Supported	Notes
Uncompressed files	gzip	When staging uncompressed files in a Snowflake stage, the files are automatically compressed using gzip, unless compression is explicitly disabled.
Already-compressed files	gzip bzip2 deflate raw_deflate	Snowflake can automatically detect any of these compression methods or you can explicitly specify the method that was used to compress the files.
	Brotli Zstandard	Auto-detection is ***not*** yet supported for these methods; when staging or loading files compressed with either of these methods, you must explicitly specify the compression method that was used.

Question 38: Skipped

COPY does not purge loaded files from the location. What option will you specify to purge files after successful load of a table from table stage

- PURGE=TRUE
(Correct)
- DELETE=TRUE
- REMOVE=TRUE
- TRUNCATE=TRUE

Explanation

Purging Files After Loading

Load files from a table's stage into the table and purge files after loading. By default, COPY does not purge loaded files from the location. To purge the files after loading:

Set `PURGE=TRUE` for the table to specify that all files successfully loaded into the table are purged after loading:

```
alter table mytable set stage_copy_options = (purge = true);


copy into mytable;
```

You can also override any of the copy options directly in the COPY command:

```
copy into mytable purge = true;
```

Question 39: Skipped

You have specified PURGE=TRUE in your COPY command. Due to a network failure the data files could not be purged after successful load. You will get an error returned by snowflake due to this failure.

- TRUE
- FALSE **(Correct)**

Explanation

`PURGE = TRUE | FALSE`

Boolean that specifies whether to remove the data files from the stage automatically after the data is loaded successfully.

If this option is set to `TRUE`, note that a best effort is made to remove successfully loaded data files. `If the purge operation fails for any reason, no error is returned currently`. We recommend that you list staged files periodically (using LIST) and manually remove successfully loaded files, if any exist.

Default: `FALSE`

https://docs.snowflake.com/en/sql-reference/sql/copy-into-table.html#copy-options-copyoptions

Question 40: Skipped

You can specify the below action to perform when an error is encountered while loading data from a file

- **CONTINUE**
  (Correct)
- SKIP_FILE
  (Correct)
- SKIP_FILE_<*num*>
  (Correct)
- SKIP_FILE_<*num*>%
  (Correct)
- ABORT_STATEMENT
  (Correct)
- ERROR_STATEMENT

Explanation

Copy Options (`copyOptions`)

You can specify one or more of the following copy options (separated by blank spaces, commas, or new lines):

```
ON_ERROR = CONTINUE | SKIP_FILE | SKIP_FILE_num | SKIP_FILE_num% |
ABORT_STATEMENT
```

String (constant) that specifies the action to perform when an error is encountered while loading data from a file:

Supported Values	Notes
`CONTINUE`	Continue loading the file. The COPY statement returns an error message for a maximum of one error encountered per data file. Note that the difference between the ROWS_PARSED and ROWS_LOADED column values represents the number of rows that include detected errors. However, each of these rows could include multiple errors. To view all errors in the data files, use the VALIDATION_MODE parameter or query the VALIDATE function.
`SKIP_FILE`	Skip file if any errors encountered in the file.
`SKIP_FILE_<num>` (e.g. `SKIP_FILE_10`)	Skip file when the number of errors in the file is equal to or exceeds the specified number.
`SKIP_FILE_<num>%` (e.g. `SKIP_FILE_10%`)	Skip file when the percentage of errors in the file exceeds the specified percentage.
`ABORT_STATEMENT`	Abort the load operation if any error is encountered in a data file. Note that the load operation is ***not*** aborted if the data file cannot be found (e.g. because it does not exist or cannot be accessed).

Question 41: Skipped

What is the default action to perform when an error is encountered while loading data from a file using SNOWPIPE?

- ABORT_STATEMENT
- SKIP_FILE
  (Correct)
- CONTINUE

Explanation
please remember this. This is different for SNOWPIPE and BULK LOAD using COPY

Default:

Bulk loading using COPY

`ABORT_STATEMENT`

Snowpipe

`SKIP_FILE`

Question 42: Skipped

Snowpipe is a serverless function

- TRUE
  (Correct)

- FALSE

Explanation
Automated data loads leverage event notifications for cloud storage to inform Snowpipe of the arrival of new data files to load. Snowpipe copies the files into a queue, from which they are loaded into the `target table in a continuous, serverless fashion` based on parameters defined in a specified pipe object.

Question 43: Skipped

Which character below identifies a table stage

- 1. @%

  (Correct)
- 1. @
- 1. %

Explanation
Unloading Data to a Table Stage

Use the COPY INTO <location> command to unload all the rows from a table into one or more files in the stage for the table. The following example unloads data files to the stage using the named `my_csv_unload_format` file format created in Preparing to Unload Data. The statement prefixes the unloaded file(s) with `unload/` to organize the files in the stage:

For example:

```
copy into @%mytable/unload/ from mytable file_format = (format_name = 'my_csv_unl
oad_format' compression = none);
```

`Note that the @% character combination identifies a table stage.`

Question 44: Skipped

Which permissions are required by Snowflake in AWS S3 to unload data to S3?

- `s3:DeleteObject`
  (Correct)
- `s3:PutObject`
  (Correct)

`s3:ListObject`

Explanation
Configuring an S3 Bucket for Unloading Data

Snowflake requires the following permissions on an S3 bucket and folder to create new files in the folder (and any sub-folders):

`s3:DeleteObject`

`s3:PutObject`

As a best practice, Snowflake recommends configuring a storage integration object to delegate authentication responsibility for external cloud storage to a Snowflake identity and access management (IAM) entity.

Question 45: Skipped

If you recreate a storage integration after it has been linked to one or more stages, you must reestablish the association between each stage by running the ALTER STAGE command.

- TRUE
(Correct)
- FALSE

Explanation
If you must recreate a storage integration after it has been linked to one or more stages, you must reestablish the association between each stage and the storage integration by executing ALTER STAGE *stage_name* SET STORAGE_INTEGRATION = *storage_integration_name*, where:

stage_name is the name of the stage.

storage_integration_name is the name of the storage integration.

Question 46: Skipped

A single storage integration can support multiple external stages

- TRUE
(Correct)
- FALSE

Explanation
Creates a new storage integration in the account or replaces an existing integration.

A storage integration is a Snowflake object that stores a generated identity and access management (IAM) entity for your external cloud storage, along with an optional set of allowed or blocked storage locations (Amazon S3, Google Cloud Storage, or Microsoft Azure). Cloud provider administrators in your organization grant permissions on the storage locations to the generated entity. This option allows users to avoid supplying credentials when creating stages or when loading or unloading data.

`A single storage integration can support multiple external stages.` The URL in the stage definition must align with the storage location specified for the STORAGE_ALLOWED_LOCATIONS parameter.

Question 47: Skipped

What is the storage hierarchy in snowflake?

- Account->Database->Schemas->Objects
(Correct)
- Organization->folder->project->resource
- Database->Schemas->Objects

Explanation
Snowflake storage hierarchy starts with an account. A customer can have as many accounts as possible, but each account will have a single URL and each account is deployed in a single cloud provider platform. Also an account cannot be multi-region, it is deployed on a singe region and a single snowflake edition is attached to the account

Databases are within accounts and each database belongs to a single account. While they can be replicated to other accounts they cannot span accounts

Schemas are contained within a database. A schema can be attached to a single database only. Again while schema can be replicated to other database and accounts it cannot span accounts or databases

All objects including tables, views, file formats, sequences, UDFs and stored procedures are contained within the schema

Question 48: Skipped

If you logon to snowflake using https://mycompany.snowflakecomputing.com/, then what is the cloud provider and region you are logging into?

- AWS and us-west-2
  (Correct)
- Azure and us-east-1
- AWS and us-east-2

Explanation
If an account has just the account name followed by snowflakecomputing.com, it means that the cloud provider is AWS and the region is us-west-2. This is the first cloud provider and region where snowflake was deployed

Question 49: Skipped

What table types are available in snowflake

- PERMANENT
  (Correct)
- TEMPORARY
  (Correct)
- TRANSIENT
  (Correct)
- EXTERNAL
  (Correct)
- INTERNAL

Explanation
Snowflake supports the below type of tables. It is important to know when to use which table type

1. PERMANENT - This is the default standard table type. It has time travel turned on. It also has a fail safe period of 7 days which is non-configurable

2. TEMPORARY - Only available for the session in which it is created. No time travel or fail safe period

3. TRANSIENT - It is same as PERMANENT except it does not have fail safe period. Time travel upto 1 day is possible

4. EXTERNAL - Stored in external stages and is read only. No cloning, time travel or fail safe is possible.

https://docs.snowflake.com/en/user-guide/tables-temp-transient.html

https://docs.snowflake.com/en/user-guide/tables-external-intro.html

Question 50: Skipped

What types of view are available in snowflake?

- REGULAR
  (Correct)
- SECURE
  (Correct)
- MATERIALIZED
  (Correct)
- HASH VIEW
- TABLE VIEW

Explanation

REGULAR Views are the default view type. The underlying DDL is accessible to any role who has access to the view

SECURE VIEWS are those where the query definition and the details are accessible to authorized users only. Please note that since the underlying DDL is not visible to the optimizer, this view type will be slower than a normal view

Materialized View - These are more like tables, results are store like it is in a table. Snowflake can auto refresh a materialized view. Please take extra care while using materialized views, it may increase your cost.

Question 51: Skipped

You are developing an ETL query. one of your ETL logic requires you to store the intermediate data temporarily. This data will be used only by your ETL. The data will not be required outside the ETL. If you want to optimize storage cost, what type of table will you create to store this data

- PERMANENT
- TRANSIENT
- TEMPORARY
  (Correct)
- EXTERNAL

Explanation
Please note that temporary tables are accessible within the session. It gets automatically destroyed when the session is destroyed. You create temporary tables if you have a use case to store data temporarily for a session only

Question 52: Skipped

Tick the statements which are true for fail safe

- Fail safe period is configurable
- Data can be recovered from a fail-safe storage only by snowflake employee
  (Correct)
- Using fail-safe you can create dev/test environments
- Tables with fail-safe turned on incurs additional storage costs compared to tables where fail-safe is not turned on
  (Correct)

Explanation
When fail-safe is turned on a table, the table requires additional storage to store historical data. That additional data storage cost is charged to the customer. Also to restore data from fail-safe storage, customer will need to contact snowflake

Question 53: Skipped

What are the various stages available in Snowflake

- USER
  (Correct)
- TABLE

(Correct)

- NAMED INTERNAL
  (Correct)
- NAMED EXTERNAL
  (Correct)
- ACCOUNT STAGE
- SCHEMA STAGE

Explanation
SNOWFLAKE supports three type of stages

1. USER

2. TABLE

3. NAMED

Named stages are again of two types internal and external

USER and Table stages are automatically available in snowflake and are not needed to be created or configured

https://docs.snowflake.com/en/user-guide/data-load-local-file-system-create-stage.html

Question 54: Skipped

Which of the following object types are stored within schema

- Tables
  (Correct)
- Views
  (Correct)
- User defined functions
  (Correct)
- File formats
  (Correct)

- Sequences
  (Correct)
- Roles

Explanation
Roles are at the account level and not at schema level

Question 55: Skipped

One of the government agencies has decided to use Snowflake for their datawarehouse. They require compliance with US federal privacy and security standards, such as FIPS 140-2and FedRAMP (Moderate Impact) . Which cloud providers can you choose to setup the snowflake instance?

- AWS
  (Correct)
- AZURE
  (Correct)
- GOOGLE
- IBM

Explanation
For government agencies that require compliance with US federal privacy and security standards, such as FIPS 140-2and FedRAMP (Moderate Impact), Snowflake supports the following regions in the US:

AWS - US East (Commercial Gov - N. Virginia)

AZURE - US Gov Virginia

https://docs.snowflake.com/en/user-guide/intro-regions.html#government-regions

Question 56: Skipped

Which AWS regions are supported by snowflake in US and Canada

- us-west-2
  (Correct)
- us-east-2
  (Correct)
-

- us-east-1
  (Correct)
- ca-central-1
  (Correct)
- us-central-2

Explanation
https://docs.snowflake.com/en/user-guide/intro-regions.html#general-regions

Question 57: Skipped

Which GCP regions are supported by Snowflake in US and Canada

- us-central1
  (Correct)
- us-central2
- us-east1

Explanation
Only US Central1 (Iowa) is supported in GCP

https://docs.snowflake.com/en/user-guide/intro-regions.html#general-regions

Question 58: Skipped

Which Azure regions are supported by Snowflake in US and Canada

- west-us-2
  (Correct)
- east-us-2
  (Correct)
- canada-central
  (Correct)
- west-us-1

Explanation
https://docs.snowflake.com/en/user-guide/intro-regions.html#general-regions

Question 59: Skipped

Your customer deals with extremely sensitive data like PHI data. They need to be compliant with HIPAA. What is the minimum edition that you will recommend

- STANDARD
- PREMIUM
- Business Critical Edition
  (Correct)
- Virtual Private Snowflake

Explanation
Business Critical Edition, formerly known as Enterprise for Sensitive Data (ESD), offers even higher levels of data protection to support the needs of organizations with extremely sensitive data, particularly PHI data that must comply with HIPAA regulations.

It includes all the features and services of Enterprise Edition, with the addition of enhanced security and data protection. In addition, database failover/failback adds support for business continuity and disaster recovery.

https://docs.snowflake.com/en/user-guide/intro-editions.html#business-critical-edition

Question 60: Skipped

Your customer required the highest level of data protection and needs a completely isolated environment for snowflake. Which edition will you recommend

- STANDARD
- PREMIER
- Virtual Private Snowflake
  (Correct)
- Business Critical Edition

Explanation
Virtual Private Snowflake offers our highest level of security for organizations that have the strictest requirements, such as financial institutions and any other large enterprises that collect, analyze, and share highly sensitive data.

It includes all the features and services of Business Critical Edition, but in a completely separate Snowflake environment, isolated from all other Snowflake accounts (i.e. VPS accounts do not share any resources with accounts outside the VPS).

https://docs.snowflake.com/en/user-guide/intro-editions.html#virtual-private-snowflake-vps

Question 61: Skipped

Which of the below semi structured data types are supported by snowflake?

- VARIANT
(Correct)
- OBJECT
(Correct)
- ARRAY
(Correct)
- GEOGRAPHY

Explanation
Snowflake supports three types of semi structured data types

VARIANT - A tagged universal type, which can store values of any other type, including OBJECT and ARRAY, up to a maximum size of 16 MB compressed.

OBJECT - Used to represent collections of key-value pairs, where the key is a non-empty string, and the value is a value of VARIANT type. Snowflake does not currently support explicitly-typed objects.

ARRAY - Used to represent dense or sparse arrays of arbitrary size, where index is a non-negative integer (up to 2^31-1), and values have VARIANT type. Snowflake does not currently support fixed-size arrays or arrays of elements of a specific non-VARIANT type.

https://www.udemy.com/instructor/course/3459188/manage/practice-tests

Question 62: Skipped

You have a multi-threaded client program that interacts with snowflake. If you are following snowflake recommendations what will you do

- Use a separate connection for each thread
(Correct)
-

Execute the threads synchronously
(Correct)

- Execute the threads asynchronously
- Use one connection for all the threads

Explanation
Although multiple sessions cannot share the same transaction, multiple ***threads*** using a single connection share the same session, and thus share the same transaction. This can lead to unexpected results, such as one thread rolling back work done in another thread.

This situation can occur when a client application using a Snowflake driver (such as the Snowflake JDBC Driver) or connector (such as the Snowflake Connector for Python) is multi-threaded. If two or more threads share the same connection, then those threads also share the current transaction in that connection. A BEGIN, COMMIT, or ROLLBACK by one thread affects all threads using that shared connection. If the threads are running asynchronously, the results can be unpredictable.

Similarly, changing the AUTOCOMMIT setting in one thread affects the AUTOCOMMIT setting in all other threads that use the same connection.

Note that changing the state of the AUTOCOMMIT setting commits and existing transaction.

Snowflake recommends that multi-threaded client programs do at least one of the following:

Use a separate connection for each thread.

Note that even with separate connections, your code can still hit race conditions that generate unpredicable output; for example, one thread might delete data before another thread tries to update it.

Execute the threads synchronously rather than asynchronously, to control the order in which steps are performed.

https://docs.snowflake.com/en/sql-reference/transactions.html#transactions-and-multi-threading

Question 63: Skipped

Which database objects can be shared using snowflake's secure data sharing

- Tables

(Correct)

- ☐ External tables
  (Correct)
- ☐ Secure views
  (Correct)
- ☐ Secure materialized views
  (Correct)
- ☐ Secure UDFs
  (Correct)
- ☐ Roles

Explanation
Secure Data Sharing enables sharing selected objects in a database in your account with other Snowflake accounts. The following Snowflake database objects can be shared:

1. Tables

2. External tables

3. Secure views

4 Secure materialized views

5. Secure UDFs

Snowflake enables the sharing of databases through *shares*, which are created by data providers and “imported” by data consumers

https://docs.snowflake.com/en/user-guide/data-sharing-intro.html#introduction-to-secure-data-sharing

Additional Explanation based on a student's question

Only secure views or secure materialized views can be shared. This is because of security reasons and that is why it is also called secure data sharing. All objects must be secure except for the tables. With replication, any objects contained within a database can be replicated to another account. The disadvantage with this over data sharing is that you are actually replicating data which will incur storage cost, compute cost to replicate data and also data egress if the replication is happening to a different region. The other thing is that data sharing will allow you to see immediate changes to the data, data replication will have latency based on the frequency of replication. Data sharing is the recommended approach if you are in the same region or deployment.

You can also share a schema. You need to include a schema in the share as the objects reside in schema.

Question 64: Skipped

Your customer needs to setup Failover/Failback mechanism in their snowflake instance. Which edition of Snowflake will you suggest?

- STANDARD
- PREMIER
- BUSINESS CRITICAL(OR HIGHER)
  (Correct)

Explanation
Failover/Failback requires Business Critical (or higher)

https://docs.snowflake.com/en/user-guide/database-replication-failover.html

Question 65: Skipped

Currently in snowflake, replication is supported for which type of objects

- Databases
  (Correct)
- Users
- Roles
- Warehouses

Explanation
Currently, replication is supported for ***databases only***. Other types of objects in an account ***cannot*** be replicated. This list includes:

1. Users

2. Roles

3. Warehouses

4. Resource monitors

5. Shares

https://docs.snowflake.com/en/user-guide/database-replication-intro.html#other-objects-in-an-account

Question 66: Skipped

You have stored the below JSON in a table named car_sales as a variant column

```
{
  "customer": [
    {
      "address": "San Francisco, CA",
      "name": "Joyce Ridgely",
      "phone": "16504378889"
    }
  ],
  "date": "2017-04-28",
  "dealership": "Valley View Auto Sales",
  "salesperson": {
    "id": "55",
    "name": "Frank Beasley"
  },
  "vehicle": [
    {
      "extras": [
        "ext warranty",
        "paint protection"
      ],
      "make": "Honda",
      "model": "Civic",
      "price": "20275",
      "year": "2017"
    }
  ]
}
```

How will you query the table to get the dealership data?

- select src:dealership from car_sales;
  (Correct)
- select src:$dealership from car_sales;
- select dealership from car_sales;

Explanation
Insert a colon `:` between the VARIANT column name and any first-level element: `<column>:<level1_element>`

https://docs.snowflake.com/en/user-guide/querying-semistructured.html#traversing-semi-structured-data

Question 67: Skipped

One of your team members accidentally truncated a table in your snowflake database? What feature of snowflake will you use to restore data into the table

- TIME TRAVEL
  (Correct)
- FAIL SAFE
- UNDROP TABLE

Explanation
Snowflake Time Travel enables accessing historical data (i.e. data that has been changed or deleted) at any point within a defined period. It serves as a powerful tool for performing the following tasks:

1. Restoring data-related objects (tables, schemas, and databases) that may have been accidentally or intentionally deleted.

2. Duplicating and backing up data from key points in the past.

3. Analyzing data usage/manipulation over specified periods of time.

https://docs.snowflake.com/en/user-guide/data-time-travel.html#understanding-using-time-travel

Additional explanation:

One of my students pointed out why is undrop not a correct option for this question. So, in the exam you may get some tricky questions. Pay careful attention to what is being asked in the question. In this question, the table is truncated and not dropped. So , your table is still there, you just lost the data. The most appropriate answer in this case is to use time travel and get the data back.

Question 68: Skipped

What happens when you execute the below query

CREATE OR REPLACE TABLE MyTable_V2 **CLONE** MyTable

- MyTable_V2 is created and all the data from MyTable is copied into MyTable_V2
- MyTable_V2 is created and a batch job is submitted to copy data from MyTable
- Snowflake creates a new entry in the metadata store to keep track of the new clone. The existing micro partitions of table MyTable are also mapped to the new table. This is called zero copy clone.
  (Correct)

Explanation
Clone creates a copy of an existing object in the system. This command is primarily used for creating zero-copy clones of databases, schemas, and non-temporary tables; however, it can also be used to quickly/easily create clones of other schema objects (i.e. external stages, file formats, and sequences).

https://docs.snowflake.com/en/sql-reference/sql/create-clone.html#create-object-clone

Question 69: Skipped

Which of the below statements are true about snowflake data sharing?

- Consumers can query shared tables in the same query as their own tables
  (Correct)
- Producers can share data to other databases such as Postgres and MySQL
- You don't need a Snowflake Virtual Warehouse to query shared data
- Data Sharing is integrated with role-based access controls
  (Correct)

Explanation
https://docs.snowflake.com/en/user-guide/data-sharing-intro.html#how-does-secure-data-sharing-work

Question 70: Skipped

What are the default roles that snowflake automatically adds for any new account?

- AccountAdmin
  (Correct)
- SecurityAdmin
  (Correct)
- SysAdmin
  (Correct)
- Public
  (Correct)
- Security Admin

Explanation
By default Snowflake offers the following default roles:

1. AccountAdmin

2. SecurityAdmin

3. SysAdmin

4. Public

These are a good starting point, but won't be sufficient for most implementations, so we end up defining custom roles. Snowflake's recommendation is to create a hierarchy of custom roles with the top-most custom role assigned to the system role SYSADMIN. That way the system administrators will be able to manage all warehouses and databases while maintaining management of user and roles restricted to users granted the SECURITYADMIN or ACCOUNTADMIN roles.

https://support.snowflake.net/s/article/Quickly-Visualize-Snowflake-s-Roles-Grants-and-Privileges

Question 71: Skipped

What are the different caching mechanisms that are available in snowflake?

- Warehouse caching
  (Correct)
- Index Caching
- Metadata caching
  (Correct)
- Query result caching
  (Correct)

Explanation
Snowflake supports three types of caching

1. Metadata cache -

2. Query result cache

3. Warehouse cache

further reading:

https://blog.ippon.tech/innovative-snowflake-features-caching/#:~:text=Metadata%20Cache,-Fully%20Managed%20in&text=Snowflake%20automatically%20collects%20and%20manages,Table%20Size%20in%20Bytes

Question 72: Skipped

One of your user complained that a query is taking a long time to return any result. What techniques would you consider to improve performance?

- Create an index on the search columns
- Create cluster keys and turn on auto clustering on the table
  (Correct)
- Enable query hints on the table
- Define the partition keys for the table

Explanation

Automatic Clustering is the Snowflake service that seamlessly and continually manages all reclustering, as needed, of clustered tables.

Automatic Clustering eliminates the need for performing any of the following tasks:

1. Monitoring the state of clustered tables.

2. Instead, as DML is performed on these tables, Snowflake monitors and evaluates the tables to determine whether they would benefit from reclustering, and automatically reclusters them, as needed.

3. Designating warehouses in your account to use for reclustering.

4. Snowflake performs automatic reclustering in the background, and you do not need to specify a warehouse to use.

All you need to do is define a clustering key for each table (if appropriate) and Snowflake manages all future maintenance.

https://docs.snowflake.com/en/user-guide/tables-auto-reclustering.html

Question 73: Skipped

Which parameter will you use while creating a STORAGE integration to limit external stages that use the integration to reference one or more storage locations.

- STORAGE_ALLOWED_LOCATIONS
  (Correct)
- STORAGE_ALLOWED_SITES
- STORAGE_OBJECTS_ALLOWED_LOCATIONS

Explanation

```
STORAGE_ALLOWED_LOCATIONS = ('<cloud_specific_url>')
```

Explicitly limits external stages that use the integration to reference one or more storage locations (i.e. S3 bucket, GCS bucket, or Azure container). Supports a comma-separated list of URLs for existing buckets and, optionally, paths used to store data files for loading/unloading. Alternatively supports the `*` wildcard, meaning “allow access to all buckets and/or paths”.

https://docs.snowflake.com/en/sql-reference/sql/create-storage-integration.html#required-parameters

Question 74: Skipped

Which are the required parameters while creating a storage integration?

- [ ] Name
(Correct)
- [ ] Type
(Correct)
- [ ] Enabled
(Correct)
- [ ] Storage_Allowed_Locations
(Correct)
- [ ] Storage_blocked_locations

Explanation
https://docs.snowflake.com/en/sql-reference/sql/create-storage-integration.html#required-parameters

Required Parameters

`name`

String that specifies the identifier (i.e. name) for the integration; must be unique in your account.

In addition, the identifier must start with an alphabetic character and cannot contain spaces or special characters unless the entire identifier string is enclosed in double quotes (e.g. `"My object"`). Identifiers enclosed in double quotes are also case-sensitive.

For more details, see Identifier Requirements.

`TYPE = EXTERNAL_STAGE`

Specify the type of integration:

`EXTERNAL_STAGE`: Creates an interface between Snowflake and an external cloud storage location.

`ENABLED = TRUE | FALSE`

Specifies whether this storage integration is available for usage in stages.

`TRUE` allows users to create new stages that reference this integration. Existing stages that reference this integration function normally.

`FALSE` prevents users from creating new stages that reference this integration. Existing stages that reference this integration cannot access the storage location in the stage definition.

`STORAGE_ALLOWED_LOCATIONS = ('cloud_specific_url')`

Explicitly limits external stages that use the integration to reference one or more storage locations (i.e. S3 bucket, GCS bucket, or Azure container). Supports a comma-separated list of URLs for existing buckets and, optionally, paths used to store data files for loading/unloading. Alternatively supports the `*` wildcard, meaning "allow access to all buckets and/or paths".

Amazon S3

`STORAGE_ALLOWED_LOCATIONS = ('s3://bucket/path/', 's3://bucket/path/')`

`bucket` is the name of an S3 bucket that stores your data files (e.g. `mybucket`).

`path` is an optional case-sensitive path for files in the cloud storage location (i.e. files have names that begin with a common string) that limits access to a set of files. Paths are alternatively called *prefixes* or *folders* by different cloud storage services.

Google Cloud Storage

`STORAGE_ALLOWED_LOCATIONS = ('gcs://bucket/path/', 'gcs://bucket/path/')`

`bucket` is the name of a GCS bucket that stores your data files (e.g. `mybucket`).

`path` is an optional case-sensitive path for files in the cloud storage location (i.e. files have names that begin with a common string) that limits access to a set of files. Paths are alternatively called *prefixes* or *folders* by different cloud storage services.

Microsoft Azure

`STORAGE_ALLOWED_LOCATIONS = ('azure://account.blob.core.windows.net/container/path/', 'azure://account.blob.core.windows.net/container/path/')`

`account` is the name of the Azure account (e.g. `myaccount`). Use the `blob.core.windows.net` endpoint for all supported types of Azure blob storage accounts, including Data Lake Storage Gen2.

`container` is the name of a Azure blob storage container that stores your data files (e.g. `mycontainer`).

`path` is an optional case-sensitive path for files in the cloud storage location (i.e. files have names that begin with a common string) that limits access to a set of files. Paths are alternatively called *prefixes* or *folders* by different cloud storage services.

Question 75: Skipped

Which command will you run to list all users and roles to which a role has been granted.

- ```
  1. SHOW GRANTS OF ROLE <role_name>
  ```

  (Correct)

- ```
  1. SHOW GRANTS TO ROLE <role_name>
  ```

- ```
  1. SHOW GRANTS ON ACCOUNT
  ```

Explanation
https://docs.snowflake.com/en/sql-reference/sql/show-grants.html#syntax

`SHOW GRANTS OF...`

`ROLE role_name`
  ```

Lists all users and roles to which the role has been granted.

Question 76: Skipped

If you plan to regularly unload similarly-formatted data, it is recommended to use named file formats

- TRUE
  (Correct)
- FALSE

Explanation
Named file formats are optional, but are ***recommended*** when you plan to regularly unload similarly-formatted data.

Question 77: Skipped

Which command will you run to view a list of files that have been unloaded to a stage named mystage.

- ```
  1. list @mystage;
  ```

  (Correct)
- ```
  1. show @mystage;
  ```
- ```
  1. display @mystage;
  ```

Explanation
Use the LIST command to view a list of files that have been unloaded to the stage:

```
list @mystage;


+---------------------------------+------+----------------------------------+---
---------------------------+

| name                              | size | md5                              | la
st_modified                |

|---------------------------------+------+----------------------------------+---
---------------------------|

| mystage/unload/data_0_0_0.csv.gz |  112 | 6f77daba007a643bdff4eae10de5bed3 | Mo
n, 11 Sep 2017 18:13:07 GMT |

+---------------------------------+------+----------------------------------+---
---------------------------+
```

Question 78: Skipped

Which character combination identifies a user stage
```

- @~
  (Correct)
- ~
- @

Explanation
Use the COPY INTO <location> command to unload all the rows from a table into one or more files in your stage. The following example unloads data files to your user stage using the named `my_csv_unload_format` file format created in Preparing to Unload Data. The statement prefixes the unloaded file(s) with `unload/` to organize the files in the stage:

For example:

```
copy into @~/unload/ from mytable file_format = (format_name = 'my_csv_unload_for
mat' compression = none);
```

`Note that the @~ character combination identifies a user stage.`

Question 79: Skipped

The Snowflake Web UI data loading wizard is only intended for loading small numbers of files of limited size (up to 50 MB).

- TRUE
  (Correct)
- FALSE

Explanation
The wizard is only intended for loading small numbers of files of limited size (up to 50 MB). This file size limit is intended to ensure better performance because browser performance varies from computer to computer and between different browser versions. Also, the memory consumption required to encrypt larger files might cause a browser to run out of memory and crash.

For loading larger files or large numbers of files, snowflake recommend using the Snowflake client, SnowSQL

Question 80: Skipped

If a query is running slow, to check if this is for an overloaded warehouse where in the Snowflake UI you will go to check the same

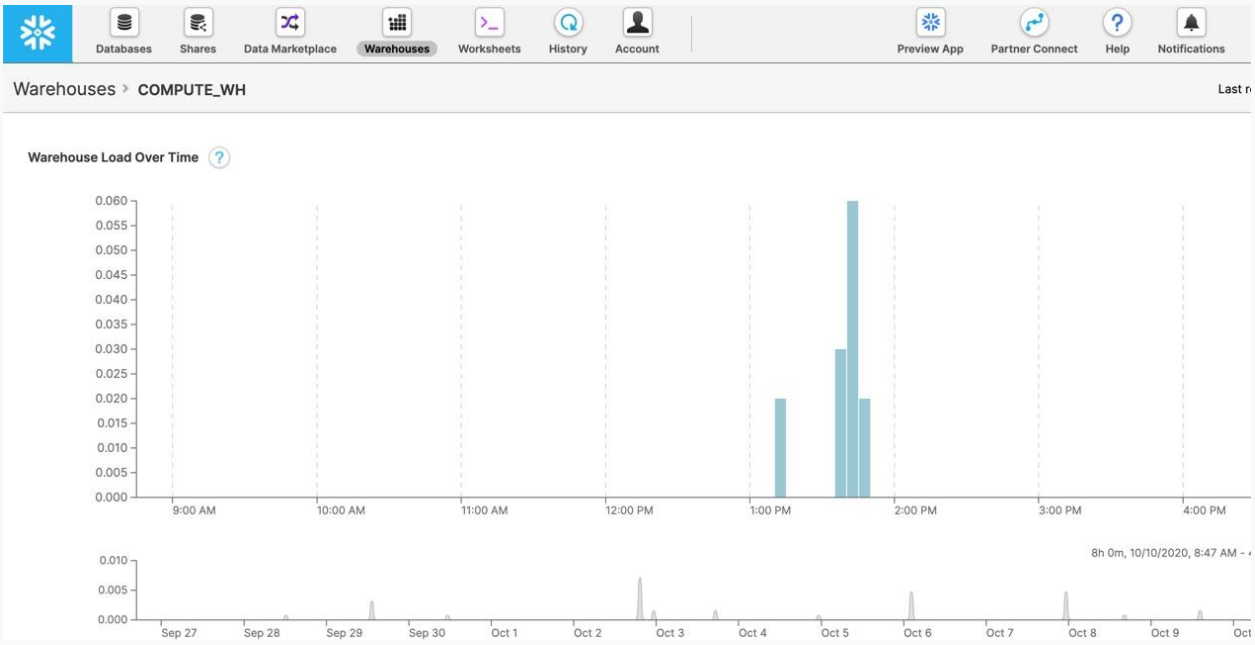
- Warehouse Tab

(Correct)

- Database Tab
- Query Tab

Explanation

It will be shown under the warehouse tab. if it is overloaded, the bars will be in yellow color



Question 81: Skipped

Which of the below mentioned clustering metadata for the micro-partitions is maintained by snowflake in a table?

- The total number of micro-partitions that comprise the table
  (Correct)
- The number of micro-partitions containing values that overlap with each other
  (Correct)
- The depth of the overlapping micro-partitions
  (Correct)
- None of the above

Explanation

Clustering Information Maintained for Micro-partitions

Snowflake maintains clustering metadata for the micro-partitions in a table, including:

1. The total number of micro-partitions that comprise the table.

2. The number of micro-partitions containing values that overlap with each other (in a specified subset of table columns).

3. The depth of the overlapping micro-partitions.

Question 82: Skipped

What would you check to see if a large table will benefit from explicitly defining a clustering key

- Clustering depth
  (Correct)
- Clustering percentage
- Clustering ratio

Explanation
Clustering Depth

The clustering depth for a populated table measures the average depth (`1` or greater) of the overlapping micro-partitions for specified columns in a table. The smaller the average depth, the better clustered the table is with regards to the specified columns.

