FROM THE AUTHOR OF 'MASTERING SNOWFLAKE SOLUTIONS'

Snowflake SnowPro Core STUDY GUIDE

Supercharge your career by starting your Snowflake journey

ADAM MORTON

INTELLIGEN

Snowflake SnowPro Core Study Guide

Supercharge your career by starting your Snowflake journey

(Book Preview - Chapter One)

Adam Morton

Compliments of 

KICK OFF YOUR LEARNING JOURNEY WITH SNOWFLAKE BY STUDYING FOR THE SNOWPRO CORE CERTIFICATION!

This certification allows you to validate your experience with one of the fastest growing cloud-based data platforms.

This chapter preview will form a key part of your preparation for the SnowPro Core certification by explaining the new concepts and terminology Snowflake introduces. The author, Adam Morton, has been working with the Snowflake data platform since 2017 and has also supported over 2,500 students in preparing for the SnowPro Core certification.

Join us on this journey as we bring to life the out of the box capabilities offered by Snowflake. Following this approach will help solidify the learning concepts in your mind and put in thr best place possible to pass your SnowPro Core certification with flying colors!

Snowflake SnowPro Core Study Guide: *Supercharge your career by starting your Snowflake journey*

Adam Morton
Sydney, NSW, Australia

ISBN-13: 979-8-8414-6687-1

Sponsorship kindly provided by: Erin Evans, CEO, Intelligen Group (www.intelligengroup.com)

Production Editing: The Expert Editor (www.experteditor.com.au)

Cover image designed by Jasmine Phillips (www.wildhoneycreative.com)

For information on reprint, paperback, or audio rights, please e-mail adam@masteringsnowflake.com.

About the Author

Adam Morton was introduced to Snowflake in 2017. Having worked with more traditional database technology for many years, he found it a breath of fresh air. Lots of things that were tedious to configure in the old world just worked straight out of the box with Snowflake! This meant he could focus on his true passion: quickly adding business value for his clients.

Before Snowflake, Adam found it time-consuming to find valuable, relevant content he could consume today and apply tomorrow. He had to research and test what worked well and what didn't. Snowflake's new and innovative features dictate a change of approach and a departure from the traditional mindset required to work with the data warehouse technologies that predated it.

Following several highly visible roles over the years, which included Head of Data for an FTSE 100 Insurer, Adam accumulated a wealth of valuable, real-world experiences in designing and implementing enterprise-scale Snowflake deployments across the UK, Europe, and Australia. As a byproduct of overcoming several business and technical challenges along the way, he found a formula that worked. In this book, he shares many of these approaches with you. In fact, a lot of these experiences formed the basis for Adam's recognition as an international leader in his field when he was awarded a Global Talent Visa by the Australian Government in 2019.

Today, Adam works in partnership with Intelligen Group, a Snowflake pureplay data and analytics consultancy in Sydney, Australia. He is dedicated to helping his clients to overcome challenges with data while extracting the most value from their data and analytics implementations.

Adam is on a mission to help as many IT professionals as possible escape dead-end jobs and supercharge their careers by breaking into cloud computing using Snowflake.

He has also developed a signature training program that includes an intensive online curriculum, weekly live consulting Q&A calls with Adam, and an exclusive mastermind of supportive data and analytics professionals helping you to become an expert in Snowflake. If you're interested in finding out more, visit www.masteringsnowflake.com.

You can also find Adam sharing his knowledge on his YouTube channel at www.youtube.com/c/AdamMortonSnowflakeDataWarehouse.

Adam will be donating a portion of the proceeds from the sale of this book to Beyond Blue, an Australian mental health and wellbeing support organization. They provide support programs to address issues related to depression, suicide, anxiety disorders, and other related mental illnesses. More details can be found here: https://www.beyondblue.org.au

Snowflake Overview and Architecture

In this chapter, I'll introduce the Snowflake architecture and some key concepts and terminology. In short, I'll explain what makes Snowflake different from legacy warehouse solutions.

What is the Snowflake Data Cloud?

Snowflake is a modern cloud-based data platform designed for the cloud from the ground up. It aims to be a one-stop shop for *all* your data, whether ***structured*** (such as Excel), ***semi-structured*** (such as JSON), or ***unstructured*** (such as PDF files). Snowflake also caters to various workloads—explained below and displayed in Figure 2.1.

- **Data Engineering:** Snowflake supports several data integration and processing tools serviced by scalable and flexible compute services. Snowflake charges its compute services on a pay-as-you-go basis, known as a consumption model. There's also very little to do from a performance tuning aspect, which keeps those tedious administrational tasks to a minimum.
- **Data Lake**: The ability to store all your data within the platform at a low cost compared to an on-premises database and make the same data accessible to ***data consumers*** using standard SQL makes Snowflake very attractive to customers. Ingesting all your data onto Snowflake reduces

the risk of data silos springing up and makes it easier to manage and govern all your corporate data.

