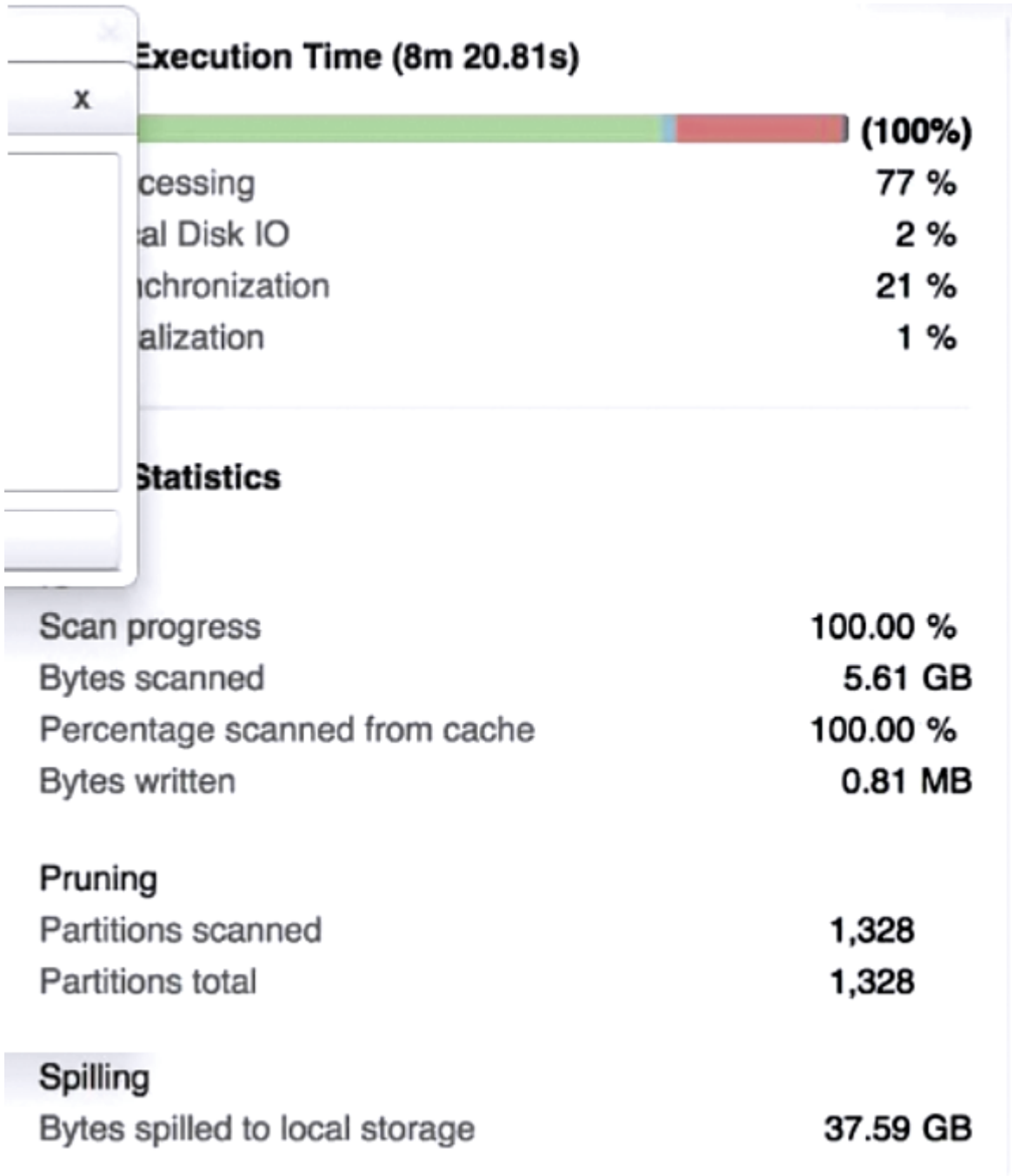
Custom View Settings

## Topic 1 - Exam A

Question #1 *Topic 1*

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.

**Correct Answer:** *B*

*Community vote distribution*

B (100%)

**BigDataBB** 2 months, 3 weeks ago

**Selected Answer: B**

What is disk spilling? When Snowflake warehouse cannot fit an operation in memory, it starts spilling (storing) data first to the local disk of a warehouse node, and then to remote storage.

In such a case, Snowflake first tries to temporarily store the data on the warehouse local disk. As this means extra IO operations, any query that requires spilling will take longer than a similar query running on similar data that is capable to fit the operations in memory.

Also, if the local disk is not sufficient to fit the spilled data, Snowflake further tries to write to the remote cloud storage, which will be shown in the query profile as "Bytes spilled to remote storage".

upvoted 2 times

Question #2 *Topic 1*

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 PARSE_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.

**Correct Answer:** *C*

*Community vote distribution*

C (83%) B (17%)

**BigDataBB** 1 month, 3 weeks ago

**Selected Answer: C**

https://docs.snowflake.com/en/sql-reference/functions/object_construct.

upvoted 1 times

**azure_bimonster** 2 months ago

**Selected Answer: C**

C is the right one.

upvoted 2 times

**stopthisnow** 2 months, 3 weeks ago

**Selected Answer: C**

Correction, C is the right answer.

https://docs.snowflake.com/en/sql-reference/functions/object_construct.

select OBJECT_CONSTRUCT(*) from PC_DBT_DB.DBT_ITAHIR.MANUAL_BOOK1

upvoted 2 times

**stopthisnow** 3 months, 1 week ago

**Selected Answer: B**

https://docs.snowflake.com/en/sql-reference/functions/object_construct.

select OBJECT_CONSTRUCT(*) from PC_DBT_DB.DBT_ITAHIR.MANUAL_BOOK1

upvoted 1 times

Question #3 *Topic 1*

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? (Choose 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.

**Correct Answer:** *EDF*

*Community vote distribution*

DEF (60%) BDE (40%)

**apace92** 5 days, 1 hour ago

**Selected Answer: DEF**

B is incorrect, results are available across virtual warehouses so it doesn't matter if you use the same warehouse or a different one

upvoted 1 times

**BigDataBB** 1 month, 3 weeks ago

**Selected Answer: DEF**

I change my previous answer to D,E,F
https://community.snowflake.com/s/article/Caching-in-the-Snowflake-Cloud-Data-Platform

upvoted 1 times

**azure_bimonster** 2 months ago

**Selected Answer: DEF**

These are the right ones. B is wrong because here it is talking about Result cache not Warehouse cache.

upvoted 2 times

**Eshkin_Kot** 2 months, 1 week ago

**Selected Answer: DEF**

https://community.snowflake.com/s/article/Understanding-Result-Caching

upvoted 1 times

**stopthisnow** 2 months, 2 weeks ago

**Selected Answer: BDE**

B is correct:
This cache is dropped when the warehouse is suspended, which may result in slower initial performance for some queries after the warehouse is resumed. As the resumed warehouse runs and processes more queries, the cache is rebuilt, and queries that are able to take advantage of the cache will experience improved performance.

upvoted 1 times

**BigDataBB** 2 months, 3 weeks ago

**Selected Answer: BDE**

https://community.snowflake.com/s/article/Caching-in-the-Snowflake-Cloud-Data-Platform
Snowflake Cache Layers
The diagram below illustrates the levels at which data and results are cached for subsequent use. These are:-

Result Cache: Which holds the results of every query executed in the past 24 hours. These are available across virtual warehouses, so query results returned to one user is available to any other user on the system who executes the same query, provided the underlying data has not changed.
Local Disk Cache: Which is used to cache data used by SQL queries. Whenever data is needed for a given query it's retrieved from the Remote Disk storage, and cached in SSD and memory.
Remote Disk: Which holds the long term storage. This level is responsible for data resilience, which in the case of Amazon Web Services, means 99.999999999% durability. Even in the event of an entire data centre failure.

upvoted 1 times

**BigDataBB** 1 month, 3 weeks ago

I correct my answer in D,E,F

upvoted 1 times

**stopthisnow** 3 months, 1 week ago

**Selected Answer: DEF**

USED_CACHED_RESULT is not the correct parameter. The correct parameter is "USE_CACHED_RESULT ". Snowflake will use the cache by default and does not require parameter

upvoted 1 times

**kevinstudyplan** 3 months, 1 week ago

**Selected Answer: BDE**

micro partitions are immutable?

upvoted 2 times

Question #4 *Topic 1*

A Data Engineer needs to load JSON output from some software into Snowflake using Snowpipe.

Which recommendations apply to this scenario? (Choose 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.

**Correct Answer:** *BDE*

*Community vote distribution*

BDE (100%)

**Snow_P** 2 months, 2 weeks ago

**Selected Answer: BDE**

https://docs.snowflake.com/en/user-guide/data-load-considerations-prepare

upvoted 2 times

**stopthisnow** 3 months, 1 week ago

**Selected Answer: BDE**

Loading data files roughly 100-250 MB in size or larger reduces the overhead charge relative to the amount of total data loaded to the point where the overhead cost is immaterial.
https://docs.snowflake.com/en/user-guide/data-load-considerations-prepare#label-snowpipe-file-size

upvoted 2 times

Question #5 *Topic 1*

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';

**Correct Answer:** *B*

*Community vote distribution*

B (100%)

**stopthisnow** 3 months, 1 week ago

**Selected Answer: B**

Both will work. Depth is more precise and only shows the depth. Information will show many other properties

upvoted 1 times

**stopthisnow** 3 months, 2 weeks ago

**Selected Answer: B**

Both A & B will give average depth. I will go with B as it only provides the average depth.
https://docs.snowflake.com/en/user-guide/tables-clustering-micropartitions

select system$clustering_depth('GOOGLE_ADS_DEMO.CLICK_PERFORMANCE','CAMPAIGN_ID', 'AD_GROUP_ID=''broccoli''');

The information schema will provide more information like histogram etc.

upvoted 1 times

**stopthisnow** 3 months, 2 weeks ago

Both A & B will give average depth. I will go with B as it only provides the average depth.
https://docs.snowflake.com/en/user-guide/tables-clustering-micropartitions