Clustering depth can be used for a variety of purposes, including:

1. Monitoring the clustering “health” of a large table, particularly over time as DML is performed on the table.

2. Determining whether a large table would benefit from explicitly defining a clustering key.

A table with no micro-partitions (i.e. an unpopulated/empty table) has a clustering depth of `0`.

Question 83: Skipped

If query performance degrades over time, the table is likely no longer well-clustered and may benefit from clustering

- TRUE
  (Correct)

- FALSE

Explanation
The clustering depth for a table is ***not*** an absolute or precise measure of whether the table is well-clustered. Ultimately, query performance is the best indicator of how well-clustered a table is:

1. If queries on a table are performing as needed or expected, the table is likely well-clustered.

2. If query performance degrades over time, the table is likely no longer well-clustered and may benefit from clustering.

Question 84: Skipped

Which system functions are available in snowflake to view/monitor the clustering metadata for a table

- SYSTEM$CLUSTERING_DEPTH
(Correct)
- SYSTEM$CLUSTERING_INFORMATION (including clustering depth)
(Correct)
- CLUSTERING_DEPTH_VIEW

Explanation
Monitoring Clustering Information for Tables

To view/monitor the clustering metadata for a table, Snowflake provides the following system functions:

SYSTEM$CLUSTERING_DEPTH

SYSTEM$CLUSTERING_INFORMATION (including clustering depth)

Question 85: Skipped

Clustering is generally most cost-effective for tables that are queried frequently and do not change frequently

- TRUE
(Correct)
- FALSE

Explanation

The more frequently a table is queried, the more benefit clustering provides. However, the more frequently a table changes, the more expensive it will be to keep it clustered. Therefore, clustering is generally most cost-effective for tables that are queried frequently and do not change frequently.

Question 86: Skipped

Which of the below will you consider while choosing a cluster key

- Columns that are most actively used in selective filters
  (Correct)
- Columns frequently used in join predicates
  (Correct)
- Columns with extremely high cardinality
- Columns with extremely low cardinality

Explanation

Snowflake recommends prioritizing keys in the order below:

1. Cluster columns that are most actively used in selective filters. For many fact tables involved in date-based queries (for example “WHERE invoice_date > x AND invoice date <= y”), choosing the date column is a good idea. For event tables, event type might be a good choice, if there are a large number of different event types. (If your table has only a small number of different event types, then see the comments on cardinality below before choosing an event column as a clustering key.)

2. If there is room for additional cluster keys, then consider columns frequently used in join predicates, for example “FROM table1 JOIN table2 ON table2.column_A = table1.column_B”.

If you typically filter queries by two dimensions (e.g. `application_id` and `user_status` columns), then clustering on both columns can improve performance.

The number of distinct values (i.e. cardinality) in a column/expression is a critical aspect of selecting it as a clustering key. It is important to choose a clustering key that has:

A large enough number of distinct values to enable effective pruning on the table.

A small enough number of distinct values to allow Snowflake to effectively group rows in the same micro-partitions.

A column with very low cardinality (e.g. a column that indicates only whether a person is male or female) might yield only minimal pruning. At the other extreme, a column with very high cardinality (e.g. a column containing UUID or nanosecond timestamp values) is also typically ***not*** a good candidate to use as a clustering key directly.

Question 87: Skipped

Reclustering in Snowflake requires manual configuration

- False
(Correct)
- True

Explanation
Reclustering in Snowflake is automatic; no maintenance is needed.

https://docs.snowflake.com/en/user-guide/tables-auto-reclustering.html

Question 88: Skipped

Automatic reclustering in snowflake is triggered only if/when the table would benefit from the operation

- TRUE
(Correct)
- FALSE

Explanation
Optimal Efficiency

With Automatic Clustering, Snowflake internally manages the state of clustered tables, as well as the resources (servers, memory, etc.) used for all automated clustering operations. This allows Snowflake to dynamically allocate resources as needed, resulting in the most efficient and effective reclustering.

```
Also, Automatic Clustering does not perform any unnecessary reclustering.
Reclustering is triggered only if/when the table would benefit from the
operation.
```

Question 89: Skipped

Which of the below privileges are required to add clustering on a table

-

USAGE or Ownership on the schema and database
(Correct)

- MONITOR USAGE on the table
- ALTER privilege on the table

Explanation
To add clustering to a table, you must also have USAGE or OWNERSHIP privileges on the schema and database that contain the table.

Question 90: Skipped

The data stored in temporary table does not contribute to storage cost

- FALSE
(Correct)

- TRUE

Explanation
Data Storage Usage for Temporary Tables

For the duration of the existence of a temporary table, the data stored in the table contributes to the overall storage charges that Snowflake bills your account. To prevent any unexpected storage changes, particularly if you create large temporary tables in sessions that you maintain for periods longer than 24 hours, Snowflake recommends explicitly dropping these tables once they are no longer needed. You can also explicitly exit the session in which the table was created to ensure no additional charges are accrued.

Question 91: Skipped

You have executed a task. Due to some coding issue, the task went into a loop. How long the task will run before it is terminated by Snowflake

- 24 hours
- 60 minutes
(Correct)

- 30 minutes

Explanation
`There is a 60 minute default limit on a single run of a task`. This limitation was implemented as a safeguard against non-terminating tasks. Query

the TASK_HISTORY table function. If the task was canceled or exceeded the window scheduled for the task, the cause is often an undersized warehouse. Review the warehouse size and consider increasing it to fit within the schedule window or the one-hour limit.

Alternatively, consider increasing the timeout limit for the task by executing ALTER TASK … SET USER_TASK_TIMEOUT_MS = *<num>*.

A larger warehouse may not help if there are query parallelization issues. Consider looking at alternate ways to rewrite the SQL statement run by the task.

Question 92: Skipped

Snowflake supports below type of streams. Please select three

- ☐ Standard
  (Correct)
- ☐ Update-only
- ☐ Append-only
  (Correct)
- ☐ Insert-only
  (Correct)

Explanation
The following stream types are available based on the metadata recorded by each:

Standard

A standard (i.e. delta) table stream tracks all DML changes to the source table, including inserts, updates, and deletes (including table truncates). This stream type performs a join on inserted and deleted rows in the change set to provide the row level delta. As a net effect, for example, a row that is inserted and then deleted between two transactional points of time in a table is removed in the delta (i.e. is not returned when the stream is queried).

Append-only

An append-only table stream tracks row inserts only. Update and delete operations (including table truncates) are not recorded. For example, if 10 rows are inserted into a table and then 5 of those rows are deleted before the offset for an append-only stream is advanced, the stream records 10 rows.

An append-only stream returns the appended rows only and therefore can be much more performant than a standard stream for extract, load, transform (ELT) and similar scenarios that depend exclusively on row inserts. For example, the source table can be truncated immediately after the rows in an append-only stream are consumed, and the record deletions do not contribute to the overhead the next time the stream is queried or consumed.

Insert-only - `This is currently a preview feature`

Supported on external tables only. An insert-only stream tracks row inserts only; they do not record delete operations that remove rows from an inserted set (i.e. no-ops). For example, in-between any two offsets, if File1 is removed from the cloud storage location referenced by the external table, and File2 is added, the stream returns records for the rows in File2 only. Unlike when tracking CDC data for standard tables, Snowflake cannot access the historical records for files in cloud storage.

Overwritten files are essentially handled as new files: The old version of the file is removed from cloud storage, but the insert-only stream does not record the delete operation. The new version of the file is added to cloud storage, and the insert-only stream records the rows as inserts. The stream does not record the diff of the old and new file versions.

Question 93: Skipped

You have a schema with a table and a stream. You cloned the schema. What will happen to the stream

- Any unconsumed records in the stream (in the clone) are inaccessible **(Correct)**
- Any unconsumed records in the stream (in the clone) will also be accessible in the clone
- Any unconsumed records in the stream (in the clone) will automatically flow into the table

Explanation
Currently, when a database or schema that contains a source table and stream is cloned, any unconsumed records in the stream (in the clone) are inaccessible. This behavior is consistent with Time Travel for tables. If a table is cloned, historical data for the table clone begins at the time/point when the clone was created.

Question 94: Skipped

You have a materialized view and you want to track changes in the view. You can use streams to track that change

-

TRUE

- FALSE
  (Correct)

Explanation
Currently, streams cannot track changes in materialized views.

Question 95: Skipped

A Stream has been created on a table. A row in the table got inserted and subsequently got updated. The stream will capture both the events.

- TRUE
- FALSE
  (Correct)

Explanation
A stream maintains a point of time into the transactional versioned timeline of the source table, called an *offset*, which starts at the transactional point when the stream contents were last consumed using a DML statement. The stream provides the set of changes from the current offset to the current *transactional time* of the source table (i.e. the current version of the table). `The stream maintains only the delta of the changes; if multiple DML statements change a row, the stream records only the latest action taken on that row.`

Question 96: Skipped

You would like to execute a merge statement in snowflake every one hour. If you want to schedule this query using a snowflake object what will you use?

- TASK
  (Correct)

- STREAM
- SNOWPIPE

Explanation
Task Scheduling

There is no event source that can trigger a task; instead, a task runs on a schedule, which can be defined when creating a task (using CREATE TASK) or later (using ALTER TASK).

Snowflake ensures only one instance of a task with a schedule (i.e. a standalone task or the root task in a tree of tasks) is executed at a given time. If a task is still running when the next scheduled execution time occurs, then that scheduled time is skipped.

Question 97: Skipped

Which roles can use SQL to view the task history within a specified date range?

- Account administrator (i.e. users with the ACCOUNTADMIN role).
(Correct)
- Task owner (i.e. role that has the OWNERSHIP privilege on a task).
(Correct)
- Any role that has the global MONITOR EXECUTION privilege.
(Correct)
- SECURITYADMIN

Explanation
Viewing the Task History for Your Account

The following roles (or roles with the specified privileges) can use SQL to view the task history within a specified date range:

1. Account administrator (i.e. users with the ACCOUNTADMIN role).

2. Task owner (i.e. role that has the OWNERSHIP privilege on a task).

3. Any role that has the global MONITOR EXECUTION privilege.

To view the task history:

SQL

Query the TASK_HISTORY table function (in the Information Schema).

Question 98: Skipped

Insert-only streams are supported on external tables only and it is currently a preview feature

- TRUE
(Correct)
- FALSE

Explanation
Insert-only

Supported on external tables only. An insert-only stream tracks row inserts only; they do not record delete operations that remove rows from an inserted set (i.e. no-ops). For example, in-between any two offsets, if File1 is removed from the cloud storage location referenced by the external table, and File2 is added, the stream returns records for the rows in File2 only. Unlike when tracking CDC data for standard tables, Snowflake cannot access the historical records for files in cloud storage.

Overwritten files are essentially handled as new files: The old version of the file is removed from cloud storage, but the insert-only stream does not record the delete operation. The new version of the file is added to cloud storage, and the insert-only stream records the rows as inserts. The stream does not record the diff of the old and new file versions.

https://docs.snowflake.com/en/user-guide/streams.html#types-of-streams

Question 99: Skipped

When the owner role of a given task (i.e. the role with the OWNERSHIP privilege on the task) is deleted, what happens to the task?

- The task is “re-possessed” by the role that dropped the owner role
  (Correct)
- The task is deleted immediately
- The task is suspended immediately

Explanation
Dropping a Task Owner Role

When the owner role of a given task (i.e. the role with the OWNERSHIP privilege on the task) is deleted, the task is “re-possessed” by the role that dropped the owner role. This ensures that ownership moves to a role that is closer to the root of the role hierarchy. When a task is re-possessed, it is automatically paused, i.e., all executions currently in flight complete processing, but new executions will not be scheduled until the task is resumed explicitly by the new owner. The rationale for this is to prevent a user with access to a particular role from leaving behind tasks that suddenly execute with higher permissions when the role is removed.

If the role that a running task is executing under is dropped while the task is running, the task completes processing under the dropped role.

Question 100: Skipped

A simple tree of tasks is limited to a maximum of 1000 tasks total (including the root task) in a resumed state

- TRUE
  (Correct)
- FALSE

Explanation

A simple tree of tasks is limited to a `maximum of 1000 tasks total` (including the root task) in a resumed state. An individual task in the tree is limited to a single predecessor task; however, a task can have a maximum of 100 *child* tasks (i.e. other tasks that identify the task as a predecessor).

Currently, we cannot guarantee that only one instance of a task with a defined ***predecessor task*** is running at a given time.

Question 1: Skipped

Cloud services layer is one of the layers in snowflake. Which services are provided by this layer

- Authentication
  (Correct)
- Infrastructure management
  (Correct)
- Metadata management
  (Correct)
- Query parsing and optimization
  (Correct)
- Access control
  (Correct)
- Query processing

Explanation

The cloud services layer is a collection of services that coordinate activities across Snowflake. These services tie together all of the different components of Snowflake in order to process user requests, from login to query dispatch. The cloud services layer also runs on compute instances provisioned by Snowflake from the cloud provider.

Among the services in this layer:

Authentication

Infrastructure management

Metadata management

Query parsing and optimization

Access control

https://docs.snowflake.com/en/user-guide/intro-key-concepts.html#cloud-services

Question 2: Skipped

Two virtual warehouses can access the same data at the same time without any contention issues

- TRUE
  (Correct)
- FALSE

Question 3: Skipped

You are speaking to a CTO of an organization and would like to define the architecture of snowflake to him in two words. What will you use?

- Shared Disk
- Shared Nothing
- Shared Data
  (Correct)
- Shared Memory

Explanation
Snowflake's architecture is a hybrid of traditional shared-disk database architectures and shared-nothing database architectures. Similar to shared-disk architectures, Snowflake uses a central data repository for persisted data that is accessible from all compute nodes in the data warehouse. But similar to shared-nothing architectures, Snowflake processes queries using MPP (massively parallel processing) compute clusters where each node in the cluster stores a portion of the entire data set locally. This approach offers the data management simplicity of a shared-disk architecture, but with the performance and scale-out benefits of a shared-nothing architecture.

https://docs.snowflake.com/en/user-guide/intro-key-concepts.html

Question 4: Skipped

Which statements are true about a micro partition in snowflake

- Contiguous unit of storage
(Correct)
- Each micro partition contains between 50 MB and 500 MB of uncompressed data
(Correct)
- Organized in a columnar fashion
(Correct)
- Micro partitions can be updated using snowflake query

Explanation
Micro partitions are immutable which means once created they cannot be changed. If a row is updated in snowflake, the micro partition holding the row is copied into a new micro partition and the updated row is inserted in that micro partition. The olde micro partition is marked for deletion

All data in Snowflake tables is automatically divided into micro-partitions, which are contiguous units of storage. Each micro-partition contains between 50 MB and 500 MB of uncompressed data (note that the actual size in Snowflake is smaller because data is always stored compressed). Groups of rows in tables are mapped into individual micro-partitions, organized in a columnar fashion. This size and structure allows for extremely granular pruning of very large tables, which can be comprised of millions, or even hundreds of millions, of micro-partitions.

Snowflake stores metadata about all rows stored in a micro-partition, including:

The range of values for each of the columns in the micro-partition.

The number of distinct values.

Additional properties used for both optimization and efficient query processing.

https://docs.snowflake.com/en/user-guide/tables-clustering-micropartitions.html

Question 5: Skipped

What are the available partner categories in snowflake?

- Data integration
  (Correct)
- Business Intelligence
  (Correct)
- Security and governance
  (Correct)
- ML and data science
  (Correct)
- Data Warehousing

Explanation

https://docs.snowflake.com/en/user-guide/ecosystem.html

Question 6: Skipped

While sharing data, consumer account and provider account must be in the same region

- TRUE
  (Correct)
- FALSE

Explanation

This is an important point to note. data sharing works within the same region. However, you can replicate your account to another region and then share data from that replicated account within that account's region. This is also true across cloud platforms.

Additional explanation

I am adding this based on a question from a student. In the exam if this questions comes, please answer 'TRUE'.

Data Sharing only works within a single region. But, it is possible to replicate your account to another region and then data share from that replicated account within that account's region. This is also true across cloud platforms.

So, technically it is true that you can share data across accounts in different region. But that is data replication. Data sharing between accounts is possible only if they are in the same region

Question 7: Skipped

Who can setup a data share

- SECURITYADMIN
- SYSADMIN
- ACCOUNTADMIN or a role granted the CREATE SHARES global privilege **(Correct)**

Explanation
You must use the ACCOUNTADMIN role (or a role granted the CREATE SHARES global privilege) to perform these tasks. For more details about the CREATE SHARES privilege

https://docs.snowflake.com/en/user-guide/data-sharing-provider.html

Question 8: Skipped

Select the statements that are true for a data shar

- Shares are named Snowflake objects **(Correct)**
- Share consists of privileges that grant access to the database(s) and the schema containing the objects to share **(Correct)**
- Share consists of privileges that grant access to the specific objects in the database **(Correct)**
- Share consists of consumer accounts with which the database and its objects are shared **(Correct)**

Explanation
Shares are named Snowflake objects that encapsulate all of the information required to share a database. Each share consists of:

The privileges that grant access to the database(s) and the schema containing the objects to share.

The privileges that grant access to the specific objects in the database.

The consumer accounts with which the database and its objects are shared.

https://docs.snowflake.com/en/user-guide/data-sharing-intro.html#what-is-a-share

Question 9: Skipped

Select the statements that are true for a shared database

- Shared databases are read-only
(Correct)
- A clone can be created from a shared database
- Time travel can be done on a shared database
- Can edit the comments for a shared database

Explanation
Shared databases have the following limitations for consumers:

Shared databases are read-only. Users in a consumer account can view/query data, but cannot insert or update data, or create any objects in the database.

The following actions are not supported:

Creating a clone of a shared database or any schemas/tables in the database.

Time Travel for a shared database or any schemas/tables in the database.

Editing the comments for a shared database.

https://docs.snowflake.com/en/user-guide/data-share-consumers.html#general-limitations-for-shared-databases

Question 10: Skipped

What are the product offerings for secure data sharing?

- DIRECT SHARE
(Correct)
- DATA MARKETPLACE

(Correct)

- DATA EXCHANGE
  (Correct)
- DATA STORE

Explanation
Snowflake provides three product offerings for data sharing that utilize Snowflake Secure Data Sharing to connect providers of data with consumers.

Direct Share

Snowflake Data Marketplace

Data Exchange

https://docs.snowflake.com/en/user-guide/data-sharing-product-offerings.html#direct-share

Question 11: Skipped

Virtual warehouses cannot be resized while they are running

- TRUE
- FALSE
  (Correct)

Explanation
Warehouses can be started and stopped at any time. They can also be resized at any time, even while running, to accommodate the need for more or less compute resources, based on the type of operations being performed by the warehouse.

https://docs.snowflake.com/en/user-guide/warehouses-overview.html

Question 12: Skipped

If a transaction is running on a session and the session disconnects abruptly, the transaction will be in a detached state. It will not allow the transaction to commit or roll back. The locks on the resources will also not be released. If you do not do anything and let snowflake abort the transaction, how long will you need to wait?

- 4 hours
  (Correct)

- 15 minutes
- 1 hour

Explanation
If a transaction is running in a session and the session disconnects abruptly, preventing the transaction from committing or rolling back, the transaction is left in a detached state, including any locks that the transaction is holding on resources. If this happens, you might need to abort the transaction.

To abort a running transaction, the user who started the transaction or an account administrator can call the system function, SYSTEM$ABORT_TRANSACTION.

If the transaction is left open, Snowflake typically rolls back the transaction after it has been idle for four hours.

https://docs.snowflake.com/en/sql-reference/transactions.html#aborting-transactions

Question 13: Skipped

Please look at the below two queries

A. SELECT * FROM TABLE1 WHERE YEAR = 2015 AND UPPER(COL1) = 'COLVAL';

B. SELECT * FROM TABLE1 WHERE YEAR = 2015 AND COL1 ='COLVAL';

Which one of these will scan more partitions?

Note: COL1 stores values both in upper and lower case

- A
(Correct)
- B
- Both will scan the same number of partitions
-

Explanation
This probably will not come in the certification exam, but it is important to know this when you are working on a project.

Snowflake stores the metadata for each column in a metadata services layer which includes the range of values of the columns in the micro partition. In this case, the range of values stored in the metadata layer for COL1 is stored in its raw form which is a mix of upper and lower case. Because of this when you apply the UPPER function, the metadata services layer cannot use the information stored in the layer to fetch the micro partitions for the range of the rows

Question 14: Skipped

You are using a 2XL warehouse which has 32 nodes and 8 cores in each node. When you submit a query to this warehouse, snowflake will try to split the query into how many partitions? (Provided that the query is eligible for splitting)

- 256
  (Correct)
- 32
- 8

Explanation
There are total 8*32 = 256 cores in 2xl ware house. When a query is submitted to this warehouse snowflake will try to split the query into 256 parts and each of the cores will try to scan 1/256 part of the table.

Please note that this is an ideal situation, the query must be splittable and the warehouse cores must be available to do these parallelization

Question 15: Skipped

What factors influence the credits charged for a warehouse?

- The number of servers per cluster (determined by warehouse size)
  (Correct)
- The number of clusters (if using multi-cluster warehouses)
  (Correct)
- The size of the query that runs on the warehouse
- The length of time each server in each cluster runs
  (Correct)

Explanation
It is very important to understand this concept

Credit charges are calculated based on:

1. The number of servers per cluster (determined by warehouse size).

2. The number of clusters (if using multi-cluster warehouses).

3. The length of time each server in each cluster runs.

https://docs.snowflake.com/en/user-guide/warehouses-considerations.html#how-are-credits-charged-for-warehouses

Question 16: Skipped

When creating a warehouse, what are the two most critical factors to consider from cost and performance persepective

- ☐ Warehouse size (i.e. number of servers per cluster)
  (Correct)
- ☐ Manual vs automated management (for starting/resuming and suspending warehouses)
  (Correct)
- ☐ Cloud provider selection

Explanation
When creating a warehouse, the two most critical factors to consider, from a cost and performance perspective, are:

1. Warehouse size (i.e. number of servers per cluster).

2. Manual vs automated management (for starting/resuming and suspending warehouses).

Ideally, bigger the warehouse faster your query will run. But this may not be true, if your query itself is not big enough to be split up into enough number of partitions to take advantage of the more number of cores in the warehouse. Please be aware of this.

Setting the auto suspend time is very important. You must calculate this based on the query patterns. If the warehouse is used for adhoc queries and most of the queries finish in less than a minute, it will be a waste of money if your auto suspend is set to 15 minutes. In this case, even if the query ends in a minute, the warehouse will keep running for 15 minutes before it is suspended

https://docs.snowflake.com/en/user-guide/warehouses-considerations.html#how-are-credits-charged-for-warehouses

Question 17: Skipped

You want to scale out query performance. Which strategy will you follow with respect to your warehouses?

- Resize the warehouse
- Add clusters to the warehouse
  (Correct)
- Create another warehouse

Explanation
Adding cluster to a multi-cluster warehouse helps scale out if the query load increases. Snowflake does this scale out automatically

https://docs.snowflake.com/en/user-guide/warehouses-considerations.html#scaling-up-vs-scaling-out

Question 18: Skipped

You need to handle a huge number of concurrent users for your snowflake data warehouse. All of them are from the same team and needs same type of access. Which is the most appropriate warehouse strategy will you recommend?

- Create one warehouse for each user
- Create one warehouse for each group of user
- Use multi-cluster warehouse
  (Correct)

Explanation
Multi-cluster warehouse is the most appropriate answer for this.

Multi-cluster warehouses are designed specifically for handling queuing and performance issues related to large numbers of concurrent users and/or queries. In addition, multi-cluster warehouses can help automate this process if your number of users/queries tend to fluctuate.

https://docs.snowflake.com/en/user-guide/warehouses-considerations.html#multi-cluster-warehouses-improve-concurrency

Question 19: Skipped

Select all the options which are true about the data that is stored in a micro partition metadata

- [ ] The range of values for each of the columns in the micro-partition
(Correct)
- [ ] Mean and standard deviation of each column
- [ ] The number of distinct values
(Correct)
- [ ] Additional properties used for both optimization and efficient query processing
(Correct)

Explanation
Snowflake stores metadata about all rows stored in a micro-partition, including:

1. The range of values for each of the columns in the micro-partition.

2. The number of distinct values.

3. Additional properties used for both optimization and efficient query processing.

https://docs.snowflake.com/en/user-guide/tables-clustering-micropartitions.html

Question 20: Skipped

You want to get the clustering information for a table T1 that is clustered on column C1. Which of the below queries will you run to get the details?

- ```
  1. select system$clustering_information('T1')
  ```
(Correct)
- ```
  1. select clustering('T1') from information_schema.clusteringinfo;
  ```
- ```
  1. Show clustering info for T1;
  ```

Explanation
Returns clustering information, including average clustering depth, for a table based on one or more columns in the table
  ```

https://docs.snowflake.com/en/sql-reference/functions/system_clustering_information.html#system-clustering-information

Question 21: Skipped

Each time a warehouse is started or resized to a larger size, the warehouse is billed for 1 minute's worth of usage.After 1 minute, all subsequent billing is per-second

- TRUE
  (Correct)
- FALSE

Explanation
Warehouse credits roll over to the first minute and then the credits are charged per second. What this means is, if your warehouse is resumed at time T0 and it was up for 30 seconds(T0+30 SECS), you will still be paying for a minute. If it was up for 62 seconds, then you will be paying for 62 seconds(60 seconds + 2 seconds).

https://docs.snowflake.com/en/user-guide/credits.html#virtual-warehouse-credit-usage

Question 22: Skipped

Cloud services also use compute. Does snowflake charge you for cloud services.

- Yes, if it exceeds 10% of the daily usage of the compute resources
  (Correct)
- Yes, irrespective of how much compute resources are used
- Snowflake never charges for cloud services

Explanation
Similar to virtual warehouse usage, Snowflake credits are used to pay for the usage of the cloud services that exceeds 10% of the daily usage of the compute resources.

Usage for cloud-services is charged only if the daily consumption of cloud services exceeds 10% of the daily usage of the compute resources. The charge is calculated daily (in the UTC time zone). This ensures that the 10% adjustment is accurately applied each day, at the credit price for that day

https://docs.snowflake.com/en/user-guide/credits.html#cloud-services-credit-usage

Question 23: Skipped

If you are on Snowflake Enterprise Edition (and higher), the time travel retention period can be set to any value from 0 up to 90 days for databases, schemas, and tables

- TRUE
  (Correct)
- FALSE

Explanation
Data Retention Period

A key component of Snowflake Time Travel is the data retention period.

When data in a table is modified, including deletion of data or dropping an object containing data, Snowflake preserves the state of the data before the update. The data retention period specifies the number of days for which this historical data is preserved and, therefore, Time Travel operations (SELECT, CREATE … CLONE, UNDROP) can be performed on the data.

The standard retention period is 1 day (24 hours) and is automatically enabled for all Snowflake accounts:

For Snowflake Standard Edition, the retention period can be set to 0 (or unset back to the default of 1 day) at the account and object level (i.e. databases, schemas, and tables).

For Snowflake Enterprise Edition (and higher):

For transient databases, schemas, and tables, the retention period can be set to 0 (or unset back to the default of 1 day). The same is also true for temporary tables.

For permanent databases, schemas, and tables, the retention period can be set to any value from 0 up to 90 days.

https://docs.snowflake.com/en/user-guide/data-time-travel.html#data-retention-period

Question 24: Skipped

Time travel can be disabled at the account level

- TRUE
-

FALSE
(Correct)

Explanation
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.

https://docs.snowflake.com/en/user-guide/data-time-travel.html#enabling-and-disabling-time-travel

---

Additional clarification

In snowflake, in one place it has been written that time-travel cannot be disabled for an account, in another place it has been written that we can specify DATA_RETENTION_TIME_IN_DAYS to 0. This part of the documentation in snowflake is little confusing. however what the document is trying to say that Time Travel is enabled by default for all accounts and cannot be taken out. It however can be set to zero at the account level which is actually not the same as disabling it for the account

Question 25: Skipped

What is the encryption algorithm used by snowflake?

- AES 256-bit encryption
(Correct)
- DES
- BLOWFISH

Explanation
All Snowflake customer data is encrypted by default using the latest security standards and best practices. Snowflake uses strong AES 256-bit encryption with a hierarchical key model rooted in a hardware security module. Keys are automatically rotated on a regular basis by the Snowflake service, and data can be automatically re-encrypted (“rekeyed”) on a regular basis. Data encryption and key management is entirely transparent and requires no configuration or management.

https://docs.snowflake.com/en/user-guide/security-encryption.html#encryption-key-management

Question 26: Skipped

Choose the answer which true about key rotation in snowflake

- Account and table master keys are automatically rotated by Snowflake when they are more than 30 days old
(Correct)
- Only account master key is rotated by snowflake when it is required
- Only table master key is rotated when it is required
- Key rotation is an expensive operation hence it is done manually when required

Explanation
Account and table master keys are automatically rotated by Snowflake when they are more than 30 days old. Active keys are retired, and new keys are created. When Snowflake determines the retired key is no longer needed, the key is automatically destroyed. When active, a key is used to encrypt data and is available for usage by the originator. When retired, the key is used solely to decrypt data and is only available for usage by the recipient. When wrapping child keys in the key hierarchy, or when inserting data into a table, only the current, active key is used to encrypt data. When a key is destroyed, it is not used for either encryption or decryption. Regular key rotation limits the lifecycle for the keys to a limited period of time.

https://docs.snowflake.com/en/user-guide/security-encryption.html#encryption-key-rotation

Question 27: Skipped

Which roles create, alter, or drop network policies?

- ACCOUNTADMIN
(Correct)
- SECURITYADMIN
(Correct)
- SYSADMIN

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

Question 28: Skipped

When a network policy includes values in both the allowed and blocked IP address lists, Snowflake applies the ***blocked*** IP address list first

- TRUE
  (Correct)
- FALSE

Explanation
When a network policy includes values in both the allowed and blocked IP address lists, Snowflake applies the ***blocked*** IP address list first.

Do ***not*** add `0.0.0.0/0` to the blocked IP address list. `0.0.0.0/0` is interpreted to be "all IPv4 addresses on the local machine". Because Snowflake resolves this list first, this would block your own access. Also, note that it is not necessary to include this IP address in the allowed IP address list.

Question 29: Skipped

What is the recommended file size to optimize the number of parallel operations for a load?

- 100-250 MB(or larger) in size *compressed*
  (Correct)
- 10 MB to 100 MB in size un *compressed*.
- less than 1 MB compressed

Explanation
The number of load operations that run in parallel cannot exceed the number of data files to be loaded. To optimize the number of parallel operations for a load, we recommend aiming to produce data files roughly 10 MB to 100 MB in size ***compressed***. Aggregate smaller files to minimize the processing overhead for each file. Split larger files into a greater number of smaller files to distribute the load among the servers in an active warehouse. The number of data files that are processed in parallel is determined by the number and capacity of servers in a warehouse. We recommend splitting large files by line to avoid records that span chunks.

If your source database does not allow you to export data files in smaller chunks, you can use a third-party utility to split large CSV files.

https://docs.snowflake.com/en/user-guide/data-load-considerations-prepare.html#file-sizing-best-practices-and-limitations

Question 30: Skipped

What is the size limit on the variant data type in snowflake

- 16 MB
  (Correct)
- 64 MB
- 32 MB

Explanation
The VARIANT data type imposes a 16 MB (compressed) size limit on individual rows.

https://docs.snowflake.com/en/user-guide/data-load-considerations-prepare.html#semi-structured-data-size-limitations

Question 31: Skipped

The interactions with data are initialized through the services layer

- TRUE
  (Correct)
- FALSE

Explanation
Cloud services tie together all of the different components of Snowflake in order to **process user requests, from login to query dispatch**

https://docs.snowflake.com/en/user-guide/intro-key-concepts.html#cloud-services

Question 32: Skipped

You have many queries that needs to run at the same time on a warehouse and you need high concurrency when they execute. You will go for a multi-cluster warehouse in this case.

- TRUE
  (Correct)
- FALSE

Explanation
This question can be framed in many ways in the exam. But please remember only two things

1. Multi Cluster warehouse can scale compute resources

2. The scaling is based on concurrent users and queries

Multi-cluster warehouses enable you to scale compute resources to manage your user and query concurrency needs as they change, such as during peak and off hours.

https://docs.snowflake.com/en/user-guide/warehouses-multicluster.html#multi-cluster-warehouses

Question 33: Skipped

MFA or multi factor authentication is part of snowflake's security and authentication

- TRUE
  (Correct)
- FALSE

Explanation
Here, you need to remember one more thing. MFA is integrated in snowflake and `powered by DUO` . Also users are not automatically enrolled to MFA, they must enroll themselves

Snowflake supports multi-factor authentication (MFA) to provide increased login security for users connecting to Snowflake. MFA support is provided as an integrated Snowflake feature, powered by the Duo Securityservice, which is managed completely by Snowflake. Users do not need to separately sign up with Duo or perform any tasks, other than installing the Duo Mobile application, which is supported on multiple smart phone platforms (iOS, Android, Windows, etc.). See the Duo User Guidefor more information about supported platforms/devices and how Duo multi-factor authentication works.

MFA is enabled on a per-user basis; however, at this time, users are not automatically enrolled in MFA. To use MFA, users must enroll themselves.

Question 34: Skipped

Snowflake has administration settings for resource consumption to help with below activities

- Help control costs associated with unexpected warehouse credit usage
  (Correct)
- Manage access to snowflake for specific users
- Manage availability of the product

Explanation
I remembered that I have not included resource monitors in any of the previous practice tests. resource monitor is important for the exam

What is a resource monitor?

To help control costs and avoid unexpected credit usage caused by running warehouses, Snowflake provides *resource monitors*.

A virtual warehouse consumes Snowflake credits while it runs.Resource monitors can be used to impose limits on the number of credits that are consumed by:

1. User-managed virtual warehouses

2. Virtual warehouses used by cloud services

https://docs.snowflake.com/en/user-guide/resource-monitors.html#working-with-resource-monitors

Question 35: Skipped

When a resource limit has hit the limit or is about to hit the limit, what can the resource monitor do

- Suspend the warehouse
  (Correct)
- Send alert notifications
  (Correct)
- Kill the query that is running
  (Correct)

Explanation
Limits can be set for a specified interval or date range. When these limits are reached and/or are approaching, the resource monitor can trigger various actions, **such as sending alert notifications and/or suspending the warehouses.**

