SNOWFLAKE

THE CLOUD DATA

PLATFORM

Introduction

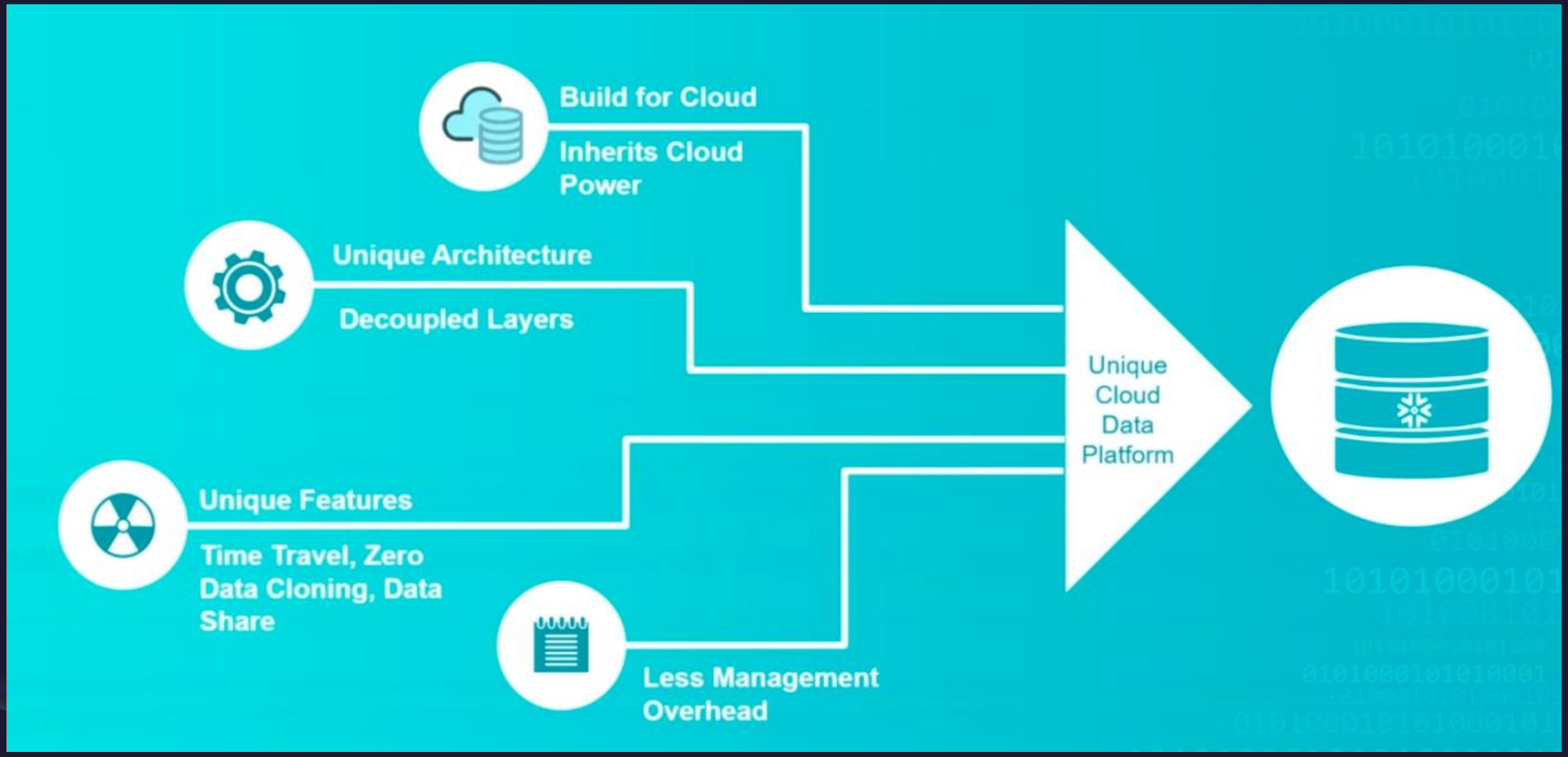
Introduction

- We are going to learn basic Snowflake concepts and fundamentals
- Doing hands-on labs, no prior knowledge of Snowflake is required
- Basic SQL and database knowledge is required
- Preferred to have knowledge on Cloud
- Take your time for the hands-on

What is Snowflake ?

- Snowflake is an analytic data warehouse provided as Software-as- Services(SaaS). Snowflake provides data warehouse that is faster, easierto use and more flexible that other traditional data warehouses.
- Snowflake data warehouse is not built on existing databases or not on big data software platform as Hadoop.
- The snowflake data warehouse uses a new SQL database engine with unique architecture designed for the cloud.

Snowflake Key Differentiator



Key Concept and Architecture

Data Warehouse as Cloud Service:

Snowflake data warehouse is true SaaS offering :

- There is no hardware (virtual or physical) for you to select, install, configure and manage.
- There is no software for you install, configure and manage.
- Ongoing maintenance, management and tuning is handled by snowflake.

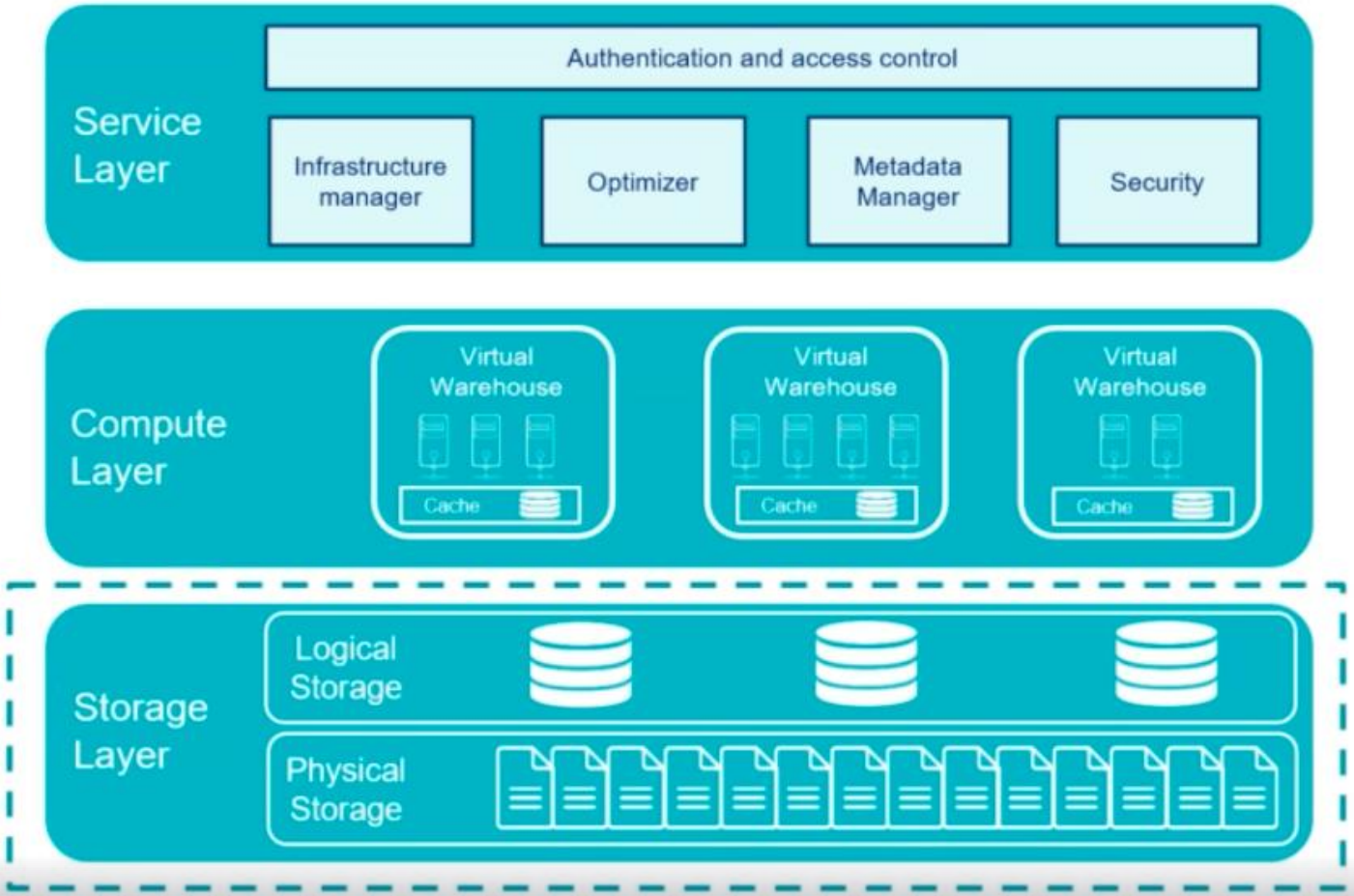
Snowflake completely runs on cloud infrastructure. All the component of the snowflake service runs on public cloud infrastructure.