- **Data Warehouse**: Snowflake provides the blank canvas to transform and model your data quickly and easily. It also provides the tools to make it straightforward to process changes in near real-time without any user intervention.
- **Data Science**: By leveraging the elastic scalability of the cloud, Snowflake can cater for large, intensive data science workloads, which often need to crunch through a lot of data! Snowflake enables data scientists to write their machine learning models and 'push down' their logic to where the data resides within the platform. This reduces data movement and latency as well as provides an uplift in performance.
- **Data Applications**: Snowflake's architecture lends itself well to supporting applications that require hundreds or thousands of concurrent connections or require resources on-demand to support any spikes in demand, such as Black Friday retail sales.
- ***Data Exchange***: This provides a secure mechanism to share data between partners or third parties. The Snowflake ***data marketplace*** provides seamless, secure, live data sharing, making it easy to tap into new data sources and potentially monetize your own data by offering it commercially on the data marketplace.

Figure 2-1: Snowflake's extensive capabilities

Software as a Service

Snowflake is offered as Software as a Service or ***SaaS***. You may have heard this term in the past but what does it mean, and how does it differ from Infrastructure as a Service (***IaaS***) and Platform as a Service (***PaaS***)?

To explain this more clearly, I'm going to call upon a useful diagram in Figure 2-2, which has been around for some time now. It uses the process of making a pizza as a metaphor to describe the different service options.

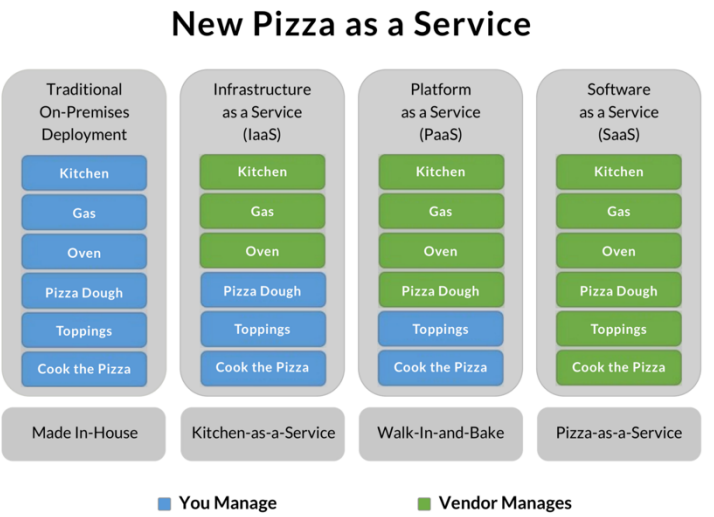


Figure 2-2: Pizza 'as a service'

The figure starts on the left with 'On-Premises Deployment,' where you, as the customer, manage all operations involved in the pizza-making process. As we progress from left to right along the diagram, we move to ***IaaS,*** where the infrastructure is provided by the vendor—in this case, it includes the kitchen, oven, and gas. The equivalent in the IT space would be a vendor handing you an environment before you go ahead and install and configure any software on top. A common example here would be signing up to Amazon Web Services or Microsoft Azure before adding and configuring the services you need to do your work.

The next step along is ***PaaS***, where you hand over more management to the vendor. In this case, they provide not only the kitchen infrastructure but also the pizza dough, essentially giving you a pizza base for you to add your preferred toppings. An example of this would be something like Windows Azure, the

operating system acting as that blank canvas or pizza base if you like. You can then install programs and configure them on top.

Finally, we get to the right-hand side of the diagram and ***SaaS***. Here you have outsourced all the pizza-making process to the vendor. All you need to worry about is eating it! Think of this as ordering your favorite pizza from Dominos. You pay a little more for the convenience while someone else takes care of the hard work behind the scenes. SaaS, then, is the arena Snowflake operates in.

This means that Snowflake runs completely on cloud infrastructure. All components of Snowflake's service (other than optional command-line clients, drivers, and connectors, which we'll cover shortly) run within the public cloud, meaning there's very little for you as the customer to do. There is no hardware (virtual or physical) to select, install, configure, or manage, and there's virtually no software to download, install, configure, or manage.

Operating in such a competitive marketplace means Snowflake must continually evolve and release new features to keep pace with the competition, all of whom are ploughing millions of dollars into development annually. To meet this commitment, Snowflake deploys new releases each week. All the required maintenance, such as management, upgrades, and tuning, are handled in the background by Snowflake.

Snowflake accounts are moved to customer accounts using a three-stage approach over two (or more) days. This staged approach enables Snowflake to monitor activity as accounts are moved and respond to any issues that may arise. Snowflake's account managers often get a schedule of planned releases to allow them to manage their customer accounts proactively. All deployments happen transparently in the background so that users experience no downtime or

disruption of service, which means Snowflake's customers are always assured of running on the most-recent release with access to the latest features.

