Staging Data into Snowflake

Data Sheet | LumenData

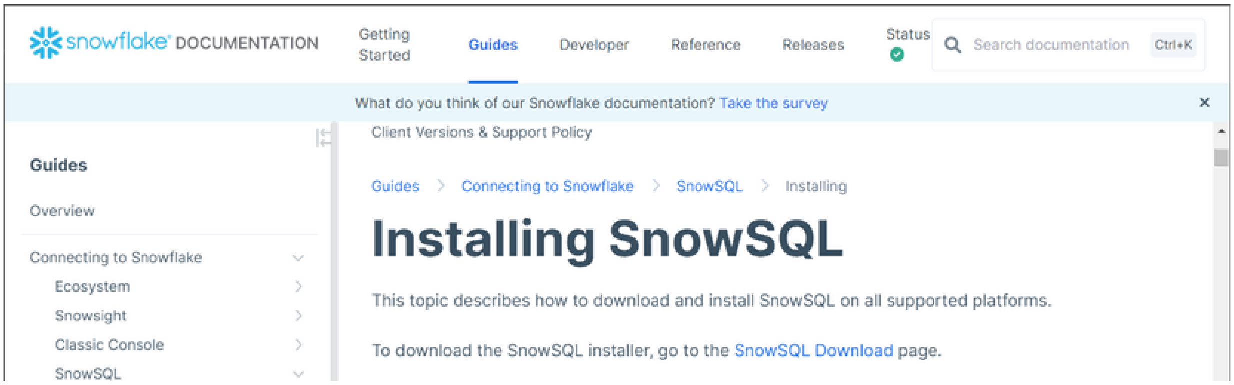
A comprehensive guide involving an introduction to internal staging, followed by a detailed explanation of how to perform external staging from AWS S3 bucket to Snowflake.

REQUIREMENTS:

1. **Snowflake account:** Click on the link to create a Snowflake free trial account.

2. **Snow SQL**

2.1 Go to the official Snowflake page to install



What do you think of our Snowflake documentation? Take the survey

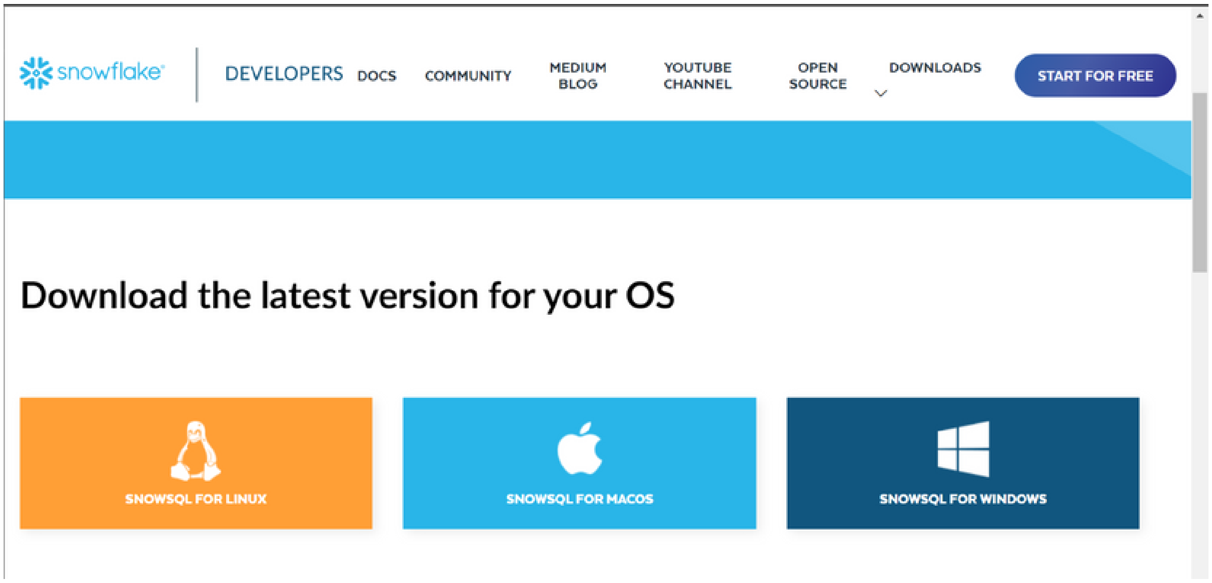
Installing SnowSQL

This topic describes how to download and install SnowSQL on all supported platforms.

To download the SnowSQL installer, go to the SnowSQL Download page.

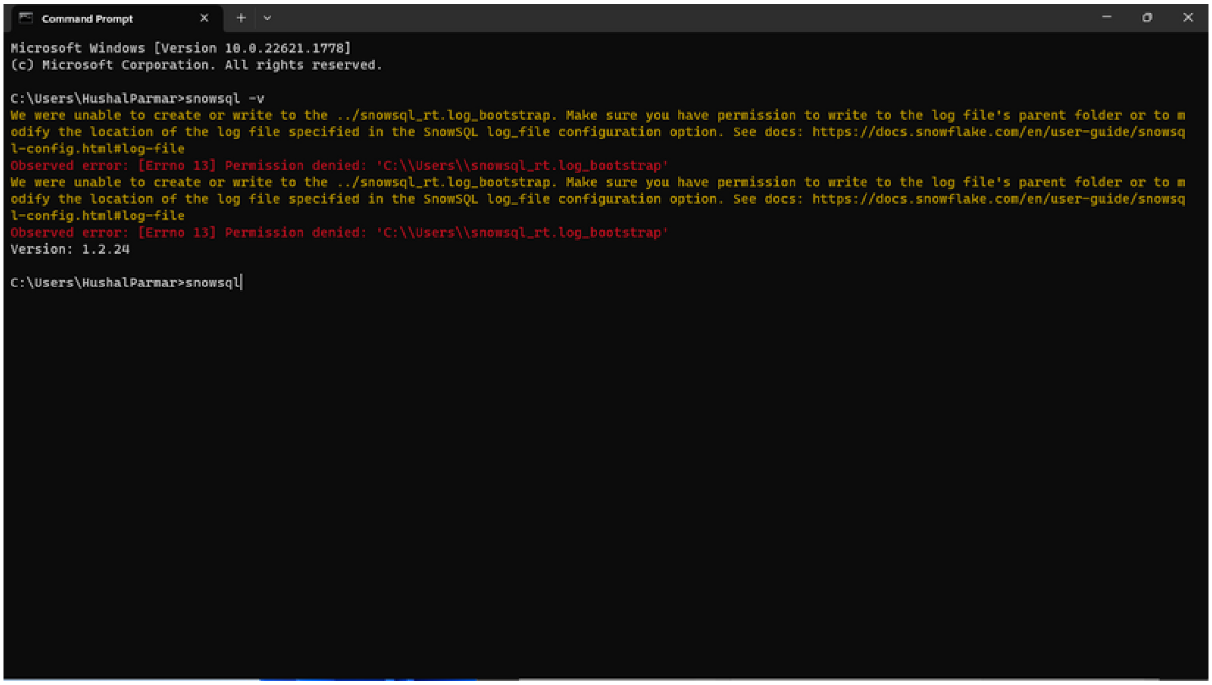
LUMENDATA
Masters in Data Management

2.2 Click the “SnowSQL Download” link. And download the latest version for your OS.



Download the latest version for your OS

2.3 You can check if Snow SQL is installed in your system by running the following command in command prompt. “SnowSQL -v”



```
Microsoft Windows [Version 10.0.22621.1778]
(c) Microsoft Corporation. All rights reserved.

C:\Users\HushalParmar>snowsql -v
We were unable to create or write to the ../snowsql_rt.log_bootstrap. Make sure you have permission to write to the log file's parent folder or to m
odify the location of the log file specified in the SnowSQL log_file configuration option. See docs: https://docs.snowflake.com/en/user-guide/snowsq
l-config.html#log-file
Observed error: [Errno 13] Permission denied: 'C:\\Users\\snowsql_rt.log_bootstrap'
We were unable to create or write to the ../snowsql_rt.log_bootstrap. Make sure you have permission to write to the log file's parent folder or to m
odify the location of the log file specified in the SnowSQL log_file configuration option. See docs: https://docs.snowflake.com/en/user-guide/snowsq
l-config.html#log-file
Observed error: [Errno 13] Permission denied: 'C:\\Users\\snowsql_rt.log_bootstrap'
Version: 1.2.24

C:\Users\HushalParmar>snowsql
```

INTERNAL STAGE:

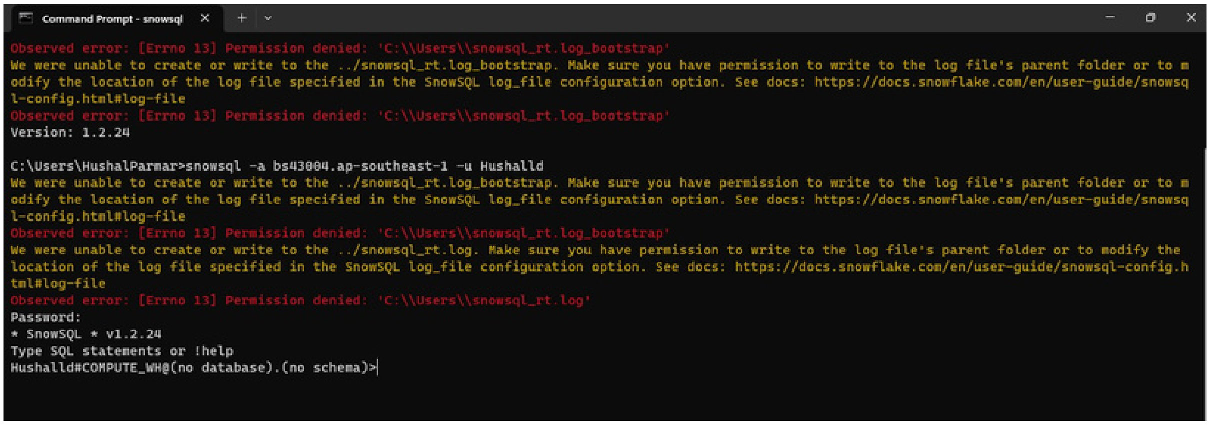
The data is stored internally. There are 3 types of Internal Stages:

- Each user’s Personal Storage Area is called the User Stage. These Stages are personal to the user, which means no one else can see them. By default, each user is assigned a User Stage, which cannot be changed or removed.
- Within a Table Object, Table Stages are Storage locations. This is useful when only a few files need to be imported into a certain table, and these Stages will get the job done the quickest. They are, however, limited to that one table, and other tables cannot access the files.
- Within a Snowflake Database/Schema, Internal Named Stages are Storage Location Objects. Since they are Database Objects, they are subject to the same Security Permissions as other Database Objects. Unlike User and Table Stages, these Stages are not created automatically. They do, however, offer more versatility when it comes to importing files into different Tables and/or allowing Multiple Users to access the same Stage.

To perform Internal Stage, you need to follow the steps given below

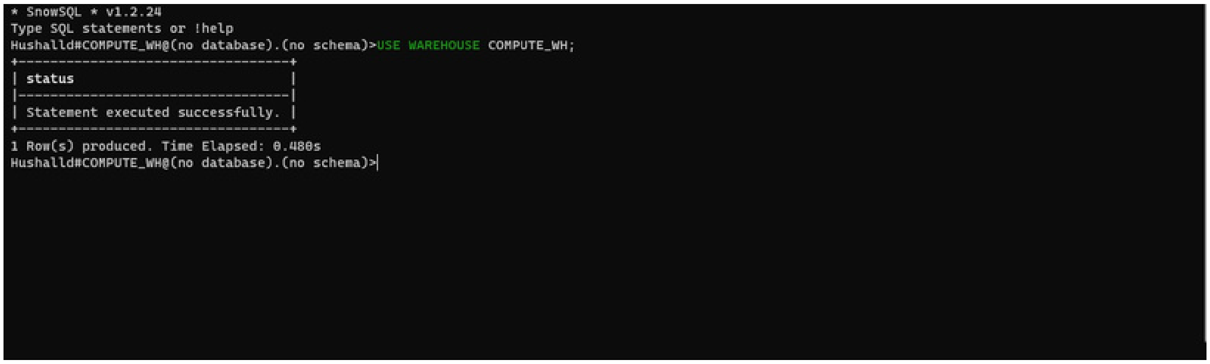
1. Sign in to our Snowflake account by following the command(example) “-a bs43004.ap-southeast-1 -u Hushalld” In this command “-a” stands for account accessing point and “-u” stands for the username of the Snowflake account. Press ‘enter’ and put the password.

2. After entering the password, you will get a prompt to select the database and schema.



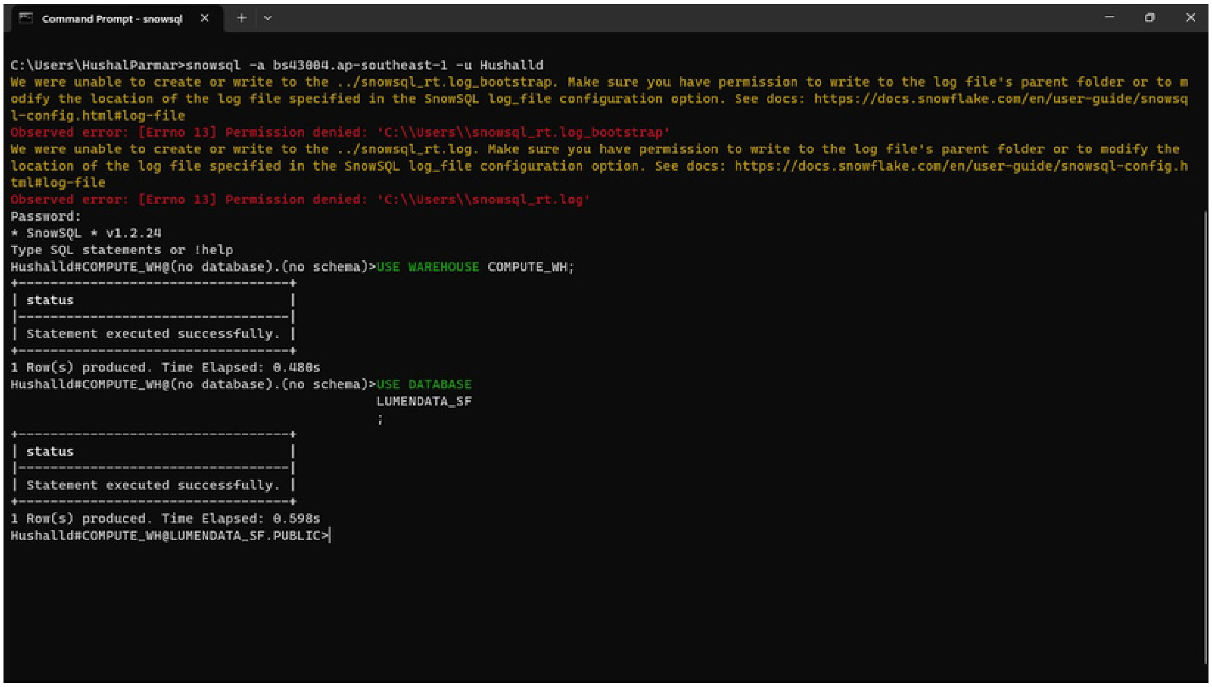
```
Observed error: [Errno 13] Permission denied: 'C:\\Users\\snowsql_rt.log_bootstrap'
We were unable to create or write to the ../snowsql_rt.log_bootstrap. Make sure you have permission to write to the log file's parent folder or to m
odify the location of the log file specified in the SnowSQL log_file configuration option. See docs: https://docs.snowflake.com/en/user-guide/snowsq
l-config.html#log-file
Observed error: [Errno 13] Permission denied: 'C:\\Users\\snowsql_rt.log_bootstrap'
Version: 1.2.24

C:\Users\HushalParmar>snowsql -a bs43004.ap-southeast-1 -u Hushalld
We were unable to create or write to the ../snowsql_rt.log_bootstrap. Make sure you have permission to write to the log file's parent folder or to m
odify the location of the log file specified in the SnowSQL log_file configuration option. See docs: https://docs.snowflake.com/en/user-guide/snowsq
l-config.html#log-file
Observed error: [Errno 13] Permission denied: 'C:\\Users\\snowsql_rt.log_bootstrap'
We were unable to create or write to the ../snowsql_rt.log. Make sure you have permission to write to the log file's parent folder or to modify the
location of the log file specified in the SnowSQL log_file configuration option. See docs: https://docs.snowflake.com/en/user-guide/snowsql-config.h
tml#log-file
Observed error: [Errno 13] Permission denied: 'C:\\Users\\snowsql_rt.log'
Password:
* SnowSQL * v1.2.24
Type SQL statements or !help
Hushalld#COMPUTE_WH@(no database).(no schema)>
```

3. You can select “USE WAREHOUSE followed by the warehouse name” to specify the warehouse name to be used.



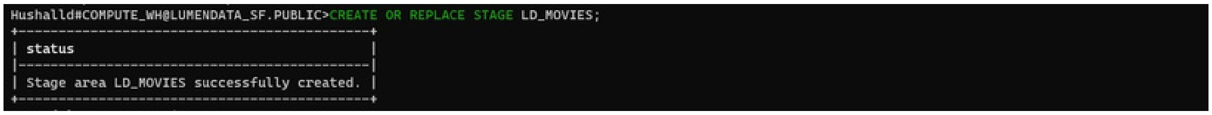
```
* SnowSQL * v1.2.24
Type SQL statements or !help
Hushalld#COMPUTE_WH@(no database).(no schema)>USE WAREHOUSE COMPUTE_WH;
+----------------------------------+
| status                           |
|----------------------------------|
| Statement executed successfully. |
+----------------------------------+
1 Row(s) produced. Time Elapsed: 0.480s
Hushalld#COMPUTE_WH@(no database).(no schema)>
```