```
DO SUSPEND | SUSPEND_IMMEDIATE | NOTIFY
```

Specifies the action performed by the trigger when the threshold is reached:

`SUSPEND`: Suspend all assigned warehouses while allowing currently running queries to complete. No new queries can be executed by the warehouses until the credit quota for the resource monitor is increased. In addition, this action sends a notification to all users who have enabled notifications for themselves.

`SUSPEND_IMMEDIATE`: Suspend all assigned warehouses immediately and cancel any currently running queries or statements using the warehouses. In addition, this action sends a notification to all users who have enabled notifications for themselves.

`NOTIFY`: Send an alert (to all users who have enabled notifications for themselves), but do not take any other action.

https://docs.snowflake.com/en/sql-reference/sql/create-resource-monitor.html#optional-parameters

Question 36: Skipped

Snowflake provides specific administration features and capabilities to support the following activities

- Managing database and warehouses within a snowflake account
(Correct)
- Managing roles and users within a snowflake account
(Correct)
- Monitor usage and manage resources to control costs in a snowflake account
(Correct)
- Manage 3rd party applications providing data to snowflake

Explanation
3rd party applications are not managed by snowflake

Question 37: Skipped

What is a point in time snapshot of data which can be updated by users is called

- Snowflakes' cloning
(Correct)
- Time travel
- Fail safe

Explanation
This is very important to know

What is cloning in snowflake?

Cloning also referred to as “zero-copy cloning” creates a copy of a database, schema or table. **A snapshot of data** present in the source object is taken when the clone is

created and is made available to the cloned object. The cloned object is **writable** and is independent of the clone source. That is, changes made to either the source object or the clone object are not part of the other. Cloning a database will clone all the schemas and tables within that database. Cloning a schema will clone all the tables in that schema. This topic provides examples of cloning tables, databases, and schemas in a script, with explanatory comments.

Question 38: Skipped

Federated authentication in snowflake is compliant with SAML2.0

- TRUE
(Correct)
- FALSE

Explanation
Federated authentication enables your users to connect to Snowflake using secure SSO (single sign-on). With SSO enabled, your users authenticate through an external, **SAML 2.0-compliant identity provider (IdP)**

https://docs.snowflake.com/en/user-guide/admin-security-fed-auth.html#federated-authentication-sso

Question 39: Skipped

All security related information is stored in which layer of snowflake architecture

- Service
(Correct)
- Storage
- Compute
- All of the above

Explanation
Service(or Cloud service) layer is the one which stores all security related information

The services included in this layer are

1. Authentication

2. Infrastructure management

3. Metadata management

4. Query parsing and optimization

5. Access control

Question 40: Skipped

What is the benefit of client-side encryption in snowflake?

- It provides a secure system for managing data in cloud storage
(Correct)
- It helps to decrypt data faster at a later period
- It reduces the cost of encryption by the cloud provider

Explanation
Client-side encryption in general, is the most secure form of managing data on a cloud provider. With client-side encryption, the data is encrypted on the client before it is uploaded. That means, the cloud provider only stores the encrypted version of the data and never sees data in the clear.

Question 41: Skipped

Multi-factor Authentication can be used for connecting to Snowflake via the Snowflake JDBC driver

- TRUE
(Correct)
- FALSE

Question 42: Skipped

Having separate accounts in snowflake enables users to have

- Different editions of snowflake and different regions
- Billing can be done at account level
- Database object deployment between environments become simpler
- All of the above
(Correct)

Question 43: Skipped

Semi structured data strings in snowflake are stored in a column with the below datatype

- VARIANT
  (Correct)
- OBJECT
- VARCHAR
- NONE OF THE ABOVE

Explanation
VARIANT is a tagged universal type, which can store values of any other type, including OBJECT and ARRAY, up to a maximum size of 16 MB compressed.

Question 44: Skipped

When you size a warehouse, which of the below factors are not required to be considered

- Number of users
  (Correct)
- Number of concurrent queries
- Number of tables queried
- Data size and composition
- All of the above

Explanation
Number of concurrent queries will tell you whether to size the warehouse bigger(or smaller) or to use multi-cluster warehouse

Warehouse size will also depend on number of tables queried, data size and composition.

Question 45: Skipped

This property of resource monitor specifies whether the resource monitor is used to monitor the credit usage for your entire Account (i.e. all warehouses in the account) or a specific set of individual warehouses

- CREDIT QUOTA
- MONITOR LEVEL
  (Correct)
-

SCHEDULE

Explanation

Please go through all the properties of resource monitor

Monitor Level

This property specifies whether the resource monitor is used to monitor the credit usage for your entire Account (i.e. all warehouses in the account) or a specific set of individual warehouses.

If this property is not set, the resource monitor doesn't monitor any credit usage. It simply remains dormant.

https://docs.snowflake.com/en/user-guide/resource-monitors.html#monitor-level

Question 46: Skipped

Resource monitors support the below actions

- ☐ Notify & Suspend
  (Correct)
- ☐ Notify & Suspend Immediately
  (Correct)
- ☐ Notify
  (Correct)
- ☐ Notify and automatically call the account administrator

Explanation

Resource monitors support the following actions:

Notify & Suspend

Send a notification (to all account administrators with notifications enabled) and suspend all assigned warehouses after all statements being executed by the warehouse(s) have completed.

Notify & Suspend Immediately

Send a notification (to all account administrators with notifications enabled) and suspend all assigned warehouses immediately, which cancels any statements being executed by the warehouses at the time.

Notify

Perform no action, but send an alert notification (to all account administrators with notifications enabled).

https://docs.snowflake.com/en/user-guide/resource-monitors.html#actions

Question 47: Skipped

A warehouse can be assigned to only a single resource monitor.

- TRUE
  (Correct)
- FALSE

Explanation
A single monitor can be set at the account level to control credit usage for all warehouses in your account.

In addition, a monitor can be assigned to one or more warehouses, thereby controlling the credit usage for each assigned warehouse. **Note, however, that a warehouse can be assigned to only a single resource monitor.**

Question 48: Skipped

Resource monitor notifications can be received by account administrators through the following means

- Web interface
  (Correct)
- Email
  (Correct)
- Pager
- Mobile

Explanation
Notifications can be received by account administrators through the web interface and/or email; however, by default, notifications are ***not*** enabled:

To receive notifications, each account administrator ***must*** explicitly enable notifications through their preferences in the web interface.

In addition, if an account administrator chooses to receive email notifications, they must provide a valid email address (and verify the address) before they will receive any emails.

Question 49: Skipped

Which access control privileges must be granted on specified resource monitors for roles other than account administrators to view and modify resource monitors as needed

- MONITOR
(Correct)
- CREATE
- MODIFY
(Correct)

Explanation
By default, resource monitors can only be created by account administrators and, therefore, can only be viewed and maintained by them.

However, Snowflake supports enabling other users to view and modify resource monitors as needed. This functionality is implemented through the following access control privileges, which can be granted on specified resource monitors to specified roles:

1. MONITOR

2. MODIFY

Question 50: Skipped

Which properties of a resource monitor can be modified?

- Increase or decrease the credit quota for the monitor
(Correct)
- If the monitor is monitoring your account, convert it to monitor individual warehouses
(Correct)
- If the monitor is monitoring individual warehouses:
1. Add or remove warehouses from the list.
2. Convert it to monitor your account

(Correct)

- Customize the schedule (frequency, start timestamp, and end timestamp) for the monitor
(Correct)

- Add or remove actions, or modify the threshold percentages for existing actions
(Correct)

Explanation
This is important to know

You can modify the following properties for an existing resource monitor:

1. Increase or decrease the credit quota for the monitor.

2. If the monitor is monitoring your account, convert it to monitor individual warehouses.

3. If the monitor is monitoring individual warehouses:

Add or remove warehouses from the list.

Convert it to monitor your account.

4. Customize the schedule (frequency, start timestamp, and end timestamp) for the monitor.

5. Add or remove actions, or modify the threshold percentages for existing actions.

Question 51: Skipped

Select the TABLE function which helps to convert semi-structured data to a relational representation

- FLATTEN
(Correct)

- CHECK_JSON
- PARSE_JSON

Explanation
Flattens (explodes) compound values into multiple rows.

FLATTEN is a table function that takes a VARIANT, OBJECT, or ARRAY column and produces a lateral view (i.e. an inline view that contains correlation referring to other tables that precede it in the FROM clause).

`FLATTEN can be used to convert semi-structured data to a relational representation.`

https://docs.snowflake.com/en/sql-reference/functions/flatten.html#flatten

Question 52: Skipped

Output of a FLATTEN query has the below columns. Select three.

- [ ] SEQ
(Correct)
- [ ] PATH
(Correct)
- [ ] VALUE
(Correct)
- [ ] RECORD_METADATA

Explanation
The returned rows from a FLATTEN table function consist of a fixed set of columns:

```
+-----+------+------+-------+-------+------+
| SEQ |  KEY | PATH | INDEX | VALUE | THIS |
|-----+------+------+-------+-------+------|
```

SEQ

A unique sequence number associated with the input record; the sequence is not guaranteed to be gap-free or ordered in any particular way.

KEY

For maps or objects, this column contains the key to the exploded value.

PATH

The path to the element within a data structure which needs to be flattened.

INDEX

The index of the element, if it is an array; otherwise NULL.

VALUE

The value of the element of the flattened array/object.

THIS

The element being flattened (useful in recursive flattening).

Question 53: Skipped

Micropartitions can be modified after creation of the table

- TRUE
- FALSE
  (Correct)

Explanation
In snowflake data is physically stored into micro-partitions in proprietary columnar format. As a used you do not have direct access to micropartitions. Snowflake manages the micropartitions. Also micro partitions are immutable which means that once a micro partition is created, it cannot be changed. When a row is updated, the micro partition containing that row is copied to a new micro partition and then the updated row is inserted into that partition. The old micro-partition is marked for deletion.

As a used, you however have control to create clustering on your table which helps to re-arrange the micro-partitions based on the cluster keys for faster query retrieval.

Question 54: Skipped

You want to update the value of a specific column in a table. But the updation is not based on any specific condition. You just need to ensure that 80% of the rows in the table are updated with the new value. What snowflake feature will you use

- SAMPLE
  (Correct)
- CONCAT
-

AVERAGE

- MAX

Explanation
SAMPLE / TABLESAMPLE

Returns a subset of rows sampled randomly from the specified table. The following sampling methods are supported:

Sample a fraction of a table, with a specified probability for including a given row. The number of rows returned depends on the size of the table and the requested probability. A seed can be specified to make the sampling deterministic.

Sample a fixed, specified number of rows. The exact number of specified rows is returned unless the table contains fewer rows.

SAMPLE and TABLESAMPLE are synonymous and can be used interchangeably.

Also try a hands on as below

```
CREATE OR REPLACE TABLE FRUITS(FRUIT_NUMBER NUMBER, FRUIT_DESCRIPTION VARCHAR, AV
AILABILITY VARCHAR);


INSERT INTO FRUITS VALUES(1,'APPLE',NULL);

INSERT INTO FRUITS VALUES(2,'MANGO',NULL);

INSERT INTO FRUITS VALUES(3,'ORANGE',NULL);

INSERT INTO FRUITS VALUES(4,'KIWI',NULL);

INSERT INTO FRUITS VALUES(5,'MELON',NULL);

INSERT INTO FRUITS VALUES(6,'PEARS',NULL);

INSERT INTO FRUITS VALUES(7,'BANANA',NULL);

INSERT INTO FRUITS VALUES(8,'PLUM',NULL);

INSERT INTO FRUITS VALUES(9,'APRICOT',NULL);

INSERT INTO FRUITS VALUES(10,'WATERMELON',NULL);


SET COUNT=(SELECT COUNT(*)*0.8 FROM FRUITS);


UPDATE FRUITS SET AVAILABILITY = 'YES'
```

```
FROM (SELECT FRUIT_NUMBER FROM FRUITS SAMPLE($COUNT ROWS)) T1

WHERE FRUITS.FRUIT_NUMBER = T1.FRUIT_NUMBER;
```

Question 55: Skipped

Select two true statements for FLATTEN ?

- ☐ Takes a VARIANT, OBJECT, or ARRAY column and produces a lateral view **(Correct)**
- ☐ Can be used to convert semi-structured data to a relational representation **(Correct)**
- ☐ Cannot be used on permanent tables

Explanation
FLATTEN

Flattens (explodes) compound values into multiple rows.

FLATTEN is a table function that **takes a VARIANT, OBJECT, or ARRAY column and produces a lateral view** (i.e. an inline view that contains correlation referring to other tables that precede it in the FROM clause).

FLATTEN can be used to convert semi-structured data to a relational representation.

Question 56: Skipped

When migrating from Teradata to snowflake, you will also need to migrate the indexes.

- ○ TRUE
- ○ FALSE **(Correct)**

Explanation
Snowflake has an optimizer built from the ground up and architected for MPP and the cloud. Snowflake understands parallel execution plans and automatically optimizes them. This means `there are no indexes in Snowflake` so you don't have to migrate Indexes from Teradata

Question 57: Skipped

Choose the continuous data loading options from below

- ☐ SNOWPIPE

(Correct)

- ☐ SNOWFLAKE CONNECTOR FOR KAFKA
(Correct)
- ☐ Third-party data integration tools
(Correct)
- ☐ BULK COPY

Explanation
Continuous data loading

Options for continuous data loading include the following:

1. Snowpipe

2. Snowflake Connector for Kafka

3. Third-party data integration tools

Question 58: Skipped

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.

- ☐ TRUE
(Correct)
- ☐ FALSE

Explanation
Change data tracking

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. A stream allows querying and consuming a set of changes to a table, at the row level, between two transactional points of time.

In a continuous data pipeline, table streams record when staging tables and any downstream tables are populated with data from business applications using continuous data loading and are ready for further processing using SQL statements.

For more information, see Change Tracking Using Table Streams.

Question 59: Skipped

Snowpipe continuously loads micro-batches of data from an external stage location (Amazon S3, Google Cloud Storage, or Microsoft Azure) into a staging table.

- TRUE
(Correct)
- FALSE

Explanation
Snowpipe

Snowpipe continuously loads micro-batches of data from an external stage location (Amazon S3, Google Cloud Storage, or Microsoft Azure) into a staging table.

Question 60: Skipped

The Kafka connector continuously loads records from one or more Apache Kafka topics into an internal (Snowflake) stage and then into a staging table using Snowpipe.

- TRUE
(Correct)
- FALSE

Explanation
Snowflake Connector for Kafka

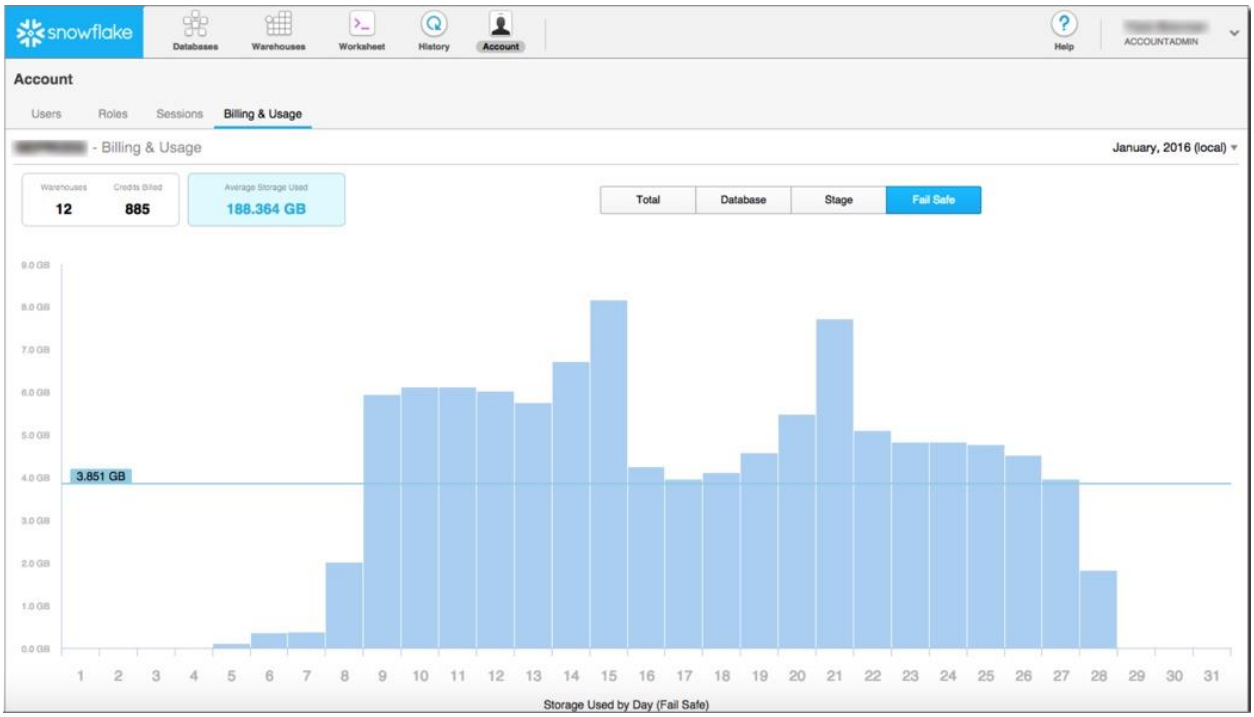
The Kafka connector continuously loads records from one or more Apache Kafka topics into an internal (Snowflake) stage and then into a staging table using Snowpipe.

Question 61: Skipped

The amount of storage used by failsafe can be viewed under Account->Billing & Usage section

- TRUE
(Correct)
- FALSE

Explanation
In the web interface, account administrators can view the total data storage for their account, including historical data in Fail-safe. The page also provides a filter for viewing only Fail-safe storage:



Question 62: Skipped

A stream stores data in the same shape as the source table (i.e. the same column names and ordering). What additional columns does it store?

- ☐ **METADATA$ACTION**
  (Correct)
- ☐ **METADATA$ISUPDATE**
  (Correct)
- ☐ **METADATA$ROW_ID**
  (Correct)
- ☐ **METADATA$STREAM_ID**

Explanation
Stream Columns

A stream stores data in the same shape as the source table (i.e. the same column names and ordering) with the following additional columns:

METADATA$ACTION

Indicates the DML operation (INSERT, DELETE) recorded.

METADATA$ISUPDATE

Indicates whether the operation was part of an UPDATE statement. Updates to rows in the source table are represented as a pair of DELETE and INSERT records in the stream with a metadata column METADATA$ISUPDATE values set to TRUE.

Note that streams record the differences between two offsets. If a row is added and then updated in the current offset, the delta change is a new row. The METADATA$ISUPDATE row records a FALSE value.

METADATA$ROW_ID

Specifies the unique and immutable ID for the row, which can be used to track changes to specific rows over time.

Question 63: Skipped

COPY transformations are supported only for named stages(both internal and external) and user stages

- TRUE
  (Correct)
- FALSE

Explanation
Supported Stages

Only named stages (internal or external) and user stages are supported for COPY transformations.

https://docs.snowflake.com/en/user-guide/data-load-transform.html#usage-notes

Question 64: Skipped

The PIPE_EXECUTION_PAUSED parameter can be set at which levels

- DATABASE
- ACCOUNT
  (Correct)
- SCHEMA
  (Correct)

PIPE
(Correct)

Explanation
Transferring Pipe Ownership

Complete the following steps to transfer ownership of a pipe:

Set the PIPE_EXECUTION_PAUSED parameter to TRUE.

This parameter enables pausing or resuming a pipe. The parameter is supported at the following levels:

Account

Schema

Pipe

At the pipe level, the object owner (or a parent role in a role hierarchy) can set the parameter to pause or resume an individual pipe.

An account administrator (user with the ACCOUNTADMIN role) can set this parameter at the account level to pause or resume all pipes in the account. Likewise, a user with the MODIFY privilege on the schema can pause or resume pipes at the schema level. Note that this larger domain control only affects pipes for which the parameter was not already set at a lower level; e.g., by the owner at the object level.

https://docs.snowflake.com/en/user-guide/data-load-snowpipe-manage.html#transferring-pipe-ownership

Question 65: Skipped

You want to delete the snowpipe file loading metadata. You decided to truncate the table to be able to do this. This will remove the metadata from the pipe

- TRUE
- FALSE
(Correct)

Explanation
Unable to Reload Modified Data, Modified Data Loaded Unintentionally

Snowflake uses file loading metadata to prevent reloading the same files (and duplicating data) in a table. Snowpipe prevents loading files with the same name even if they were later modified (i.e. have a different eTag).

`The file loading metadata is associated with the pipe object rather than the table. As a result:`

Staged files with the same name as files that were already loaded are ignored, even if they have been modified, e.g. if new rows were added or errors in the file were corrected.

`Truncating the table using the TRUNCATE TABLE command does not delete the Snowpipe file loading metadata.`

However, note that pipes only maintain the load history metadata for ***14 days***.

Question 66: Skipped

Which cache type gets purged in a predictable way so that it ends up empty of all cached infromation?

- Warehouse Cache
  (Correct)
- Results Cache
- Metadata Cache

Explanation
Lets investigate why this answer is Warehouse cache

First what is metadata cache, this is the cache which is in the global service layer and stores Object Information and statistics of the objects(like count, max and min). As a user you do not have any control on this cache

Results Cache - This stores the exact results from the exact queries and last for 24 hours. But any time the query is run, this time gets extended by another 24 hours so there is no way you can predict when it will be empty

Warehouse cache - This contains the raw data from the table, this is cached in the SSD disk attached to the warehouse. When a warehouse is suspended, this data is lost and you completely loose the cache.

Now you see why the warehouse cache is a predictable way to empty the cache

Question 67: Skipped

Which cache runs on a 24 hour clock?

- Result Cache
  (Correct)
- Metadata cache
- Warehouse cache

Question 68: Skipped

You are a snowflake architect and you are investigating a long running query to optimize the query. You want to know if the query retrieved data from long-term centralized storage. If you are looking at the Query History table in the history area of the WebUI, where will you look for this information?

- Look in the Bytes Scanned Column for a green bar
  (Correct)
- Look in the Bytes scanned column for the acronym LTCS
- Look in the Rows column for a red bar
- Look in the Rows column for the acronym LTCS

Explanation

Status	Query ID	SQL Text	U...	Warehouse	Clust...	Size	Session ID	Start Time	End Time	Total Duration	Bytes Scanned	Client Info	Rows
✔	0197e598-...	SELECT * FROM inve...	J...	ETL_WH	1	Small	6215552118...	1:08:46 PM	1:09:44 PM	58.0s	3.9GB	Snowflake UI 202010...	

Question 69: Skipped

Which is the preferred way to store a generated identity and access management (IAM) entity for external cloud storage

- Access Token
- Stage
- Storage integration
  (Correct)
- File Format

Explanation
A storage integration is a Snowflake object that stores a generated identity and access management (IAM) entity for your external cloud storage, along with an optional set of allowed or blocked storage locations (Amazon S3, Google Cloud Storage, or Microsoft Azure). Cloud provider administrators in your organization grant permissions on the storage locations to the generated entity. This option allows users to avoid supplying credentials when creating stages or when loading or unloading data.

A single storage integration can support multiple external stages. The URL in the stage definition must align with the storage location specified for the STORAGE_ALLOWED_LOCATIONS parameter.

https://docs.snowflake.com/en/sql-reference/sql/create-storage-integration.html#create-storage-integration

Question 70: Skipped

What is the default option for ON_ERROR clause of COPY command for bulk loading?

- `CONTINUE`
- `SKIP_FILE`
- `SKIP_FILE_10`
- `ABORT_STATEMENT`
  (Correct)

Explanation
`ON_ERROR = CONTINUE | SKIP_FILE | SKIP_FILE_num | SKIP_FILE_num% | ABORT_STATEMENT`

Definition

String (constant) that specifies the action to perform when an error is encountered while loading data from a file:

Values

Supported Values

Notes

`CONTINUE`

Continue loading the file. The COPY statement returns an error message for a maximum of one error encountered per data file. Note that the difference between the ROWS_PARSED and ROWS_LOADED column values represents the number of rows that include detected errors. However, each of these rows could include multiple errors. To view all errors in the data files, use the VALIDATION_MODE parameter or query the VALIDATE function.

`SKIP_FILE`

Skip file if any errors encountered in the file.

`SKIP_FILE_num` (e.g. `SKIP_FILE_10`)

Skip file when the number of errors in the file is equal to or exceeds the specified number.

`SKIP_FILE_num%` (e.g. `SKIP_FILE_10%`)

Skip file when the percentage of errors in the file exceeds the specified percentage.

`ABORT_STATEMENT`

Abort the load operation if any error is encountered in a data file. Note that the load operation is ***not*** aborted if the data file cannot be found (e.g. because it does not exist or cannot be accessed).

Default

Bulk loading using COPY : `ABORT_STATEMENT`

Snowpipe: `SKIP_FILE`

https://docs.snowflake.com/en/sql-reference/sql/copy-into-table.html#copy-options-copyoptions

Question 71: Skipped

Select the limitations with materialized views(Select 3)

- ☐ Snowflake's "Time Travel" feature is not supported on materialized views **(Correct)**
- ☐ using CURRENT_TIME or CURRENT_TIMESTAMP is not permitted **(Correct)**

- A materialized view cannot include Window functions
  (Correct)
- Clustering keys cannot be defined in Materialized views

Explanation
https://docs.snowflake.com/en/user-guide/views-materialized.html#limitations-on-creating-materialized-views

The following limitations apply to creating materialized views:

A materialized view can query only a single table.

Joins, including self-joins, are not supported.

A materialized view cannot query:

1. A materialized view.

2. A non-materialized view.

3. A UDTF (user-defined table function).

A materialized view cannot include:

1. UDFs (this limitation applies to all types of user-defined functions, including external functions).

2. Window functions.

3. HAVING clauses.

4. ORDER BY clause.

5. LIMIT clause.

6. GROUP BY keys that are not within the SELECT list. All GROUP BY keys in a materialized view must be part of the SELECT list.

7. GROUP BY GROUPING SETS.

8. GROUP BY ROLLUP.

9. GROUP BY CUBE.

10. Nesting of subqueries within a materialized view.

Many aggregate functions are not allowed in a materialized view definition.

The aggregate functions that are ***supported*** in materialized views are:

APPROX_COUNT_DISTINCT (HLL).

AVG (except when used in PIVOT).

BITAND_AGG.

BITOR_AGG.

BITXOR_AGG.

COUNT.

MIN.

MAX.

STDDEV.

STDDEV_POP.

STDDEV_SAMP.

SUM.

VARIANCE (VARIANCE_SAMP, VAR_SAMP).

VARIANCE_POP (VAR_POP).

The other aggregate functions are ***not supported*** in materialized views.

Note

Aggregate functions that are allowed in materialized views still have some restrictions:

Aggregate functions cannot be nested.

Aggregate functions used in complex expressions (e.g. `(sum(salary)/10)`) can only be used in the outer-most level of a query, not in a subquery or an in-line view.

For example, the following is allowed:

```
create materialized view mv1 as
    select
        sum(x) + 100
      from t;
```

The following is ***not*** allowed:

```
create materialized view mv2 as
    select
        y + 10
      from (
        select
          sum(x) as y
        from t
      );
```

DISTINCT cannot be combined with aggregate functions.

In a materialized view, the aggregate functions `AVG`, `COUNT`, `MIN`, `MAX`, and `SUM` can be used as aggregate functions but not as window functions. In a materialized view, these functions cannot be used with the `OVER` clause:

```
over ( [ partition by <expr1> ] [ order by <expr2> ] )
```

If an aggregate function is in a subquery, the materialized view cannot create an expression on top of the aggregated column(s) from that subquery. For example, consider the following materialized view definition:

```
create or replace materialized view mv1 as
    select c1 + 10 as c1new, c2
        from (select sum(c1) as c1, c2 from t group by c2);
```

The expression “c1 + 10” is an expression on top of an aggregate function in a subquery, and therefore causes an error message.

Note that even an equality operator counts as an expression, which means that `CASE` expressions using columns that represent aggregate functions in a subquery are also prohibited.

To work around this limitation, create a materialized view without the expression, and then create a non-materialized view that includes the expression, for example:

```
create or replace materialized view mv1 as
    select c1, c2
        from (select sum(c1) as c1, c2 from t group by c2);


create or replace view expr_v1 as
    select c1 + 10 as c1new, c2
        from (select c1, c2 from mv1);
```

Functions used in a materialized view must be deterministic. For example, using CURRENT_TIME or CURRENT_TIMESTAMP is not permitted.

A materialized view should not be defined using a function that produces different results for different settings of parameters, such as the session-level parameter TIMESTAMP_TYPE_MAPPING.

Question 72: Skipped

Which of the below columns are usually a good choice for clustering key.

- UUID column in a Customer fact table
- Gender with Male/Female values in a multi-terabyte table
- Timestamp column in a 10TB order fact table
- Store_Id in a 2 TB Item Store Sales fact table
  (Correct)

Explanation
What is the reason behind this?

Remember that a cluster key should not have a very high cardinality nor it should have a very low cardinality. If it has high cardinality, there will be lot of partitions that will be created, with low cardinality very few partitions will be created and hence query perfromance will not be able take advantage of partition pruning. We need to take a middle ground.

Question 73: Skipped

An empty table with no micro-partition will have a clustering depth of

- 1
- 0
  (Correct)
- -1
- 99

Explanation
The clustering depth for a populated table measures the average depth (`1` or greater) of the overlapping micro-partitions for specified columns in a table. The smaller the average depth, the better clustered the table is with regards to the specified columns.

Clustering depth can be used for a variety of purposes, including:

1. Monitoring the clustering “health” of a large table, particularly over time as DML is performed on the table.

2. Determining whether a large table would benefit from explicitly defining a clustering key.

A table with no micro-partitions (i.e. an unpopulated/empty table) has a clustering depth of `0`.

Question 74: Skipped

Why would you not recommend a high cardinality column for including in the clustering key? (Select 2)

- Extremely high cardinality column results in skewed distribution across micro-partitions and hence will impact query pruning
-

High cardinality will lead to low clustering depth, so as you add more data query performance will degrade

- [ ] Cluster maintenance with high cardinality keys is expensive
(Correct)

- [ ] Point lookup queries will only benefit, whereas range queries will not be effectively utilize the clustering key
(Correct)

Explanation
In general, if a column (or expression) has higher cardinality, then maintaining clustering on that column is more expensive.

The cost of clustering on a unique key might be more than the benefit of clustering on that key, especially if point lookups are not the primary use case for that table.

Question 75: Skipped

Temporary tables cannot be created with a clustering key defined

- ( ) TRUE
- ( ) FALSE
(Correct)

Explanation
Run the below command in snowflake and see if you can create it

```
CREATE TEMPORARY TABLE TEMP_TABLE(ITEM_ID INT) CLUSTER BY(ITEM_ID);
```

Question 76: Skipped

In which of the below scenarios, you will use an external table.

- [ ] You have data on the cloud providers's object store(AWS/GCS/AZURE Blob) but the data cannot be copied or moved to any other location due to compliance regulations
(Correct)

- [ ] You have high volume of data on the cloud providers's object store(AWS/GCS/AZURE Blob) but only a part of the data is needed in snowflake
(Correct)

- [ ] You have data on the cloud providers's object store(AWS/GCS/AZURE Blob) which needs to be updated by snowflake after applying required transformations
- [ ]

You have XML data on cloud provider's object store

Explanation

Please use elimination method to answer this questions

1. External tables are read-only, hence #3 is not correct

2. External tables do not support XML and hence #4 is not correct

Question 77: Skipped

What is the recommended maximum data size limitations for Parquet files?

- 100 MB
- 10 MB
- 1 GB
  (Correct)
- 3 GB

Explanation

Parquet Data Size Limitations

Currently, data loads of large Parquet files (e.g. greater than 3 GB) could time out. Split large files into files 1 GB in size (or smaller) for loading.

Question 78: Skipped

Integration object is a first-class database object

- True
- False
  (Correct)

Explanation

Integration object is an account level object and not a database object

Question 79: Skipped

Which of the below function returns the name of the warehouse of the current session?

- ACTIVE_WAREHOUSE()
- SESSION_WAREHOUSE()
- CURRENT_WAREHOUSE()
  (Correct)

- WAREHOUSE()

Explanation
Run the below query to see it yourself

SELECT CURRENT_WAREHOUSE();

Question 80: Skipped

Warehouse cache size changes with warehouse size

- True
(Correct)
- False

Explanation
The size of the cache is determined by the number of servers in the warehouse (i.e. the larger the warehouse and, therefore, the number of servers in the warehouse), the larger the cache.

https://docs.snowflake.com/en/user-guide/warehouses-considerations.html#how-does-warehouse-caching-impact-queries

Question 81: Skipped

A task can execute multiple SQL statement.

- TRUE
- FALSE
(Correct)

Explanation
`Currently, a task can execute a single SQL statement, including a call to a stored procedure.`

Tasks can be combined with table streams for continuous ELT workflows to process recently changed table rows. Streams ensure exactly once semantics for new or changed data in a table.

Tasks can also be used independently to generate periodic reports by inserting or merging rows into a report table or perform other periodic work.

Question 82: Skipped

What are the recommendations from Snowflake to effectively and efficiently use a warehouse.

- Experiment with different types of queries and different warehouse sizes to determine the combinations that best meet your specific query needs and workload **(Correct)**
- Don’t focus on warehouse size. Snowflake utilizes per-second billing, so you can run larger warehouses (Large, X-Large, 2X-Large, etc.) and simply suspend them when not in use **(Correct)**
- Be very careful on what size of warehouse you are picking up. The credits charged are based on number of queries running on the warehouse and the size of the warehouse

Explanation
The keys to using warehouses effectively and efficiently are:

1. Experiment with different types of queries and different warehouse sizes to determine the combinations that best meet your specific query needs and workload.

2. Don’t focus on warehouse size. Snowflake utilizes per-second billing, so you can run larger warehouses (Large, X-Large, 2X-Large, etc.) and simply suspend them when not in use.

https://docs.snowflake.com/en/user-guide/warehouses-considerations.html#warehouse-considerations

Remember one more thing. The credits are not dependent on number of queries running in your warehouse, it depends on how long your warehouse is active.

Question 83: Skipped

In a stream, if a row is added and then updated in the current offset, how will the delta change be represented

- The delta change is a new row **(Correct)**
- The delta change will be a pair of delete and insert record
- The delta change will be a pair of insert and upate record

Explanation
METADATA$ISUPDATE

Indicates whether the operation was part of an UPDATE statement. Updates to rows in the source table are represented as a pair of DELETE and INSERT records in the stream with a metadata column METADATA$ISUPDATE values set to TRUE.

