Snowflake Overview

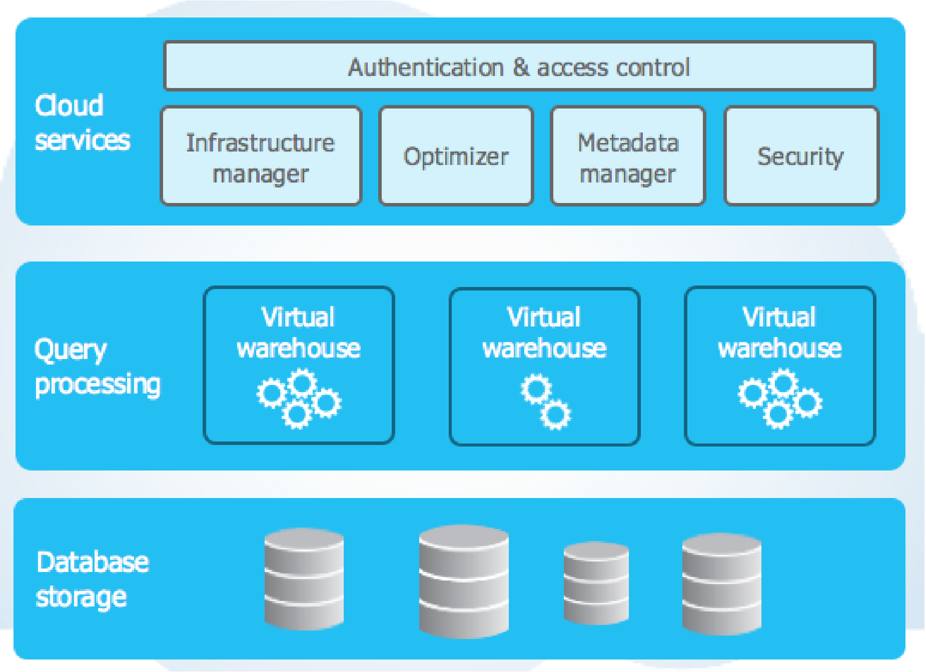
Snowflake is a cloud data platform which provided as Software-as-a-Service (SaaS). Snowflake enables data storage, processing, and analytic solutions that are faster, easier to use, and far more flexible than traditional offerings.

Benefits:

- 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.

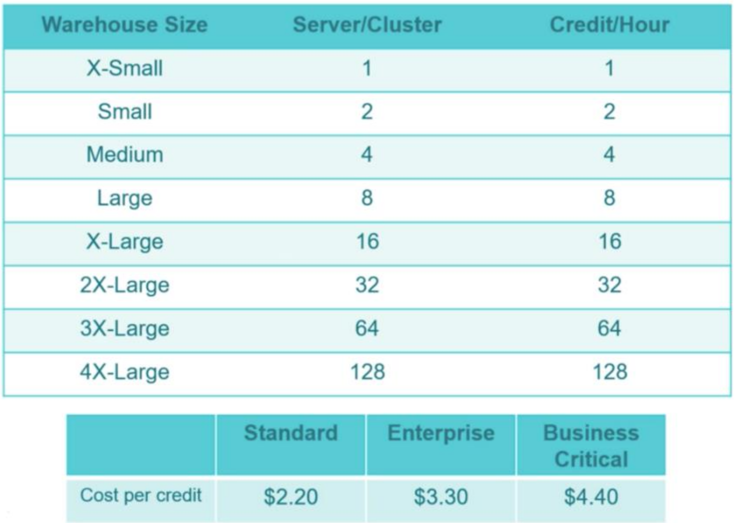
Before we dive into our SQL queries, some key concept we have to know about snowflake platform:
Snowflake's unique architecture consists of three key layers:

- Database Storage
- Query Processing/Compute
- Cloud Services



Database Storage Layer is the layer which is independent of Compute layer, has virtually unlimited storage capacity and data physically stored as micro-partition. Each micro-partitioned file is of size 16 MB and 3 replicas get created to ensure high availability. File format is propriety to snowflake and are immutable. Table definition are logical and stored in its metadata layer.

Query Processing/Compute Layer is the only way to access data from the Database Storage layer. We can create warehouses and adjust the warehouse size at runtime. Every warehouse works independently hence it will not impact another warehouse. Best part is that when idle, compute layer gets auto-suspended saving cost for running the warehouse.



Warehouse Size	Server/Cluster	Credit/Hour
X-Small	1	1
Small	2	2
Medium	4	4
Large	8	8
X-Large	16	16
2X-Large	32	32
3X-Large	64	64
4X-Large	128	128

	Standard	Enterprise	Business Critical
Cost per credit	$2.20	$3.30	$4.40

Cloud Services Layer is the brain of the snowflake. All interaction with snowflake established via this layer, stores query output in a result cache. It's a secure layer which don't allow users to get insights and access metadata.

Creating Snowflake account:

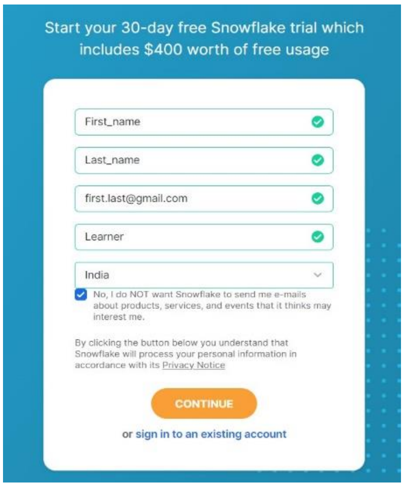
1.Go to Snowflake - https://www.snowflake.com/en/

2.Click on START FOR FREE

3.Fill in the personal details and you are ready to go.

Put your first name, last name, email id for verification process and country.

Click on CONTINUE



Start your 30-day free Snowflake trial which includes $400 worth of free usage

First_name ✓

Last_name ✓

first.last@gmail.com ✓

Learner ✓

India

☑ No, I do NOT want Snowflake to send me e-mails about products, services, and events that it thinks may interest me.

