

Exam Information

Average salary	\$90.501
Duration	120 minutes
Exam Guide	https://learn.snowflake.com/en/certifications/snowpro-core/
Format	Multiple Select, Multiple Choice, True/False questions
Language	English & Japanese
Number of Questions	100
Passing Score	80% or more
Price	175 USD
Recommended course	https://plazagonzalo.medium.com
Validity	2 years
Version	COF-C02

What is Snowflake?

Snowflake is a Data Solution provided as Software-as-a-Service (SaaS). **It's not available on-premise.** It combines a new SQL query engine with an innovative architecture natively designed for the cloud.

Use cases

Snowflake is optimal as Data Warehouse, Data Lake, Data Exchange, Data Apps, Data Science, and Data Engineering.

Snowflake Use Cases



Architecture

Shared-disk architectures Central data repository for persisted data accessible from all compute nodes in the platform.

Architecture (cont)

Shared-nothing architectures Each node in the virtual warehouse cluster stores a portion of the entire data set locally.

Hybrid of traditional shared-disk and shared-nothing database architectures.

Architecture Layers

Centralized Storage Layer Snowflake reorganizes that data into its internal optimized, compressed, columnar format.

Compute Layer Query execution is performed using virtual warehouses in the compute/processing layer

Cloud Services Layer Collection of services that coordinate activities across Snowflake, including Authentication, Access Control, etc.

Cloud Agnostic Layer It is only used the first time when we choose a provider.

Costs

Storage Costs Average amount of storage used per month, after compression

Compute Costs Billed by seconds with a one-minute minimum

Cloud Service Up to 10% of daily compute credits is included for free

Data Transfer Move or copy their data between regions or cloud providers

The Snowflake edition, warehouse size, number of clusters, and time that each server in each cluster runs determine the number of credits that a data warehouse consumes.



By cabanasj486

Not published yet.

Last updated 6th September, 2023.

Page 1 of 6.

Sponsored by **CrosswordCheats.com**

Learn to solve cryptic crosswords!

<http://crosswordcheats.com>

Capacity Options

On-Demand	Fixed rate for the consumed services
Pre-paid	Cheaper, but commitment to Snowflake

Micro-partitions

All data in Snowflake tables are automatically divided into micro-partitions, contiguous units of storage between 50 and 500MB of uncompressed data, organized in a columnar way. They are immutable, meaning they cannot be changed once created.

Pruning process

Technique to analyze the smallest number of micro-partitions to solve a query. It retrieves all the necessary data to give a solution without looking at all the micro-partitions, saving a lot of time to return the result. [You can find a real example here](#)

Load Data

Bulk Load	Loading batches of data from files already available at any stage into Snowflake tables
Continuous Load	Load small volumes of data (micro-batches) and incrementally make them available for analysis.

Default Roles

ACCOUNTADMIN	Top-level role
SECURITYADMIN	Manage users and roles
SYSADMIN	Create warehouses and databases (and other objects)
USERADMIN	User and role management
PUBLIC	Automatically granted to every user an role
CUSTOM	Create your own roles and assign the privileges that you want

Continuos Load

Snowpipe	Loading data when the files are available in any (internal/external) stage. 14 days of metadata.
----------	--

Continuos Load (cont)

Snowflake Connector for Kafka	Reads data from Apache Kafka topics and loads the data into a Snowflake table
-------------------------------	---

Third-Party Data Integration Tools	You can see the list at the following link .
------------------------------------	--

The most important part of this section is Snowpipe. You should use it for **small volume of frequent data, and you load it continuously** (micro-batches). It's **serverless**, which means that it doesn't use Virtual Warehouses.

It can detect new files by automating Snowpipe using cloud messaging, or by calling the Snowpipe REST endpoints.

Bulk Load

COPY INTO	Load data from any stage to an existing table. 64 days of metadata.
-----------	---

Some important considerations:

- 1) You cannot Load/Unload files from your Local Drive
- 2) Using the Snowflake UI, you can only Load 50MB files
- 3) Organizing input data by granular path can improve load performance
- 4) FORCE=True to copy the files again and ommit the 64 days of metadata.
- 5) PURGE = True removes the data files from the stage.
- 5) If there is any error, you can specify different options: ABORT_STATEMENT, CONTINUE, SKIP_FILE, SKIP_FILE_num, SKIP_FILE_num%.

Cache Strategies

Metadata Cache	Objects Information & Statistics.
Warehouse Cache	Attached SSD storage to a Warehouse. Information lost when the Warehouse is suspended.
Query Result Cache	It stores the results of our queries for 24 hours. If we perform the same query and the data hasn't changed, it will return the same result without using the Warehouse.

You can find a complete example of how to use the different cache strategies in the [following link](#).



By cabanasj486

Not published yet.

Last updated 6th September, 2023.

Page 2 of 6.

