Vendor: Snowflake

Exam Code: DEA-C02

Exam Name: SnowPro Advanced: Data Engineer Certification Exam

Version: DEMO

QUESTION 1
A Data Engineer made modifications to staged files and re-staged them. A pipe is running but the changed data is not appearing in the table.

What step should the Engineer take in order to see the changed data without duplicating data?

A. Run the `CREATE` or `REPLACE PIPE` syntax.
B. Re-run the pipe by using the `RELOAD PIPE` syntax.
C. Re-run the pipe by using the `ONLY_CHANGES` syntax.
D. Create new files with the changed data.

Answer: D
Explanation:
Snowpipe **tracks filenames** to prevent duplicate loads. If a file with the same name is re-staged, **Snowpipe ignores it**. To load the modified data, the engineer must **create new files with unique names**.

QUESTION 2
A Data Engineer has been asked by the development team to create a snapshot of production data from the ABC database into the developer account. None of the data is Personal Identifiable Information (PII) or sensitive. Both accounts are in the same region and use the same cloud provider.

The Engineer performs these steps:

In the production account:
```
use role accountadmin;
create database ABC_CLONE clone ABC;
create share ABC_SHARE;
grant usage on database ABC_CLONE to share ABC_SHARE;
alter share ABC_share add accounts = dev_account;
```

In the developer account:
```
use role accountadmin;
create database ABC_DEV FROM share prod_account.ABC_SHARE;
grant imported privileges on database ABC_DEV to role developers;
```

What is the reason the development team cannot see the data in the tables in the developer account?

A. The Data Engineer did not grant `usage` on the schema and `select` on the tables in the production account to the share.
B. The Data Engineer did not grant `imported privileges` to the schemas and tables in the developer account.
C. The Data Engineer did not grant `select` on the schema and tables in the share in the developer account.
D. The Data Engineer did not grant `usage` on the database in the developer account.

Answer: A
Explanation:
In Snowflake, **granting database usage alone is not enough** for shared data access. The **schema and tables within the shared database also need explicit privileges**. The engineer should have added:

```
GRANT USAGE ON SCHEMA ABC_CLONE.SCHEMA_NAME TO SHARE ABC_SHARE;
GRANT SELECT ON ALL TABLES IN SCHEMA ABC_CLONE.SCHEMA_NAME TO SHARE
ABC_SHARE;
```

Without these grants, the developer account **cannot see the tables** in the shared database.

QUESTION 3
A company would like to get data from a Snowflake Marketplace provider for data enrichment purposes. A Data Engineer needs to build a data pipeline that would blend existing data in the company's account with the provider's data, with some transformations. The provider's data is not available in the region where the company account resides.

How can the Data Engineer make the data available to the company for processing?

A. Create a new reader account in the region where the provider's data is available and set up a transformation job to unload transformed data into a stage. Set up a copy job between this stage and the company's stage and then ingest data into the company's account to blend with any existing data.
B. Create a new reader account in the region where the provider's data is available and set up a `COPY` job to unload data into a stage. Set up a data copy job between this stage and the company's stage. Then ingest the data into the company's account to transform and blend it with any existing data.
C. Create a new account in the region where the provider's data is available. Get data from the Marketplace, and create a share to the company's account. Then build a data pipeline to blend and transform the data.
D. Create a new account in the region where the provider's data is available. Get data from the Marketplace and replicate the data to the company's account. Then build a data pipeline to blend and transform the data.

Answer: D
Explanation:
Snowflake Marketplace data is only accessible within the same region as the provider's account. The best approach is to **create a new Snowflake account in the provider's region** and access the **Marketplace data** from there.

Cross-region data sharing is not supported, so **replicating the data** to the company's account ensures availability for transformation and blending.

QUESTION 4
A Data Engineer is creating a pipe by executing the following statement:

```
create or replace pipe mypipe auto_ingest=true as copy into mytable
from @mystage;
```

The Data Engineer adds a file data.csv.gz to the stage and it successfully loads into the table `mytable`.

