



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

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 3

Which methods will trigger an action that will evaluate a DataFrame? (Select TWO)

- A. DataFrame.random_split ()
- B. DataFrame.collect ()
- C. DataFrame.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 new customer table is created by a data pipeline in a Snowflake schema where MANAGED ACCESS enabled. Can grant 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 5

A Data Engineer ran a stored procedure containing various transactions. During the execution, the session abruptly disconnected preventing one transaction from committing or rolling back. 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_TIMEOUT to 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 6

The following is returned from SYSTEMCLUSTERING_INFORMATION () for a table named orders with a date column named O_ORDERDATE:

```
{
  "cluster_by_keys" : "LINEAR(YEAR(O_ORDERDATE))",
  "total_partition_count" : 536,
  "total_constant_partition_count" : 493,
  "average_overlaps" : 0.1716,
  "average_depth" : 1.0914,
  "partition_depth_histogram" : {
    "00000" : 0,
    "00001" : 491,
    "00002" : 41,
    "00003" : 4,
    "00004" : 0,
    "00005" : 0,
    "00006" : 0,
    "00007" : 0,
    "00008" : 0,
    "00009" : 0,
    "00010" : 0,
    "00011" : 0,
    "00012" : 0,
    "00013" : 0,
    "00014" : 0,
    "00015" : 0,
    "00016" : 0
  }
}
```

What does the total_constant_partition_count value indicate about this table?

- A. The table is clustered very well on O_ORDERDATE, as there are 493 micro-partitions that could not be significantly improved by reclustered.
- B. The table is not clustered well on O_ORDERDATE, as there are 493 micro-partitions where the range of values in that column overlap with every other micro-partition in the table.
- C. The data in O_ORDERDATE does not change very often as there are 493 micro-partitions containing rows where that column has not been modified since the row was created.
- D. The data in O_ORDERDATE has a very low cardinality as there are 493 micro-partitions where there is only a single distinct value in that column for all rows in the micro-partition.

Answer: B

Explanation:

The total_constant_partition_count value indicates the number of micro-partitions where the clustering key column has a constant value across all rows in the micro-partition. However, this does not necessarily mean that the table is clustered well on that column, as there could be other micro-partitions where the range of values in that column overlap with each other. This is the case for the orders table, as the clustering depth is 1, which means that every micro-partition overlaps with every other micro-partition on O_ORDERDATE. This indicates that the table is not clustered well on O_ORDERDATE and could benefit from reclustered.

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 ktask, 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 and set 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 by Snowpipe, 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

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 not support 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 10

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 in Snowflake and runs a COPY INTO 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 stage
- 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 12

What is a characteristic of the operations of streams in Snowflake?

- A. Whenever a stream is queried, the offset is automatically advanced.
- B. When a stream is used to update a target table the offset is advanced to the current time.
- C. Querying a stream returns all change records and table rows from the current offset to the current time.
- D. Each committed and uncommitted transaction on the source table automatically puts a change record in the stream.

Answer: C

Explanation:

A stream is a Snowflake object that records the history of changes made to a table. A stream has an offset, which is a point in time that marks the beginning of the change records to be returned by the stream. Querying a stream returns all change records and table rows from the current offset to the current time. The offset is not automatically advanced by querying the stream, but it can be manually advanced by using the ALTER STREAM command. When a stream is used to update a target table, the offset is advanced to the current time only if the ON UPDATE clause is specified in the stream definition. Each committed transaction on the source table automatically puts a change record in the stream, but uncommitted transactions do not.

NEW QUESTION 15

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 20

The following code is executed in a Snowflake environment with the default settings:

```

able customer;

transaction;

table customer
integer,
varchar

into customer values ('1', 'John');

ck;

$1 from customer;

```

What will be the result of the select statement?

