Snowflake

Exam Questions DEA-C01

SnowPro Advanced: Data Engineer Certification Exam

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
A Data Engineer defines the following masking policy:

```
arrent_role() IN ('ADMIN') THEN val
******'
```

….
must be applied to the full_name column in the customer table:

```
TABLE customer(

ame VARCHAR,
ame VARCHAR AS CONCAT(first_name, ' ', last_name)
```

Which query will apply the masking policy on the full_name column?

A. ALTER TABLE customer MODIFY COLUMN full_name Set MASKING POLICY name_policy;
B. ALTER TABLE customer MODIFY COLUMN full_nam ADD MASKING POLICYname_poiicy;
C. ALTER TABLE customer MODIFY COLUMN first_nane SET MASKING POLICY name_policy; lasT_name SET MASKING POLICY name_pclicy;
D. ALTER TABLE customer MODIFY COLUMN first_name ADD MASKING POLICY name_policy,

Answer: A

Explanation:
The query that will apply the masking policy on the full_name column is ALTER TABLE customer MODIFY COLUMN full_name SET MASKING POLICY name_policy;. This query will modify the full_name column and associate it with the name_policy masking policy, which will mask the first and last names of the customers with asterisks. The other options are incorrect because they do not follow the correct syntax for applying a masking policy on a column. Option B is incorrect because it uses ADD instead of SET, which is not a valid keyword for modifying a column. Option C is incorrect because it tries to apply the masking policy on two columns, first_name and last_name, which are not part of the table structure. Option D is incorrect because it uses commas instead of dots to separate the database, schema, and table names

NEW QUESTION 3
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 4
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 5
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 6
A new customer table is created by a data pipeline in a Snowflake schema where MANAGED ACCESSenabled.
…. Can gran access to the CUSTOMER table? (Select THREE.)

A. The role that owns the schema
B. The role that owns the database
C. The role that owns the customer table
D. The SYSADMIN role
E. The SECURITYADMIN role
F. The USERADMIN role with the manage grants privilege

Answer: ABE

Explanation:
The roles that can grant access to the CUSTOMER table are the role that owns the schema, the role that owns the database, and the SECURITYADMIN role. These roles have the ownership or the manage grants privilege on the schema or the database level, which allows them to grant access to any object within them. The other options are incorrect because they do not have the necessary privilege to grant access to the CUSTOMER table. Option C is incorrect because the role that owns the customer table cannot grant access to itself or to other roles. Option D is incorrect because the SYSADMIN role does not have the manage grants privilege by default and cannot grant access to objects that it does not own. Option F is incorrect because the USERADMIN role with the manage grants privilege can only grant access to users and roles, not to tables.

NEW QUESTION 7
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 8
Company A and Company B both have Snowflake accounts. Company A's account is hosted on a different cloud provider and region than Company B's account Companies A and B are not in the same Snowflake organization.
How can Company A share data with Company B? (Select TWO).

A. Create a share within Company A's account and add Company B's account as a recipient of that share
B. Create a share within Company A's account, and create a reader account that is a recipient of the share Grant Company B access to the reader account
C. Use database replication to replicate Company A's data into Company B's account Create a share within Company B's account and grant users within Company B's account access to the share
D. Create a new account within Company A's organization in the same cloud provider and region as Company B's account Use database replication to replicate CompanyA's data to the new account Create a share within the new account and add Company B's account as a recipient of that share
E. Create a separate database within Company A's account to contain only those data setsthey wish to share with Company B Create a share within Company A's account and add all the objects within this separate database to the share Add Company B's account as a recipient of the share

Answer: AE

Explanation:
The ways that Company A can share data with Company B are:
? Create a share within Company A’s account and add Company B’s account as a recipient of that share: This is a valid way to share data between different accounts on different cloud platforms and regions. Snowflake supports cross-cloud and cross-region data sharing, which allows users to create shares and grant access to other accounts regardless of their cloud platform or region. However, this option may incur additional costs for network transfer and storage replication.
? Create a separate database within Company A’s account to contain only those data sets they wish to share with Company B Create a share within Company A’saccount and add all the objects within this separate database to the share Add Company B’s account as a recipient of the share: This is also a valid way to share data between different accounts on different cloud platforms and regions. This option is similar to the previous one, except that it uses a separate database to isolate the data sets that need to be shared. This can improve security and manageability of the shared data. The other options are not valid because:
? Create a share within Company A’s account, and create a reader account that is a recipient of the share Grant Company B access to the reader account: This option is not valid because reader accounts are not supported for cross-cloud or cross- region data sharing. Reader accounts are Snowflake accounts that can only consume data from shares created by their provider account. Reader accounts must be on the same cloud platform and region as their provider account.
? Use database replication to replicate Company A’s data into Company B’s account Create a share within Company B’s account and grant users within Company B’s account access to the share: This option is not valid because database replication cannot be used for cross-cloud or cross-region data sharing. Database replication is a feature in Snowflake that allows users to copy databases across accounts within the same cloud platform and region. Database replication cannot copy databases across different cloud platforms or regions.
? Create a new account within Company A’s organization in the same cloud provider and region as Company B’s account Use database replication to replicate Company A’s data to the new account Create a share within the new account and add Company B’s account as a recipient of that share: This option is not valid because it involves creating a new account within Company A’s organization, which may not be feasible or desirable for Company A. Moreover, this option is unnecessary, as Company A can directly share data with Company B without creating an intermediate account.