By clicking the button below you understand that Snowflake will process your personal information in accordance with its Privacy Notice

CONTINUE

or sign in to an existing account

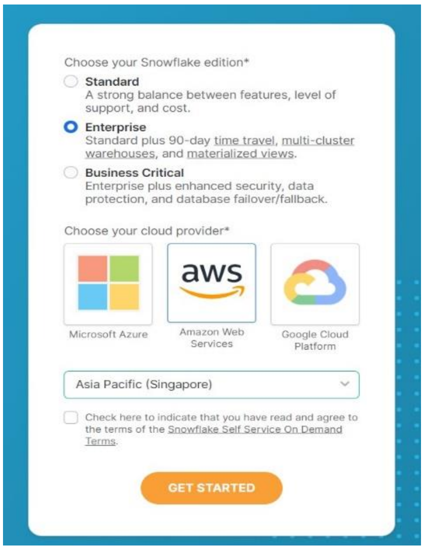
4. In next process select Snowflake edition.

- Standard
- Enterprise
- Business Critical

Then choose cloud provider.

- Microsoft Azure
- Amazon Web Services (AWS)
- Google Cloud Platform (GCP)

For our need we select Enterprise edition and AWS cloud platform.



Choose your Snowflake edition*

○ Standard
A strong balance between features, level of support, and cost.

● Enterprise
Standard plus 90-day time travel, multi-cluster warehouses, and materialized views.

○ Business Critical
Enterprise plus enhanced security, data protection, and database failover/fallback.

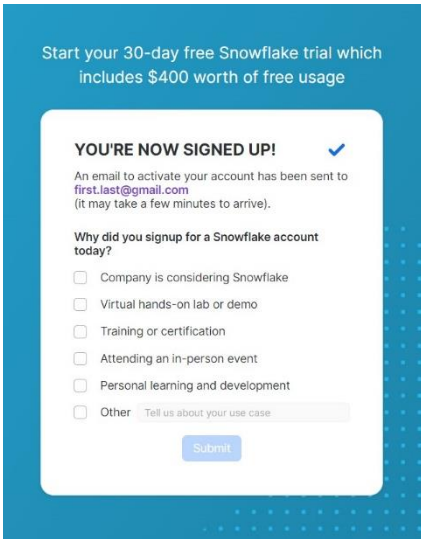
Choose your cloud provider*

Asia Pacific (Singapore)

□ Check here to indicate that you have read and agree to the terms of the Snowflake Self Service On Demand Terms.

5. Next step is verifying the account, setup the user id and password for login to the snowflake.

- Go to email and activate the link.



Start your 30-day free Snowflake trial which includes $400 worth of free usage

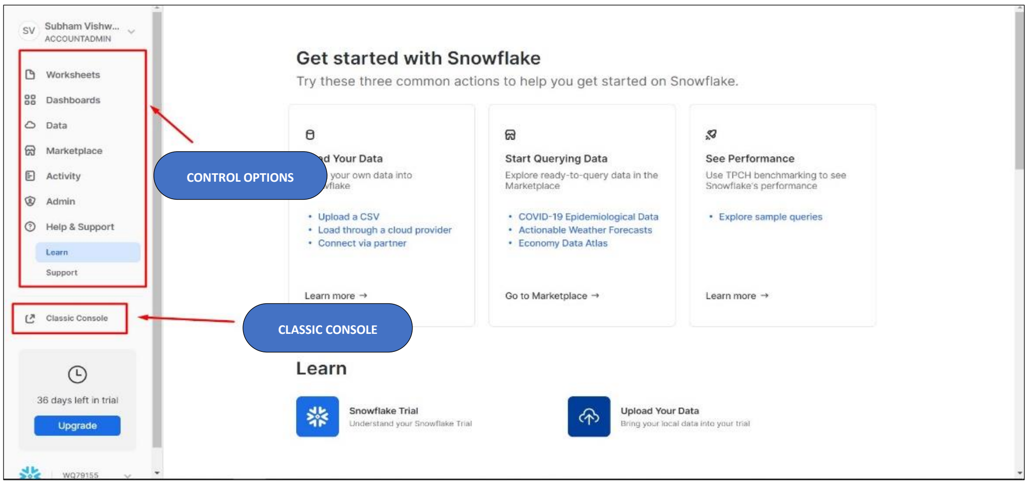
YOU'RE NOW SIGNED UP! ✓

An email to activate your account has been sent to first.last@gmail.com
(it may take a few minutes to arrive).

Why did you signup for a Snowflake account today?

□ Company is considering Snowflake
□ Virtual hands-on lab or demo
□ Training or certification
□ Attending an in-person event
□ Personal learning and development
□ Other Tell us about your use case

6. After this done, we have to login to the snowflake. The window that comes up having bunch of control options on left hand side including Dashboard, Admin, Worksheets, Help & Support etc. And just below there is "Classic Console" which opens a new window of Snowflake classic console where we can write, execute and store our SQL Queries/Codes.



Get started with Snowflake

Try these three common actions to help you get started on Snowflake.

...d Your Data
... your own data into ...wflake
- Upload a CSV
- Load through a cloud provider
- Connect via partner

Learn more →

Start Querying Data
Explore ready-to-query data in the Marketplace
- COVID-19 Epidemiological Data
- Actionable Weather Forecasts
- Economy Data Atlas

Go to Marketplace →

See Performance
Use TPCH benchmarking to see Snowflake's performance
- Explore sample queries

