snowflake®

TEST-DRIVING
SNOWFLAKE:
THE DEFINITIVE GUIDE TO MAXIMIZING
YOUR DATA WAREHOUSE FREE TRIAL

**SELF-SERVICE,
PROOF-OF-CONCEPT
GUIDE**

USE YOUR FREE CREDITS WISELY

This guide will help you use our free, self-service, proof-of-concept (POC) trial to evaluate Snowflake, the data warehouse built for the cloud. Discover how Snowflake can help your organization get the most out of its data. This guide offers tips and advice, and it's aligned with key test milestones to maximize your 200 free credits, a $400 value.

Since Snowflake provides usage-based, per-second pricing, you already have the most efficient data warehouse available. But you'll still want to use your self-service credits effectively. It's possible to use all of your free credits in a day simply by leaving an idle warehouse up and running. Therefore, it's best to spend a few hours a day for a week loading data and executing queries with Snowflake's auto-suspend feature set to 10 minutes or less to conserve credits. 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 cases, following the advice in this guide will give you a strong indication of how Snowflake can fulfill your business requirements by the time you have consumed all your free, self-service credits. If you'll need more access or help to investigate your business case, please **contact Snowflake**.

GETTING STARTED

First things first...

Before you spin up your first virtual warehouse (compute cluster), get a sense of how Snowflake works and what your data warehouse 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 a Snowflake data warehouse:

1. Read about Snowflake architecture **here**.

2. Think about which of Snowflake's capabilities most closely map to the needs of your business case 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 (extract, transform, load) and BI (business intelligence) tools.

3. Also, think about how you want to divide workloads to cover all the functions you want to test, because loading data, querying, and running external tools (ETL, BI) all consume credits.

   Rule of thumb: No more than 25% of your free credits should be spent on data loading. If you think data loading may consume more than 50% of your credit allocation, **please contact Snowflake for assistance**.

4. View the "Snowflake in 20 Minutes" getting started tutorial **here**.
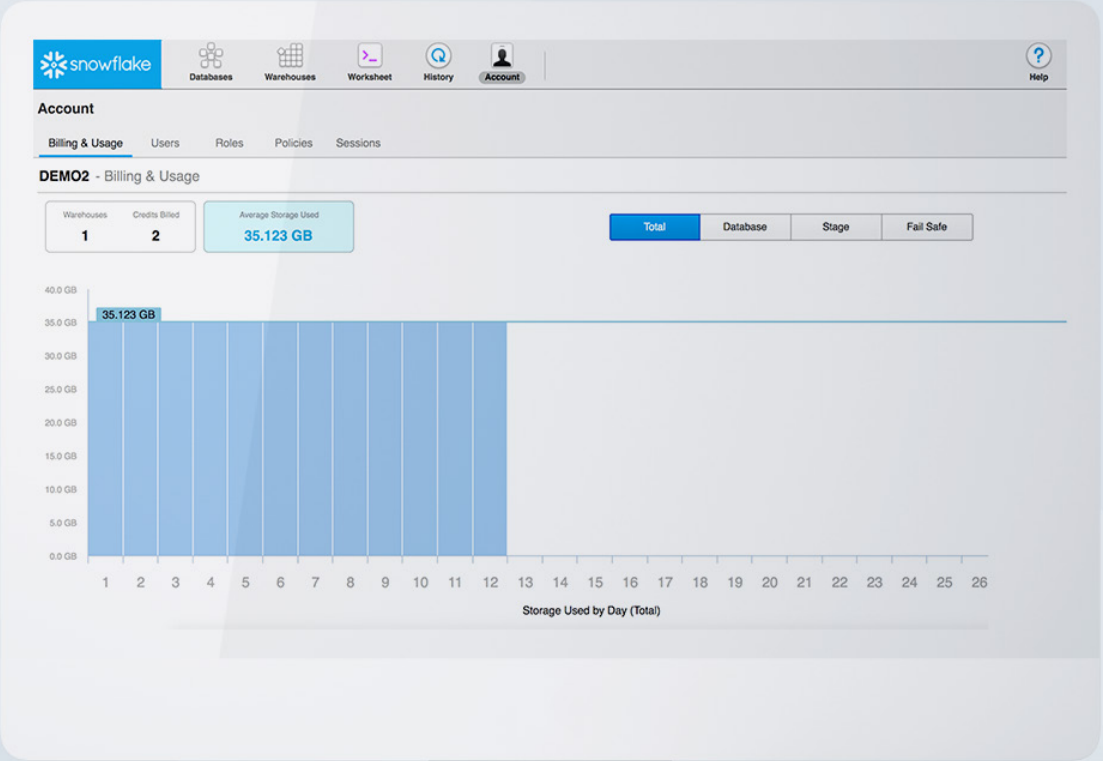
Account billing and usage is easy to access within Snowflake. Always check usage during the trial period to monitor your credits.