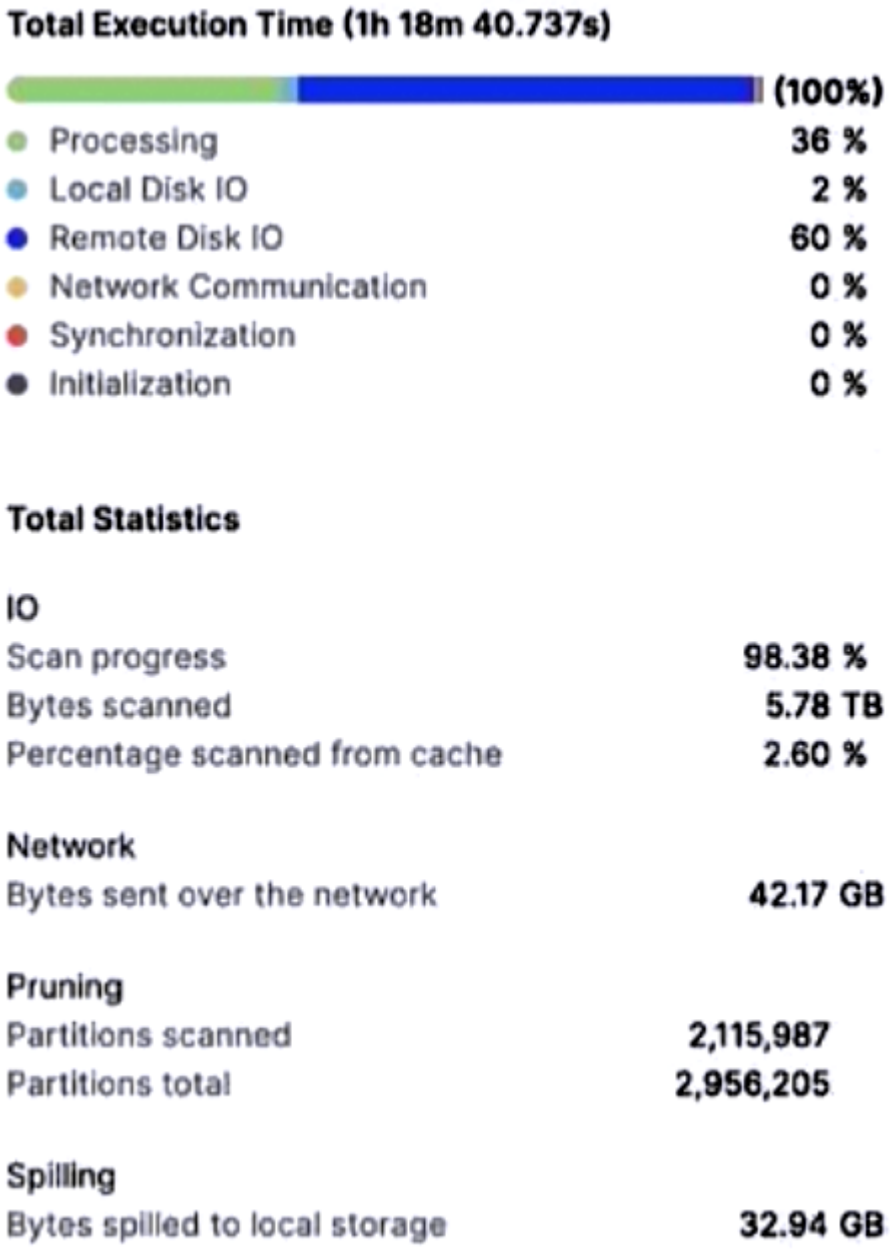
select system$clustering_depth('GOOGLE_ADS_DEMO.CLICK_PERFORMANCE','CAMPAIGN_ID', 'AD_GROUP_ID=''broccoli''');

The information schema will provide more information like histogram etc.

upvoted 2 times

Question #6 *Topic 1*

A large table with 200 columns contains two years of historical data. When queried, the table is filtered on a single day. Below is the Query Profile:




Using a size 2XL virtual warehouse, this query took over an hour to complete.

What will improve the query performance the MOST?

A. Increase the size of the virtual warehouse.

B. Increase the number of clusters in the virtual warehouse.

C. Implement the search optimization service on the table.

D. Add a date column as a cluster key on the table.

**Correct Answer:** *D*

*Community vote distribution*

D (100%)

**prshntdxt7** 1 month ago

are these 71 questions enough to prepare for advanced DE exam ?

upvoted 1 times

**stopthisnow** 3 months, 1 week ago

**Selected Answer: D**

D makes sense

upvoted 2 times

**stopthisnow** 3 months, 2 weeks ago

**Selected Answer: D**

D

Clustering key will not scan the full table which is what is required here.

upvoted 2 times

Question #7 *Topic 1*

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? (Choose 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. SSD attached to an Amazon EC2 instance on AWS eu-west-1 (Ireland)

**Correct Answer:** *BDE*

*Community vote distribution*

CDE (100%)

**stopthisnow** 3 months, 1 week ago

**Selected Answer: CDE**

External stages can be in any region and can be created on any cloud provider. Internal stages are linked to the actual location of Snowflake account

upvoted 2 times

**stopthisnow** 3 months, 2 weeks ago

**Selected Answer: CDE**

Internal stages are created in the same region as the Snowflake account so their location can't be different. External stages can be in any region since they are external.

upvoted 3 times

Question #8 *Topic 1*

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 PROD;

D. CREATE DATABASE DEV -

CLONE PROD -
DATA_RETENTION_TIME_IN DAYS = 0;

**Correct Answer:** *C*

*Community vote distribution*

C (100%)

**apace92** 5 days, 1 hour ago

**Selected Answer: C**

Transient and Temp tables have 0 day Fail-Safe

upvoted 1 times

**stopthisnow** 3 months, 1 week ago

**Selected Answer: C**

Both Transient and Temporary tables have 0 day Fail-safe and 0/1 day time travel. https://docs.snowflake.com/en/user-guide/tables-temp-transient

upvoted 3 times

**stopthisnow** 3 months, 2 weeks ago

**Selected Answer: C**

Because transient tables do not have a Fail-safe period, they provide a good option for managing the cost of very large tables used to store transitory data; however, the data in these tables cannot be recovered after the Time Travel retention period passes.

upvoted 3 times

Question #9 *Topic 1*

Which query will show a list of the 20 most recent executions of a specified task, MYTASK, that have been scheduled within the last hour that have ended or are still running?

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')
```

**Correct Answer:** *C*

*Community vote distribution*

B (100%)

**BigDataBB** 2 months, 3 weeks ago

**Selected Answer: B**

This is a really a Bast*** Inside question, because all the stases are:
SCHEDULED: scheduled for execution.
EXECUTING: currently executing.
SUCCEEDED: execution successful.
FAILED: execution failed.
FAILED_AND_AUTO_SUSPENDED: task failed, and was automatically suspended.
CANCELLED: execution cancelled.
SKIPPED

upvoted 1 times

**Gemnidhi17** 3 months, 1 week ago

B because of reason listed by @stopthisnow (c point)

upvoted 1 times

**stopthisnow** 3 months, 2 weeks ago

**Selected Answer: B**

B
To query only those tasks that have already completed or are currently running, include WHERE query_id IS NOT NULL as a filter. The QUERY_ID column in the TASK_HISTORY output is populated only when a task has started running.
https://docs.snowflake.com/en/sql-reference/functions/task_history
A - will give all the schedules even the ones that have not run yet
C - A schedule could be skipped, cancelled so it won't give all the runs
D - It won't return the most recent tasks.

upvoted 3 times

Question #10 *Topic 1*

Which methods can be used to create a DataFrame object in Snowpark? (Choose three.)

A. session.jdbc_connection()

B. session.read.json()

C. session.table()

D. DataFrame.write()

E. session.builder()

F. session.sql()

**Correct Answer:** *BCF*

*Community vote distribution*

BCF (100%)

**stopthisnow** 3 months, 1 week ago

**Selected Answer: BCF**

df = session.sql("select 1/2")
https://docs.snowflake.com/en/developer-guide/snowpark/reference/python/latest/api/snowflake.snowpark.DataFrame

upvoted 4 times

Question #11 *Topic 1*

A new CUSTOMER table is created by a data pipeline in a Snowflake schema where MANAGED ACCESS is enabled.
Which roles can grant access to the CUSTOMER table? (Choose 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

**Correct Answer:** *ABC*

*Community vote distribution*

AEF (100%)

**djoekje** Highly Voted 3 months, 2 weeks ago

Wrong solution, should be A-E-F:
In managed access schemas (i.e. schemas created using the CREATE SCHEMA ... WITH MANAGED ACCESS syntax), object owners lose the ability to make grant decisions. Only the schema owner (i.e. the role with the OWNERSHIP privilege on the schema) or a role with the global MANAGE GRANTS privilege can grant privileges on objects in the schema.

upvoted 7 times

**stopthisnow** Most Recent 3 months, 1 week ago

**Selected Answer: AEF**

https://docs.snowflake.com/en/user-guide/security-access-control-configure
-SECURITYADMIN or higher
-Schema owner
-Any role with the MANAGE GRANTS privilege

upvoted 4 times

**acapone001** 3 months, 2 weeks ago

Thank you! I was struggling with this solution because it doesn't match the Snowflake documentation.
AEF should be correct.

upvoted 2 times

Question #12 *Topic 1*

What is the purpose of the BUILD_STAGE_FILE_URL function in Snowflake?

A. It generates an encrypted URL for accessing a file in a stage.

B. It generates a staged URL for accessing a file in a stage.

C. It generates a permanent URL for accessing files in a stage.

D. It generates a temporary URL for accessing a file in a stage.

**Correct Answer:** *C*

*Community vote distribution*

C (100%)

**stopthisnow** 3 months, 1 week ago

**Selected Answer: C**

Generates a Snowflake file URL to a staged file using the stage name and relative file path as inputs. A file URL permits prolonged access to a specified file. That is, the file URL does not expire.
https://docs.snowflake.com/en/sql-reference/functions/build_stage_file_url

upvoted 2 times

Question #13 *Topic 1*

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 with 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;
```

B.
```
select
    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 => v:teamMembers) t2;
```

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;
```

**Correct Answer:** *B*

*Community vote distribution*

B (100%)

**BigDataBB** 2 months, 3 weeks ago

**Selected Answer: B**

https://docs.snowflake.com/user-guide/semistructured-considerations#using-flatten-to-list-distinct-key-names

upvoted 1 times

**stopthisnow** 3 months, 1 week ago

**Selected Answer: B**

Technically, none of the statements work properly. The nearest valid option is B.