- A. SQL compilation error object 'CUSTOMER' does not exist or is not authorized.
- B. John
- C. 1
- D. 1John

Answer: C

NEW QUESTION 25

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 27

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

```

id": "6282638561cf48544e2ef7e9",
company": "FLYBOYZ",
isActive": true,
name": "Dean Head",
teamMembers": [
  {
    "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
  lateral flatten(v) t2
select
  Name
  ,t2.value:name::varchar AS memberName
  ,t2.value:registered::timestamp AS registeredDttm
  ,t2.value:age::number AS age
  ,t2.value:eyeColor::varchar AS eyeColor
from jCustRaw t1
  lateral 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

Given the table sales which has a clustering key of column CLOSED_DATE which table function will return the average clustering depth for the SALES_REPRESENTATIVE column for the North American region?

- A)
- ```
select system$clustering_information('Sales', 'sales_representative', 'region = 'North America');
```
- B)
- ```
select system$clustering_depth('Sales', 'sales_representative', 'region = 'North America');
```
- C)
- ```
select system$clustering_depth('Sales', 'sales_representative') where region = 'North America';
```
- D)
- ```
select system$clustering_information('Sales', 'sales_representative') where region = 'North America';
```

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

Answer: B

Explanation:

The table function SYSTEM\$CLUSTERING_DEPTH returns the average clustering depth for a specified column or set of columns in a table. The function takes two arguments: the table name and the column name(s). In this case, the table name is sales and the column name is SALES_REPRESENTATIVE. The function also supports a WHERE clause to filter the rows for which the clustering depth is calculated. In this case, the WHERE clause is REGION = 'North America'. Therefore, the function call in Option B will return the desired result.

NEW QUESTION 33

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 37

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 'REVENUE_BY_REGION' IN SCHEMA SALES;
- B. SELECT IS_SECURE FROM SNOWFLAKE
- C. INFORMATION_SCHEMA
- D. FUNCTIONS WHERE FUNCTION_SCHEMA = 'SALES' AND FUNCTION_NAME = 'REVENUE_BY_REGION';
- E. SELECT IS_SECURE FROM INFORMATION_SCHEMA
- F. FUNCTIONS WHERE FUNCTION_SCHEMA = 'SALES' AND FUNCTION_NAME = 'REVENUE_BY_REGION';

- G. SHOW EXTERNAL FUNCTIONS LIKE 'REVENUE_BY_REGION' IN SCHEMA SALES;
- H. SHOW SECURE FUNCTIONS LIKE 'REVENUE 3Y REGION' IN SCHEMA SALES;

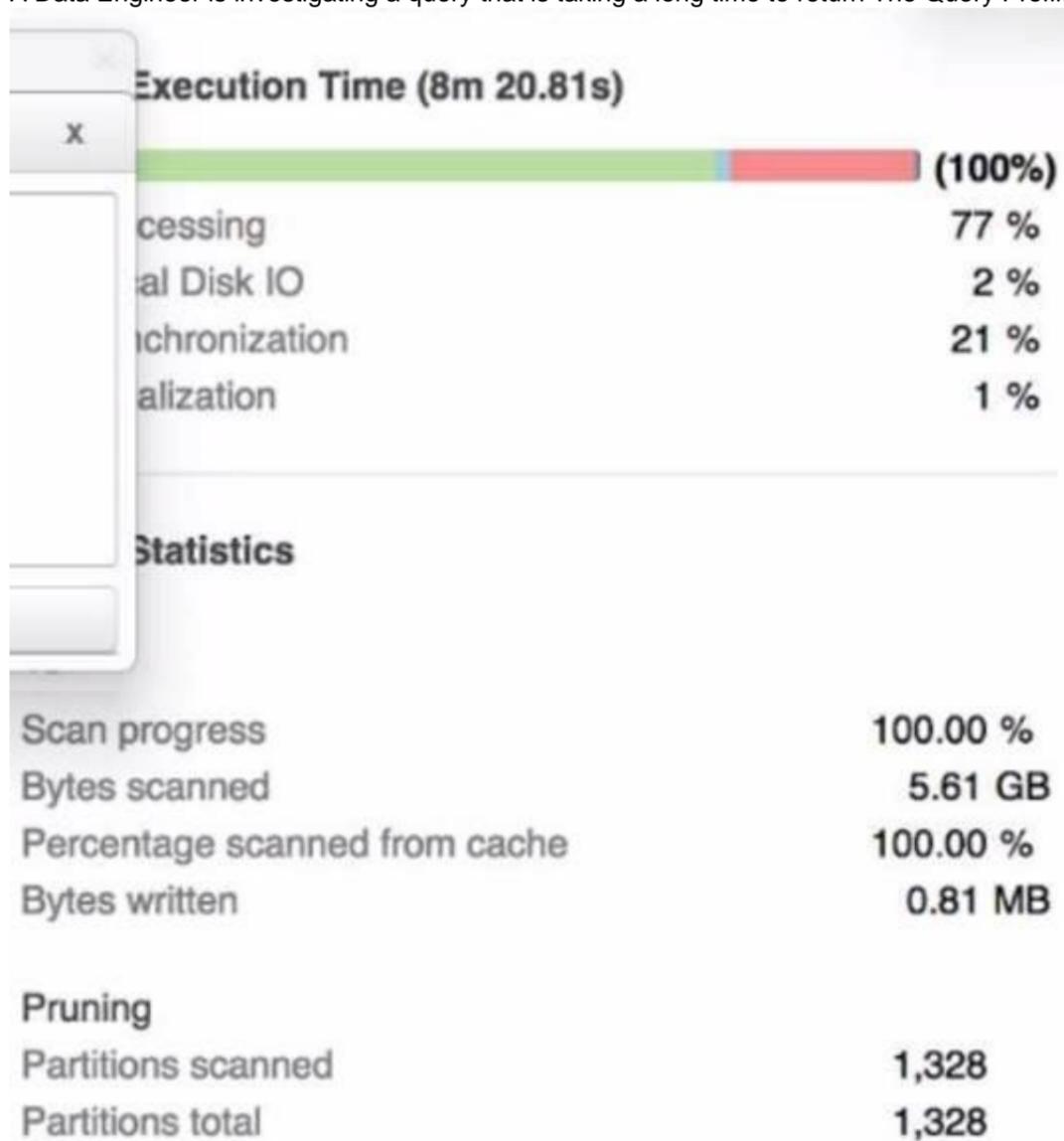
Answer: AB