Learn more →

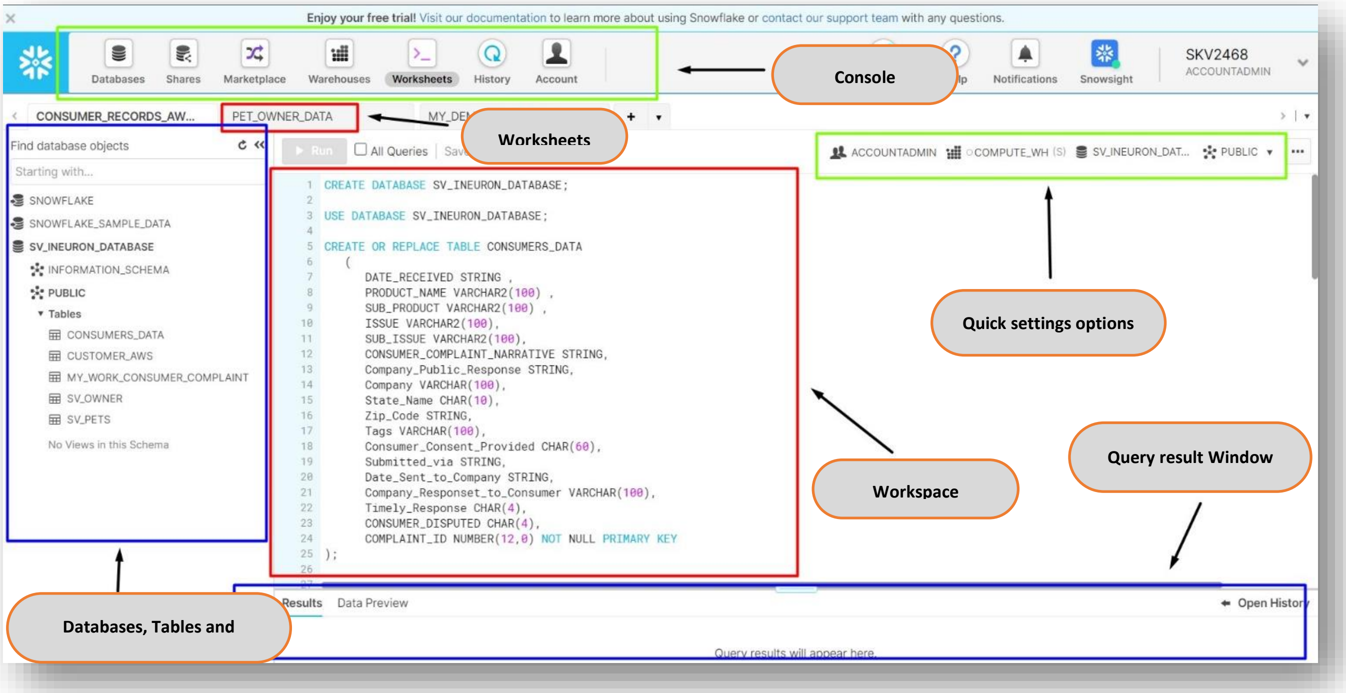
Learn

Snowflake Trial
Understand your Snowflake Trial

Upload Your Data
Bring your local data into your trial

CLASSIC CONSOLE:

This classic console interface is having different options which is used to create, load and share databases maintain warehouse and grant the role to the users as well.



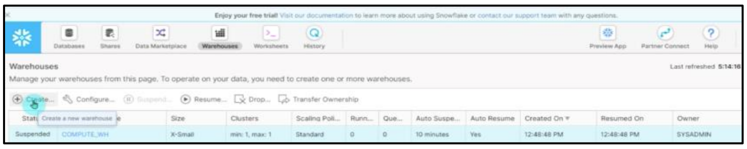
```
CREATE DATABASE SV_INEURON_DATABASE;

USE DATABASE SV_INEURON_DATABASE;

CREATE OR REPLACE TABLE CONSUMERS_DATA
    (
        DATE_RECEIVED STRING ,
        PRODUCT_NAME VARCHAR2(100) ,
        SUB_PRODUCT VARCHAR2(100) ,
        ISSUE VARCHAR2(100),
        SUB_ISSUE VARCHAR2(100),
        CONSUMER_COMPLAINT_NARRATIVE STRING,
        Company_Public_Response STRING,
        Company VARCHAR(100),
        State_Name CHAR(10),
        Zip_Code STRING,
        Tags VARCHAR(100),
        Consumer_Consent_Provided CHAR(60),
        Submitted_via STRING,
        Date_Sent_to_Company STRING,
        Company_Responset_to_Consumer VARCHAR(100),
        Timely_Response CHAR(4),
        CONSUMER_DISPUTED CHAR(4),
        COMPLAINT_ID NUMBER(12,0) NOT NULL PRIMARY KEY
);
```

CREATING WAREHOUSE:

- Go to Warehouses option on the left top bar of snowflake console options.
- Click on Create option as shown below.
- Create new warehouse and fill the following details like name, size, clusters, policy, auto-suspension time etc.

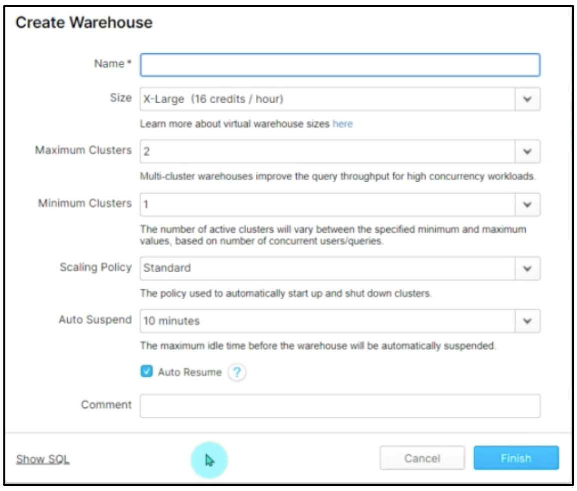


Warehouses

Manage your warehouses from this page. To operate on your data, you need to create one or more warehouses.

Last refreshed 5:14:16

Status	Name	Size	Clusters	Scaling Poli...	Runn...	Que...	Auto Suspe...	Auto Resume	Created On	Resumed On	Owner
Suspended	COMPUTE_WH	X-Small	min: 1, max: 1	Standard	0	0	10 minutes	Yes	12:48:48 PM	12:48:48 PM	SYSADMIN



Create Warehouse

Name *

Size: X-Large (16 credits / hour)

Learn more about virtual warehouse sizes here

Maximum Clusters: 2

Multi-cluster warehouses improve the query throughput for high concurrency workloads.