TIPS TO USE YOUR CREDITS WISELY

> Our sample data is free to browse but not to query. If you don't have a data set ready to test on your own, use our sample data set. Keep in mind our sample data set is large (up to 10TB compressed).

> Pick a reasonable batch of data to load for testing—enough to run a few sophisticated queries and test ETL (ingestion), BI, and concurrency, but not enough to consume too many credits. If you are trying to run production-type workloads, this will generally exceed your free credits.

> As a baseline, try to keep loaded data **under 150GB**.

> **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 clusters 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 warehouses in your account. To view warehouse credit use for your account, click Account >Billing & Usage.
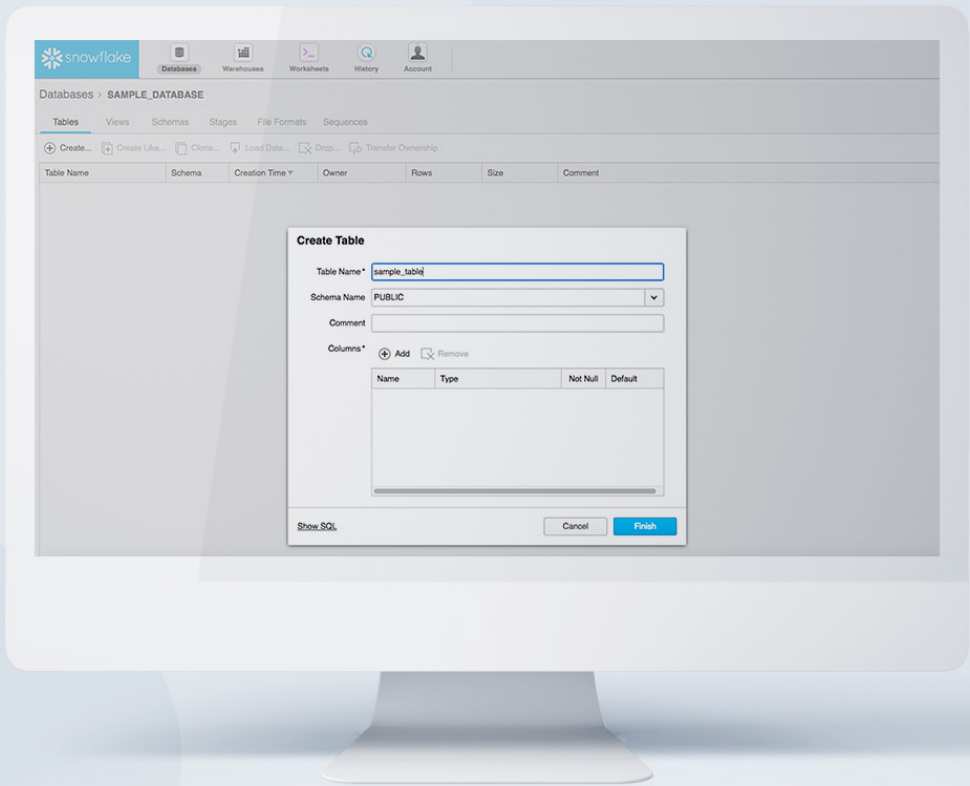
LOADING DATA

REQUIRES 50-100 CREDITS DEPENDING ON THE DATA SET SIZE

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. **Click here** for more detailed instructions.

STEP 1
Easily create tables inside the Snowflake UI.

