# Snowflake

## Exam Questions DEA-C01

SnowPro Advanced: Data Engineer Certification Exam

# About Exambible

## *Your Partner of IT Exam*

## Found in 1998

Exambible is a company specialized on providing high quality IT exam practice study materials, especially Cisco CCNA, CCDA, CCNP, CCIE, Checkpoint CCSE, CompTIA A+, Network+ certification practice exams and so on. We guarantee that the candidates will not only pass any IT exam at the first attempt but also get profound understanding about the certificates they have got. There are so many alike companies in this industry, however, Exambible has its unique advantages that other companies could not achieve.

## Our Advances

* 99.9% Uptime

All examinations will be up to date.

* 24/7 Quality Support

We will provide service round the clock.

* 100% Pass Rate

Our guarantee that you will pass the exam.

* Unique Gurantee

If you do not pass the exam at the first time, we will not only arrange FULL REFUND for you, but also provide you another exam of your claim, ABSOLUTELY FREE!

**NEW QUESTION 1**
The following chart represents the performance of a virtual warehouse over time:

A DataEngineer notices that the warehouse is queueing queries The warehouse is size X- Smallthe minimum and maximum cluster counts are set to 1 the scaling policy is set to i and auto-suspend is set to 10 minutes.
How canthe performance be improved?

A. Change the cluster settings
B. Increase the size of the warehouse
C. Change the scaling policy to economy
D. Change auto-suspend to a longer time frame

**Answer:** B

**Explanation:**
The performance can be improved by increasing the size of the warehouse. The chart shows that the warehouse is queueing queries, which means that there are more queries than the warehouse can handle at its current size. Increasing the size of the warehouse will increase its processing power and concurrency limit, which could reduce the queueing time and improve the performance. The other options are not likely to improve the performance significantly. Option A, changing the cluster settings, will not help unless the minimum and maximum cluster countsare increased to allow for multi-cluster scaling. Option C, changing the scaling policy to economy, will not help because it will reduce the responsiveness of the warehouse to scale up or down based on demand. Option D, changing auto-suspend to a longer time frame, will not help because it will only affect how long the warehouse stays idle before suspending itself.

**NEW QUESTION 2**
Which methods will trigger an action that will evaluate a DataFrame? (Select TWO)

A. DataFrame.random_split ( )
B. DataFrame.collect ()
C. DateFrame.select ()
D. DataFrame.col ( )
E. DataFrame.show ()

**Answer:** BE

**Explanation:**
The methods that will trigger an action that will evaluate a DataFrame are DataFrame.collect() and DataFrame.show(). These methods will force the execution of any pending transformations on the DataFrame and return or display the results. The other options are not methods that will evaluate a DataFrame. Option A, DataFrame.random_split(), is a method that will split a DataFrame into two or more DataFrames based on random weights. Option C, DataFrame.select(), is a method that will project a set of expressions on a DataFrame and return a new DataFrame. Option D, DataFrame.col(), is a method that will return a Column object based on a column name in a DataFrame.

**NEW QUESTION 3**
A company is using Snowpipe to bring in millions of rows every day of Change Data Capture (CDC) into a Snowflake staging table on a real-time basis The CDC needs to get processedand combined with other data in Snowflake and land in a final table as part of the full data pipeline.
How can a Data engineer MOST efficiently process the incoming CDC on an ongoing basis?

A. Create a stream on the staging table and schedule a task that transforms data from the stream only when the stream has data.
B. Transform the data during the data load with Snowpipe by modifying the related copy into statement to include transformation steps such as case statements andJOIN'S.
C. Schedule a task that dynamically retrieves the last time the task was run from information_schema-rask_hiSwOry and use that timestamp to process the delta of the new rows since the last time the task was run.
D. Use a create ok replace table as statement that references the staging table and includes all the transformation SQ
E. Use a task to run the full create or replace table as statement on a scheduled basis

**Answer:** A

**Explanation:**
The most efficient way to process the incoming CDC on an ongoing basis is to create a stream on the staging table and schedule a task that transforms data from the stream only when the stream has data. A stream is a Snowflake object that records changes made to a table, such as inserts, updates, or deletes. A stream can be queried like a table and can provide information about what rows have changed since the last time the stream was consumed. A task is a Snowflake object that can execute SQL statements on a schedule without requiring a warehouse. A task can be configured to run only when certain conditions are met, such as when a stream has data or when another task has completed successfully. By creating a stream on the staging table and scheduling a task that transforms data from the stream, the Data Engineer can ensure that only new or modified rows are processed and that no unnecessary computations are performed.

