

Chapter 2: Snowflake Architecture

Second Chapter of the Snowflake SnowPro Core Certification Complete Course.

This chapter will study in detail the architecture that Snowflake uses to provide this amazing service, apart from typical exam questions to prepare you for the Snowflake SnowPro Core Certification Exam. These are the key concepts that we are going to study:

1. Shared-Disk & Shared-Nothing
2. Snowflake Layers
3. Snowflake Objects
4. Typical Exam Questions about Snowflake Architecture

SHARED-DISK & SHARED-NOTHING

Snowflake's architecture is a hybrid of traditional shared-disk and shared-nothing database architectures. This is of utmost importance; **IT ALWAYS APPEARS IN THE EXAM.**

- **Similar to shared-disk architectures** → Snowflake uses a central data repository for persisted data accessible from all compute nodes in the platform.
- **Similar to shared-nothing architectures** → Snowflake processes queries using virtual warehouses where each node in the cluster stores a portion of the entire data set locally.



Shared Disk



Shared Nothing

Shared Disk & Shared Nothing Architectures.

This approach offers the data management simplicity of a shared-disk architecture but with the performance and scale-out benefits of a shared-nothing architecture.

SNOWFLAKE LAYERS

Snowflake's unique architecture consists of three essential layers, all of them with High Availability. The price is also charged separately for each layer. Let's see them:

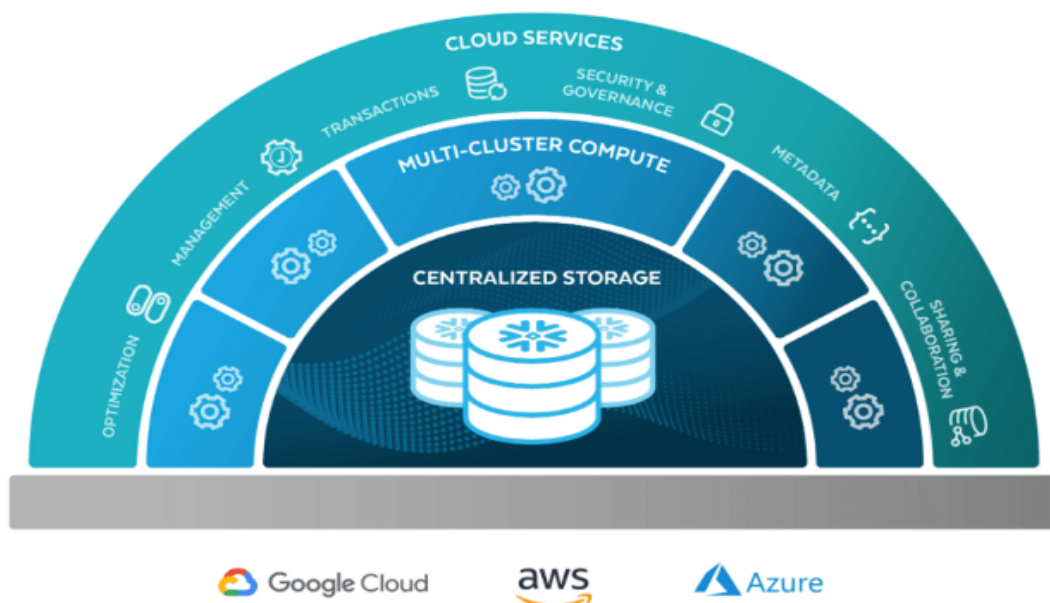
1) **Centralized Storage** → When data is loaded into Snowflake, Snowflake reorganizes that data into its internal optimized, compressed, columnar format in this layer.

Snowflake manages all aspects of how this data is stored.

2) **Compute** → Query execution is performed in the compute/processing layer using "virtual warehouses". Each virtual warehouse is an MPP compute cluster composed of multiple compute nodes allocated by Snowflake from a cloud provider.

3) **Cloud Services** → Collection of services that coordinate activities across Snowflake. It includes: Authentication, Infrastructure management, Metadata management, Query parsing and optimization and Access control.

4) **Cloud Agnostic Layer** → We could also talk about the fourth layer, also known as the Cloud Agnostic Layer. It is used only the first time when we choose a provider.



SNOWFLAKE OBJECTS

In this part of the chapter, we will review some of the most important objects part of the Snowflake system. We will see them in detail later in this course; this is just a brief explanation:

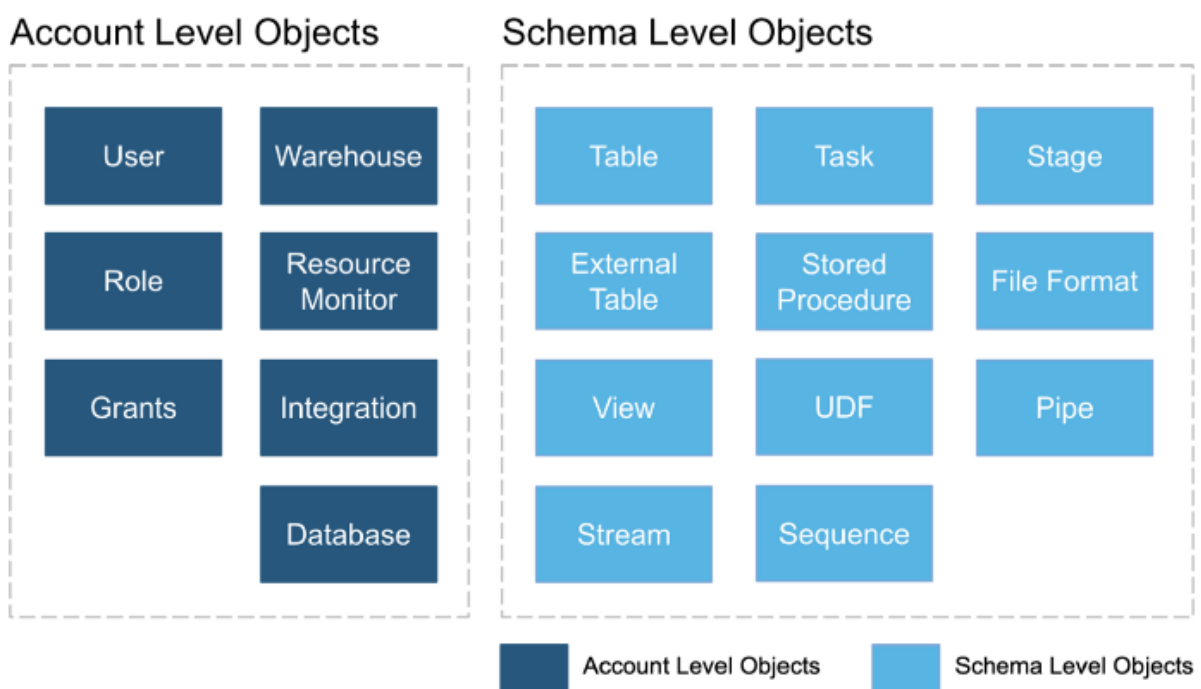
- **Account** → The name of the account must be unique within an organization, regardless of which Snowflake Region the account is in. An account could be, for example, "tn10000.eu-west-1.snowflakecomputing.com", although you can ask for a vanity address to convert the identifier into the name of your company, for example, "myrealcode.eu-west-1.snowflakecomputing.com".
- **Warehouse** → Warehouses are the compute part of the Snowflake engine. They are a set of virtual machines provided at the runtime to help execute a given query.
- **Database** → Logical collection of Schemas. Like any other Database that we already know. It's a logical container for storing other objects such as Tables and Views, organized in the user-defined Schemas.
- **Schema** → Logical collection of objects that provides a good facility for managing objects and their respective accesses. One schema belongs to a single Database. When a database is created, by default, two schemas are created with it: the Public schema, which defaults to any object created without specifying any schema, and the Information_Schema, which stores the metadata information.

Among other objects that we will see later in the course, Schemas contain the following objects:

- **Tables** → Database objects that contain all the data in a database.
- **Views** → Virtual table whose contents are defined by a query. We can perform a query and save the result in a view.
- **Stages** → Stage is the location of data files in cloud storage.
- **File Formats** → Pre-defined format structure that describes a set of staged data to access or load into Snowflake tables for CSV, JSON, AVRO, ORC, PARQUET, and XML input types.
- **Sequences** → You can generate unique numbers with sequences. They are like counters.

- **Pipes** → Special and unique type of object in Snowflake which enables automatic loading of data from files as soon as they are available in a Stage.
- **Stored Procedures & User Defined Functions (UDF)** → They let you extend the system to perform operations in SQL and JavaScript.

Objects can be at the Account level or the Schema level. In this image, I attach the types that exist, just in case they appear in the exam. What is the difference? The difference is where you store the objects. For example, if we create a new Role, it doesn't belong to a specific Schema; they are created for the entire account. On the other hand, when we create a table, they are stored inside a Schema. It's a straightforward and logical question that can appear in the SnowPro exam.



Snowflake Objects divided by Account or Schema Level.

TYPICAL EXAM QUESTIONS ABOUT SNOWFLAKE ARCHITECTURE

The three main layers of SnowFlare are

1. Extraction, Ingestion, Load

1. Database Storage, Compute & Cloud Services
2. Database, Virtual Warehouse, Data Experience

Solution: 1.

2. Select the term that is associated with the compute layer:

1. Query optimization
2. Query planning
3. Query processing

Solution: 3. You can find the name of the layers in different ways, like Query Processing for the Compute Layer.

3. Which of the following services are provided by the Cloud Services Layer?

1. Metadata Management
2. Authentication
3. Storage
4. Infrastructure Management
5. Query Execution

Solution: 1, 2, 4

4. What statement is true about Snowflake's unique architecture?

1. One Node Shared Data
2. Multi-Cluster Shared Data
3. One Node Private Data

Solution: 2.

5. What is the storage hierarchy in Snowflake?

1. Account → DB → Schemas → Objects
2. Account → Schemas → DB → Objects
3. Account → DB → Objects → Schemas

Solution: 1.

6. Can two Virtual Warehouses access the same data simultaneously without any contention issues?

1. True
2. False

Solution: 1. Indeed, we haven't seen this topic in-depth, but this is the main reason for the "Shared Data" concept. All the data is in the same place, and the Virtual Warehouse has access to it.

7. The interactions with data are initialized through the services layer?

1. True

2. ¹False

Solution: 1. Cloud services tie together all of the different components of Snowflake to process user requests, from login to query dispatch.

8. In which layer of Snowflake architecture is stored all security-related information?

1. Storage
2. Compute
3. Cloud Services
4. All of the above.

Solution: 3.

9. Can the table functions in INFORMATION_SCHEMA be used to return account-level usage and historical information for storage, warehouses, user logins, and queries?

1. True
2. False

Solution: 1.

Thanks for Reading!

Thanks for going through the doc.

21 Chapters notes with Q&A and 500 question sample sheet, cheat sheet for end moment preparation & dumps with detailed solution @1299 (500+ Q&As)(questions and detailed solutions) Get Interview assistance referrals and study materials.

Sample dumps :

https://drive.google.com/file/d/1epWk_iuaY4eKutNrg0dUyG_tFEvN30fA/view?usp=sharing
Let me know if this was helpful. If you have any concerns and suggestions please feel free to reach out.

LinkedIn: <https://www.linkedin.com/in/avinash-sharma-553378151/>