STEP 2: FIRE UP 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 warehouse or a cluster of warehouses depending on data size, query speed and query or user concurrency needs. You can move your cluster size up and down, and in real time, based on your needs at any given moment. Start with one node (one credit), which will fire 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 save node additions to speed query and concurrency testing later on. For more information on adding nodes, **click here**.

QUICK TIP

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: PREPARE TO LOAD DATA

You can load data into Snowflake in several different ways:

- **Use Partner Connect.** Partner Connect simplifies data loading through pre-built integrations with Snowflake's technology partners such as Fivetran, Alooma and Stitch. Using Partner Connect, you can avoid the manual work associated with creating databases, creating warehouses and separately developing your data pipeline. **Here's a 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 UI.

- **Manually load bulk data.** You can also load bulk data from **Amazon S3** by using **SnowSQL**, a command-line client. You can find more information about loading bulk data **here**.

QUICK TIP FOR DATA LOADING

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

QUERYING DATA

REQUIRES APPROXIMATELY 75–100 CREDITS

Once you load your data, you're ready for the fun part: running queries. Using the Snowflake sample data set or your own loaded data, you can start running test queries to see the power of the Snowflake data warehouse. Here are six tasks for querying in Snowflake.

STEP 1: GET COMFORTABLE

Browse around and look at sample query results. Once you fire up a warehouse, you can also run basic queries on sample data. But keep in mind, though, that sample data queries 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.
- **OpenWeatherMap** is a repository of historical and intraday-updated current and forecasted weather data.
- **TPC-H** is a decision-support benchmark. It consists of a suite of business-oriented ad hoc queries and concurrent data modifications.
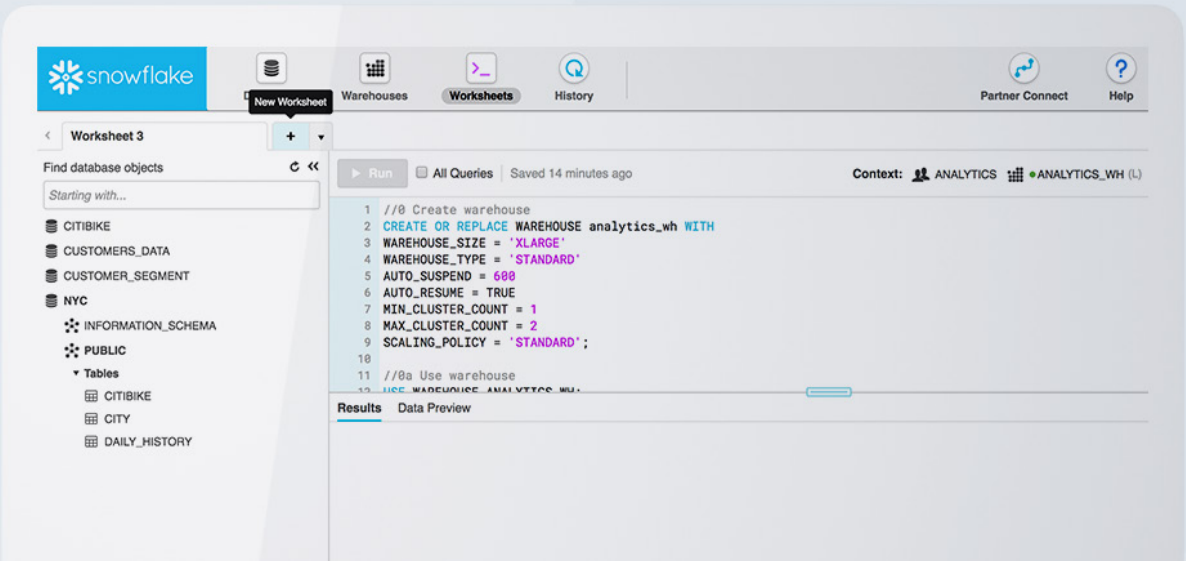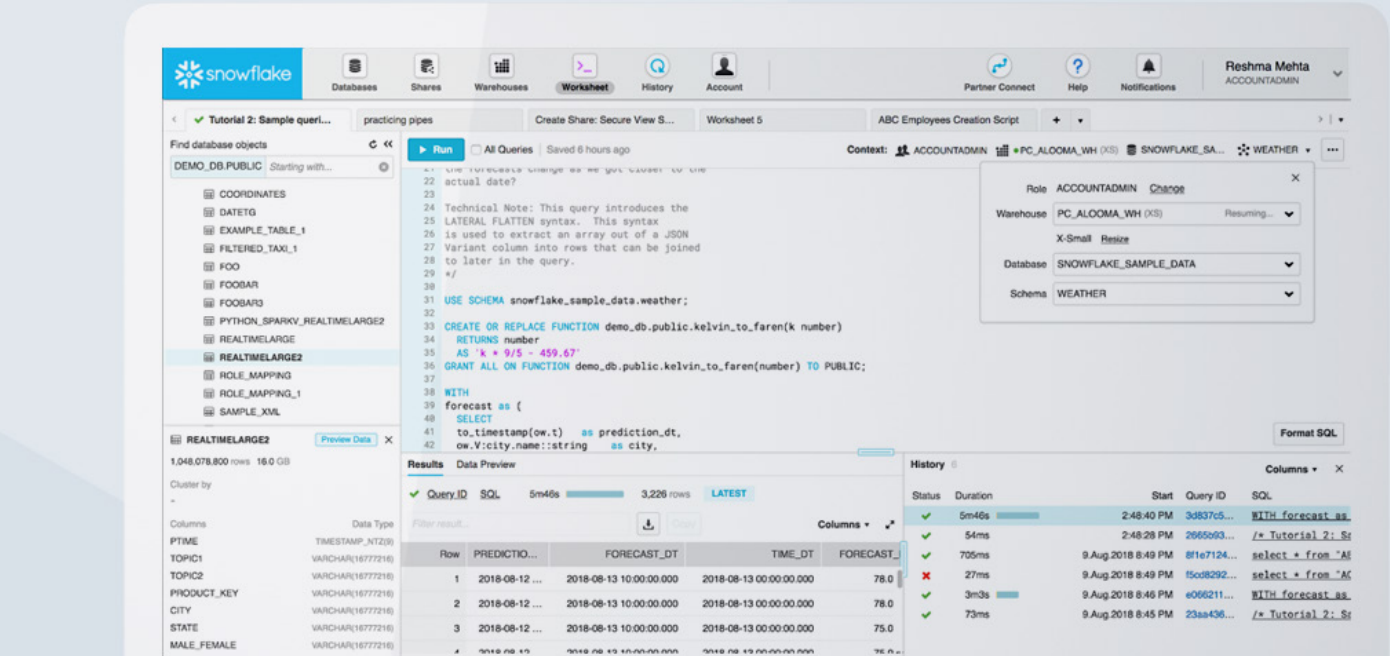
> **"** **The new Snowflake Worksheet removes the need for third-party tools to work within the platform—including all administrative operations, most SELECT queries, and any type of explorative work—and makes working within the Snowflake environment even more seamless."**
>
> **NICK ELSER,** Director of Engineering at Instacart

STEP 2: START RUNNING QUERIES ON YOUR OWN DATA INSIDE SNOWFLAKE

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

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

- **Overview**
- **Using worksheets**

STEP 3: TEST QUERIES WITH THE GOAL OF IMPROVING CURRENT PERFORMANCE

Since you are evaluating SaaS data warehouse options, you are likely searching for improvements over your current processes. Make sure you save enough credits to test for 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 to speed up querying time and you can also choose which 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/or user concurrency results. Similar to the point above, you can add or subtract nodes to change the speed of concurrent queries.

