Snowflake

Exam Questions DEA-C01

SnowPro Advanced: Data Engineer Certification Exam

NEW QUESTION 1
Which use case would be BEST suited for the search optimization service?

A. Analysts who need to perform aggregates over high cardinality columns
B. Business users who need fast response times using highly selective filters
C. Data Scientists who seek specific JOIN statements with large volumes of data
D. Data Engineers who create clustered tables with frequent reads against clustering keys

Answer: B

Explanation:
The use case that would be best suited for the search optimization service is business users who need fast response times using highly selective filters. The search optimization service is a feature that enables faster queries on tables with high cardinality columns by creating inverted indexes on those columns. High cardinality columns are columns that have a large number of distinct values, such as customer IDs, product SKUs, or email addresses. Queries that use highly selective filters on high cardinality columns can benefit from the search optimization service because they can quickly locate the relevant rows without scanning the entire table. The other options are not best suited for the search optimization service. Option A is incorrect because analysts who need to perform aggregates over high cardinality columns will not benefit from the search optimization service, as they will still need to scan all the rows that match the filter criteria. Option C is incorrect because data scientists who seek specific JOIN statements with large volumes of data will not benefit from the search optimization service, as they will still need to perform join operations that may involve shuffling or sorting data across nodes. Option D is incorrect because data engineers who create clustered tables with frequent reads against clustering keys will not benefit from the search optimization service, as they already have an efficient way to organize and access data based on clustering keys.

NEW QUESTION 2
Within a Snowflake account permissions have been defined with custom roles and role hierarchies.
To set up column-level masking using a role in the hierarchy of the current user, what command would be used?

A. CORRECT_ROLE
B. IKVOKER_ROLE
C. IS_RCLE_IN_SESSION
D. IS_GRANTED_TO_INVOKER_ROLE

Answer: C

Explanation:
The IS_ROLE_IN_SESSION function is used to set up column-level masking using a role in the hierarchy of the current user. Column-level masking is a feature in Snowflake that allows users to apply dynamic data masking policies to specific columns
based on the roles of the users who access them. The IS_ROLE_IN_SESSION function takes a role name as an argument and returns true if the role is in the current user's session, or false otherwise. The function can be used in a masking policy expression to determine whether to mask or unmask a column value based on the role of the user. For example:
CREATE OR REPLACE MASKING POLICY email_mask AS (val string) RETURNS string -
> CASE WHEN IS_ROLE_IN_SESSION('HR') THEN val ELSE REGEXP_REPLACE(val, '(.).(.@.)', '\1****\2') END;
In this example, the IS_ROLE_IN_SESSION function is used to create a masking policy for an email column. The masking policy returns the original email value if the user has the HR role in their session, or returns a masked email value with asterisks if not.

NEW QUESTION 3
What is a characteristic of the use of binding variables in JavaScript stored procedures in Snowflake?

A. All types of JavaScript variables can be bound
B. All Snowflake first-class objects can be bound
C. Only JavaScript variables of type number, string and sf Date can be bound
D. Users are restricted from binding JavaScript variables because they create SQL injection attack vulnerabilities

Answer: C

Explanation:
A characteristic of the use of binding variables in JavaScript stored procedures in Snowflake is that only JavaScript variables of type number, string and sf Date can be bound. Binding variables are a way to pass values from JavaScript variables to SQL statements within a stored procedure. Binding variables can improve the security and performance of the stored procedure by preventing SQL injection attacks and reducing the parsing overhead. However, not all types of JavaScript variables can be bound. Only the primitive types number and string, and the Snowflake-specific type sf Date, can be bound. The other options are incorrect because they do not describe a characteristic of the use of binding variables in JavaScript stored procedures in Snowflake. Option A is incorrect because authenticator is not a type of JavaScript variable, but a parameter of the snowflake.connector.connect function. Option B is incorrect because arrow_number_to_decimal is not a type of JavaScript variable, but a parameter of the
snowflake.connector.connect function. Option D is incorrect because users are not restricted from binding JavaScript variables, but encouraged to do so.