What will cause the file data.csv.gz to be loaded a second time? (Choose two.)

A. `copy into mytable from @mystage;`
B.
```
truncate table mytable;
alter pipe mypipe refresh;
```

C. `drop table mytable;`
`create table mytable (coll string,...)`
D. `create or replace pipe mypipe auto_ingest=true as copy into mytable from @mystage;`
`alter pipe mypipe refresh;`
E. The file will automatically be reloaded a second time after eight days.

Answer: CD
Explanation:
Dropping and recreating `mytable` removes its metadata, including **tracking of previously loaded files**. When the new table is created, Snowpipe does **not recognize that `data.csv.gz` was already processed**, so it gets **reloaded**.

Replacing the pipe recreates it, resetting its metadata. Running `ALTER PIPE mypipe REFRESH` **forces Snowpipe to scan all staged files again**, causing `data.csv.gz` to be **reloaded**.

QUESTION 5
How can a directory table be used when working with unstructured data? (Choose two.)

A. File access URLs can be shared with collaborators and partners.
B. Data stored in an external stage can be queried as if the data was inside a table in Snowflake.
C. All unstructured files on a stage can be queried using the `SELECT * FROM DIRECTORY(@<stage_name>);` command.
D. All unstructured files on a stage can be queried using the `SELECT * FROM @<stage_name>;` command.
E. A directory table can be joined with other tables to produce a view of unstructured data that combines the file URLs with the file metadata.

Answer: AE
Explanation:
A directory table stores **metadata** (e.g., filenames, sizes, and URLs) of unstructured data in a **stage**, allowing file URLs to be **shared externally**.

A directory table can be **joined with structured tables** to combine file metadata with other data, enabling better organization and retrieval.

QUESTION 6
A Data Engineer needs to obtain details about data loading history for tables in a specific database and schema. The tables were loaded using Snowpipe in the last 24 hours.

How should the Engineer accomplish these tasks?

A. Use the `USAGE` privilege on the database and schema to use the copy history view in the information schema.
B. Use the `LOAD HISTORY` view using the usage privilege on the database and schema, and any privilege on the table.
C. Use the `MONITOR` privilege on the account and access the table details to use the copy history view in the information schema.
D. Use the `USAGE` privilege on the database and schema, select the database and the schema, and access the table details to use the copy history tab.

Answer: B
Explanation:
The LOAD_HISTORY view provides Snowpipe load details, and querying it requires USAGE on the database and schema plus any privilege on the target table.

QUESTION 7
A Data Engineer is developing a data pipeline to ingest data from an Amazon S3 bucket into a Snowflake table.

Files will be loaded into the S3 bucket randomly, with most files loaded during working hours. The files are in a compressed CSV format (`*.csv.gz`) and average 100 MB per file.

The data ingestion requirements are listed in order of priority:

1. MOST cost-effective method
2. LEAST amount of operational overhead
3. New data MUST be available within 10 minutes of the creation of a file

Which ingestion method will meet these requirements?

A. Snowpipe
B. Snowpipe Streaming
C. Snowflake Kafka Connector
D. The `COPY INTO <location>` command

Answer: A
Explanation:
Snowpipe provides serverless, event-driven file ingestion with minimal operational overhead and low per-file cost, delivering data well within the 10-minute requirement.

QUESTION 8
How should a Data Engineer build a pipeline that executes a stored procedure to parse and load unstructured data files as the files arrive in a stage?

A. Create a Snowpipe with the parameter `AUTO_INGEST = TRUE`.
B. Create a Snowpipe that includes the `INTEGRATION = '<procedure name>'` parameter.
C. Create a stage with parameter `ENCODING = '<procedure name>'` to call the procedure when flies land in the stage.
D. Create a stage with a directory table, a stream on the directory table, and a task referencing the stream.