Snowflake uses virtual compute instance for its compute need and storage service for storage of data.
Snowflake can not be run on private cloud infrastructure(on primises)

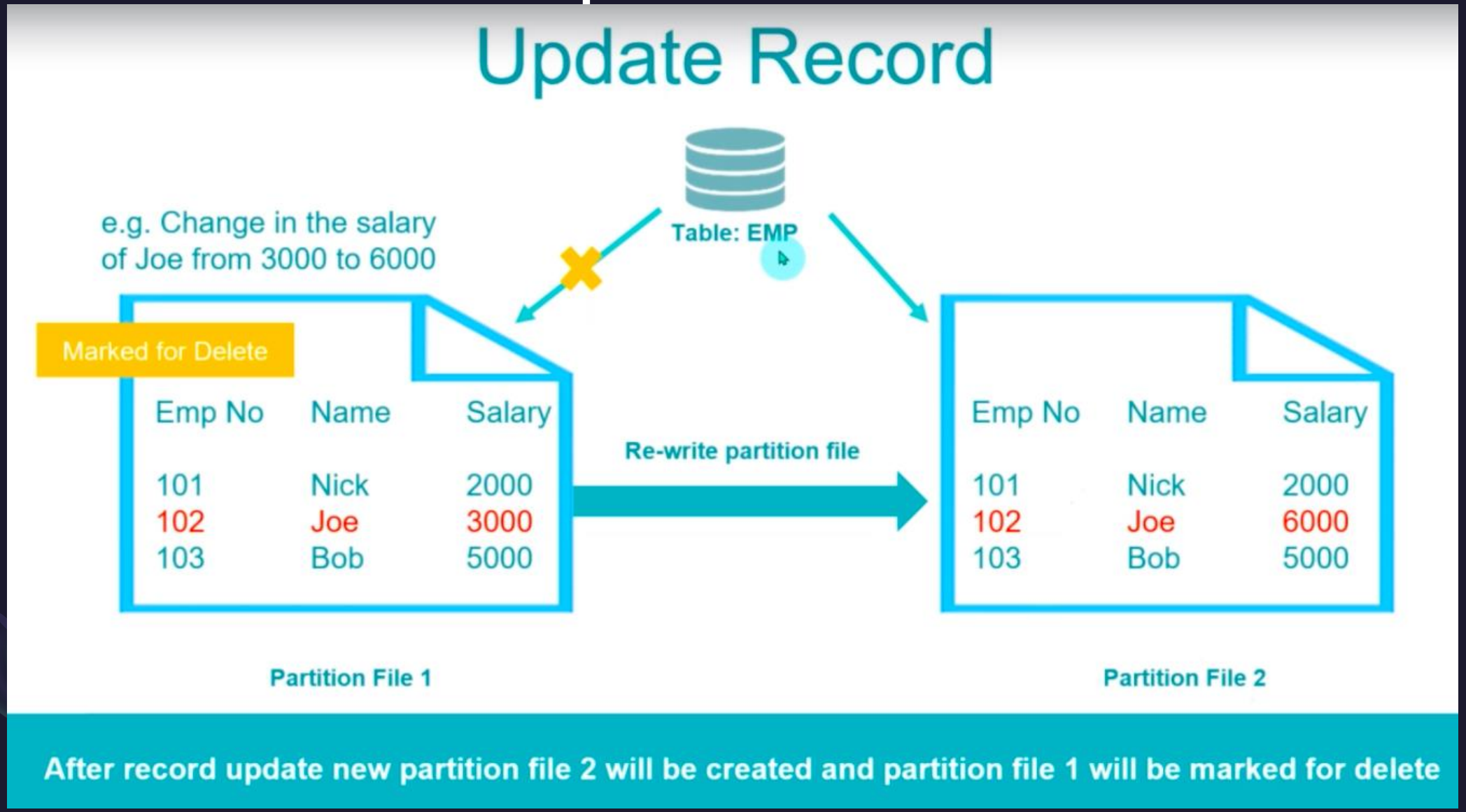
Snowflake Architecture

Storage Layer

- Storage is independent from compute
- Virtually unlimited storage capacity
- Data physically stored as micro-partition files
- Each micro-partitioned file is of size 16 MB
- Every file gets replicated 3 times to ensure high availability
- File format is propriety to Snowflake
- Files are immutable
- Table definition are logical and stored in its metadata layer