NEW QUESTION 4
A Data Engineer is building a pipeline to transform a 1 TD tab e by joining it with supplemental tables The Engineer is applying filters and several aggregations leveraging Common TableExpressions (CTEs) using a size Medium virtual warehouse in a single query in Snowflake.
After checking the Query Profile, what is the recommended approach to MAXIMIZE performance of this query if the Profile shows data spillage?

A. Enable clustering on the table
B. Increase the warehouse size
C. Rewrite the query to remove the CTEs.
D. Switch to a multi-cluster virtual warehouse

Answer: B

Explanation:
The recommended approach to maximize performance of this query if the Profile shows data spillage is to increase the warehouse size. Data spillage occurs

when the query requires more memory than the warehouse can provide and has to spill some intermediate results to disk. This can degrade the query performance by increasing the disk IO time. Increasing the warehouse size can increase the amount of memory available for the query and reduce or eliminate data spillage.

NEW QUESTION 5
Which callback function is required within a JavaScript User-Defined Function (UDF) for it to execute successfully?

A. initialize ()
B. processRow ()
C. handler
D. finalize ()

Answer: B

Explanation:
The processRow () callback function is required within a JavaScript UDF for it to execute successfully. This function defines how each row of input data is processed and what output is returned. The other callback functions are optional and can be used for initialization, finalization, or error handling.

NEW QUESTION 6
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 7
A Data Engineer would like to define a file structure for loading and unloading data Where can the file structure be defined? (Select THREE)

A. copy command
B. MERGE command
C. FILE FORMAT Object
D. pipe object
E. stage object
F. INSERT command

Answer: ACE

Explanation:
The places where the file format can be defined are copy command, file format object, and stage object. These places allow specifying or referencing a file format that defines how data files are parsed and loaded into or unloaded from Snowflake tables. A file format can include various options, such as field delimiter, field enclosure, compression type, date format, etc. The other options are not places where the file format can be defined. Option B is incorrect because MERGE command is a SQL command that can merge data from one table into another based on a join condition, but it does not involve loading or unloading data files. Option D is incorrect because pipe object is a Snowflake object that can load data from an external stage into a Snowflake table using COPY statements, but it does not define or reference a file format. Option F is incorrect because INSERT command is a SQL command that can insert data into a Snowflake table from literal values or subqueries, but it does not involve loading or unloading data files.

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 is working on a continuous data pipeline which receives data from Amazon Kinesis Firehose and loads the data into a staging table which will later be used in the data transformation process The average file size is 300-500 MB.
The Engineer needs to ensure that Snowpipe is performant while minimizing costs. How can this be achieved?

A. Increase the size of the virtual warehouse used by Snowpipe.
B. Split the files before loading them andset the SIZE_LIMIT option to 250 MB.
C. Change the file compression size and increase the frequency of the Snowpipe loads
D. Decrease the buffer size to trigger delivery of files sized between 100 to 250 MB in Kinesis Firehose

Answer: B

Explanation:
This option is the best way to ensure that Snowpipe is performant while minimizing costs. By splitting the files before loading them, the Data Engineer can reduce the size of each file and increase the parallelism of loading. By setting the SIZE_LIMIT option to 250 MB, the Data Engineer can specify the maximum file size that can be loaded bySnowpipe, which can prevent performance degradation or errors due to large files. The other options are not optimal because:
? Increasing the size of the virtual warehouse used by Snowpipe will increase the
performance but also increase the costs, as larger warehouses consume more credits per hour.
? Changing the file compression size and increasing the frequency of the Snowpipe
loads will not have much impact on performance or costs, as Snowpipe already supports various compression formats and automatically loads files as soon as they are detected in the stage.
? Decreasing the buffer size to trigger delivery of files sized between 100 to 250 MB
in Kinesis Firehose will not affect Snowpipe performance or costs, as Snowpipe does not depend on Kinesis Firehose buffer size but rather on its own SIZE_LIMIT option.