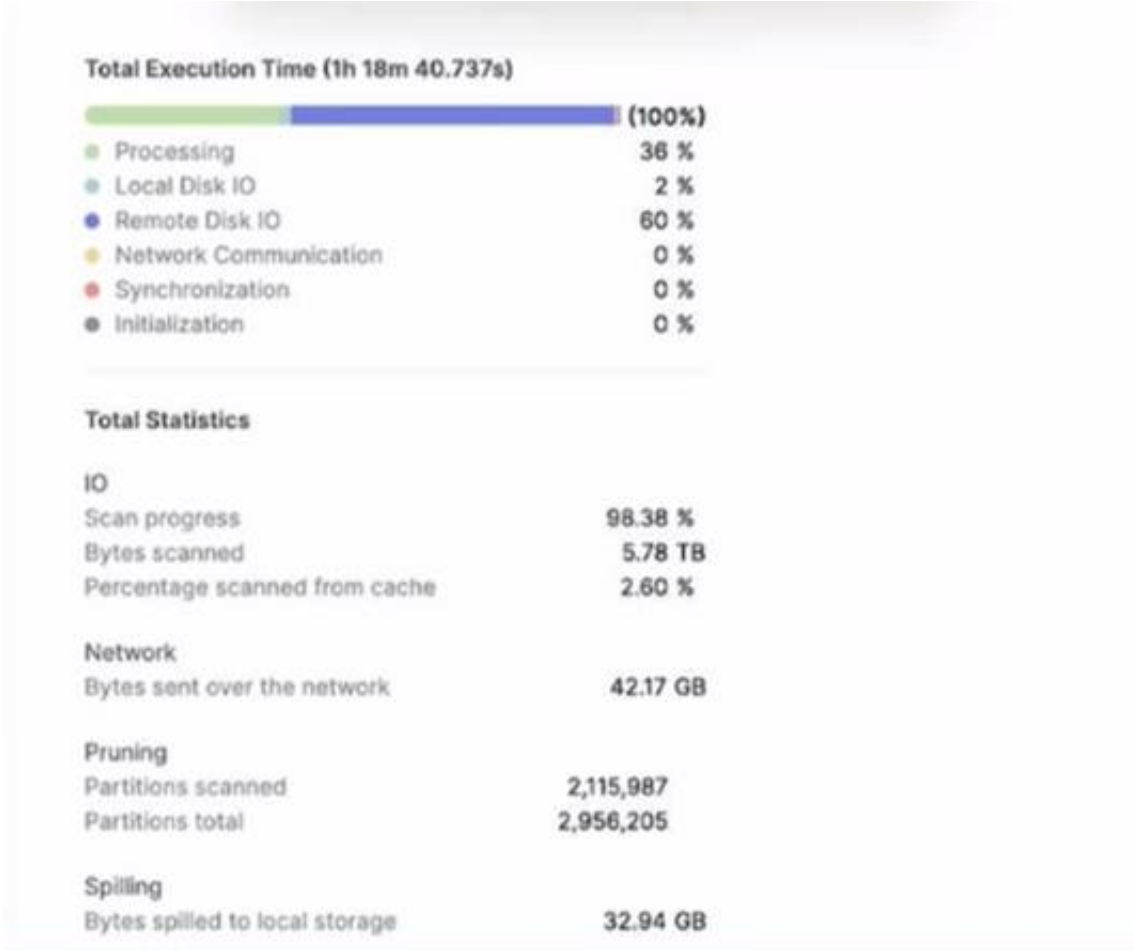
**NEW QUESTION 4**
Database XYZ has the data_retention_time_in_days parameter set to 7 days and table xyz.public.ABC has the data_retention_time_in_daysset to 10 days.
A Developer accidentally dropped the database containing this single table 8 days ago and just discovered the mistake.
How can the table be recovered?

A. undrop database xyz;
B. create -able abc_restore as select * from xyz.public.abc at {offset => -60*60*24*8};
C. create table abc_restore clone xyz.public.abc at (offset => -3€0G*24*3);
D. Create a Snowflake Support case lo restore the database and tab e from "a i-safe

**Answer:** A

**Explanation:**
The table can be recovered by using the undrop database xyz; command. This command will restore the database that was dropped within the last 14 days, along with all its schemas and tables, including the customer table. The data_retention_time_in_days parameter does not affect this command, as it only applies to time travel queries that reference historical data versions of tables or databases. The other options are not valid ways to recover the table. Option B is incorrect because creating a table as select * from xyz.public.ABC at {offset => -6060248} will not work, as this query will try to access a historical data version of the ABC table that does not exist anymore after dropping the database. Option C is incorrect because creating a table clone xyz.public.ABC at {offset => -360024*3} will not work, as this query will try to clone a historical data version of the ABC table that does not exist anymore after dropping the database. Option D is incorrect because creating a Snowflake Support case to restore the database and table from fail-safe will not work, as fail-safe is only available for disaster recovery scenarios and cannot be accessed by customers.

**NEW QUESTION 5**
A large table with 200 columns contains two years of historical data. When queried. the table is filtered on a single day Below is the Query Profile:




Using a size 2XL virtual warehouse, this query look over an hour to complete What will improve the query performance the MOST?

A. increase the size of the virtual warehouse.
B. Increase the number of clusters in the virtual warehouse
C. Implement the search optimization service on the table
D. Add a date column as a cluster key on the table

**Answer:** D

**Explanation:**
Adding a date column as a cluster key on the table will improve the query performance by reducing the number of micro-partitions that need to be scanned. Since the table is filtered on a single day, clustering by date will make the query more selective and efficient.

**NEW QUESTION 6**
While running an external function, me following error message is received: Error:function received the wrong number of rows
What iscausing this to occur?

A. External functions do not support multiple rows
B. Nested arrays are not supported in the JSON response
C. The JSON returned by the remote service is not constructed correctly
D. The return message did not produce the same number of rows that it received

**Answer:** D

**Explanation:**
The error message “function received the wrong number of rows” is caused by the return message not producing the same number of rows that it received. External functions require that the remote service returns exactly one row for each input row that it receives from Snowflake. If the remote service returns more or fewer rows than expected, Snowflake will raise an error and abort the function execution. The other options are not causes of this error message. Option A is

incorrect because external functions do support multiple rows as long as they match the input rows. Option B is incorrect because nested arrays are supported in the JSON response as long as they conform to the return type definition of the external function. Option C is incorrect because the JSON returned by the remote service may be constructed correctly but still produce a different number of rows than expected.