4. Then, you need to specify the Database name using “Use Database and followed by the database name”.



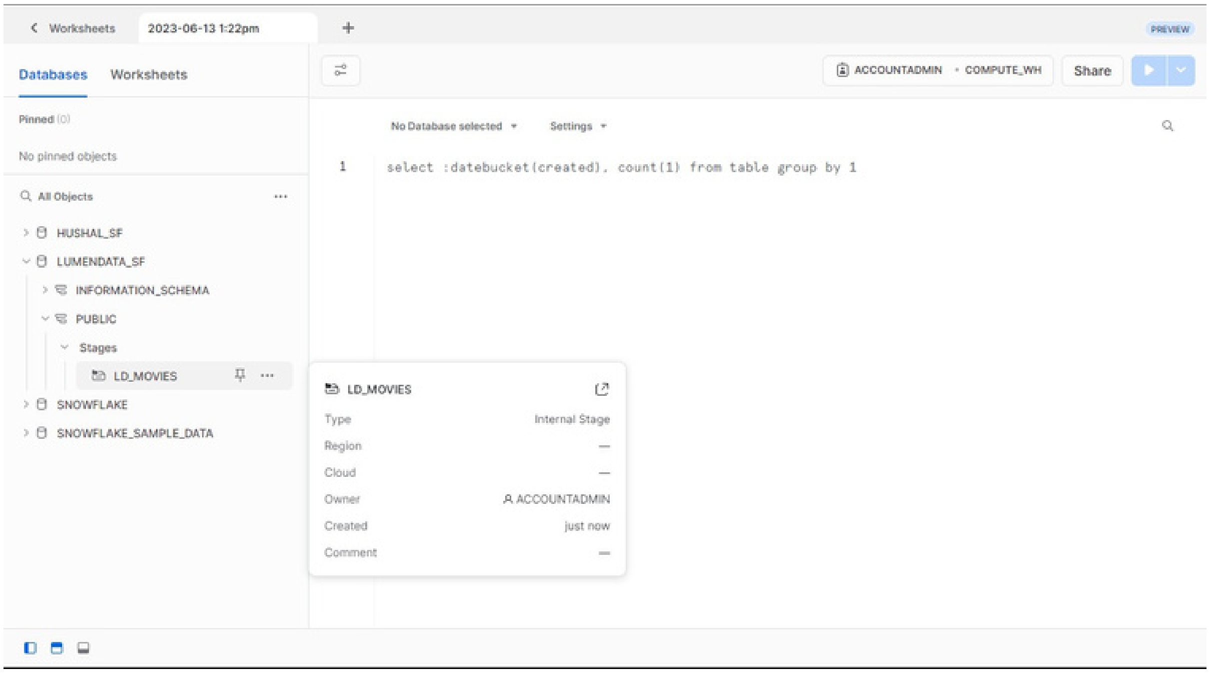
```
C:\Users\HushalParmar>snowsql -a bs43004.ap-southeast-1 -u Hushalld
We were unable to create or write to the ../snowsql_rt.log_bootstrap. Make sure you have permission to write to the log file's parent folder or to m
odify the location of the log file specified in the SnowSQL log_file configuration option. See docs: https://docs.snowflake.com/en/user-guide/snowsq
l-config.html#log-file
Observed error: [Errno 13] Permission denied: 'C:\\Users\\snowsql_rt.log_bootstrap'
We were unable to create or write to the ../snowsql_rt.log. Make sure you have permission to write to the log file's parent folder or to modify the
location of the log file specified in the SnowSQL log_file configuration option. See docs: https://docs.snowflake.com/en/user-guide/snowsql-config.h
tml#log-file
Observed error: [Errno 13] Permission denied: 'C:\\Users\\snowsql_rt.log'
Password:
* SnowSQL * v1.2.24
Type SQL statements or !help
Hushalld#COMPUTE_WH@(no database).(no schema)>USE WAREHOUSE COMPUTE_WH;
+----------------------------------+
| status                           |
|----------------------------------|
| Statement executed successfully. |
+----------------------------------+
1 Row(s) produced. Time Elapsed: 0.480s
Hushalld#COMPUTE_WH@(no database).(no schema)>USE DATABASE
                                              LUMENDATA_SF
                                              ;
+----------------------------------+
| status                           |
|----------------------------------|
| Statement executed successfully. |
+----------------------------------+
1 Row(s) produced. Time Elapsed: 0.598s
Hushalld#COMPUTE_WH@LUMENDATA_SF.PUBLIC>
```

5. Here, you can create the stages by using the following command “CREATE OR REPLACE STAGE FOLLOWED BY STAGE NAME”.



```
Hushalld#COMPUTE_WH@LUMENDATA_SF.PUBLIC>CREATE OR REPLACE STAGE LD_MOVIES;
+-------------------------------------------+
| status                                    |
|-------------------------------------------|
| Stage area LD_MOVIES successfully created. |
+-------------------------------------------+
```

6. You can check the same in Snowflake web UI after signing in to web UI.



```
select :datebucket(created), count(1) from table group by 1
```

LD_MOVIES

Type Internal Stage

Region —

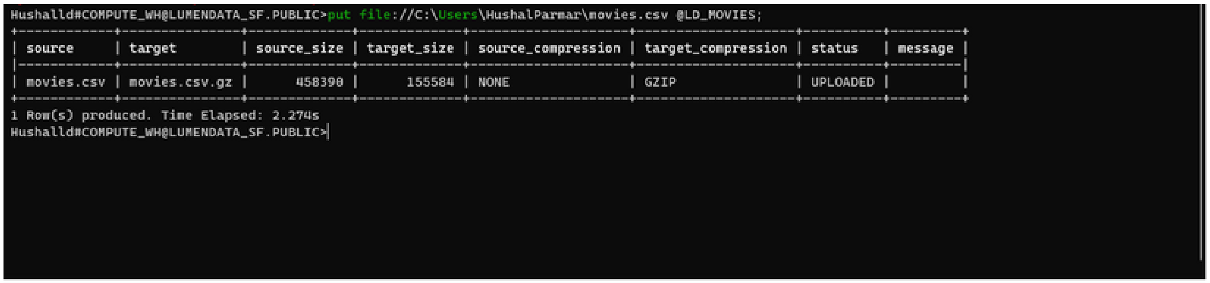
Cloud —

Owner ACCOUNTADMIN

Created just now

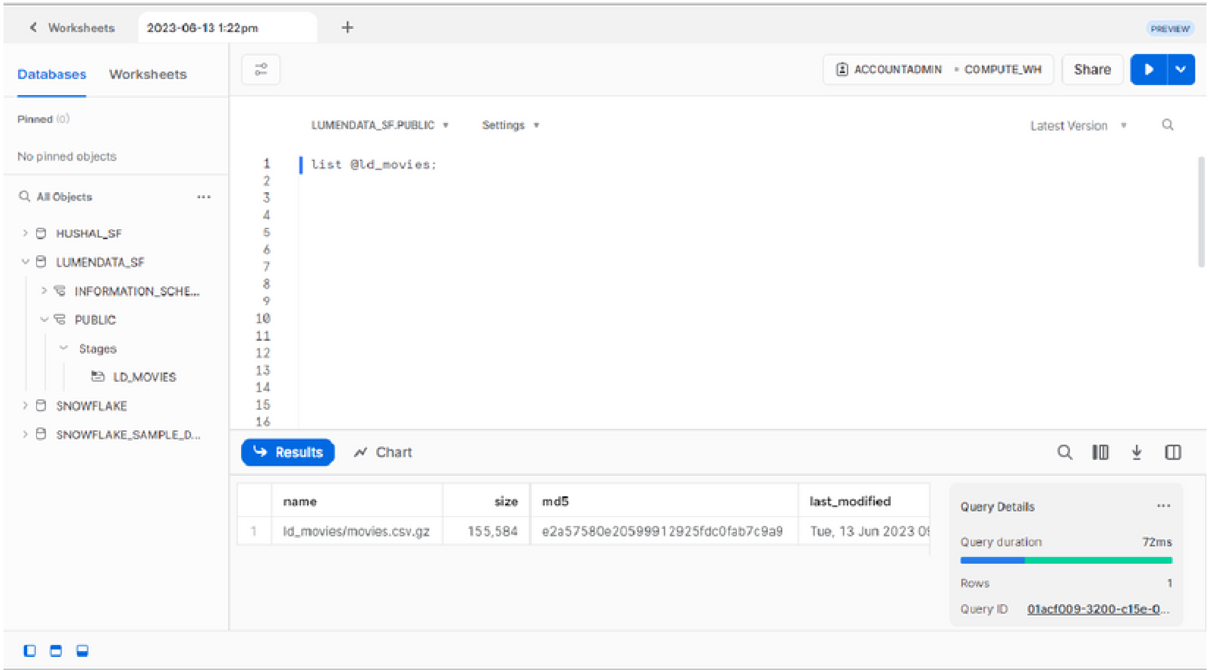
Comment —

7. You need to “Put” API call in command prompt with the CSV file location form your local system.
” put file://C: \Users\HushalParmar\movies.csv @LD_MOVIES;”



```
Hushalld#COMPUTE_WH@LUMENDATA_SF.PUBLIC>put file://C:\Users\HushalParmar\movies.csv @LD_MOVIES;
+------------+---------------+-------------+-------------+--------------------+--------------------+----------+---------+
| source     | target        | source_size | target_size | source_compression | target_compression | status   | message |
|------------+---------------+-------------+-------------+--------------------+--------------------+----------+---------|
| movies.csv | movies.csv.gz |      458398 |      155584 | NONE               | GZIP               | UPLOADED |         |
+------------+---------------+-------------+-------------+--------------------+--------------------+----------+---------+
1 Row(s) produced. Time Elapsed: 2.274s
Hushalld#COMPUTE_WH@LUMENDATA_SF.PUBLIC>
```

