



**PROOF-OF-CONCEPT
GUIDE**

TEST-DRIVING SNOWFLAKE FOR DATA LAKE

**THE DEFINITIVE GUIDE TO
MAXIMIZING YOUR FREE TRIAL**



**CHAMPION
GUIDES**

USE YOUR FREE CREDITS WISELY

This guide will help you use the free, self-service, proof-of-concept trial to evaluate Snowflake's cloud data platform for your data lake. It offers tips and advice, and it's aligned with key test milestones to maximize your 200 free credits, a \$400 value. Use this guide to discover how Snowflake can help your organization get the most out of its data.

Since Snowflake provides usage-based, per-second pricing, you already have the most efficient data platform available. But you'll still want to use your self-service credits effectively. It's possible to use all your free credits in a day simply by leaving an idle virtual warehouse up and running. To conserve credits, it's best to spend a few hours a day for a week executing queries, loading data, and transforming data with Snowflake's auto-suspend feature set to 10 minutes or less. With this approach, you'll get a good feel for the system and still have credits in reserve to get started on your next project.

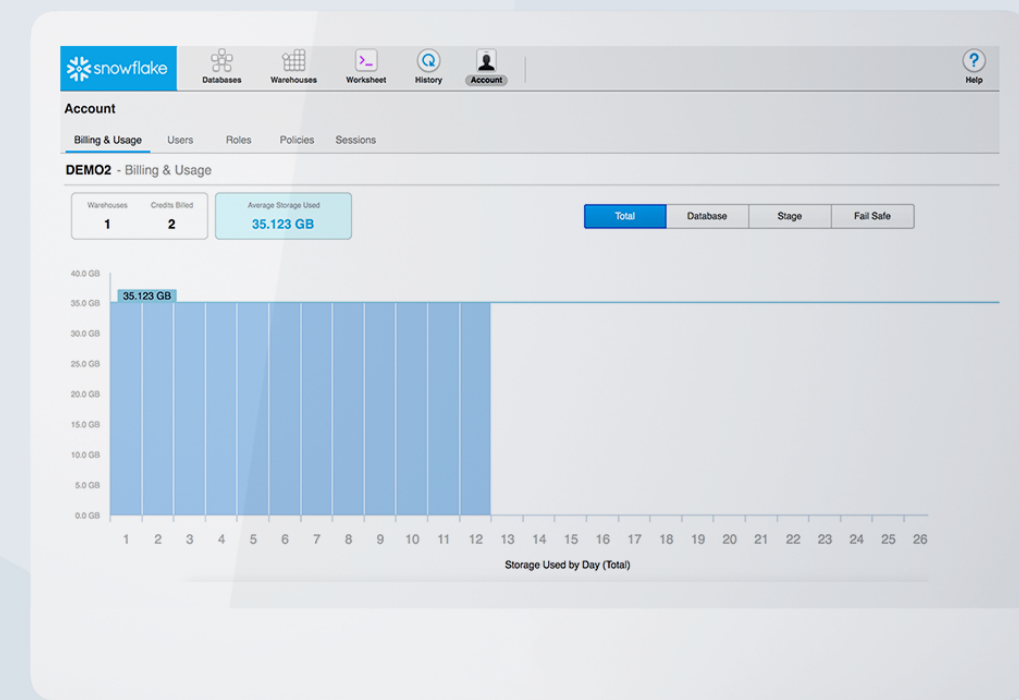
For most business requirements, following the advice in this guide will give you a strong indication of how Snowflake can fulfill your data lake use cases by the time you have consumed all your free, self-service credits. If you need more time or help to investigate your use cases, please **contact Snowflake**.

GETTING STARTED

Before you spin up your first virtual warehouse (compute cluster), get a sense of how Snowflake works and what your data platform priorities are. Plan how you will use your credits wisely to ensure you get a good look at Snowflake's strengths and how it can transform your data analytics and your organization. If possible, consider planning test scenarios using a subset of your data.