```
create or replace table persons as
select parse_json(
'{
"_id": "4534554654640",
"company": "Boo",
"name": "Dean",
"teamMembers": [
{
"age":29,
"eyeColor": "green"
},
{
"age":39,
"eyeColor": "green2"
}
]
}'
) as c ;

select f.value:eyeColor::varchar as eyeColor,
f.value:age
from TASK_DB.PUBLIC.PERSONS as p,
lateral flatten(input => p:teamMembers) f;
```

upvoted 2 times

Question #14 *Topic 1*

A company has an extensive script in Scala that transforms data by leveraging DataFrames. A Data Engineer needs to move these transformations to Snowpark.
What characteristics of data transformations in Snowpark should be considered to meet this requirement? (Choose 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.

**Correct Answer:** *AB*

*Community vote distribution*

AB (100%)

**stopthisnow** 3 months, 1 week ago

**Selected Answer: AB**

Referring to Columns in Different DataFrames
When referring to columns in two different DataFrame objects that have the same name (for example, joining the DataFrames on that column), you can use the DataFrame.col method in one DataFrame object to refer to a column in that object (for example, df1.col("name") and df2.col("name")).

To retrieve and manipulate data, you use the DataFrame class. A DataFrame represents a relational dataset that is evaluated lazily: it only executes when a specific action is triggered. In a sense, a DataFrame is like a query that needs to be evaluated in order to retrieve data.

https://docs.snowflake.com/en/developer-guide/snowpark/python/working-with-dataframes

upvoted 4 times

Question #15 *Topic 1*

The following is returned from SYSTEM$CLUSTERING_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 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 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.

**Correct Answer:** *A*

*Community vote distribution*

A (100%)

**stopthisnow** 3 months, 1 week ago

**Selected Answer: A**

total_constant_partition_count
Total number of micro-partitions for which the value of the specified columns have reached a constant state (i.e. the micro-partitions will not benefit significantly from reclustering). The number of constant micro-partitions in a table has an impact on pruning for queries. The higher the number, the more micro-partitions can be pruned from queries executed on the table, which has a corresponding impact on performance.

https://docs.snowflake.com/en/sql-reference/functions/system_clustering_information

upvoted 3 times

Question #16 *Topic 1*

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 topic 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.

**Correct Answer:** *B*

*Community vote distribution*

B (100%)

**stopthisnow** 3 months, 1 week ago

**Selected Answer: B**

This mode is enabled by specifying the same value for both maximum and minimum number of clusters (note that the specified value must be larger than 1). In this mode, when the warehouse is started, Snowflake starts all the clusters so that maximum resources are available while the warehouse is running.

This mode is effective for statically controlling the available compute resources, particularly if you have large numbers of concurrent user sessions and/or queries and the numbers do not fluctuate significantly.

The key term here is : "The top priority is the responsiveness".

A is also valid but it will have impact on the responsiveness due to economy option. The queries will queue up before scale-up happens.

upvoted 2 times

Question #17 *Topic 1*

A Data Engineer has developed a dashboard that will issue the same SQL select clause to Snowflake every 12 hours.
How long 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

**Correct Answer:** *B*

*Community vote distribution*

D (100%)

**stopthisnow** Highly Voted 3 months, 1 week ago

**Selected Answer: D**

Each time the persisted result for a query is reused, Snowflake resets the 24-hour retention period for the result, up to a maximum of 31 days from the date and time that the query was first executed. After 31 days, the result is purged and the next time the query is submitted, a new result is generated and persisted.
https://docs.snowflake.com/en/user-guide/querying-persisted-results

upvoted 6 times

**Eshkin_Kot** Most Recent 2 months, 1 week ago

**Selected Answer: D**

Each time the persisted result for a query is reused, Snowflake resets the 24-hour retention period for the result, up to a maximum of 31 days

upvoted 2 times

**Gemnidhi17** 3 months, 1 week ago

correct ans is D 31 days because it can extends upto 31 days if executed again within 24 hours.

upvoted 4 times

Question #18 *Topic 1*

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.
What step 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.

**Correct Answer:** *A*

*Community vote distribution*

A (100%)

**stopthisnow** 3 months, 1 week ago

**Selected Answer: A**

Aborting Transactions
If a transaction is running in a session and the session disconnects abruptly, preventing the transaction from committing or rolling back, the transaction is left in a detached state, including any locks that the transaction is holding on resources. If this happens, you might need to abort the transaction.

To abort a running transaction, the user who started the transaction or an account administrator can call the system function, SYSTEM$ABORT_TRANSACTION.

If the transaction is not aborted by the user:

If it blocks another transaction from acquiring a lock on the same table and is idle for 5 minutes, it is automatically aborted and rolled back.

If it does not block other transactions from modifying the same table and is older than 4 hours, it is automatically aborted and rolled back.

upvoted 3 times

Question #19 *Topic 1*

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.

**Correct Answer:** *D*

*Community vote distribution*

D (100%)

**randreag** Highly Voted 3 months, 1 week ago

**Selected Answer: D**

in the documentation, it says: RETURNS NULL ON NULL INPUT will not call the stored procedure if any input is null, so the statements inside the stored procedure will not be executed.

upvoted 5 times

**chinese1250** Most Recent 2 weeks, 3 days ago

RETURNS NULL ON NULL INPUT (or its synonym STRICT) will not call the UDF if any input is null. Instead, a null value will always be returned for that row. Note that the UDF might still return null for non-null inputs.

see: https://docs.snowflake.com/en/sql-reference/sql/create-function

upvoted 1 times

**kevinstudyplan** 3 months, 1 week ago

"return null on null input"

upvoted 2 times

**acapone001** 3 months, 2 weeks ago

why does the log table contain no records? Wouldn't the SQL statement execute and one record with a null value be inserted?

upvoted 1 times

Question #20 *Topic 1*

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.

**Correct Answer:** *B*

*Community vote distribution*

B (60%) D (40%)

**BigDataBB** 2 weeks, 4 days ago

**Selected Answer: B**

as wrote here: https://docs.snowflake.com/en/user-guide/json-basics-tutorial-flatten

You can flatten the event objects in the events array into separate rows using the FLATTEN function. The function output includes a VALUE column that stores these individual events.

You can then use the LATERAL modifier to join the FLATTEN function output with any information outside of the object

upvoted 1 times

**prshntdxt7** 3 weeks, 2 days ago

**Selected Answer: D**

D. When TABLE with FLATTEN is acting like a sub-query executed for each returned row.

The TABLE with FLATTEN combination is typically used when you want to treat the result of the FLATTEN function as a table that can be further joined or manipulated. This is particularly useful when you need to perform additional operations on each row returned by the FLATTEN function. The LATERAL FLATTEN combination is used when the FLATTEN function references other tables or sources in the FROM clause, and it allows for correlated subqueries.

upvoted 1 times

**BigDataBB** 1 month, 3 weeks ago

**Selected Answer: D**

By my opinion,
as is wrote here: https://docs.snowflake.com/en/sql-reference/functions/flatten
the answer should be: Flattens (explodes) compound values into multiple rows.

upvoted 1 times

**Snow_P** 2 months, 2 weeks ago

**Selected Answer: B**

https://docs.snowflake.com/en/sql-reference/functions/flatten

upvoted 2 times

Question #21 *Topic 1*

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

**Correct Answer:** *A*

*Community vote distribution*

A (100%)

**Snow_P** 2 months, 2 weeks ago

**Selected Answer: A**

https://docs.snowflake.com/en/sql-reference/functions/system_clustering_information

https://docs.snowflake.com/en/sql-reference/functions/system_clustering_depth

upvoted 3 times

**stopthisnow** 3 months, 1 week ago

**Selected Answer: A**

average depth is common

upvoted 1 times

Question #22 *Topic 1*

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. What is the 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.

**Correct Answer:** *D*

*Community vote distribution*

D (100%)

**Snow_P** 2 months, 2 weeks ago

**Selected Answer: D**

Using python or java libraries so I guess D

upvoted 2 times

Question #23 *Topic 1*

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.

**Correct Answer:** *C*

*Community vote distribution*

D (60%) C (40%)

**prshntdxt7** 3 weeks, 2 days ago

**Selected Answer: C**

C. Set the FORCE=TRUE option in the Snowpipe COPY INTO command.

Using the FORCE=TRUE option in the Snowpipe COPY INTO command allows reloading files even if they have been loaded previously. This can be useful in situations where you need to reload data, and the metadata of the pipe is preventing the same files from being loaded again. It minimizes operational overhead by avoiding the need to wait for metadata expiration or modifying the files.

upvoted 1 times

**5effea7** 3 weeks, 3 days ago

**Selected Answer: C**

According to https://docs.snowflake.com/en/sql-reference/sql/copy-into-table FORCE = TRUE would "load all files, regardless of whether they've been loaded previously and have not changed since they were loaded". The question asked for the least impact. C is the least impact

upvoted 1 times

**stopthisnow** 3 months, 1 week ago

**Selected Answer: D**

Snowpipe ignores modified files that are staged again. To reload modified data files, it is currently necessary to recreate the pipe object using the CREATE OR REPLACE PIPE syntax.
https://docs.snowflake.com/en/user-guide/data-load-snowpipe-ts

upvoted 3 times

Question #24 *Topic 1*

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

A. TRANSACTIONS_STM will keep working as expected.

B. TRANSACTIONS_STM will be stale and will need to be re-created.

C. TRANSACTIONS_STM will be automatically renamed TRANSACTIONS_RAW_STM.

D. Reading from the TRANSACTIONS_STM stream will succeed for some time after the expected STALE_TIME.

**Correct Answer:** *B*

*Community vote distribution*

A (100%)

**drunk_goat82** Highly Voted 3 months, 2 weeks ago

I think it's A.
Renaming a source object does not break a stream or cause it to go stale. In addition, if a source object is dropped and a new object is created with the same name, any streams linked to the original object are not linked to the new object

upvoted 8 times

**randreag** Most Recent 3 months, 1 week ago

but will the stream work as expected if there is no more changes in the table(old) to record? After a while it will become stale, doesn't it?

upvoted 1 times

**stopthisnow** 3 months, 1 week ago

**Selected Answer: A**

Renaming a source object does not break a stream or cause it to go stale. In addition, if a source object is dropped and a new object is created with the same name, any streams linked to the original object are not linked to the new object.
https://docs.snowflake.com/en/user-guide/streams-intro

upvoted 2 times

**acapone001** 3 months, 1 week ago

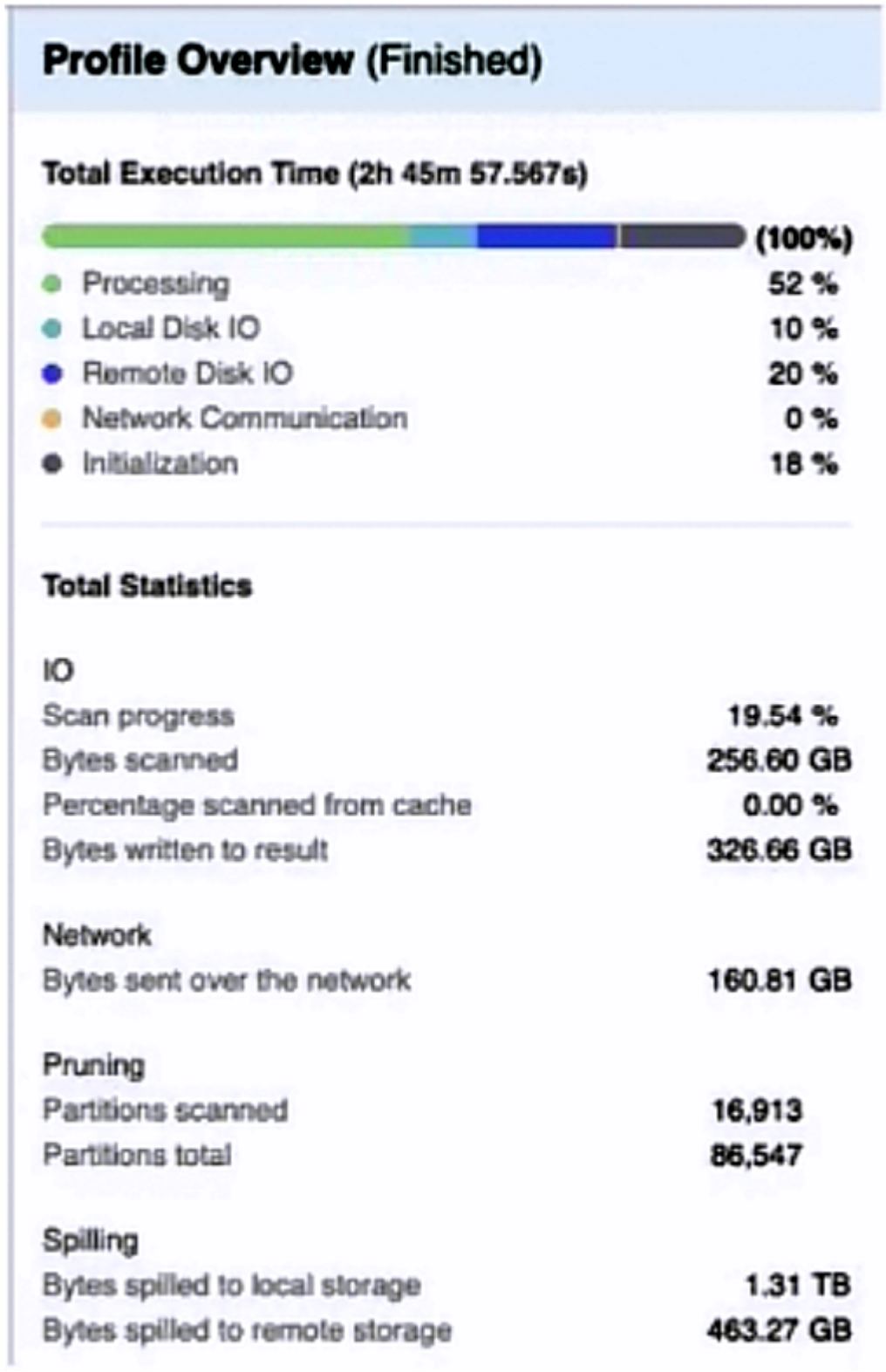
A should be the correct answer here. From Snowflake: "Renaming a source object does not break a stream or cause it to go stale. In addition, if a source object is dropped and a new object is created with the same name, any streams linked to the original object are not linked to the new object."

upvoted 3 times

Question #25 *Topic 1*

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;
```