```
Note that streams record the differences between two offsets. If a row is
added and then updated in the current offset, the delta change is a new row.
The METADATA$ISUPDATE row records a FALSE value.
```

In the above highlighted, please also note how METADATA$ISUPDATE will be handled, it will record a FALSE value. This may be asked in the exam.

Question 84: Skipped

What are the two types of snowflake data listings which define how data is shared and consumed via data marketplace?

- ☐ Standard Data Listings
  (Correct)
- ☐ Personalized Data Listings
  (Correct)
- ☐ Business Critical Data Listings
- ☐ Enterprise Data Listings

Explanation
Types of Data Listings

The Snowflake Data Marketplace offers two types of data listings which define how data is shared and consumed.

Once published, both types of listings are displayed to consumers in the selected regions. The key difference between standard and personalized listings is the ability to access the data share. Personalized listings allow you to control which consumers can access the data.

Standard Data Listings

A standard listing provides instant access to a published data set. This type of listing is best for providing generic, aggregated, or non-customer-specific data. Each listing includes details about the shared data, sample queries, and information about the data provider.

To access data from a standard data listing, consumers click **Get**, opening a prompt where they name the database for Snowflake users, permission access among Snowflake users, and agree to the provider's terms of use and Snowflake's consumer terms. Clicking **Create Database** immediately creates the database inside their Snowflake account in the Databases tab.

Only users that have a role with the IMPORT SHARE privilege can create databases from the Snowflake Data Marketplace.

Personalized Data Listings

A personalized listing allows customers to request specific data sets.

This can be premium data that a provider charges for or data that is specific to each consumer. Each listing includes details about the shared data, sample usage examples, and information about the data provider.

To access data from a personalized listing, consumers must submit a request by clicking the **Request** button, and provide their contact information. Once a request is submitted, the data provider is notified. The provider then contacts the consumer.

Each data provider can have different commercial terms. Once those are agreed to, then the personalized data set is created and shared with the consumer.

Question 85: Skipped

Snowflake is a SaaS offering

- TRUE
  (Correct)
- FALSE

Explanation
SaaS is software as a service. So, snowflake is not IaaS like Amazon Redshift since you do not need to manage infrastructure, it is not a PaaS as well as it is not a platform offered by Snowflake. It is a data warehousing software offered as a service on all three leading cloud providers(Google, AWS, Azure)

Question 86: Skipped

How will you check if a share is inbound or outbound?

- Execute SHOW SHARES and look at kind column
  (Correct)
- Execute describe share and look at location column

- Execute DISPLAY SHARE and look at kind column

Explanation

Use SHOW SHARES to confirm the share. The output of the command lists the details of the share. The `kind` column indicates that the share is OUTBOUND, meaning this share is sharing a database with other Snowflake accounts. The `to` column lists all accounts to which the share has been made available:

Question 87: Skipped

A share cannot be shared

- TRUE
(Correct)

- FALSE

Question 88: Skipped

What are the two types of data consumer accounts available in snowflake

- FULL ACCOUNT
(Correct)

- READER ACCOUNT
(Correct)

- SHARED ACCOUNT
- QUERY ONLY ACCOUNT

Explanation

Reader Accounts

Data sharing is only supported between Snowflake accounts. As a data provider, you might wish to share data with a consumer who does not already have a Snowflake account and/or is not ready to become a licensed Snowflake customer.

To facilitate sharing data with these consumers, Snowflake supports providers creating reader accounts. Reader accounts (formerly known as “read-only accounts”) provide a quick, easy, and cost-effective way to share data without requiring the consumer to become a Snowflake customer.

Users in a `reader account` can query data that has been shared with it, but cannot perform any of the DML tasks that are allowed in a `full account (data loading, insert, update, etc.).`

Question 89: Skipped

What is the maximum number of child tasks can a task have?

- 50
- 100
  (Correct)
- 1000
- 10

Explanation

A task is limited to a single predecessor task; however, `a task can have a maximum of 100 child tasks` (i.e. other tasks that identify the task as a predecessor); in addition, a simple tree of tasks is limited to a maximum of 1000 tasks ***total*** (including the root task) in either resumed (i.e. in a 'Started' state) or suspended.

For example, task T2 is limited to a single predecessor task, e.g. T1; however, T1 can serve as the predecessor task for tasks T2, T3, T4, etc.

Read the task chapter thoroughly

https://docs.snowflake.com/en/sql-reference/sql/create-task.html

Question 90: Skipped

A simple tree of tasks is limited to a maximum of 1000 tasks ***total***

- TRUE
  (Correct)
- FALSE

Explanation

A simple tree of tasks is limited to a maximum of `1000 tasks total` (including the root task) in either resumed (i.e. in a 'Started' state) or suspended.

Question 91: Skipped

What are the three features provided by Snowflake for continuous data pipeline

- Continuous data loading
  (Correct)
- Change data tracking
  (Correct)
- Recurring tasks
  (Correct)
- Cloud functions

Explanation
Snowflake provides the following features to enable continuous data pipelines:

Continuous data loading

Options for continuous data loading include the following:

Snowpipe

Snowflake Connector for Kafka

Third-party data integration tools

Change data tracking

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. A stream allows querying and consuming a set of changes to a table, at the row level, between two transactional points of time.

In a continuous data pipeline, table streams record when staging tables and any downstream tables are populated with data from business applications using continuous data loading and are ready for further processing using SQL statements.

For more information, see Change Tracking Using Table Streams.

Recurring tasks

A *task* object defines a recurring schedule for executing a SQL statement, including statements that call stored procedures. Tasks can be chained together for successive execution to support more complex periodic processing.

Tasks may optionally use table streams to provide a convenient way to continuously process new or changed data. A task can transform new or changed rows that a stream surfaces. Each time a task is scheduled to run, it can verify whether a stream contains change data for a table (using SYSTEM$STREAM_HAS_DATA) and either consume the change data or skip the current run if no change data exists.

Users can define a simple tree-like structure of tasks that executes consecutive SQL statements to process data and move it to various destination tables.

Question 92: Skipped

Multiple streams can be created for the same table and consumed by different tasks

- TRUE
  (Correct)
- FALSE

Explanation
This concept is very important for understand not just for the exam but for real world implementations as well.

Multiple tasks that consume change data `from a single table stream retrieve different deltas`. When a task consumes the change data in a stream using a DML statement, the stream advances the offset. The change data is no longer available for the next task to consume. Currently, we recommend that only a single task consumes the change data from a stream. `Multiple streams can be created for the same table and consumed by different tasks.`

https://docs.snowflake.com/en/user-guide/streams.html#stream-consumption-using-tasks

Question 93: Skipped

The cron expression in a task definition supports specifying a time zone

- TRUE
  (Correct)
- FALSE

Explanation
Task Scheduling and Daylight Saving Time

`The cron expression in a task definition supports specifying a time zone.` A scheduled task runs according to the specified cron expression in the local time for a given time zone. Special care should be taken with regard to scheduling tasks for time zones that recognize daylight saving time. Tasks scheduled during specific times on days when the transition from standard time to daylight saving time (or the reverse) occurs can have unexpected behaviors.

For example:

During the autumn change from daylight saving time to standard time, a task scheduled to start at 1 AM in the America/Los_Angeles time zone (i.e. `0 1 * * *` `America/Los_Angeles`) would run ***twice***: once at 1 AM and then again when 1:59:59 AM shifts to 1:00:00 AM local time. That is, there are two points in time when the local time is 1 AM.

During the spring change from standard time to daylight saving time, a task scheduled to start at 2 AM in the America/Los_Angeles time zone (i.e. `0 2 * * *` `America/Los_Angeles`) would ***not run at all*** because the local time shifts from 1:59:59 AM to 3:00:00 AM. That is, there is no point during that day when the local time is 2 AM.

To avoid unexpected task executions due to daylight saving time, ***either***:

Do not schedule tasks to run at a specific time between 1 AM and 3 AM (daily, or on days of the week that include Sundays), ***or***

Manually adjust the cron expression for tasks scheduled during those hours twice each year to compensate for the time change due to daylight saving time.

https://docs.snowflake.com/en/user-guide/tasks-intro.html#task-scheduling-and-daylight-saving-time

Question 94: Skipped

In snowflake data storage cost depends on which factors

- ☐ Cloud service provider
  (Correct)
- ☐ Region
  (Correct)
- ☐ Type of account we chose (on-demand/capacity)
  (Correct)

- Edition

Explanation
The storage cost is determined by cloud provider, region and the type of account. Please go the below site and try out the options

https://www.snowflake.com/pricing/

Question 95: Skipped

Your business team wants to load JSON data into a table in snowflake. What will you recommend them?

- Create a table with a variant column and load the JSON data in that column **(Correct)**
- Tell the team to convert JSON to CSV
- Construct a python program to convert the JSON into CSV and then load the data.

Explanation
This is very important to note and this is also a performance optimization technique. I have seen people creating columns as VARCHAR and then load JSON data in it. Never do that. Always create a variant columns and then load JSON data. Why? If you use VARCHAR, you will have to explicitly use parse_json for every row in the table which is additional processing. If you use VARIANT, you can use DOT or BRACKET notation to retrieve the data

https://docs.snowflake.com/en/user-guide/semistructured-intro.html#loading-semi-structured-data

https://docs.snowflake.com/en/sql-reference/data-types-semistructured.html#variant

Question 96: Skipped

JSON data is a hierarchical collection of name/value pairs grouped into objects and arrays

- TRUE

(Correct)

- FALSE

Explanation
Basic JSON Syntax

JSON data is a hierarchical collection of name/value pairs grouped into objects and arrays:

Colons `:` separate names and values in name/value pairs.

Curly braces `{}` denote objects.

Square brackets `[]` denote arrays.

Commas `,` separate entities in objects and arrays.

Name/Value Pairs

JSON name/value pairs consist of a field name (in double quotes), followed by a colon, then a value.

For example:

```
{"firstName":"John", "empid":45611}
```

Question 97: Skipped

Which parameter controls the maximum number of days for which Snowflake can extend the data retention period for tables to prevent streams on the tables from becoming stale.

- **MAX_DATA_EXTENSION_TIME_IN_DAYS**
  (Correct)
- **MAX_DATA_RETENTION_TIME_IN_DAYS**
- **MAX_DATA_EXTENSION_TIME_IN_MONTHS**

Explanation
MAX_DATA_EXTENSION_TIME_IN_DAYS

Type

Object (for databases, schemas, and tables) — Can be set for Account » Database » Schema » Table

Data Type

Integer

Description

Maximum number of days for which Snowflake can extend the data retention period for tables to prevent streams on the tables from becoming stale. By default, if the DATA_RETENTION_TIME_IN_DAYS setting for a source table is less than 14 days, and a stream has not been consumed, Snowflake temporarily extends this period to the stream's offset, up to a maximum of 14 days, regardless of the Snowflake Edition for your account. The MAX_DATA_EXTENSION_TIME_IN_DAYS parameter enables you to limit this automatic extension period to control storage costs for data retention or for compliance reasons.

This parameter can be set at the account, database, schema, and table levels. Note that setting the parameter at the account or schema level only affects tables for which the parameter has not already been explicitly set at a lower level (e.g. at the table level by the table owner). A value of `0` effectively disables the automatic extension for the specified database, schema, or table. For more information about streams and staleness, see Change Tracking Using Table Streams.

Values

`0` to `14` (i.e. 14 days) — a value of `0` disables the automatic extension of the data retention period. To increase the maximum value for tables in your account, contact Snowflake Support.

Default

`14`

https://docs.snowflake.com/en/sql-reference/parameters.html#max-data-extension-time-in-days

Question 98: Skipped

What is the size limit imposed by VARIANT data type.

- 100 MB compressed

- 16 MB compressed
  (Correct)
- 16 MB uncompressed
- 100 MB compressed

Explanation
VARIANT

A tagged universal type, which can store values of any other type, including OBJECT and ARRAY, u`p to a maximum size of 16 MB compressed.`

Question 99: Skipped

Non-native values such as dates and timestamps are stored as strings when loaded into a VARIANT column

- TRUE
  (Correct)
- FALSE

Explanation
For data that is mostly regular and uses only native JSON types (strings and numbers, not timestamps), both storage and query performance for operations on relational data and data in a VARIANT column is very similar. `Non-native values such as dates and timestamps are stored as strings when loaded into a VARIANT column`, so operations on these values could be slower and also consume more space than when stored in a relational column with the corresponding data type.

Question 100: Skipped

Streams cannot track changes in materialized views

- TRUE
  (Correct)
- FALSE

Explanation
Overview of Table Streams

When created, a table stream logically takes an initial snapshot of every row in the source table by initializing a point in time (called an *offset*) as the current transactional version of the table. The change tracking system utilized by the stream then records information about the DML changes after this snapshot was taken. Change records

provide the state of a row before and after the change. Change information mirrors the column structure of the tracked source table and includes additional metadata columns that describe each change event.

Note that a stream itself does ***not*** contain any table data. A stream only stores the offset for the source table and returns CDC records by leveraging the versioning history for the source table. When the first stream for a table is created, a pair of hidden columns are added to the source table and begin storing change tracking metadata. These columns consume a small amount of storage. The CDC records returned when querying a stream rely on a combination of the *offset* stored in the stream and the *change tracking metadata* stored in the table.

It may be useful to think of a stream as a bookmark, which indicates a point in time in the pages of a book (i.e. the source table). A bookmark may be thrown away and other bookmarks inserted in different places in a book. So too, a stream may be dropped and other streams created at the same or different points of time (either by creating the streams consecutively at different times or by using Time Travel) to consume the change records for a table at the same or different offsets.

One example of a consumer of CDC records is a data pipeline, in which only the data in staging tables that has changed since the last extraction is transformed and copied into other tables.

```
Currently, streams cannot track changes in materialized views.
```

Question 1: Skipped

Which role in Snowflake allows a user to administer users and manage all database objects?

- ACCOUNTADMIN
  (Correct)
- SYSADMIN
- SECURITYADMIN
- ROOT

Explanation
The account administrator (ACCOUNTADMIN) role is the most powerful role in the system. This role alone is responsible for configuring parameters at the account level. Users with the ACCOUNTADMIN role can view and operate on all objects in the account, can view and manage Snowflake billing and credit data, and can stop any running SQL statements.

In the default access control hierarchy, both of the other administrator roles are owned by this role:

The security administrator (SECURITYADMIN) role includes the privileges to create and manage users and roles.

The system administrator (SYSADMIN) role includes the privileges to create warehouses, databases, and all database objects (schemas, tables, etc.).

https://docs.snowflake.com/en/user-guide/security-access-control-considerations.html#using-the-accountadmin-role

Question 2: Skipped

Which transformations are available when using the COPY INTO command to load data files into Snowflake from a stage? (select all that apply)

- ☐ Column data type conversion **(Correct)**
- ☐ Column concatenation **(Correct)**
- ☐ Filters
- ☐ Aggregates

Explanation
Filtering the results of a FROM clause using a WHERE clause is not supported.

The VALIDATION_MODE parameter does not support COPY statements that transform data during a load.

https://docs.snowflake.com/en/user-guide/data-load-transform.html#supported-functions

Question 3: Skipped

Snowflake offers tools to extract data from source systems

- ○ TRUE
- ○

FALSE
(Correct)

Question 4: Skipped

Select the layers which are part of snowflake(select 3)

- STORAGE
(Correct)
- DATA CATALOG
- VIRTUAL WAREHOUSE
(Correct)
- CLOUD SERVICES
(Correct)

Explanation
Snowflake's novel design physically separates but logically integrates storage, compute and services like security and metadata; we call it multi-cluster, shared data and it consists of 3 components:

1. Storage: the persistent storage layer for data stored in Snowflake

2. Compute: a collection of independent compute resources that execute data processing tasks required for queries

3. Services: a collection of system services that handle infrastructure, security, metadata, and optimization across the entire Snowflake system

https://www.snowflake.com/wp-content/uploads/2014/10/A-Detailed-View-Inside-Snowflake.pdf

Question 5: Skipped

When data is staged to a Snowflake internal staging area using the PUT command, the data is encrypted on the client's machine

- TRUE
(Correct)
- FALSE

Explanation

Uploaded files are automatically encrypted with 128-bit or 256-bit keys. The CLIENT_ENCRYPTION_KEY_SIZE parameter specifies the size key used to encrypt the files

https://docs.snowflake.com/en/sql-reference/sql/put.html#usage-notes

Question 6: Skipped

Which snowflake features are available for enabling continuous data pipelines

- Continuous data loading
  (Correct)
- Change data tracking
  (Correct)
- Recurring tasks
  (Correct)
- Table Pipes

Explanation
Snowflake provides the following features to enable continuous data pipelines:

Continuous data loading

Options for continuous data loading include the following:

Snowpipe

Snowflake Connector for Kafka

Third-party data integration tools

Change data tracking

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. A stream allows querying and consuming a set of changes to a table, at the row level, between two transactional points of time.

In a continuous data pipeline, table streams record when staging tables and any downstream tables are populated with data from business applications using continuous data loading and are ready for further processing using SQL statements.

For more information, see Change Tracking Using Table Streams.

Recurring tasks

A *task* object defines a recurring schedule for executing a SQL statement, including statements that call stored procedures. Tasks can be chained together for successive execution to support more complex periodic processing.

Tasks may optionally use table streams to provide a convenient way to continuously process new or changed data. A task can transform new or changed rows that a stream surfaces. Each time a task is scheduled to run, it can verify whether a stream contains change data for a table (using SYSTEM$STREAM_HAS_DATA) and either consume the change data or skip the current run if no change data exists.

Users can define a simple tree-like structure of tasks that executes consecutive SQL statements to process data and move it to various destination tables.

https://docs.snowflake.com/en/user-guide/data-pipelines-intro.html#introduction-to-data-pipelines

Question 7: Skipped

Select the statements which are true for an external table

- ☐ External tables are read-only
  (Correct)
- ☐ External tables can be used for query and join operations
  (Correct)
- ☐ Views can be created against external tables
  (Correct)
- ☐ Data can be updated in external tables

Explanation
In a typical table, the data is stored in the database; however, in an external table, the data is stored in files in an external stage. External tables store file-level metadata about the data files, such as the filename, a version identifier and related properties.

This enables querying data stored in files in an external stage as if it were inside a database. External tables can access data stored in any format supported by COPY INTO <table> statements.

External tables are read-only, therefore no DML operations can be performed on them; however, external tables can be used for query and join operations. Views can be created against external tables.

Querying data stored external to the database is likely to be slower than querying native database tables; however, materialized views based on external tables can improve query performance.

https://docs.snowflake.com/en/user-guide/tables-external-intro.html

Question 8: Skipped

What are the two data loading approaches in snowflake

- BULK LOADING
  (Correct)
- CONTINUOUS LOADING
  (Correct)
- INGEST LOADING

Explanation
BULK LOADING and CONTINUOUS LOADING are the two approaches. For bulk loading you can use the COPY command and select a relevant warehouse to perform the COPY. In this case compute is user managed

For continuous loading you can use snowpipe which is a serverless way of loading data in micro batches. Compute is managed by snowflake.

Question 9: Skipped

Which of the below are considered as best practices while loading data into snowflake?

- Isolate data loading workload into its own virtual warehouse
  (Correct)
- Split large files into smaller files
  (Correct)

- Compress the source files
  (Correct)
- If format is in CSV, convert them into ORC

Explanation
Having a dedicated warehouse for load workload ensures that it will not be interrupted by any other workloads. Splitting the large files is a good practice because it enables snowflake to parallize the load operation. Since the files are sent over the wire, it is always better to compress them first.

Question 10: Skipped

Load performance in snowflake is fastest for which file format

- CSV
  (Correct)
- ORC
- AVRO
- PARQUET

Question 11: Skipped

COPY and INSERT operations in snowflake are non-blocking

- TRUE
  (Correct)
- FALSE

Explanation
COPY and INSERT do not block any other operations on the table

Question 12: Skipped

Organizing input data by granular path can improve load performance

- TRUE
  (Correct)
- FALSE

Question 13: Skipped

Which are the key concepts that will need to be considered while loading data into snowflake

-

STAGE OBJECT
(Correct)

- FILE FORMAT
(Correct)
- TRANSFORMATION AND ERROR VALIDATION
(Correct)
- FILE SIZE

Explanation
Copying the file to a stage object is recommended while loading data into snowflake. File format is used to identify the data format (CSV, JSON etc) of the source file. Minor transformation and validations can be done as part of loading data

Question 14: Skipped

Which approach would result in improved performance through linear scaling of data ingestion workload?

- Resize virtual warehouse
- Consider practice of organizing data by granular path
- Consider practice of splitting input file batch within the recommended size of 10 MB to 100 MB
- All of the above
(Correct)

Question 15: Skipped

Which are the two variant columns available in a snowflake table loaded by kafka connector

- RECORD_CONTENT
(Correct)
- RECORD_METADATA
(Correct)
- RECORD_KEY

Explanation
Every Snowflake table loaded by the Kafka connector has a schema consisting of two VARIANT columns:

RECORD_CONTENT. This contains the Kafka message.

RECORD_METADATA. This contains metadata about the message, for example, the topic from which the message was read.

If Snowflake creates the table, then the table contains only these two columns. If the user creates the table for the Kafka Connector to add rows to, then the table can contain more than these two columns (any additional columns must allow NULL values because data from the connector does not include values for those columns).

https://docs.snowflake.com/en/user-guide/kafka-connector-overview.html#schema-of-topics-for-kafka-topics

Question 16: Skipped

The RECORD_METADATA contains which information

- ☐ Topic **(Correct)**
- ☐ Partition **(Correct)**
- ☐ Key **(Correct)**
- ☐ CreateTime / LogAppendTime **(Correct)**
- ☐ Value

Explanation
The RECORD_METADATA column contains the following information by default:

Topic

Partition

Offset

CreateTime / LogAppendTime

key

schema_id

headers

The value of the message is in RECORD_CONTENT

https://docs.snowflake.com/en/user-guide/kafka-connector-overview.html#schema-of-topics-for-kafka-topics

Question 17: Skipped

If multiple instances of the kafka connector is started on the same topic or partitions, duplicate records may flow into snowflake table

- TRUE
(Correct)
- FALSE

Explanation
Instances of the Kafka connector do not communicate with each other. If you start multiple instances of the connector on the same topics or partitions, then multiple copies of the same row might be inserted into the table. This is not recommended; each topic should be processed by only one instance of the connector.

Question 18: Skipped

Select the ones that are true for snowflake kafka connector

- Kafka connector guarantees exactly-once delivery
(Correct)
- Kafka connector guarantees that rows are inserted in the order
- Kafka connector guarantees reprocessing of messages

Explanation
Although the Kafka connector guarantees exactly-once delivery, it does ***not*** guarantee that rows are inserted in the order that they were originally published.

Question 19: Skipped

Which are true with respect to SMT(simple message transformation) when neither `key.converter` or `value.converter` is set

- All SMTs are supported
- Most SMTs are supported
(Correct)

- regex.router is not supported
  (Correct)

Explanation
Single Message Transformations (SMTs) are applied to messages as they flow through Kafka Connect. When you configure the Kafka Configuration Properties, if you set ***either*** `key.converter` or `value.converter` to one of the following values, then SMTs are not supported on the corresponding key or value:

`com.snowflake.kafka.connector.records.SnowflakeJsonConverter`

`com.snowflake.kafka.connector.records.SnowflakeAvroConverter`

`com.snowflake.kafka.connector.records.SnowflakeAvroConverterWithoutSchemaRegistry`

When neither `key.converter` or `value.converter` is set, then most SMTs are supported, with the current exception of `regex.router`.

https://docs.snowflake.com/en/user-guide/kafka-connector-overview.html#kafka-connector-limitations

Question 20: Skipped

Which are the supported data types for a JSON name/value pair?

- A number (integer or floating point)
  (Correct)
- A string (in double quotes)
  (Correct)
- A Boolean (true or false)
  (Correct)
- An array (in square brackets)
  (Correct)
- An object (in curly braces)
  (Correct)

- Null
  (Correct)
- complex datatype

Explanation
A value in a name/value pair can be:

1. A number (integer or floating point)

2. A string (in double quotes)

3. A Boolean (true or false)

4. An array (in square brackets)

5. An object (in curly braces)

6. Null

https://docs.snowflake.com/en/user-guide/semistructured-intro.html#supported-data-types

Question 21: Skipped

Which of the below are binary formats?

- AVRO
  (Correct)
- JSON
- ORC
  (Correct)
- PARQUET
  (Correct)

Explanation
What is ORC?

Used to store Hive data, the ORC (Optimized Row Columnar) file format was designed for efficient compression and improved performance for reading, writing, and processing data over earlier Hive file formats. For more information about ORC, see https://orc.apache.org/.

Snowflake reads ORC data into a single VARIANT column. You can query the data in a VARIANT column just as you would JSON data, using similar commands and functions.

Alternatively, you can extract select columns from a staged ORC file into separate table columns using a CREATE TABLE AS SELECT statement.

ORC is a binary format.

What is parquet?

Parquet is a compressed, efficient columnar data representation designed for projects in the Hadoop ecosystem. The file format supports complex nested data structures and uses Dremel record shredding and assembly algorithms. For more information, see parquet.apache.org/documentation/latest/ .

Snowflake reads Parquet data into a single VARIANT column. You can query the data in a VARIANT column just as you would JSON data, using similar commands and functions.

Alternatively, you can extract select columns from a staged Parquet file into separate table columns using a CREATE TABLE AS SELECT statement.

Parquet is a binary format.

What is AVRO?

Avro is an open-source data serialization and RPC framework originally developed for use with Apache Hadoop. It utilizes schemas defined in JSON to produce serialized data in a compact binary format. The serialized data can be sent to any destination (i.e. application or program) and can be easily deserialized at the destination because the schema is included in the data.

An Avro schema consists of a JSON string, object, or array that defines the type of schema and the data attributes (field names, data types, etc.) for the schema type. The attributes differ depending on the schema type. Complex data types such as arrays and maps are supported.

Snowflake reads Avro data into a single VARIANT column. You can query the data in a VARIANT column just as you would JSON data, using similar commands and functions.

https://docs.snowflake.com/en/user-guide/semistructured-intro.html#what-is-parquet

https://docs.snowflake.com/en/user-guide/semistructured-intro.html#what-is-orc

https://docs.snowflake.com/en/user-guide/semistructured-intro.html#what-is-avro

Question 22: Skipped

Which constrains are enforced in snowflake

- Referential integrity constraints
- NOT NULL constraint
  (Correct)
- UNIQUE Constraint

Explanation
Referential integrity constraints in Snowflake are informational and, with the exception of NOT NULL, not enforced. Constraints other than NOT NULL are created as disabled.

However, constraints provide valuable metadata. The primary keys and foreign keys enable members of your project team to orient themselves to the schema design and familiarize themselves with how the tables relate with one another.

https://docs.snowflake.com/en/user-guide/table-considerations.html#referential-integrity-constraints

Question 23: Skipped

You want to get the DDL statement of a snowflake table. What is the command that you will use?

- ```
  1. select get_ddl('table', 'mydb.public.salesorders');
  ```
  (Correct)
- ```
  1. show table 'mydb.public.salesorders';
  ```
- ```
  1. show table like 'mydb.public.salesorders';
  ```

Explanation
Query the GET_DDL function to retrieve a DDL statement that could be executed to recreate the specified table. The statement includes the constraints currently set on a table.

Question 24: Skipped

You have a small table in snowflake which has only 10,000 rows. Specifying a clustering key will further improved the queries that run on this table

- FALSE
  (Correct)
-
  ```

TRUE

Explanation

Specifying a clustering key is not necessary for most tables. Snowflake performs automatic tuning via the optimization engine and micro-partitioning. In many cases, data is loaded and organized into micro-partitions by date or timestamp, and is queried along the same dimension.

When should you specify a clustering key for a table? First, note that clustering a small table typically doesn't improve query performance significantly.

For larger data sets, you might consider specifying a clustering key for a table when:

The order in which the data is loaded does not match the dimension by which it is most commonly queried (e.g. the data is loaded by date, but reports filter the data by ID). If your existing scripts or reports query the data by both date ***and*** ID (and potentially a third or fourth column), you may see some performance improvement by creating a multi-column clustering key.

Query Profile indicates that a significant percentage of the total duration time for typical queries against the table is spent scanning. This applies to queries that filter on one or more specific columns.

Note that reclustering rewrites existing data with a different order. The previous ordering is stored for 7 days to provide Fail-safe protection. Reclustering a table incurs compute costs that correlate to the size of the data that is reordered.

https://docs.snowflake.com/en/user-guide/table-considerations.html#when-to-set-a-clustering-key

Question 25: Skipped

There is no query performance difference between a column with a maximum length declaration (e.g. `VARCHAR(16777216)`), and a smaller precision. Still it is recommended to define an appropriate column length because of the below reasons

- Data loading operations are more likely to detect issues such as columns loaded out of order, e.g. a 50-character string loaded erroneously into a VARCHAR(10) column. Such issues produce errors
  (Correct)

- When the column length is unspecified, some third-party tools may anticipate consuming the maximum size value, which can translate into increased client-side memory usage or unusual behavior

(Correct)

- Data unloading will be performant if appropriate column lengths are defined

Explanation
Snowflake compresses column data effectively; therefore, creating columns larger than necessary has minimal impact on the size of data tables. Likewise, there is no query performance difference between a column with a maximum length declaration (e.g. `VARCHAR(16777216)`), and a smaller precision.

However, when the size of your column data is predictable, we do recommend defining an appropriate column length, for the following reasons:

Data loading operations are more likely to detect issues such as columns loaded out of order, e.g. a 50-character string loaded erroneously into a VARCHAR(10) column. Such issues produce errors.

When the column length is unspecified, some third-party tools may anticipate consuming the maximum size value, which can translate into increased client-side memory usage or unusual behavior.

https://docs.snowflake.com/en/user-guide/table-considerations.html#when-to-specify-column-lengths

Question 26: Skipped

You want to convert an existing permanent table to a transient table (or vice versa) while preserving data and other characteristics such as column defaults and granted privileges. What is the best way to do it?

- Run an ALTER TABLE command to convert the tables
- Unload the data from the existing table into a CSV file. Create the new table and then load the data back in
- Create a new table and use the `COPY GRANTS` clause

(Correct)

Explanation
Currently, it isn't possible to change a permanent table to a transient table using the ALTER TABLE command. The TRANSIENT property is set at table creation and cannot be modified.

Similarly, it isn't possible to directly change a transient table to a permanent table.

To convert an existing permanent table to a transient table (or vice versa) while preserving data and other characteristics such as column defaults and granted privileges, you can create a new table and use the `COPY GRANTS` clause, then copy the data:

```
create transient table my_new_table like my_old_table copy grants;

insert into my_new_table select * from my_old_table;
```

https://docs.snowflake.com/en/user-guide/table-considerations.html#converting-a-permanent-table-to-a-transient-table-or-vice-versa

Question 27: Skipped

A transient table can be cloned to a permanent table

- TRUE
- FALSE
  (Correct)

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

Question 28: Skipped

If you clone a permanent table(bar) into a transient table(foo) using the below command

```
1. create transient table foo clone bar copy grants;
```

What will happen to the partitions?

- Old partitions will ***not*** be affected,but new partitions added to the clone will follow the transient lifecycle
  (Correct)
- All the partitions will be affected
- Only the old partitions will be affected

Explanation
Another way to make a copy of a table (but change the lifecycle from permanent to transient) is to `CLONE` the table, for example:

```
create transient table foo clone bar copy grants;
```

Old partitions will ***not*** be affected (i.e. won't become transient), but new partitions added to the clone will follow the transient lifecycle.

Question 29: Skipped

You want to identify the potential performance bottlenecks and improvement opportunities of a query. What will you do?

- Use Query Profile
  (Correct)
- Use Explain plan
- Call snowflake support

Explanation
Query Profile is a powerful tool for understanding the mechanics of queries. It can be used whenever you want or need to know more about the performance or behavior of a particular query. It is designed to help you spot typical mistakes in SQL query expressions to identify potential performance bottlenecks and improvement opportunities.

https://docs.snowflake.com/en/user-guide/ui-query-profile.html#when-to-use-query-profile

Question 30: Skipped

What does snowflake use for monitoring network traffic and use activity?

- Lacework
  (Correct)
- Sumo logic
- Threat Stack

Explanation
Snowflake uses lacework for behavioral monitoring of production infrastructure which includes network traffic and user activity. It uses Sumo Logic and Threat Stack to monitor failed logins, file integrity monitoring and unauthorized system modifications.

Question 31: Skipped

In which scenarios would you consider to use materialized views

- The query results contain a small number of rows and/or columns relative to the base table
  (Correct)
- Query results contain results that require significant processing
  (Correct)

- Query is on an external table
  (Correct)
- View's base table does not change frequently
  (Correct)
- None of the above

Explanation
Materialized views are particularly useful when:

1. Query results contain a small number of rows and/or columns relative to the base table (the table on which the view is defined).

2. Query results contain results that require significant processing, including:

3. Analysis of semi-structured data.

4. Aggregates that take a long time to calculate.

5. 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.