Sponsored by [CrosswordCheats.com](https://crosswordcheats.com)

Learn to solve cryptic crosswords!

<http://crosswordcheats.com>

Access Management Approaches

Discretionary Access Control (DAC)	Each object has an owner who can, in turn, grant access to that object
Role-Based Access Control (RBAC)	Access privileges are assigned to roles, which are, in turn, given to users

Access Management in Snowflake

User	Person or program
Role	Entity to which we grant privileges
Securable Object	Entity to which we can grant access
Privilege	Defined level of access to an object

Other Concepts

Partner Connect	Technology & Solution partners
Compliance	HITRUST / HIPAA, ISO/IEC 27001, FedRAMP Moderate, PCI-DSS, etc
Data Marketplace	For providers to buy or sell their datasets. Free, Personalized, and Paid Listings
Column Level Security	Dynamic Data Masking & External Tokenization

Data Sharing

Share	Snowflake objects that encapsulate all information required to share a database
Types of Shares	Outbound & Inbound
Producer	Snowflake account that creates shares and makes them available to other Snowflake accounts to consume
Consumer	Accounts that receive the share/data.

Data Sharing (cont)

Types of Consumers	Full account (existing Snowflake account), and Reader Account (share data with someone without Snowflake account).
--------------------	--

Shared data is instantaneous for consumers as no actual data is copied or transferred between accounts. For this reason, shared data is always up-to-date, and consumers don't pay for storage.

Streams

Definition Snowflake objects that record data manipulation language (DML) changes made to tables and views, including INSERTS, UPDATES, and DELETES, as well as metadata about each change

Storage They don't contain table data; they only store offsets

Types Standard, Append Only, and Insert Only

Columns METADATA\$ACTION, METADATA\$ISUPDATE, METADATA\$ROW_ID

Function that indicates whether a stream contains change data capture (CDC) records

Another important function is `SYSTEM$STREAM_HAS_DATA`, which that indicates whether a stream contains change data capture (CDC) records. You can see a example of how streams work in the [following link](#).

File Formats

Structured Data CSV. Fastest way to load data.

Semi-structured Data JSON, Parquet, XML, Avro, ORC

FLATTEN Convert semi-structured data to a relational representation

Sequences

Use case Generate unique numbers across sessions and statements

`nextval` Function to generate a set of distinct values



By cabanasj486

Not published yet.

Last updated 6th September, 2023.

Page 3 of 6.

Sponsored by **CrosswordCheats.com**

Learn to solve cryptic crosswords!

<http://crosswordcheats.com>

Editions

- Standard
- Enterprise
- Business Critical
- Virtual Private Snowflake

Snowflake Editions

The image contains two tables. The top table, titled 'Snowflake Features by Edition (B)', compares Standard, Enterprise, and Business Critical editions across features like Multi-Cluster, Full Lake, Streaming, Custom Default View Properties, Materialized Views, and Replication. The bottom table, titled 'Snowflake Features by Edition (B)', compares the same editions across features like Column Level Security, Query Statement Encryption, Failover/Block, Federated Authentication, Privilege Lock, and TDE (Secure Encryption). A 'Full Certified' logo is visible at the bottom left of the second table.

Cloud Providers

- Amazon Web Services
- Azure
- Google Cloud Platform

Connect to Snowflake

- Web Interface
- SnowSQL (CLI Client)
- ODBC
- JDBC
- SDK for Node, Python, Kafka, Go, and more!

Snowflake Objects

- Account: Must be unique.
- Warehouse: Virtual Machine to execute queries. Compute Part.
- Database: Logical Collection of Schemas.
- Schema: Logical Collection of Objects. The Public schema and the Information_Schema are created when creating a Database.

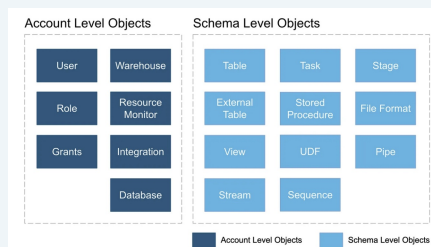
Main Objects contained within a Schema

- Tables: Contains all the data in the DB.
- Views: Virtual table defined by a query.

Main Objects contained within a Schema (cont)

- Stages: Location of data files in the Cloud Storage
- File Formats: Describes a set of staged data to access or load into Snowflake tables.
- Sequences: Like counters to create unique numbers.
- Pipes: Object that enables automatic loading of data from files as soon as they are available in a Stage.
- Stored Procedures & User-Defined Functions (UDF): Extend the system to perform operations in different programming languages.

Snowflake Objects



Types of tables

- Permanent
- Transient
- Temporary
- External

Types of views

- Regular
- Materialized
- Secure



Stages

Named External Stage
Named Internal Stage
User Internal Stage (@~)
Table Internal Stage (@%)

Storage Integrations will enable users to avoid supplying credentials when creating stages or when loading or unloading data. It's an object that stores a generated identity and access management (IAM) entity for your external cloud storage.