Snowflake Industry Compliance

Snowflake has been awarded several industry-compliance awards. It is worthwhile familiarizing yourself with these because questions relating to this topic may come up in the certification exam. Let's take a brief look at the primary industry compliance certifications Snowflake has in place currently.

The Health Insurance Portability and Accountability Act (HIPAA)

The certification looks at safeguarding protected healthcare information (PHI). Examples of PHI are data typically found within patient records, such as name, address, medical reference number, date of birth, and social security number. Such rich information allows hackers to obtain credit cards or loans fraudulently and is why a single medical record is worth 10 to 20 times more than a credit card number! It's big business, with the cost of a typical healthcare breach spiraling to an average of $9.4 million in 2021, an increase of $2 million over the previous year, according to a report by IBM.[1]

SOC1, SOC2 Type 2, and PCI DSS

SOC stands for System and Organization Controls. A SOC1 report looks at an organization's internal control over financial reporting, while a SOC2 report focuses on how to secure and protect customer data.

---

[1] https://newsroom.ibm.com/2021-07-28-IBM-Report-Cost-of-a-Data-Breach-Hits-Record-High-During-Pandemic

The Payment Card Industry Data Security Standard or PCI DSS is a set of standards to ensure companies manage, store, process, and transmit credit card information in a secure way. This may include replacing the credit card number with a 'token,' which can be exchanged at execution time when a payment needs to be processed. We'll touch on this example when we look at ***external functions*** within Snowflake later in this book.

FEDRAMP

The Federal Risk and Authorization Management Program is a US government-wide program that provides a standardized approach to security assessment, authorization, and continuous monitoring for cloud-based products and services.

Note: There are some other, less well-known certifications. I have purposely omitted these from this section because they aren't included in the certification exam. If you wish to see an up-to-date and exhaustive list, head here: https://www.snowflake.com/snowflakes-security-compliance-reports

---

Additionally, if you are considering introducing Snowflake into your own environment, you'll no doubt have to address the concerns of your data security teams. The reports mentioned above are available to potential and existing Snowflake customers under a non-disclosure agreement (NDA) with Snowflake.

The Snowflake Architecture

No doubt you would have heard the term ‘big data’ over the past few years, along with technologies such as Apache Hadoop. Hadoop is a collection of open-source tools to help solve problems associated with the challenges of working with big data. Some vendors decided to build solutions on top of Hadoop and offer those products to their customers. But not Snowflake.

When compared with different flavors of Hadoop and legacy warehouse solutions, Snowflake is quite different. Don’t forget it was designed from the ground up for the cloud, while many other database offerings started life as on-premises technologies and had to be adapted to work in the cloud. This naturally created a level of complexity for those vendors, immediately placing Snowflake at a distinct advantage.

Interestingly, Snowflake runs on cloud platforms that can be considered both as partners and competitors. As displayed in Figure 2.3, Snowflake can run on Microsoft Azure, Google Cloud Platform, and Amazon Web Services Cloud Infrastructure—all of which have their own data analytics cloud-based offering that can be considered in direct competition with Snowflake. The fact that Snowflake can run on any of the three main cloud provider platforms further extends the appeal Snowflake has in the marketplace. It cannot, however, run on private cloud infrastructures, which include cloud and on-premises.

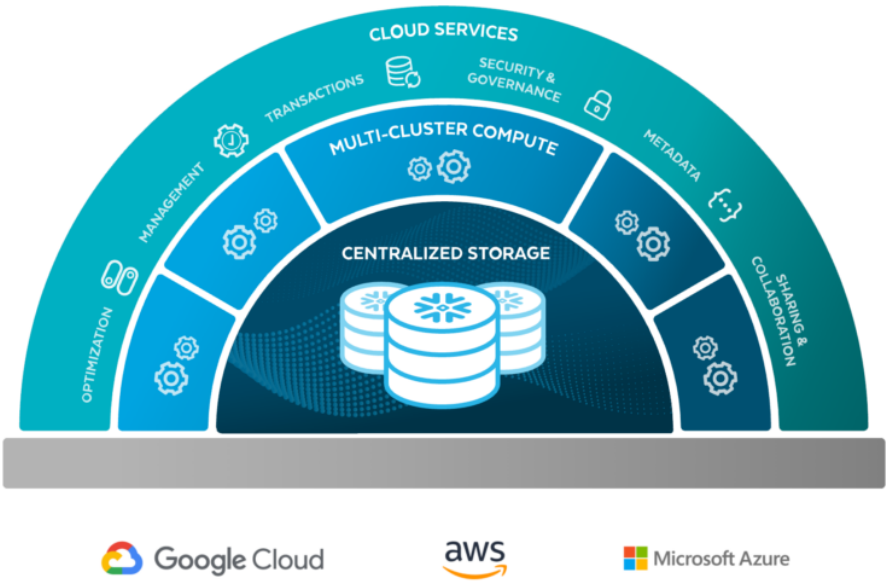


Figure 2-3: Snowflake's layers

As the end user, you only interact with the ***cloud services layer***. The cloud services layer coordinates activities between the ***compute layer*** and the ***storage layer***, which, in turn, interact with the underlying cloud platform resources.