- **Multi-cluster 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 time 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 will consume more credits. For more information on multi-cluster warehousing, **click here.**
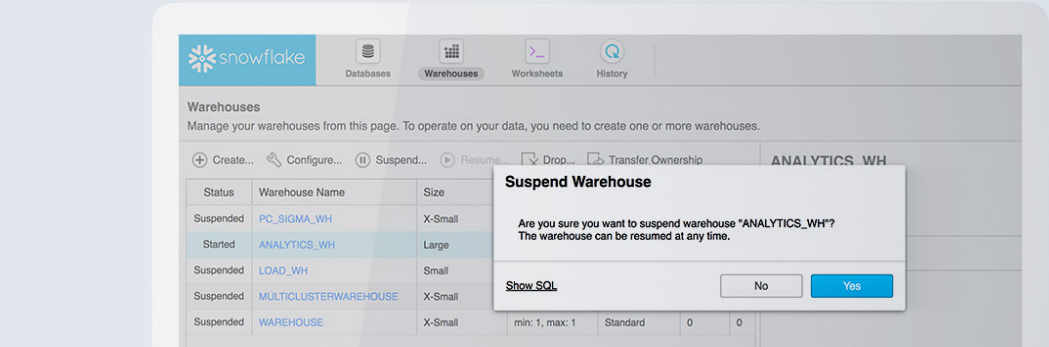
STEP 4: SET UP AND RUN ETL TOOLS INSIDE SNOWFLAKE

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 EXTERNAL BI TOOLS

Save some credits for connecting and testing any BI tools you're currently using (such as Tableau). **Click here** 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 works 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 a only handful of credits.
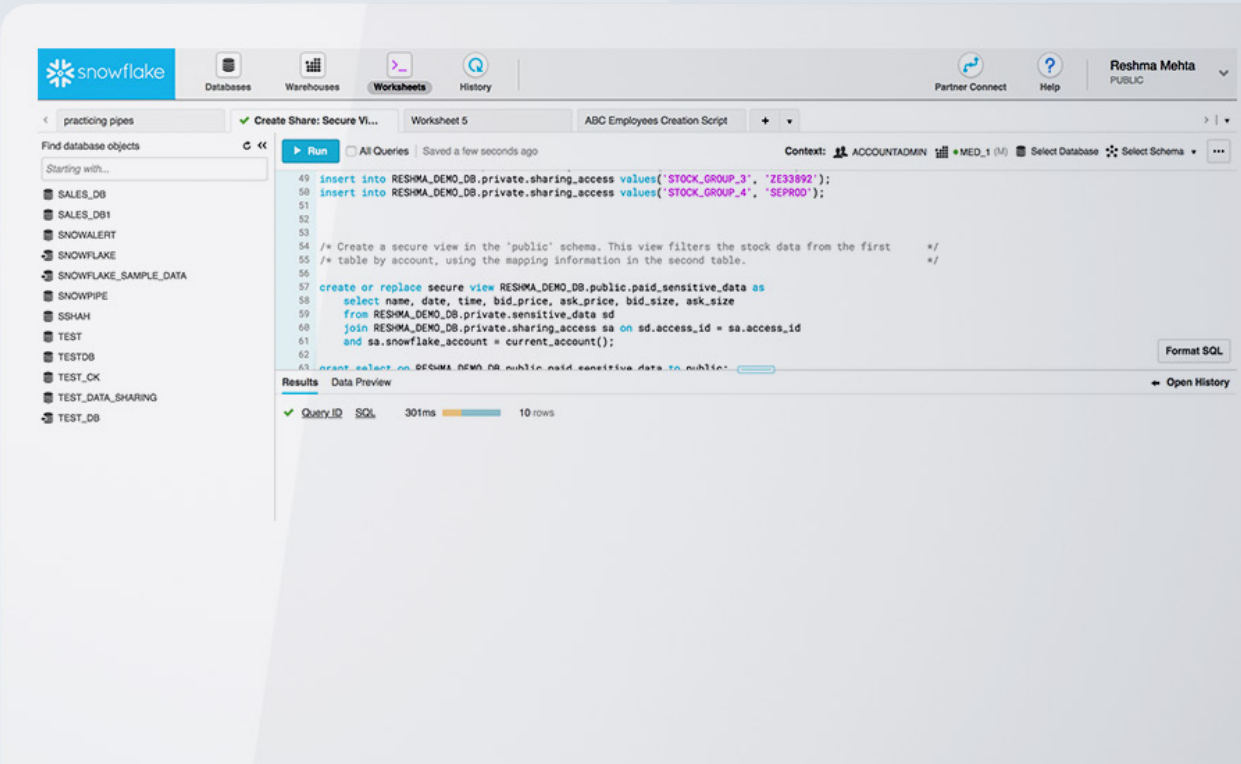
QUICK TIP

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

STEP 6: TEST DRIVE PROGRAMMATIC ACCESS AND SPARK/PYTHON

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 and 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 start, but if you are more adventurous (or have code you would like to try with Snowflake), feel free to download Spark or Python connectors as well.