8. Check the Snowflake dashboard if the file is loaded in the right stage. “list @ld_movies;



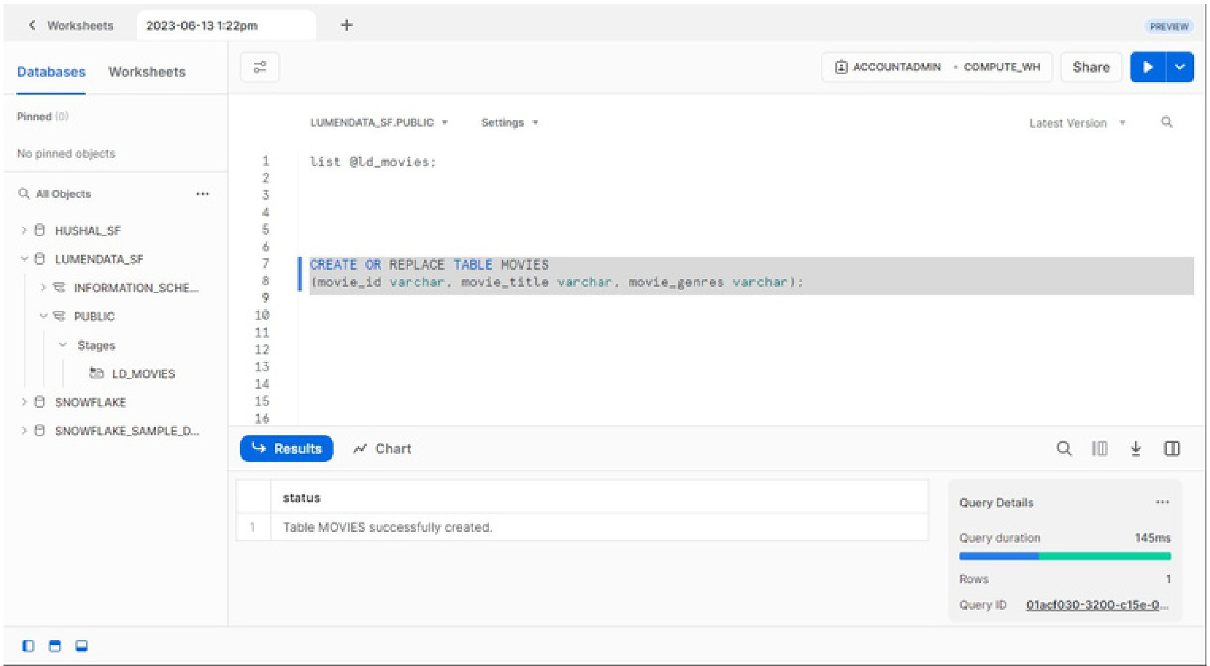
```
list @ld_movies;
```

	name	size	md5	last_modified
1	ld_movies/movies.csv.gz	155,584	e2a57580e20599912925fdc0fab7c9a9	Tue, 13 Jun 2023 0

Query Details
Query duration 72ms
Rows 1
Query ID 01acf009-3200-c15e-0...

9. You will have to create a table where you can store data and utilize it. After internal staging, it is a convenient option if your files need to be accessible to multiple users and only need to be copied into a single table.

To support bulk loading of the data into tables, Snowflake utilizes stages where the files containing the data to be loaded are stored. Snowflake supports both internal stages and external stages.
” CREATE OR REPLACE TABLE MOVIES
(movie_id varchar, movie_title varchar, movie_genres varchar);”



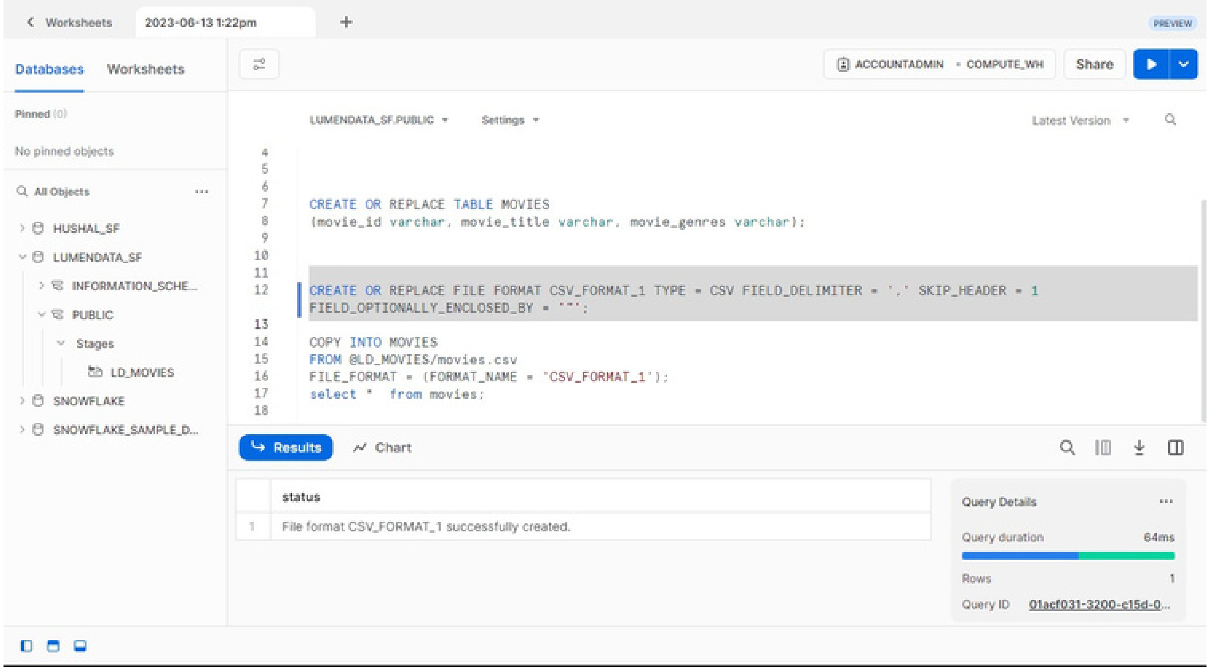
```
list @ld_movies;



CREATE OR REPLACE TABLE MOVIES
(movie_id varchar, movie_title varchar, movie_genres varchar);
```

	status
1	Table MOVIES successfully created.

Query Details
Query duration 145ms
Rows 1
Query ID 01acf030-3200-c15e-0...

10. Next, you can create a file format. Creating file formats in Snowflake is useful when you have data in a specific format that you want to load into Snowflake tables.

CREATE OR REPLACE FILE FORMAT CSV_FORMAT_1 TYPE = CSV
FIELD_DELIMITER = ',' SKIP_HEADER = 1
FIELD_OPTIONALLY_ENCLOSED_BY = '"';"

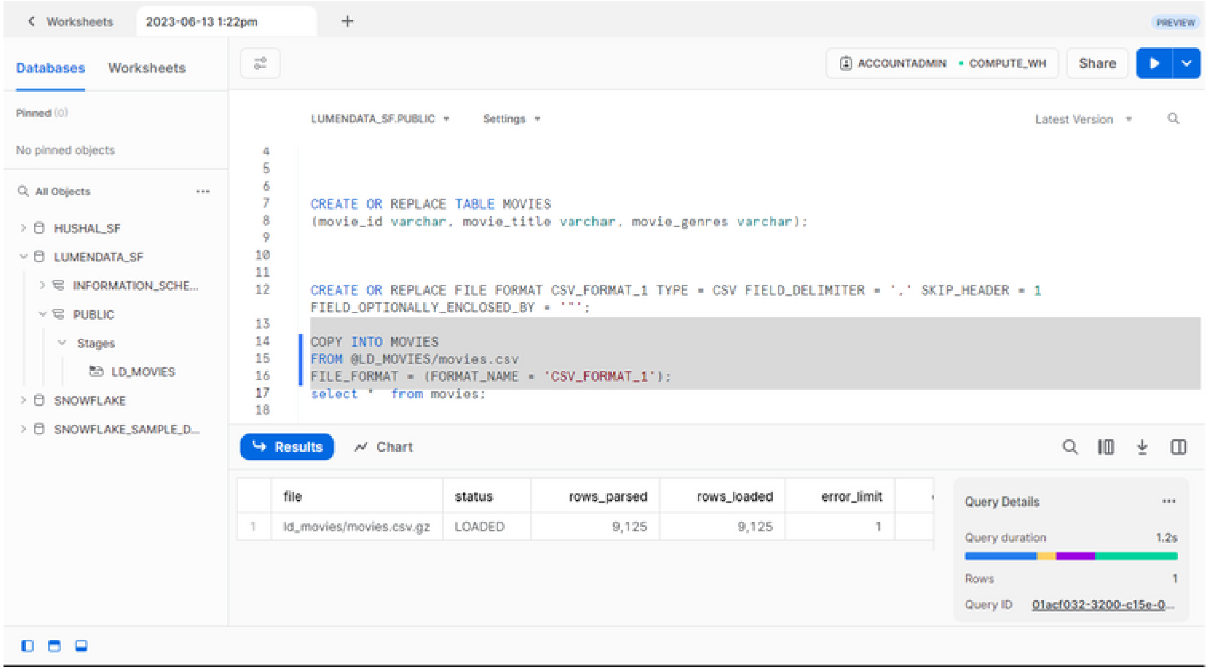


```
CREATE OR REPLACE TABLE MOVIES
(movie_id varchar, movie_title varchar, movie_genres varchar);


CREATE OR REPLACE FILE FORMAT CSV_FORMAT_1 TYPE = CSV FIELD_DELIMITER = ',' SKIP_HEADER = 1
FIELD_OPTIONALLY_ENCLOSED_BY = '"';

COPY INTO MOVIES
FROM @LD_MOVIES/movies.csv
FILE_FORMAT = (FORMAT_NAME = 'CSV_FORMAT_1');
select * from movies;
```

	status
1	File format CSV_FORMAT_1 successfully created.