Minimum Clusters: 1

The number of active clusters will vary between the specified minimum and maximum values, based on number of concurrent users/queries.

Scaling Policy: Standard

The policy used to automatically start up and shut down clusters.

Auto Suspend: 10 minutes

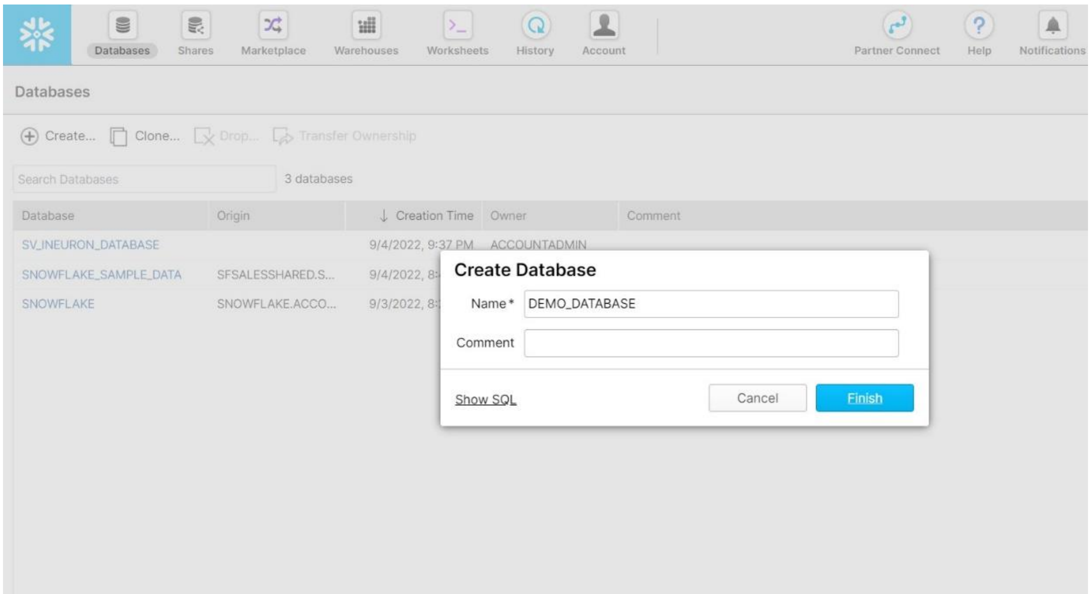
The maximum idle time before the warehouse will be automatically suspended.

☑ Auto Resume

Comment

CREATING DATABASES, TABLES AND SCHEMAS:

- Go to Database option from console options.
- Click on Create option as shown below.
- Enter the name for the database and then click finish.



Databases

3 databases

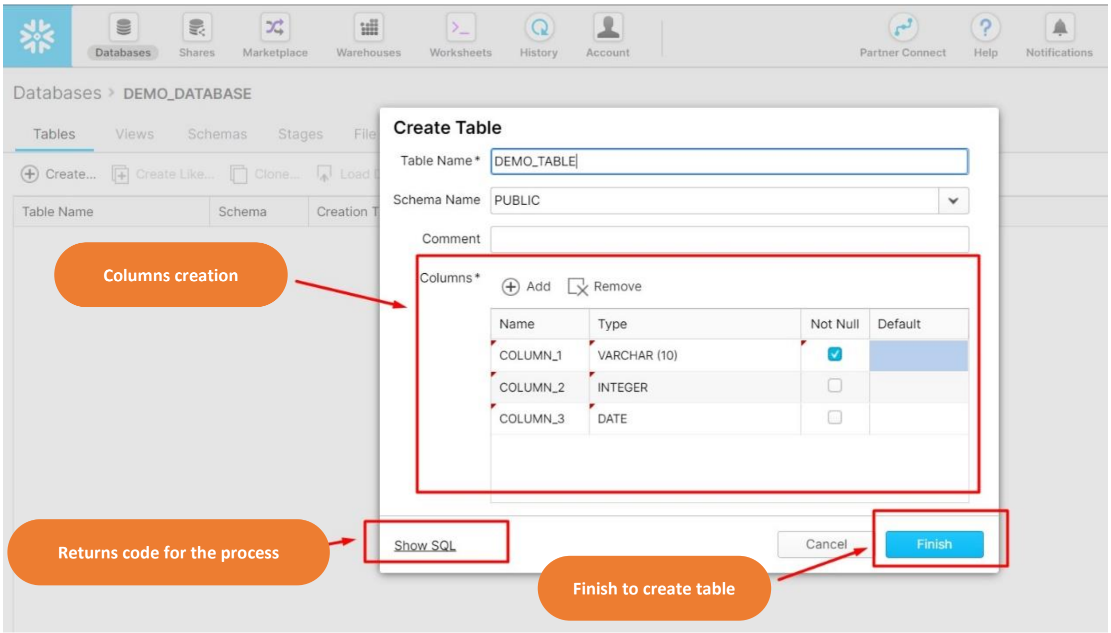
Database	Origin	Creation Time	Owner	Comment
SV_INEURON_DATABASE		9/4/2022, 9:37 PM	ACCOUNTADMIN	
SNOWFLAKE_SAMPLE_DATA	SFSALESSHARED.S...	9/4/2022, 8:		
SNOWFLAKE	SNOWFLAKE.ACCO...	9/3/2022, 8:		

Create Database

Name * DEMO_DATABASE

Comment

- Select DEMO_DATABASE as created earlier from database console option.
- Click on Table tab in database.
- Click on Create as shown below, add columns as per requirements and create the table.