**NEW QUESTION 7**
A Data Engineer ran a stored procedure containing various transactions During the execution, the session abruptly disconnected preventing one transactionfrom committing or rolling hark.The transaction was left in a detached state and created a lock on resources
...must the Engineer take to immediately run a new transaction?

A. Call the system function SYSTEM$ABORT_TRANSACTION.
B. Call the system function SYSTEM$CANCEL_TRANSACTION.
C. Set the LOCK_TIMEOUTto FALSE in the stored procedure
D. Set the transaction abort on error to true in the stored procedure.

**Answer:** A

**Explanation:**
The system function SYSTEM$ABORT_TRANSACTION can be used to abort a detached transaction that was left in an open state due to a session disconnect or termination. The function takes one argument: the transaction ID of the detached transaction. The function will abort the transaction and release any locks held by it. The other options are incorrect because they do not address the issue of a detached transaction. The system function SYSTEM$CANCEL_TRANSACTION can be used to cancel a running transaction, but not a detached one. The LOCK_TIMEOUT parameter can be used to set a timeout period for acquiring locks on resources, but it does not affect existing locks. The TRANSACTION_ABORT_ON_ERROR parameter can be used to control whether a transaction should abort or continue when an error occurs, but it does not affect detached transactions.

**NEW QUESTION 8**
Which query will show a list of the 20 most recent executions of a specified task kttask, that have been scheduled within the last hour that have ended or are stillrunning's.
A)
```
select * from table(information_schema.task_history(scheduled_time_range_start
=>dateadd('hour',-1,current_timestamp()), result_limit => 20,
task_name=>'MYTASK'))
```
B)
```
select * from table(information_schema.task_history(scheduled_time_range_start
=>dateadd('hour',-1,current_timestamp()), result_limit => 20,
task_name=>'MYTASK')) where query_id IS NOT NULL;
```
C)
```
select * from table(information_schema.task_history(scheduled_time_range_start
=>dateadd('hour',-1,current_timestamp()), result_limit => 20,
task_name=>'MYTASK')) where STATE IN ('EXECUTING', 'SUCCEEDED', 'FAILED')
```
D)
```
select * from table(information_schema.task_history(scheduled_time_range_end
=>dateadd('hour',-1,current_timestamp()), result_limit => 10,
task_name=>'MYTASK')) where STATE IN ('EXECUTING', 'SUCCEEDED')
```

A. Option A
B. Option B
C. Option C
D. Option D

**Answer:** B

**NEW QUESTION 9**
A Data Engineer enables a result cache at the session level with the following command: ALTER SESSION SET USE CACHED RESULT = TRUE;
The Engineer then runs the following select query twice without delay:

```
SELECT *
FROM SNOWFLAKE_SAMPLE_DATA.TPCH_SF1.CUSTOMER
SAMPLE(10) SEED (99);
```

The underlying table does not change between executions What are the results of both runs?

A. The first and second run returned the same results because sample is deterministic
B. The first and second run returned the same results, because the specific SEEDvalue was provided.
C. The first and second run returned different results because the query is evaluated each time it is run.
D. The first and second run returned differentresults because the query uses *instead of an explicit column list

**Answer:** B

**Explanation:**
The result cache is enabled at the session level, which means that repeated queries will return cached results if there is no change in the underlying data or session parameters. However, in this case, the result cache is not relevant because the query uses a specific SEED value for sampling, which makes it deterministic. Therefore, both runs will return the same results regardless of caching.

**NEW QUESTION 10**
A Data Engineer has written a stored procedure that will run with caller's rights. The Engineer has granted ROLEA right to use this stored procedure.

What is a characteristic of the stored procedure being called using ROLEA?

A. The stored procedure must run with caller's rights it cannot be converted later to run with owner's rights
B. If the stored procedure accesses an object that ROLEA does not have access to the stored procedure will fail
C. The stored procedure will run in the context (database and schema) where the owner created the stored procedure
D. ROLEA will not be able to see the source code for the stored procedure even though the role has usage privileges on the stored procedure

**Answer:** B

**Explanation:**
A stored procedure that runs with caller's rights executes with the privileges of the role that calls it. Therefore, if the stored procedure accesses an object that ROLEA does not have access to, such as a table or a view, the stored procedure will fail with an insufficient privileges error. The other options are not correct because:
? A stored procedure can be converted from caller's rights to owner's rights by using the ALTER PROCEDURE command with the EXECUTE AS OWNER option.
? A stored procedure that runs with caller's rights executes in the context (database and schema) of the caller, not the owner.
? ROLEA will be able to see the source code for the stored procedure by using the GET_DDL function or the DESCRIBE command, as long as it has usage privileges on the stored procedure.

**NEW QUESTION 10**
Which stages support external tables?

A. Internal stages only; within a single Snowflake account
B. internal stages only from any Snowflake account in the organization
C. External stages only from any region, and any cloud provider
D. External stages only, only on the same region and cloud provider as the Snowflake account

**Answer:** C

**Explanation:**
External stages only from any region, and any cloud provider support external tables. External tables are virtual tables that can query data from files stored in external stages without loading them into Snowflake tables. External stages are references to locations outside of Snowflake, such as Amazon S3 buckets, Azure Blob Storage containers, or Google Cloud Storage buckets. External stages can be created from any region and any cloud provider, as long as they have a valid URL and credentials. The other options are incorrect because internal stages do notsupport external tables. Internal stages are locations within Snowflake that can store files for loading or unloading data. Internal stages can be user stages, table stages, or named stages.