Query Details
Query duration 64ms
Rows 1
Query ID 01acf031-3200-c15d-0...

11. Copy the data from internal stage into table.
"COPY INTO MOVIES
FROM @LD_MOVIES/movies.csv
FILE_FORMAT = (FORMAT_NAME = 'CSV_FORMAT_1');"

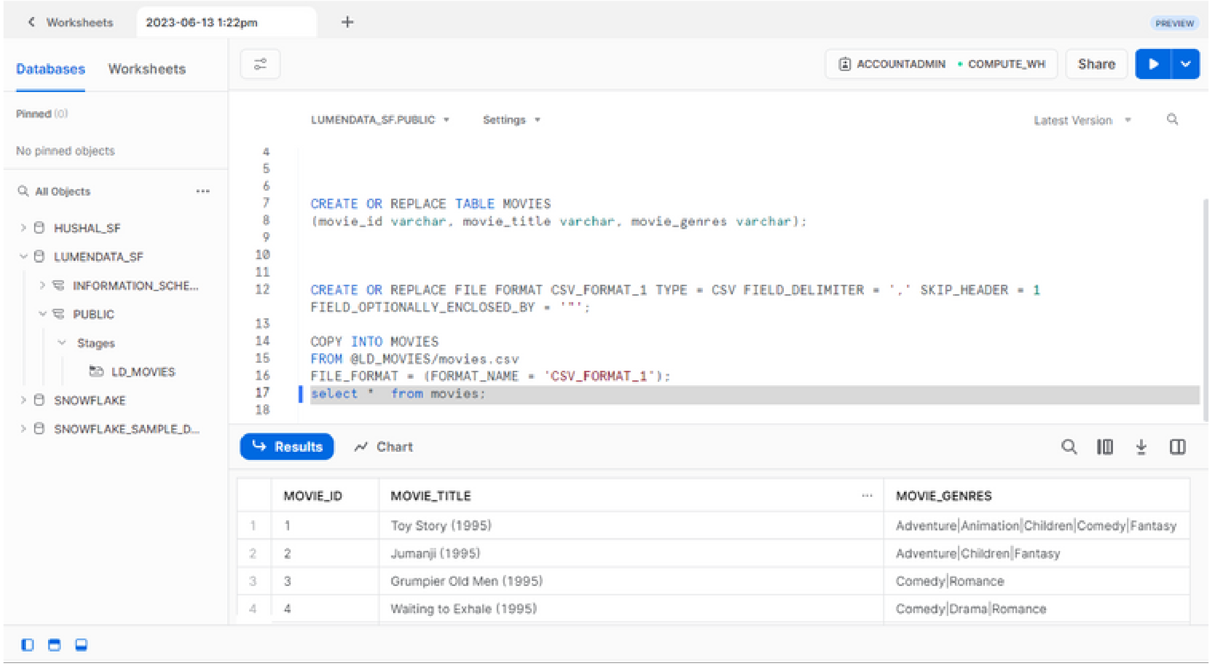


```
CREATE OR REPLACE TABLE MOVIES
(movie_id varchar, movie_title varchar, movie_genres varchar);


CREATE OR REPLACE FILE FORMAT CSV_FORMAT_1 TYPE = CSV FIELD_DELIMITER = ',' SKIP_HEADER = 1
FIELD_OPTIONALLY_ENCLOSED_BY = '"';

COPY INTO MOVIES
FROM @LD_MOVIES/movies.csv
FILE_FORMAT = (FORMAT_NAME = 'CSV_FORMAT_1');
select * from movies;
```

	file	status	rows_parsed	rows_loaded	error_limit
1	ld_movies/movies.csv.gz	LOADED	9,125	9,125	1

Query Details
Query duration 1.2s
Rows 1
Query ID 01acf032-3200-c15e-0...

12. Verify if the data is loaded correctly. “Select * from movies;”



```
CREATE OR REPLACE TABLE MOVIES
(movie_id varchar, movie_title varchar, movie_genres varchar);



CREATE OR REPLACE FILE FORMAT CSV_FORMAT_1 TYPE = CSV FIELD_DELIMITER = ',' SKIP_HEADER = 1
FIELD_OPTIONALLY_ENCLOSED_BY = '"';

COPY INTO MOVIES
FROM @LD_MOVIES/movies.csv
FILE_FORMAT = (FORMAT_NAME = 'CSV_FORMAT_1');
select *  from movies;
```

	MOVIE_ID	MOVIE_TITLE	MOVIE_GENRES
1	1	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy
2	2	Jumanji (1995)	Adventure|Children|Fantasy
3	3	Grumpier Old Men (1995)	Comedy|Romance
4	4	Waiting to Exhale (1995)	Comedy|Drama|Romance

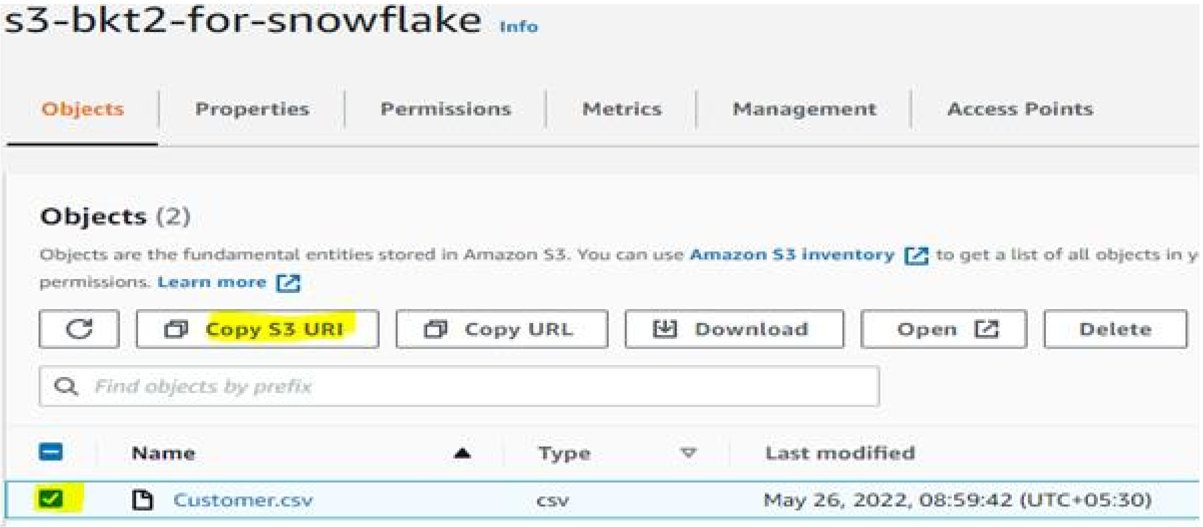
EXTERNAL STAGE:

When stages are located outside of Snowflake, they are External Stages.
Ex: The files are with external cloud vendors such as AWS S3, Azure BLOB and GCP storage. Unlike internal stages, loading and unloading the data can be directly done using COPY command.

To load files from Amazon S3 into a Snowflake table, you can use External Snowflake stages. All you need to do is create an Amazon S3 bucket, upload files on S3, and use S3 Keys to generate external Snowflake stages for the same.

Requirement: AWS Amazon Account.