Answer: D
Explanation:
Use a directory table on the stage to capture arriving files, create a stream on that table to detect new rows, and define a task that invokes the stored procedure whenever new files appear.

QUESTION 9
A Data Engineer needs to configure a task in Snowflake to automatically send error notifications in case of a data loading failure. The Engineer wants to use an existing task that inserts data into a table every 5 minutes. The Engineer set up a notification integration with Amazon Secure

Notification Service (SNS) named `my_notification_int`.

How should the task be updated to automatically send error notifications if it fails?

A. Add an `ALERT` clause in the task definition to link it to the notification service.
B. Add the command: `ALTER TASK MYTASK SET ERROR_INTEGRATION = 'my_notification_int';` to the notification.
C. Use the `NOTIFICATION INTEGRATION = 'my_notification_int'` parameter in the task configuration.
D. Add `NOTIFICATION_PROVIDER = 'my_notificacion_int'` to the task schedule.

Answer: B
Explanation:
Use ALTER TASK … SET ERROR_INTEGRATION = 'my_notification_int' so that SNS will be notified whenever the task encounters a failure.

QUESTION 10
A company using Snowflake works with a third party that provides economic data. The third-party provider recently moved their data to Snowflake.

How can the company access the third-party data in Snowflake with the LEAST amount of operational overhead?

A. Have the data provider enable database replication so the company can create a replica of the provider's database.
B. Have the data provider create a share object and add the company's Snowflake account to the share.
C. Have the data provider create a listing on the Snowflake Marketplace so the company can request access to the data.
D. Have the data provider create a new user and role specifically for the company so the company can log into the data provider's Snowflake account to access the data.

Answer: B
Explanation:
The most efficient way to share data with minimal operational overhead is for the provider to create a share object and add the consumer's Snowflake account to it. This uses Snowflake's secure data sharing feature without needing replication or user management.

QUESTION 11
A Data Engineer is trying to clone a very large table from a source environment that has set `DATA_RETENTION_TIME_IN_DAYS = 0`.

During the cloning operation, Data Manipulation Language (DML) transactions are executed on the source table, and the user receives the error "Data is not available."

How can this error be prevented, while ensuring successful cloning of the table? (Choose two.)

A. Increase the warehouse size to speed up the cloning process and avoid DML conflicts.
B. Set the `DATA_RETENTION_TIME_IN_DAYS = 1` for the table in the source environment before starting the cloning process, and revert it to `= 0` after the clone is completed.
C. Avoid all DML transactions on the source table until the cloning operation is complete.

D. Disable Time Travel on the cloned table to reduce storage costs once the cloning is complete.
E. Use a `CREATE TABLE AS SELECT STATEMENT` to create a new table, instead of creating a clone.

Answer: BC
Explanation:
Cloning relies on Time Travel to capture a consistent snapshot; setting DATA_RETENTION_TIME_IN_DAYS = 1 ensures the snapshot remains available through the clone operation.

Preventing DML on the source table during the clone avoids conflicts and "Data is not available" errors when Time Travel retention is minimal.

QUESTION 12
A company has deployed a data pipeline that streams customer transaction data from an on-premises PostgreSQL database to Snowflake for real-time analytics. A Data Engineer needs to configure a connector that will continuously and reliably transfer data to Snowflake.

Which connector configuration will meet these requirements, while providing MINIMAL latency, and OPTIMAL performance?

A. Configure the Snowflake Connector for Kafka to continuously stream data from the PostgreSQL database into Snowflake by configuring PostgreSQL as a Kafka producer.
B. Deploy the Snowflake JDBC driver to connect the PostgreSQL database directly to Snowflake, setting up a continuous query on the PostgreSQL database to push data to Snowflake.
C. Install the Snowflake Connector for Python on the PostgreSQL server to stream data directly to Snowflake, using Python scripts to manage the data transfer.
D. Configure the Snowflake Connector for Kafka on a Kafka Connect cluster to stream data from the PostgreSQL database to Kafka topics, which will be ingested using the Kafka connector.