**Profile Overview (Finished)**

**Total Execution Time (2h 45m 57.567s)** (100%)

| | |
|---|---|
| Processing | 52 % |
| Local Disk IO | 10 % |
| Remote Disk IO | 20 % |
| Network Communication | 0 % |
| Initialization | 18 % |

**Total Statistics**

| | |
|---|---|
| **IO** | |
| Scan progress | 19.54 % |
| Bytes scanned | 256.60 GB |
| Percentage scanned from cache | 0.00 % |
| Bytes written to result | 326.66 GB |
| **Network** | |
| Bytes sent over the network | 160.81 GB |
| **Pruning** | |
| Partitions scanned | 16,913 |
| Partitions total | 86,547 |
| **Spilling** | |
| Bytes spilled to local storage | 1.31 TB |
| Bytes spilled to remote storage | 463.27 GB |

Based on the Query Profile, what are some performance tuning options the Engineer can use? (Choose 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

**Correct Answer:** *AC*

*Community vote distribution*

AC (100%)

**BigDataBB** 1 month, 2 weeks ago

Snowflake don't support the create of index, You can cluster a table by a cluster key, but this option is not in the list of option. in addition to this in the question is write that You are in DEV environment. So You don't need to have all the data but to know how SF process them.

upvoted 1 times

**Eshkin_Kot** 2 months, 1 week ago

Why not D?
It doesn't say we don't need all the data, so why LIMIT (option A)?
Agree with C
upvoted 2 times

**stopthisnow** 3 months, 1 week ago

**Selected Answer: AC**

AC makes sense
upvoted 1 times

Question #26 *Topic 1*

Which methods will trigger an action that will evaluate a DataFrame? (Choose two.)

A. DataFrame.random_split()

B. DataFrame.collect()

C. DataFrame.select()

D. DataFrame.col()

E. DataFrame.show()

**Correct Answer:** *BE*

*Community vote distribution*

BE (100%)

**Snow_P** 2 months, 2 weeks ago

**Selected Answer: BE**

collect and show are actions, for the rest Spark is just writing down what it needs to do when an action occurs
upvoted 1 times

**stopthisnow** 3 months, 1 week ago

**Selected Answer: BE**

https://docs.snowflake.com/en/developer-guide/snowpark/python/working-with-dataframes#performing-an-action-to-evaluate-a-dataframe
As mentioned earlier, the DataFrame is lazily evaluated, which means the SQL statement isn't sent to the server for execution until you perform an action. An action causes the DataFrame to be evaluated and sends the corresponding SQL statement to the server for execution.
upvoted 3 times

Question #27 *Topic 1*

Which Snowflake objects does the Snowflake Kafka connector use? (Choose three.)

A. Pipe

B. Serverless task

C. Internal user stage

D. Internal table stage

E. Internal named stage

F. Storage integration

**Correct Answer:** *ADE*

*Community vote distribution*

ADE (67%) ABD (17%) ABE (17%)

**BigDataBB** 23 hours, 7 minutes ago

**Selected Answer: ABD**

https://docs.snowflake.com/en/user-guide/kafka-connector-overview
One table for each topic.
One internal stage to temporarily store data files for each topic

upvoted 1 times

**BigDataBB** 4 days, 8 hours ago

**Selected Answer: ABE**

I finally find the solution, a pipe use a "Snowflake-provided virtual warehouse" so a serverless task.
For this reason the answer are A, B, E

upvoted 1 times

**BigDataBB** 23 hours, 7 minutes ago

Sorry I correct my self A,B,D
https://docs.snowflake.com/en/user-guide/kafka-connector-overview
One table for each topic.
One internal stage to temporarily store data files for each topic

So table stage

upvoted 1 times

**BigDataBB** 1 month, 2 weeks ago

I think that there is an error, kafka connector uses these snowflake objects:
- pipe
- internal named stage
- tables

And not internal table stage

upvoted 1 times

**BigDataBB** 1 month, 2 weeks ago

https://docs.snowflake.com/en/user-guide/kafka-connector-manage

upvoted 1 times

**Snow_P** 2 months, 2 weeks ago

**Selected Answer: ADE**

pipe and internal stages

upvoted 2 times

**stopthisnow** 3 months, 1 week ago

**Selected Answer: ADE**

The connector creates the following objects for each topic:

One internal stage to temporarily store data files for each topic.

One pipe to ingest the data files for each topic partition.