Databases > DEMO_DATABASE

Create Table

Table Name * DEMO_TABLE

Schema Name PUBLIC

Comment

Columns *

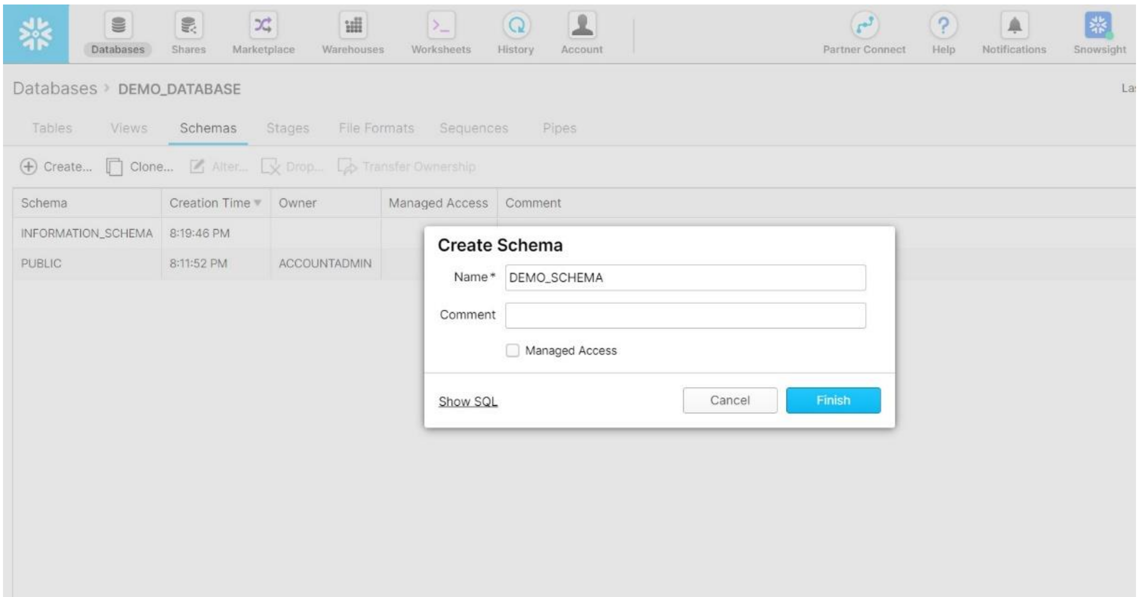
Name	Type	Not Null	Default
COLUMN_1	VARCHAR (10)	☑	
COLUMN_2	INTEGER	☐	
COLUMN_3	DATE	☐	

Columns creation

Returns code for the process

Finish to create table

Similarly, we can create schema in the same fashion.



Databases › DEMO_DATABASE

Schema	Creation Time	Owner	Managed Access	Comment
INFORMATION_SCHEMA	8:19:46 PM			
PUBLIC	8:11:52 PM	ACCOUNTADMIN		

Create Schema

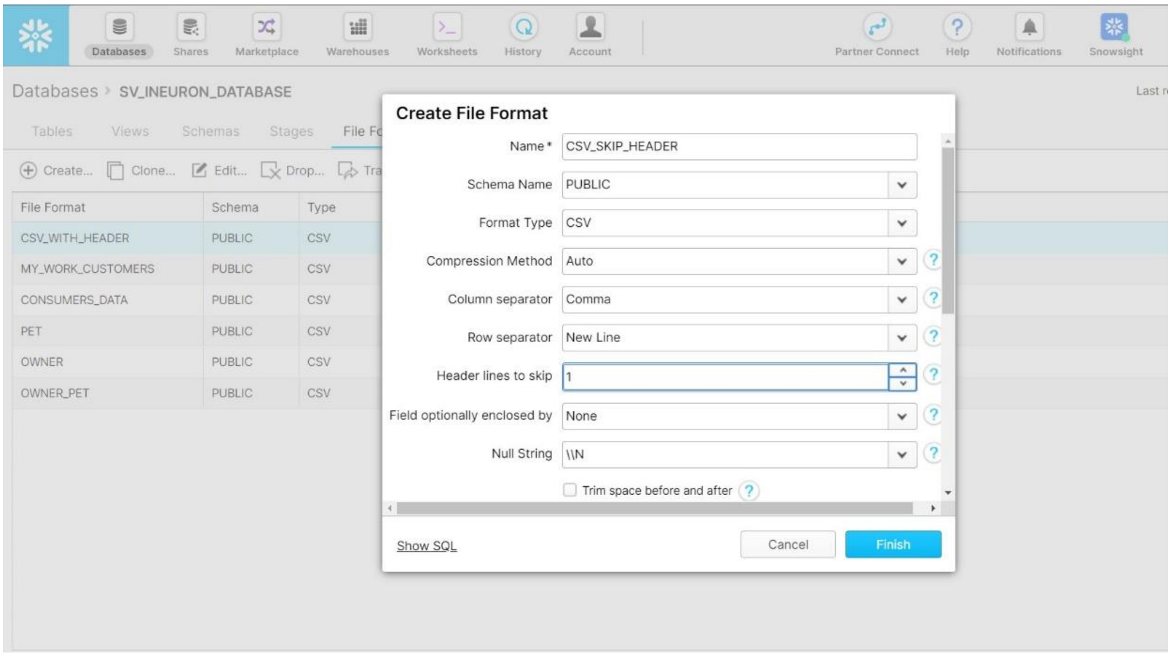
Name* DEMO_SCHEMA

Comment

□ Managed Access

CREATING FILE FORMATS:

We have to create file format, so to load data from different file formats like csv, json, xml etc. If we have header(features) available in the file then we have to skip headers lines.