NEW QUESTION 9
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 10

The JSON below is stored in a variant column named v in a table named jCustRaw:

```
id": "6282638561cf48544e2ef7e9",
ompany": "FLYBOYZ",
sActive": true,
ame": "Dean Head",
eamMembers": [
  {
      "age": 29,
      "eyeColor": "green",
      "name": "Dominique Grimes",
      "registered": "2017-02-19T06:12:36 +06:00"
  },
  {
      "age": 39,
      "eyeColor": "green",
      "name": "Pearl Dunlap",
      "registered": "2018-05-12T09:21:42 +05:00"
  },
  {
      "age": 22,
      "eyeColor": "blue",
      "name": "Cardenas Warren",
      "registered": "2019-04-08T01:24:29 +05:00"

      }
  ]
}
```

Which query will return one row per team member (stored in the teamMembers array) along all of the attributes of each team member?

A)

```
select
     t2.name AS memberName
     ,t2.registered AS registeredDttm
     ,t2.age AS age
     ,t2.eyeColor AS eyeColor
from jCustRaw t1
  ateral flatten(v) t2
se
                    Name
  lect
     t2.value:name::va
     ,t2.value:registe
     ,t2.value:age::nu
fr   ,t2.value:eyeColo
,l
om jCustRaw t1
  ateral flatten(input
```

C)

```
select
     v:teamMembers.name::varchar AS memberName
     ,v:teamMembers.registered::timestamp AS
     registeredDttm
     ,v:teamMembers.age::number AS age
     ,v:teamMembers.eyeColor::varchar AS eyeColor
from jCustRaw;
```

D)

```
select
     v:teamMembers[0].name::varchar AS memberName
     ,v:teamMembers[0].registered::timestamp AS registeredDttm
     ,v:teamMembers[0].age::number AS age
     ,v:teamMembers[0].eyeColor::varchar AS eyeColor
from jCustRaw;
```

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

Answer: B

NEW QUESTION 10
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 12
Assuming a Data Engineer has all appropriate privileges and context which statements would be used to assess whether the User-Defined Function (UDF), MTBATA3ASZ. SALES .REVENUE_BY_REGION, exists and is secure? (Select TWO)

A. SHOW DS2R FUNCTIONS LIKE 'REVEX'^BYJIESION' IN SCHEMA SALES;
B. SELECT IS_SECURE FROM SNOWFLAK
C. INFCRXATION_SCKZM
D. FUNCTIONS WHERE FUNCTI0N_3CHEMA = 'SALES' AND FUNCTI CN_NAXE =•ftEVEXUE_BY_RKXQH4;
E. SELECT IS_SEC"JRE FROM INFOR>LVTICN_SCHEM
F. FUNCTIONS WHERE FUNCTION_SCHEMA = 'SALES1 AND FUNGTZON_NAME = ' REVENUE_BY_REGION';
G. SHOW EXTERNAL FUNCTIONS LIKE ‘REVENUE_BY_REGION’IB SCHEMA SALES;
H. SHOW SECURE FUNCTIONS LIKE 'REVENUE 3Y REGION' IN SCHEMA SALES;

Answer: AB

Explanation:
The statements that would be used to assess whether the UDF, MTBATA3ASZ. SALES .REVENUE_BY_REGION, exists and is secure are:
? SHOW DS2R FUNCTIONS LIKE ‘REVEX’^BYJIESION’ IN SCHEMA SALES;:
This statement will show information about the UDF, including its name, schema, database, arguments, return type, language, and security option. If the UDF does not exist, the statement will return an empty result set.
? SELECT IS_SECURE FROM SNOWFLAKE. INFCRXATION_SCKZMA.
FUNCTIONS WHERE FUNCTI0N_3CHEMA = ‘SALES’ AND FUNCTI CN_NAXE
= •ftEVEXUE_BY_RKXQH4;: This statement will query the SNOWFLAKE.INFORMATION_SCHEMA.FUNCTIONS view, which contains metadata about the UDFs in the current database. The statement will return the IS_SECURE column, which indicates whether the UDF is secure or not. If the UDF does not exist, the statement will return an empty result set. The other statements are not correct because:
? SELECT IS_SEC"JRE FROM INFOR>LVTICN_SCHEMA. FUNCTIONS WHERE
FUNCTION_SCHEMA = ‘SALES1 AND FUNGTZON_NAME = ’
REVENUE_BY_REGION’;: This statement will query the INFORMATION_SCHEMA.FUNCTIONS view, which contains metadata about the UDFs in the current schema. However, the statement has a typo in the schema name (‘SALES1’ instead of ‘SALES’), which will cause it to fail or return incorrect results.
? SHOW EXTERNAL FUNCTIONS LIKE ‘REVENUE_BY_REGION’ IB SCHEMA
SALES;: This statement will show information about external functions, not UDFs. External functions are Snowflake functions that invoke external services via HTTPS requests and responses. The statement will not return any results for the UDF.
? SHOW SECURE FUNCTIONS LIKE ‘REVENUE 3Y REGION’ IN SCHEMA
SALES;: This statement is invalid because there is no such thing as secure functions in Snowflake. Secure functions are a feature of some other databases, such as PostgreSQL, but not Snowflake. The statement will cause a syntax error.