To accomplish this, here are four steps to take before you fire up Snowflake:

- 1 Read about [Snowflake's architecture](#).
- 2 Think about which of Snowflake's capabilities most closely maps to your data lake needs and make sure you test against those needs. Here are some examples: concurrency, ease of use, instant elasticity, usage-based pricing, SQL-based query capability, flexibility, speed, and compatibility with existing ETL and BI tools. You can use Snowflake to complement your existing data lake by using it as a query engine, a transformation engine, or both. Or you can make Snowflake your data lake if that fits your needs.
- 3 Also, think about how you want to divide workloads to cover all the functions you want to test, because querying data, loading data, and running external ETL and BI tools all consume credits.
- 4 View the [Snowflake in 20 Minutes](#) getting started tutorial.



QUICK TIP

Account billing and usage information is easy to access within Snowflake. Always check usage during the trial period to monitor your credits.

Figure 1: Billing & Usage screen

QUICK TIPS FOR USING YOUR CREDITS WISELY

- > Snowflake sample data is free to browse but not to query. If you don't have your own data set ready to test, use the Snowflake sample data set. Keep in mind the sample data set is large (up to 10 TB compressed).
- > Query your test data in your cloud storage directly via Snowflake external tables, or pick a reasonable batch of data to load for testing. A reasonable batch requires enough data to run a few sophisticated queries and test ETL (ingestion), BI, and concurrency but not enough to consume too many credits. Running production-type workloads will generally exceed the number of free credits.
- > As a baseline, if you choose to load data into Snowflake, try to keep loaded data under **150 GB**.
- > **Do not** turn off auto-suspend. By default, the auto-suspend threshold is set to 10 minutes. You should consider **reducing the auto-suspend threshold to 5 minutes** during the trial period, especially if you are running regular ETL jobs.
- > Account metering and billing runs in real time as you size virtual warehouses up and down and run queries. Users with the ACCOUNTADMIN role can use the Snowflake web interface or SQL to view monthly and daily credit use for all virtual warehouses in your account. To view virtual warehouse credit use for your account, click **Account > Billing & Usage**.

QUERYING DATA

There are various options for how you query data as part of the test-drive. Follow Step 1 to test external tables if you want to query data in your existing data lake. Or skip that and go to Step 2 to get familiar with the Snowflake sample data before you run queries.