One table for each topic.
https://docs.snowflake.com/en/user-guide/kafka-connector-overview

upvoted 2 times

**randreag** 3 months, 1 week ago

Is this answer correct?

upvoted 1 times

Question #28 *Topic 1*

A Data Engineer has created table t1 with one column c1 with datatype VARIANT: create or replace table t1 (c1 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 t2, where t2 has string data type columns model_id, model, manufacturer, and model_name, 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);
```

**Correct Answer:** *B*

*Community vote distribution*

B (100%)

**dr_klauf** 1 month ago

**Selected Answer: B**

Definitely B.
D gives a similar result but with double quotes.

upvoted 2 times

**stopthisnow** 3 months, 1 week ago

**Selected Answer: B**

Possibly B

upvoted 2 times

Question #29 *Topic 1*

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 the preservation of analytical values after de-identification.

**Correct Answer:** *D*

*Community vote distribution*

D (73%) C (27%)

**prshntdxt7** 3 weeks, 1 day ago

**Selected Answer: D**

**correction- changing answer to D, @Maicas you're right.

D. External tokenization allows the preservation of analytical values after de-identification.

External tokenization typically involves using a separate service or system to tokenize sensitive data. This method allows for the preservation of certain analytical values even after de-identification, making it possible to perform some types of analysis on the tokenized data without compromising individual identities.

upvoted 4 times

**prshntdxt7** 1 month ago

**Selected Answer: C**

option-C
https://docs.snowflake.com/en/user-guide/security-column-ext-token-intro#external-tokenization-benefits
"Using a tokenization provider, tokenized data is pre-loaded into Snowflake. Therefore, even without applying a masking policy to a column in a table or view, users never see the real data value. This provides enhanced data security to the most sensitive data in your organization."

upvoted 1 times

**Maicas** 4 weeks ago

Tokenized data is protected, preventing users from seing the real value. Just the opposite of answer C, which is UNmasked data.

upvoted 1 times

**Maicas** 1 month ago

**Selected Answer: D**

Can't be C because it says "pre-loading of UNmasked data".
The correct answer is D

https://docs.snowflake.com/en/user-guide/security-column-intro#benefits

upvoted 4 times

**stopthisnow** 2 months, 3 weeks ago

**Selected Answer: C**

https://docs.snowflake.com/en/user-guide/security-column-ext-token-intro
Using a tokenization provider, tokenized data is pre-loaded into Snowflake. Therefore, even without applying a masking policy to a column in a table or view, users never see the real data value. This provides enhanced data security to the most sensitive data in your organization.

upvoted 2 times

Question #30 *Topic 1*

A Data Engineer is implementing a near real-time ingestion pipeline to load data into Snowflake using the Snowflake Kafka connector. There will be three Kafka topics created.
Which Snowflake objects are created automatically when the Kafka connector starts? (Choose three.)

A. Tables

B. Tasks

C. Pipes

D. Internal stages

E. External stages

F. Materialized views

**Correct Answer:** *ACD*

*Community vote distribution*

ACD (100%)

**prshntdxt7** 1 month ago

**Selected Answer: ACD**

The three Snowflake objects created automatically when the Snowflake Kafka connector starts are:

Tables (A): The connector creates tables to store the actual data ingested from Kafka. These tables can be used to query and analyze the data.

Pipes (C): Pipes are automatically created to move data from the staging area (staging table) to the final destination tables in Snowflake. Pipes are part of Snowflake's data loading mechanism.

Internal stages (D): The connector may create internal stages to facilitate the movement of data between the Kafka topics and the staging tables or final destination tables in Snowflake.

Therefore, the correct choices are A (Tables), C (Pipes), and D (Internal stages).

upvoted 2 times

**stopthisnow** 2 months, 3 weeks ago

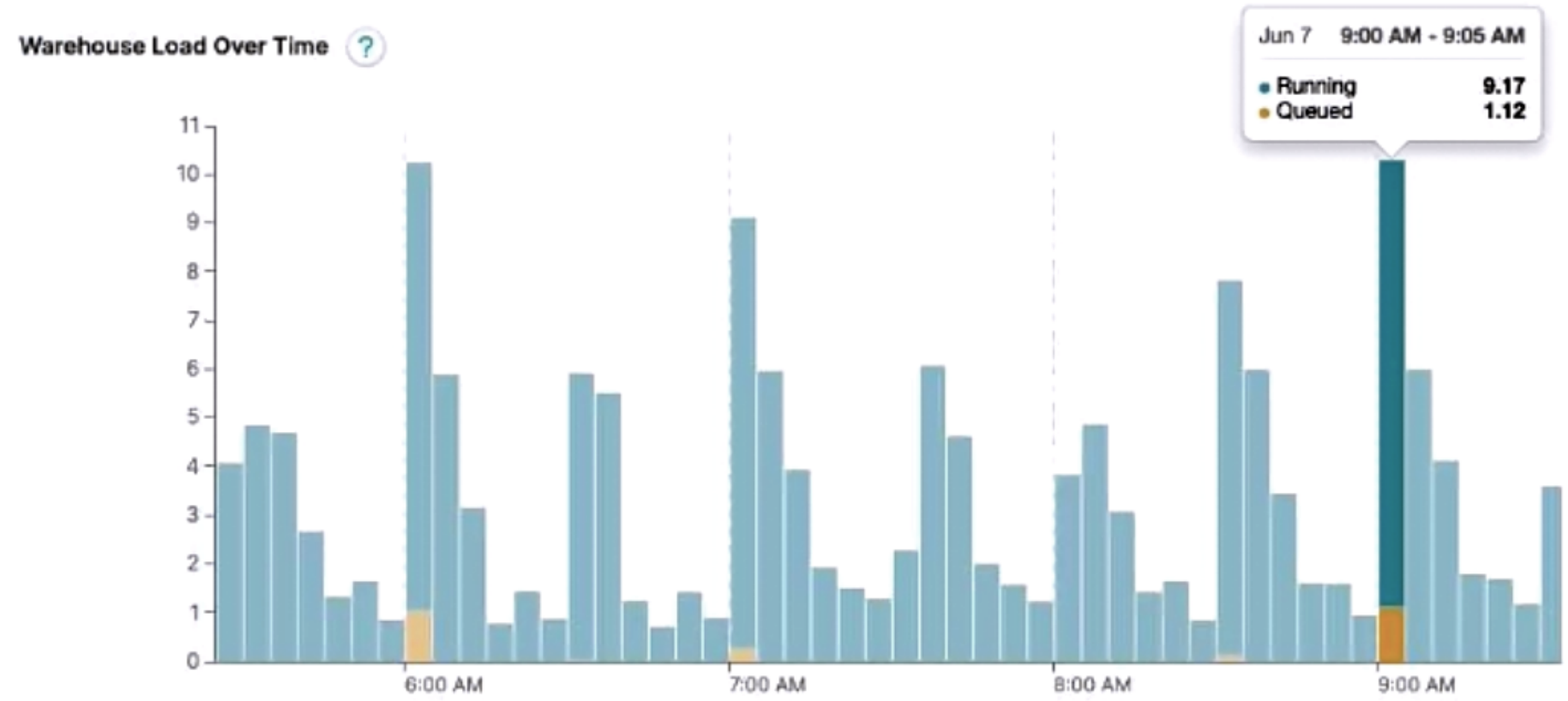
**Selected Answer: ACD**

Kafka topics can be mapped to existing Snowflake tables in the Kafka configuration. If the topics are not mapped, then the Kafka connector creates a new table for each topic using the topic name.
https://docs.snowflake.com/en/user-guide/kafka-connector-overview#:~:text=Kafka%20topics%20can%20be%20mapped%20to%20existing%20Snowflake%20tables%20in%20the%20Kafka%20configuration.%20If%20the%20topics%20are%20not%20mapped%2C%20then%20the%20Kafka%20connector%20creates%20a%20new%20table%20for%20each%20topic%20using%20the%20topic%20name.

upvoted 2 times

Question #31 *Topic 1*

The following chart represents the performance of a virtual warehouse over time:




A Data Engineer notices that the warehouse is queueing queries. The warehouse is size X-Small, the minimum and maximum cluster counts are set to 1, the scaling policy is set to standard, and auto-suspend is set to 10 minutes.
How can the 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.

**Correct Answer:** *A*

*Community vote distribution*

A (100%)

**stopthisnow** 2 months, 3 weeks ago

**Selected Answer: A**

A could be correct. The cluster size is 1 and the queries are taking a long time to complete

upvoted 1 times

Question #32 *Topic 1*

A secure function returns data coming through an inbound share.
What will happen if a Data Engineer tries to assign USAGE privileges on this function to an outbound share?

A. An error will be returned because the Engineer cannot share data that has already been shared.

B. An error will be returned because only views and secure stored procedures can be shared.

C. An error will be returned because only secure functions can be shared with inbound shares.

D. The Engineer will be able to share the secure function with other accounts.

**Correct Answer:** *A*

*Community vote distribution*

A (100%)

**azure_bimonster** 2 months ago

**Selected Answer: A**

Correct answer would be A.
A share cannot be shared again.

upvoted 1 times

Question #33 *Topic 1*

Which functions will compute a 'fingerprint' over an entire table, query result, or window to quickly detect changes to table contents or query results? (Choose two.)

A. HASH(*)

B. HASH_AGG(*)

C. HASH_AGG(<expr>, <expr>)

D. HASH_AGG_COMPARE(*)

E. HASH_COMPARE(*)

**Correct Answer:** *BD*

*Community vote distribution*

BC (100%)

**whiteshiba** 1 month, 3 weeks ago

No hash_agg_compare function

upvoted 1 times

**stopthisnow** 2 months, 3 weeks ago

**Selected Answer: BC**

https://docs.snowflake.com/en/sql-reference/functions/hash_agg
There is no hash_agg_compare function.

upvoted 3 times

Question #34 *Topic 1*

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

**Correct Answer:** *C*

*Community vote distribution*

C (100%)

**BigDataBB** 2 weeks, 3 days ago

**Selected Answer: C**

An external table is a Snowflake feature that allows you to query data stored in an external stage as if the data were inside a table in Snowflake. The external stage is not part of Snowflake, so Snowflake does not store or manage the stage.
https://docs.snowflake.com/en/user-guide/tables-external-intro

So the correct answer is C

upvoted 1 times

**prshntdxt7** 3 weeks, 2 days ago

**Selected Answer: C**

C. External stages only; from any region, and any cloud provider

External stages in Snowflake allow you to load or unload data from cloud-based storage locations (e.g., Amazon S3, Azure Blob Storage, Google Cloud Storage) without using a Snowflake stage. External stages provide flexibility in terms of storage locations and cloud providers.

upvoted 1 times

**stopthisnow** 2 months, 3 weeks ago

**Selected Answer: C**

C makes sense. External stage can retrieve from any region and different cloud providers

upvoted 1 times

Question #35 *Topic 1*

A Data Engineer wants to check the status of a pipe named my_pipe. The pipe is inside a database named test and a schema named Extract (case-sensitive).
Which query will provide the status of the pipe?

A. SELECT SYSTEM$PIPE_STATUS("test.'extract'.my_pipe");

B. SELECT SYSTEM$PIPE_STATUS('test."Extract".my_pipe');

C. SELECT * FROM SYSTEM$PIPE_STATUS('test."Extract".my_pipe');

D. SELECT * FROM SYSTEM$PIPE_STATUS("test.'extract'.my_pipe");

**Correct Answer:** *B*

*Community vote distribution*

B (100%)

**prshntdxt7** 3 weeks, 2 days ago

**Selected Answer: B**

The correct query to check the status of a pipe named my_pipe inside a database named test and a case-sensitive schema named Extract is:

B. SELECT SYSTEM$PIPE_STATUS('test."Extract".my_pipe');

This query uses double quotes around the schema name "Extract" to ensure that the case sensitivity is preserved. Additionally, the single quotes around the pipe name my_pipe and the dot notation for specifying the database and schema are correct in Snowflake SQL syntax.

upvoted 1 times

**stopthisnow** 2 months, 3 weeks ago

**Selected Answer: B**

SELECT SYSTEM$PIPE_STATUS('mydb.myschema.mypipe');

upvoted 2 times

**SV1122** 1 month, 3 weeks ago

https://docs.snowflake.com/en/sql-reference/functions/system_pipe_status

upvoted 1 times

Question #36 *Topic 1*

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

A. Create a share within Company A's account and add Company B's account as a recipient of that share.

B. Create a share within Company A's account, and create a reader account that is a recipient of the share. Grant Company B access to the reader account.

C. Use database replication to replicate Company A's data into Company B's account. Create a share within Company B's account and grant users within Company B's account access to the share.

D. Create a new account within Company A's organization in the same cloud provider and region as Company B's account. Use database replication to replicate Company A's data to the new account. Create a share within the new account, and add Company B's account as a recipient of that share.

E. Create a separate database within Company A's account to contain only those data sets they wish to share with Company B. Create a share within Company A's account and add all the objects within this separate database to the share. Add Company B's account as a recipient of the share.

**Correct Answer:** *AB*

*Community vote distribution*

AD (100%)

**BigDataBB** 4 days, 7 hours ago

https://docs.snowflake.com/en/guides-overview-sharing

Data Sharing Mechanism Share With Whom?
Listing One or more accounts in any region
Direct share One or more accounts in your region

upvoted 1 times

**BigDataBB** 2 weeks, 3 days ago

**Selected Answer: AD**

By documentation: https://docs.snowflake.com/guides-overview-sharing#options-for-sharing
CASE A: by listing
CASE D: by replication

upvoted 1 times

**stopthisnow** 2 months, 3 weeks ago

**Selected Answer: AD**

Company A must replicate data to the same region as Company B first. Since company B already has a Snowflake account, data can be imported from a Share setup by Company A. "Reader Account" option is not valid

upvoted 1 times

Question #37 *Topic 1*

A Data Engineer is trying to load the following rows from a CSV file into a table in Snowflake with the following structure:

```
CUSTOMERID,ADDRESS,REGISTERDT
      30 Ford Walk, Dante, Rhode Island, 366",2014-02-08
      14 Monroe Street, Kersey, Nevada, 6384",2021-04-19
789,"783 Gate Ave, Edgewater, New York, 1757",2020-07-03
```

| | |
|---|---|
| CUSTOMERID | NUMBER(38,0) |
| ADDRESS | VARCHAR(255) |
| REGISTERDT | DATE |

The engineer is using the following COPY INTO statement:

```
copy into stgCustomer
from @csv_stage/address.csv.gz
file_format = (type = CSV skip_header = 1);
```

However, the following error is received:

Number of columns in file (6) does not match that of the corresponding table (3), use file format option error_on_column_count_mismatch=false to ignore this error File 'address.csv.gz', line 3, character 1 Row 1 starts at line 2, column "STGCUSTOMER"[6] If you would like to continue loading when an error is encountered, use other values such as 'SKIP_FILE' or 'CONTINUE' for the ON_ERROR option.

Which file format option should be used to resolve the error and successfully load all the data into the table?

A. ESCAPE_UNENCLOSED FIELD = '\\'

B. ERROR_ON_COLUMN_COUNT_MISMATCH = FALSE

C. FIELD_DELIMITER = ','

D. FIELD_OPTIONALLY_ENCLOSED_BY = ""

**Correct Answer:** *D*

*Community vote distribution*

D (100%)

**ecvdata** 2 weeks, 6 days ago

https://docs.snowflake.com/en/user-guide/data-unload-considerations
FIELD_OPTIONALLY_ENCLOSED_BY = 'character' | NONE
Use this option to enclose strings in the specified character: single quote ('), double quote ("), or NONE.

upvoted 1 times

**stopthisnow** 2 months, 3 weeks ago

**Selected Answer: D**

D makes sense. Due to commas in the column value, the column needs to be enclosed by " to be treated as a single column.

upvoted 1 times

Question #38 *Topic 1*

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.

**Correct Answer:** *D*

*Community vote distribution*

D (100%)

**Snow_P** 2 months, 2 weeks ago

**Selected Answer: D**

https://docs.snowflake.com/en/user-guide/data-load-considerations-prepare

upvoted 1 times

**stopthisnow** 2 months, 3 weeks ago

**Selected Answer: D**

Various tools can aggregate and batch data files. One convenient option is Amazon Kinesis Firehose. Firehose allows defining both the desired file size, called the buffer size, and the wait interval after which a new file is sent (to cloud storage in this case), called the buffer interval. For more information, see the Kinesis Firehose documentation. If your source application typically accumulates enough data within a minute to populate files larger than the recommended maximum for optimal parallel processing, you could decrease the buffer size to trigger delivery of smaller files. Keeping the buffer interval setting at 60 seconds (the minimum value) helps avoid creating too many files or increasing latency.

upvoted 1 times

Question #39 *Topic 1*

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

A. CURRENT_ROLE

B. INVOKER_ROLE

C. IS_ROLE_IN_SESSION

D. IS_GRANTED_TO_INVOKER_ROLE

**Correct Answer:** *D*

*Community vote distribution*

D (100%)

**Snow_P** 2 months, 2 weeks ago

**Selected Answer: D**

More info here https://docs.snowflake.com/en/user-guide/security-column-intro

upvoted 1 times

**stopthisnow** 2 months, 3 weeks ago

**Selected Answer: D**

https://docs.snowflake.com/en/sql-reference/functions/is_granted_to_invoker_role

upvoted 1 times

Question #40 *Topic 1*

Assuming a Data Engineer has all appropriate privileges and context, which statements would be used to assess whether the User-Defined Function (UDF), MYDATABASE.SALES.REVENUE_BY_REGION, exists and is secure? (Choose two.)

A. SHOW USER FUNCTIONS LIKE 'REVENUE_BY_REGION' IN SCHEMA SALES;

B. SELECT IS_SECURE FROM SNOWFLAKE.INFORMATION_SCHEMA.FUNCTIONS WHERE FUNCTION_SCHEMA = 'SALES' AND FUNCTION_NAME = 'REVENUE_BY_REGION';

C. SELECT IS_SECURE FROM INFORMATION_SCHEMA.FUNCTIONS WHERE FUNCTION_SCHEMA = 'SALES' AND FUNCTION_NAME = 'REVENUE_BY_REGION';

D. SHOW EXTERNAL FUNCTIONS LIKE 'REVENUE_BY_REGION' IN SCHEMA SALES;

E. SHOW SECURE FUNCTIONS LIKE 'REVENUE_BY_REGION' IN SCHEMA SALES;

**Correct Answer:** *AC*

*Community vote distribution*

AC (100%)

**stopthisnow** 2 months, 3 weeks ago

**Selected Answer: AC**

SHOW USER FUNCTIONS LIKE 'DAY_NAME_ON' IN SCHEMA DEMO_SCHEMA;
SELECT IS_SECURE FROM INFORMATION_SCHEMA.FUNCTIONS WHERE FUNCTION_SCHEMA = 'DEMO_SCHEMA' AND FUNCTION_NAME = 'DAY_NAME_ON';

upvoted 1 times

Question #41 *Topic 1*

A Data Engineer has written a stored procedure that will run with caller's rights. The Engineer has granted ROLEA the right to use this stored procedure.
What is a characteristic of the stored procedure being called using ROLEA?

A. The stored procedure must run with caller's rights; it cannot be converted later to run with owner's rights.

B. If the stored procedure accesses an object that ROLEA does not have access to, the stored procedure will fail.

C. The stored procedure will run in the context (database and schema) where the owner created the stored procedure.

D. ROLEA will not be able to see the source code for the stored procedure, even though the role has usage privileges on the stored procedure.

**Correct Answer:** *B*

*Community vote distribution*

B (100%)

**stopthisnow** 2 months, 3 weeks ago

**Selected Answer: B**

A caller's rights stored procedure runs with the database privileges of the role that called the stored procedure. Any statement that the caller could not execute outside the stored procedure cannot be executed inside the stored procedure, either. For example, if the role named "Nurse" does not have privileges to delete rows from the medical_records table, then if a user with the role "Nurse" calls a caller's rights stored procedure that tries to delete rows from that table, the stored procedure will fail.

https://docs.snowflake.com/en/developer-guide/stored-procedure/stored-procedures-rights#caller-s-rights-stored-procedures

upvoted 1 times

Question #42 *Topic 1*

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 SfDate can be bound.

D. Users are restricted from binding JavaScript variables because they create SQL injection attack vulnerabilities.

**Correct Answer:** *C*

*Community vote distribution*

C (100%)

**prshntdxt7** 3 weeks, 2 days ago

**Selected Answer: C**

C. Only JavaScript variables of type number, string, and SfDate can be bound.

In Snowflake, when using binding variables in JavaScript stored procedures, only JavaScript variables of certain types can be bound. These types are typically limited to number, string, and SfDate.

Options A and B are incorrect because not all types of JavaScript variables or all Snowflake first-class objects can be bound. There are restrictions on the types of variables that can be used as binding variables.

Option D is incorrect because users are not restricted from binding JavaScript variables due to SQL injection concerns. The use of binding variables is a good practice to prevent SQL injection vulnerabilities as it helps separate SQL logic from user input, making it more secure.

upvoted 1 times

**Snow_P** 2 months, 2 weeks ago

**Selected Answer: C**

https://docs.snowflake.com/en/developer-guide/stored-procedure/stored-procedures-javascript

upvoted 3 times

**stopthisnow** 2 months, 3 weeks ago

**Selected Answer: C**

The data type of the variable should be appropriate for the use of the value in the SQL statement. Currently, only JavaScript variables of type number, string, and SfDate can be bound.

upvoted 3 times

Question #43 *Topic 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.

**Correct Answer:** *B*

*Community vote distribution*

B (100%)

**prshntdxt7** 3 weeks, 2 days ago

**Selected Answer: B**

B. Business users who need fast response times using highly selective filters.

The search optimization service in Snowflake is designed to improve query performance, especially for scenarios involving highly selective filters and search operations. Business users who require fast response times and need to perform searches with highly selective filters would benefit the most from the search optimization service. This feature helps optimize queries with complex filtering conditions, making it well-suited for scenarios where users need to efficiently retrieve specific subsets of data.

upvoted 1 times

**Snow_P** 2 months, 2 weeks ago

**Selected Answer: B**

https://docs.snowflake.com/en/user-guide/search-optimization-service

upvoted 2 times

**stopthisnow** 2 months, 3 weeks ago

**Selected Answer: B**

B makes sense

upvoted 1 times

Question #44 *Topic 1*

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.

**Correct Answer:** *B*

*Community vote distribution*

B (100%)

**stopthisnow** 2 months, 3 weeks ago

**Selected Answer: B**

B is correct. Running SELECT statement on does not change the offset. A DML statement must be run to update the offset

upvoted 2 times

Question #45 *Topic 1*

At what isolation level are Snowflake streams?

A. Snapshot

B. Repeatable read

C. Read committed

D. Read uncommitted

**Correct Answer:** *B*

*Community vote distribution*

B (100%)

**Snow_P** 2 months, 2 weeks ago

**Selected Answer: B**

https://docs.snowflake.com/en/user-guide/streams-intro

upvoted 1 times

**stopthisnow** 2 months, 3 weeks ago

**Selected Answer: B**

Streams support repeatable read isolation. In repeatable read mode, multiple SQL statements within a transaction see the same set of records in a stream. This differs from the read committed mode supported for tables, in which statements see any changes made by previous statements executed within the same transaction, even though those changes are not yet committed.

upvoted 2 times

Question #46 *Topic 1*

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

**Correct Answer:** *A*

*Community vote distribution*

A (100%)

**stopthisnow** 2 months, 3 weeks ago

**Selected Answer: A**

https://docs.snowflake.com/en/sql-reference/external-functions-security

upvoted 1 times

Question #47 *Topic 1*

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 requirements are:
Raise an exception if the specified database, schema, or warehouse does not exist

Improve download performance -
Which parameters of the connect function should be used? (Choose two.)

A. authenticator

B. arrow_number_to_decimal

C. client_prefetch_threads

D. client_session_keep_alive

E. validate_default_parameters

**Correct Answer:** *CE*

*Community vote distribution*

CE (100%)

**stopthisnow** 2 months, 3 weeks ago

**Selected Answer: CE**

client_prefetch_threads: Number of threads used to download the results sets (4 by default). Increasing the value improves fetch performance but requires more memory.

validate_default_parameters: Raise an exception if the specified database, schema, or warehouse doesn't exist.

upvoted 2 times

Question #48 *Topic 1*

A Data Engineer wants to centralize grant management to maximize security. A user needs OWNERSHIP on a table in 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.

**Correct Answer:** *D*

*Community vote distribution*

D (100%)

**stopthisnow** 2 months, 3 weeks ago

**Selected Answer: D**

D is correct. The owner of the table won't be able to grant access to other roles etc

upvoted 1 times

Question #49 *Topic 1*

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. 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.

D. 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.

**Correct Answer:** *B*

*Community vote distribution*

B (100%)

**stopthisnow** 2 months, 3 weeks ago

**Selected Answer: B**

B seems like a good option

upvoted 1 times

Question #50 *Topic 1*

What are characteristics of Snowpark Python packages? (Choose three.)

A. Third-party packages can be registered as a dependency to the Snowpark session using the session.import() method.

B. Python packages can access any external endpoints.

C. Python packages can only be loaded in a local environment.

D. Third-party supported Python packages are locked down to prevent hitting.

E. The SQL command DESCRIBE FUNCTION will list the imported Python packages of the Python User-Defined Function (UDF).

F. Querying information_schema.packages will provide a list of supported Python packages and versions.

**Correct Answer:** *AEF*

*Community vote distribution*

AEF (100%)

**BigDataBB** 2 weeks ago

**Selected Answer: AEF**

https://docs.snowflake.com/en/developer-guide/udf/python/udf-python-packages?utm_source=legacy&utm_medium=serp&utm_term=%22DESCRIBE+FUNCTION%22+python+package

upvoted 1 times

**stopthisnow** 2 months, 3 weeks ago

**Selected Answer: AEF**