Everest Guide

As I begin to introduce the Snowflake architecture in this chapter, I'm going to use an asset I have created known as the *Everest Guide*, displayed in Figure 2-4. The aim of this guide is to give you a visual representation of the Snowflake architecture. I have included screenshots from the guide within this book, but for the best interactive experience, access the *Everest Guide* at the following link: https://tinyurl.com/2p8u8fwv.

Note: You may have to allow pop-ups in your browser and authorize the draw.io application in Google Drive to be able to view the *Everest Guide*.

Figure 2-4: Everest Guide

Snowflake editions

Snowflake offers four different editions:

- Standard
- Enterprise
- Business Critical
- Virtual Private Snowflake.

The edition you choose, based on your specific requirements, determines the unit costs attached to the credits and the data storage you use. Each edition builds on the features of the previous one, so Enterprise includes all the features of Standard plus some extra functionality, whereas Business Critical includes all Enterprise features and so on. Let's break these down so you can appreciate the features of each edition:

- **Standard**—offers a balance between features, support, and cost. This is the most basic level of offering.
- **Enterprise**—is aimed at large enterprise-scale clients, as the name suggests. It comes with all the features of Standard, plus what I would say are critical features required for large organizations. This includes data protection and performance features, which we'll cover later in this book.
- **Business Critical**—adds enhanced security features for those organizations dealing with highly-sensitive data, particularly PHI data that must comply with HIPAA and HITRUST CSF regulations. This edition also includes advanced data protection capabilities as well as database ***failover***/rollback, business continuity, and disaster recovery.
- **Virtual Private Snowflake (VPS)**—This edition offers the highest level of security for those organizations with highly-sensitive data that operate in a tightly governed sector. It contains all the features of Business Critical, with the added benefit that the Snowflake account is isolated from all other Snowflake accounts. This means that outside of the VPS, there are no shared resources.

For a break-down at a granular level, head to this link on the official Snowflake site for the latest Feature/Edition matrix:

https://docs.snowflake.com/en/user-guide/intro-editions.html#security-governance-data-protection

Other factors that influence cost is the region where your Snowflake account is located and whether it is an on-demand or capacity account:

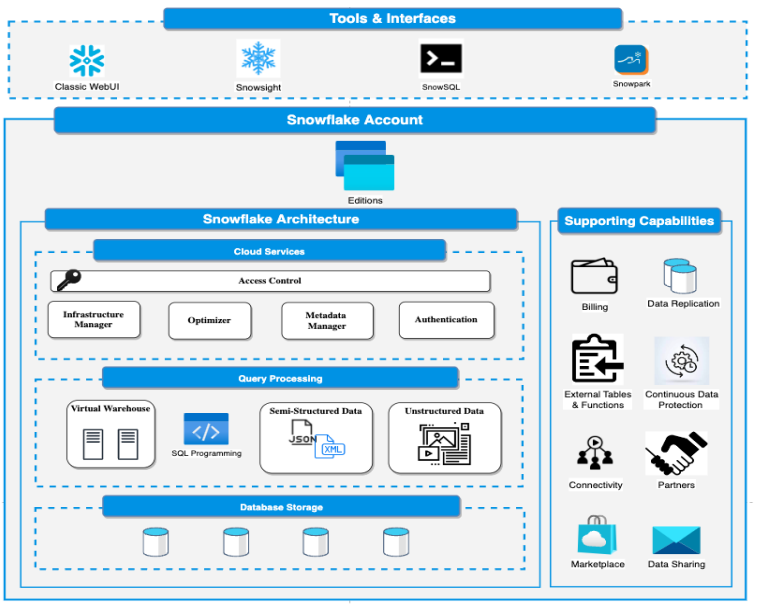
- **On Demand:** Usage-based pricing with no long-term licensing requirements
- **Capacity**: Discounted pricing based on an up-front capacity commitment.

In my experience with customers, it's typically a decision between Enterprise and Business Critical. All my customers have opted for a capacity account model.

Snowflake's three-tier architecture

Three tiers make up Snowflake's architecture, as shown in Figure 2-5.

Figure 2-5: Snowflake's three-tier architecture



Let's take a moment to understand the role each of these layers plays within the Snowflake service.

- ***Cloud Services layer***—This layer looks after the overall management of the Snowflake account. Essentially, it handles a collection of services to support the use of Snowflake, such as user authentication, metadata management, and infrastructure management.
- ***Query Processing layer***—This layer manages all query execution. The amount of compute is 'elastic,' meaning it can scale up and down in response to demand. This elasticity is provided by something Snowflake calls a ***virtual warehouse***, which we'll cover in detail in the next chapter.
- ***Database Storage layer***—This layer manages, organizes, and stores data of all formats—***structured***, ***semi-structured***, and ***unstructured***—in an optimized format. Data sits on the underlying cloud platform, meaning it can take advantage of the huge scalability offered by the cloud.