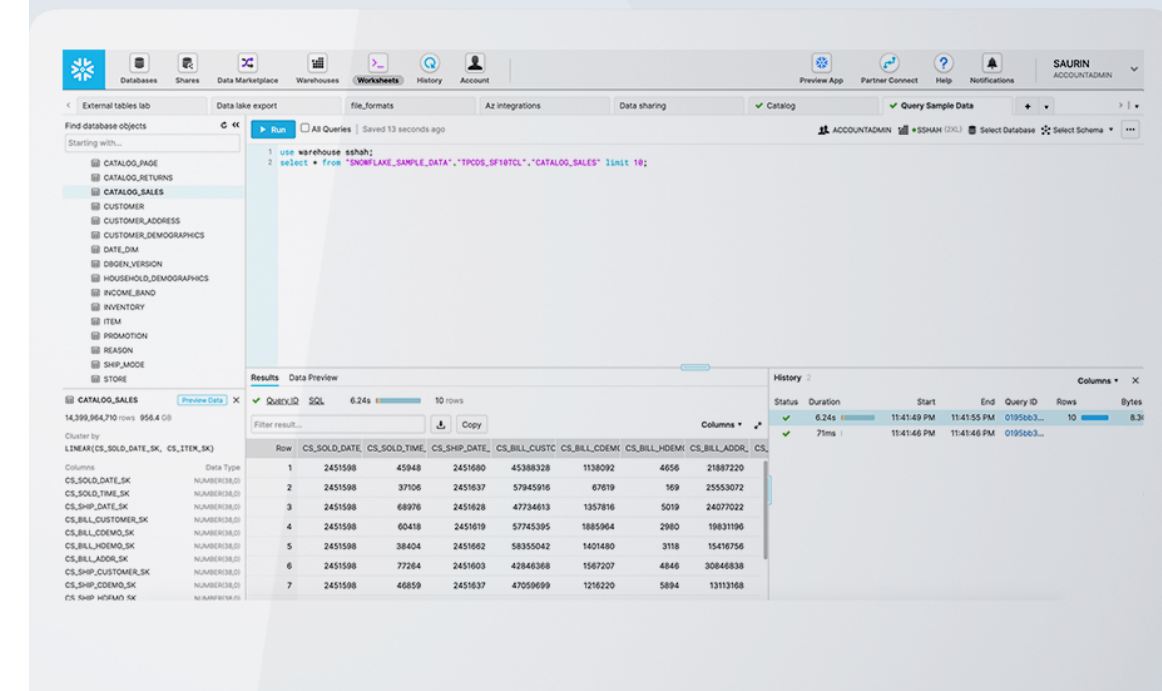
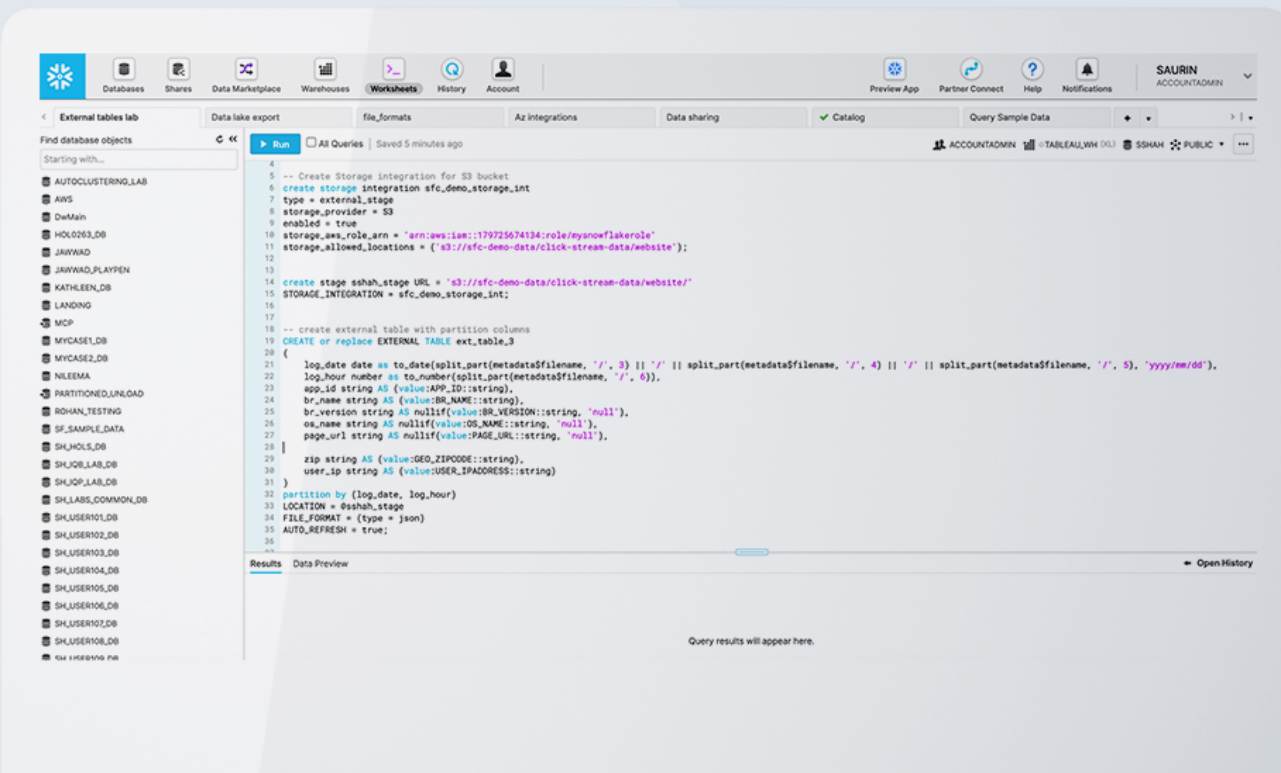
STEP 1: TEST-DRIVE EXTERNAL DATA

You can use Snowflake **external tables** to query data in your data lake directly. External tables store file-level metadata about the data files such as the file path, a version identifier, and partitioning information. This enables you to query data stored in files in a data lake as if it were inside a database.