NEW QUESTION 10
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 stream called TRANSACTIONS_STM is created on top of a transactions table in a continuous pipeline running in Snowflake. After a couple of months, the TRANSACTIONS table is renamed transactiok3_raw to comply with new naming standards
What will happen to the TRANSACTIONS _STM object?

A. TRANSACTIONS _STMwill keep working as expected
B. TRANSACTIONS _STMwill be stale and will need to be re-created
C. TRANSACTIONS _STMwill be automatically renamedTRANSACTIONS _RAW_STM.
D. Reading from the traksactioks_3T>: stream will succeed for some time after the expected STALE_TIME.

Answer: B

Explanation:
A stream is a Snowflake object that records the history of changes made to a table. A stream is associated with a specific table at the time of creation, and it cannot be altered to point to a different table later. Therefore, if the source table is renamed, the stream will become stale and will need to be re-created with the new table name. The other options are not correct because:
? TRANSACTIONS _STM will not keep working as expected, as it will lose track of
the changes made to the renamed table.
? TRANSACTIONS _STM will not be automatically renamed TRANSACTIONS

_RAW_STM, as streams do not inherit the name changes of their source tables.
? Reading from the transactions_stm stream will not succeed for some time after the expected STALE_TIME, as streams do not have a STALE_TIME property.

NEW QUESTION 15
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).

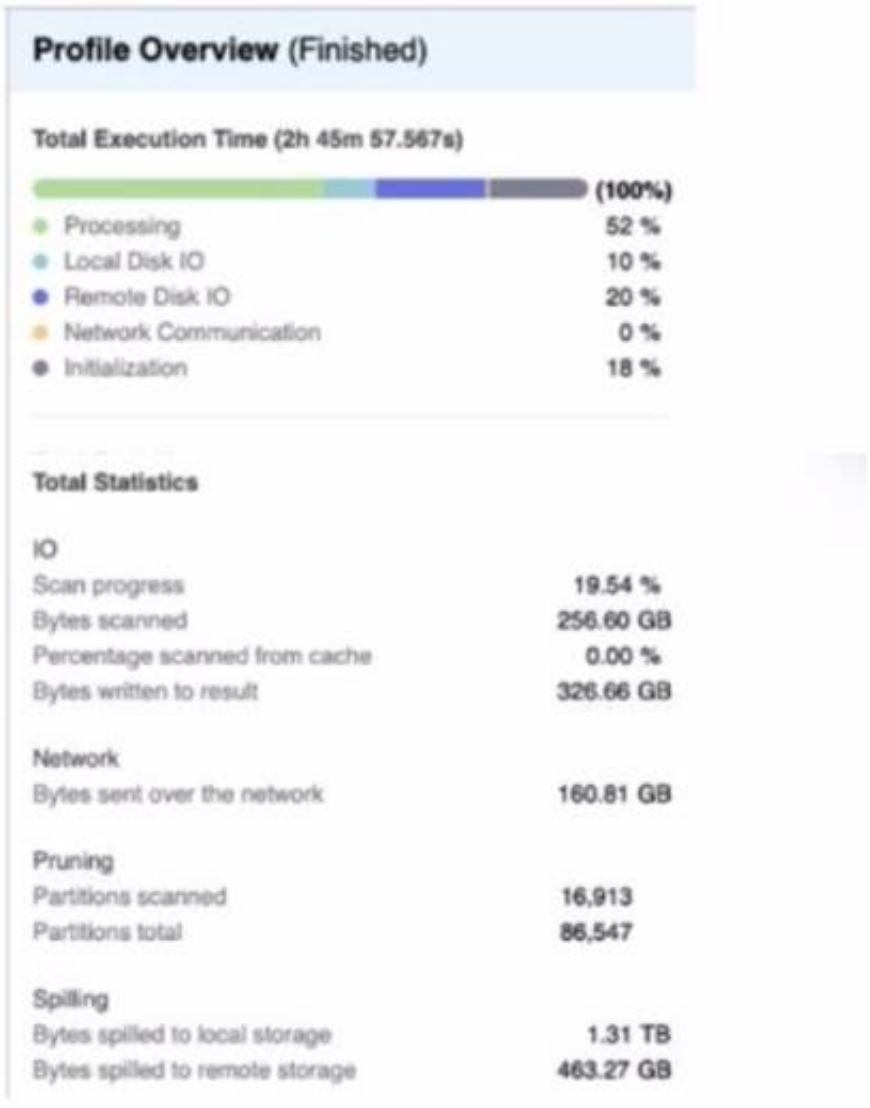
A. Create a share within Company A's account and add Company B's account as a recipient of that share
B. Create a share within Company A's account, and create a reader account that is a recipient of the share Grant Company B access to the reader account
C. Use database replication to replicate Company 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 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 21
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 23
A company built a sales reporting system with Python, connecting to Snowflake using the Python Connector. Based on the user's selections, the system generates the SQL queries needed to fetch the data for the report First it gets the customers that meet the given query parameters (on average 1000 customer records for each report run) and then it loops the customer records sequentially Inside that loop it runs the generated SQL clause for the current customer to get the detailed data for that customer number from the sales data table
When the Data Engineer tested the individual SQL clauses they were fast enough (1 second to get the customers 0 5 second to get the sales data for one customer) but the total runtime of the report is too long
How can this situation be improved?