Databases › SV_INEURON_DATABASE

File Format	Schema	Type
CSV_WITH_HEADER	PUBLIC	CSV
MY_WORK_CUSTOMERS	PUBLIC	CSV
CONSUMERS_DATA	PUBLIC	CSV
PET	PUBLIC	CSV
OWNER	PUBLIC	CSV
OWNER_PET	PUBLIC	CSV

Create File Format

Name* CSV_SKIP_HEADER

Schema Name PUBLIC

Format Type CSV

Compression Method Auto

Column separator Comma

Row separator New Line

Header lines to skip 1

Field optionally enclosed by None

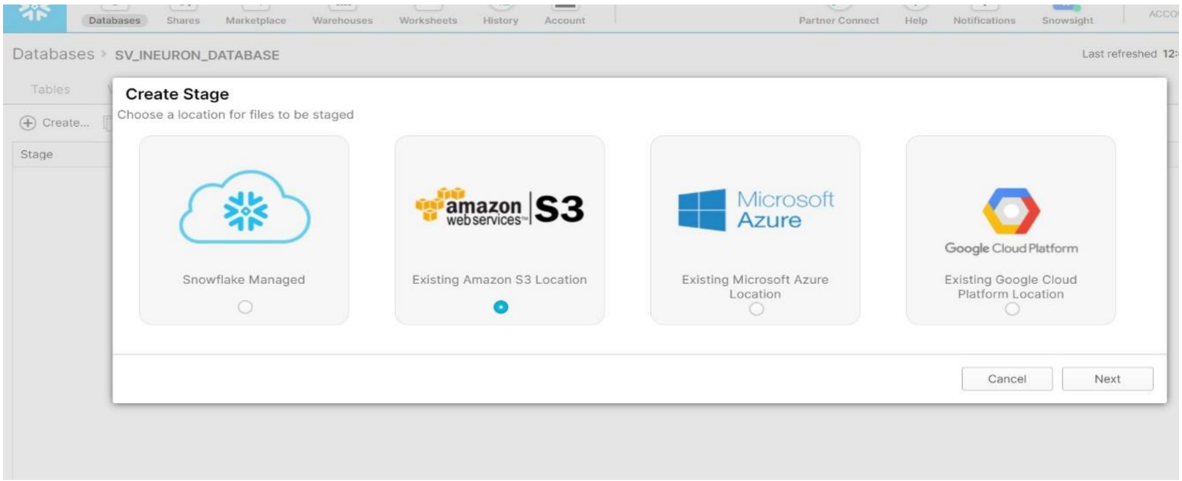
Null String \\N

□ Trim space before and after

CREATING STAGES:

Stages in snowflake is the area where data gets loaded before ingesting into the tables. Stages can be internal or externally located on the cloud or local system. There are 4 staging options we can observe as of now.

- Snowflake Managed
- Amazon S3
- Microsoft Azure
- Google Cloud Platform



Create Stage

Choose a location for files to be staged

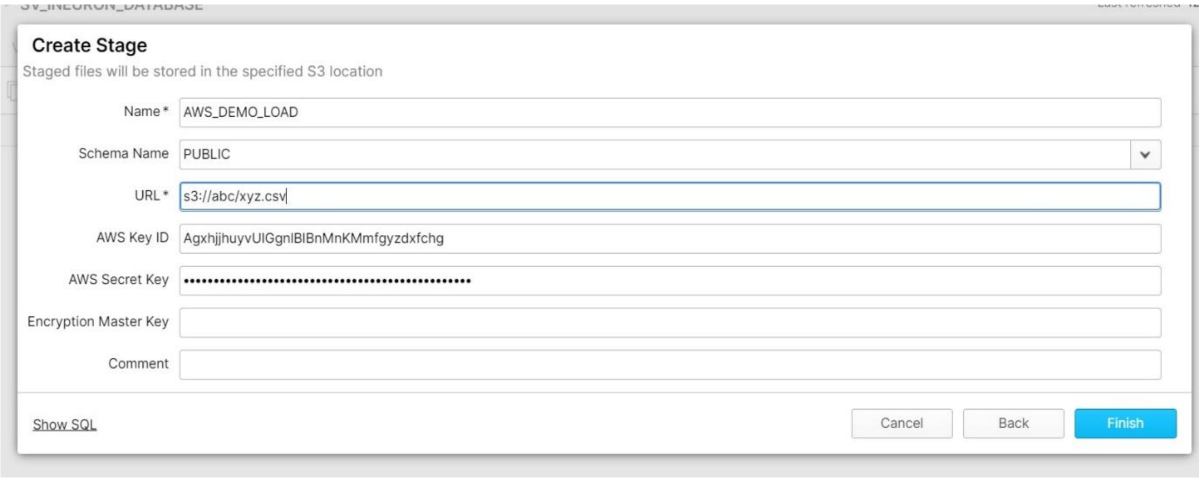
○ Snowflake Managed

● Existing Amazon S3 Location

○ Existing Microsoft Azure Location

○ Existing Google Cloud Platform Location

- Click on cloud stage as per need say we need Amazon S3 service.
- Stage creation window will pop up.
- Fill in the details as needed (maybe provided by client who handles the cloud environment) and click finish.