For the purpose of this test drive, you can create an **external stage** and external tables to query your data lake directly. Once you have your data ready, simply run your queries through the external tables. See Figure 2.

STEP 2: GET FAMILIAR WITH SNOWFLAKE SAMPLE DATA

Figure 2: worksheets tab



QUICK TIP

You can find Snowflake sample data sets in the Worksheets tab in the UI.

Figure 3: Sample worksheet

Browse around and look at sample query results. Once you fire up a virtual warehouse, you can also run basic queries on sample data. But keep in mind that queries on sample data still consume credits.

The following Snowflake sample data sets are available:

- **TPC-DS** models the decision-support functions of a retail product supplier. The supporting schema contains vital business information, such as customer, order, and product data.
- **TPC-H** is a decision-support benchmark. It consists of a suite of business-oriented ad hoc queries and concurrent data modifications.

“The new Snowflake worksheet removes the need to use third-party tools to work within the platform—including doing all administrative operations, most SELECT queries, and any type of explorative work—and makes working within the Snowflake environment even more seamless.”
—NICK ELSER, Director of Engineering at Instacart

STEP 3: RUN QUERIES ON DATA

Identify workloads and key business-critical or highly common queries you're currently running. Then set and run those queries in Snowflake.

For information on how to run queries, see the following:

- [Query documentation](#)
- [Using Worksheets for Queries/DML/DDL](#)

It's likely you are searching for improvements over your current processes. Make sure you save enough credits to test the following:

- **Ease of use and speed-testing:** How easy is it to set up a query and how fast will it run? Remember, with Snowflake, you can add nodes (servers) to a virtual warehouse to speed up querying and you can also choose which virtual warehouse to use to run specific queries. If results are coming in too slow, add some power to the query but, as always, be credit-conscious when you do.
- **Concurrency:** Run multiple queries at once to test query and user concurrency results. You can add or subtract nodes to change the speed of concurrent queries.
- **Multi-cluster virtual warehouses:** Scale compute resources to manage your user and query concurrency needs as they change, such as during peak and off hours. For example, if you are testing ETL feeds or want to emulate periods where concurrent use is very high, you can set up a multi-cluster warehouse to accommodate these extra loads and then **automatically scale back down** when peak use subsides. Keep in mind that a multi-cluster configuration consumes more credits. For more information on multi-cluster warehousing, see the [multi-cluster warehouse documentation](#).

QUICK TIP

Be sure to test business-critical queries during your trial.