**NEW QUESTION 15**
A Data Engineer wants to centralize grant management to maximize security. A user needs ownership on a table m a new schema However, this user should not have the ability to make grant decisions
What is the correct way to do this?

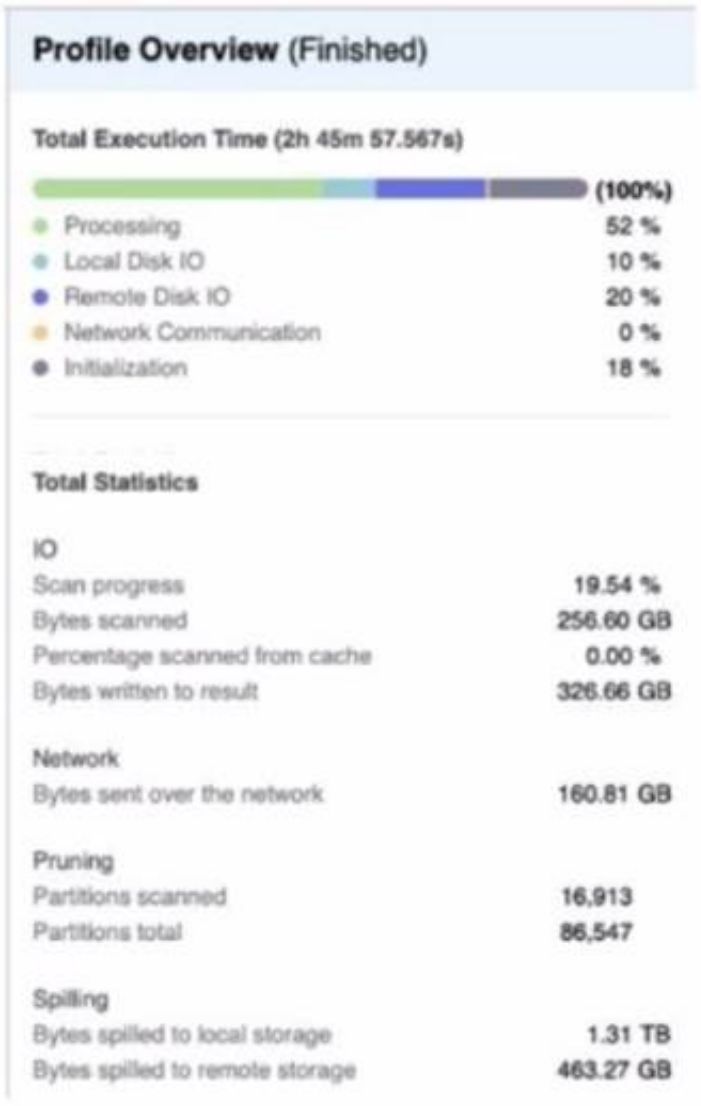
A. Grant ownership to the user on the table
B. Revoke grant decisions from the user on the table
C. Revoke grant decisions from the user on the schema.
D. Add the with managed access parameter on the schema

**Answer:** D

**Explanation:**
The with managed access parameter on the schema enables the schema owner to control the grant and revoke privileges on the objects within the schema. This way, the user who owns the table cannot make grant decisions, but only the schema owner can. This is the best way to centralize grant management and maximize security.

**NEW QUESTION 20**
A Data Engineer is evaluating the performance of a query in a development environment.

```
select *
from
      sample_data.tpcds_sf10tcl.store_sales,
      order by ss_item_sk;
```




Based on the Query Profile what are some performance tuning options the Engineer can use? (Select TWO)

A. Add a LIMIT to the ORDER BY If possible
B. Use a multi-cluster virtual warehouse with the scaling policy set to standard
C. Move the query to a larger virtual warehouse
D. Create indexes to ensure sorted access to data
E. Increase the max cluster count

**Answer:** AC

**Explanation:**
The performance tuning options that the Engineer can use based on the Query Profile are:
? Add a LIMIT to the ORDER BY If possible: This option will improve performance by reducing the amount of data that needs to be sorted and returned by the query. The ORDER BY clause requires sorting all rows in the input before returning them, which can be expensive and time-consuming. By adding a LIMIT clause, the query can return only a subset of rows that satisfy the order criteria, which can reduce sorting time and network transfer time.
? Create indexes to ensure sorted access to data: This option will improve performance by reducing the amount of data that needs to be scanned and filtered by the query. The query contains several predicates on different columns, such as o_orderdate, o_orderpriority, l_shipmode, etc. By creating indexes on these columns, the query can leverage sorted access to data and prune unnecessary micro-partitions or rows that do not match the predicates. This can reduce IO time and processing time.
The other options are not optimal because:
? Use a multi-cluster virtual warehouse with the scaling policy set to standard: This option will not improve performance, as the query is already using a multi-cluster virtual warehouse with the scaling policy set to standard. The Query Profile shows that the query is using a 2XL warehouse with 4 clusters and a standard scaling policy, which means that the warehouse can automatically scale up or down based on the load. Changing the warehouse size or the number of clusters will not affect the performance of this query, as it is already using the optimal resources.
? Increase the max cluster count: This option will not improve performance, as the query is not limited by the max cluster count. The max cluster count is a parameter that specifies the maximum number of clusters that a multi-cluster virtual warehouse can scale up to. The Query Profile shows that the query is using a 2XL warehouse with 4 clusters and a standard scaling policy, which means that the warehouse can automatically scale up or down based on theload. The default max cluster count for a 2XL warehouse is 10, which means that the warehouse can scale up to 10 clusters if needed. However, the query does not need more than 4 clusters, as it is not CPU-bound or memory-bound. Increasing the max cluster count will not affect the performance of this query, as it will not use more clusters than necessary.