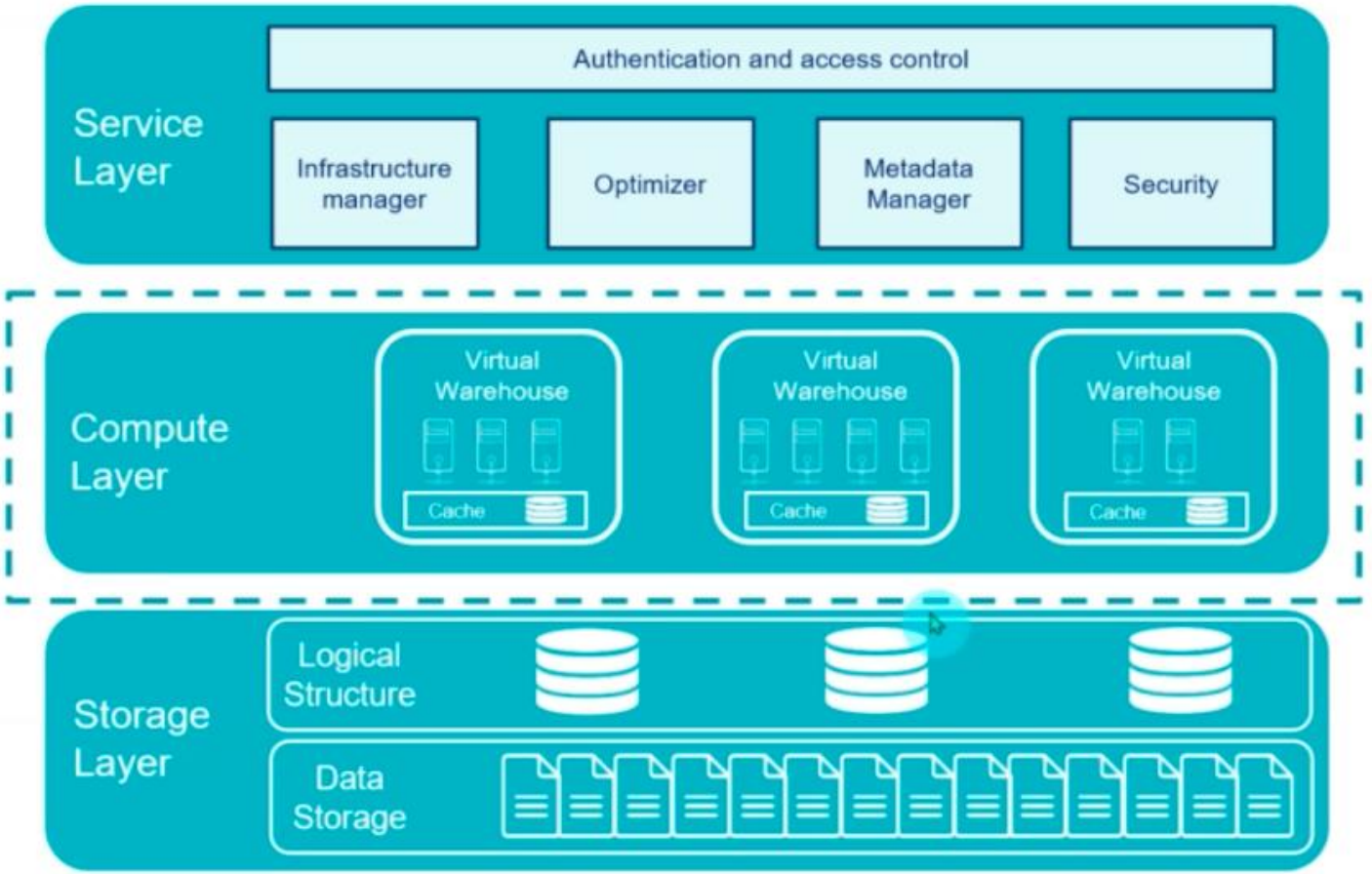
Update Record



Compute Layer

Compute (a.k.a Warehouse) Layer

- Only way to access data from storage layer
- Horizontally and vertically scalable
  - Change the size of the warehouse at runtime
  - Add more nodes at the runtime
- Warehouse size vary from X-small (1 server/ cluster) to 4X-Large (128 server/cluster)
- Every warehouse works independently and performance of one warehouse will not impact another warehouse
- Every warehouse have its own small storage, which it use to cache query data for better performance
- Compute can be auto-suspended when idle

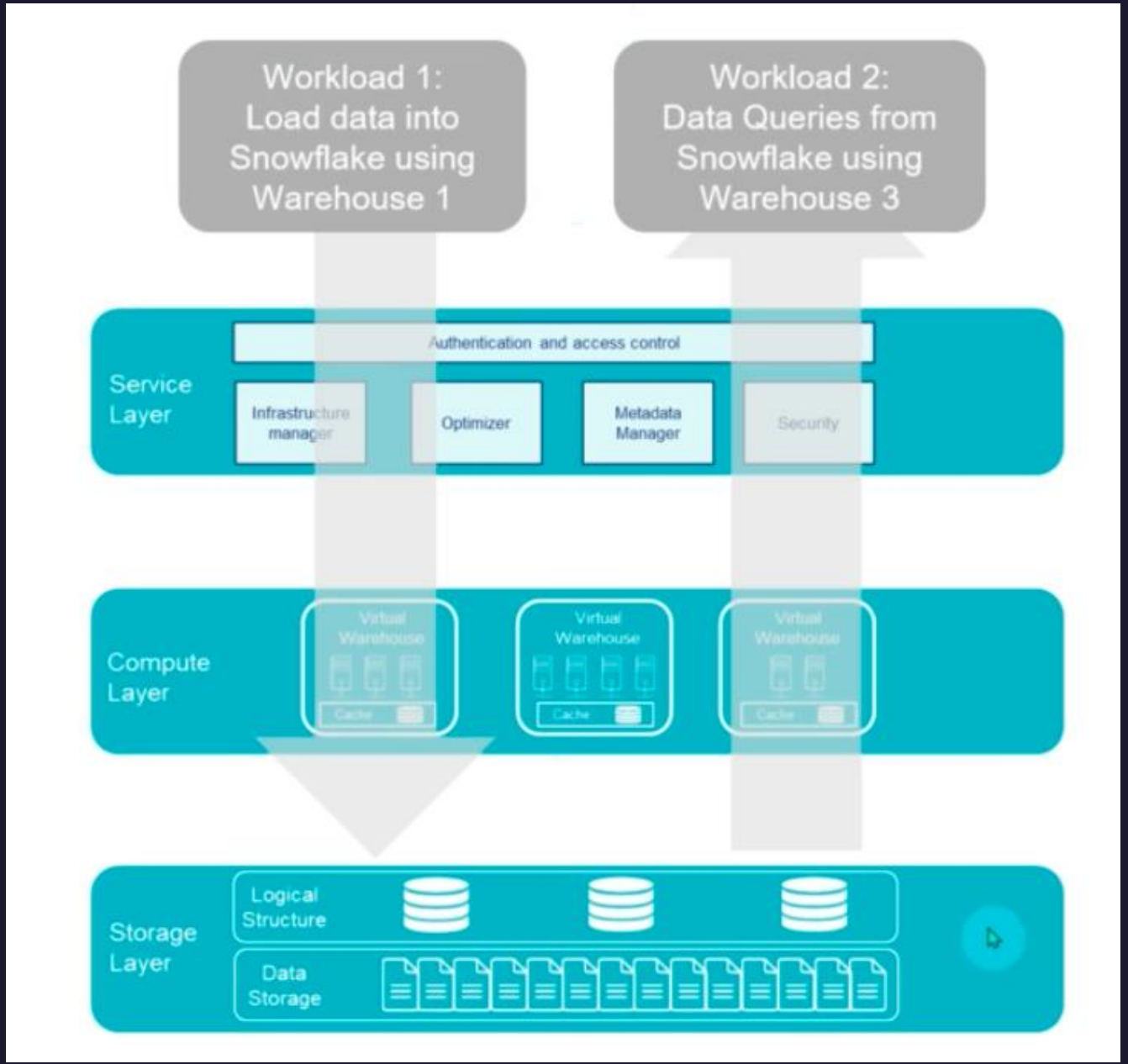


Warehouse Size & Pricing

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

Separate Warehouse for Different Workloads



Service Layer

Service Layer

- Brain of the Snowflake
- All the interaction with Snowflake established via this layer
- Stores query output in a result cache
- Snowflake doesn't expose its service layer and users can't get insight and access their metadata
- During data load Service layer keeps track of which data stored in which partition file
- Maintain transaction consistency across Warehouse

