



## TUTORIAL

# Snowflake Tutorial For Beginners: From Architecture to Running Databases



Learn the fundamentals of cloud data warehouse management using Snowflake. Snowflake is a cloud-based platform that offers significant benefits for companies wanting to extract as much insight from their data as quickly and efficiently as possible.

 March 2025  12 min read



**Bex Tychiev**

Course Instructor at DataCamp

## What Is Snowflake?

If someone asked me to describe Snowflake in as few words as possible, I would choose these:

- Data Warehouses
- Large-Scale Data
- Multi-Cloud
- Separation
- Scalable
- Flexible
- Simple

If they wanted me to elaborate, I would string together the words like this:

Snowflake is a massively popular AI data cloud platform. It stands out from competitors due to its ability to handle large-scale data and workloads more rapidly and efficiently. Its superior performance comes from its unique architecture, which uses separate storage and compute layers, allowing it to be incredibly flexible and scalable. Additionally, it natively integrates with multiple cloud providers. Despite these advanced features, it remains simple to learn and implement.

If they request even more details, well, then I would write this tutorial. If you're totally new to the subject, our [Introduction to Snowflake](#) course is an excellent place to start.

## Why Use Snowflake?

Snowflake serves more than 8,900 customers worldwide and processes 3.9 billion queries every day. That kind of usage statistics isn't a coincidence by any means.

Below are the best benefits of Snowflake that have so much appeal:

## **1. Cloud-based architecture**

Snowflake operates in the clouds, allowing companies to scale up and down resources based on demand without worrying about physical infrastructure (hardware). The platform also handles routine maintenance tasks such as software updates, hardware management, and performance tuning. This relieves the burden of maintenance overhead, allowing organizations to focus on what matters: deriving value from data.

## **2. Elasticity and scalability**

Snowflake separates storage and compute layers, allowing users to scale their computing resources independently of their storage needs. This elasticity enables efficient handling of diverse workloads with optimal performance and without unnecessary costs.

## **3. Concurrency and performance**

Snowflake easily handles high concurrency: multiple users can access and query the data without performance loss.

## **4. Data sharing**

Snowflake's security safeguards enable data sharing across other organizations, internal departments, external partners, customers, or other stakeholders. No need for complex data transfers.

## **5. Time travel**

Snowflake uses a fancy term "Time Travel" for data versioning. Whenever a change is made to the database, Snowflake takes a snapshot. This allows users to access historical data at various points in time.

## **6. Cost efficiency**

Snowflake offers a pay-as-you-go model due to its ability to scale resources dynamically. You will only pay for what you use.

All these benefits combined make Snowflake a highly desirable data AI data cloud platform. Now, let's take a look at the underlying architecture of Snowflake that unlocks these features.

# **What Is a Data Warehouse?**

Before we dive into Snowflake's architecture, let's review data warehouses to ensure we are all on the same page.

A data warehouse is a centralized repository that stores large amounts of structured and organized data from various sources for a company. Different personas (employees) in organizations use the data within to derive different insights.

For example, data analysts, in collaboration with the marketing team, may run an A/B test for a new marketing campaign using the sales table. HR specialists may query the employee information to track performance.

These are some of the examples of how companies globally use data warehouses to drive growth. But without proper implementation and management using tools like Snowflake, data warehouses remain as elaborate concepts.

You can learn more about the subject with our [Data Warehousing](#) course.

## Snowflake Architecture

Snowflake's unique architecture, designed for faster analytical queries, comes from its separation of the storage and compute layers. This distinction contributes to the benefits we've mentioned earlier.

### Storage layer

In Snowflake, the storage layer is a critical component, storing data in an efficient and scalable manner. Here are some key features of this layer:

1. **Cloud-based:** Snowflake seamlessly integrates with major cloud providers such as AWS, GCP, and Microsoft Azure.
2. **Columnar format:** Snowflake stores data in a columnar format, optimized for analytical queries. Unlike the traditional row-based formats used by tools like Postgres, the columnar format is well-suited for data aggregation. In columnar storage, queries access only the specific columns they need, making it more efficient. On the other hand, row-based formats require accessing all rows in memory for simple operations like calculating averages.
3. **Micro-partitioning:** Snowflake uses a technique called micro-partitioning that stores tables in memory in small chunks. Each chunk is typically immutable and only a few megabytes in size, which makes query optimization and execution much faster.
4. **Zero-copy cloning:** Snowflake has a unique feature that allows it to create virtual clones of data. Cloning is instantaneous and doesn't consume additional memory until changes are made to the new copy.
5. **Scale and elasticity:** The storage layer scales horizontally, which means it can handle increasing data volumes by adding more servers to distribute the load. Also, this scaling happens independently of compute resources, which is ideal when you desire to store large volumes of data but analyze only a small fraction.