**NEW QUESTION 22**
A CSV file around 1 TB in size is generated daily on an on-premise server A corresponding table. Internal stage, and file format have already been created in Snowflake to facilitate the data loading process
How can the process of bringing the CSV file into Snowflake be automated using the LEAST amount of operational overhead?

A. Create a task in Snowflake that executes once a day and runs a copy into statement that references the internal stage The internal stage will read the files directly from the on-premise server and copy the newest file into the table from the on-premise server to the Snowflake table
B. On the on-premise server schedule a SQL file to run using SnowSQL that executes a PUT to push a specific file to the internal stage Create a task that executes once a day m Snowflake and runs a OOPY WTO statement that references the internal stage Schedule the task to start after the file lands in the internal stage
C. On the on-premise server schedule a SQL file to run using SnowSQL that executes a PUT to push a specific file to the internal stag

D. Create a pipe that runs a copy into statement that references the internal stage Snowpipe auto-ingest will automatically load the file from the internal stage when the new file lands in the internal stage.
E. On the on premise server schedule a Python file that uses the Snowpark Python library.The Python script will read the CSV data into a DataFrame and generate an insert into statement that will directly load into the table The script will bypass the need to move a file into an internal stage

**Answer:** C

**Explanation:**
This option is the best way to automate the process of bringing the CSV file into Snowflake with the least amount of operational overhead. SnowSQL is a command- line tool that can be used to execute SQL statements and scripts on Snowflake. By scheduling a SQL file that executes a PUT command, the CSV file can be pushed from the on-premise server to the internal stage in Snowflake. Then, by creating a pipe that runs a COPY INTO statement that references the internal stage, Snowpipe can automatically load the file from the internal stage into the table when it detects a new file in the stage. This way, there is no need to manually start or monitor a virtual warehouse or task.

**NEW QUESTION 27**
A Data Engineer has developed a dashboard that will issue the same SQL select clause to Snowflake every 12 hours.
---will Snowflake use the persisted query results from the result cache provided that the underlying data has not changed^

A. 12 hours
B. 24 hours
C. 14 days
D. 31 days

**Answer:** C

**Explanation:**
Snowflake uses the result cache to store the results of queries that have been executed recently. The result cache is maintained at the account level and is shared across all sessions and users. The result cache is invalidated when any changes are made to the tables or views referenced by the query. Snowflake also has a retention policy for the result cache, which determines how long the results are kept in the cache before they are purged. The default retention period for the result cache is 24 hours, but it can be changed at the account, user, or session level. However, there is a maximum retention period of 14 days for the result cache, which cannot be exceeded. Therefore, if the underlying data has not changed, Snowflake will use the persisted query results from the result cache for up to 14 days.

**NEW QUESTION 29**
What is the purpose of the BUILD_FILE_URL function in Snowflake?

A. It generates an encrypted URL foe accessing a file in a stage.
B. It generates a staged URL for accessing a file in a stage.
C. It generates a permanent URL for accessing files in a stage.
D. It generates a temporary URL for accessing a file in a stage.

**Answer:** B

**Explanation:**
The BUILD_FILE_URL function in Snowflake generates a temporary URL for accessing a file in a stage. The function takes two arguments: the stage name and the file path. The generated URL is valid for 24 hours and can be used to download or view the file contents. The other options are incorrect because they do not describe the purpose of the BUILD_FILE_URL function.

**NEW QUESTION 32**
What are characteristics of Snowpark Python packages? (Select THREE).
Third-party packages can be registered as a dependency to the Snowpark session using the session, import () method.

A. Python packages can access any external endpoints
B. Python packages can only be loaded in a local environment
C. Third-party supported Python packages are locked down to prevent hitting
D. The SQL command DESCRIBE FUNCTION will list the imported Python packages of the Python User-Defined Function (UDF).
E. Querying information schema .packages will provide a list of supported Python packages and versions

**Answer:** ADE

**Explanation:**
The characteristics of Snowpark Python packages are:
? Third-party packages can be registered as a dependency to the Snowpark session using the session.import() method.
? The SQL command DESCRIBE FUNCTION will list the imported Python packages of the Python User-Defined Function (UDF).
? Querying information_schema.packages will provide a list of supported Python packages and versions.
These characteristics indicate how Snowpark Python packages can be imported, inspected, and verified in Snowflake. The other options are not characteristics of Snowpark Python packages. Option B is incorrect because Python packages can be loaded in both local and remote environments using Snowpark. Option C is incorrect because third-party supported Python packages are not locked down to prevent hitting external endpoints, but rather restricted by network policies and security settings.

**NEW QUESTION 34**
A Data Engineer is building a set of reporting tables to analyze consumer requests by region for each of the Data Exchange offerings annually, as well as click-through rates for each listing
Which views are needed MINIMALLY as data sources?