Create Stage

Staged files will be stored in the specified S3 location

Name * AWS_DEMO_LOAD

Schema Name PUBLIC

URL * s3://abc/xyz.csv

AWS Key ID AgxhjjhuyvUIGgnIBIBnMnKMmfgyzdxfchg

AWS Secret Key ••

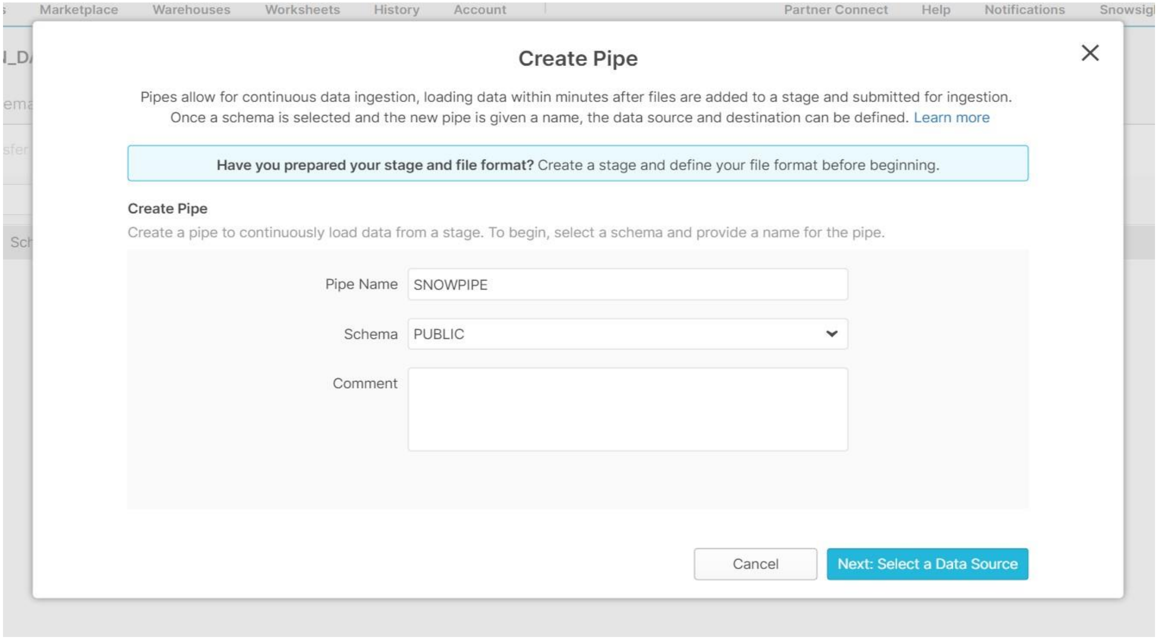
Encryption Master Key

Comment

CREATING SNOWPIPES:

Snowpipe/pipe is created to load continuous data from the cloud environment. Also, we have to configure the cloud environment to make it work efficiently like access permissions for S3 bucket, notifications to automate pipe, identity and access management (IAM) in case of AWS, configure Simple Notification Service (SNS) and Simple Queue Service (SQS) etc.

Pipes are be created under database, in pipe tab as shown below:



Create Pipe

Pipes allow for continuous data ingestion, loading data within minutes after files are added to a stage and submitted for ingestion. Once a schema is selected and the new pipe is given a name, the data source and destination can be defined. Learn more

Have you prepared your stage and file format? Create a stage and define your file format before beginning.

Create Pipe

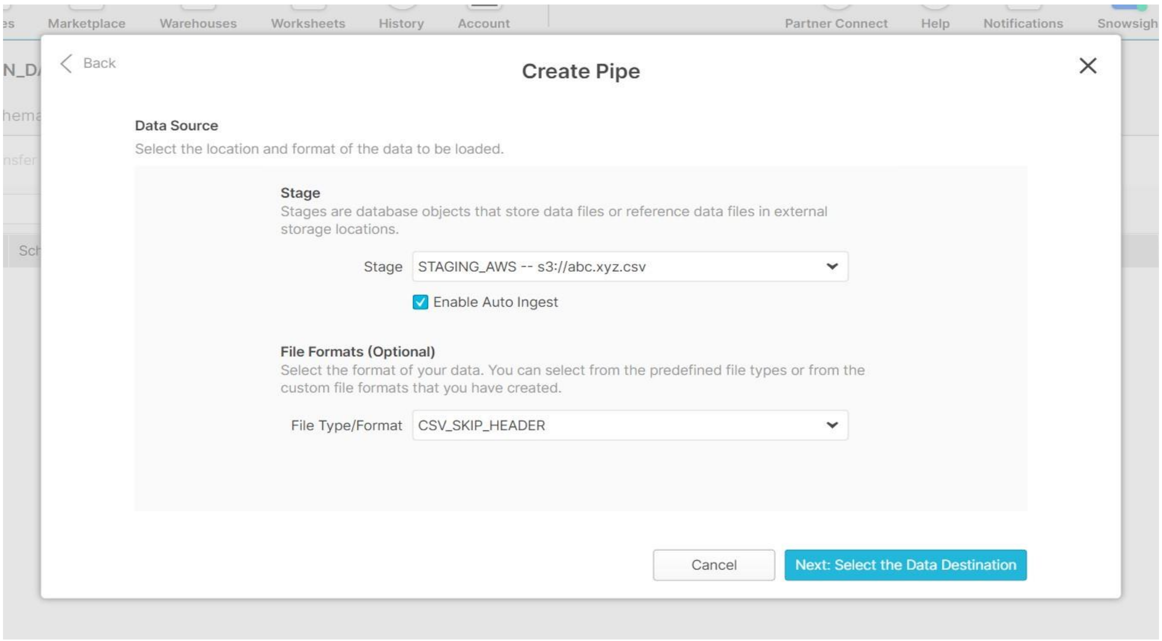
Create a pipe to continuously load data from a stage. To begin, select a schema and provide a name for the pipe.

Pipe Name: SNOWPIPE

Schema: PUBLIC

Comment:

Select the stage and file type as shown below and check mark the Enable Auto Ingest.



Create Pipe

Data Source

Select the location and format of the data to be loaded.

Stage

Stages are database objects that store data files or reference data files in external storage locations.