Now, let's look at the compute layer.

## Compute layer

As the name suggests, the compute layer is the engine that executes your queries. It works in conjunction with the storage layer to process the data and perform various computational tasks. Below are some more details about how this layer operates:

1. **Virtual warehouses:** You can think of Virtual Warehouses as teams of computers (compute nodes) designed to handle query processing. Each member of the team handles a different part of the query, making execution impressively fast and parallel. Snowflake offers Virtual Warehouses in different sizes, and subsequently, at different prices (the sizes include XS, S, M, L, XL).
2. **Multi-cluster, multi-node architecture:** The compute layer uses multiple clusters with multiple nodes for high concurrency, allowing several users to access and query the data simultaneously.
3. **Automatic query optimization:** Snowflake's system analyzes all queries and identifies patterns to optimize using historical data. Common optimizations include pruning unnecessary data, using metadata, and choosing the most efficient execution path.
4. **Results cache:** The compute layer includes a cache that stores the results of frequently executed queries. When the same query is run again, the results are returned almost instantaneously.

These design principles of the compute layer all contribute to Snowflake's ability to handle different and demanding workloads in the cloud.

## Cloud services layer

The final layer is cloud services. As this layer integrates into every component of Snowflake's architecture, there are many details on its operation. In addition to the features related to other layers, it has the following additional responsibilities:

1. **Security and access control:** This layer enforces security measures, including authentication, authorization, and encryption. Administrators use Role-Based Access Control (RBAC) to define and manage user roles and permissions.
2. **Data sharing:** This layer implements secure data sharing protocols across different accounts and even third-party organizations. Data consumers can access the data without the need for data movement, promoting collaboration and data monetization.
3. **Semi-structured data support:** Another unique benefit of Snowflake is its ability to handle semi-structured data, such as JSON and Parquet, despite being a data warehouse management platform. It can easily query semi-structured data and integrate the results with existing tables. This flexibility is not seen in other RDBMS tools.

Now that we have a high-level picture of Snowflake's architecture, let's write some SQL on the platform.

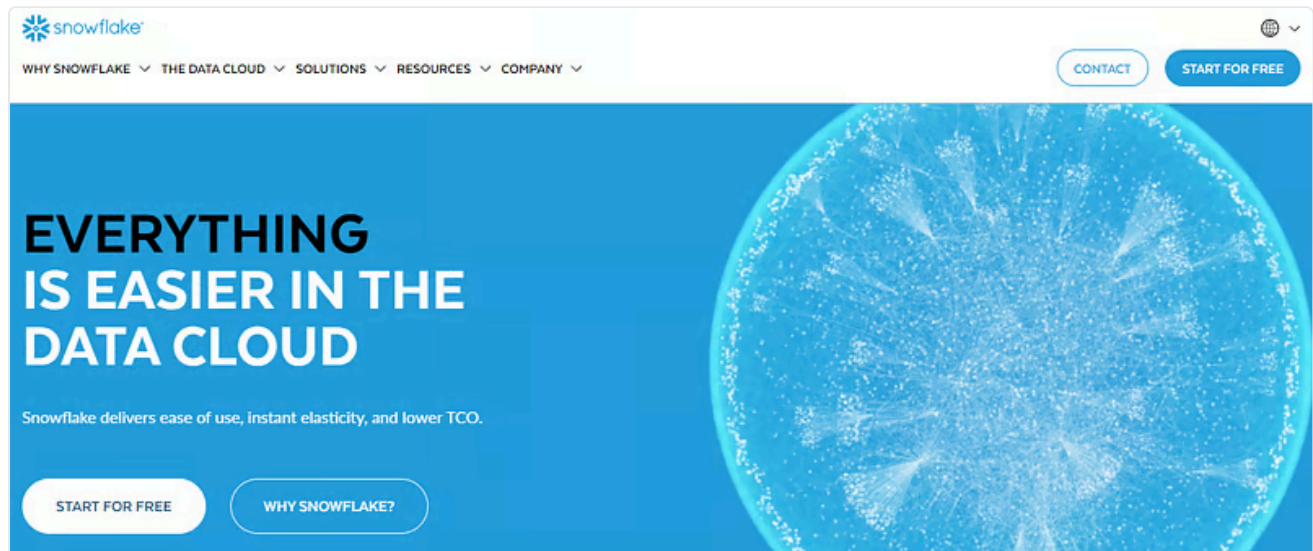
# Setting Up Snowflake SQL

Snowflake has its own version of SQL called Snowflake SQL. The difference between it and other SQL dialects is akin to the difference between English accents.

Many of the analytical queries you perform in dialects like PostgreSQL don't change, but there are some discrepancies in DDL (Data Definition Language) commands.

Now, let's see how to run some queries!