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 'REVENUE_BY_REGION' 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. INFORMATION_SCHEMA. FUNCTIONS WHERE FUNCTION_SCHEMA = 'SALES' AND FUNCTION_NAME = 'REVENUE_BY_REGION';: 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_SECURE FROM INFORMATION_SCHEMA. FUNCTIONS WHERE FUNCTION_SCHEMA = 'SALES1 AND FUNCTION_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' IN 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 42

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. increase the 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 47

A company is building a dashboard for thousands of Analysts. The dashboard presents the results of a few summary queries on tables that are regularly updated. The query conditions vary by tope according to what data each Analyst needs Responsiveness of the dashboard queries is a top priority, and the data cache should be preserved.

How should the Data Engineer configure the compute resources to support this dashboard?

- A. Assign queries to a multi-cluster virtual warehouse with economy auto-scaling Allow the system to automatically start and stop clusters according to demand.
- B. Assign all queries to a multi-cluster virtual warehouse set to maximized mode Monitor to determine the smallest suitable number of clusters.
- C. Create a virtual warehouse for every 250 Analysts Monitor to determine how many of these virtual warehouses are being utilized at capacity.
- D. Create a size XL virtual warehouse to support all the dashboard queries Monitor query runtimes to determine whether the virtual warehouse should be resized.

Answer: B

Explanation:

This option is the best way to configure the compute resources to support this dashboard. By assigning all queries to a multi-cluster virtual warehouse set to maximized mode, the Data Engineer can ensure that there is enough compute capacity to handle thousands of concurrent queries from different analysts. A multi-cluster virtual warehouse can scale up or down by adding or removing clusters based on the load. A maximized scaling policy ensures that there is always at least one cluster running and that new clusters are added as soon as possible whenneeded. By monitoring the utilization and performance of the virtual warehouse, the Data Engineer can determine the smallest suitable number of clusters that can meet the responsiveness requirement and minimize costs.

NEW QUESTION 52

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 55

How can the following relational data be transformed into semi-structured data using the LEAST amount of operational overhead?