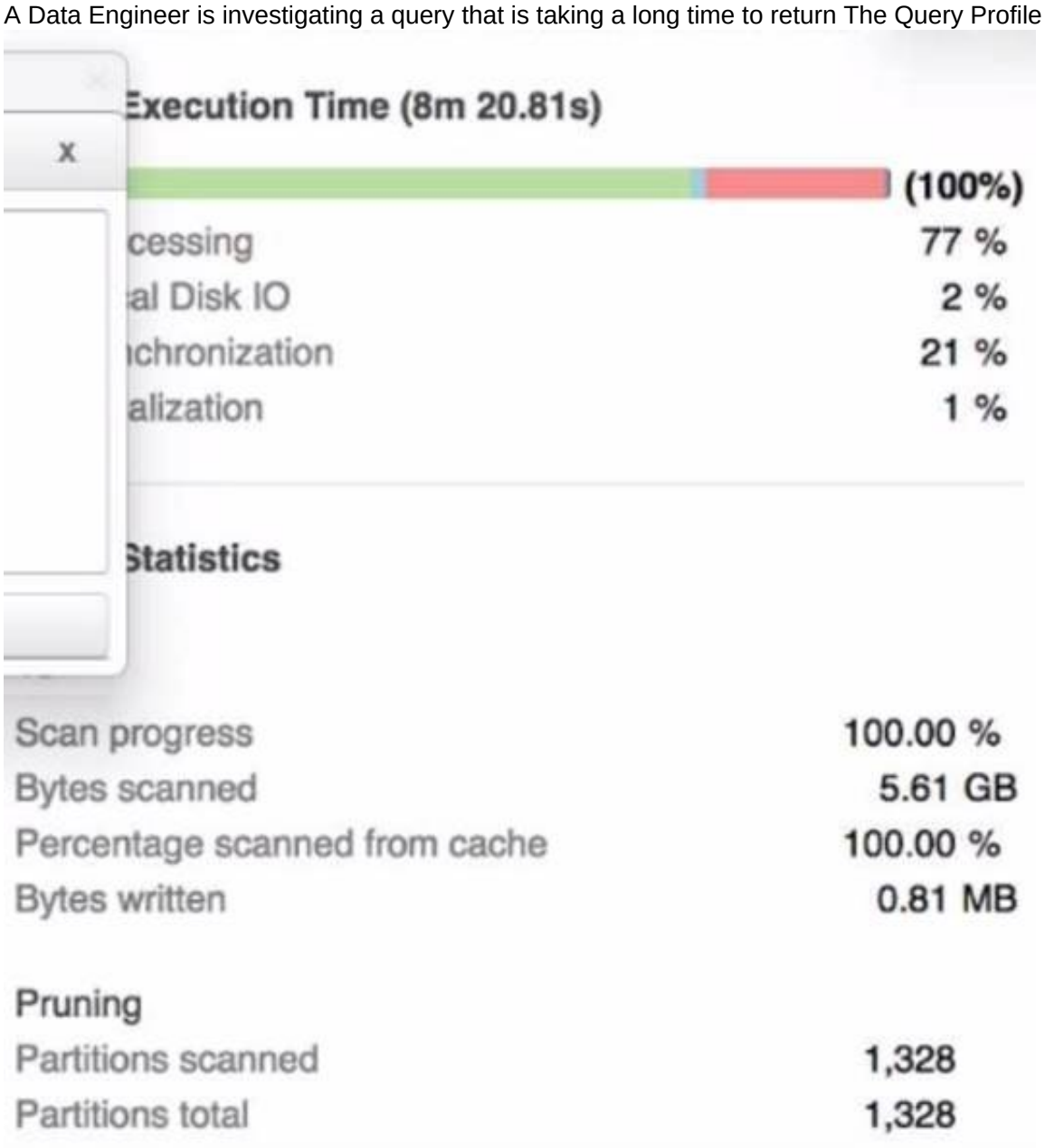
A. SNOWFLAKE- DATA_SHARING_USAGE - LISTING_EVENTS_BAILY
B. SNOWFLAKE.DATA_SHARING_USAGE.LISTING_CONSOKE>TION_DAILY
C. SNOWFLAK
D. DATA_SHARING_USAG

E. LISTING_TELEMETRY_DAILy
F. SNOWFLAKE.ACCOUNT_USAGE.DATA _TRANSFER_HISTORY

**Answer:** B

**Explanation:**
The SNOWFLAKE.DATA SHARING _USAGE.LISTING_CONSOKE>TION_DAILY view provides information about consumer requests by region for each of the Data Exchange offeringsannually, as well as click- through rates for each listing. This view is the minimal data source needed for building the reporting tables. The other views are not relevant for this use case.

**NEW QUESTION 39**
A Data Engineer is investigating a query that is taking a long time to return The Query Profile shows the following:




What step should the Engineer take to increase the query performance?

A. Add additional virtual warehouses.
B. increasethe size of the virtual warehouse.
C. Rewrite the query using Common Table Expressions (CTEs)
D. Change the order of the joins and start with smaller tables first

**Answer:** B

**Explanation:**
The step that the Engineer should take to increase the query performance is to increase the size of the virtual warehouse. The Query Profile shows that most of the time was spent on local disk IO, which indicates that the query was reading a lot of data from disk rather than from cache. This could be due to a large amount of data being scanned or a low cache hit ratio. Increasing the size of the virtual warehouse will increase the amount of memory and cache available for the query, which could reduce the disk IO time and improve the query performance. The other options are not likely to increase the query performance significantly. Option A, adding additional virtual warehouses, will not help
unless they are used in a multi-cluster warehouse configuration or for concurrent queries. Option C, rewriting the query using Common Table Expressions (CTEs), will not affect the amount of data scanned or cached by the query. Option D, changing the order of the joins and starting with smaller tables first, will not reduce the disk IO time unless it also reduces the amount of data scanned or cached by the query.