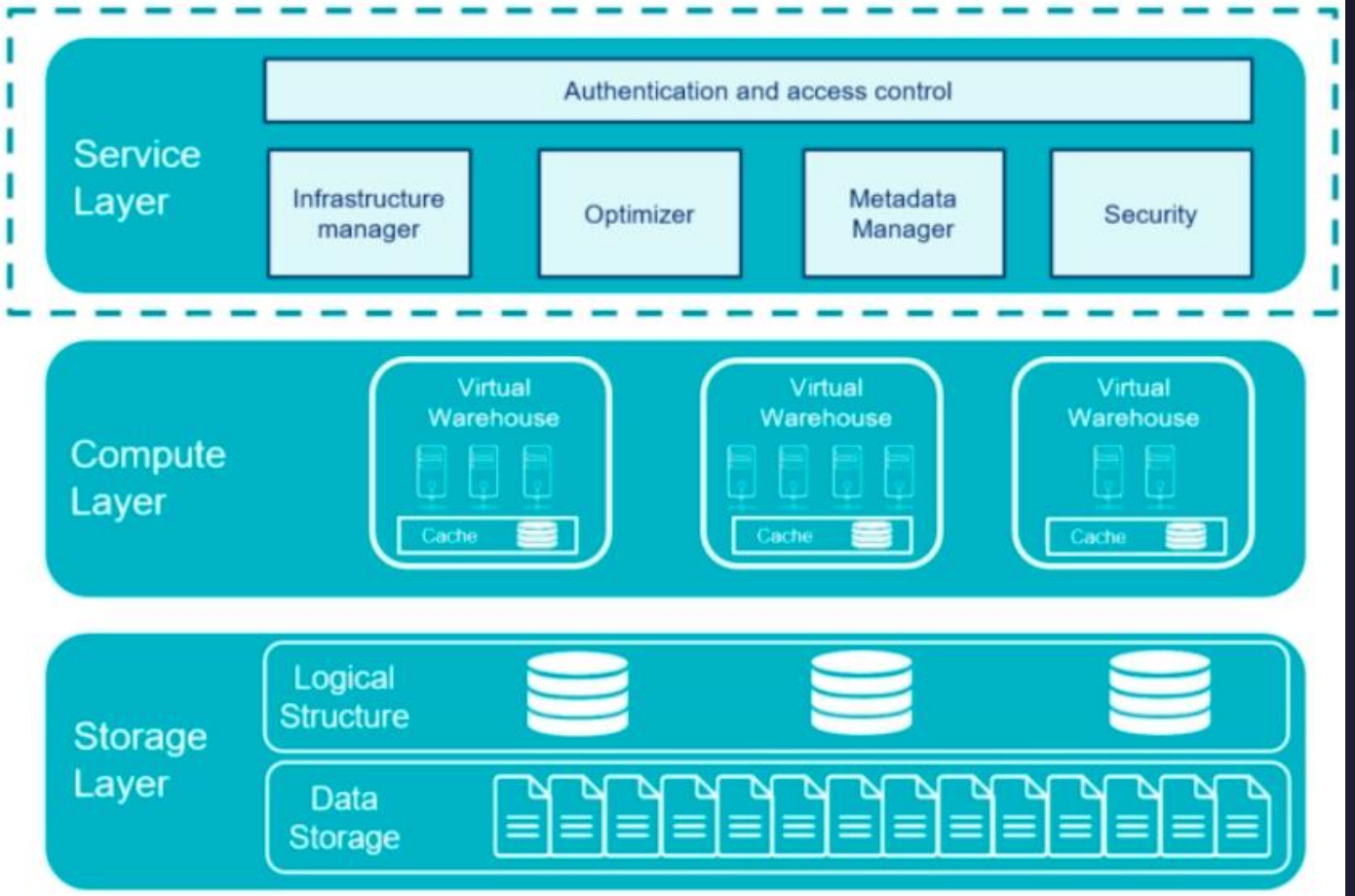
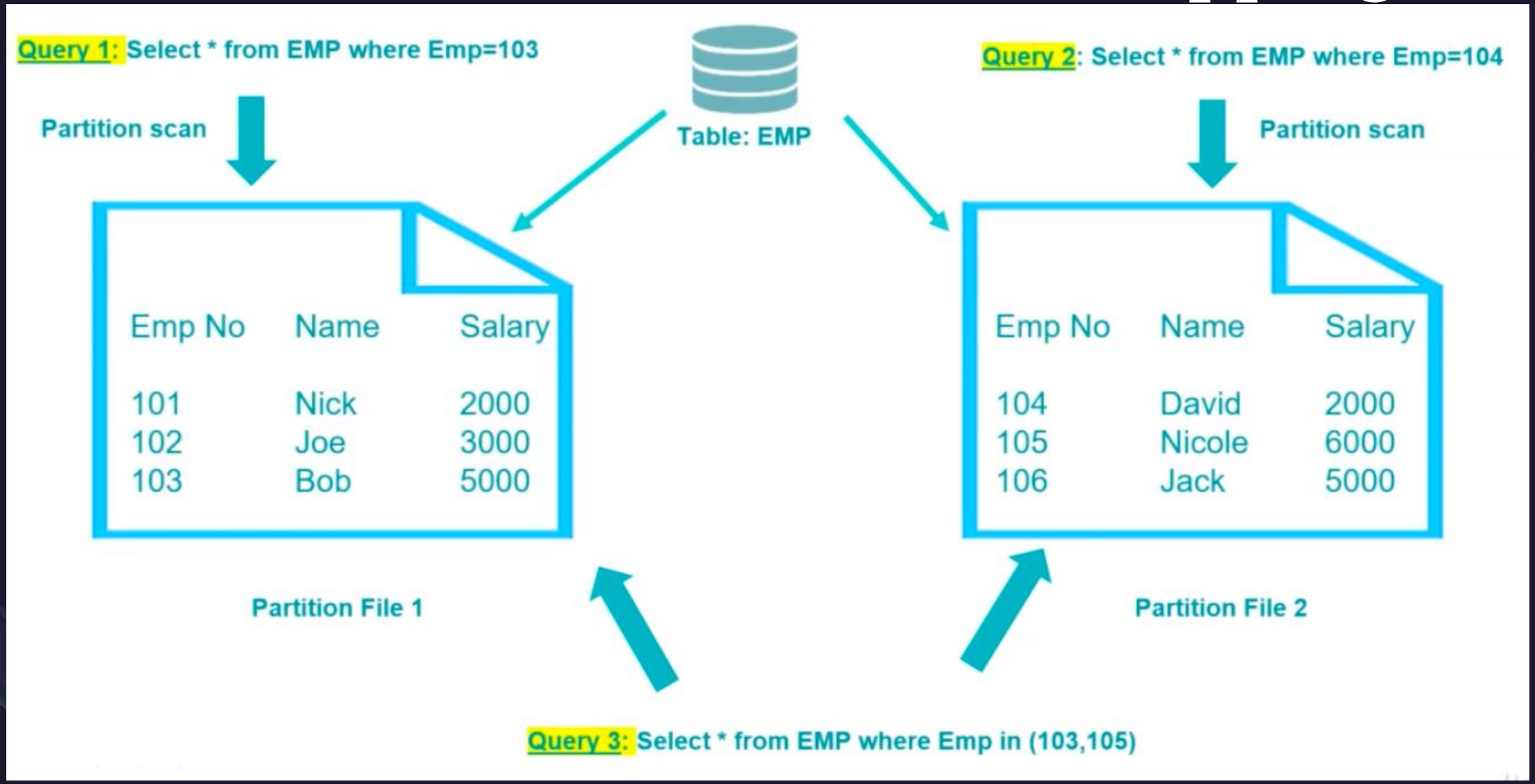


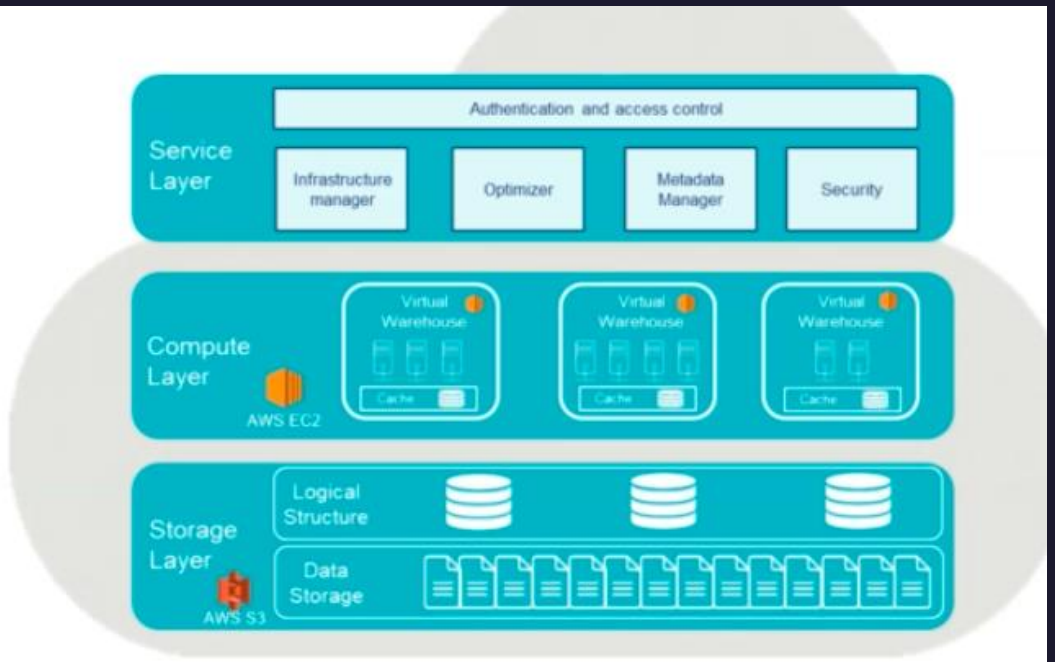
Table and Partition File Mapping



Why Only For Cloud(SaaS) ?