The 'mailroom'

To help you remember the architecture, the roles each of the tiers play, and how they interact, I like to use this analogy based on a mailroom, as shown in Figure 2-6 below.

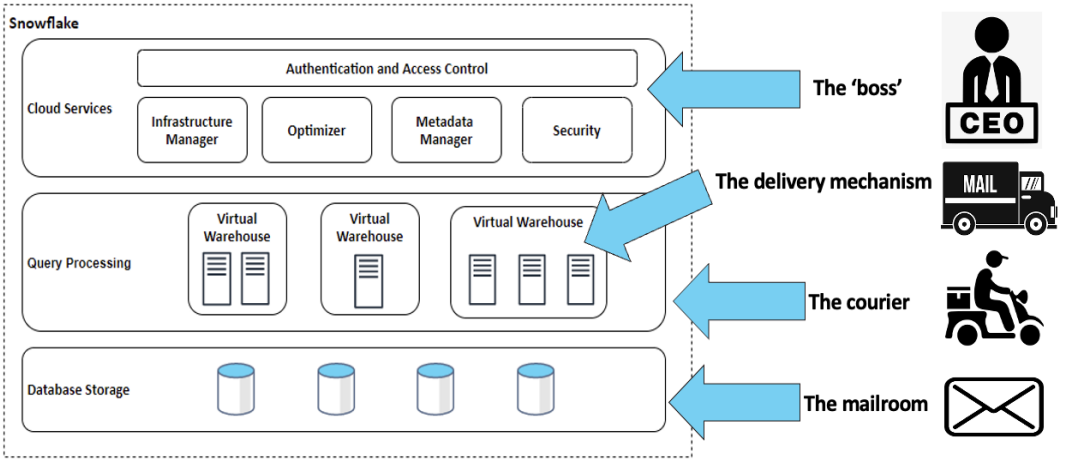


Figure 2-6: The 'mailroom'

The ***Cloud Services layer*** is the HQ, the big boss. It's the end-to-end tracking system, providing oversight across all services. It ensures the post is secured for access when in transit and keeps the lights on to ensure the post reaches its destination within the guaranteed time scale.

The ***Query Processing layer*** acts as the courier. It handles the logistics of taking the mail from the sender and working out the quickest and most efficient route to deliver it. Once the mail is delivered, it can obtain a response and return it to the original sender.

The ***Database Storage layer*** acts like a physical mailroom, designed to efficiently organize and store all the different letters and parcels arriving from many different places. Some parcels are very large, some small, while others are just regular letters, but they must all find their place within the mailroom.

The ***virtual warehouse*** provides the resources to allow these deliveries to take place. Sometimes all that might be required is a bicycle, at other times, a van or truck. A plane might be required for international items.

Keep this metaphor in mind as we look at each of these services in detail in the following pages. Understanding how these features complement each other will be a great help when thinking about how to design efficient solutions on Snowflake.

Cloud Services layer

This layer is the head office of our postal service, the brains of the operation. This collection of services—which we break down below—ties together everything we have discussed so far.

- **Authentication**—This aspect manages the service that allows users and applications to log on to the Snowflake platform.
- **Infrastructure Management**—This aspect looks after the management of the underlying cloud infrastructure.
- **Metadata Management**—This service collects metadata when various operations are carried out on the Snowflake platform.
- **Query Parsing and Execution**—This service takes care of the ***query planning*** to work out the most efficient way to process queries. It compiles the queries to ensure no syntactical errors and manages the query execution. Snowflake uses a cost-based optimizer, which determines the fastest path to access the data by generating a range of query plans—based on the metadata available—before assessing which one is optimal to use. This is referred to as ***query optimization***.

When you run a query on Snowflake, the query optimizer takes this plan, which contains the information enabling it to pinpoint only those ***micro-partitions*** that store the data required to satisfy the query. This process is known as ***query pruning*** and aims to improve query performance by skipping unnecessary ***micro-partitions***.

- ***Access Control***—This service ensures that users can only access or carry out operations on the objects and data they are permitted to, based on their privileges.

Query Processing layer

The ***Query Processing layer*** provides the resources to execute the query by using a ***virtual warehouse***. A virtual warehouse is essentially the basket of compute (CPU) and memory (RAM) required to run pretty much any SQL data manipulation language (***DML***) operations against the data in Snowflake. This also includes loading data into Snowflake tables. As we'll see in the next chapter, virtual warehouses come in a range of t-shirt sizes from extra small to large, with variations in between. Virtual warehouses can access any of the underlying data in Snowflake. You can also start, stop, drop, and create them as you need them, with no impact on any other data or operations within Snowflake.