Figure 4: Sample worksheet

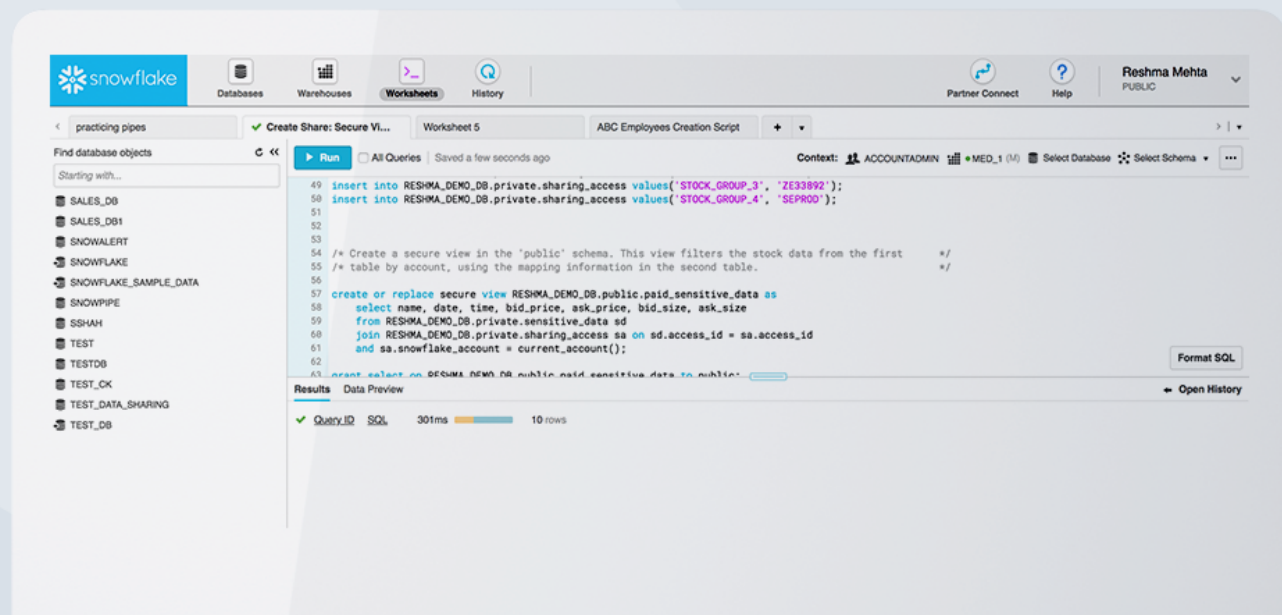
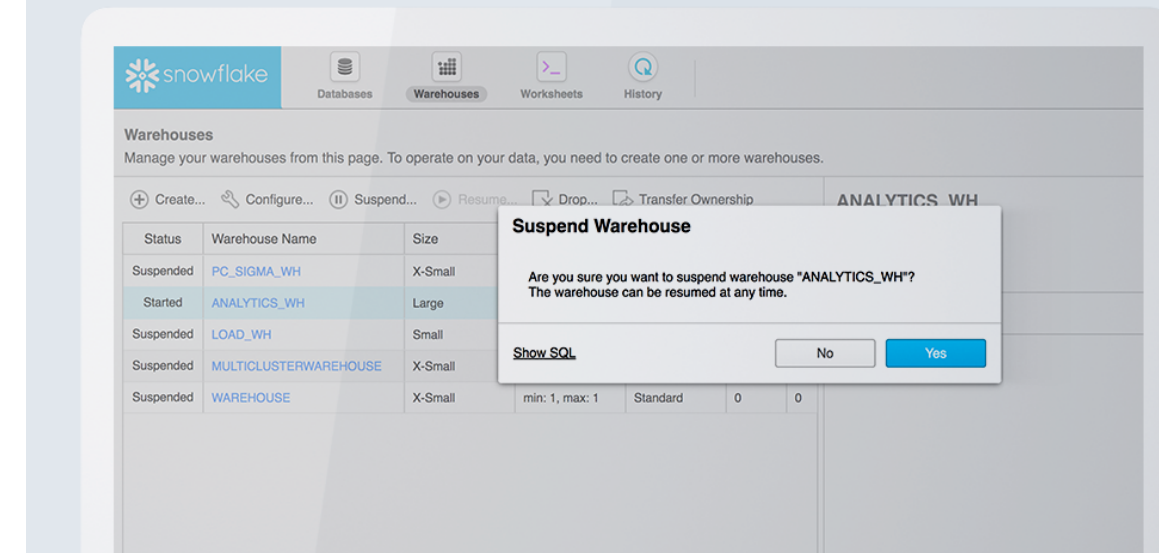


Figure 5: Auto-suspend dialog box



STEP 4: TEST EXTERNAL BI TOOLS

Save some credits for connecting and testing any BI tools you're currently using (such as Tableau). See the [BI documentation](#) for more details.

Put aside some time for connector setup, keeping in mind that every data warehouse vendor or architecture has slight variations in BI syntax. Convert a couple of queries to BI visualized reports to get a feel for how your BI tools work inside Snowflake. But remember, it takes credits to configure and run BI reports so don't run too many reports; you can run a few test reports using only a handful of credits.

QUICK TIP FOR TESTING EXTERNAL BI TOOLS

Convert a couple of queries to BI visualized reports to get a feel for how your BI tools work inside Snowflake.

LOADING AND TRANSFORMING DATA

When you're ready, you can simply load your own data into Snowflake and use it to test running queries and doing data transformations.