```
create table provinces (province varchar, created_date date);
```

Row	PROVINCE	CREATED_DATE
2	Alberta	2020-01-19
1	Manitoba	2020-01-18

- A. Use the to_json function
- B. Use the PAESE_JSON function to produce a variant value
- C. Use the OBJECT_CONSTRUCT function to return a Snowflake object
- D. Use the TO_VARIANT function to convert each of the relational columns to VARIANT.

Answer: C

Explanation:

This option is the best way to transform relational data into semi-structured data using the least amount of operational overhead. The OBJECT_CONSTRUCT function takes a variable number of key-value pairs as arguments and returns a Snowflake object, which is a variant type that can store JSON data. The function can be used to convert each row of relational data into a JSON object with the column names as keys and the column values as values.

NEW QUESTION 58

A Data Engineer needs to ingest invoice data in PDF format into Snowflake so that the data can be queried and used in a forecasting solution. recommended way to ingest this data?

- A. Use Snowpipe to ingest the files that land in an external stage into a Snowflake table
- B. Use a COPY INTO command to ingest the PDF files in an external stage into a Snowflake table with a VARIANT column.
- C. Create an external table on the PDF files that are stored in a stage and parse the data into structured data
- D. Create a Java User-Defined Function (UDF) that leverages Java-based PDF parser libraries to parse PDF data into structured data

Answer: D

Explanation:

The recommended way to ingest invoice data in PDF format into Snowflake is to create a Java User-Defined Function (UDF) that leverages Java-based PDF parser libraries to parse PDF data into structured data. This option allows for more flexibility and control over how the PDF data is extracted and transformed. The other options are not suitable for ingesting PDF data into Snowflake. Option A and B are incorrect because Snowpipe and COPY INTO commands can only ingest files that are in supported file formats, such as CSV, JSON, XML, etc. PDF files are not supported by Snowflake and will cause errors or unexpected results. Option C is incorrect because external tables can only query files that are in supported file formats as well. PDF files cannot be parsed by external tables and will cause errors or unexpected results.

NEW QUESTION 62

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

- A. When TABLE with FLATTEN requires another source in the from clause to refer to
- B. When TABLE with FLATTEN requires no additional source in the from clause to refer to
- C. When the LATERAL FLATTEN combination requires no other source in the from clause to refer to
- D. When table with FLATTEN is 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 66

Which output is provided by both the SYSTEM\$CLUSTERING_DEPTH function and the SYSTEM\$CLUSTERING_INFORMATION function?

- 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 combination of 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 69

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 73

Which Snowflake feature facilitates access to external API services such as geocoders, data transformation, machine Learning models and other custom code?

- A. Security integration
- B. External tables
- C. External functions
- D. Java User-Defined Functions (UDFs)

Answer: C

Explanation:

External functions are Snowflake functions that facilitate access to external API services such as geocoders, data transformation, machine learning models and other custom code. External functions allow users to invoke external services from within SQL queries and pass arguments and receive results as JSON values. External functions require creating an API integration object and an external function object in Snowflake, as well as deploying an external service endpoint that can communicate with Snowflake via HTTPS.

NEW QUESTION 77

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 79

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 80

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 84

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 85

.....

Thank You for Trying Our Product

We offer two products:

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

2nd - Questions 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

100% Actual & Verified — Instant Download, Please Click
[Order The DEA-C01 Practice Test Here](#)