Database Storage layer

Behind the scenes, Snowflake stores data in a compressed, native format known as ***micro-partitions***. These are small blocks of storage that are 50–500 MB uncompressed—note that once in Snowflake, they're even smaller, as all data in Snowflake is compressed. If you have a very large table, you'll have millions of these micro-partitions scattered across the storage layer. How is data organized across them? Well, Snowflake automatically selects a ***clustering key*** based on when the data is loaded. Metadata is collected and stored, which allows the

query optimizer to understand what data is contained within each micro-partition. We'll get into ***micro-partitions*** in more detail in the chapter on performance concepts later in this book.

Tools and Interfaces

Snowflake has several tools available as part of the service.

- ***Classic Web UI***—A browser-based web interface allowing users to work with objects and data in Snowflake. This was the original web interface that came out with Snowflake.
- ***Snowsight***—This is positioned as the next generation web interface, the successor to the Classic Web UI if you like. It contains all the features of the Classic Web UI but packaged in a different way. It also comes with a range of new capabilities, such as the ability to visualize data using graphs and charts. Snowsight allows you to create dashboards and filters and share them with other users, all within the web interface. It won't replace a dedicated data visualization tool, but it will satisfy the need for efficient analysis of data without needing to move the data out of Snowflake.
- ***SnowSQL***—This is the command line interface (CLI) for Snowflake. It allows users to execute SQL queries, including unloading out of and loading data into Snowflake. You need to download and install this on your own machine to use it.
- ***Snowpark***—This is a recent addition and a rapidly evolving area of Snowflake. *Snowpark* is aimed at data engineers and data scientists who wish to interact with querying and processing data in a data pipeline on Snowflake. It allows users to interact with the data using Scala or Java,

with support for Python coming very soon. It provides the ability to write custom code and push this logic down to where the data lives in Snowflake.

Data Sharing

Many of the customers I work with have a business model that involves sending data or receiving data, or both, from third parties. Often this is handled by sending files to each other or querying an ***API*** (Application Programming Interface) to extract the data for ingestion into their own data platform. Data received using files is often delayed or corrupt, resulting in teams spending countless hours backing out data loads and identifying the issues before finally getting a good copy of the data. When users need to query an API, they still need to learn the structure of the API and how to query it before physically copying data over the network so that it can be analyzed or provided to downstream ***data consumers*** or applications.

Snowflake aims to eradicate a lot of the drawbacks associated with these processes. In my opinion, data sharing is one of the primary innovations Snowflake has developed. This capability allows for the seamless sharing of data between different Snowflake accounts *without* the need for any data movement. This means zero latency, no need for duplicate datasets or additional storage, and less risk of errors than any process that involves the physical movement of data.

Shares

Let's say you have a database, ***schema,*** or table in your Snowflake account that contains information you need to share with others. To do this, you need to create a ***share***.

You can create a share and choose to add the database object(s) to the share before providing access to another Snowflake account that you'd like to share the data with. In practice, all you're really doing is sharing a pointer to your data within another account.

When you create a share, it creates an object in your database that contains all the information required to share a database. This information consists of privileges that grant access to the database, schema, and specific objects (within that schema) that are to be shared.

Account types

When it comes to sharing data in Snowflake, there are two main types of account: a ***data provider*** and a ***data consumer***.

A ***data provider*** is the Snowflake account that makes the data available to other Snowflake accounts to consume. You're able to share a database with one or more Snowflake accounts. When a database is shared, you can still use the granular ***access control*** functionality to manage access to specific objects within a shared database.

A ***data consumer*** is any account that chooses to create a database from a ***share*** made available by a ***data provider***. Once a data consumer creates a database, it acts and behaves just like any other Snowflake database.

When a ***data consumer*** creates a database from the share, it'll create a read-only version of the database containing all the objects you, as the provider, have granted access to.

Data exchange

The ***data marketplace*** allows you to create and host your own ***data exchange***. This allows you to publish and share datasets securely between a select group of

members you choose to invite. Setting up your own data exchange can remove the need for transferring files over the network and cumbersome ***ETL*** processes to create the files for your ***data consumers***. Adopting this approach goes some way to breaking down data silos in your organization. We go into more detail about ***data sharing*** later in this book.

Snowflake Partner Ecosystem

The power of Snowflake can only be realized by connecting it to other data sources and technologies. Working within Snowflake is all well and good, but you'll need to connect it to other technologies to get the most from it. Examples here are data integration and transformation, business intelligence, security and governance, and machine learning. To facilitate this process, Snowflake has an impressive and growing array of partners and connectors.

Let's look at these elements in a bit more detail in this section.

Cloud partners

Cloud partners are those companies who have developed products that work well with Snowflake. There's a vast list of these, as displayed in Figure 2-7 below:

Figure 2-7: Snowflake's partner ecosystem

One thing I really like about the partner ecosystem is that these partners offer a quick way of getting up and running to trial the various technologies called a ***partner connect portal***. Often, the portal is coupled with a tutorial that walks you through a practical example. This is helpful if you have heard of a technology but want to take a hands-on look at it to see what it's all about. Those partners who offer partner connect are denoted by a special symbol to the right-hand side of the vendor's logo (a blue connected line in a circle), as shown below.