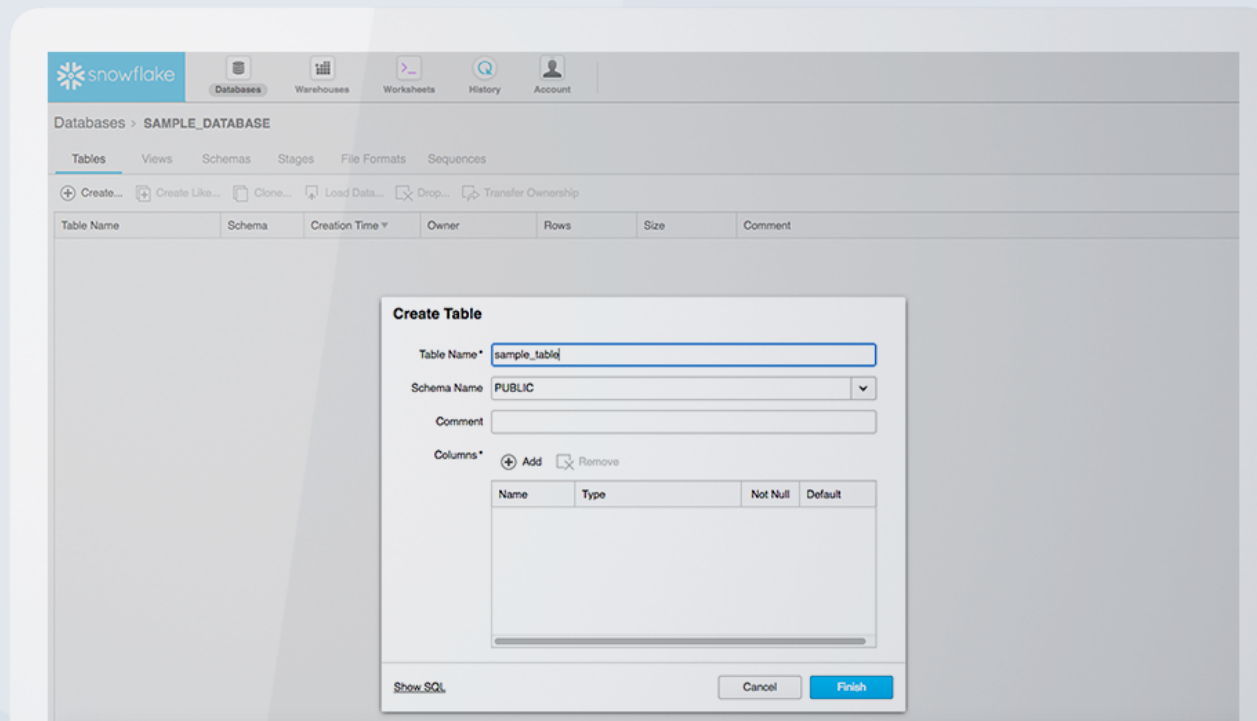
STEP 1: CREATE A DATABASE AND TABLE

First, create a database and table with the Snowflake SQL worksheet. The columns and data types in your table need to match the text file you are loading. For more information, see [Create Stage Objects](#).

QUICK TIP

You can easily create tables inside the Snowflake UI.

Figure 6: Create Tables dialog box



STEP 2: START ONE OR MORE WAREHOUSES

To load data, you need to activate a warehouse into which you will load the data. Here's how to spend your credits wisely when loading and eventually querying data:

- **Start small.** With Snowflake, you can activate one virtual warehouse or a cluster of virtual warehouses depending on data size, query speed, and query or user concurrency needs. You can change your cluster size up and down in real time, based on your needs at any given moment. Start with one node (one credit), which will start up instantly.
- **Grow as needed.** Add or subtract cluster nodes as needed (each active node consumes one credit per hour, so be judicious). If possible, keep the cluster size for the proof-of-concept period to around eight nodes or less, and add nodes later to speed query and concurrency testing. For more information on adding nodes, see [Elasticity & Separation of Compute and Storage](#).

QUICK TIP FOR STARTING WITH A SMALL WAREHOUSE

Pick a reasonable batch of data to load for testing (enough to run a few sophisticated queries and test ETL, BI, and concurrency), but not enough to consume too many credits.