A. Increase the size of the virtual warehouse
B. Increase the number of maximum clusters of the virtual warehouse
C. Define a clustering key for the sales data table
D. Rewrite the report to eliminate the use of the loop construct

Answer: D

Explanation:
This option is the best way to improve the situation, as using a loop construct to run SQL queries for each customer is very inefficient and slow. Instead, the report should be rewritten to use a single SQL query that joins the customer and sales data tables and applies the query parameters as filters. This way, the report can leverage Snowflake's parallel processing and optimization capabilities and reduce the network overhead and latency.

NEW QUESTION 25
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
```

```
  lect                        Name
      t2.value:name::va        registeredDttm
      ,t2.value:registe
      ,t2.value:age::nu      yeColor
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 29
A database contains a table and a stored procedure defined as.

```
CREATE OR REPLACE TABLE log_table(col1 VARCHAR);

CREATE OR REPLACE PROCEDURE insert_log(input VARCHAR)
RETURNS FLOAT
LANGUAGE JAVASCRIPT
RETURNS NULL ON NULL INPUT
AS
'
var rs = snowflake.execute({sqlText:`INSERT INTO log_table(col1) VALUES(:1);`
,binds:[INPUT]});

return 1;
';
```

The `log_table` is initially empty and a Data Engineer issues the following command:

```
CALL insert_log(NULL::VARCHAR);
```

No other operations are affecting the log_table. What will be the outcome of the procedure call?

A. The log_table contains zero records and the stored procedure returned 1 as a return value
B. The log_table contains one record and the stored procedure returned 1 as a return value
C. The log_table contains one record and the stored procedure returned NULL as a return value
D. The log_table contains zero records and the stored procedure returned NULL as a return value

Answer: B

Explanation:
The stored procedure is defined with a FLOAT return type and a JavaScript language. The body of the stored procedure contains a SQL statement that inserts a row into the log_table with a value of '1' for col1. The body also contains a return statement that returns 1 as a float value. When the stored procedure is called with any VARCHAR parameter, it will execute successfully and insert one record into the log_table and return 1 as a return value. The other options are not correct because:
? The log_table will not be empty after the stored procedure call, as it will contain

one record inserted by the SQL statement.
? The stored procedure will not return NULL as a return value, as it has an explicit return statement that returns 1.