NEW QUESTION 16
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 18
Assuming that the session parameter USE_CACHED_RESULT is set to false, what are characteristics of Snowflake virtual warehouses in terms of the use of Snowpark?

A. Creating a DataFrame from a table will start a virtual warehouse
B. Creating a DataFrame from a staged file with the read () method will start a virtual warehouse
C. Transforming a DataFrame with methods like replace () will start a virtual warehouse -
D. Calling a Snowpark stored procedure to query the database with session, call () will start a virtual warehouse

Answer: A

Explanation:
Creating a DataFrame from a table will start a virtual warehouse because it requires reading data from Snowflake. The other options will not start a virtual warehouse because they either operate on local data or use an existing session to query Snowflake.

NEW QUESTION 21
Which output is provided by both theSYSTEM$CLUSTERING_DEPTHfunction and theSYSTEM$CLUSTERING_INFORMATIONfunction?

A. average_depth
B. notes
C. average_overlaps
D. total_partition_count

Answer: A

Explanation:
The output that is provided by both the SYSTEM$CLUSTERING_DEPTH function and the SYSTEM$CLUSTERING_INFORMATION function is average_depth. This output indicates the average number of micro-partitions that contain data for a given column value or combinationof column values. The other outputs are not common to both functions. The notes output is only provided by the SYSTEM$CLUSTERING_INFORMATION function and it contains additional information or recommendations about the clustering status of the table. The average_overlaps output is only provided by the SYSTEM$CLUSTERING_DEPTH function and it indicates the average number of micro-partitions that overlap with other micro-partitions for a given column value or combination of column values. The total_partition_count output is only provided by the SYSTEM$CLUSTERING_INFORMATION function and it indicates the total number of micro-partitions in the table.

NEW QUESTION 25
At what isolation level are Snowflake streams?

A. Snapshot
B. Repeatable read
C. Read committed
D. Read uncommitted

Answer: B

Explanation:
The isolation level of Snowflake streams is repeatable read, which means that each transaction sees a consistent snapshot of data that does not change during its execution. Streams use time travel internally to provide this isolation level and ensure that queries on streams return consistent results regardless of concurrent transactions on their source tables.

NEW QUESTION 28
A Data Engineer wants to create a new development database (DEV) as a clone of the permanent production database (PROD) There is a requirement to disable Fail-safe for all tables.
Which command will meet these requirements?

A. CREATE DATABASE DEV CLONE PROD FAIL_SAFE=FALSE;
B. CREATE DATABASE DEV CLONE PROD;
C. CREATE TRANSIENT DATABASE DEV CLONE RPOD
D. CREATE DATABASE DEV CLOSE PRODDATA_RETENTION_TIME_IN_DAYS =0L

Answer: C

Explanation:
This option will meet the requirements of creating a new development database (DEV) as a clone of the permanent production database (PROD) and disabling Fail-safe for all tables. By using the CREATE TRANSIENT DATABASE command, the Data Engineer can create a transient database that does not have Fail-safe enabled by default. Fail-safe is a feature in Snowflake that provides additional protection against data loss by retaining historical data for seven days beyond the time travel retention period. Transient databases do not have Fail-safe enabled, which means that they do not incur additional storage costs for historical data beyond their time travel retention period. By using the CLONE option, the Data Engineer can create an exact copy of the PROD database, including its schemas, tables, views, and other objects.

NEW QUESTION 30
A Data Engineer is writing a Python script using the Snowflake Connector for Python. The Engineer will use the snowflake. Connector.connect function to connect to Snowflake The requirementsare:
*Raise an exception if the specified database schema or warehouse does not exist
*improve download performance
Whichparameters of the connect function should be used? (Select TWO).

A. authenticator
B. arrow_nunber_to_decimal
C. client_prefetch_threads
D. client_session_keep_alivs
E. validate_default_parameters

Answer: CE

Explanation:
The parameters of the connect function that should be used are client_prefetch_threads and validate_default_parameters. The client_prefetch_threads parameter controls the number of threads used to download query results from Snowflake. Increasing this parameter can improve download performance by parallelizing the download process. The validate_default_parameters parameter controls whether an exception should be raised if the specified database, schema, or warehouse does not exist or is not authorized. Setting this parameter to True can help catch errors early and avoid unexpected results.

NEW QUESTION 32
.....

Thank You for Trying Our Product

We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questons and Answers in PDF Format

DEA-C01 Practice Exam Features:

* DEA-C01 Questions and Answers Updated Frequently

* DEA-C01 Practice Questions Verified by Expert Senior Certified Staff

* DEA-C01 Most Realistic Questions that Guarantee you a Pass on Your FirstTry

* DEA-C01 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year