STEP 3: LOAD DATA

You can load data into Snowflake in several different ways:

- **Use Partner Connect.** Partner Connect simplifies data loading through prebuilt integrations with Snowflake technology partners such as Fivetran, Alouma, and Stitch. Using Partner Connect, you can avoid the manual work associated with creating databases, creating warehouses, and separately developing your data pipeline. Watch this [video](#) to learn more.
- **Use the Snowflake UI.** To manually load small files and even spreadsheets from your desktop, the easiest method is to use the [Snowflake web interface](#).
- **Manually load bulk data.** You can also load bulk data from **Amazon S3** by using **SnowSQL**, a command-line client. For more information, see [Data Loading Considerations](#).

QUICK TIP FOR DATA LOADING

While data loading is in progress, take a quick spin around the Snowflake web interface to start getting comfortable with Snowflake.

STEP 4: SET UP AND RUN ETL TOOLS INSIDE SNOWFLAKE

Test your ETL processes. If you set up regular interval ETL runs, make sure you change the [auto-suspend default](#) from 10 minutes to 5 minutes to save credits.

STEP 5: TEST DRIVE PROGRAMMATIC ACCESS

Once you have mastered executing queries using the web interface and you have tested some standard BI tools, it is time to venture outside the database to test programmatic access. All requests to connect to Snowflake are made over a secure connection, and all executed queries are translated to REST calls.

Using the Snowflake command-line interface (CLI) is a great place to try programmatic access, but if you are more adventurous (or have code you would like to try with Snowflake), feel free to download and use [Snowflake Connector for Spark](#) and [Snowflake Connector for Python](#).

IF YOU HAVE MORE CREDITS TO BURN, TRY THESE IDEAS!

> USE SOME SEMI-STRUCTURED DATA

Snowflake is unique in that it natively supports JSON (and other semi-structured data) alongside relational data. Most databases and data stores support only one format.

With Snowflake, you can choose to “flatten” nested objects into a relational table or store the objects and arrays in their native format within the VARIANT data type. JSON and other semi-structured data can be manipulated using **ANSI-standard SQL** with the addition of dot notation.

To learn more about processing JSON data, watch this [video](#).