Stage: STAGING_AWS -- s3://abc.xyz.csv

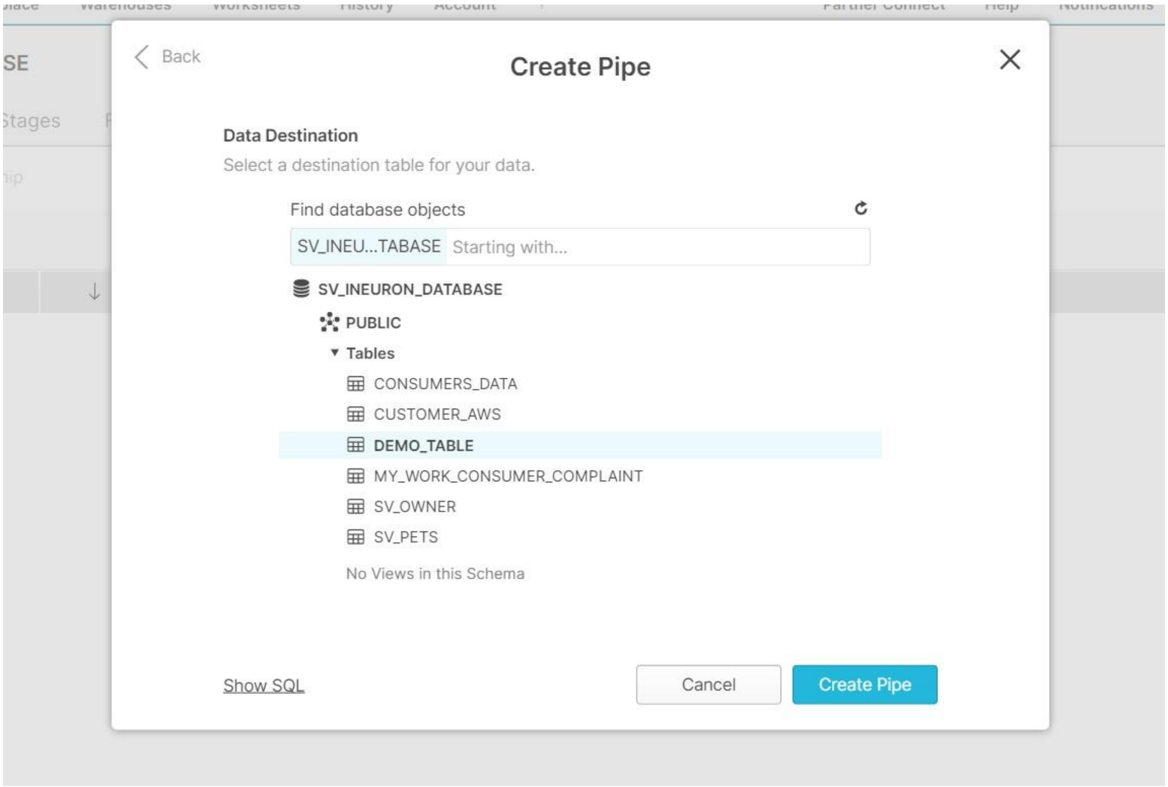
☑ Enable Auto Ingest

File Formats (Optional)

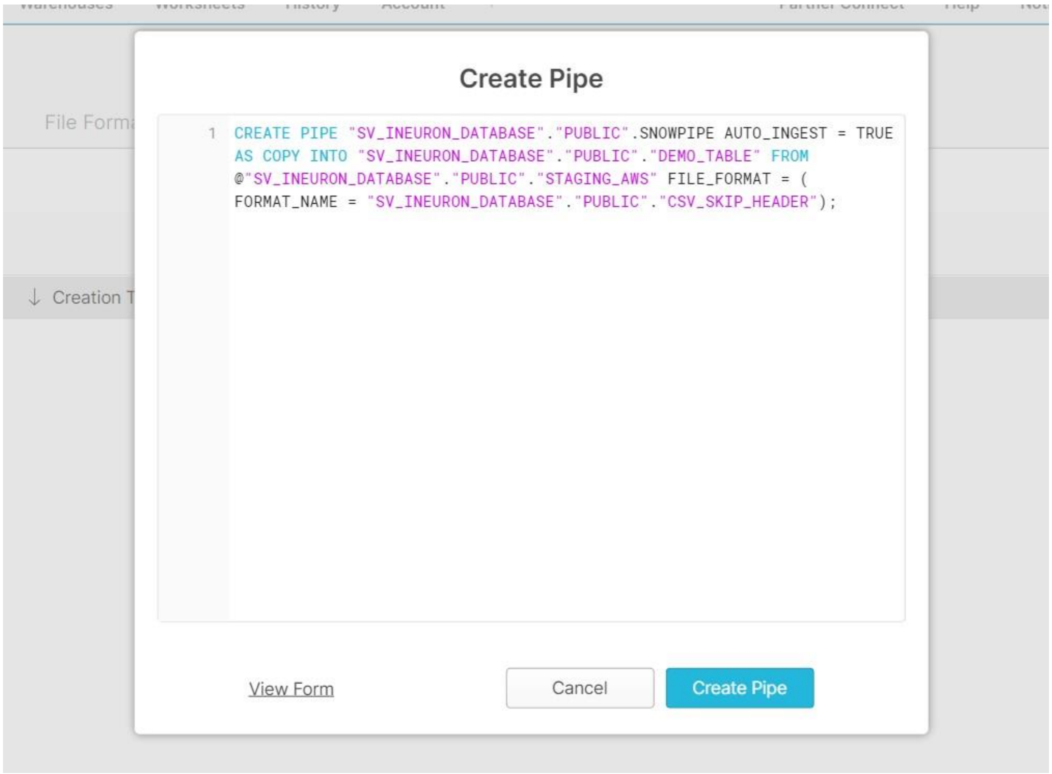
Select the format of your data. You can select from the predefined file types or from the custom file formats that you have created.

File Type/Format: CSV_SKIP_HEADER

Then in next step we have to select the table where we have to load the continuous data and then our pipe is created.



If we like to view or edit the entire pipe creation code then we can go to the Show SQL option on the left bottom of the pop-up window and do the changes if required.



```
CREATE PIPE "SV_INEURON_DATABASE"."PUBLIC".SNOWPIPE AUTO_INGEST = TRUE
AS COPY INTO "SV_INEURON_DATABASE"."PUBLIC"."DEMO_TABLE" FROM
@"SV_INEURON_DATABASE"."PUBLIC"."STAGING_AWS" FILE_FORMAT = (
FORMAT_NAME = "SV_INEURON_DATABASE"."PUBLIC"."CSV_SKIP_HEADER");
```

So, to load the data we have two options i.e., Bulk Loading and Continuous Loading.

Bulk Loading is used to load the data from Stage using copy command, it relies on the user provisioned compute resources i.e., Data Warehouse and supports simple transformations.

Continuous Loading is used to load data by creating the Snowpipe, it doesn’t rely on the user provisioned compute resource but provisioned internally by Snowflake and both simple and complex transformations can be done.

So, this is the short overview of the snowflake. Hope you like this work and if you want to go in depth of snowflake structure and data staging and creating pipe for different cloud platform then please refer the references attached below.

References:

https://docs.snowflake.com/en/index.html