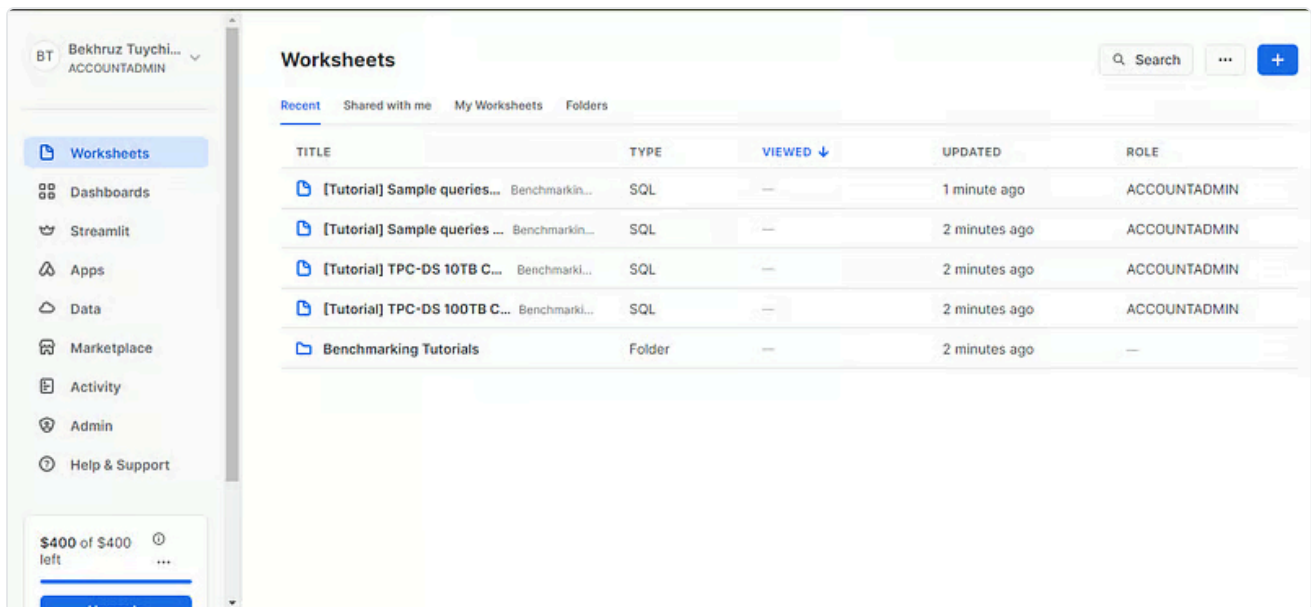
## Snowsight: Web interface



To begin with Snowsight, navigate to the [Snowflake 120-day free trial](#) page and create an account. Input your personal information and select any listed cloud provider. This gives a **120-day free trial** rather than the more standard 30-day trial you find elsewhere. The trial also includes \$400 worth of credits.

When signing up for a trial, it's recommended that users pick AWS and the US-Oregon West region. Among other reasons, Oregon is one of the lowest-cost regions for AWS infrastructure and, as a result, trial credits last longer.

After verifying your email, you'll be redirected to the **Worksheets** page. Worksheets are interactive, live-coding environments where you can write, execute, and view the results of your SQL queries.



To run some queries, we need a database and a table (we won't be using the sample data in Snowsight). To get started, I suggest you try to create a new database (you could name it something like `test_db`) and a table named using a local CSV file. You can download the CSV file by running the code in [this GitHub gist](#) in your terminal.

Afterward, you will be directed to a new worksheet where you can run any SQL query you want. I find the worksheet interface is quite straightforward and highly functional. Take a few minutes to familiarize yourself with the panels, the buttons, and their respective locations.

## Conclusion and Further Learning

Whew! We started off with some simple concepts, but towards the end, we really dove into the gnarly details. Well, that's my idea of a decent tutorial.

You've probably guessed that there is much more to Snowflake than what we've covered. In fact, the Snowflake documentation includes quickstart guides that are actually 128 minutes long! But before you tackle those, I recommend getting your hands wet with some other resources. How about these:

- [Introduction to Snowflake](#) course
- [A webinar on modernizing sales analytics with Snowflake](#)
- [Data analysis in Snowflake using Python](#) code-along
- [Official Snowflake user guides](#)
- [Snowflake Northstar](#) to learn more about Snowflake's long-term vision.

Thank you for reading!

### TOPICS



**Bex Tychiev**

Course Instructor at DataCamp

I am a data science content creator with over 2 years of experience and one of the largest followings on Medium. I like to write detailed articles on AI and ML with a bit of a sarcastic style because you've got to do something to make them a bit less dull. I have produced over 130 articles and a DataCamp course to boot, with another one in the making. My content has been seen by over 5 million pairs of eyes, 20k of whom became followers on both Medium and LinkedIn.