6. The view's base table does not change frequently.

https://docs.snowflake.com/en/user-guide/views-materialized.html#when-to-use-materialized-views

Question 32: Skipped

What will the below query return

```
SELECT TOP 10 GRADES FROM STUDENT;
```

- The top 10 highest grades
- The 10 lowest grades
- Non-deterministic list of 10 grades
  (Correct)

Explanation

An ORDER BY clause is not required; however, without an ORDER BY clause, the **results are non-deterministic** because results within a result set are not necessarily in any particular order. To control the results returned, use an ORDER BY clause.

n must be a non-negative integer constant.

https://docs.snowflake.com/en/sql-reference/constructs/top_n.html#usage-notes

Question 33: Skipped

Loading data using snowpipe REST API is supported for external stage only

- TRUE
- FALSE
  (Correct)

Explanation
Snowpipe supports loading from the following stage types:

1. Named internal (Snowflake) or external (Amazon S3, Google Cloud Storage, or Microsoft Azure) stages

2. Table stages

https://docs.snowflake.com/en/user-guide/data-load-snowpipe-rest-gs.html#step-1-create-a-stage-if-needed

Question 34: Skipped

With default settings, how long will a query run on snowflake

- Snowflake will cancel the query if it runs more than 48 hours
  (Correct)
- Snowflake will cancel the query if it runs more than 24 hours
- Snowflake will cancel the query if the warehouse runs out of memory
- Snowflake will cancel the query if the warehouse runs out of memory and hard disk storage

Explanation
STATEMENT_TIMEOUT_IN_SECONDS

This parameter tells Snowflake how long can a SQL statement run before the system cancels it. The default value is 172800 seconds (48 hours)

This is both a session and object type parameter. As a session type, it can be applied to the account, a user or a session. As an object type, it can be applied to warehouses. If set at both levels, the lowest value is used.

Question 35: Skipped

You have created a TASK in snowflake. How will you resume it?

- No need to resume, the creation operation automatically enables the task
- ALTER TASK mytask1 RESUME;
  (Correct)
- ALTER TASK mytask1 START;

Explanation
It is important to remember that a Task that has just been created will be suspended by default. It is necessary to manually enable this task by “altering” the task as follows:

ALTER TASK mytask1 RESUME;

Question 36: Skipped

What technique does snowflake use to limit the number of micro-partitions scanned by each query

- Indexing
- Pruning
  (Correct)
- Map Reduce
- B-tree

Explanation
The micro-partition metadata maintained by Snowflake enables precise pruning of columns in micro-partitions at query run-time, including columns containing semi-structured data. In other words, a query that specifies a filter predicate on a range of values that accesses 10% of the values in the range should ideally only scan 10% of the micro-partitions.

For example, assume a large table contains one year of historical data with date and hour columns. Assuming uniform distribution of the data, a query targeting a particular

hour would ideally scan 1/8760th of the micro-partitions in the table and then only scan the portion of the micro-partitions that contain the data for the hour column; Snowflake uses columnar scanning of partitions so that an entire partition is not scanned if a query only filters by one column.

In other words, the closer the ratio of scanned micro-partitions and columnar data is to the ratio of actual data selected, the more efficient is the pruning performed on the table.

For time-series data, this level of pruning enables potentially sub-second response times for queries within ranges (i.e. “slices”) as fine-grained as one hour or even less.

Not all predicate expressions can be used to prune. For example, Snowflake does not prune micro-partitions based on a predicate with a subquery, even if the subquery results in a constant.

https://docs.snowflake.com/en/user-guide/tables-clustering-micropartitions.html#query-pruning

Question 37: Skipped

Which of the two statements are true about the variant data type in SnowFlake?

- Optimized storage based on repeated elements
(Correct)
- Stored in a seperate file format from structured data
- Can be queried using json path notation
(Correct)
- Requires a custom mapping for each record type

Explanation
When Snowflake loads semi-structured data, it optimizes how it stores that data internally by automatically discovering the attributes and structure that exist in the data, and using that knowledge to optimize how the data is stored. Snowflake also looks for repeated attributes across records, organizing and storing those repeated attributes separately. This enables better compression and faster access, similar to the way that a columnar database optimizes storage of columns of data.

Question 38: Skipped

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

-

A pre-defined mapping

- A view
(Correct)

- Leveraging a json parser
- BI tool cannot access json

Question 39: Skipped

You can map snowflake to any s3 bucket and can query the data directly as long as the data is in Parquet or ORC format

- TRUE
- FALSE
(Correct)

Question 40: Skipped

The following factors affect data load rates

- Physical location of the stage
(Correct)

- Virtual warehouse RAM
- Gzip compression efficiency
(Correct)

- Thread size

Question 41: Skipped

What are the two mechanisms to detect if new stage file is there in a snowpipe?

- Automating Snowpipe using cloud messaging
(Correct)

- Calling Snowpipe REST endpoints
(Correct)

- Calling the custom APIs exposed through AWS EKS

Explanation
https://docs.snowflake.com/en/user-guide/data-load-snowpipe-intro.html#how-does-snowpipe-work

Snowpipe loads data from files `as soon as they are available in a stage`. The data is loaded according to the COPY statement defined in a referenced pipe.

A pipe is a named, first-class Snowflake object that contains a COPY statement used by Snowpipe. The COPY statement identifies the source location of the data files (i.e., a stage) and a target table. All data types are supported, including semi-structured data types such as JSON and Avro.

Different mechanisms for detecting the staged files are available:

Automating Snowpipe using cloud messaging

Automated data loads leverage event notifications for cloud storage to inform Snowpipe of the arrival of new data files to load. Snowpipe copies the files into a queue, from which they are loaded into the target table in a continuous, serverless fashion based on parameters defined in a specified pipe object.

Snowflake currently supports the following storage account types:

Amazon Web Services (AWS)

Amazon S3

Microsoft Azure

Blob storage

Data Lake Storage Gen2 — Supported as a preview feature.

General-purpose v2

For more information, see Automating Continuous Data Loading Using Cloud Messaging.

Calling Snowpipe REST endpoints

Your client application calls a public REST endpoint with the name of a pipe object and a list of data filenames. If new data files matching the list are discovered in the stage referenced by the pipe object, they are queued for loading. Snowflake-provided compute resources load data from the queue into a Snowflake table based on parameters defined in the pipe.

Amazon Web Services (AWS)

Amazon S3

Google Cloud Platform

Cloud Storage

Microsoft Azure

Blob storage

Data Lake Storage Gen2 — Supported as a preview feature.

General-purpose v2

For more information, see Calling Snowpipe REST Endpoints to Load Data.

Question 42: Skipped

When you load data using Snowpipe, loads are always performed in a single transaction

- FALSE
  (Correct)
- TRUE

Explanation
It is important to know the difference between the two techniques, so not just on this transaction related questions, please also go through the other differences here

`https://docs.snowflake.com/en/user-guide/data-load-snowpipe-intro.html#how-is-snowpipe-different-from-bulk-data-loading`

Transactions

Bulk data load

Loads are always performed in a single transaction. Data is inserted into table alongside any other SQL statements submitted manually by users.

Snowpipe

Loads are combined or split into a single or multiple transactions based on the number and size of the rows in each data file. Rows of partially loaded files (based on the ON_ERROR copy option setting) can also be combined or split into one or more transactions.

Question 43: Skipped

Snowpipe does not guarantee loading of files in the order that they are staged

- FALSE
- TRUE
  (Correct)

Explanation

`Very important to remember this`

Load Order of Data Files

For each pipe object, Snowflake establishes a single queue to sequence data files awaiting loading. As new data files are discovered in a stage, Snowpipe appends them to the queue. However, multiple processes pull files from the queue; and so, while Snowpipe generally loads older files first, there is no guarantee that files are loaded in the same order they are staged.

Question 44: Skipped

What command will you run to pause a pipe?

- ```
  1.  alter pipe <pipe name> set pipe_execution_paused = true;
  ```
  (Correct)
- ```
  1.  alter pipe <pipe name> set pipe_execution_paused = stop;
  ```
- ```
  1.  alter pipe <pipe name> set pipe_execution_paused = halt;
  ```

Explanation

Pause the `mypipe` pipe:

```
alter pipe mypipe set pipe_execution_paused = true;
```

https://docs.snowflake.com/en/sql-reference/sql/alter-pipe.html#examples

Question 45: Skipped

All of the below are valid executionState of a snowpipe except:

- RUNNING
- STOPPED_FEATURE_DISABLED
-
```

STALLED_EXECUTION_ERROR

- PAUSED
- STOPPED
  (Correct)

Explanation

`executionState`

Current execution state of the pipe; could be any one of the following:

`RUNNING` (i.e. everything is normal; Snowflake may or may not be actively processing files for this pipe)

`STOPPED_FEATURE_DISABLED`

`STOPPED_STAGE_DROPPED`

`STOPPED_FILE_FORMAT_DROPPED`

`STOPPED_MISSING_PIPE`

`STOPPED_MISSING_TABLE`

`STALLED_COMPILATION_ERROR`

`STALLED_INITIALIZATION_ERROR`

`STALLED_EXECUTION_ERROR`

`STALLED_INTERNAL_ERROR`

`PAUSED`

`PAUSED_BY_SNOWFLAKE_ADMIN`

`PAUSED_BY_ACCOUNT_ADMIN`

Question 46: Skipped

How do you set a return value in a task?

-
```
create task set_return_value
  warehouse=return_task_wh
  schedule='1 minute' as
  call system$set_return_value('The quick brown fox jumps over the lazy dog');
```

(Correct)

-
```
create task set_return_value
  warehouse=return_task_wh
  schedule='1 minute' as
  call system$set_return_code('The quick brown fox jumps over the lazy dog');
```

-
```
create task set_return_value
  warehouse=return_task_wh
  schedule='1 minute' as
  call set_return_value('The quick brown fox jumps over the lazy dog');
```

Explanation
SYSTEM$SET_RETURN_VALUE

Explicitly sets the return value for a task.

In a tree of tasks, a task can call this function to set a return value. Another task that identifies this task as the predecessor task (using the `AFTER` keyword in the task definition) can retrieve the return value set by the predecessor task.

https://docs.snowflake.com/en/sql-reference/functions/system_set_return_value.html#examples

Question 47: Skipped

Query load is calculated by dividing the execution time (in seconds) of all queries in an interval by the total time (in seconds) for the interval.

- TRUE
(Correct)

- FALSE

Explanation
https://docs.snowflake.com/en/user-guide/warehouses-load-monitoring.html#how-query-load-is-calculated

Question 48: Skipped

Resource monitors can be used to control credit usage for the Snowflake-provided warehouses, including the snowpipe warehouse

- TRUE

- FALSE
  (Correct)

Explanation
Resource monitors provide control over virtual warehouse credit usage; however, `you cannot use them` to control credit usage for the Snowflake-provided warehouses, including the **SNOWPIPE** warehouse.

Question 49: Skipped

This snowpipe rest API Fetches a report about ingested files whose contents have been added to table

- `loadHistoryScan and insertReport`
  (Correct)
- `insertPipeReport`
- `insertFiles`

Explanation
Endpoint: `loadHistoryScan`

Fetches a report about ingested files whose contents have been added to table. Note that for large files, this may only be part of the file. This endpoint differs from `insertReport` in that it views the history between two points in time. There is a maximum of 10,000 items returned, but multiple calls can be issued to cover the desired time range.

Additional explanation

Please note there was a mistake in this question earlier. The earlier question's answer selected only loadHistoryScan but for this question both `loadHistoryScan` and `insertReport` are correct. Please see the explanation above to see the difference between `loadHistoryScan` and `insertReport`

Thanks to Rupa who caught this mistake

Question 50: Skipped

To help avoid exceeding the rate limit (error code 429), snowflake recommends relying more heavily on `insertReport` than `loadHistoryScan`

- TRUE
  (Correct)
- FALSE

Explanation
loadHistoryScan endpoint is rate limited to avoid excessive calls. To help avoid exceeding the rate limit (error code 429), we recommend relying more heavily on `insertReport` than `loadHistoryScan`. When calling `loadHistoryScan`, specify the most narrow time range that includes a set of data loads. For example, reading the last 10 minutes of history every 8 minutes would work well. Trying to read the last 24 hours of history every minute will result in 429 errors indicating a rate limit has been reached. The rate limits are designed to allow each history record to be read a handful of times.

https://docs.snowflake.com/en/user-guide/data-load-snowpipe-rest-apis.html#endpoint-loadhistoryscan

Question 51: Skipped

Select two true statement related to streams

- Stream itself does ***not*** contain any table data
  (Correct)
- A stream only stores the offset for the source table
  (Correct)
- The hidden columns used by a stream does not consume any storage

Explanation
Note that a stream itself does ***not*** contain any table data. A stream only stores the offset for the source table and returns CDC records by leveraging the versioning history for the source table. When the first stream for a table is created, a pair of hidden columns are added to the source table and begin storing change tracking metadata. `These columns consume a small amount of storage`. The CDC records returned when querying a stream rely on a combination of the *offset* stored in the stream and the *change tracking metadata* stored in the table.

Question 52: Skipped

You have a Snowflake table which is defined as below

```
CREATE OR REPLACE TABLE FRUITS(FRUIT_NUMBER NUMBER, FRUIT_DESCRIPTION
VARCHAR, AVAILABILITY VARCHAR);
```

If you would like to convert the fruit_number column to be a decimal with a certain precision and scale, which command will you run?

- SELECT FRUIT_NUMBER::DECIMAL(10,5) FROM FRUITS;
  (Correct)
- SELECT FRUIT_NUMBER(DECIMAL(10,5)) FROM FRUITS;
- SELECT FRUIT_NUMBER AS DECIMAL(10,5) FROM FRUITS;
- SELECT FRUIT_NUMBER.DECIMAL(10,5) FROM FRUITS;

Explanation
This question on CAST may come in may forms, but just remember that you can CAST to a specific datatype using '::' or by using the CAST command

CAST , `::`

Converts a value of one data type into another data type. The semantics of CAST are the same as the semantics of the corresponding TO_ *`datatype`* conversion functions. If the cast is not possible, an error is raised. For more details, see the individual TO_ *`datatype`* conversion functions.

The `::` operator provides alternative syntax for CAST.

I could have also written the query as below

```
SELECT CAST(FRUIT_NUMBER AS DECIMAL(10,5)) FROM FRUITS;
```

Question 53: Skipped

You ran a query in snowflake and went to query history tab. The query history shows you the below columns

1. QueryID

2. SQL TEXT

3. WAREHOUSE NAME

4. WAREHOUSE SIZE

5. SESSION ID

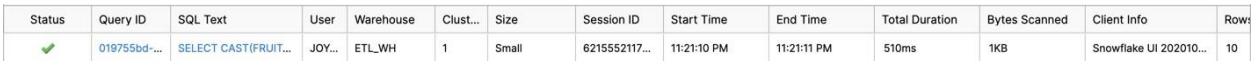
6. START TIME

7. END TIME

Which of the above column will indicate if an compute cost was incurred to run the query?

- WAREHOUSE NAME
- CREDIT
- WAREHOUSE SIZE
  (Correct)
- SESSION ID

Explanation
Anytime a query incurs compute cost, you will see the warehouse size mentioned as shown below



Status	Query ID	SQL Text	User	Warehouse	Clust...	Size	Session ID	Start Time	End Time	Total Duration	Bytes Scanned	Client Info	Rows
✔	019755bd-...	SELECT CAST(FRUIT...	JOY...	ETL_WH	1	Small	6215552117...	11:21:10 PM	11:21:11 PM	510ms	1KB	Snowflake UI 202010...	10

Just fo an experiment as below

1. Run the query `SHOW TABLES`

2. Go to query history

Do you see the ware house size? No, because SHOW TABLES is a metadata query and does not incur any compute cost

Question 54: Skipped

The FLATTEN command in snowflake has two version. One version uses a join and the other version uses an object keyword. Please select two words that represent the options used with the command?

- OBJECT_CONSTRUCT
- TABLE
  (Correct)
- TRY_CAST
- LATERAL
  (Correct)

Explanation
An example of the command is as below

```
select * from table(flatten(input => parse_json('[1, ,77]'))) f;
```

Question 55: Skipped

You have two types of named stages, one is external stage and the other one is internal stage. External stage will always require a cloud storage provider

- TRUE
  (Correct)
- FALSE

Explanation
Ok, this is an easy question. But when you are working for a customer, what will you suggest him? When should we use external stage? Please read below

It is preferred to use internal stage because SNOWFLAKE automatically encrypts the data in internal stage. SNOWFLAKE is responsible for that encryption. If you use external stage, it will be your responsibility to encrypt the data. But does it mean that you will never use external stage. Well no, there may be use cases where the data is not much and is coming from an external stakeholder and you do not want to store that data into your snowflake tables, in such cases go ahead and use external stage. But always do an analysis of the use case in hand

Question 56: Skipped

When cloning a table, if the COPY GRANTS keywords are not included in the CREATE <*object*> statement, then the new object does ***not*** inherit any explicit access privileges granted on the original table but does inherit any future grants defined for the object type in the schema

- TRUE
  (Correct)
- FALSE

Explanation
General Usage Notes

A clone is writable and is independent of its source (i.e. changes made to the source or clone are not reflected in the other object).

To create a clone, your current role must have the following privilege(s) on the source object:

Tables

SELECT

Pipes, Streams, Tasks

OWNERSHIP

Other objects

USAGE

In addition, to clone a schema or an object within a schema, your current role must have required privileges on the container object(s) for both the source and the clone.

For tables, Snowflake only supports cloning permanent and transient tables; temporary tables cannot be cloned.

For databases and schemas, cloning is recursive:

Cloning a database clones all the schemas and other objects in the database.

Cloning a schema clones all the contained objects in the schema.

However, the following object types are ***not*** cloned:

External tables

Internal (Snowflake) stages

For databases, schemas, and tables, a clone does ***not*** contribute to the overall data storage for the object until operations are performed on the clone that modify existing data or add new data, such as:

Adding, deleting, or modifying rows in a cloned table.

Creating a new, populated table in a cloned schema.

Cloning a table replicates the structure, data, and certain other properties (e.g. `STAGE FILE FORMAT`) of the source table. A cloned table does ***not*** include the load history of the source table. Data files that were loaded into a source table can be loaded again into its clones.

When cloning tables, the CREATE *<object>* command syntax includes the COPY GRANTS keywords:

```
If the COPY GRANTS keywords are not included in the CREATE <object> statement, th
en the new object does not inherit any explicit access privileges granted on the
original table but does inherit any future grants defined for the object type in
the schema (using the GRANT <privileges> … TO ROLE … ON FUTURE syntax).
```

If the COPY GRANTS option is specified in the CREATE *<object>* statement, then the new object inherits any explicit access privileges granted on the original table but does ***not*** inherit any future grants defined for the object type in the schema.

Question 57: Skipped

Scaling down a virtual warehouse(e.g from a large warehouse to a small one) is an automated process.

- TRUE
- FALSE
  (Correct)

Explanation
To use a different warehouse, you will need to use the below command

```
USE WAREHOUSE <WAREHOUSE_NAME>
```

Question 58: Skipped

When working with a cloned table, you can use the below SQL statements

- SELECT
- DROP
- SHOW
- ALL OF THE ABOVE
(Correct)

Explanation
Clone table is just another table, only thing is that it shares micro partitions with the table it has been cloned from.

Question 59: Skipped

Select the options that differentiates a Partner Connect partner from a regular partner

- Connect with snowflake through a wizard
(Correct)
- Includes a partner trial account signup
(Correct)
- Can be connected from the WEB UI
(Correct)
- Includes automated role, user and staging database setup
(Correct)
- None of the above

Explanation
Partner Connect lets you easily create trial accounts with selected Snowflake business partners and integrate these accounts with Snowflake. This feature provides a convenient option for trying additional tools and services, and then adopting the ones that best meet your business needs.

https://docs.snowflake.com/en/user-guide/ecosystem-partner-connect.html#snowflake-partner-connect

Question 60: Skipped

To improve the query performance, which of the below techniques can be used in snowflake

- Indexes

- Distribution keys
- Query hints
- Cluster keys/Reclustering
  (Correct)

Question 61: Skipped

Which of the following best describes Snowflake's processing engine

- EMR(Elastic Map Reduce)
- Spark Engine
- Presto
- Native SQL Database engine
  (Correct)

Explanation
Snowflake’s data warehouse is not built on an existing database or “big data” software platform such as Hadoop. The Snowflake data warehouse uses a new SQL database engine with a unique architecture designed for the cloud. To the user, Snowflake has many similarities to other enterprise data warehouses, but also has additional functionality and unique capabilities

https://docs.snowflake.com/en/user-guide/intro-key-concepts.html#key-concepts-architecture

Question 62: Skipped

Which of the below are automatically provided by snowflake compared to other databases?

- Installation and Hardware Configurations
  (Correct)
- Patch releases
  (Correct)
- Physical Security
  (Correct)
- Metadata and Collection statistics documentation

Question 63: Skipped

Client has ODBC or JDBC available in their system but they do not have snowflake drivers. Client will still be able to connect to snowflake?

- TRUE
- FALSE
  (Correct)

Explanation
You will need the Snowflake ODBC or JDBC driver to connect to Snowflake, just having JDBC and ODBC will not solve the problem

JDBC

https://docs.snowflake.com/en/user-guide/jdbc.html#jdbc-driver

Snowflake provides a JDBC type 4 driver that supports core JDBC functionality. The JDBC driver must be installed in a 64-bit environment and requires Java 1.8 (or higher).

ODBC

https://docs.snowflake.com/en/user-guide/odbc.html#odbc-driver

Snowflake provides a driver for connecting to Snowflake using ODBC-based client applications.

The ODBC driver has different prerequisites depending on the platform where it is installed. For details, see the individual installation and configuration instructions for each platform.

In addition, different versions of the ODBC driver support the GET and PUT commands, depending on the cloud service that hosts your Snowflake account:

Amazon Web Services: Version 2.17.5 (and higher)

Google Cloud Platform: Version 2.21.5 (and higher)

Microsoft Azure: Version 2.20.2 (and higher)

Question 64: Skipped

Auto clustering can be switched off at an account level

- FALSE
  (Correct)

- TRUE

Explanation
Auto clustering cannot be switched off at database of account level, it will need to be done at the table level.

Question 65: Skipped

A user can be defaulted to a role which user does not have access to

- TRUE
- FALSE
  (Correct)

Explanation
You will be able to create a user with a default role to which the user does not have access. However, the user will nit be able to logon to snowflake if he/she does not have access to the default role. Hence a user cannot be defaulted to a role which he/she does not have access to

Question 66: Skipped

With respect to Snowflake UI, which of the following is true?

- A single session can be shared between multiple worksheets
- Every worksheet can have a different role, warehouse and a database
  (Correct)
- Worksheets cannot have different role, warehouse and database
- Every worksheet has its own session
  (Correct)

Question 67: Skipped

A network policy includes values in both allowed and blocked IP address lists, snowflake applies the blocked IP address list first.

- TRUE
  (Correct)
- FALSE

Explanation
When a network policy includes values in both the allowed and blocked IP address lists, Snowflake applies the ***blocked*** IP address list first.

Do ***not*** add `0.0.0.0/0` to the blocked IP address list. `0.0.0.0/0` is interpreted to be “all IPv4 addresses on the local machine”. Because Snowflake resolves this list first, this would block your own access. Also, note that it is not necessary to include this IP address in the allowed IP address list.

https://docs.snowflake.com/en/user-guide/network-policies.html#managing-account-level-network-policies

Question 68: Skipped

How are virtual warehouse credits charged?

- per minute
- per second
- per-second, with a 60-second (i.e. 1-minute) minimum:
  (Correct)
- per hour

Question 69: Skipped

Tri-Secret Secure option is available in which snowflake edition

- Business critical or higher
  (Correct)
- Enterprise Edition
- All editions

Question 70: Skipped

A warehouse can be assigned to a single resource monitor only

- TRUE
  (Correct)
- FALSE

Explanation
Assignment of Resource Monitors

A single monitor can be set at the account level to control credit usage for all warehouses in your account.

In addition, a monitor can be assigned to one or more warehouses, thereby controlling the credit usage for each assigned warehouse. Note, however, that a warehouse can be assigned to only a single resource monitor.

https://docs.snowflake.com/en/user-guide/resource-monitors.html#assignment-of-resource-monitors

Question 1: Skipped

With respect to snowflake security, Tri-secret requires that customers manage their keys

- TRUE
  (Correct)
- FALSE

Explanation
Tri-Secret Secure lets you control access to your data using a master encryption key that you maintain in the key management service for the cloud provider that hosts your Snowflake account:

AWS: AWS Key Management Service (KMS)

Google Cloud: Cloud Key Management Service (Cloud KMS)

Microsoft Azure: Azure Key Vault

Please note that to use tri-secret, you will need business critical edition or more

https://docs.snowflake.com/en/user-guide/security-encryption.html#tri-secret-secure-and-customer-managed-keys

Question 2: Skipped

The security information of snowflake are store in which layer

- Storage
- Compute
- Service
  (Correct)

Explanation

Remember, what the cloud services layer is for? Authentication and access control.

The cloud services layer is a collection of services that coordinate activities across Snowflake. These services tie together all of the different components of Snowflake in order to process user requests, from login to query dispatch. The cloud services layer also runs on compute instances provisioned by Snowflake from the cloud provider.

Among the services in this layer:

Authentication

Infrastructure management

Metadata management

Query parsing and optimization

Access control

https://docs.snowflake.com/en/user-guide/intro-key-concepts.html#cloud-services

Question 3: Skipped

Snowflake supports multi-factor authentication (MFA) to provide increased login security for users connecting to Snowflake. Which statements are true about MFA security?

- ☐ MFA is not enabled by default. it is available to all, but user need to activate it **(Correct)**
- ☐ MFA is an integrated feature powered by DUO Security Service **(Correct)**
- ☐ MFA can be used to connect to SNOWFLAKE using Snowflake JDBC Driver **(Correct)**
- ☐ MFA Login is designed to connect to snowflake only through the web interface

Explanation

Snowflake supports multi-factor authentication (MFA) to provide increased login security for users connecting to Snowflake. MFA support is provided as an integrated Snowflake feature, powered by the Duo Securityservice, which is managed completely by Snowflake. Users do not need to separately sign up with Duo or perform any tasks, other than installing the Duo Mobile application, which is supported on multiple smart phone platforms (iOS, Android, Windows, etc.). See the Duo User Guidefor more

information about supported platforms/devices and how Duo multi-factor authentication works.

MFA is enabled on a per-user basis; however, at this time, users are not automatically enrolled in MFA. To use MFA, users must enroll themselves.

Previously, users could only be enrolled in MFA by submitting a request to Snowflake Support. This is no longer required.

https://docs.snowflake.com/en/user-guide/security-mfa.html#multi-factor-authentication-mfa

Question 4: Skipped

Which of the below activities can be done leveraging time-travel feature of snowflake?

- ☐ Restoring data-related objects (tables, schemas, and databases) that may have been accidentally or intentionally deleted
(Correct)
- ☐ Duplicating and backing up data from key points in the past
(Correct)
- ☐ Analyzing data usage/manipulation over specified periods of time
(Correct)
- ☐ To backup data for compliance and regulatory requirements

Explanation
Time travel is a beautiful concept in snowflake and you must understand this feature in details. A key component of Snowflake Time Travel is the data retention period. The standard retention period is 1 day (24 hours) and is automatically enabled for all Snowflake accounts

For **Snowflake Standard Edition**, the **retention period can be set to 0** (or unset back to the default of 1 day) at the account and object level (i.e. databases, schemas, and tables).

For Snowflake Enterprise Edition (and higher):

1. For **transient databases, schemas, and tables, the retention period can be set to 0** (or unset back to the default of 1 day). The same is also true for temporary tables.

2. For permanent databases, schemas, and tables, the retention period **can be set to any value from 0 up to 90 days.**

You can use this feature to restore objects if you have accidentally deleted one, you can duplicate or back up data from a certain point in the past. You can also analyze data usage over a period of time

https://docs.snowflake.com/en/user-guide/data-time-travel.html#understanding-using-time-travel

Question 5: Skipped

One of your users have accidentally dropped a table T1 in production. What can you do to restore the table?

- Execute `UNDROP TABLE T1;`
  (Correct)
- Execute `UNDELETE TABLE T1;`
- Call Snowflake team and request them to recover the data from cloud storage
- Nothing can be done, whatever is lost is lost

Explanation
UNDROP is a very useful feature in Snowflake. It makes it easier to retrieve your data and that is because snowflake stores data and objects in encrypted form for 24 hours by default. You have an option to purchase even longer retention period, if required.

Any lost data can be instantaneously recovered with the UNDROP features. It is a single short command

UNDROP TABLE <tablename>

In addition to this, you can also do **UNDROP SCHEMA** and **UNDROP DATABASE** if someone messes it up even bigger.

https://docs.snowflake.com/en/sql-reference/sql/undrop-table.html#undrop-table

Question 6: Skipped

In your organization, you use a 2XL warehouse to load data into Snowflake as part of your ETL workload. Your BI team reached out to you and wanted access to the data stored in Snowflake. You gave them access to the tables required for them to query the data. What will be your recommendation with respect to the warehouse that they will need to use?

- Since you already have a 2XL warehouse for loading data, you will let them use that warehouse
- Warehouse is not required for querying data in snowflake
- It is recommended to use a multi-warehouse approach for workload isolation in snowflake. Since you would not like the load workload to interfere with the read workload of the BI team, you will create a separate warehouse for them **(Correct)**

Explanation
Snowflake's unique multi-cluster, shared data architecture makes it possible to allocate multiple independent, isolated clusters for processing while sharing the same data. In snowflake, each cluster can both read and write data complying to transactional consistency which is ensured by the cloud services layer. The compute clusters in Snowflake are called virtual warehouses. The size and resources of each virtual warehouse can be chosen independently by the user based on the characteristics and performance requirements for the workload(s) that will run on each virtual warehouse. It is, therefore, recommended to isolate workloads and have each workload use its own warehouse. It helps to track usage better and also through this means you will ensure that one workload does not interfere with the other workload. Hence you will create a separate workload for the BI TEAM

Question 7: Skipped

When you clone a table in snowflake, the storage requirements are doubled as it needs to now store two copies of data.

- TRUE
- FALSE
  (Correct)

Explanation
A massive benefit of zero-copy cloning of snowflake is that the underlying data is never duplicated. Only the metadata/pointers of the micropartitons are replicated for the new cloned table. Hence, when you clone a table actual data is never copied over to the new cloned table.

Question 8: Skipped

To ingest semi structured data in snowflake, which data type will you use?

- NUMBER
- VARCHAR
- VARIANT
  (Correct)

Explanation
VARIANT data type allows Snowflake to ingest semi-structured data without having to pre-define the schema.

https://docs.snowflake.com/en/sql-reference/data-types-semistructured.html#variant

Question 9: Skipped

Select the ones that are true about data sharing through data shares

- There is only one copy of data, which lives in the data provider's account
  (Correct)
- Shared data is always live, real-time and immediately available to consumers
  (Correct)
- Providers can establish revocable, fine-grained access grants to shares
  (Correct)
- Data Sharing is supported between accounts in the same or different Snowflake Provider and Region

Explanation

Data sharing is currently supported between accounts in the same Snowflake Provider and Region. If the accounts are in different region, please consider data replication

Question 10: Skipped

Which factors influence the unit cost of snowflake credits and data storage?

- [ ] Snowflake Edition
(Correct)
- [ ] The region where your Snowflake account is located
(Correct)
- [ ] Whether it is an *On Demand* or *Capacity* account
(Correct)
- [ ] Number of users of snowflake

Explanation

The Snowflake Edition that your organization chooses determines the unit costs for the credits and the data storage you use. Other factors that impact unit costs are the region where your Snowflake account is located and whether it is an *On Demand* or *Capacity* account:

1. On Demand: Usage-based pricing with no long-term licensing requirements.

2. Capacity: Discounted pricing based on an up-front Capacity commitment.

Question 11: Skipped

What was Business Critical Edition formerly known as

- Standard Edition
- Enterprise Edition
- Business Edition
- Enterprise for Sensitive Data (ESD)
(Correct)

Explanation

Business Critical Edition, **formerly known as Enterprise for Sensitive Data (ESD)**, offers even higher levels of data protection to support the needs of organizations with extremely sensitive data, particularly PHI data that must comply with HIPAA regulations.

It includes all the features and services of Enterprise Edition, with the addition of enhanced security and data protection. In addition, database failover/failback adds support for business continuity and disaster recovery.

Question 12: Skipped

Before any PHI data can be stored in Snowflake, a signed business associate agreement (BAA) must be in place between your agency/organization and Snowflake Inc

- TRUE
  (Correct)
- FALSE

Explanation
As required by HIPAA regulations, before any PHI data can be stored in Snowflake, a signed business associate agreement (BAA) must be in place between your agency/organization and Snowflake Inc.

Question 13: Skipped

Your organization requires a snowflake instance that is a completely separate Snowflake environment, isolated from all other Snowflake accounts. Which edition will you suggest?

- Standard
- Premier
- Virtual Private Snowflake (VPS)
  (Correct)

Explanation
Virtual Private Snowflake offers our highest level of security for organizations that have the strictest requirements, such as financial institutions and any other large enterprises that collect, analyze, and share highly sensitive data.

It includes all the features and services of Business Critical Edition, but in a completely separate Snowflake environment, isolated from all other Snowflake accounts (i.e. VPS accounts do not share any resources with accounts outside the VPS).

https://docs.snowflake.com/en/user-guide/intro-editions.html#virtual-private-snowflake-vps

Question 14: Skipped

Extended Time Travel (up to 90 days) is available in which editions

- Standard
- Enterprise
  (Correct)
- Business Critical
  (Correct)
- VPS
  (Correct)

Explanation
Please go through this link. there may be questions on feature availability

https://docs.snowflake.com/en/user-guide/intro-editions.html#security-data-protections

Question 15: Skipped

Your organization has a need for Column-level security by applying masking policies to columns in tables or views. Which editions will be suitable for this requirement?

- Standard
- Enterprise
  (Correct)
- Business Critical
  (Correct)
- VPS
  (Correct)

Explanation
Dynamic masking is currently a preview feature and is available to all accounts(as of Aug 2020), it requires enterprise edition or above.

Question 16: Skipped

Which of the below features are only available to Business Critical edition and above?

- Customer-managed encryption keys through Tri-Secret Secure.
  (Correct)

- Support for secure, direct proxy to your other virtual networks or on-premises data centers using AWS PrivateLink or Azure Private Link.
  (Correct)
- Support for PHI data (in accordance with HIPAA regulations).
  (Correct)
- Support for PCI DSS.
  (Correct)
- Extended Time Travel (up to 90 days).

Explanation
Extended Time Travel (up to 90 days) is available in Enterprise and above editions

Question 17: Skipped

Which feature is only available in VPS edition?

- Dedicated metadata store and pool of virtual servers (used in virtual warehouses).
  (Correct)
- SOC 2 Type II certification.
- Automatic encryption of all data.

Question 18: Skipped

In load metadata, Snowflake maintains detailed metadata for each table into which data is loaded. What else is stored in this metadata

- Name of each file from which data was loaded
  (Correct)
- File size
  (Correct)
- Number of columns in the file
- ETag for the file
  (Correct)
- Number of rows parsed in the file
  (Correct)
-

Timestamp of the last load for the file
(Correct)

- Information about any errors encountered in the file during loading
(Correct)

Explanation
Snowflake maintains detailed metadata for each table into which data is loaded, including:

1. Name of each file from which data was loaded

2. File size

3. ETag for the file

4. Number of rows parsed in the file

5. Timestamp of the last load for the file

6. Information about any errors encountered in the file during loading

This load metadata expires after 64 days. If the LAST_MODIFIED date for a staged data file is less than or equal to 64 days, the COPY command can determine its load status for a given table and prevent reloading (and data duplication). The LAST_MODIFIED date is the timestamp when the file was initially staged or when it was last modified, whichever is later.

https://docs.snowflake.com/en/user-guide/data-load-considerations-load.html#load-metadata

Question 19: Skipped

After how many days does the load metadata expire

- 30
- 60
- 64
(Correct)

Explanation
This load metadata expires after 64 days. If the LAST_MODIFIED date for a staged data file is less than or equal to 64 days, the COPY command can determine its load status for a given table and prevent reloading (and data duplication). The LAST_MODIFIED

date is the timestamp when the file was initially staged or when it was last modified, whichever is later.

Question 20: Skipped

Which OPTION will you set on the copy if you would like to ignore the load metadata?

- FORCE
  (Correct)
- LOAD_UNCERTAIN_FILES
- IGNORE METADATA

Explanation

To load files whose metadata has expired, set the LOAD_UNCERTAIN_FILES copy option to true. The copy option references load metadata, if available, to avoid data duplication, but also attempts to load files with expired load metadata.

Alternatively, set the FORCE option to load all files, ignoring load metadata if it exists. Note that this option reloads files, potentially duplicating data in a table.

https://docs.snowflake.com/en/user-guide/data-load-considerations-load.html#workarounds

Question 21: Skipped

Standard Time Travel (up to 1 day) for accessing/restoring modified and deleted data is available in all editions

- FALSE
- TRUE
  (Correct)

Question 22: Skipped

Each object has an owner, who can in turn grant access to that object. What is this model called?

- Discretionary Access Control (DAC)
  (Correct)
- Role-based Access Control (RBAC)
- Object ownership model

Explanation

Snowflake's approach to access control combines aspects from both of the following models:

1. Discretionary Access Control (DAC): Each object has an owner, who can in turn grant access to that object.

2. Role-based Access Control (RBAC): Access privileges are assigned to roles, which are in turn assigned to users.

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

Question 23: Skipped

An entity to which access can be granted. Unless allowed by a grant, access will be denied. in snowflake terms, what is this called?

- Securable object
  (Correct)
- Role
- Privilege
- User

Explanation
The key concepts to understanding access control in Snowflake are:

Securable object: An entity to which access can be granted. Unless allowed by a grant, access will be denied.

Role: An entity to which privileges can be granted. Roles are in turn assigned to users. Note that roles can also be assigned to other roles, creating a role hierarchy.

Privilege: A defined level of access to an object. Multiple distinct privileges may be used to control the granularity of access granted.

User: A user identity recognized by Snowflake, whether associated with a person or program.

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

Question 24: Skipped

Which are the system defined roles in snowflake?

- ☐ ACCOUNTADMIN
  (Correct)
- ☐ SECURITYADMIN
  (Correct)
- ☐ USERADMIN
  (Correct)
- ☐ SYSADMIN
  (Correct)
- ☐ POWERUSER

Explanation
Below are all the system defined roles

ACCOUNTADMIN

(aka Account Administrator)

Role that encapsulates the SYSADMIN and SECURITYADMIN system-defined roles. It is the top-level role in the system and should be granted only to a limited/controlled number of users in your account.

SECURITYADMIN

(aka Security Administrator)

Role that can manage any object grant globally, as well as create, monitor, and manage users and roles. More specifically, this role:

Is granted the MANAGE GRANTS security privilege to be able to modify any grant, including revoking it.

Inherits the privileges of the USERADMIN role via the system role hierarchy (e.g. USERADMIN role is granted to SECURITYADMIN).

USERADMIN

(aka User and Role Administrator)

Role that is dedicated to user and role management only. More specifically, this role:

Is granted the CREATE USER and CREATE ROLE security privileges.

Can create and manage users and roles in the account (assuming that ownership of those roles or users has not been transferred to another role).

SYSADMIN

(aka System Administrator)

Role that has privileges to create warehouses and databases (and other objects) in an account.

If, as recommended, you create a role hierarchy that ultimately assigns all custom roles to the SYSADMIN role, this role also has the ability to grant privileges on warehouses, databases, and other objects to other roles.

PUBLIC

Pseudo-role that is automatically granted to every user and every role in your account. The PUBLIC role can own securable objects, just like any other role; however, the objects owned by the role are, by definition, available to every other user and role in your account.

This role is typically used in cases where explicit access control is not needed and all users are viewed as equal with regard to their access rights.

Question 25: Skipped

In snowflake, you want to compare whether two expressions are equal. The expressions can have null values. What is the best way to compare in this case

- Use **EQUAL_NULL**
  (Correct)
- Run a query with **WHERE EXPR1 = EXPR2**
- Run a query with **WHERE EXPR2=EXPR1**

Explanation
This is an important concept to understand in Snowflake. This has caused me pains in production. Even we do not follow this, you will have incorrect data in production.

EQUAL comparison operator (`=`), treats NULLs as unknown values. Like most SQL languages, comparing `NULL = NULL` does not return `TRUE`. In SnowFlake, it returns `NULL`

EQUAL_NULL function is NULL-safe, meaning it treats NULLs as known values for comparing equality.

https://docs.snowflake.com/en/sql-reference/functions/equal_null.html#equal-null

Question 26: Skipped

PUT command in snowflake can upload files from a local folder/directory of a client machines to the below locations

- Named internal stage
(Correct)
- Internal stage for a specified table
(Correct)
- Internal stage for the current user
(Correct)
- External Stage

Explanation
PUT does ***not*** support uploading files to external stages. To upload files to external stages, use the utilities provided by the cloud service.

The following Snowflake clients do not support PUT:

1. Go Snowflake Driver

2. .NET Driver

3. Node.js Driver

The ODBC driver supports PUT with Snowflake accounts hosted on the following platforms:

1. Amazon Web Services (using ODBC Driver Version 2.17.5 and higher).

2. Google Cloud Platform (using ODBC Driver Version 2.21.5 and higher).

3. Microsoft Azure (using ODBC Driver Version 2.20.2 and higher).

Question 27: Skipped

If you attempt to PUT a file but cannot because a file with the same name already exists in the stage, what can you do to load the file?

- [ ] Wait until the existing file's data is loaded, then retry PUT
(Correct)
- [ ] Rename the file that you want to PUT
(Correct)
- [ ] Set `OVERWRITE = TRUE`
(Correct)
- [ ] Set `FORCE COPY = TRUE`

Explanation

`OVERWRITE = TRUE | FALSE`

Specifies whether Snowflake overwrites an existing file with the same name during upload:

`TRUE`: An existing file with the same name is overwritten.

`FALSE`: An existing file with the same name is not overwritten.

If you attempt to PUT a file but cannot because a file with the same name already exists in the stage, you can do any of the following:

Wait until the existing file's data is loaded, then retry PUT.

Rename the file that you want to PUT.

Set `OVERWRITE = TRUE`. Do this only if it is actually safe to overwrite a file with data that might not yet have been loaded into Snowflake.

Question 28: Skipped

OVERWRITE option for PUT is supported in which cloud providers

- [ ] Amazon AWS
(Correct)
- [ ] Microsoft Azure
(Correct)

- Google cloud platform

Explanation
The following platforms support the OVERWRITE option:

Amazon AWS

Microsoft Azure.

GCP, the Google cloud platform, does not support `OVERWRITE=FALSE`. On GCP, the PUT command always uploads the file, even if a file with the same name exists and OVERWRITE is set to false.

Question 29: Skipped

PUT command can be run from Snowflake worksheets

- TRUE
- FALSE
  (Correct)

Explanation

The PUT command cannot be executed from the **Worksheets** page in the Snowflake web interface; instead, use the SnowSQL client to upload data files, or check the documentation for the specific Snowflake client to verify support for this command.

https://docs.snowflake.com/en/sql-reference/sql/put.html#usage-notes

Question 30: Skipped

A COUNT(*) and COUNT(<COLUMN_NAME>) on a table will always give you the same result

- TRUE
- FALSE
  (Correct)

Explanation
A COUNT(*) will return a total count of rows in the table, while COUNT(<column_name>) will return a count of rows with a non-NULL value in that particular column. COUNT(A,B) only counts the rows that have no NULL values in either the A or B column while COUNT(<ALIAS>.*) can be used to count all the rows containing no NULL columns.

Question 31: Skipped

How did you validate the data after unloading it using COPY INTO command?

- Load the data in a relational table and validate the rows
- Load the data in a CSV and validate the rows
- Use validation_mode='RETURN_ROWS'
  (Correct)

Explanation
Validating Data to be Unloaded (from a Query)

Execute COPY in validation mode to return the result of a query and view the data that will be unloaded from the `orderstiny` table if COPY is executed in normal mode:

```
copy into @my_stage

from (select * from orderstiny limit 5)

validation_mode='RETURN_ROWS';



----+--------+----+-----------+------------+----------+-----------------+----+---
-------------------------------------------------------------------------+

 C1 |   C2   | C3 |    C4     |     C5     |    C6    |       C7        | C8 |
C9                                    |

----+--------+----+-----------+------------+----------+-----------------+----+---
-------------------------------------------------------------------------+

 1  | 36901  | O  | 173665.47 | 1996-01-02 | 5-LOW    | Clerk#000000951 | 0  | ns
tructions sleep furiously among                                          |

 2  | 78002  | O  | 46929.18  | 1996-12-01 | 1-URGENT | Clerk#000000880 | 0  |  f
oxes. pending accounts at the pending\, silent asymptot                  |

 3  | 123314 | F  | 193846.25 | 1993-10-14 | 5-LOW    | Clerk#000000955 | 0  | sl
y final accounts boost. carefully regular ideas cajole carefully. depos  |

 4  | 136777 | O  | 32151.78  | 1995-10-11 | 5-LOW    | Clerk#000000124 | 0  | si
ts. slyly regular warthogs cajole. regular\, regular theodolites acro    |

 5  | 44485  | F  | 144659.20 | 1994-07-30 | 5-LOW    | Clerk#000000925 | 0  | qu
ickly. bold deposits sleep slyly. packages use slyly                     |

----+--------+----+-----------+------------+----------+-----------------+----+---
-------------------------------------------------------------------------+
```

https://docs.snowflake.com/en/sql-reference/sql/copy-into-location.html#validating-data-to-be-unloaded-from-a-query

Question 32: Skipped

How do you remove files from a stage?

- REMOVE
  (Correct)
- DELETE
- PURGE
- TRUNCATE

Explanation
REMOVE

Removes files that have been staged (i.e. uploaded from a local file system or unloaded from a table) in one of the following Snowflake internal stages:

Named internal stage.

Stage for a specified table.

Stage for the current user.

Note that using the command to remove files from an external stage might work but is not officially supported.

REMOVE can be abbreviated to RM.

https://docs.snowflake.com/en/sql-reference/sql/remove.html#remove

Question 33: Skipped

To improve performance materialized view can created on top of external tables without any additional cost

- TRUE
- False
  (Correct)

Explanation
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 re-using these results frequently enough.

Question 34: Skipped

Which COPY INTO <table> copy options are not supported by SNOWPIPE?

- [ ] `FILES = ( 'file_name1' [ , 'file_name2', ... ] )`
(Correct)

- [ ] `ON_ERROR = ABORT_STATEMENT`
(Correct)

- [ ] `PURGE = TRUE | FALSE`
(Correct)

- [ ] `FORCE = TRUE | FALSE`
(Correct)

- [ ] `FILE_FORMAT =( { FORMAT_NAME = '[<namespace>.]<file_format_name>' | TYPE = { CSV | JSON | AVRO | ORC | PARQUET | XML } [ formatTypeOptions ] } ) ]`

Explanation
Usage Notes

All COPY INTO <table> copy options are supported ***except for*** the following:

`FILES = ( 'file_name1' [ , 'file_name2', ... ] )`

`ON_ERROR = ABORT_STATEMENT`

`SIZE_LIMIT = num`

`PURGE = TRUE | FALSE` (i.e. automatic purging while loading)

`MATCH_BY_COLUMN_NAME = CASE_SENSITIVE | CASE_INSENSITIVE | NONE`

`FORCE = TRUE | FALSE`

Note that you can manually remove files from an internal (i.e. Snowflake) stage (after they've been loaded) using the REMOVE command.

`RETURN_FAILED_ONLY = TRUE | FALSE`

`VALIDATION_MODE = RETURN_n_ROWS | RETURN_ERRORS | RETURN_ALL_ERRORS`

Please go through this link thoroughly

https://docs.snowflake.com/en/sql-reference/sql/create-pipe.html#usage-notes

Please note that `PATTERN = 'regex_pattern'` copy option is provided as a preview feature . If this is listed as one of the option, it will fall under not supported option.

Question 35: Skipped

A stream itself does ***not*** contain any table data

- TRUE
(Correct)
- FALSE

Explanation

Note that a stream itself does ***not*** contain any table data. A stream only stores the offset for the source table and returns CDC records by leveraging the versioning history for the source table. When the first stream for a table is created, a pair of hidden columns are added to the source table and begin storing change tracking metadata. These columns consume a small amount of storage. The CDC records returned when querying a stream rely on a combination of the *offset* stored in the stream and the *change tracking metadata* stored in the table.

Question 36: Skipped

A DML statement that selects from a stream consumes all of the change data in the stream as long as the transaction commits successfully

- TRUE
(Correct)
- FALSE

Explanation

`A DML statement` that selects from a stream consumes all of the change data in the stream as long as the transaction commits successfully. To ensure multiple statements access the same change records in the stream, surround them with an explicit transaction statement (BEGIN .. COMMIT). This locks the stream. DML updates to the source table in parallel transactions are tracked by the change tracking system but do not update the stream until the explicit transaction statement is committed and the existing change data is consumed.

This is very important to know, if you are given a TASK which consumes from the stream and a set of query statements and then if you are asked which query statement

will consume all the change data then please select the DML query statement(like INSERT query)

Question 37: Skipped

By executing the 'SHOW TABLES' command, we can list all the tables in all the schemas even if we do not have access to all the tables

- TRUE
- FALSE **(Correct)**

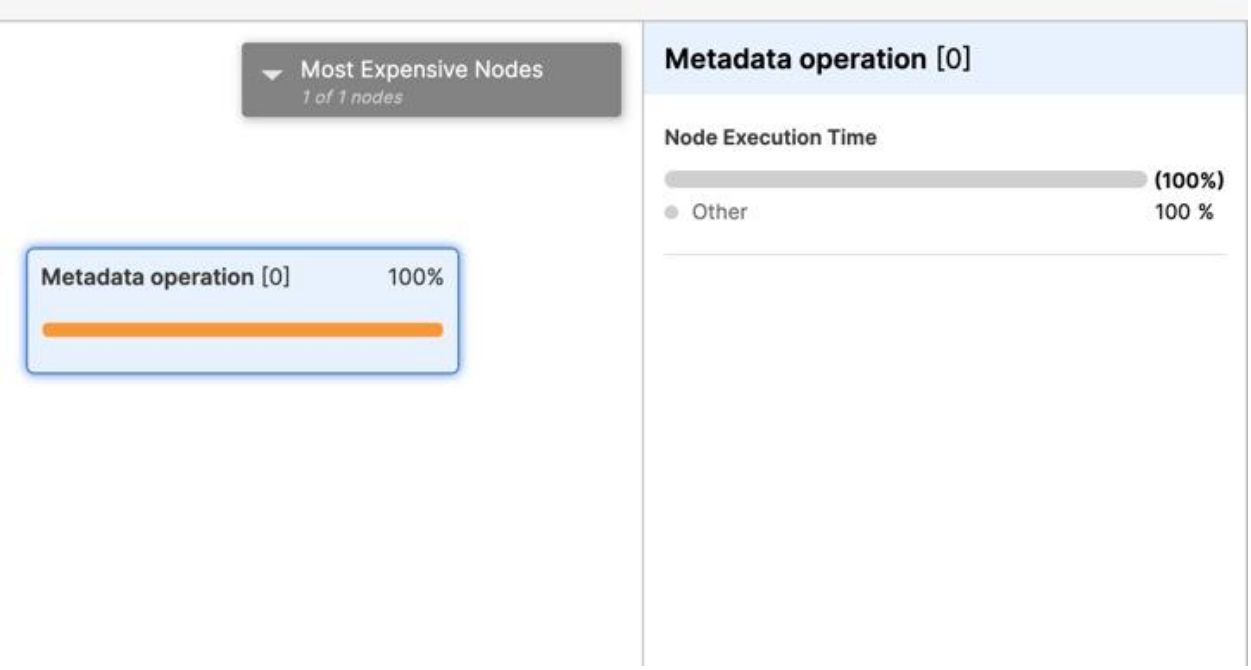
Explanation
SHOW TABLES

Lists the tables for `which you have access privileges`, including dropped tables that are still within the Time Travel retention period and, therefore, can be undropped. The command can be used to list tables for the current/specified database or schema, or across your entire account.

`The output returns table metadata and properties,` ordered lexicographically by database, schema, and table name (see Output in this topic for descriptions of the output columns). This is important to note if you wish to filter the results using the provided filters.

Please also note the comment on `table metadata`

The SHOW COMMAND does not cost you any compute resource as it does not use warehouse. It gets the result from metadata store. Please try it yourself. Run this command and go to query profile and see yourself



Question 38: Skipped

Let's say that you have two JSONs as below

1. {"stuId":2000, "stuName":"Amy"}

2.{"stuId":2000,"stuCourse":"Snowflake"}

How will you write a query that will check if stuId in JSON in #1 is also there in JSON in#2

- ```
  with stu_demography
  as
  (select parse_json(column1) as src, src:stuId as ID
  from values('{"stuId":2000, "stuName":"Amy"}')),
  stu_course
  as
  (select parse_json(column1) as src, src:stuId as ID
  from values('{"stuId":2000,"stuCourse":"Snowflake"}'))
  select
  case
  when stdemo.ID in(select ID from stu_course)
  then 'True'
  else 'False'
  end as result
  from stu_demography stdemo;
  ```

  (Correct)

-
  ```

```
with stu_demography
as
(select parse_json(column1) as src, src['stuId'] as ID
from values('{"stuId":2000, "stuName":"Amy"}')),
stu_course
as
(select parse_json(column1) as src, src['stuId'] as ID
from values('{"stuId":2000,"stuCourse":"Snowflake"}'))
select
case
when stdemo.ID in(select ID from stu_course)
then 'True'
else 'False'
end as result
from stu_demography stdemo;
```

(Correct)

- ```
  SELECT CONTAINS('{"stuId":2000,
  "stuName":"Amy"}','{"stuId":2000,"stuCourse":"Snowflake"}');
  ```
- ```
  with stu_demography
  as
  (select parse_json(column1) as src, src['STUID'] as ID
  from values('{"stuId":2000, "stuName":"Amy"}')),
  stu_course
  as
  (select parse_json(column1) as src, src['stuId'] as ID
  from values('{"stuId":2000,"stuCourse":"Snowflake"}'))
  select
  case
  when stdemo.ID in(select ID from stu_course)
  then 'True'
  else 'False'
  end as result
  from stu_demography stdemo;
  ```

Explanation
I would like you to try this out in your snowflake instance and find that out

Please note that this may not be the way the question will appear in the certification exam, but why we are still learning this?

1. When you take this course, I prepare you not just for the certification but also to be prepared to work on a project

2. With this hands-on, what will you learn

a. That there are two ways to query JSON, dot notation(src:stuId) and bracket notation src['stuId']

b. The JSON Value is case sensitive, if you run the 4th query with src['STUID'], it will return false because src['STUID'] will return you a NULL since there is a mismatch in the case.

In the certification you may be asked question on a), b) - but do you agree if you learn this way you will never forget the concept:)

Question 39: Skipped

Create a task and a stream following the below steps. So, when the system$stream_has_data('rawstream1') condition returns false, what will happen to the task ?

```
1.  -- Create a landing table to store raw JSON data.
2.  -- Snowpipe could load data into this table.
3.  create or replace table raw (var variant);
4.
5.  -- Create a stream to capture inserts to the landing table.
6.  -- A task will consume a set of columns from this stream.
7.  create or replace stream rawstream1 on table raw;
8.
9.  -- Create a second stream to capture inserts to the landing table.
10. -- A second task will consume another set of columns from this stream.
11. create or replace stream rawstream2 on table raw;
12.
13. -- Create a table that stores the names of office visitors identified in the raw data.
14. create or replace table names (id int, first_name string, last_name string);
15.
16. -- Create a table that stores the visitation dates of office visitors identified in the
    raw data.
17. create or replace table visits (id int, dt date);
18.
19. -- Create a task that inserts new name records from the rawstream1 stream into the name
    s table
20. -- every minute when the stream contains records.
21. -- Replace the 'etl_wh' warehouse with a warehouse that your role has USAGE privilege o
    n.
22. create or replace task raw_to_names
23. warehouse = etl_wh
24. schedule = '1 minute'
25. when
26. system$stream_has_data('rawstream1')
27. as
28. merge into names n
29.   using (select var:id id, var:fname fname, var:lname lname from rawstream1) r1 on n.id
    = to_number(r1.id)
30.   when matched then update set n.first_name = r1.fname, n.last_name = r1.lname
31.   when not matched then insert (id, first_name, last_name) values (r1.id, r1.fname, r1.
    lname)
32. ;
33.
34. -- Create another task that merges visitation records from the rawstream1 stream into t
    he visits table
```

```
-- every minute when the stream contains records.
-- Records with new IDs are inserted into the visits table;
-- Records with IDs that exist in the visits table update the DT column in the table.
-- Replace the 'etl_wh' warehouse with a warehouse that your role has USAGE privilege o
n.
create or replace task raw_to_visits
warehouse = etl_wh
schedule = '1 minute'
when
system$stream_has_data('rawstream2')
as
merge into visits v
  using (select var:id id, var:visit_dt visit_dt from rawstream2) r2 on v.id = to_numbe
r(r2.id)
  when matched then update set v.dt = r2.visit_dt
  when not matched then insert (id, dt) values (r2.id, r2.visit_dt)
;

-- Resume both tasks.
alter task raw_to_names resume;
alter task raw_to_visits resume;

-- Insert a set of records into the landing table.
insert into raw
  select parse_json(column1)
  from values
  ('{"id": "123","fname": "Jane","lname": "Smith","visit_dt": "2019-09-17"}'),
  ('{"id": "456","fname": "Peter","lname": "Williams","visit_dt": "2019-09-17"}');

-- Query the change data capture record in the table streams
select * from rawstream1;
select * from rawstream2;
```

- Task will be executed but no rows will be merged
- Task will return an warning message
- Task will be skipped
  (Correct)

Explanation

I would like you to complete the above steps and then run the below query to see it yourself. This way you will never forget this:)

```
select *

  from table(information_schema.task_history())

  order by scheduled_time;
```

Do you see that the task has been skipped?

MA_NAME	QUERY_TEXT	CONDITION_TEXT	STATE	ERROR_CODE	ERROR_MESSAGE	SCHEDULED_TI	QUERY_STAR
IY_SC...	merge into v...	system$stream_has_data('rawstream2')	SUCCEEDED	NULL	NULL	2020-09-16 ...	2020-09-16
IY_SC...	merge into n...	system$stream_has_data('rawstream1')	SKIPPED	0040003	Conditional expression for task evaluated to false.	2020-09-16 ...	NUI
IY_SC...	merge into v...	system$stream_has_data('rawstream2')	SKIPPED	0040003	Conditional expression for task evaluated to false.	2020-09-16 ...	NUI
IY_SC...	merge into n...	system$stream_has_data('rawstream1')	SKIPPED	0040003	Conditional expression for task evaluated to false.	2020-09-16 ...	NUI
V_SCHEMA	merge into v...	system$stream_has_data('rawstream2')	SKIPPED	0040003	Conditional expression for task evaluated to false.	2020-09-16 ...	NUI
IY_SC...	merge into n...	system$stream_has_data('rawstream1')	SKIPPED	0040003	Conditional expression for task evaluated to false.	2020-09-16 ...	NUI
IY_SC...	merge into v...	system$stream_has_data('rawstream2')	SKIPPED	0040003	Conditional expression for task evaluated to false.	2020-09-16 ...	NUI

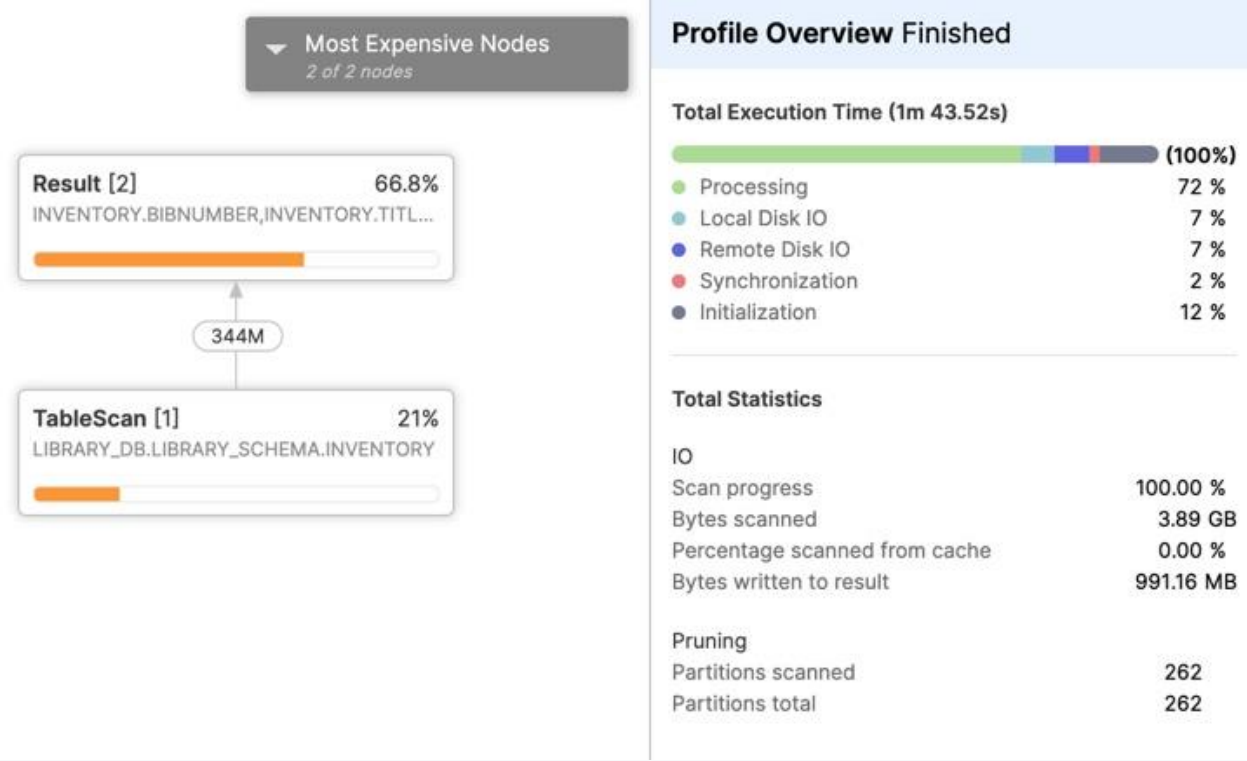
Question 40: Skipped

You ran the below query. I have a warehouse with auto suspend set at 5 seconds. Also result cache is disabled.

```
SELECT * FROM INVENTORY;
```

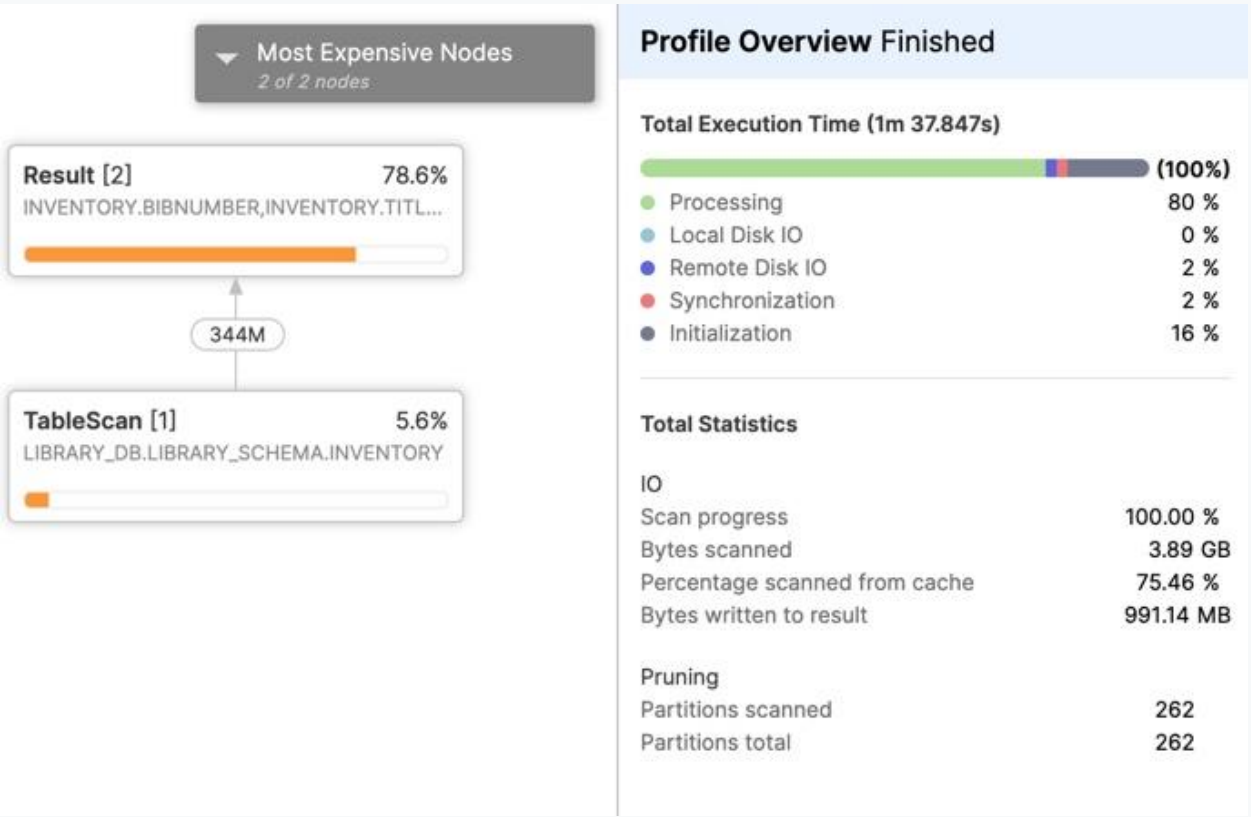
The query profile looks like as below. Please see below 'Percentage scanned from cache' is 0%

Larger image



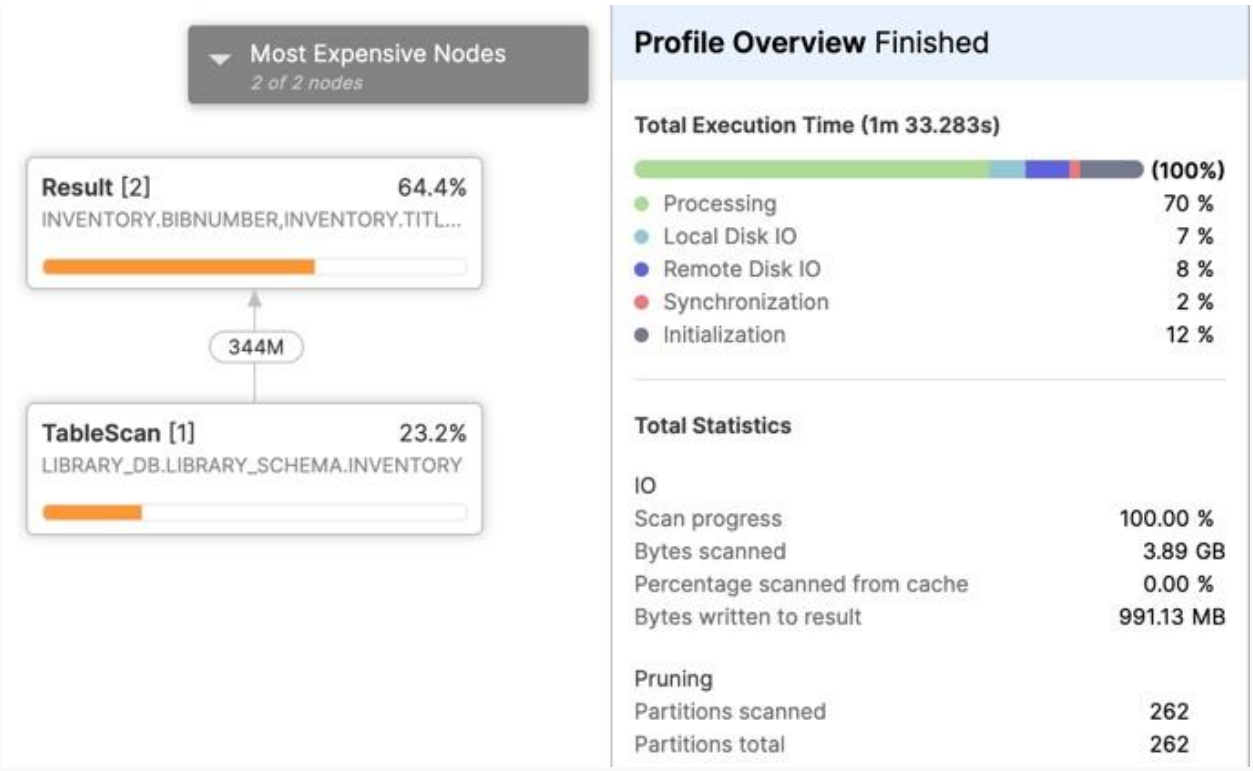
You ran the query again before 5 seconds has elapsed and the query profile looks as below. Look at the 'Percentage scanned for cache', it is 75%

Larger image



You ran the query again after 5 seconds. The query profile looks as below. Look at the 'Percentage scanned from cache', it is zero again.