**NEW QUESTION 40**
A Data Engineer is trying to load the following rows from a CSV file into a table in Snowflake with the following structure:

```
ERID,ADDRESS,REGISTERDT
30 Ford Walk, Dante, Rhode Island, 366",2014-02-08
14 Monroe Street, Kersey, Nevada, 6384",2021-04-19
33 Gate Ave, Edgewater, New York, 1757",2020-07-03
```

| | type |
|---|---|
| MERID | NUMBER(38,0) |
| SS | VARCHAR(255) |
| ERDT | DATE |

....engineer is using the following COPY INTO statement:

```
copy into stgCustomer
from @csv_stage/address.csv.gz
file_format = (type = CSV skip_header = 1);
```

However, the following error is received.

```
Number of columns in file (6) does not match that of the corresponding table (3), use file format option
error_on_column_count_mismatch=false to ignore this error File 'address.csv.gz', line 3, character 1 Row 1 starts at line 2, column
"STGCUSTOMER"[6] If you would like to continue loading when an error is encountered, use other values such as 'SKIP_FILE' or
'CONTINUE' for the ON_ERROR option.
```

Which file format option should be used to resolve the error and successfully load all the data into the table?

A. ESC&PE_UNENGLO9ED_FIELD = '\\'
B. ERROR_ON_COLUMN_COUKT_MISMATCH = FALSE
C. FIELD_DELIMITER = ","
D. FIELD OPTIONALLY ENCLOSED BY = " "

**Answer:** D

**Explanation:**
The file format option that should be used to resolve the error and successfully load all the data into the table is FIELD_OPTIONALLY_ENCLOSED_BY = ‘”’. This option specifies that fields in the file may be enclosed by double quotes, which allows for fields that contain commas or newlines within them. For example, in row 3 of the file, there is a field that contains a comma within double quotes: “Smith Jr., John”. Without specifying this option, Snowflake will treat this field as two separate fields and cause an error due to column count mismatch. By specifying this option, Snowflake will treat this field as one field and load it correctly into the table.

**NEW QUESTION 45**
A table is loaded using Snowpipe and truncated afterwards Later, a Data Engineer finds that the table needs to be reloaded but the metadata of the pipe will not allow the same files to be loaded again.
How can this issue be solved using the LEAST amount of operational overhead?

A. Wait until the metadata expires and then reload the file using Snowpipe
B. Modify the file by adding a blank row to the bottom and re-stage the file
C. Set the FORCE=TRUE option in the Snowpipe COPY INTO command
D. Recreate the pipe by using the create or replace pipe command

**Answer:** C

**Explanation:**
The FORCE=TRUE option in the Snowpipe COPY INTO command allows Snowpipe to load files that have already been loaded before, regardless of the metadata. This is the easiest way to reload the same files without modifying them or recreating the pipe.

**NEW QUESTION 50**
A Data Engineer is working on a Snowflake deployment in AWS eu-west-1 (Ireland). The Engineer is planning to load data from staged files into target tables using the copy into command
Which sources are valid? (Select THREE)

A. Internal stage on GCP us-central1 (Iowa)
B. Internal stage on AWS eu-central-1 (Frankfurt)
C. External stage on GCP us-central1 (Iowa)
D. External stage in an Amazon S3 bucket on AWS eu-west-1 (Ireland)
E. External stage in an Amazon S3 bucket on AWS eu-central 1 (Frankfurt)
F. SSO attached to an Amazon EC2 instance on AWS eu-west-1 (Ireland)

**Answer:** CDE

**Explanation:**
The valid sources for loading data from staged files into target tables using the copy into command are:
? External stage on GCP us-central1 (Iowa): This is a valid source because Snowflake supports cross-cloud data loading from external stages on different cloud platforms and regions than the Snowflake deployment.
? External stage in an Amazon S3 bucket on AWS eu-west-1 (Ireland): This is a valid source because Snowflake supports data loading from external stages on the same cloud platform and region as the Snowflake deployment.
? External stage in an Amazon S3 bucket on AWS eu-central 1 (Frankfurt): This is a valid source because Snowflake supports cross-region data loading from external stages on different regions than the Snowflake deployment within the same cloud platform. The invalid sources are:
? Internal stage on GCP us-central1 (Iowa): This is an invalid source because internal stages are always located on the same cloud platform and region as the Snowflake deployment. Therefore, an internal stage on GCP us-central1 (Iowa) cannot be used for a Snowflake deployment on AWS eu-west-1 (Ireland).

? Internal stage on AWS eu-central-1 (Frankfurt): This is an invalid source because internal stages are always located on the same region as the Snowflake deployment. Therefore, an internal stage on AWS eu-central-1 (Frankfurt) cannot be used for a Snowflake deployment on AWS eu-west-1 (Ireland).
? SSO attached to an Amazon EC2 instance on AWS eu-west-1 (Ireland): This is an invalid source because SSO stands for Single Sign-On, which is a security integration feature in Snowflake, not a data staging option.

**NEW QUESTION 52**
A secure function returns data coming through an inbound share
What will happen if a Data Engineer tries to assign usage privileges on this function to an outbound share?

A. An error will be returned because the Engineer cannot share data that has already been shared
B. An error will be returned because only views and secure stored procedures can be shared
C. An error will be returned because only secure functions can be shared with inboundshares
D. The Engineer will be able to share the secure function with other accounts

**Answer:** A

**Explanation:**
An error will be returned because the Engineer cannot share data that has already been shared. A secure function is a Snowflake function that can access data from an inbound share, which is a share that is created by another account and consumed by the current account. A secure function can only be shared with an inbound share, not an outbound share, which is a share that is created by the current account and shared with other accounts. This is to prevent data leakage or unauthorized access to the data from the inbound share.

**NEW QUESTION 53**
Which functions will compute a 'fingerprint' over an entire table, query result, or window to quickly detect changes to table contents or query results? (Select TWO).