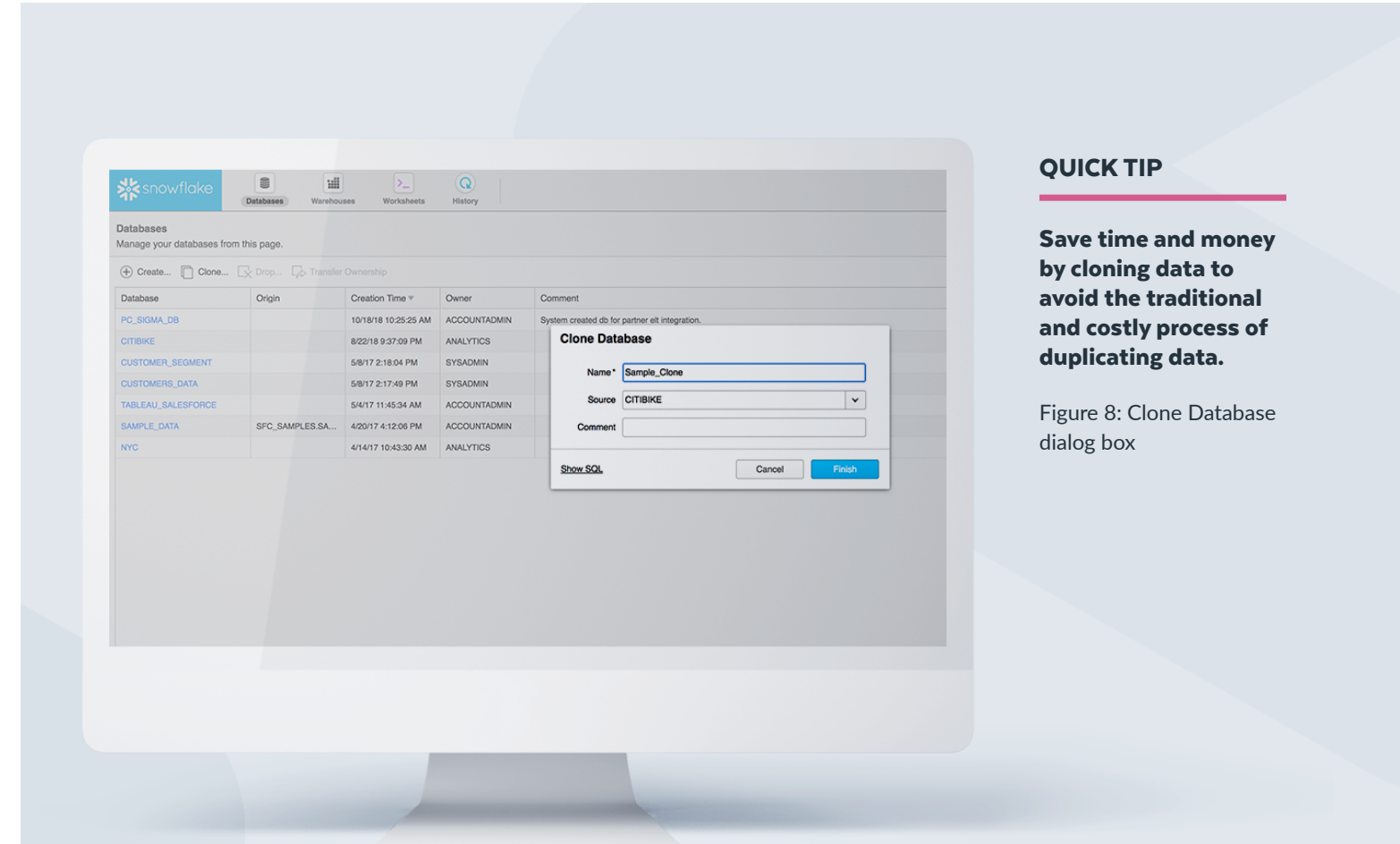
> CREATE MULTIPLE COPIES OF DATA

Using the simple Snowflake CLONE command, you can create multiple copies of data tables, schemas, and databases without replicating the data itself. This provides you with the ability to almost instantly make the data available for multiple user groups, without the additional cost (or time) of replicating the data.

For more information on cloning, watch this [video](#).

Figure 7: JSON semi-structured data

```
▶ Run All Queries Saved a month ago
43 // Show JSON data
44 // Create a table for semi-structured data and load it in
45 //4
46 CREATE OR REPLACE TABLE DAILY_HISTORY
47 ( CITY_ID INT,
48 T timestamp,
49 MAIN variant,
50 WEATHER variant);
51
52 //5 List data in staging area
53 list @weather_stage;
54
```



QUICK TIP

Save time and money by cloning data to avoid the traditional and costly process of duplicating data.

Figure 8: Clone Database dialog box

> ACCESS HISTORICAL DATA

If you’ve been carefully managing your credits over time, you can use the Snowflake **Time Travel** feature to access historical data. This includes accessing data that has been changed or deleted at any point within a defined period.

With Time Travel, you can:

- Restore data-related objects (tables, schemas, and databases) that were accidentally or intentionally deleted
- Duplicate and back up data from key points in the past
- Analyze data use and manipulation over specified periods

See [Understanding & Using Time Travel](#) for more information.

NEXT STEPS

After you have used all your credits and are satisfied with what Snowflake can provide, you can continue to use Snowflake with pay-as-you-go billing. You can also continue to test larger data sets and workloads, if needed, with pay-as-you-go billing.

If you still have questions or need more time for the trial evaluation, [contact Snowflake](#).

Also, see the following resources:

- [Snowflake documentation](#)
- [Snowflake blog](#)
- [Snowflake Community](#), the online community and support portal for Snowflake users and partners



ABOUT SNOWFLAKE

Snowflake's cloud data platform shatters barriers that have prevented organizations of all sizes from unleashing the true value from their data. Thousands of customers deploy Snowflake to advance their businesses beyond what was once possible by deriving insights from their data by all their business users. Snowflake equips organizations with a single, integrated platform that offers the data warehouse built for the cloud; instant, secure and governed access to their network of data; and a core architecture to enable many types of data workloads, including a single platform for developing modern data applications. Snowflake: Data without limits.

Find out more at [snowflake.com](https://www.snowflake.com)