Stage Metadata

METADATA\$FILE-ENAME	Name of the staged data file the current row belongs to.
METADATA\$FILE_ROW_NUMBER	Row number for each record in the container staged data file.

Fail-Safe

Use Cases	It ensures historical data is protected in the event of a system failure or other catastrophic event
Retention Period	NON-CONFIGURABLE 7-day period
Example	No, you cannot recover this data alone; you MUST ask Snowflake support

Note: Fail-Safe requires additional storage, which will be reflected in your monthly storage charges

Zero-Copy Cloning

Use Cases	Create a snapshot of any table, schema, or Database
Cost	FREE, it doesn't consume storage. It does NOT duplicate data; it duplicates the metadata of the micro-partitions.
Other considerations	Privileges are not cloned. Data History is not cloned.

Note: When you modify some cloned data, it will consume storage because Snowflake has to recreate the micro-partitions, which will cost money.

Time Travel

Use Cases	Access historical data at any point within a defined period. Useful to restore tables.
Objects that we can restore	Databases, Schemas, and tables.
Retention Period	1 day by default, with a maximum of 90 days (Enterprise edition).
Ways to restore	By offset, query statement ID, or timestamp.
Example	<i>UNDROP TABLE mytable;</i>

Note: Time Travel requires additional storage, which will be reflected in your monthly storage charges

Extend Snowflake Functionality

Store Procedures	Extend Snowflake SQL by combining it with JavaScript
User-Defined Functions (UDFs)	Perform operations that are not available through Snowflake's built-in, system-defined functions. SQL, JavaScript, Java, and Python. It returns a single row
User-Defined Table Functions (UDTFs)	They can multiple rows for each input row (only difference with UDFs)
External Functions	They call code that is executed outside Snowflake

Tasks

Definition	Schedulable scripts that are run inside your Snowflake environment
When they run	Task run on a schedule
Execution	They execute a single SQL statement, including a call to a Stored Procedure
Duration	Maximum duration of 60 minutes by default



By cabanasj486

Not published yet.

Last updated 6th September, 2023.

Page 5 of 6.

Sponsored by **CrosswordCheats.com**

Learn to solve cryptic crosswords!

<http://crosswordcheats.com>

Tasks (cont)

Tree of tasks	Each task can have a maximum of 100 children tasks. A tree of tasks can have a maximum of 1000 tasks, including the root one.
Task History	Query the history of task usage within a specified date range
Serverless Tasks	Compute resources automatically scale up or down by Snowflake as required for each workload

Note: Snowflake ensures only one instance of a task with a schedule is executed at a given time. If a task is still running when the next scheduled execution time occurs, that scheduled time is skipped.

Transactions

ACID. Sequence of SQL statements that are committed or rolled back as a unit. Things we need to know for the exam:

- 1) Snowflake takes 4 hours to abort it if we do not abort it with the `SYSTEM$ABORT_TRANSACTION`
- 2) Each transaction has independent scope.
- 3) Snowflake does not support Nested Transactions

Resource Monitors

Credit Quota	Snowflake credits allocated to the monitor for the specified frequency interval
Monitor Level	Monitor the credit usage for your entire Account or individual Warehouses
Schedule	When the monitor is going to start monitoring
Actions	What to do when the threshold is reached. Notify (send notification), Notify & Suspend (suspend warehouse), or Notify & Suspend Immediately (kill query).

Resource monitors help you control costs and avoid unexpected credit usage caused by running data warehouses. You can impose limits on the number of credits that warehouses consume. They can only be created by AccountAdmins.

Data Warehouses Properties

Size	Impact the amount of time required to execute queries
Multi-Cluster Warehouses	Scale compute resources to manage query concurrency
Multi-Cluster Warehouses Modes	Maximized & Auto-scale
Scaling	Scale up/down to increase performance. Scale out/in to improve concurrency for users/queries.
Scaling policies	Standard & Economy
Auto Suspend & Auto Resume	Enabled by default.

A Data Warehouse is a cluster of computing resources in Snowflake that provides CPU, memory, and temporary storage to perform queries and DML operations. While a warehouse is running, it consumes Snowflake credits. It utilizes per-second billing (with a 60-second minimum each time the warehouse starts).

Other commands

PUT	UPLOAD files from a local directory/folder on a client machine into internal stages.
GET	DOWNLOAD files from a Snowflake internal stage into a directory/folder on a client machine
Example	<code>GET @my_int_stage file:///tmp/data/;</code>
You cannot use both of these commands with the Snowflake Web UI.	



By [cabanasj486](#)

Not published yet.

Last updated 6th September, 2023.

Page 6 of 6.

Sponsored by [CrosswordCheats.com](#)

Learn to solve cryptic crosswords!

<http://crosswordcheats.com>