Figure 2-8: The partner connect symbol

You can access the ***partner connect portal*** by logging into your Snowflake account and setting your current role to ACCOUNTADMIN, before clicking the partner connect button in the ribbon.

Connectors

Snowflake comes with a range of connectors for various applications. These can be accessed in the ***Classic Web UI*** by clicking on the 'help' icon, then selecting 'downloads.' See Figure 2-9.

The download area provides access to ***SnowSQL***, which is the command line interface (CLI) for Snowflake. It allows the user to execute SQL queries, including unloading out of and loading data into Snowflake.

The rest of the connectors are specific to the programming language you may wish to use against Snowflake—Python, Spark, etc.—with the ODBC driver providing general connectivity.

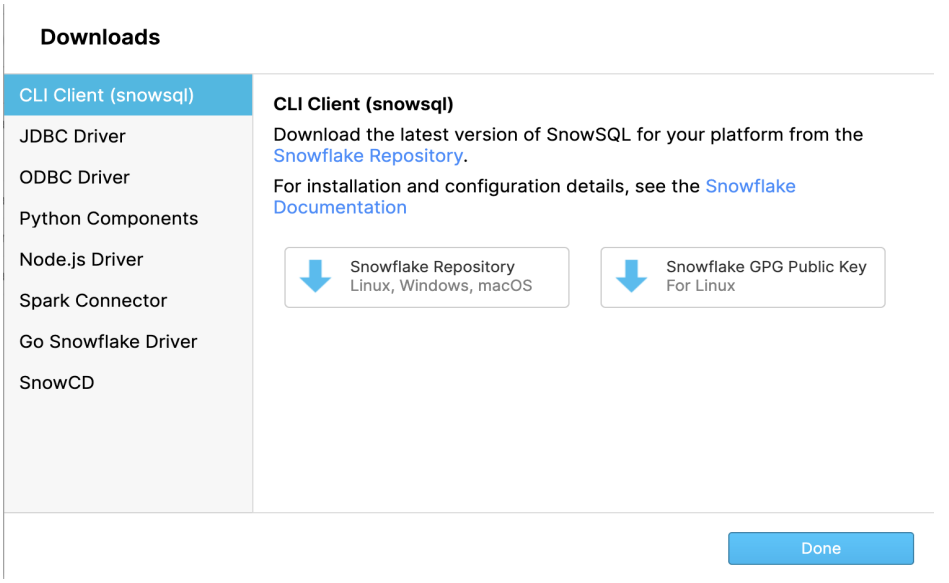


Figure 2-9: The download area in the Classic Web UI

Note: ***SnowCD*** is a connectivity diagnostic tool used for troubleshooting networking connection issues. It's highly unlikely that this will be covered in the certification, but it's good to be aware that it exists.

Chapter Revision Questions

You can find the answers to these questions at the back of this book.

1. **What makes Snowflake a true SaaS offering? (Select 3 answers.)**
    a. There is no hardware (virtual or physical) to select, install, configure, or manage.
    b. Snowflake manages all the hardware, but users are responsible for installing and patching the software.
    c. There is virtually no software to install, configure, or manage.
    d. Ongoing maintenance, management, upgrades, and tuning are handled by Snowflake.
    e. Everything other than tuning is managed by Snowflake. The users are responsible for tuning.
2. **Which cloud platforms does Snowflake run on?** (Select all that apply.)
    a. Microsoft Azure
    b. Google Cloud Platform
    c. Amazon Web Services
    d. Private Cloud Instances
3. **Which one of these is NOT a Snowflake edition?**
    a. Standard

b. Basic

c. Enterprise

d. Business Critical

4. What are the three key layers that make up the Snowflake architecture?

a. Database Storage, Query Processing, Cloud Services

b. Database Storage, Metadata Management, Cloud Services

c. Database Storage, Query Processing, Virtual Warehouse

d. Database Storage, Metadata Management, Virtual Warehouse

5. Which layer takes care of query execution?

a. Database Storage

b. Query Processing

c. Cloud Services

6. Which layer takes care of access control?

a. Database Storage

b. Query Processing

c. Cloud Services

7. Which layer takes care of metadata management?

a. Database Storage

b. Query Processing

c. Cloud Services

8. How frequently does Snowflake release new features?

a. Daily

b. Weekly

c. Monthly

d. Annually

9. How does Snowflake apply full releases?

a. Snowflake applies the release to all accounts at the same time.

b. Snowflake applies the release to all Early Access accounts, Regular Access accounts, and all remaining Enterprise and VPS accounts.

c. Snowflake applies the release to all Regular Access accounts and all remaining Enterprise and VPS accounts.