/* Creating a stage object by pointing to AWS S3 bucket
Here URL -> Go to S3 bucket and select the file, select Copy S3 URI */



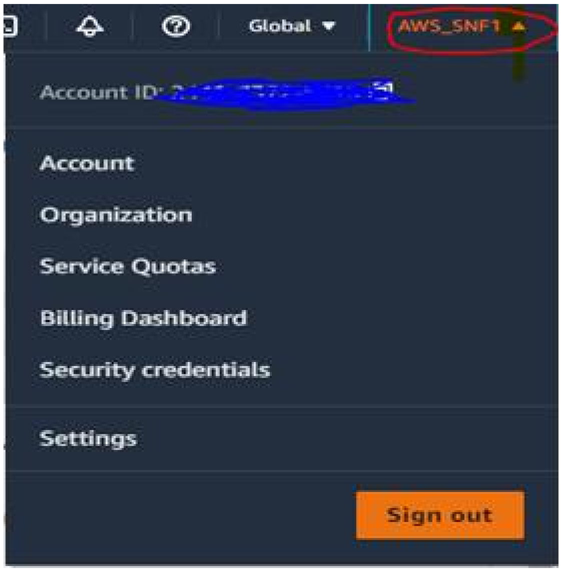
Objects (2)

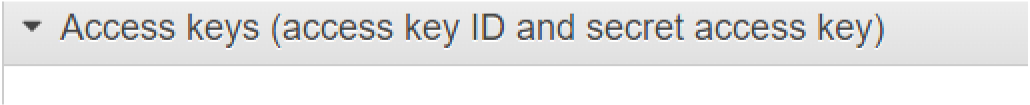
Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 inventory to get a list of all objects in y
permissions. Learn more

☑	Name	Type	Last modified
☑	Customer.csv	csv	May 26, 2022, 08:59:42 (UTC+05:30)

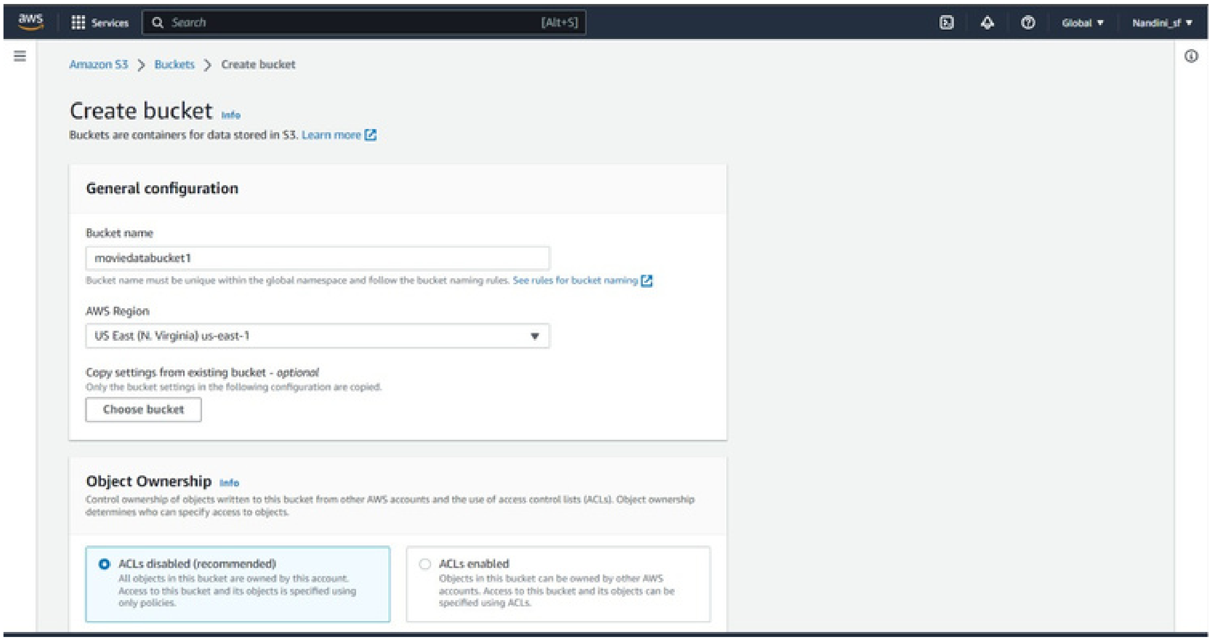
AWS_KEY_ID and AWS_SECRET_KEY we have to get using below steps

Go to AWS account-> Security Credentials -> Go to Access Keys -> Click on Create New Access Key, this will download rootkey.csv file which contains the details.


Account ID: [redacted]


▾ Access keys (access key ID and secret access key)

Create a bucket in AWS to store the file:



Amazon S3 > Buckets > Create bucket

Create bucket Info

Buckets are containers for data stored in S3. Learn more

General configuration

Bucket name: moviedatabucket1

Bucket name must be unique within the global namespace and follow the bucket naming rules. See rules for bucket naming

AWS Region: US East (N. Virginia) us-east-1 ▾

Copy settings from existing bucket - *optional*

Only the bucket settings in the following configuration are copied.

Choose bucket

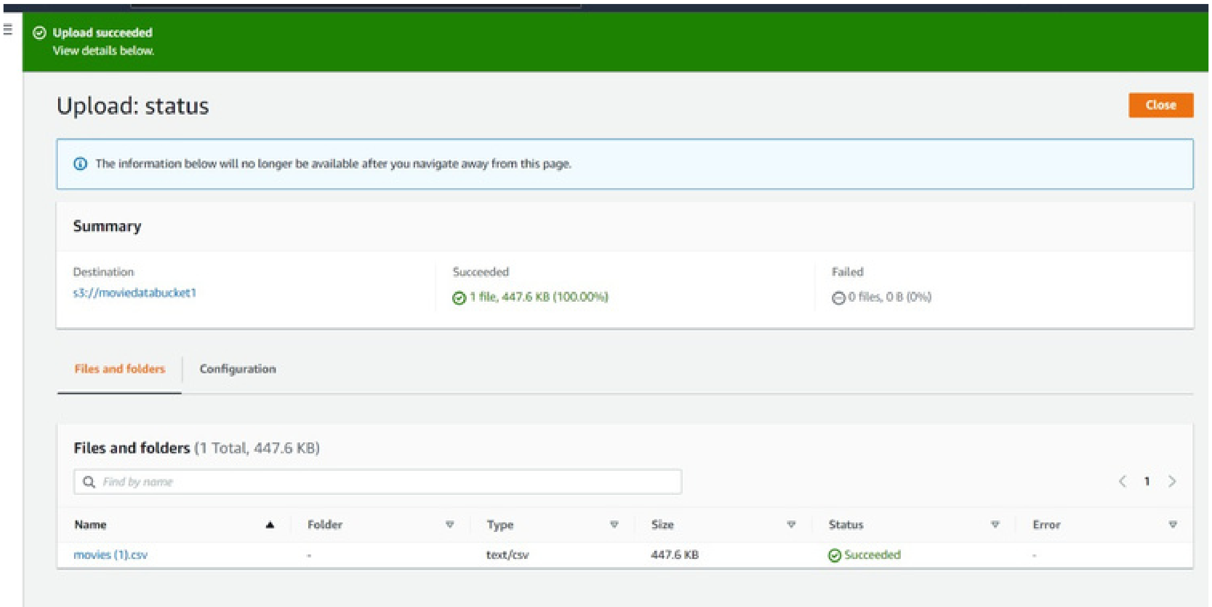
Object Ownership Info

Control ownership of objects written to this bucket from other AWS accounts and the use of access control lists (ACLs). Object ownership determines who can specify access to objects.

● ACLs disabled (recommended)
All objects in this bucket are owned by this account. Access to this bucket and its objects is specified using only policies.

○ ACLs enabled
Objects in this bucket can be owned by other AWS accounts. Access to this bucket and its objects can be specified using ACLs.

Upload a file into the S3 bucket that is created.



Upload succeeded
View details below.

Upload: status

The information below will no longer be available after you navigate away from this page.

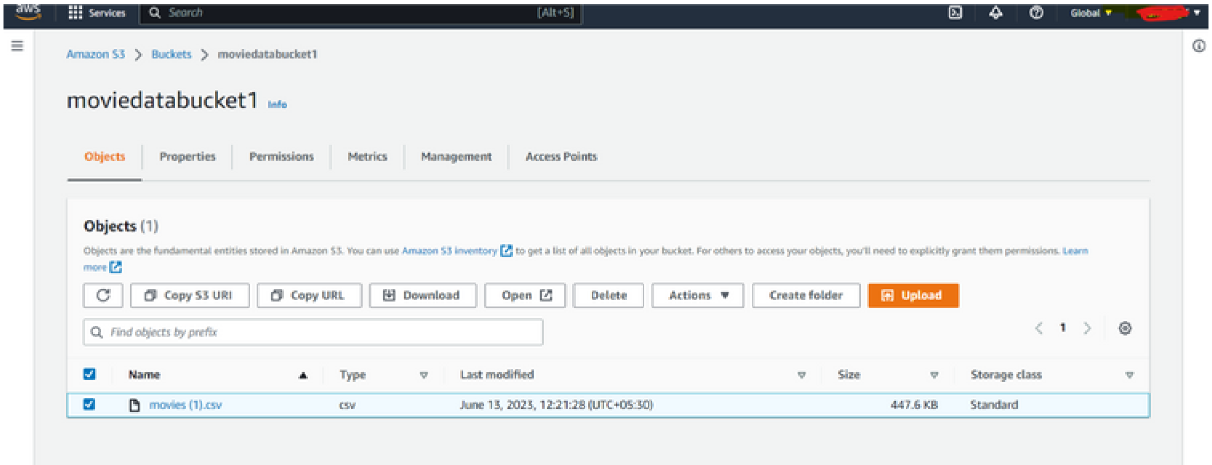
Summary

Destination	Succeeded	Failed
s3://moviedatabucket1	✓ 1 file, 447.6 KB (100.00%)	0 files, 0 B (0%)

Files and folders (1 Total, 447.6 KB)

Name	Folder	Type	Size	Status	Error
movies (1).csv	-	text/csv	447.6 KB	✓ Succeeded	-

Now, the File is available in AWS.