Larger image



Why is this happening?

- The second run of the query used data cache to retrieve part of the result since it ran before the warehouse was suspended
(Correct)
- The second run of the query used query result cache
- The third run of the query used query result cache

Explanation
This is a very important concept to understand. There may be many different questions on this concept.

Lets, understand what is going on in here

Virtual warehouses are an abstraction on the compute instances of the cloud provider(in case of AWS, it is EC2 instances). Each Virtual warehouse is a cluster of these compute instances(or EC2 in case of AWS). The compute instances has local SSD attached to them. When you ran the query for the first time, the results of the query were retrieved from the remote storage(which is the object store of the cloud provider, S3 in case of AWS), part of the results also got cached in the local SSD storage of the

compute instance. So, when we ran the query second time, part of the results got retrieved from the SSD cache also known as **Data Cache.**

Ok, if that is the case why did not it retrieve from data cache the third time. The third run of the query happened after 5 seconds. The virtual warehouse had a auto suspend setting of 5 seconds. So, since there were no activity for 5 seconds, the warehouse suspended itself. When the warehouse is suspended, it loses the data cache. Why? because when it resumes it may be a completely new set of compute instances which were not used in the earlier runs.

Hope this clarifies and you probably will never forget this concept now. Please remember SSD cache is also called **Data cache**

In real work scenario, how can you leverage the understanding of this concept. Well if you a nightly ETL that runs against several tables, group the ETLs which use similar tables to run at the same time. This will help you to utilize the data cache effectively. This strategy, I named it as ETL-Colocation:)

Question 41: Skipped

Snowflake adds the system defined roles to each new account. Check the system defined roles below.

- ☐ SYSADMIN
  (Correct)
- ☐ ACCOUNTADMIN
  (Correct)
- ☐ SECURITYADMIN
  (Correct)
- ☐ PUBLIC
  (Correct)
- ☐ USERADMIN
  (Correct)
- ☐ SUPERUSER

Explanation
There are a small number of system-defined roles in a Snowflake account. Users with appropriate access can alter the system-defined roles and can also create custom roles.

System-Defined Roles

ACCOUNTADMIN

(aka Account Administrator)

Role that encapsulates the SYSADMIN and SECURITYADMIN system-defined roles. It is the top-level role in the system and should be granted only to a limited/controlled number of users in your account.

SECURITYADMIN

(aka Security Administrator)

Role that can manage any object grant globally, as well as create, monitor, and manage users and roles. More specifically, this role:

Is granted the MANAGE GRANTS security privilege to be able to modify any grant, including revoking it.

Inherits the privileges of the USERADMIN role via the system role hierarchy (e.g. USERADMIN role is granted to SECURITYADMIN).

USERADMIN

(aka User and Role Administrator)

Role that is dedicated to user and role management only. More specifically, this role:

Is granted the CREATE USER and CREATE ROLE security privileges.

Can create and manage users and roles in the account (assuming that ownership of those roles or users has not been transferred to another role).

SYSADMIN

(aka System Administrator)

Role that has privileges to create warehouses and databases (and other objects) in an account.

If, as recommended, you create a role hierarchy that ultimately assigns all custom roles to the SYSADMIN role, this role also has the ability to grant privileges on warehouses, databases, and other objects to other roles.

PUBLIC

Pseudo-role that is automatically granted to every user and every role in your account. The PUBLIC role can own securable objects, just like any other role; however, the objects owned by the role are, by definition, available to every other user and role in your account.

This role is typically used in cases where explicit access control is not needed and all users are viewed as equal with regard to their access rights.

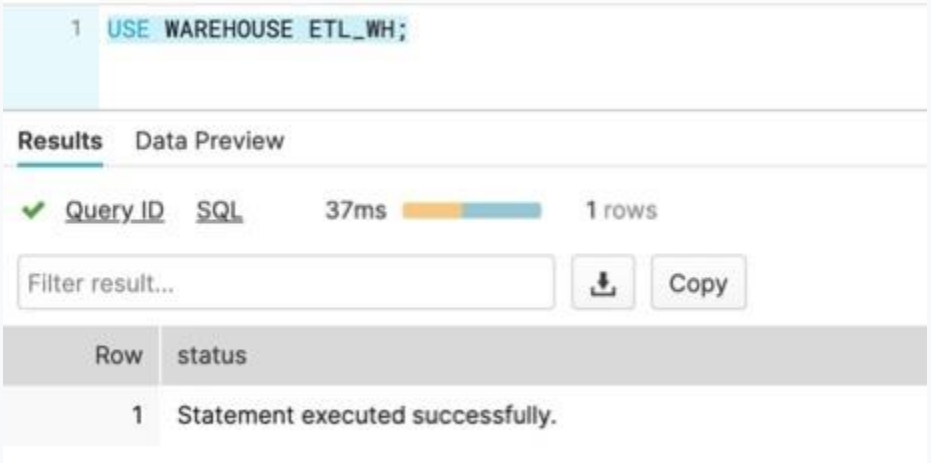
Question 42: Skipped

You are running your query in the worksheet using a large warehouse. You want to change the warehouse. How do you do that? Choose 2 answers

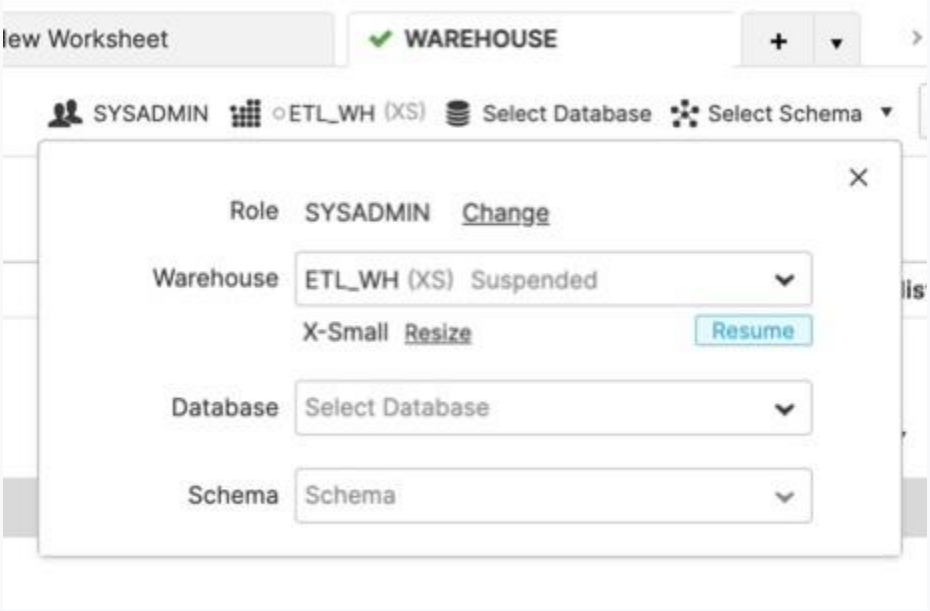
- ☐ USE WAREHOUSE ETL_WH;
  (Correct)
- ☐ SET WAREHOUSE CONTEXT=ETL_WH
- ☐ Update the Warehouse field in the Context Menu located above the worksheet
  (Correct)
- ☐ Go to the Worksheet properties page and set the Warehouse field to ETL_WH

Explanation
Let us not mug this up. Logon to your snowflake account and open a worksheet

Option#1



Option#2



Question 43: Skipped

The COPY command is more performant than the INSERT statement?

- TRUE
(Correct)
- FALSE

Explanation
Let us understand why the answer is TRUE here.

Do you remember that Snowflake stores data physically into immutable micro-partitions? So, how does insert/update work in snowflake?

Since the micropartitons are immutable, any insert or update will have to copy the entire micro-partition into a new copy and then insert the new record in that partition. For update the old record will be marked for deletion and the new record will be added. so, imagine if you do insert one by one, how many micropartitons will get created.

Snowflake's architecture is **optimized for bulk load**. Otherwise, small DMLs (and at high frequency) cause issue of triggering the engine to create a lot of small partition files (this is because each DML create a new (micro)partition file; and a lot of small partition files are bad for data processing and the database engine has to regularly compact the small partition files into larger ones for more optimal query performance).

Hope this clarifies the concept.

Question 44: Skipped

To help control the usage of credits in Auto-scale mode, Snowflake provides a property, _____, that determines the scaling policy to use when automatically starting or shutting down additional clusters. Select the property

- Auto Scale
- Scaling Policy
  (Correct)
- Max no of cluster
- Min no of cluster

Explanation
Setting the Scaling Policy for a Multi-cluster Warehouse

To help control the credits consumed by a multi-cluster warehouse running in Auto-scale mode, Snowflake provides scaling policies, which are used to determine when to start or shut down a cluster.

The scaling policy for a multi-cluster warehouse only applies if it is running in Auto-scale mode. In Maximized mode, all clusters run concurrently so there is no need to start or shut down individual clusters.

https://docs.snowflake.com/en/user-guide/warehouses-multicluster.html#setting-the-scaling-policy-for-a-multi-cluster-warehouse

Question 45: Skipped

Which three objects did we explicitly refer to using the COPY INTO command to load data on using external stages?

- Table
  (Correct)
- File Format
  (Correct)
- View
- Stage
  (Correct)

Explanation
Try this out and you will never forget

1. create a table as below

```
CREATE OR REPLACE TABLE EMPLOYEE(FIRST_NAME VARCHAR, LAST_NAME VARCHAR,EMPLOYEE_ID NUMBER, SALARY NUMBER,JOINING_DATE TIMESTAMP);
```

2. Create a file format and then a stage

```
CREATE OR REPLACE file format load_csv_format
  TYPE = 'CSV'
  FIELD_DELIMITER = ','
  FIELD_OPTIONALLY_ENCLOSED_BY = '"'
  ;
 CREATE OR REPLACE stage employee_stage
  file_format = load_csv_format
;
```

3. Use SNOWSQL to load a file to a stage using PUT

```
put file:///Users/Udemy/Documents/EMPLOYEE.csv @employee_stage; --replace with location of your csv


Populate the CSV with at least 1 or 2 rows(First Name, Last Name, EMPLOYEE ID, SALARY, JOINING DATE
```

4. Run the below command. I had to do SKIP_HEADER because my CSV had the column headers also.

```
COPY INTO EMPLOYEE
FROM @EMPLOYEE_STAGE
FILE_FORMAT=(TYPE=CSV SKIP_HEADER=1);
```

Question 46: Skipped

Mark all the true statements

- [ ] Snowflake Stages can be defined as either External or Internal.
(Correct)

- [ ] Snowflake Stages can be defined as either FTP or SFTP.

- External Stages require a cloud storage provider.
  (Correct)
- FTP Stages require three cloud storage providers.
- Cloud storage providers must be linked to internal stage objects.

Question 47: Skipped

What is the column datatype used for ingesting semi-structured data?

- VARCHAR
- BINARY
- VARIANT
  (Correct)

Explanation
Try this out

```
create or replace table mytable
(
  src variant -- this is the datatype
);


insert into mytable
  select parse_json(column1)
  from values
  ('{
  "a": "1",
  "b": "2",
  "c": null
  }')
  , ('{
  "a": "1",
  "b": "2",
  "c": "3"
  }');
```

A bonus one:) Try this out as well

```
select strip_null_value(src:c) from mytable;
```

strip_null_value converted json null to SQL NULL. Remember this.

Question 48: Skipped

JSON and PARQUET files can be loaded to columns in the same table

- TRUE
- FALSE
  (Correct)

Explanation
Let us try to answer some logical questions to arrive at the answer for this question

1. How do you specify the file format when you try to load a file to a snowflake table

```
[ FILE_FORMAT = ( { FORMAT_NAME = '[<namespace>.]<file_format_name>' |
                    TYPE = { CSV | JSON | AVRO | ORC | PARQUET | XML } [ formatTy
peOptions ] } ) ]
```

2. Can you specify more than one file format while loading a table?

No

And there is your answer:)

Question 49: Skipped

```
create or replace table result_scan_table_1 like result_scan_table;
```

Will the above query cost compute credits?

- No, since it is a metadata operation only
  (Correct)
- Yes, it will need compute as the table structure is also getting created
- It is hard to say without looking at query profile

Explanation
Why do you think it will not cost any compute credit? The reason behind this is that the `LIKE clause` only creates the definition of the table(or the structure of the table) in the metadata repository. It does not require any compute credit for any type of metadata operations.

How can you prove that it did not cost you any compute. Look at the history tab after you run the query.

Status	Query ID	SQL Text	U...	Warehouse	Clust...	Size	Session ID	Start Time	End Time	Total Duration	Bytes Scanned
✔	0196fa14-...	create or replace tabl...	J...	ETL_WH			94841851937	4:16:45 PM	4:16:45 PM	146ms	

The size column is empty. If it would have cost you compute, the size column would be populated with the warehouse size.

Now please try this out. Run the below query

```
SHOW TABLES LIKE 'result_scan%';
```

Will this cost you compute? (HINT - It is a metadata operation)

If you are not able to answer this, please ping me on UDEMY.

Question 50: Skipped

To avoid unexpected task executions due to daylight saving time, you will do one of the below activities.

- ☐ Do not schedule tasks to run at a specific time between 1 AM and 3 AM (daily, or on days of the week that include Sundays)
(Correct)

- ☐ Manually adjust the cron expression for tasks scheduled during those hours twice each year to compensate for the time change due to daylight saving time
(Correct)

- ☐ Add the USE DAYLIGHT SAVINGS parameter for task

Explanation
The cron expression in a task definition supports specifying a time zone. A scheduled task runs according to the specified cron expression in the local time for a given time zone. Special care should be taken with regard to scheduling tasks for time zones that recognize daylight saving time. Tasks scheduled during specific times on days when the transition from standard time to daylight saving time (or the reverse) occurs can have unexpected behaviors.

For example:

During the autumn change from daylight saving time to standard time, a task scheduled to start at 1 AM in the America/Los_Angeles time zone (i.e. `0 1 * * *`

`America/Los_Angeles`) would run ***twice***: once at 1 AM and then again when 1:59:59 AM shifts to 1:00:00 AM local time. That is, there are two points in time when the local time is 1 AM.

During the spring change from standard time to daylight saving time, a task scheduled to start at 2 AM in the America/Los_Angeles time zone (i.e. `0 2 * * *` `America/Los_Angeles`) would ***not run at all*** because the local time shifts from 1:59:59 AM to 3:00:00 AM. That is, there is no point during that day when the local time is 2 AM.

To avoid unexpected task executions due to daylight saving time, ***either***:

```
1. Do not schedule tasks to run at a specific time between 1 AM and 3 AM (daily,
or on days of the week that include Sundays), or

2. Manually adjust the cron expression for tasks scheduled during those hours twi
ce each year to compensate for the time change due to daylight saving time.
```

Question 51: Skipped

The query profiler view is only available for completed queries

- TRUE
- FALSE
  (Correct)

Explanation
The query profiler is available for query in progress also.

Question 52: Skipped

You have created a secure view view1. Now you want to convert it back to a regular view. Which of the below can be used for the conversion?

- ```
  1.  alter view view1 set type = unsecure;
  ```
- ```
  1.  alter view view1 unset secure;
  ```

  (Correct)

- ```
  1.  alter view view1 set secure;
  ```
- None of the above

Explanation
Revert a secure view to a regular view:

```
alter view view1 unset secure;
```

Please look at the other examples here
```

https://docs.snowflake.com/en/sql-reference/sql/alter-view.html#examples

Question 53: Skipped

You want to load a JSON file but would like to remove the outer array structure and load the records into separate table rows. Which file format option would you use?

- STRIP_OUTER_ARRAY
  (Correct)
- TRUNCATE_OUTER_ARRAY
- UNSET_OUTER_ARRAY
- DROP_OUTER_ARRAY

Explanation
In general, JSON and Avro data sets are a simple concatenation of multiple documents. The JSON or Avro output from some software is composed of a single huge array containing multiple records. There is no need to separate the documents with line breaks or commas, though both are supported.

Instead, snowflake recommends enabling the STRIP_OUTER_ARRAY file format option for the COPY INTO <table> command to remove the outer array structure and load the records into separate table rows:

```
copy into <table>
from @~/<file>.json
file_format = (type = 'JSON' strip_outer_array = true);
```

Question 54: Skipped

Which two options mentioned below will you choose to concat a column named 'FRUIT_DESCRIPTION' between two % signs?

- '%'||FRUIT_DESCRIPTION||'%'
  (Correct)
- CONCAT('%',CONCAT(FRUIT_DESCRIPTION,'%'))
  (Correct)
- %+FRUIT_DESCRIPTION+%
- SUBSTR(%,FRUIT_DESCRIPTION,%)

Explanation

Try it yourself

```
CREATE OR REPLACE TABLE FRUITS(FRUIT_NUMBER NUMBER, FRUIT_DESCRIPTION VARCHAR, AV
AILABILITY VARCHAR);


INSERT INTO FRUITS VALUES(1,'APPLE',NULL);

INSERT INTO FRUITS VALUES(2,'MANGO',NULL);

INSERT INTO FRUITS VALUES(3,'ORANGE',NULL);

INSERT INTO FRUITS VALUES(4,'KIWI',NULL);

INSERT INTO FRUITS VALUES(5,'MELON',NULL);

INSERT INTO FRUITS VALUES(6,'PEARS',NULL);

INSERT INTO FRUITS VALUES(7,'BANANA',NULL);

INSERT INTO FRUITS VALUES(8,'PLUM',NULL);

INSERT INTO FRUITS VALUES(9,'APRICOT',NULL);

INSERT INTO FRUITS VALUES(10,'WATERMELON',NULL);


SELECT '%'||FRUIT_DESCRIPTION||'%' FROM FRUITS;

SELECT CONCAT('%',CONCAT(FRUIT_DESCRIPTION,'%')) FROM FRUITS;
```

Question 55: Skipped

Snowflake has three types of stages USER, TABLE, NAMED. Named stage is again divided into External and Internal. Which of the below two stages are automatically created and does not need explicit configuration by the user

- ☐ TABLE
  (Correct)
- ☐ EXTERNAL
- ☐ INTERNAL
- ☐ USER
  (Correct)

Question 56: Skipped

When you create internal stage, you must link it to the cloud storage providers

- ○ FALSE
  (Correct)

- TRUE

Explanation
That is the beauty of internal stage. Snowflake manages it for you. You do not have to do anything with the underlying cloud provider. Snowflake also automatically encrypts the data for you in internal stage.

Question 57: Skipped

Data providers must sign an agreement before their data is made public on the data marketplace of snowflake

- TRUE
(Correct)

- FALSE

Explanation
Before you can list a data set on the Snowflake Data Marketplace, you must become an approved data provider.

The approval is a one-time process that ensures that:

1. Provider is a Snowflake customer in the supported regions.

2. Provider has signed the Snowflake Provider agreement.

3. Provider's data meets Snowflake's data requirements. To meet the requirements, shared data must be:

a. Fresh — the data is near real-time or updated on a regular basis.

b. Real — not sample data.

c. Legally shareable — the provider must own the data or have the right to share it.

Question 58: Skipped

Snowflake supports the below security features

- TRI-SECRET SECURE ENCRYPTION
(Correct)

- ROLE-BASED ACCESS
(Correct)

- MULTI-FACTOR AUTHENTICATION

(Correct)

- NONE OF THE ABOVE

Question 59: Skipped

Snowflake supports the following performance optimization techniques

- SSD CACHING/DATA CACHING
(Correct)
- QUERY RESULT CACHING
(Correct)
- B-TREE INDEXES

Explanation
Please read the below article

https://community.snowflake.com/s/article/Caching-in-Snowflake-Data-Warehouse

Question 60: Skipped

Select two options which are true about variant data type in snowflake?

- Supports querying using JSON path notation
(Correct)
- Optimizes storage based on repeated elements
(Correct)
- A separate file format is used to store variant data
- Requires custom mapping for each type of record

Question 1: Skipped

According to snowflake architecture, data storage is independent of compute

- TRUE
(Correct)
-

FALSE

Explanation

Snowflake's architecture has been defined with the belief that a tight coupling of compute and storage will not be an effective approach of leveraging the scale of cloud. The idea is to be able scale each of this component independent of each other. Snowflake's **multi-cluster, shared data** architecture separates compute resource scaling from storage resources, thus enabling seamless, non-disruptive scaling.

For data storage, Snowflake leverages the power of the respective cloud providers object store(AWS S3, Azure BLOB, GCS) and for compute it leverages the cloud VM instances which is called '**VIRTUAL WAREHOUSE**' in snowflake. Virtual warehouses are nothing but a set of compute clusters abstracted using the underlying cloud VM machines.

Question 2: Skipped

All virtual warehouses in snowflake has access to all data

- TRUE
  (Correct)
- FALSE

Explanation

Snowflake was built with a unique new architecture that makes it possible to solve this challenge. Snowflake's unique multi-cluster, shared data architecture makes it possible to allocate multiple independent, isolated clusters for processing while sharing the same data. Each cluster (we call them "virtual warehouses") can both read and write data, with full transactional consistency ensured by Snowflake's cloud services layer. The size and resources of each virtual warehouse can be chosen independently by the user based on the characteristics and performance requirements for the workload(s) that will run on each virtual warehouse.

https://www.snowflake.com/blog/snowflake-challenge-concurrent-load-and-query/

Question 3: Skipped

Which layer in the snowflake architecture contains virtual warehouse

- Cloud services
- Query processing
  (Correct)
- Database Storage
-

None

Explanation

Virtual warehouses are used to provide compute resources for processing queries. Hence it falls under query processing

Question 4: Skipped

Which layer in snowflake contains data in compressed, columnar format

- Cloud services
- Query processing
- Database storage
  (Correct)
- None

Explanation

Snowflake stores data in compressed and encrypted format by default in the database storage layer. Data is store in micropartitions in a columnar format.

Question 5: Skipped

Which of the below approach would result in a better performance through linear scaling of data ingestion workload

- Resize virtual warehouse
- Consider the practice of organizing data by granular path
- Consider the practice of splitting input file batch within the recommended range of 10 MB and 100 MB
- All of the above
  (Correct)

Explanation

Resizing a warehouse to a larger size is useful when the operations being performed by the warehouse will benefit from more compute resources, including:

1. Improving the performance of large, complex queries against large data sets.

2. Improving performance while loading and unloading significant amounts of data.

Note that, as a general practice, while performing data loads for Snowflake ETL, it is important to partition the data in your Snowflake or external locations like S3 buckets or Azure containers using logical, granular paths. By creating a partition using details like location or application along with the date when this data was written, you are

optimizing for a later data loading activity. When you load the data, you can simply copy any fraction of the partitioned data into Snowflake with a single command. You can copy data into Snowflake by the hour, day, month, or even year when you initially populate tables.

The best practice is to load smaller files of 10-100 MB to maximize parallel loading i.e, number of files will be distributed between available nodes for faster processing and performance.

Due to all these, the answer to this question is "All of the above"

Question 6: Skipped

Snowflake provides specific administration features and capabilities to support the following activities except?

- Manage databases and warehouses within snowflake account
- Manage roles and users within a snowflake account
- Monitor usage and manage resources to control costs in a snowflake account
- Manage 3rd party applications providing data to a snowflake account **(Correct)**

Question 7: Skipped

Select all characteristics of snowflake's multi-cluster environment

- Multiple virtual warehouses in a deployment **(Correct)**
- Individual warehouses scale out/in based on query activity **(Correct)**
- Multi-cluster warehouses support the same properties and actions as single-cluster warehouse **(Correct)**
- User must specify which cluster each query will use

Explanation

Multiple virtual warehouse in a deployment means specifying more than 1 warehouse while creating a multi-cluster warehouse. In a multi-cluster warehouse, you can specify the maximum number of server cluster(or virtual warehouse) greater than 1 and upto 10 and also a minimum number of clusters(equal to or less than the maximum(up to 10).

Additionally, multi-cluster warehouses `support all the same properties and actions as single-cluster warehouses`, including:

1. Specifying a warehouse size.

2. Resizing a warehouse at any time.

3. Auto-suspending a running warehouse due to inactivity; note that this does not apply to individual clusters, but rather the entire warehouse.

4. Auto-resuming a suspended warehouse when new queries are submitted.

In a multi-cluster warehouse, individual warehouses can scale up or down based on query load which is not possible in a single cluster warehouse. However as mentioned above, `it supports the same properties and actions like the single cluster warehouse`

Question 8: Skipped

What are the two modes in which a multi-cluster warehouse can run?

- ☐ Static
- ☐ Dynamic
- ☐ Maximized **(Correct)**
- ☐ Auto-scale **(Correct)**
- ☐ None of the above

Explanation
You can choose to run a multi-cluster warehouse in either of the following modes:

Maximized

This mode is enabled by specifying the ***same*** value for both maximum and minimum clusters (note that the specified value must be larger than 1). In this mode, when the warehouse is started, Snowflake starts all the clusters so that maximum resources are available while the warehouse is running.

This mode is effective for statically controlling the available resources (i.e. servers), particularly if you have large numbers of concurrent user sessions and/or queries and the numbers do not fluctuate significantly.

Auto-scale

This mode is enabled by specifying ***different*** values for maximum and minimum clusters. In this mode, Snowflake starts and stops clusters as needed to dynamically manage the load on the warehouse:

As the number of concurrent user sessions and/or queries for the warehouse increases, and queries start to queue due to insufficient resources, Snowflake automatically starts additional clusters, up to the maximum number defined for the warehouse.

Similarly, as the load on the warehouse decreases, Snowflake automatically shuts down clusters to reduce the number of running servers and, correspondingly, the number of credits used by the warehouse.

To help control the usage of credits in Auto-scale mode, Snowflake provides a property, SCALING_POLICY, that determines the scaling policy to use when automatically starting or shutting down additional clusters. For more information, see Setting the Scaling Policy for a Multi-cluster Warehouse (in this topic).

https://docs.snowflake.com/en/user-guide/warehouses-multicluster.html#maximized-vs-auto-scale

Question 9: Skipped

What does snowflake offer to help control the credits consumed by a multi-cluster warehouse running in Auto-scale mode?

- Auto scale
- Scaling policy
  (Correct)
- Maximum_number_of_server_clusters
- Minimum_number_of_server_clusters
- None of the above

Explanation

To help control the usage of credits in Auto-scale mode, Snowflake provides a property, **SCALING_POLICY,** that determines the scaling policy to use when automatically starting or shutting down additional clusters

Question 10: Skipped

What scaling policies are supported by snowflake?

- [ ] Standard
(Correct)
- [ ] Economy
(Correct)
- [ ] Premium

Explanation
Standard (default)

Prevents/minimizes queuing by favoring starting additional clusters over conserving credits.

Economy

Conserves credits by favoring keeping running clusters fully-loaded rather than starting additional clusters, which may result in queries being queued and taking longer to complete.

https://docs.snowflake.com/en/user-guide/warehouses-multicluster.html#setting-the-scaling-policy-for-a-multi-cluster-warehouse

Question 11: Skipped

You are sizing a snowflake warehouse, what factors would you consider?

- [ ] Number of users
- [ ] Number of concurrent queries
(Correct)
- [ ] Number of tables being queried
(Correct)
- [ ] Data size and composition
(Correct)

Explanation

Warehouse size does not depend on number of users. Number of concurrent queries in important, because if the warehouse does not have enough resource, queries will start queueing. Same is the case with number of tables and data size.

Question 12: Skipped

Compute usage billing in snowflake is based on

- Per-second basis, with a minimum of 60 seconds
  (Correct)
- Per-second basis only
- By number of users

Explanation

Snowflake compute utilizes per second billing after the first minute. what it means is that when a warehouse starts or resumes, it bills for the first minute whether you use it or not, after that the billing is based on per second. For example if a warehouse is resumed and suspended after 30 seconds, you will still have to pay for 60 seconds because that is the first minute after the warehouse resumed.

Question 13: Skipped

SQL functionality in snowflake can be extended by

- USER DEFINED FUNCTIONS USING SQL
  (Correct)
- USER DEFINED FUNCTIONS USING JAVASCRIPT
  (Correct)
- SESSION VARIABLES
  (Correct)
- NONE OF THE ABOVE

Explanation

Snowflake supports SQL/SESSION variables declared by the user. They have many uses, such as storing application-specific environment settings.

https://docs.snowflake.com/en/sql-reference/session-variables.html#sql-variables

User-defined functions (UDFs) let you extend the system to perform operations that are not available through the built-in, system-defined functions provided by Snowflake. Snowflake currently supports two types of UDFs, *SQL* and *JavaScript*

Example of SQL UDF

```
create function area_of_circle(radius float)
  returns float
  as
  $$
    pi() * radius * radius
  $$
  ;
```

Example of JAVASCRIPT UDF

```
CREATE OR REPLACE FUNCTION array_sort(a array)
  RETURNS array
  LANGUAGE JAVASCRIPT
AS
$$
  return A.sort();
$$
;


-- Call the UDF with a small array.
SELECT ARRAY_SORT(PARSE_JSON('[2,4,5,3,1]'));
```

Question 14: Skipped

To provide failover protection, how many availability zones does snowflake replicate to

- one
- two
  (Correct)
- three
- as configured by user

Explanation

Providing an even higher degree of data protection and service resilience, within the same deployment region, Snowflake provides standard failover protection **across three availability zones (including the primary active zone**). Your data and business are protected. As you ingest your data, it is synchronously and transparently replicated across availability zones. This protection is automatically extended from Snowflake to customers, at no added charge.

https://www.snowflake.com/blog/how-to-make-data-protection-and-high-availability-for-analytics-fast-and-easy/

Additional note

The answer to this question has been corrected(from 3 to 2) based on the feedback from a student. The explanation above is correct, but the replication happens to two additional availability zones only, and the primary zone is the original location of the data.

Question 15: Skipped

Which of the following is not a characteristic of micro-partitions in snowflake?

- New partitions are created in logical properties
  (Correct)
- Avoids skews between partitions
- Partitioning is automatically completed on the natural ingestion order of the data
- The natural ingestion order maintains correlations between columns which could be useful for pruning

Explanation
Snowflake uses physical properties and partitions can overlap ranges

Question 16: Skipped

The storage architecture of snowflake has two key features, they are

- Time travel
  (Correct)
- Replication
- Zero-copy cloning
  (Correct)

- Query tag

Explanation
Snowflake Time Travel enables accessing historical data (i.e. data that has been changed or deleted) at any point within a defined period. It serves as a powerful tool for performing the following tasks:

1. Restoring data-related objects (tables, schemas, and databases) that may have been accidentally or intentionally deleted.

2. Duplicating and backing up data from key points in the past.

3. Analyzing data usage/manipulation over specified periods of time

Cloning also referred to as “zero-copy cloning” creates a copy of a database, schema or table. **A snapshot of data** present in the source object is taken when the clone is created and is made available to the cloned object. The cloned object is writable and is independent of the clone source. That is, changes made to either the source object or the clone object are not part of the other. Cloning a database will clone all the schemas and tables within that database. Cloning a schema will clone all the tables in that schema

Question 17: Skipped

Which is not a characteristic of time travel?

- Protection from accidental data operations
- Previous versions of data automatically retained
- Retention period selected by customers (up to 90 days for Enterprise edition)
- Recover data with the cost of running backups
(Correct)

Question 18: Skipped

The commands to load data into snowflake are

- COPY AND PUT
(Correct)
- COPY AND INSERT
(Correct)

- INSERT AND PUT
  (Correct)
- NONE OF THE ABOVE

Explanation
PUT in snowflake is used to load data from your local system to snowflake stage

```
put file:///tmp/data/mydata.csv @my_int_stage;
```

COPY and INSERT are the operations through which you can load data into snowflake tables

Question 19: Skipped

In order to load data into snowflake, which of the below are required?

- Virtual Warehouse
  (Correct)
- Predefined target table
  (Correct)
- Staging location with data staged
  (Correct)
- File format
  (Correct)

Explanation
Virtual warehouse is required to provide compute resources

Target table must be pre-defined load does not have feature to auto create tables

Data must be staged through the PUT command

File format must be defined to indicate whether CSV, JSON, XML etc is being loaded

Question 20: Skipped

Which of the following is not a snowpipe feature?

Snowpipe can load data from any internal or external stage

- It is a server-less compute model
- The service provides REST end points and uses snowflake provided compute resources to load the data and retrieve history reports
- Snowpipe loads data after it is in stage and use executes the LOADDATA command **(Correct)**

Explanation
There is no LOADDATA command and data is automatically loaded by snowpipe after the files are added to the stage

https://docs.snowflake.com/en/user-guide/data-load-snowpipe-intro.html#introduction-to-snowpipe

Question 21: Skipped

Which feature does not come with query profile?

- Graphical representation of the main components of the processing plan of the query
- Details and statistics for the overall query
- Hints for improving the query performance **(Correct)**
- Statistics for each component of the query

Explanation
Go to query profile in snowflake web UI and see if you can find HINT:)

Question 22: Skipped

Snowflake mechanism to limit the number of micro-partitions scanned by a query is called what?

- Pruning **(Correct)**
- Selection
- Filtering
- None of the above

Explanation

Snowflake makes extensive use of pruning to reduce the amount of data that has to be read from storage. In summary, this means that a query like

SELECT SUM(x) FROM T1 WHERE y=42

will not read Columns x and y completely from storage but will make use of the predicate y=42 and limit the table scan to the subset of all partitions that can potentially match this condition. Snowflake can do this because it stores the min and max value of each of the columns in the metadata store

Question 23: Skipped

Which type of data integration tools leverage snowflake's scalable compute for data transformation?

- ELT
  (Correct)
- ETL
- Database replication
- Streaming

Explanation
Commonly referred to as ETL, data integration encompasses the following primary operations:

Extract

Exporting data from specified data sources.

Transform

Modifying the source data (as needed), using rules, merges, lookup tables or other conversion methods, to match the target.

Load

Importing the resulting transformed data into a target database.

More recent usage **references the term** `ELT`**, emphasizing that the transformation part of the process** does not necessarily need to be performed before loading, particularly in systems such as Snowflake that support transformation during or after loading.

In addition, the definition of data integration has expanded to include a wider range of operations, including:

Data preparation.

Data migration/movement and management.

Data warehouse automation.

https://docs.snowflake.com/en/user-guide/ecosystem-etl.html

Question 24: Skipped

Snowflake offers tools to extract data from source systems

- TRUE
- FALSE
(Correct)

Question 25: Skipped

Which of the following are options when creating a virtual warehouse

- Auto Suspend
(Correct)
- Auto Resume
(Correct)
- Storage Size
- Server count

Explanation

Syntax