Click here for more details on programmatic access.

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

> UTILIZE SOME SEMI-STRUCTURED DATA

Snowflake is unique in that it can natively support **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 with **ANSI-standard SQL**, with the addition of dot notation.

To learn more about JSON, check out our video: **Processing JSON data in Snowflake**.

> 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 actually replicating the data.

Watch this video on zero-copy cloning.

> ACCESS HISTORICAL DATA

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

With **Time Travel**, you can:

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

Click here for more details on Time Travel.

```
      ▶ 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
```
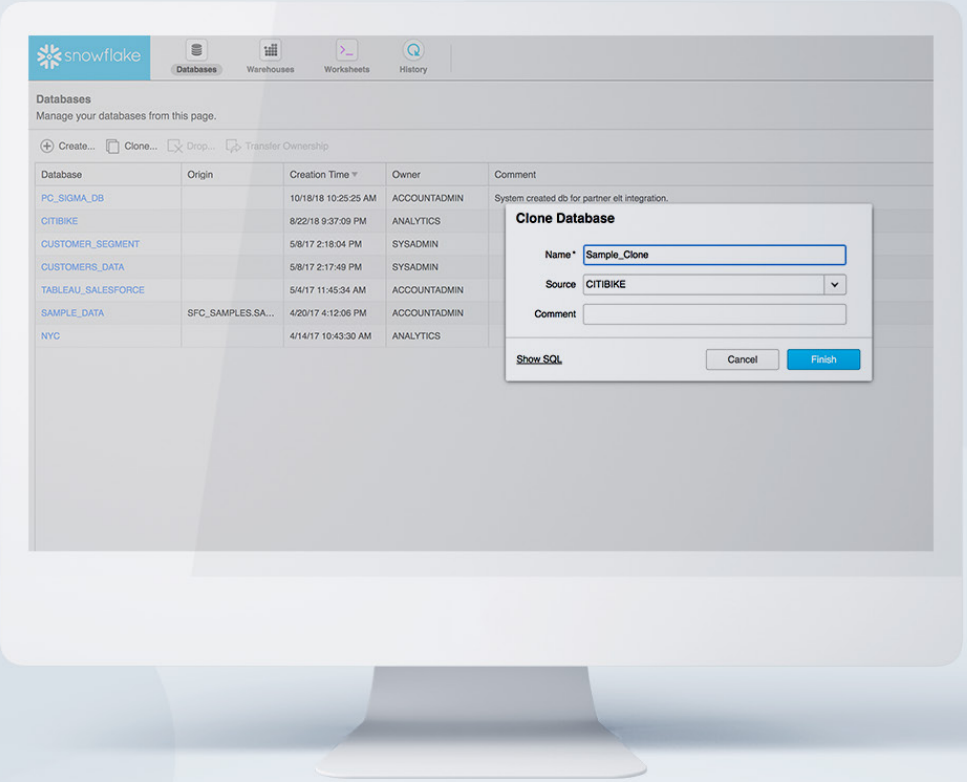
QUICK TIP

Save time and money **cloning data** to avoid the traditional and costly forms of duplicating data.

NEXT
STEPS

After you have used all your credits and are satisfied with what Snowflake can provide, you can continue to use your Snowflake data warehouse right away 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 evaluation, **contact Snowflake** to get answers to your questions.

Also, see the following resources:

- **Snowflake documentation**
- **Snowflake blog**
- **Snowflake Lodge**, the online community and support portal for Snowflake users and partners

ABOUT SNOWFLAKE

Snowflake is the only data warehouse built for the cloud, enabling the data-driven enterprise with instant elasticity, secure data sharing and per-second pricing, across multiple clouds. Snowflake combines the power of data warehousing, the flexibility of big data platforms and the elasticity of the cloud at a fraction of the cost of traditional solutions. Snowflake: Your data, no limits. Find out more at **snowflake.com**