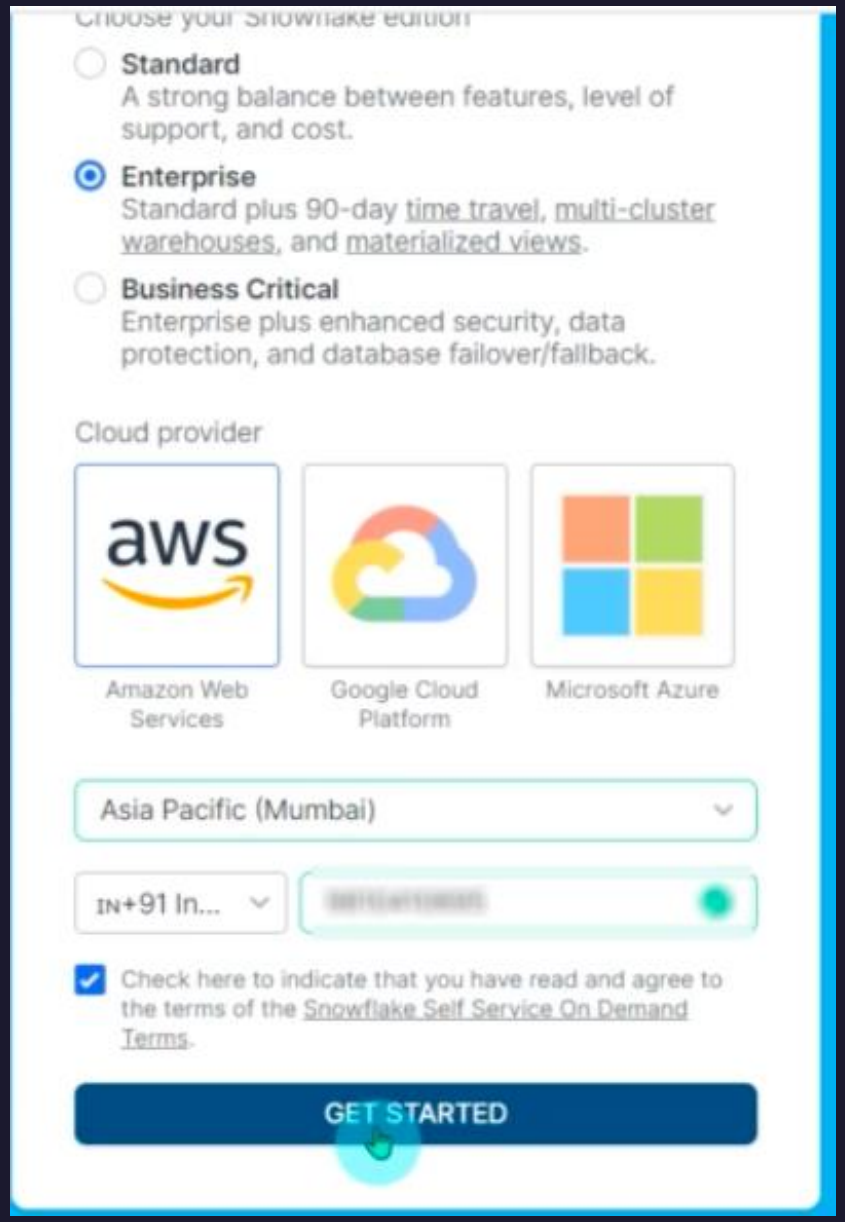
Considering AWS based Snowflake account

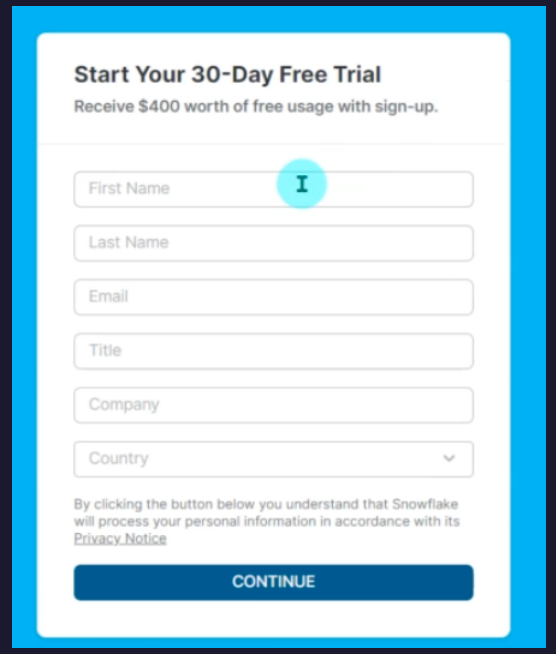
- Storage Layer hosted on AWS S3 and offers unlimited storage and 3 copies of the data by default
- Compute/Warehouse leverage AWS EC2 where local SSD will act as the cache
- Inherits Cloud elasticity, scalability and high availability
- Snowflake has a large pool of EC2 instances, which makes it very fast to scale up