Answer: D
Explanation:
Deploying the Snowflake Connector for Kafka on a Kafka Connect cluster to ingest from PostgreSQL CDC topics into Snowflake ensures a fault-tolerant, low-latency stream with built-in batching and exactly-once delivery semantics, delivering optimal performance.

QUESTION 13
How can a Data Engineer return all the errors encountered during a previous file upload that used the `COPY` command?

A. Query the `VALIDATE` information schema view.
B. Call the `VALIDATE` table function.
C. Call the `VALIDATE_PIPE_LOAD` table function.
D. Query the `VALIDATE` account usage schema view.

Answer: B
Explanation:
The VALIDATE table function returns detailed error rows for a prior COPY operation, including file name, row number, and error message.

QUESTION 14

A Data Engineer is cloning a production database for use as a development environment using these commands:

```
USE ROLE CLONE_ADMIN;
CREATE DATABASE DB_DEV CLONE DE_PROD AT (OFFSET => -1800);
```

What must the Engineer do to ensure the cloned environment is ready to use?

A. Run this command on all pipes in the DB_DEV to resume loading from them:
   `ALTER PIPE <name> SET PIPE_EXECUTION_PAUSED = FALSE;`
B. Run this command on all external stages in the DB_DEV to point them to development files:
   `ALTER STAGE <name> SET STORAGE_INTEGRATION = <dev_integration>;`
C. Run this command on all tables in the DB_DEV to revoke permissions from all production roles:
   `GRANT OWNERSHIP ON ALL TABLES IN DATABASE DB_DEV TO ROLE CLONE_ADMIN`
   `REVOKE CURRENT GRANTS;`
D. Run this command on all tables with sequences in the DB_DEV to reattach the sequences to the cloned tables:
   `ALTER TABLE <DB_DEV>.<PUBLIC>.<name>`
   `ALTER COLUMN <column_using_sequence> SET DEFAULT <DV_DEV>.<PUBLIC>.`
   `<name>.NEXTVAL;`

Answer: A
Explanation:
Cloned pipes are created in a paused state. You must run ALTER PIPE … SET PIPE_EXECUTION_PAUSED = FALSE to resume each pipe so that data loading continues in the dev clone.

QUESTION 15
A Data Engineer is setting up a data processing pipeline in Snowflake. The pipeline needs to process PDF files that are uploaded to a stage with a directory table enabled. The data must be extracted from the files, and then the processed data will be loaded into a table.

What should be done to ensure that the pipeline will automatically process any newly-added PDF files?

A. Configure Snowpipe to automatically load the PDF data directly into the table without the need for a stream or task.
B. Implement a custom Python script to continuously scan the stage and process any newly-created PDF files.
C. Use a materialized view to monitor the directory table for changes and trigger the processing task when new PDF files are added.
D. Create a stream on the directory table to track changes and use the stream to trigger a task to process any new PDF files.

Answer: D
Explanation:
By creating a stream on the directory table, Snowflake will track each new file arrival. A scheduled task can then query the stream and invoke your PDF-processing logic for only those new files, providing a fully automated pipeline.

Thank You for Trying Our Product

Passleader Certification Exam Features:

★ More than **99,900** Satisfied Customers Worldwide.

★ Average **99.9%** Success Rate.

★ **Free Update** to match latest and real exam scenarios.

★ **Instant Download** Access! No Setup required.

★ Questions & Answers are downloadable in **PDF** format and **VCE** test engine format.

★ Multi-Platform capabilities - **Windows, Laptop, Mac, Android, iPhone, iPod, iPad**.

★ **100%** Guaranteed Success or **100%** Money Back Guarantee.

★ **Fast**, helpful support **24x7**.

View list of all certification exams: http://www.passleader.com/all-products.html

10% Discount Coupon Code: **ASTR14**