10. When Snowflake releases new features, how much downtime is there?

a. A set maintenance window at the same time each week

b. Zero downtime

c. You get to choose when to accept the new features

d. One hour anytime during the month

11. If you have multiple virtual warehouses in your environment, you also must have multiple copies of the data.

a. True

b. False

12. Which of the following terms are associated with the Cloud Services layer? (Select all that apply.)

a. Query processing

b. Query planning

c. Query optimization

d. Query design

e. Query compilation

13. Which of the following are performed by the Cloud Services layer? (Select all that apply.)

a. Metadata management

b. User authentication

c. Metadata storage

d. Access control

e. Availability zone management

14. Which of the following terms are associated with the Compute/Warehouse Layer?

a. Query processing

b. Query planning

c. Query optimization

d. Query design

e. Query compilation

15. Which Snowflake partner would you associate with data cataloging?

a. Alteryx

b. dbt

c. Alation

d. Data Robot

16. Which Snowflake partner would you associate with machine learning and data science?

a. Data Robot

b. dbt

c. Domo

d. Erwin

17. Which is the lowest Snowflake edition to offer support for Private Connectivity to the Snowflake Service using AWS PrivateLink, Azure Private Link, or Google Cloud Private Service Connect?

a. Standard

b. Business Critical

c. Enterprise

d. VPS

Chapter Revision Answers

Snowflake Overview and Architecture

1. A, C, D—Snowflake runs completely on cloud infrastructure. All components of Snowflake's service (other than optional command line clients, drivers, and connectors) run in public cloud infrastructures. There is no hardware (virtual or physical) to select, install, configure, or manage. There is virtually no software to install, configure, or manage. Ongoing maintenance, management, upgrades, and tuning are handled by Snowflake.
2. A, B, C—Snowflake can run on Microsoft Azure, Google Cloud Platform, and Amazon Web Services Cloud Infrastructure. It cannot, however, run on private cloud infrastructures, which includes cloud and on-premises. For more information, see https://docs.snowflake.com/en/user-guide/intro-key-concepts.html#data-platform-as-a-cloud-service
3. B—Basic is NOT a Snowflake edition. Snowflake offers four different editions: Standard, Enterprise, Business Critical, and Virtual Private Snowflake. For more information, see: https://www.snowflake.com/pricing/

4. A—The three key aspects that make up Snowflake's architecture are:
    a. Database Storage—manages, organizes, and stores data in an optimized format
    b. Query Processing—manages all query execution
    c. Cloud Services—handles a collection of services to support the use of Snowflake, such as user authentication, metadata management, and infrastructure management. For more information, see: https://docs.snowflake.com/en/user-guide/intro-key-concepts.html#snowflake-architecture
5. B—Query Execution is performed in the processing layer. Snowflake processes queries using ***virtual warehouses***, see: https://docs.snowflake.com/en/user-guide/intro-key-concepts.html#query-processing
6. C—Cloud Services takes care of access control. The cloud services layer is a collection of services that coordinate activities across Snowflake. These services tie together all the different components of Snowflake to process user requests from login to query dispatch. The cloud services layer also runs on compute instances provisioned by Snowflake from the cloud provider.
7. C—Cloud Services takes care of metadata management. The cloud services layer is a collection of services that coordinate activities across Snowflake. These services tie together all the different components of Snowflake to process user requests, from login to query dispatch. The cloud services layer also runs on compute instances provisioned by Snowflake from the cloud provider.
8. B—Snowflake is committed to providing a seamless, always up-to-date experience for users while delivering ever-increasing value through rapid

development and continual innovation. To meet this commitment, Snowflake deploys new releases each week. https://docs.snowflake.com/en/user-guide/intro-releases.html#snowflake-releases

9. B—Accounts are moved to the release using a three-stage approach over two (or more) days. This staged approach enables Snowflake to monitor activity as accounts are moved and respond to any issues that may arise.
10. B—The deployments happen transparently in the background. Users experience no downtime or disruption of service and are always assured of running on the most-recent release with access to the latest features. https://docs.snowflake.com/en/user-guide/intro-releases.html#snowflake-releases
11. B—False. Snowflake's unique architecture promotes one copy of the data, which can be shared and accessed by multiple users.
12. B, C, E—The cloud services layer looks after query planning, optimization, and compilation. Query processing is handled by the compute layer.
13. A, B, C, D—The cloud services layer looks after metadata management, user authentication, metadata storage, and access control. It does not look after availability zones, as this responsibility sits with the cloud provider.
14. A—The term 'query processing' is associated with the compute/warehouse layer.
15. C—Alation provides Data Cataloging functionality. It states it is the 'One Place to Find, Understand, & Govern Data Across an Enterprise.'
16. A—Data Robot is an AI cloud platform.
17. B—Both Business Critical and VPS Snowflake editions provide support for Private Connectivity to the Snowflake Service using AWS PrivateLink,

Azure Private Link, or Google Cloud Private Service Connect. Therefore, the lowest available edition to support this is Business Critical.