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 33
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 36
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 41
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 44
When would a Data engineer use table with the flatten function instead of the lateral flatten combination?

A. When TABLE with FLATTENrequires another source in the from clause to refer to
B. WhenTABLE with FLATTENrequires no additional source m the from clause to refer to
C. Whenthe LATERALFLATTENcombination requires no other source m the from clause to refer to
D. When table withFLATTENis acting like a sub-query executed for each returned row

Answer: A

Explanation:
The TABLE function with the FLATTEN function is used to flatten semi- structured data, such as JSON or XML, into a relational format. The TABLE function returns a table expression that can be used in the FROM clause of a query. The TABLE function with the FLATTEN function requires another source in the FROM clause to refer to, such as a table, view, or subquery that contains the semi-structured data. For example: SELECT t.value:city::string AS city, f.value AS population FROM cities t, TABLE(FLATTEN(input => t.value:population)) f;
In this example, the TABLE function with the FLATTEN function refers to the cities table in the FROM clause, which contains JSON data in a variant column named value. The FLATTEN function flattens the population array within each JSON object and returns a table expression with two columns: key and value. The query then selects the city and population values from the table expression.

NEW QUESTION 46
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 47
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 48
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 52
A company has an extensive script in Scala that transforms data by leveraging DataFrames. A Data engineer needs to move these transformations to Snowpark. …characteristics of data transformations in Snowpark should be considered to meet this requirement? (Select TWO)

A. It is possible to join multiple tables using DataFrames.
B. Snowpark operations are executed lazily on the server.
C. User-Defined Functions (UDFs) are not pushed down to Snowflake
D. Snowpark requires a separate cluster outside of Snowflake for computations
E. Columns in different DataFrames with the same name should be referred to with squared brackets

Answer: AB

Explanation:
The characteristics of data transformations in Snowpark that should be considered to meet this requirement are:
? It is possible to join multiple tables using DataFrames.
? Snowpark operations are executed lazily on the server.
These characteristics indicate how Snowpark can perform data transformations using DataFrames, which are similar to the ones used in Scala. DataFrames are distributed collections of rows that can be manipulated using various operations, such as joins, filters, aggregations, etc. DataFrames can be created from different sources, such as tables, files, or SQL queries. Snowpark operations are executed lazily on the server, which means that they are not performed until an action is triggered, such as a write or a collect operation. This allows Snowpark to optimize the execution plan and reduce the amount of data transferred between the client and the server.
The other options are not characteristics of data transformations in Snowpark that should be considered to meet this requirement. Option C is incorrect because User-Defined Functions (UDFs) are pushed down to Snowflake and executed on the server. Option D is incorrect because Snowpark does not require a separate cluster outside of Snowflake for computations, but rather uses virtual warehouses within Snowflake. Option E is incorrect because columns in different DataFrames with the same name should be referred to with dot notation, not squared brackets.

NEW QUESTION 57
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 58
Which methods can be used to create a DataFrame object in Snowpark? (Select THREE)

A. session.jdbc_connection()
B. session.read.json{)
C. session,table()
D. DataFraas.writeO
E. session.builder()
F. session.sql()

Answer: BCF

Explanation:
The methods that can be used to create a DataFrame object in Snowpark are session.read.json(), session.table(), and session.sql(). These methods can create a DataFrame from different sources, such as JSON files, Snowflake tables, or SQL queries.
The other options are not methods that can create a DataFrame object in Snowpark. Option A, session.jdbc_connection(), is a method that can create a JDBC connection object to connect to a database. Option D, DataFrame.write(), is a method that can write a DataFrame to a destination, such as a file or a table. Option E, session.builder(), is a method that can create a SessionBuilder object to configure and build a Snowpark session.

NEW QUESTION 59
.....

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