```
select * from information_schema.packages where language = 'python'
and PACKAGE_NAME LIKE '%pandas%'

DESC FUNCTION multiply(number, number);

-----------+---------------------------------+
property | value |
-----------+---------------------------------+
signature | (a NUMBER(38,0), b NUMBER(38,0)) |
returns | NUMBER(38,0) |
language | SQL |
body | a * b |
-----------+---------------------------------+
```

upvoted 1 times

Question #51 *Topic 1*

While running an external function, the following error message is received:
Error: Function received the wrong number of rows
What is causing this to occur?

A. External functions do not support multiple rows.

B. Nested arrays are not supported in the JSON response.

C. The JSON returned by the remote service is not constructed correctly.

D. The return message did not produce the same number of rows that it received.

**Correct Answer:** *D*

*Community vote distribution*

D (100%)

**stopthisnow** 2 months, 3 weeks ago

**Selected Answer: D**

The remote service tried to return more or fewer rows than it received. Even though the function is nominally scalar, it might receive multiple rows in the body field of the event parameter, and should return exactly as many rows as it received.

upvoted 2 times

Question #52 *Topic 1*

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 SEED value 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 different results, because the query uses * instead of an explicit column list.

**Correct Answer:** *B*

*Community vote distribution*

B (100%)

**stopthisnow** 2 months, 3 weeks ago

**Selected Answer: B**

If a table does not change, and the same seed and probability are specified, SAMPLE generates the same result.

upvoted 1 times

Question #53 *Topic 1*

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.

**Correct Answer:** *D*

*Community vote distribution*

D (100%)

**stopthisnow** 2 months, 3 weeks ago

**Selected Answer: D**

D makes sense. Shouldn't have to use loops in SQL

upvoted 2 times

Question #54 *Topic 1*

A company is using Snowpipe to bring in millions of rows every day of Change Data Capture (CDC) into a Snowflake staging table on a real-time basis. The CDC needs to get processed and combined with other data in Snowflake and land in a final table as part of the full data pipeline.
How can a Data Engineer MOST efficiently process the incoming CDC on an ongoing basis?

A. Create a stream on the staging table and schedule a task that transforms data from the stream, only when the stream has data.

B. Transform the data during the data load with Snowpipe by modifying the related COPY INTO statement to include transformation steps such as CASE statements and JOINS.

C. Schedule a task that dynamically retrieves the last time the task was run from information_schema.task_history and use that timestamp to process the delta of the new rows since the last time the task was run.

D. Use a CREATE OR REPLACE TABLE AS statement that references the staging table and includes all the transformation SQL. Use a task to run the full CREATE OR REPLACE TABLE AS statement on a scheduled basis.

**Correct Answer:** *A*

*Community vote distribution*

A (100%)

**stopthisnow** 2 months, 3 weeks ago

**Selected Answer: A**

A makes sense.
D is definitely wrong. Can't drop and create table every time, it will also loose previous data. Needs to be incremental

upvoted 1 times

Question #55 *Topic 1*

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

A. Enable clustering on the table.

B. Increase the warehouse size.

C. Rewrite the query to remove the CTEs.

D. Switch to a multi-cluster virtual warehouse.

**Correct Answer:** *B*

*Community vote distribution*

B (100%)

**stopthisnow** 2 months, 3 weeks ago

**Selected Answer: B**

B makes sense

upvoted 2 times

Question #56 *Topic 1*

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

A. ACCOUNTADMIN

B. SECURITYADMIN

C. SYSADMIN

D. USERADMIN

**Correct Answer:** *C*

*Community vote distribution*

C (100%)

**stopthisnow** 2 months, 3 weeks ago

**Selected Answer: C**

Snowflake recommends creating a hierarchy of custom roles, with the top-most custom role assigned to the system role SYSADMIN

upvoted 1 times

Question #57 *Topic 1*

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

A. initialize()

B. processRow()

C. handler()

D. finalize()

**Correct Answer:** *B*

*Community vote distribution*

B (100%)

**Snow_P** 2 months, 2 weeks ago

**Selected Answer: B**

https://docs.snowflake.com/en/developer-guide/udf/javascript/udf-javascript-tabular-functions

upvoted 2 times

**stopthisnow** 2 months, 3 weeks ago

**Selected Answer: B**

The defined object must include a callback function named processRow()

upvoted 2 times

Question #58 *Topic 1*

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)

**Correct Answer:** *C*

*Community vote distribution*

C (100%)

**Snow_P** 2 months, 2 weeks ago

**Selected Answer: C**

https://docs.snowflake.com/en/sql-reference/external-functions

upvoted 1 times

**stopthisnow** 2 months, 3 weeks ago

**Selected Answer: C**

c makes sense

upvoted 1 times

Question #59 *Topic 1*

A Data Engineer needs to know the details regarding the micro-partition layout for a table named Invoice using a built-in function.
Which query will provide this information?

A. SELECT SYSTEM$CLUSTERING_INFORMATION('Invoice');

B. SELECT $CLUSTERING_INFORMATION('Invoice');

C. CALL SYSTEM$CLUSTERING_INFORMATION('Invoice');

D. CALL $CLUSTERING_INFORMATION('Invoice');

**Correct Answer:** *A*

*Community vote distribution*

A (100%)

**stopthisnow** 2 months, 3 weeks ago

**Selected Answer: A**

SYSTEM$CLUSTERING_INFORMATION
Returns clustering information, including average clustering depth, for a table based on one or more columns in the table.

upvoted 1 times

Question #60 *Topic 1*

A Data Engineer would like to define a file structure for loading and unloading data.
Where can the file structure be defined? (Choose three.)

A. COPY command

B. MERGE command

C. FILE FORMAT object

D. PIPE object

E. STAGE object

F. INSERT command

**Correct Answer:** *ACE*

**Snow_P** 2 months, 2 weeks ago

and stage object

upvoted 1 times

**Snow_P** 2 months, 2 weeks ago

Copy into and File format but not sure which is the 3rd option

upvoted 2 times

Question #61 *Topic 1*

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.

**Correct Answer:** *A*

*Community vote distribution*

D (100%)

**prshntdxt7** 3 weeks, 2 days ago

**Selected Answer: D**

A. Creating a DataFrame from a table will start a virtual warehouse.

Incorrect. Creating a DataFrame from a table does not necessarily start a virtual warehouse. It's a common operation that doesn't trigger the initiation of a virtual warehouse.
B. Creating a DataFrame from a staged file with the read() method will start a virtual warehouse.