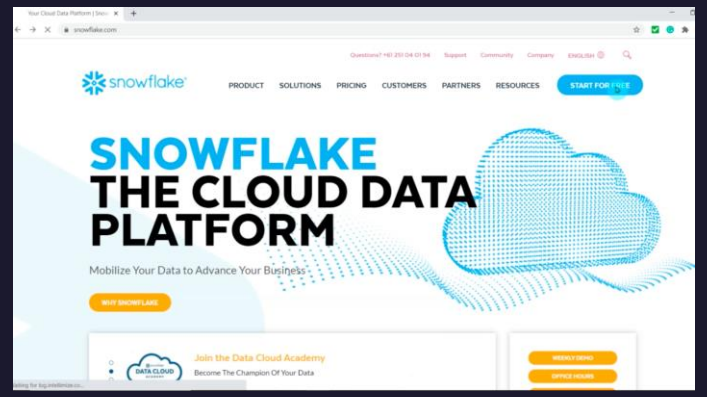


This is the reason Snowflake is not available on-premise

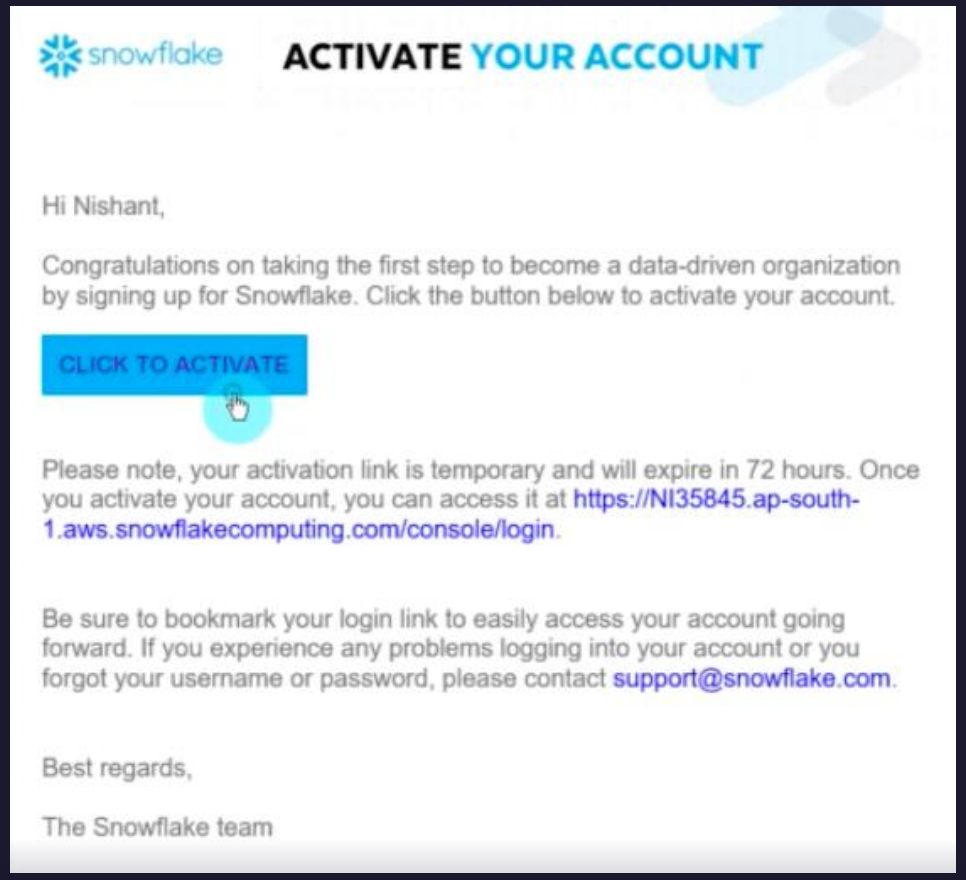
CREATE YOUR SNOWFLAKE ACCOUNT : SIGNUP

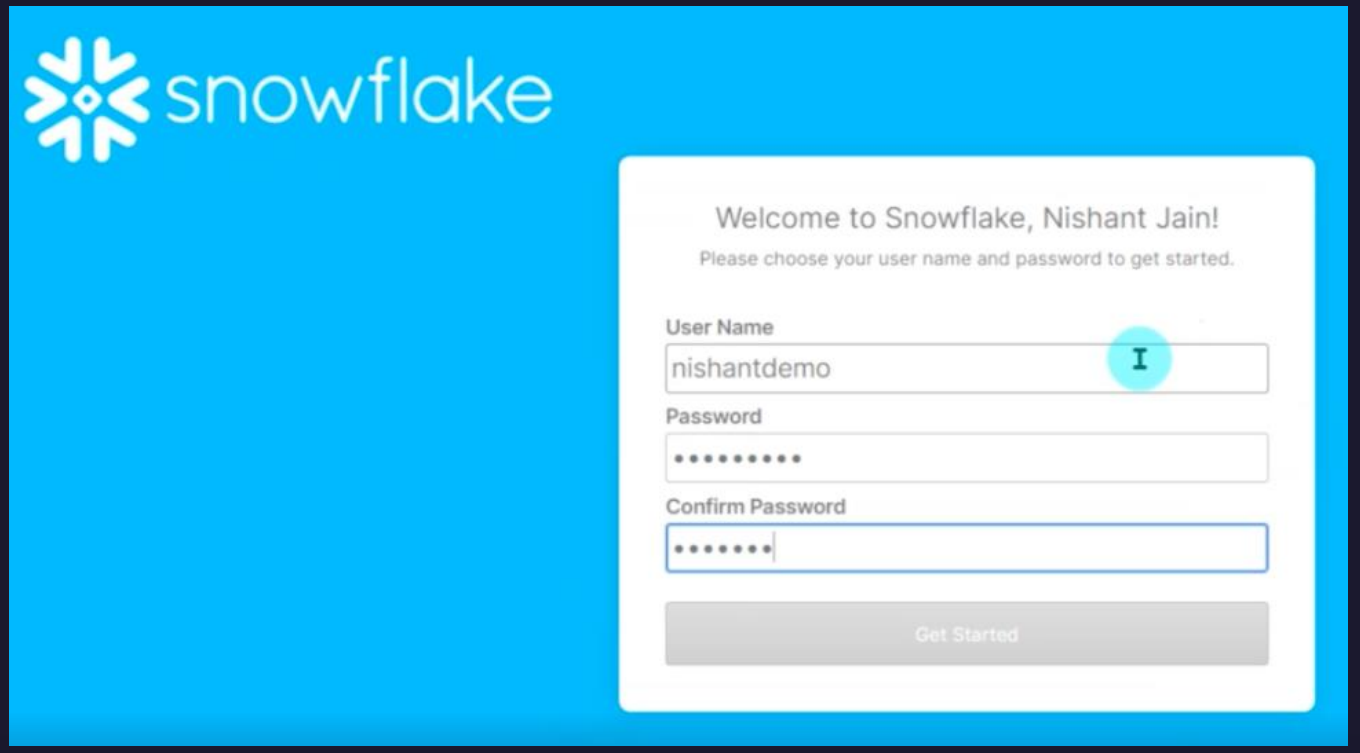




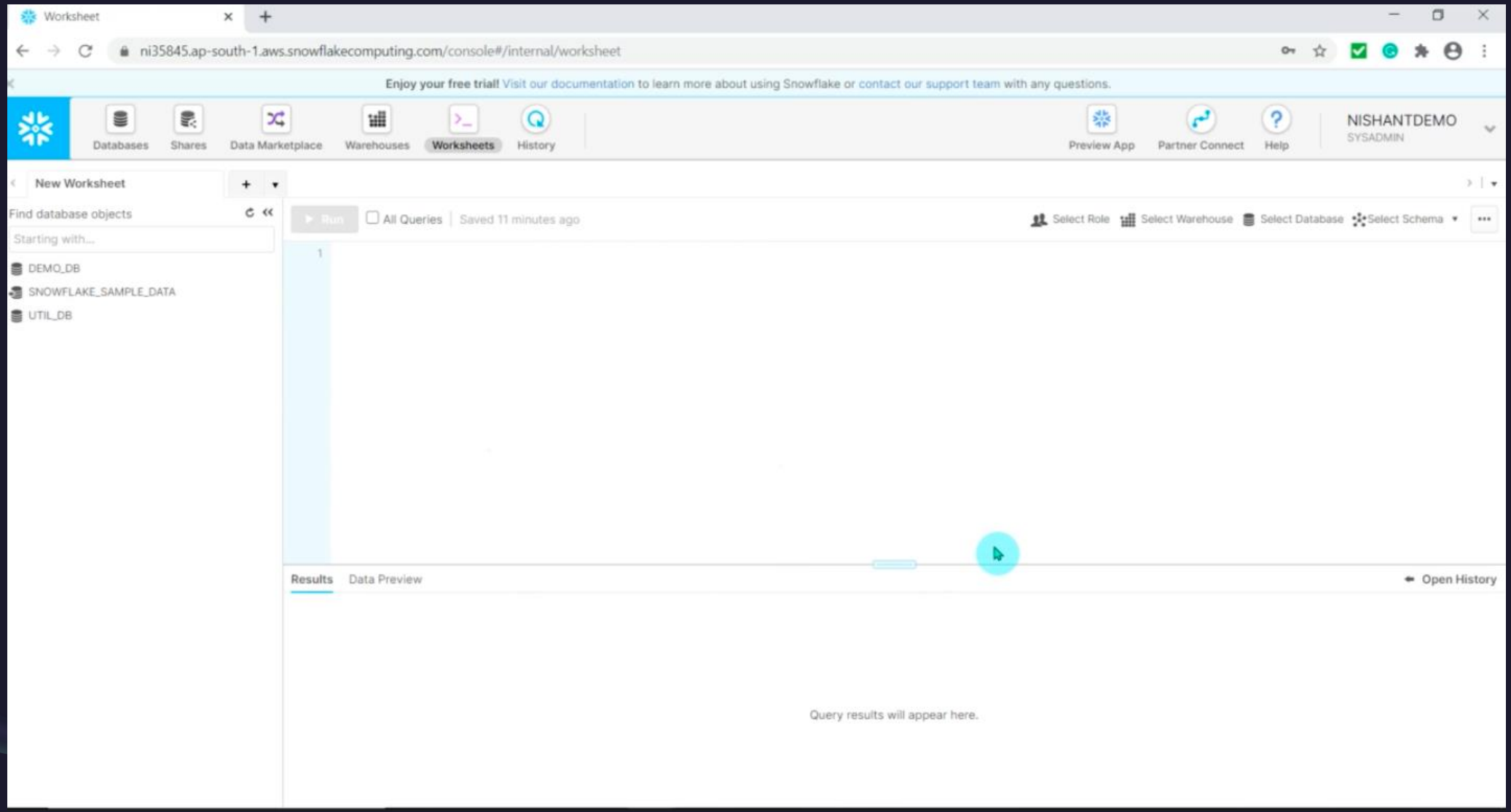


GET STARTED

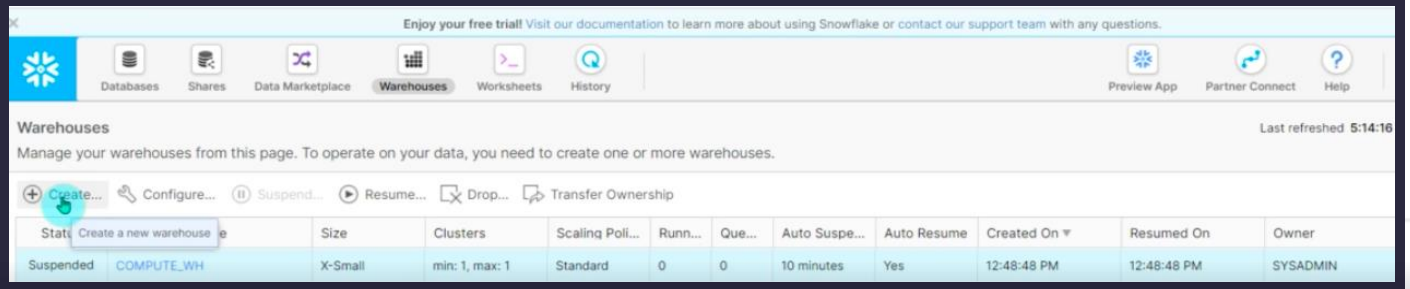

snowflake
ACTIVATE YOUR ACCOUNT
Hi Nishant,
Congratulations on taking the first step to become a data-driven organization by signing up for Snowflake. Click the button below to activate your account.
CLICK TO ACTIVATE