[redacted]

Amazon S3 > Buckets > moviedatabucket1

moviedatabucket1 Info

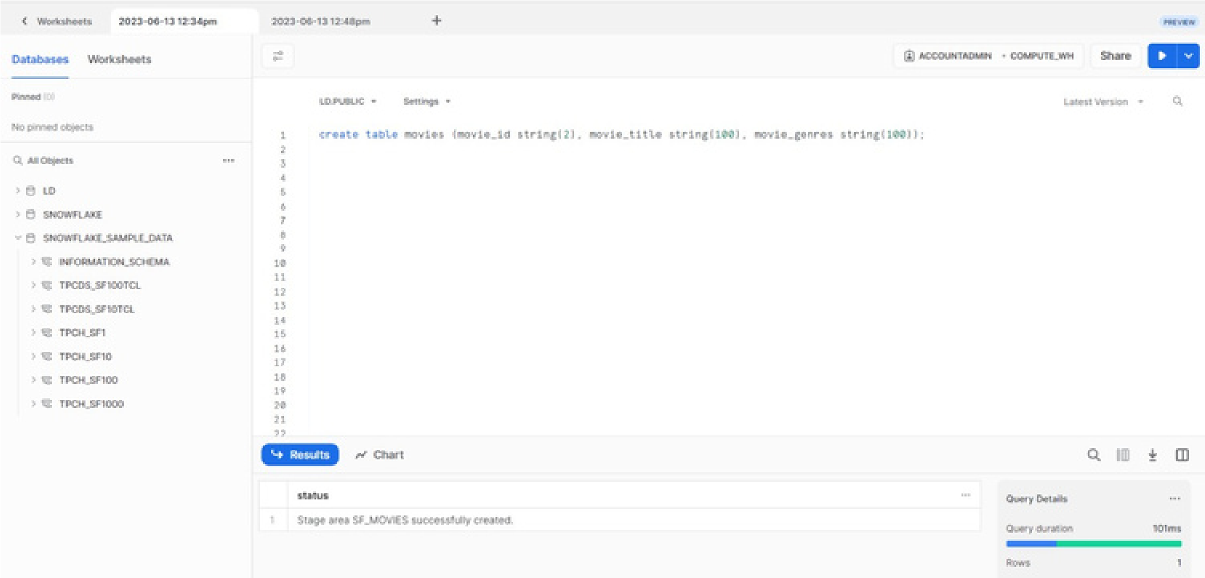
Objects (1)

Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 inventory to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. Learn more

☑	Name	Type	Last modified	Size	Storage class
☑	movies (1).csv	csv	June 13, 2023, 12:21:28 (UTC+05:30)	447.6 KB	Standard

STAGING THE AWS FILE INTO SNOWFLAKE:

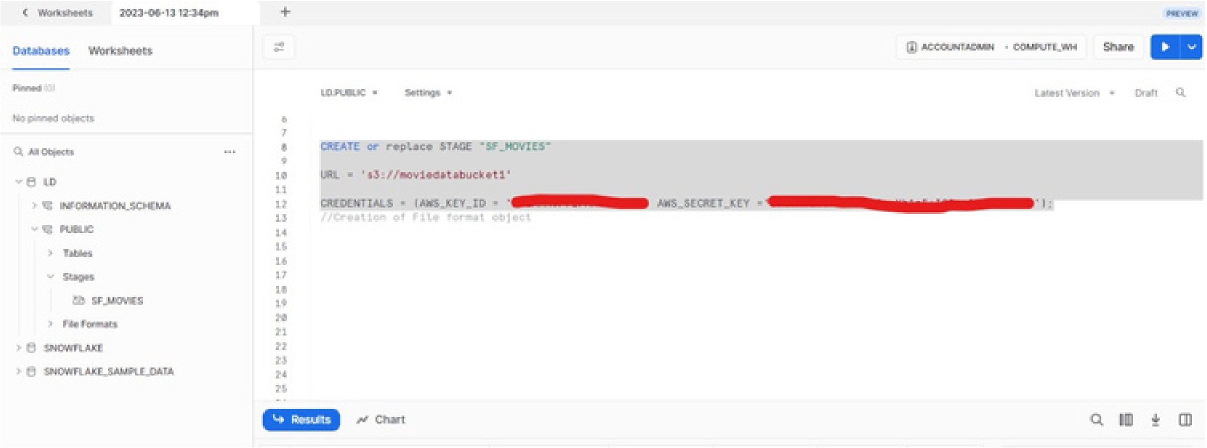
Create a table for Movies: create table movies (movie_id string(100), movie_title string(100), movie_genres string(100));



```
create table movies (movie_id string(2), movie_title string(100), movie_genres string(100));
```

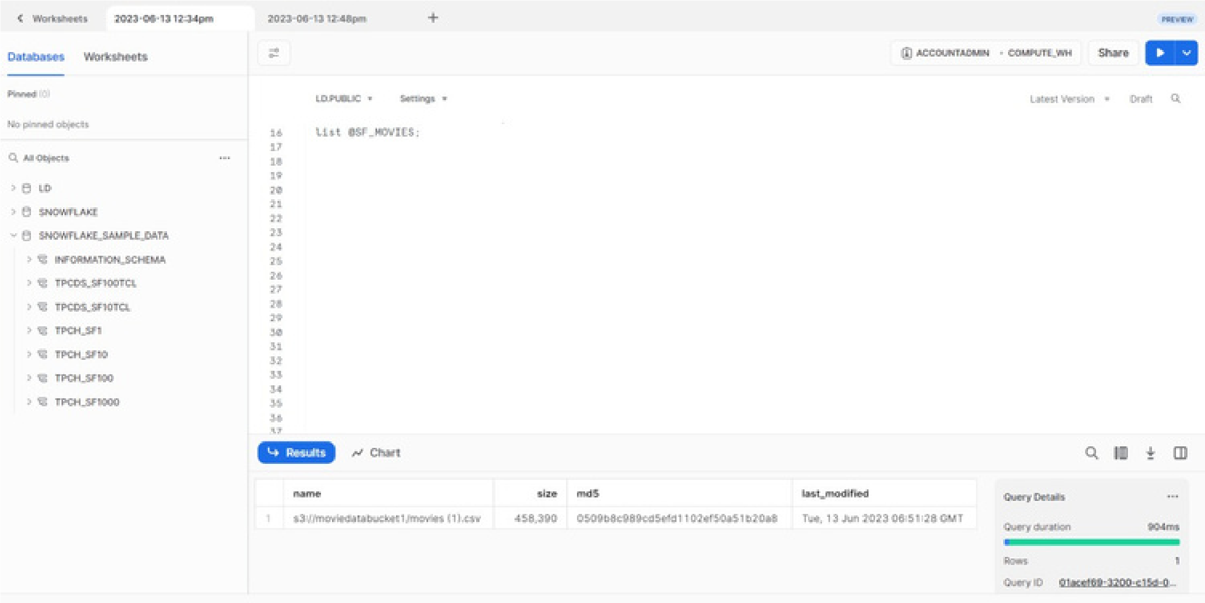
```
   status
1  Stage area SF_MOVIES successfully created.
```

Create a Stage:
CREATE or replace STAGE "SF_MOVIES"
URL = 's3://moviedatabucket1'
CREDENTIALS = (AWS_KEY_ID = '--------------' AWS_SECRET_KEY ='------ ----------------------');



```
CREATE or replace STAGE "SF_MOVIES"

URL = 's3://moviedatabucket1'

CREDENTIALS = (AWS_KEY_ID = '[redacted] AWS_SECRET_KEY ='[redacted]');
//Creation of File format object
```

As you have created the stage, you now have access to the entire bucket on AWS: list @SF_MOVIES;



```
list @SF_MOVIES;
```

	name	size	md5	last_modified
1	s3://moviedatabucket1/movies (1).csv	458,390	0509b8c989cd5efd1102ef50a51b20a8	Tue, 13 Jun 2023 06:51:28 GMT

Query Details

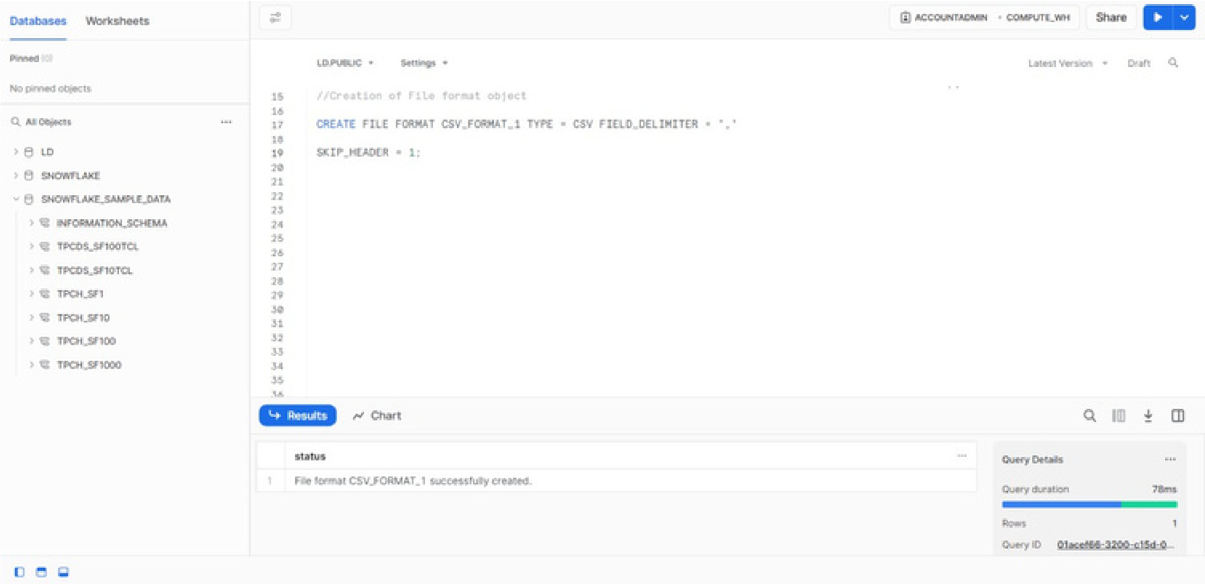
Query duration 904ms

Rows 1

Query ID 01acef69-3200-c15d-0...