Incorrect. Similar to option A, creating a DataFrame from a staged file using the read() method does not trigger the initiation of a virtual warehouse.
C. Transforming a DataFrame with methods like replace() will start a virtual warehouse.

Incorrect. DataFrame transformations, such as using the replace() method, do not start a virtual warehouse. These operations are performed locally on the DataFrame.
D. Calling a Snowpark stored procedure to query the database with session.call() will start a virtual warehouse.

upvoted 2 times

**Snow_P** 2 months, 2 weeks ago

**Selected Answer: D**

ABC are all transformations so D

upvoted 2 times

**Snow_P** 2 months, 2 weeks ago

**Selected Answer: D**

I think D

upvoted 1 times

Question #62 *Topic 1*

Database XYZ has the data_retention_time_in_days parameter set to 7 days and table XYZ.public.ABC has the data_retention_time_in_days set to 10 days.
A Developer accidentally dropped the database containing this single table 8 days ago and just discovered the mistake.
How can the table be recovered?

A. undrop database xyz;

B. create table abc_restore as select * from xyz.public.abc at (offset => -60*60*24*8);

C. create table abc_restore clone xyz.public.abc at (offset => -3600*24*8);

D. Create a Snowflake Support case to restore the database and table from Fail-safe.

**Correct Answer:** *A*

*Community vote distribution*

D (100%)

**prshntdxt7** 3 weeks, 2 days ago

**Selected Answer: D**

A.Snowflake does not provide a direct undrop command for databases or tables.

B.The at (offset => -60*60*24*8) syntax is not a valid way to recover dropped tables in Snowflake.
The at clause is used for temporal queries, and it doesn't have the capability to recover dropped tables.

C.However, it does not have the ability to recover dropped tables.
The at (offset => -3600*24*8) part is not a valid syntax for recovering dropped tables.

D. Create a Snowflake Support case to restore the database and table from Fail-safe.
This is the correct option. Snowflake's Fail-safe feature is designed for recovering dropped tables
or databases within a specified retention period.
By creating a support case, you can request assistance from Snowflake support to restore the
dropped database and table from Fail-safe, ensuring data recovery within the allowed retention time.

upvoted 1 times

**Maicas** 1 month ago

**Selected Answer: D**

Just like Snow_P said, D is the correct answer.

https://docs.snowflake.com/en/user-guide/data-time-travel#dropped-containers-and-object-retention-inheritance

upvoted 1 times

**Snow_P** 2 months, 2 weeks ago

**Selected Answer: D**

I think D, the retention period for the table although longer should not be honored when dropping the DB, the table should have been dropped first

upvoted 4 times

Question #63 *Topic 1*

A Data Engineer is building a set of reporting tables to analyze consumer requests by region for each of the Data Exchange offerings annually, as well as click-through rates for each listing.
Which views are needed MINIMALLY as data sources?

A. SNOWFLAKE.DATA_SHARING_USAGE.LISTING_EVENTS_DAILY

B. SNOWFLAKE.DATA_SHARING_USAGE.LISTING_CONSUMPTION_DAILY

C. SNOWFLAKE.DATA_SHARING_USAGE.LISTING_TELEMETRY_DAILY

D. SNOWFLAKE.ACCOUNT_USAGE.DATA_TRANSFER_HISTORY

**Correct Answer:** *AB*

*Community vote distribution*

BC (86%) 14%

**BigDataBB** 2 weeks ago

**Selected Answer: B**

In LISTING_CONSUMPTION_DAILY there is the column SNOWFLAKE_REGION, so this is the necessart view: https://docs.snowflake.com/en/sql-reference/data-sharing-usage/listing-consumption-daily

upvoted 1 times

**prshntdxt7** 3 weeks, 2 days ago

**Selected Answer: BC**

The minimal data sources needed for analyzing consumer requests by region for each Data Exchange offering annually, as well as click-through rates for each listing, would likely include:

B. SNOWFLAKE.DATA_SHARING_USAGE.LISTING_CONSUMPTION_DAILY
C. SNOWFLAKE.DATA_SHARING_USAGE.LISTING_TELEMETRY_DAILY

These two tables provide information related to listing consumption and telemetry, which are crucial for understanding consumer requests and click-through rates for each listing. The other options (A and D) seem less relevant to the specific requirements mentioned.

upvoted 1 times

**stopthisnow** 2 months, 2 weeks ago

**Selected Answer: BC**

A looks wrong. It gives info about installations and permission, not consumption.
https://docs.snowflake.com/en/sql-reference/data-sharing-usage/listing-events-daily

upvoted 2 times

**Snow_P** 2 months, 2 weeks ago

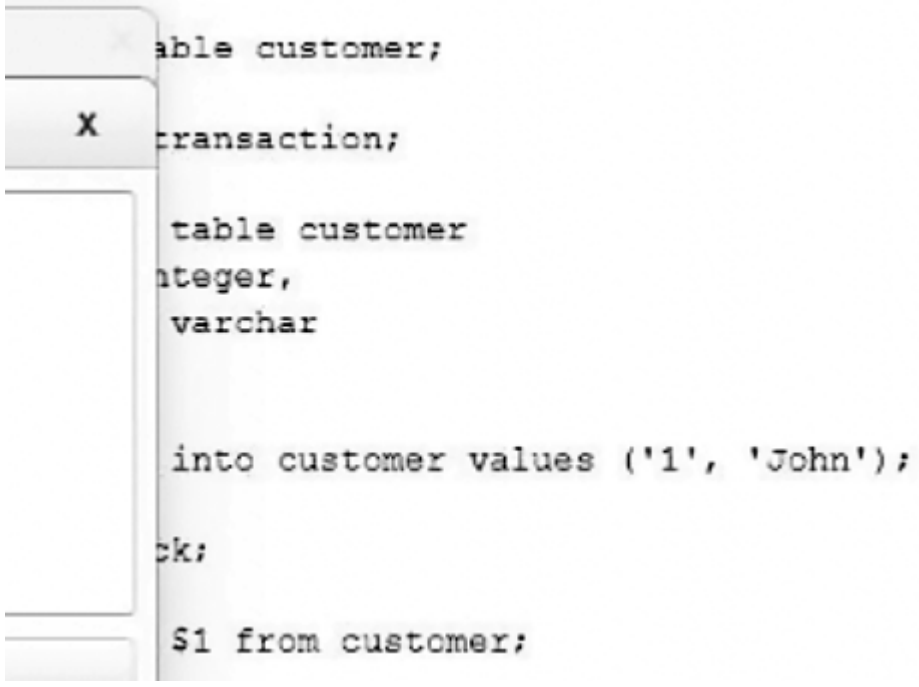
**Selected Answer: BC**

For sure C but I am wondering between A & B. Overall I think B&C

upvoted 3 times

Question #64 *Topic 1*

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




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

**Correct Answer:** *A*

*Community vote distribution*

C (50%) A (50%)

**BigDataBB** 2 weeks ago

is impossible to read the code : (((((

upvoted 1 times

**Maicas** 1 month, 1 week ago

**Selected Answer: C**

Asumming the table is created after the transaction has begun, the DDL operation will autocommit the transaction and the Rollback won't have any effect.

https://docs.snowflake.com/en/sql-reference/transactions#ddl

upvoted 1 times

**azure_bimonster** 2 months ago

**Selected Answer: A**

I assume there is a Rollback in that scripting, if that's the case A would be correct.

upvoted 1 times

Question #65 *Topic 1*

A Data Engineer defines the following masking policy:

```
CREATE MASKING POLICY name_policy AS (val string) RETURNS string ->
  CASE WHEN current_role() IN ('ADMIN') THEN val
  ELSE '******'
```

The policy must be applied to the full_name column in the customer table:

```
CREATE TABLE customer(
  first_name VARCHAR,
  last_name VARCHAR,
  full_name 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_name
ADD MASKING POLICY name_policy;

C. ALTER TABLE customer MODIFY COLUMN
first_name SET MASKING POLICY name_policy,
last_name SET MASKING POLICY name_policy;

D. ALTER TABLE customer MODIFY COLUMN
first_name ADD MASKING POLICY name_policy,
last_name SET MASKING POLICY name_policy;

**Correct Answer:** *B -*

*Community vote distribution*

C (100%)

**azure_bimonster** 2 months ago

**Selected Answer: C**

A virtual column is similar to a normal table column but it is defined by an expression. The result of evaluation of this expression becomes the value of the column. If we look at full_name column it came with concat function that was used making it as a virtual column. So, C would be correct here.

upvoted 2 times

**sajeski** 2 months, 1 week ago

**Selected Answer: C**

Masking policy cannot be attached to a VIRTUAL_COLUMN column. so it has to be C.

upvoted 2 times

**Snow_P** 2 months, 2 weeks ago

For sure it's SET not ADD but wondering between A and C

upvoted 1 times