Please note, your activation link is temporary and will expire in 72 hours. Once you activate your account, you can access it at https://NI35845.ap-south-1.aws.snowflakecomputing.com/console/login.
Be sure to bookmark your login link to easily access your account going forward. If you experience any problems logging into your account or you forgot your username or password, please contact support@snowflake.com.
Best regards,
The Snowflake team

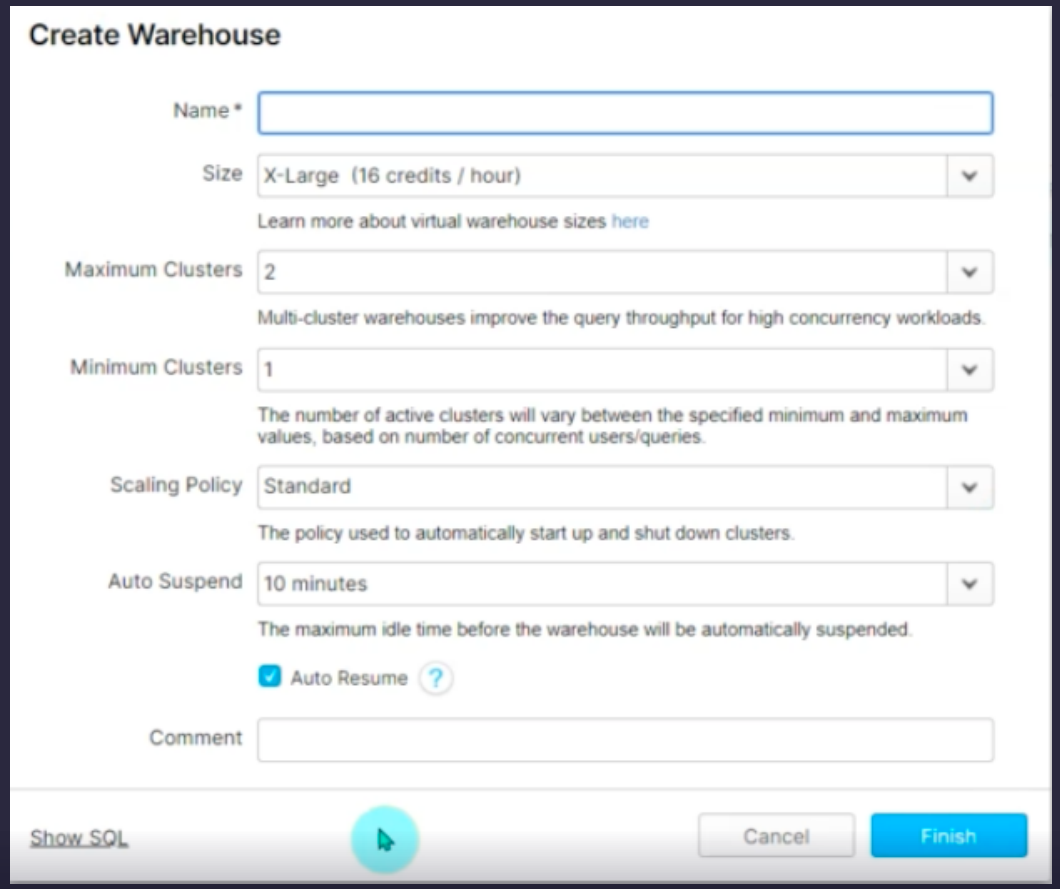

snowflake
Welcome to Snowflake, Nishant Jain!
Please choose your user name and password to get started.
User Name
nishantdemo
Password
Confirm Password
Get Started