```
CREATE [ OR REPLACE ] WAREHOUSE [ IF NOT EXISTS ] <name>
       [ [ WITH ] objectProperties ]
       [ objectParams ]
```

Where:

```
objectProperties ::=
  WAREHOUSE_SIZE = XSMALL | SMALL | MEDIUM | LARGE | XLARGE | XXLARGE | XXXLARGE
| X4LARGE
  MAX_CLUSTER_COUNT = <num>
```

```
  MIN_CLUSTER_COUNT = <num>
  SCALING_POLICY = STANDARD | ECONOMY
  AUTO_SUSPEND = <num> | NULL
  AUTO_RESUME = TRUE | FALSE
  INITIALLY_SUSPENDED = TRUE | FALSE
  RESOURCE_MONITOR = <monitor_name>
  COMMENT = '<string_literal>'
objectParams ::=
  MAX_CONCURRENCY_LEVEL = <num>
  STATEMENT_QUEUED_TIMEOUT_IN_SECONDS = <num>
  STATEMENT_TIMEOUT_IN_SECONDS = <num>
```

Question 26: Skipped

The warehouse cache may be reset if a running warehouse is suspended and then resumes

- TRUE
  (Correct)
- FALSE

Explanation
You do not need to remember this, just understand the concept. Warehouses are nothing but VM instances underneath. And the cache is nothing but SSD cache attached to the instances, If the warehouse is suspended and resumed again it may not get teh same instance.

Question 27: Skipped

A table in snowflake can only be queried using the virtual warehouse used to load the data

- TRUE
- FALSE
  (Correct)

Explanation
Again apply concept, storage and compute are isolated in snowflake. That is why we can run different workloads in different warehouses but they all can refer to the same data

Question 28: Skipped

Snowflake caches are automatically invalidated if the underlying data changes

- TRUE
  (Correct)
- FALSE

Question 29: Skipped

The query profiler view is only available for completed queries

- TRUE
- FALSE
  (Correct)

Question 30: Skipped

Which best describes zero-copy cloning?

- Metadata only operation
- No replication of data
- Unmodified data stored once and modified data stored as new micro partitions
- All of the above
  (Correct)

Explanation
When a clone is created of a table, the clone utilizes no data storage because it shares all the existing micro-partitions of the original table at the time it was cloned; however, rows can then be added, deleted, or updated in the clone independently from the original table. Each change to the clone results in new micro-partitions that are owned exclusively by the clone and are protected through CDP(continuous data protection).

Question 31: Skipped

Your business team runs a set of identical queries every day after the batch ETL run is complete. From the following actions, what is the best action that you will recommend.

- After the ETL run, execute the identical queries so that they remain in the result cache
  (Correct)
- After the ETL run, resize the warehouse to a larger warehouse
- After the ETL run, copy the tables to another schema for the business users to query

Explanation

Please note the key word here which is `IDENTICAL` queries. When a query is run the first time, and the same query is run the second time it picks up the results from the query result cache and it does not cost you any compute. The query result cache is valid for 24 hours and it gets extended for another 24 hours, every time you access it(even if you access it at 23 hours 59 seconds).

Try this out yourself

Using worksheet create a table and insert the rows as below

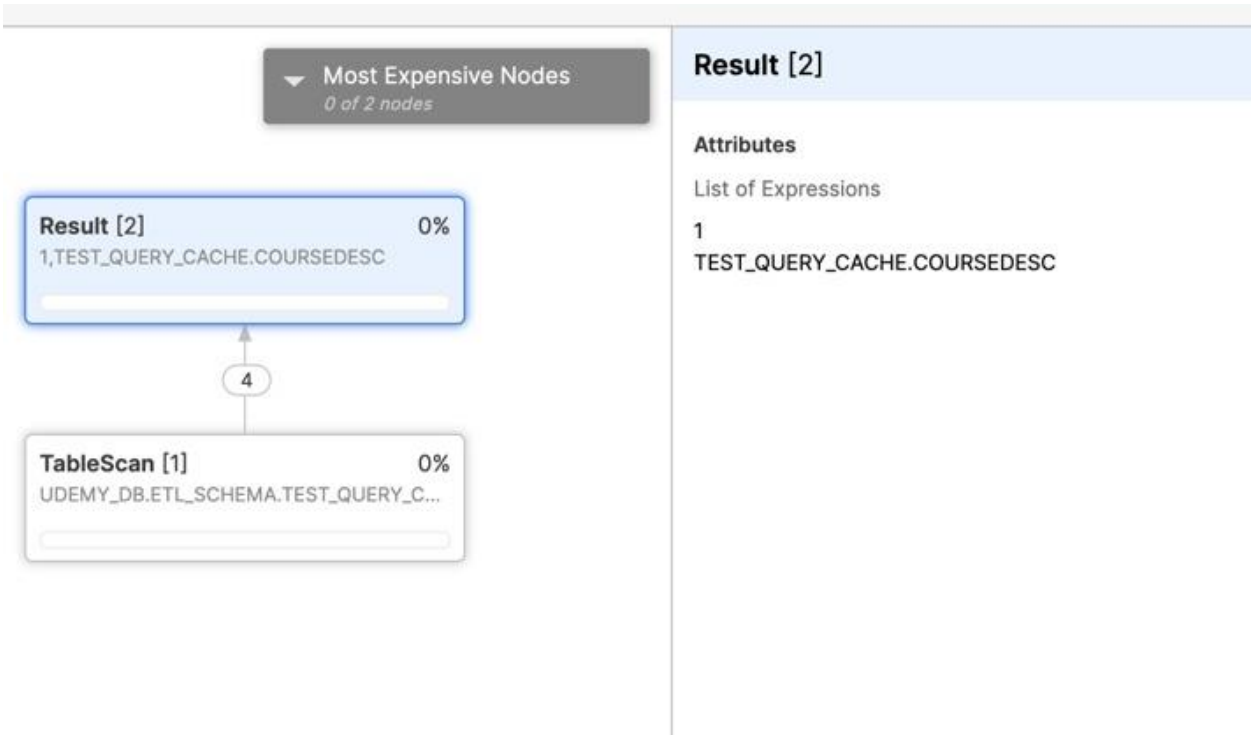
```
CREATE TABLE TEST_QUERY_CACHE(COURSEID NUMBER, COURSEDESC VARCHAR);


INSERT INTO TEST_QUERY_CACHE VALUES(1, 'SNOWFLAKE');
INSERT INTO TEST_QUERY_CACHE VALUES(1, 'C++');
INSERT INTO TEST_QUERY_CACHE VALUES(1, 'PYTHON');
```

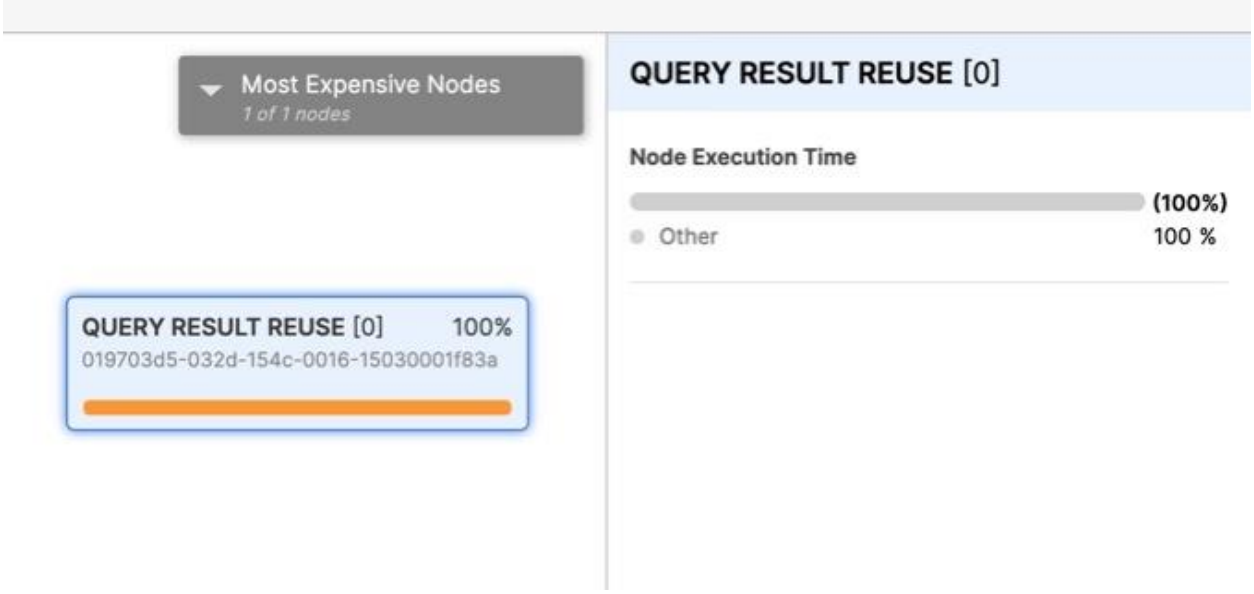
Run the below query and go to query profile

```
SELECT * FROM TEST_QUERY_CACHE;
```

You see that it is doing a table scan



Run the query again, and you will see that it has picked up from query cache



Next question:) How do you disable query cache. to know the answer look at the next question

Question 32: Skipped

If you would like to disable the query cache, what will you do?

- ALTER SESSION SET USE_CACHED_RESULT = FALSE;
  (Correct)
- ALTER SESSION SET USE_CACHED_RESULT = OFF;
- ALTER SESSION SET USE_QUER_CACHE = TRUE;

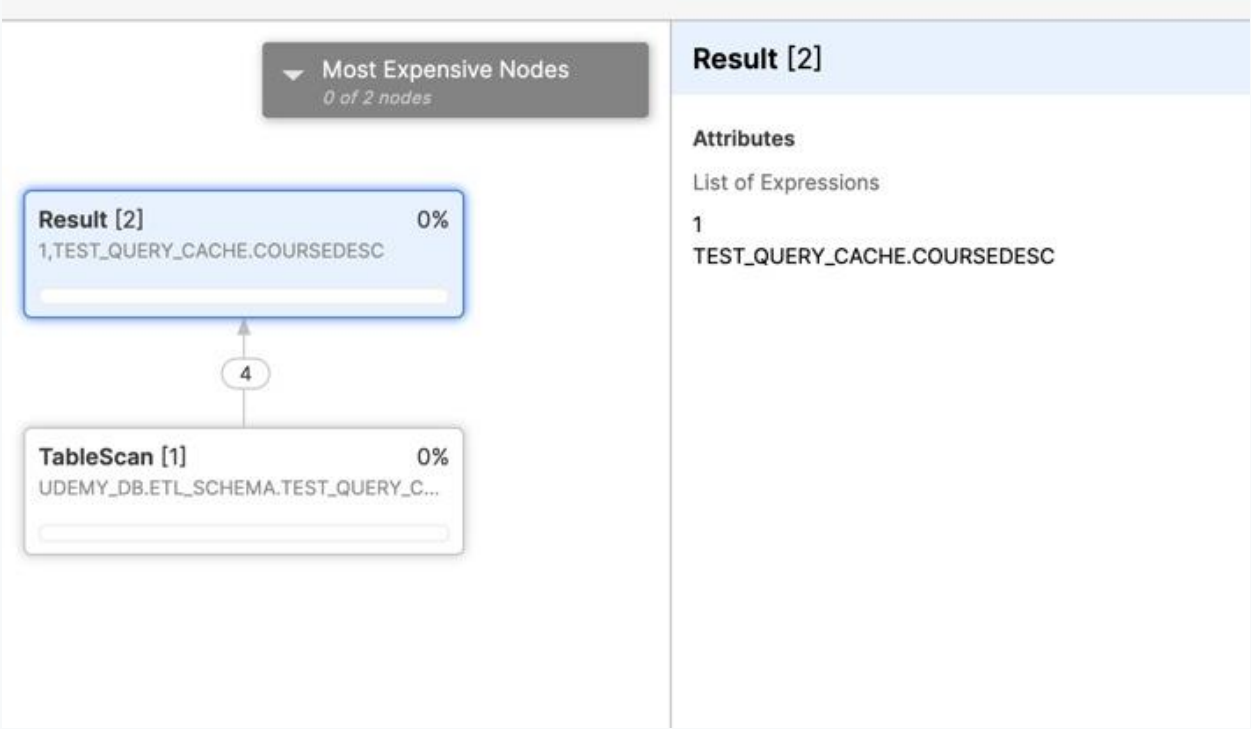
Explanation
Try it again, run the below command

```
ALTER SESSION SET USE_CACHED_RESULT = FALSE;
```

Then the run the query one more time, what do you see

```
SELECT * FROM TEST_QUERY_CACHE;
```

You will see the table scan coming back:)



Question 33: Skipped

A task can have a maximum of 100 *child* tasks

- TRUE
  (Correct)
- FALSE

Explanation

This is something that you will need to remember

```
A simple tree of tasks is limited to a maximum of 1000 tasks total
(including the root task) in a resumed state. An individual task in the tree
is limited to a single predecessor task; however, a task can have a maximum
of 100 child tasks (i.e. other tasks that identify the task as a
predecessor).
```

Question 34: Skipped

Snowflake guarantees that only one instance of a task with a defined ***predecessor task*** is running at a given time

-

TRUE

- FALSE
  (Correct)

Explanation
Snowflake cannot guarantee that only one instance of a task with a defined ***predecessor task*** is running at a given time.

Question 35: Skipped

You created an warehouse(ETL_WH) which is sized at XSMALL, you want to resize the Warehouse to SMALL. How will you do that?

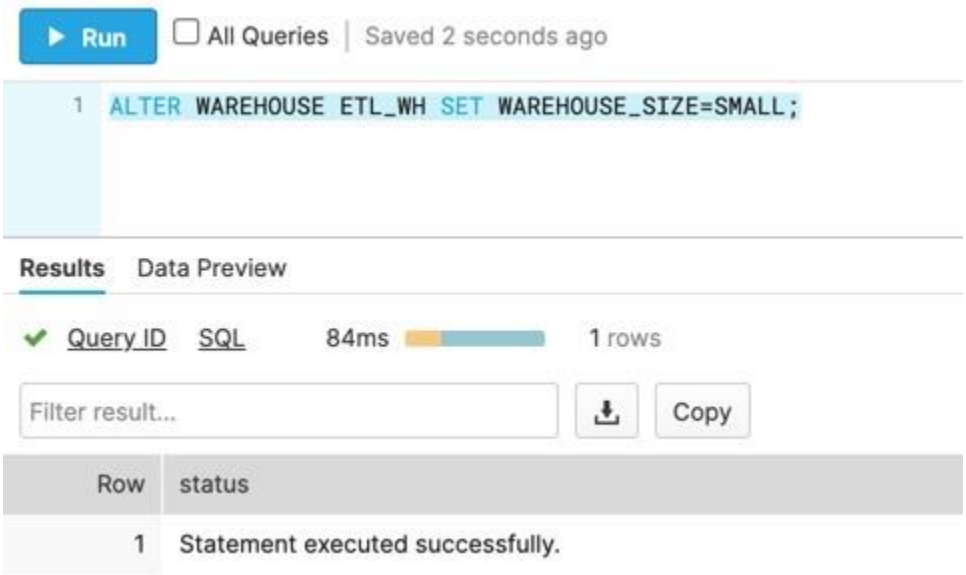
- ```
  1.  alter warehouse etl_wh set warehouse_size=SMALL;
  ```

  (Correct)

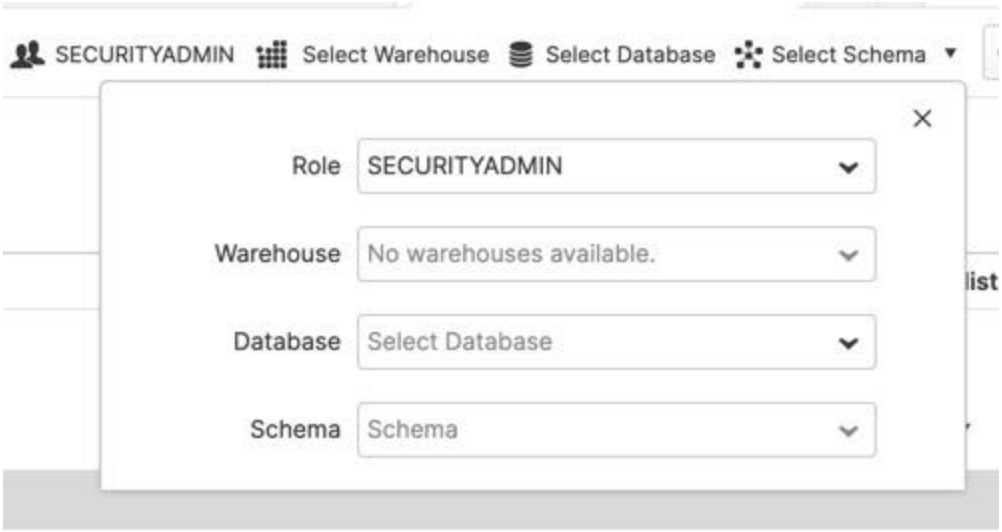
- You will need to recreate the warehouse with the new size
- Once created an warehouse can never be re-sized

Explanation
Ok, to remember it for the rest of your life, please log on to SNOWFLAKE and execute

```
ALTER WAREHOUSE ETL_WH SET WAREHOUSE_SIZE=SMALL;
```



Ok just for fun, change your role to SECURITYADMIN, and try to resize the WAREHOUSE. Will you be able to do it?
```



No, by default SECURITYADMIN or any other role cannot change the warehouse size. You need to be SYSADMIN or ACCOUNTADMIN. Recommended approach is to use SYSADMIN .

Question 36: Skipped

You have just loaded a file named student_data.csv from your snowflake stage to snowflake table. You try to reload again using the COPY INTO command. You will be able to load the file.

- TRUE
- FALSE
  (Correct)

Explanation
Let's understand why we will not be able to load the file again. Please also look at the `64 Days` mentioned below

Snowflake maintains detailed metadata for each table into which data is loaded, including:

1. Name of each file from which data was loaded

2. File size

3. ETag for the file

4. Number of rows parsed in the file

5. Timestamp of the last load for the file

6. Information about any errors encountered in the file during loading

`This load metadata expires after 64 days`. If the LAST_MODIFIED date for a staged data file is less than or equal to 64 days, the COPY command can determine its load status for a given table and `prevent reloading (and data duplication).` The LAST_MODIFIED date is the timestamp when the file was initially staged or when it was last modified, whichever is later.

So, is there no way to reload it?

There is a way, you can use FORCE=TRUE

Reloading Files

Add `FORCE = TRUE` to a COPY command to reload (duplicate) data from a set of staged data files that have not changed (i.e. have the same checksum as when they were first loaded).

In the following example, the first command loads the specified files and the second command forces the same files to be loaded again (producing duplicate rows), even though the contents of the files have not changed:

```
copy into load1 from @%load1/data1/
    files=('test1.csv', 'test2.csv')


copy into load1 from @%load1/data1/
```

```
    files=('test1.csv', 'test2.csv')
    force=true;
```

Question 37: Skipped

For a particular snowflake session, you can specify more than one virtual warehouse

- TRUE
- FALSE
  (Correct)

Explanation
A session can only use one warehouse but you can change the warehouse using USE WAREHOUSE <WAREHOUSE_NAME>

Warehouse Usage in Sessions

When a session is initiated in Snowflake, the session does not, by default, have a warehouse associated with it. Until a session has a warehouse associated with it, queries cannot be submitted within the session.

Default Warehouse for Users

To facilitate querying immediately after a session is initiated, Snowflake supports specifying a default warehouse for each individual user. The default warehouse for a user is used as the warehouse for all sessions initiated by the user.

A default warehouse can be specified when creating or modifying the user, either through the web interface or using CREATE USER/ALTER USER.

Default Warehouse for Client Utilities/Drivers/Connectors

In addition to default warehouses for users, any of the Snowflake clients (SnowSQL, JDBC driver, ODBC driver, Python connector, etc.) can have a default warehouse:

SnowSQL supports both a configuration file and command line option for specifying a default warehouse.

The drivers and connectors support specifying a default warehouse as a connection parameter when initiating a session.

For more information, see Connecting to Snowflake.

Precedence for Warehouse Defaults

When a user connects to Snowflake and start a session, Snowflake determines the default warehouse for the session in the following order:

Default warehouse for the user,

» **overridden by…**

Default warehouse in the configuration file for the client utility (SnowSQL, JDBC driver, etc.) used to connect to Snowflake (if the client supports configuration files),

» **overridden by…**

Default warehouse specified on the client command line or through the driver/connector parameters passed to Snowflake.

Note

```
In addition, the default warehouse for a session can be changed at any time
by executing the USE WAREHOUSE command within the session.
```

Question 38: Skipped

Time travel and fail safe requires additional storage which will cost you storage cost

- TRUE
  (Correct)
- FALSE

Explanation

```
Storage fees are incurred for maintaining historical data during both the Time Tr
avel and Fail-safe periods.
```

Storage Usage and Fees

The fees are calculated for each 24-hour period (i.e. 1 day) from the time the data changed. The number of days historical data is maintained is based on the table type and the Time Travel retention period for the table.

Also, Snowflake minimizes the amount of storage required for historical data by maintaining only the information required to restore the individual table rows that were updated or deleted. As a result, storage usage is calculated as a percentage of the table that changed. Full copies of tables are only maintained when tables are dropped or truncated.

Question 39: Skipped

From which of the following stage locations can snowflake load data

- [ ] AWS S3
(Correct)
- [ ] AZURE BLOB STORAGE
(Correct)
- [ ] GOOGLE CLOUD STORAGE
(Correct)
- [ ] INTERNAL STAGE
(Correct)
- [ ] FILE SERVER
- [ ] IBM CLOUD STORAGE

Explanation
External Stages

Loading data from any of the following cloud storage services is supported regardless of the cloud platform that hosts your Snowflake account:

```
Amazon S3

Google Cloud Storage

Microsoft Azure
```

Upload (i.e. *stage*) files to your cloud storage account using the tools provided by the cloud storage service.

A named external stage is a database object created in a schema. This object stores the URL to files in cloud storage, the settings used to access the cloud storage account, and convenience settings such as the options that describe the format of staged files. Create stages using the CREATE STAGE command.

`Internal Stages`

Snowflake maintains the following stage types in your account:

User

A user stage is allocated to each user for storing files. This stage type is designed to store files that are staged and managed by a single user but can be loaded into multiple tables. User stages cannot be altered or dropped.

Table

A table stage is available for each table created in Snowflake. This 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.

Note that a table stage is not a separate database object; rather, it is an implicit stage tied to the table itself. A table stage has no grantable privileges of its own. To stage files to a table stage, list the files, query them on the stage, or drop them, you must be the table owner (have the role with the OWNERSHIP privilege on the table).

Named

A named internal stage is a database object created in a schema. This stage type can store files that are staged and managed by one or more users and loaded into one or more tables. Because named stages are database objects, the ability to create, modify, use, or drop them can be controlled using security access control privileges. Create stages using the CREATE STAGE command.

Upload files to any of the internal stage types from your local file system using the PUT command.

Question 40: Skipped

ACCOUNTADMIN ROLE encapsulates the SYSADMIN and SECURITYADMIN system-defined roles and ACCOUNTADMIN sits on top of the role hierarchy

- TRUE
(Correct)

- FALSE

Explanation
ACCOUNTADMIN

(aka Account Administrator)

Role that encapsulates the SYSADMIN and SECURITYADMIN system-defined roles. It is the top-level role in the system and should be granted only to a limited/controlled number of users in your account.

Question 41: Skipped

Select the two true statements about multi cluster warehouse

- These type of warehouses cannot be resumed or suspended
- Multi cluster warehouses can be created on snowflake hosted on AWS only
- As query demand decreases, clusters are removed from this warehouse **(Correct)**
- When query workload increases, clusters are automatically added **(Correct)**

Explanation
In Auto-scale mode, a multi-cluster warehouse eliminates the need for resizing the warehouse or starting and stopping additional warehouses to handle fluctuating workloads. `Snowflake automatically starts and stops additional clusters as needed.`

In Maximized mode, you can control the capacity of the warehouse by increasing or decreasing the number of clusters as needed.

Question 42: Skipped

If you want to clone a table, your role must have which access on the Table

- SELECT **(Correct)**
- USAGE
- READ
- CLONE

Explanation
To create a clone, your current role must have the following privilege(s) on the source object:

Tables - SELECT

Pipes, Streams, Tasks - OWNERSHIP

Other objects - USAGE

In addition, to clone a schema or an object within a schema, your current role must have required privileges on the container object(s) for both the source and the clone.

Question 43: Skipped

You have three worksheets open in Snowflake Web UI. You can have three different roles in these worksheets.

- TRUE
  (Correct)
- FALSE

Explanation
Please do not believe me:) Try it out

Log on to snowflake, open three worksheets and try to use a different role. Run the below in worksheet and see if it works

Worksheet 1

```
USE ROLE SYSADMIN
```

Worksheet 2

```
USE ROLE ACCOUNTADMIN
```

Worksheet 3

```
USE ROLE PUBLIC
```

Question 44: Skipped

When will you consider creating a clustering key on a table. Please select all that applies

- The table houses multi-terabyte data
  (Correct)
- The table has variant columns
- The query performance on the table has degraded over time
  (Correct)

Explanation
No need to mug this up

1. Clustering will be effective only when you have enough partitions in the table and the table is more than 1 TB

2. As you insert/update rows in the table, the partitions are moved around and hence it no longer remains clustered. With Auto clustering, Snowflake automatically reclusters the table based on a proprietary algorithm

Question 45: Skipped

Which of the below objects will contribute to your account level storage usage?

- Secure Views
- Internal Stage
  (Correct)
- External stage
  (Correct)
- Database and all tables inside it
  (Correct)
- Functions

Explanation
SECURE VIEWS and FUNCTION definitions are stored in metadata store. They do not store any data physically.

Update to the explanation

One of my students asked a valid question that external stage is a pointer to the cloud provider's storage and hence from a snowflake perspective there is no cost. This is absolutely true. But when this question is presented to you and it is not explicitly asking 'snowflake account', you should tick external stage also. If the question specifically asks about 'snowflake account', then you should not tick this option. This is kind of a trick question so please pay attention to the question text.

More on external stage

External stage

References data files stored in `a location outside of Snowflake`. Currently, the following cloud storage services are supported:

1. Amazon S3 buckets

2. Google Cloud Storage buckets

3. Microsoft Azure containers

The storage location can be either private/protected or public.

Question 46: Skipped

A consumer of shared database can add as many tables and views they want in the shared database

- FALSE
(Correct)
- TRUE

Explanation
Remember, we discussed earlier that SHARES are read-only and cannot be changed by consumer

Question 47: Skipped

To have multi cluster data virtual warehouse, what should be the minimum snowflake subscription

- ENTERPRISE
(Correct)
- BUSINESS CRITICAL
- STANDARD

Question 48: Skipped

Once a new release has been deployed, Snowflake does not move all accounts to the release at the same time. It follows a staged release approach.

- TRUE
(Correct)

- FALSE

Explanation
Staged Release

Once a new release has been deployed, Snowflake does not move all accounts to the release at the same time. Accounts are moved to the release using a three stage approach over two days. Accounts are moved to the new release in the following order, based on their Snowflake Edition:

Day 1

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

Day 1 or 2

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

Day 2

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

The minimum amount of elapsed time between the early access and final stages is 24 hours. This staged approach enables Snowflake to monitor activity as accounts are moved and respond to any issues that may occur.

Question 49: Skipped

When you are loading data through the COPY command, the table and the file being loaded must have the same order of columns

- FALSE
(Correct)

- TRUE

Explanation
Note that the actual field/column order in the data files can be different from the column order in the target table. It is only important that the SELECT list maps fields/columns in the data files to the ***corresponding*** columns in the table.

Question 50: Skipped

A reader account user will not be able to do the following

- LOAD DATA
(Correct)

- INSERT
  (Correct)
- SELECT
- UPDATE
  (Correct)

Explanation
A reader account is intended primarily for querying data shared by the provider of the account. Adding new data to the account and/or updating shared data in the account is not supported. As such, the following DML and DDL commands are not allowed:

INSERT

UPDATE

DELETE

MERGE

COPY INTO <table>

CREATE PIPE

CREATE SHARE

All other operations are allowed.

Question 51: Skipped

Which two modes are available in Multi cluster warehouse

- Auto-scale
  (Correct)
- Peak Scale
- Flexible
- Maximized
  (Correct)

Explanation
Maximized vs. Auto-scale

You can choose to run a multi-cluster warehouse in either of the following modes:

Maximized

This mode is enabled by specifying the ***same*** value for both maximum and minimum clusters (note that the specified value must be larger than 1). In this mode, when the warehouse is started, Snowflake starts all the clusters so that maximum resources are available while the warehouse is running.

This mode is effective for statically controlling the available resources (i.e. servers), particularly if you have large numbers of concurrent user sessions and/or queries and the numbers do not fluctuate significantly.

Auto-scale

This mode is enabled by specifying ***different*** values for maximum and minimum clusters. In this mode, Snowflake starts and stops clusters as needed to dynamically manage the load on the warehouse:

As the number of concurrent user sessions and/or queries for the warehouse increases, and queries start to queue due to insufficient resources, Snowflake automatically starts additional clusters, up to the maximum number defined for the warehouse.

Similarly, as the load on the warehouse decreases, Snowflake automatically shuts down clusters to reduce the number of running servers and, correspondingly, the number of credits used by the warehouse.

To help control the usage of credits in Auto-scale mode, Snowflake provides a property, SCALING_POLICY, that determines the scaling policy to use when automatically starting or shutting down additional clusters. For more information, see Setting the Scaling Policy for a Multi-cluster Warehouse (in this topic).

Question 52: Skipped

Which of the below operations are allowed on an inbound share data?

- ☐ MERGE
- ☐ CREATE TABLE
- ☐ ALTER TABLE
- ☐ SELECT WITH GROUP BY **(Correct)**
- ☐

SELECT WITH JOIN
(Correct)

- INSERT INTO

Explanation
This is a trick question:) remember a share is read only, so you can only select data from a share

Important

All database objects shared between accounts are ***read-only*** (i.e. the objects cannot be modified or deleted, including adding or modifying table data).

Question 53: Skipped

Select the most accurate statement regarding Snowflake's transaction support

- ACID Compliant
(Correct)
- Requires special user configuration
- BASE compliant
- Works like a NoSQL database

Question 54: Skipped

Select the table type that is automatically deleted after the session is closed and hence it has no fail-safe or time travel option post closure of the session

- TRANSIENT
- TEMPORARY
(Correct)
- PERMANENT
- EXTERNAL

Explanation
Please create a temporary table and try it out

```
CREATE TEMPORARY TABLE T1(COL1 VARCHAR);

INSERT INTO T1 VALUES('TESTING TIME TRAVEL FOR T1');
```

Close the worksheet and open another worksheet and run the below query, are you able to time tarvel

```
SELECT * FROM T1 AT(OFFSET => -60*1);
```

Question 55: Skipped

What does snowflake recommend for making a variant column accessible in a BI Tool?

- Create a View
(Correct)
- Create a separate table
- Use Parse_Json
- Convert into external table

Question 56: Skipped

When you load data into snowflake, snowflake reorganizes the data into its internal optimized compressed columnar format

- TRUE
(Correct)
- FALSE

Explanation
When data is loaded into Snowflake, Snowflake reorganizes that data into its internal optimized, compressed, columnar format. Snowflake stores this optimized data in cloud storage.

Snowflake manages all aspects of how this data is stored — the organization, file size, structure, compression, metadata, statistics, and other aspects of data storage are handled by Snowflake. The data objects stored by Snowflake are not directly visible nor accessible by customers; they are only accessible through SQL query operations run using Snowflake.

Question 57: Skipped

When you clone a table, it does not include the load history of the source table, and hence the data files that were already loaded in source table can be reloaded in the clone table

- TRUE
(Correct)
- FALSE

Question 58: Skipped

A snowflake account hosted on AWS cannot load data from a file that is staged in GCP or Azure

- FALSE
(Correct)
- TRUE

Question 59: Skipped

If you clone a database, the internal snowflake stages in that database are also cloned

- FALSE
(Correct)
- TRUE

Explanation
Cloning and Stages

The following rules apply to cloning stages or objects that contain stages (i.e. databases and schemas):

Individual external named stages can be cloned; internal named stages cannot be cloned.

When cloning a database or schema:

External named stages that were present in the source when the cloning operation started are cloned.

Tables are cloned, which means 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.

Question 60: Skipped

Which command is used to download data from Snowflake stage

- PUT
- GET
  (Correct)
- COPY
- DUMP

Explanation
GET

Downloads data files from one of the following Snowflake stages to a local directory/folder on a client machine:

Named internal stage.

Internal stage for a specified table.

Internal stage for the current user.

Typically, this command is executed after using the COPY INTO <location> command to unload data from a table into a Snowflake stage.