Snowflake

Exam Questions DEA-C01

SnowPro Advanced: Data Engineer Certification Exam

NEW QUESTION 1
A Data Engineer needs to load JSON output from some software into Snowflake using Snowpipe.
Which recommendations apply to this scenario? (Select THREE)

A. Load large files (1 GB or larger)
B. Ensure that data files are 100-250 MB (or larger) in size compressed
C. Load a single huge array containing multiple records into a single table row
D. Verify each value of each unique element stores a single native data type (string or number)
E. Extract semi-structured data elements containing null values into relational columns before loading
F. Create data files that are less than 100 MB and stage them in cloud storage at a sequence greater than once each minute

Answer: BDF

Explanation:
The recommendations that apply to this scenario are:
? Ensure that data files are 100-250 MB (or larger) in size compressed: This recommendation will improve Snowpipe performance by reducing the number of files that need to be loaded and increasing the parallelism of loading. Smallerfiles can cause performance degradation or errors due to excessive metadata operations or network latency.
? Verify each value of each unique element stores a single native data type (string or number): This recommendation will improve Snowpipe performance by avoiding data type conversions or errors when loading JSON data into variant columns. Snowflake supports two native data types for JSON elements: string and number. If an element has mixed data types across different files or records, such as string and boolean, Snowflake will either convert them to string or raise an error, depending on the FILE_FORMAT option.
? Create data files that are less than 100 MB and stage them in cloud storage at a sequence greater than once each minute: This recommendation will minimize Snowpipe costs by reducing the number of notifications that need to be sent to Snowpipe for auto-ingestion. Snowpipe charges for notifications based on the number of files per notification and the frequency of notifications. By creating smaller files and staging them at a lower frequency, fewer notifications will be needed.

NEW QUESTION 2
The following is returned fromSYSTEMCLUSTERING_INFORMATION () for a tablenamed orders with adate 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_ORDERDATE, as there are 493 micro-partitions that could not be significantly improved by reclustering
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 inO_ORDERDATEdoes not change very often as there are 493 micro- partitionscontaining rows where that column has not been modified since the row was created
D. The data inO_ORDERDATEhas 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 reclustering.

NEW QUESTION 3
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 4
What kind of Snowflake integration is required when defining an external function in Snowflake?

A. API integration
B. HTTP integration
C. Notification integration
D. Security integration

Answer: A

Explanation:
An API integration is required when defining an external function in Snowflake. An API integration is a Snowflake object that defines how Snowflake communicates with an externalservice via HTTPS requests and responses. An API integration specifies parameters such as URL, authentication method, encryption settings, request headers, and timeout values. An API integration is used to create an external function object that invokes the external service from within SQL queries.

NEW QUESTION 5
What is a characteristic of the use of external tokenization?

A. Secure data sharing can be used with external tokenization
B. External tokenization cannot be used with database replication
C. Pre-loading of unmasked data is supported with external tokenization
D. External tokenization allows (he preservation of analytical values after de-identification

Answer: D

Explanation:
External tokenization is a feature in Snowflake that allows users to replace sensitive data values with tokens that are generated and managed by an external service. External tokenization allows the preservation of analytical values after de-identification, such as preserving the format, length, or range of the original values. This way, users can perform analytics on the tokenized data without compromising the security or privacy of the sensitive data.

NEW QUESTION 6
A Data Engineer wants to centralize grant management to maximize security. A user needs ownership on a table m a new schema However, this user should not have the ability to make grant decisions
What is the correct way to do this?

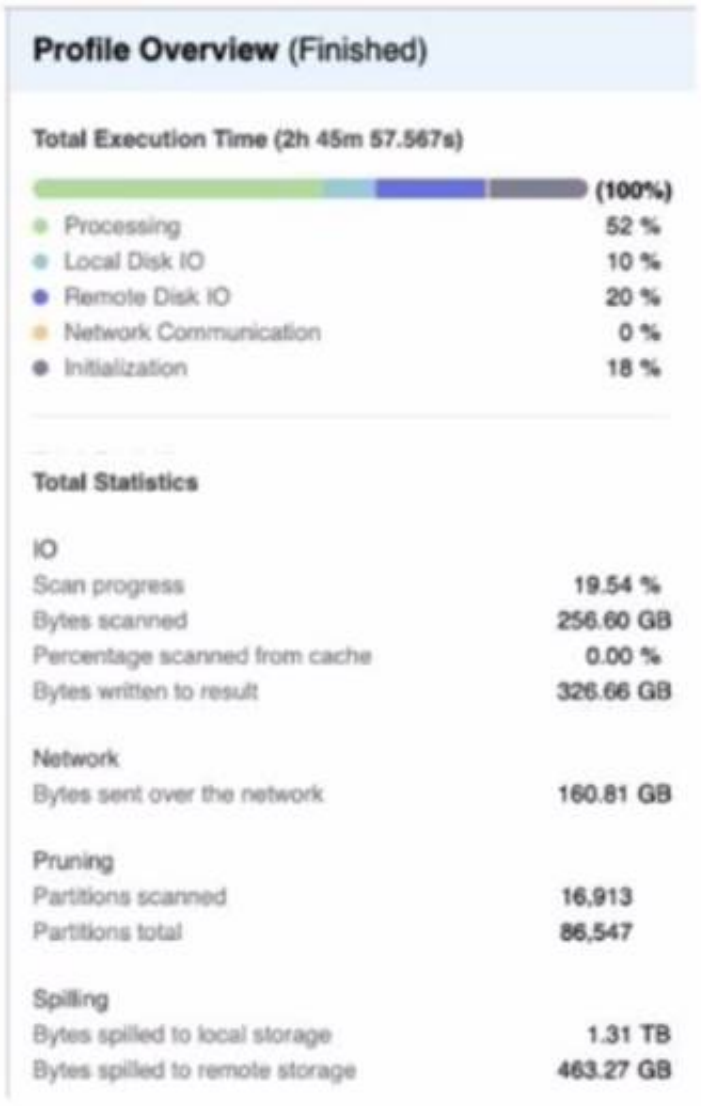
A. Grant ownership to the user on the table
B. Revoke grant decisions from the user on the table
C. Revoke grant decisions from the user on the schema.
D. Add the with managed access parameter on the schema

Answer: D

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

NEW QUESTION 7
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 8
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 9
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_REPRESENTATIVEcolumn for the North American region?
A)

```
select system$clustering_information('Sales', 'sales_representative', 'region = ''North America''');

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 10
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 10
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 13
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 17
Which output is provided by both theSYSTEM$CLUSTERING_DEPTHfunction and theSYSTEM$CLUSTERING_INFORMATIONfunction?

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

Answer: A

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

NEW QUESTION 22
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 26
.....

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