CREATION OF FILE FORMAT OBJECT

CREATE FILE FORMAT CSV_FORMAT_1 TYPE = CSV FIELD_DELIMITER = ','
SKIP_HEADER = 1 FIELD_OPTIONALLY_ENCLOSED_BY = ' " ';



```
//Creation of File format object

CREATE FILE FORMAT CSV_FORMAT_1 TYPE = CSV FIELD_DELIMITER = ','

SKIP_HEADER = 1;
```

	status
1	File format CSV_FORMAT_1 successfully created.

Query Details

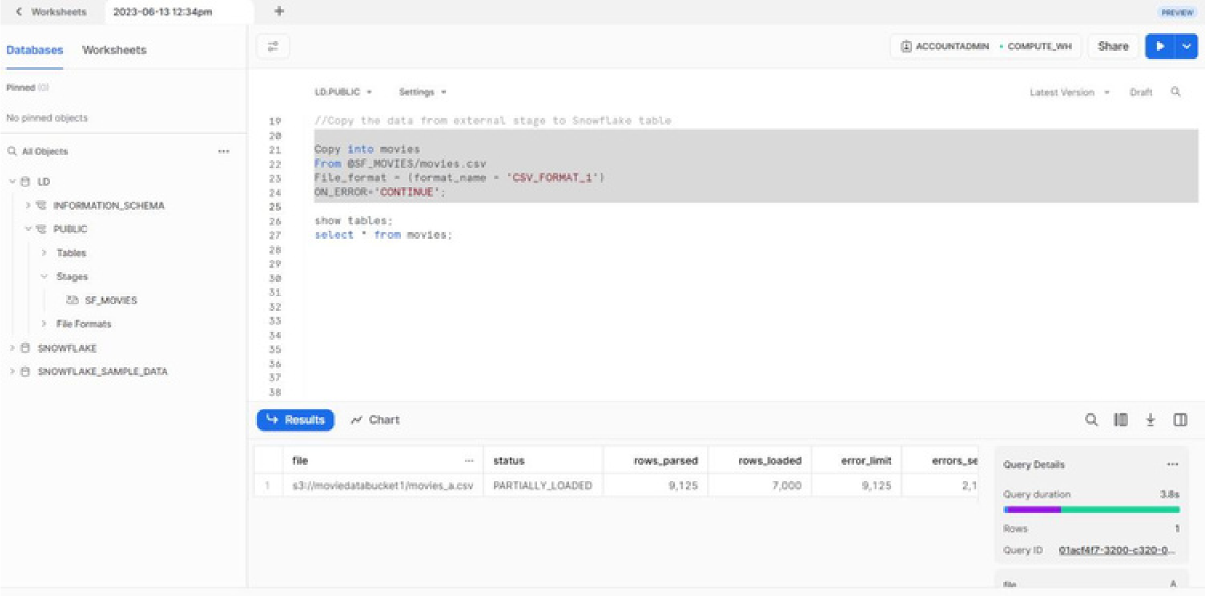
Query duration 78ms

Rows 1

Query ID 01acef66-3200-c15d-0...

COPY THE DATA FROM EXTERNAL STAGE TO SNOWFLAKE TABLE.

Copy into movies
From @SF_MOVIES/movies.csv
File_format = (format_name = 'CSV_FORMAT_1');

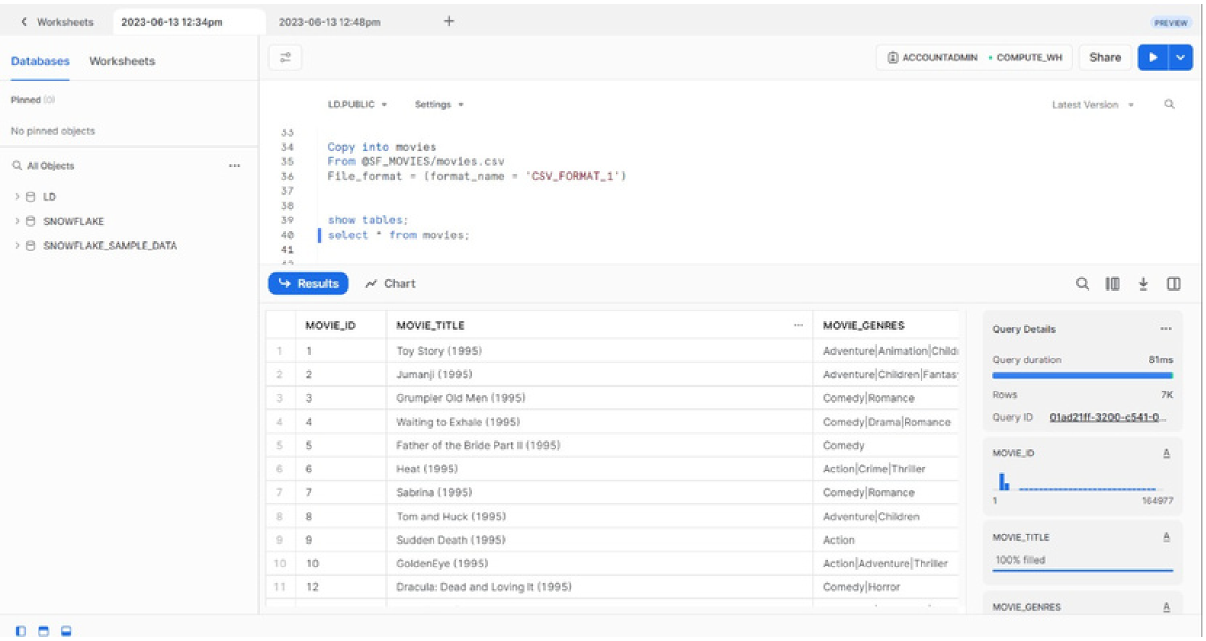


```
//Copy the data from external stage to Snowflake table

Copy into movies
From @SF_MOVIES/movies.csv
File_format = (format_name = 'CSV_FORMAT_1')
ON_ERROR='CONTINUE';

show tables;
select * from movies;
```

	file	status	rows_parsed	rows_loaded	error_limit	errors_se
1	s3://moviedatabucket1/movies_a.csv	PARTIALLY_LOADED	9,125	7,000	9,125	2,1

Query Details
Query duration 3.8s
Rows 1
Query ID 01acf4f7-3200-c320-0...

After running the following query “select * from movies;” you can see the loaded data in Snowflake.



```
Copy into movies
From @SF_MOVIES/movies.csv
File_format = (format_name = 'CSV_FORMAT_1')


show tables;
select * from movies;
```

	MOVIE_ID	MOVIE_TITLE	MOVIE_GENRES
1	1	Toy Story (1995)	Adventure|Animation|Child
2	2	Jumanji (1995)	Adventure|Children|Fantas
3	3	Grumpier Old Men (1995)	Comedy|Romance
4	4	Waiting to Exhale (1995)	Comedy|Drama|Romance
5	5	Father of the Bride Part II (1995)	Comedy
6	6	Heat (1995)	Action|Crime|Thriller
7	7	Sabrina (1995)	Comedy|Romance
8	8	Tom and Huck (1995)	Adventure|Children
9	9	Sudden Death (1995)	Action
10	10	GoldenEye (1995)	Action|Adventure|Thriller
11	12	Dracula: Dead and Loving It (1995)	Comedy|Horror

Query Details
Query duration 81ms
Rows 7K
Query ID 01ad21ff-3200-c541-0...
MOVIE_ID
1 164977
MOVIE_TITLE
100% filled
MOVIE_GENRES

The following are some best practices for working with Snowflake Stages:

- If you need to load a file into Numerous Tables, the User Stage is the way to go.
- If you only need to load into one Table, use the Table Stage.
- Numerous Users can access the Internal Named Stages, which can be utilized to Load Multiple Tables. If you want to share files with different users and load them into Multiple Tables, the Internal Named Stage is the way to go.
- For huge files, it is always recommended to use an External Named Stage.

ABOUT LUMENDATA:

LumenData is a leading provider of Enterprise Data Management, Cloud & Analytics solutions. We help businesses navigate their data visualization and analytics anxieties and enable them to accelerate their innovation journeys. Founded in 2008, with locations in multiple countries, LumenData is privileged to serve over 100 leading companies, including KwikTrip, Versant Health, US Food & Drug Administration, US Department of Labor, Cummins Engine, BCG, and others. LumenData is SOC2 certified and has instituted extensive controls to protect client data, including adherence to GDPR and CCPA regulations.

Get in touch to discuss how we can facilitate data-driven transformation for your organization.

MEET OUR AUTHORS

Sai Bharadwaja Reddy
Consultant

Hushal Parmar
Associate Consultant

Nandini K
Software Consultant