Snowflake

Exam Questions DEA-C01

SnowPro Advanced: Data Engineer Certification Exam

About Exambible

Your Partner of IT Exam

Found in 1998

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

Our Advances

* 99.9% Uptime

All examinations will be up to date.

* 24/7 Quality Support

We will provide service round the clock.

* 100% Pass Rate

Our guarantee that you will pass the exam.

* Unique Gurantee

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

NEW QUESTION 1
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
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 3
The following chart represents the performance of a virtual warehouse over time:

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

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

Answer: B

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

NEW QUESTION 4
A Data Engineer defines the following masking policy:

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

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

```
TABLE customer(

name VARCHAR,
me VARCHAR,
me VARCHAR AS CONCAT(first_name, ' ', last_name)
```

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

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

Answer: A

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

NEW QUESTION 5
A Data Engineer has created table t1 with datatype VARIANT: create or replace table t1 (cl variant);
The Engineer has loaded the following JSON data set. which has information about 4 laptop models into the table:

```
{
"device_model": [
{
"manufacturer": "HP",
"model": "HP 240 G8",
"model_id": "hp 240 g8",
"model_name": "240 G8"
},
{
"manufacturer": "HP",
"model": "HP EliteBook 1030 G1",
"model_id": "hp elitebook 1030 g1",
"model_name": "EliteBook 1030 G1"
},
{
"manufacturer": "HP",
"model": "HP ZBook 15 G2",
"model_id": "hp zbook 15 g2",
"model_name": "ZBook 15 G2"
},
{
"manufacturer": "Lenovo",
"model": "Lenovo B50-70",
"model_id": "lenovo b50-70",
"model_name": "B50-70"
}
]
}
```

The Engineer now wants to query that data set so that results are shown as normal structured data. The result should be 4 rows and 4 columns without the double quotes surrounding the data elements in the JSON data.
The result should be similar to the use case where the data was selected from a normal
relational table z2 where t2 has string data type columns model id. model, manufacturer, and =iccisi_r.an=. and is queried with the SQL clause select * from t2;
Which select command will produce the correct results?
A)

```
select value:model_id::string
, value:model::string
, value:manufacturer::string
, value:model_name::string
from t1
, lateral flatten(input => c1);
```

B)

```
select value:model_id::string
, value:model::string
, value:manufacturer::string
, value:model_name::string
from t1
, lateral flatten(input => c1:device_model);
```

C)

```
select model_id::string
, model::string
, manufacturer::string
, model_name::string
from t1
, lateral flatten(input => c1:device_model);
```

D)

```
select value:model_id
, value:model
, value:manufacturer
, value:model_name
from t1
, lateral flatten(input => c1:device_model);
```

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

Answer: B

NEW QUESTION 6
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 7
A Data Engineer enables a result cache at the session level with the following command: ALTER SESSION SET USE CACHED RESULT = TRUE;
The Engineer then runs the following select query twice without delay:

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

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

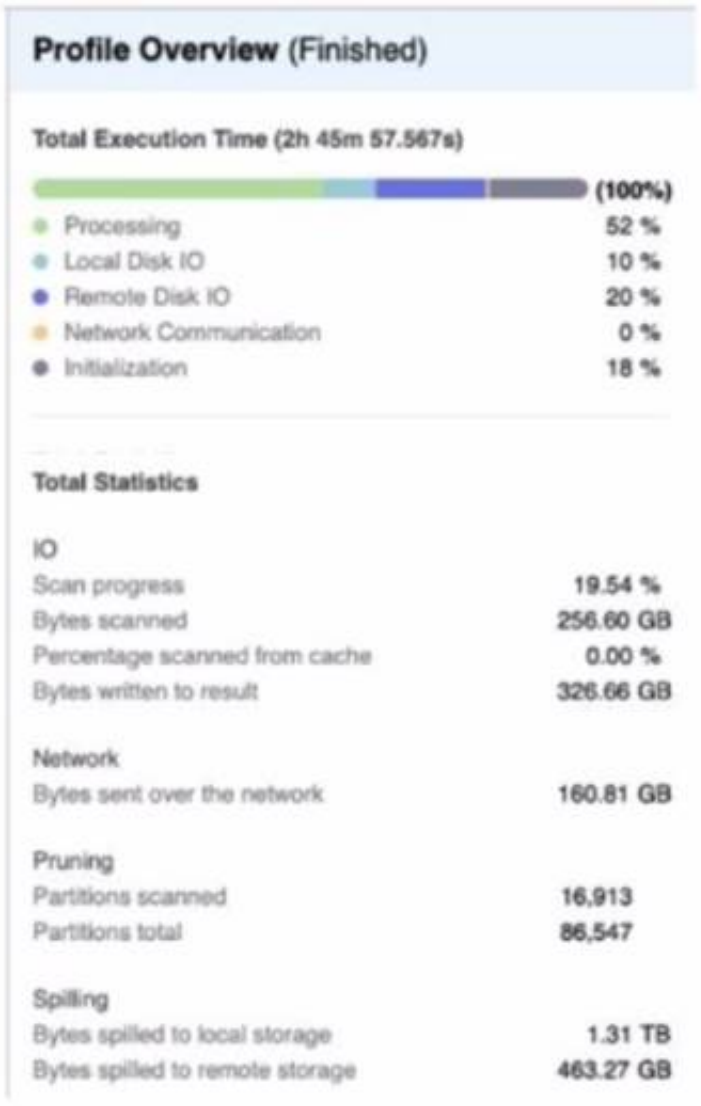
A. The first and second run returned the same results because sample is deterministic
B. The first and second run returned the same results, because the specific SEEDvalue was provided.
C. The first and second run returned different results because the query is evaluated each time it is run.
D. The first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

Answer: B

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

NEW QUESTION 8
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 9
A Data Engineer has developed a dashboard that will issue the same SQL select clause to Snowflake every 12 hours.
---will Snowflake use the persisted query results from the result cache provided that the underlying data has not changed^

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

Answer: C

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

NEW QUESTION 10
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 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 15
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
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 nto 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 20
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 23
A Data Engineer is implementing a near real-time ingestionpipeline to toad data into Snowflake using the Snowflake Kafka connector. There will be three Kafka topics created.
……snowflake objects are created automatically when the Kafka connector starts? (Select THREE)

A. Tables
B. Tasks
C. Pipes
D. internal stages
E. External stages
F. Materialized views

Answer: ACD

Explanation:
The Snowflake objects that are created automatically when the Kafka connector starts are tables, pipes, and internal stages. The Kafka connector will create one table, one pipe, and oneinternal stage for each Kafka topic that is configured in the connector properties. The table will store the data from the Kafka topic, the pipe will load the data from the stage to the table using COPY statements, and the internal stage will store the files that are produced by the Kafka connector using PUT commands. The other options are not Snowflake objects that are created automatically when the Kafka connector starts. Option B, tasks, are objects that can execute SQL statements on a schedule without requiring a warehouse. Option E, external stages, are objects that can reference locations outside of Snowflake, such as cloud storage services. Option F, materialized views, are objects that can store the precomputed results of a query and refresh them periodically.

NEW QUESTION 24
Which system role is recommended for a custom role hierarchy to be ultimately assigned to?

A. ACCOUNTADMIN
B. SECURITYADMIN
C. SYSTEMADMIN
D. USERADMIN

Answer: B

Explanation:
The system role that is recommended for a custom role hierarchy to be ultimately assigned to is SECURITYADMIN. This role has the manage grants privilege on all objects in an account, which allows it to grant access privileges to other roles or revoke them as needed. This role can also create or modify custom roles and assign them to users or other roles. By assigning custom roles to SECURITYADMIN, the role hierarchy can be managed centrally and securely. The other options are not recommended system roles for a custom role hierarchy to be ultimately assigned to. Option A is incorrect because ACCOUNTADMIN is the most powerful role in an account, which has full access to all objects and operations. Assigning custom roles to ACCOUNTADMIN can pose a security risk and should be avoided. Option C is incorrect because SYSTEMADMIN is a role that has full access to all objects in the public schema of the account, but not to other schemas or databases. Assigning custom roles to SYSTEMADMIN can limit the scope and flexibility of the role hierarchy. Option D is incorrect because USERADMIN is a role that can manage users and roles in an account, but not grant access privileges to other objects. Assigning custom roles to USERADMIN can prevent the role hierarchy from controlling access to data and resources.

NEW QUESTION 26
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 28
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 31
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 35
.....

Relate Links

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

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

Contact us

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

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