A. HASH (*)
B. HASH_AGG(*)
C. HASH_AGG(<expr>, <expr>)
D. HASH_AGG_COMPARE (*)
E. HASH COMPARE(*)

**Answer:** BC

**Explanation:**
The functions that will compute a ‘fingerprint’ over an entire table, query result, or window to quickly detect changes to table contents or query results are:
? HASH_AGG(*): This function computes a hash value over all columns and rows in
a table, query result, or window. The function returns a single value for each group defined by a GROUP BY clause, or a single value for the entire input if no GROUP BY clause is specified.
? HASH_AGG(<expr>, <expr>): This function computes a hash value over two
expressions in a table, query result, or window. The function returns a single value for each group defined by a GROUP BY clause, or a single value for the entire input if no GROUP BY clause is specified. The other functions are not correct because:
? HASH (*): This function computes a hash value over all columns in a single row.
The function returns one value per row, not one value per table, query result, or window.
? HASH_AGG_COMPARE (): This function compares two hash values computed by
HASH_AGG() over two tables or query results and returns true if they are equal or false if they are different. The function does not compute a hash value itself, but rather compares two existing hash values.
? HASH COMPARE(): This function compares two hash values computed by
HASH() over two rows and returns true if they are equal or false if they are different. The function does not compute a hash value itself, but rather compares two existing hash values.

**NEW QUESTION 57**
A Data Engineer needs to know the details regarding the micro-partition layout for a table named invoice using a built-in function.
Which query will provide this information?

A. SELECT SYSTEM$CLUSTERING_INTFORMATICII (‘Invoice' ) ;
B. SELECT $CLUSTERXNG_INFQRMATION ('Invoice')'
C. CALL SYSTEM$CLUSTERING_INFORMATION (‘Invoice’);
D. CALL $CLUSTERINS_INFORMATION('Invoice’);

**Answer:** A

**Explanation:**
The query that will provide information about the micro-partition layout for a table named invoice using a built-in function is SELECT SYSTEM$CLUSTERING_INFORMATION(‘Invoice’);. The
SYSTEM$CLUSTERING_INFORMATION function returns information about the clustering status of a table, such as the clustering key, the clustering depth, the clustering ratio, the partition count, etc. The function takes one argument: the table name in a qualified or unqualified form. In this case, the table name is Invoice and it is unqualified, which means that it will use the current database and schema as the context. The other options are incorrect because they do not use a valid built-in function for providing information about the micro-partition layout for a table. Option B is incorrect because it uses $CLUSTERING_INFORMATION instead of SYSTEM$CLUSTERING_INFORMATION,
which is not a valid function name. Option C is incorrect because it uses CALL instead of SELECT, which is not a valid way to invoke a table function. Option D is incorrect because it uses CALL instead of SELECT and $CLUSTERING_INFORMATION instead of SYSTEM$CLUSTERING_INFORMATION, which are both invalid.

**NEW QUESTION 58**
Which Snowflake objects does the Snowflake Kafka connector use? (Select THREE).

A. Pipe
B. Serverless task
C. Internal user stage

D. Internal table stage
E. Internal named stage
F. Storage integration

**Answer:** ADE

**Explanation:**
The Snowflake Kafka connector uses three Snowflake objects: pipe, internal table stage, and internal named stage. The pipe object is used to load data from an external stage into a Snowflake table using COPY statements. The internal table stage is used to store files that are loaded from Kafka topics into Snowflake using PUT commands. The internal named stage is used to store files that are rejected by the COPY statements due to errors or invalid data. The other options are not objects that are used by the Snowflake Kafka connector. Option B, serverless task, is an object that can execute SQL statements on a schedule without requiring a warehouse. Option C, internal user stage, is an object that can store files for a specific user in Snowflake using PUT commands. Option F, storage integration, is an object that can enable secure access to external cloud storage services without exposing credentials.

**NEW QUESTION 63**
A Data Engineer executes a complex query and wants to make use of Snowflake s query results caching capabilities to reuse the results.
Which conditions must be met? (Select THREE).

A. The results must be reused within 72 hours.
B. The query must be executed using the same virtual warehouse.
C. The USED_CACHED_RESULT parameter must be included in the query.
D. The table structure contributing to the query result cannot have changed
E. The new query must have the same syntax as the previously executed query.
F. The micro-partitions cannot have changed due to changes to other data in the table

**Answer:** ADE

**Explanation:**
Snowflake’s query results caching capabilities allow users to reuse the results of previously executed queries without re-executing them. For this to happen, the following conditions must be met:
? The results must be reused within 24 hours (not 72 hours), which is the default
time-to-live (TTL) for cached results.
? The query must be executed using any virtual warehouse (not necessarily the same one), as long as it is in the same region and account as the original query.
? The USED_CACHED_RESULT parameter does not need to be included in the query, as it is enabled by default at the account level. However, it can be disabled or overridden at the session or statement level.
? The table structure contributing to the query result cannot have changed, such as adding or dropping columns, changing data types, or altering constraints.
? The new query must have the same syntax as the previously executed query, including whitespace and case sensitivity.
? The micro-partitions cannot have changed due to changes to other data in the table, such as inserting, updating, deleting, or merging rows.

**NEW QUESTION 68**
.....

# Relate Links

**100% Pass Your DEA-C01 Exam with Exambible Prep Materials**

https://www.exambible.com/DEA-C01-exam/

# Contact us

**We are proud of our high-quality customer service, which serves you around the clock 24/7.**

**Viste -** https://www.exambible.com/