SNOWFLAKE CONSONE

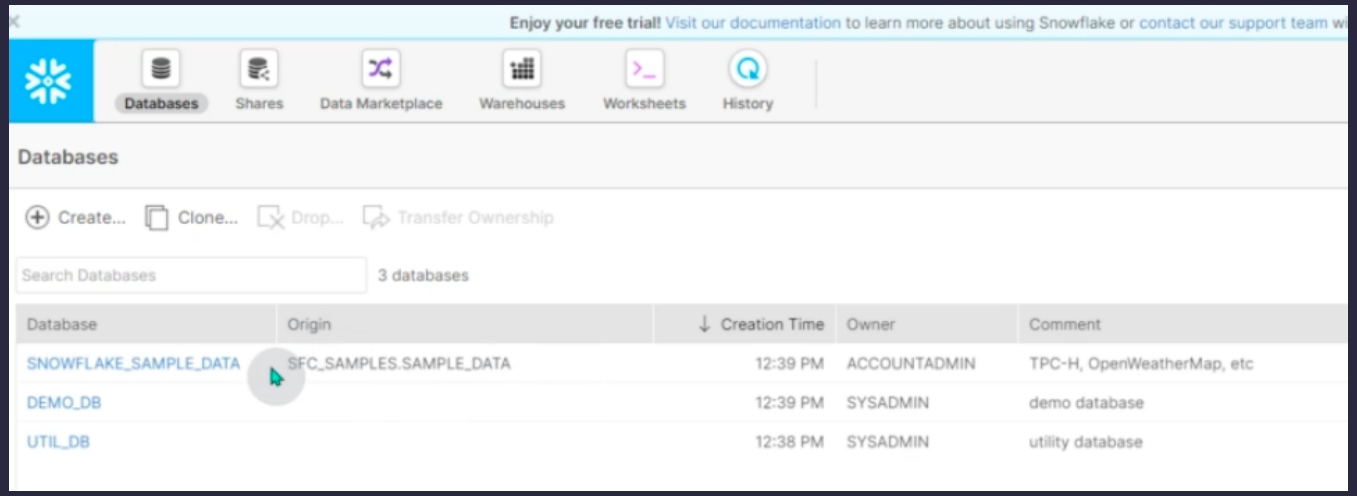


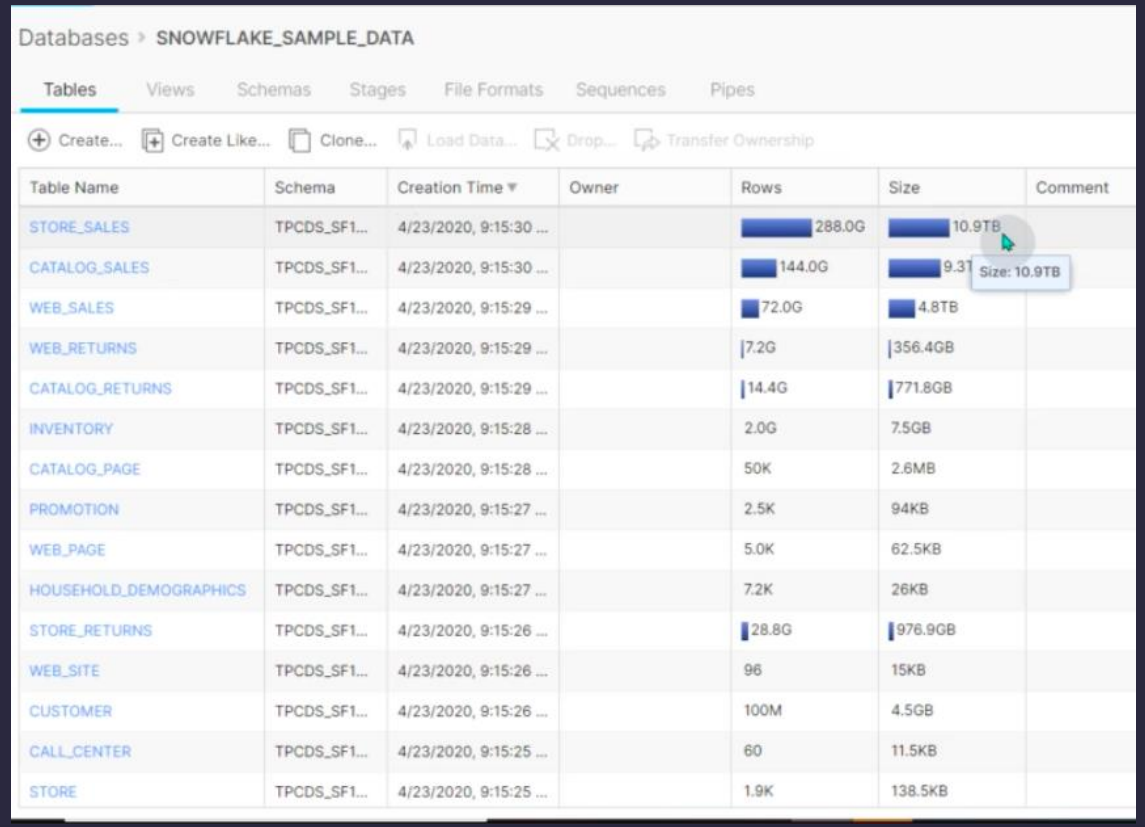
CREATING WAREHOUSE



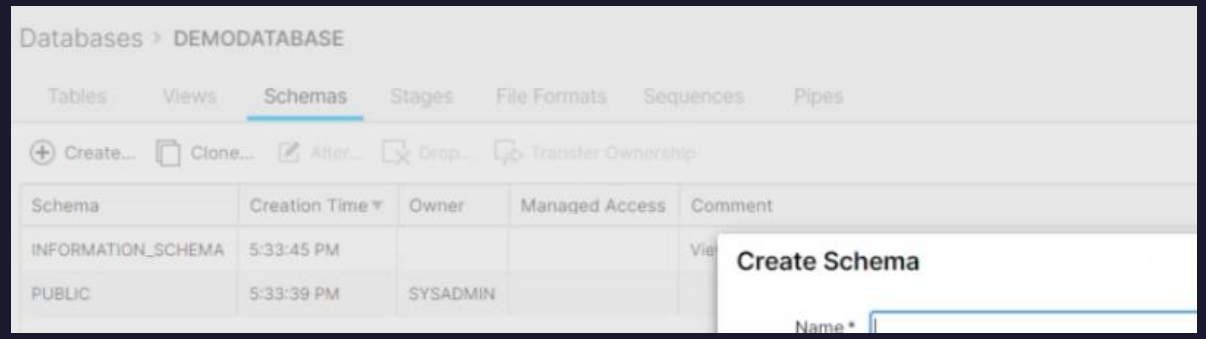


CREATE DATABASE AND TABLES





CREATE SCHEMAS



Create... Clone... Alter... Drop... Transfer Ownership

Schema	Creation Time	Owner	Managed Access	Comment
INFORMATION_SCHEMA	5:32:23 PM			
TPCDS_SF100TCL	4/23/2020, 9:15:20 AM			
TPCDS_SF10TCL	4/23/2020, 6:59:10 AM			
TPCH_SF10000	4/23/2020, 6:34:53 AM			
TPCH_SF1000	4/23/2020, 6:34:35 AM			
TPCH_SF100	4/23/2020, 6:34:16 AM			
TPCH_SF10	4/23/2020, 6:33:58 AM			
TPCH_SF1	4/23/2020, 6:33:39 AM			
TPCH_SF001	4/23/2020, 6:33:15 AM			

Create Schema

Unable to create schema TEST.
SQL execution error: Creating schema on shared database 'SNOWFLAKE_SAMPLE_DATA' is not allowed.

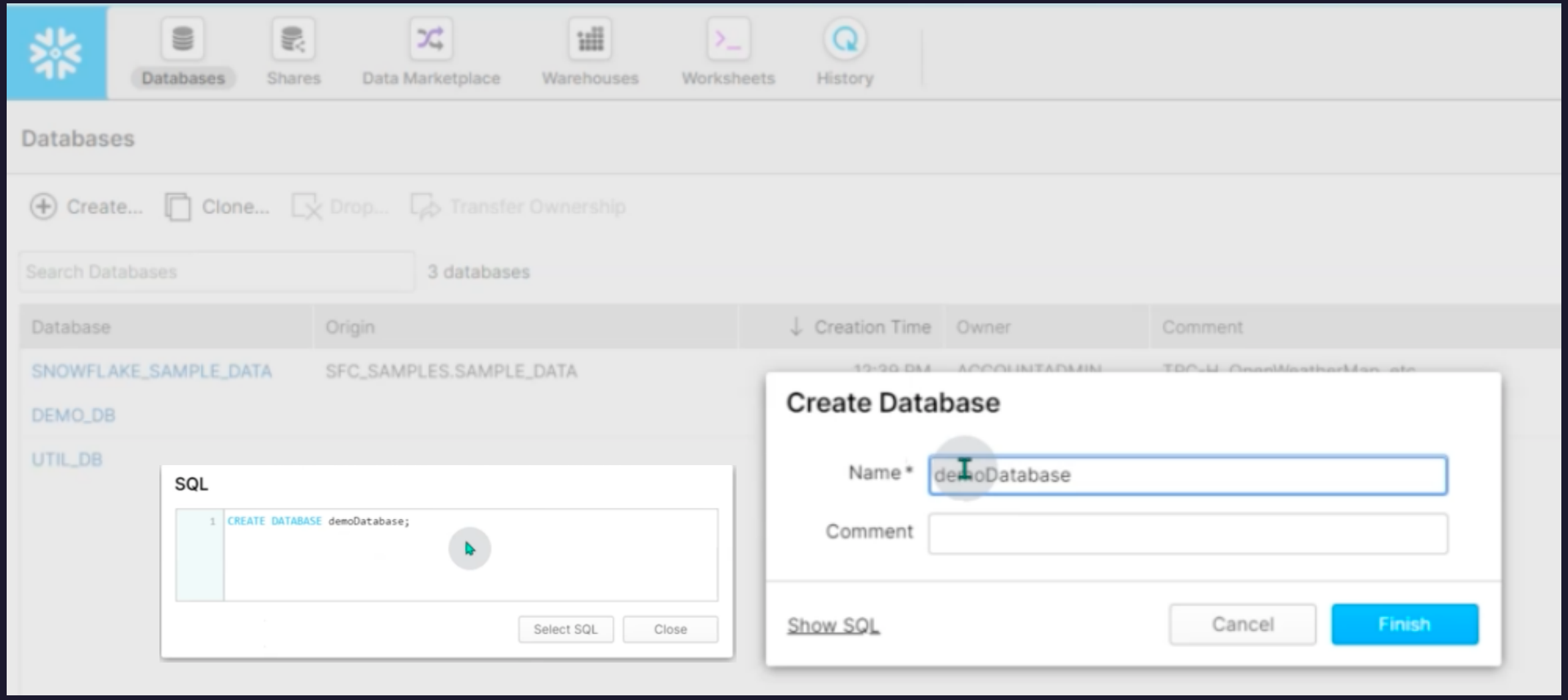
Name * test

Comment

Managed Access

Show SQL | Cancel | Finish

CREATE DATABASE



CREATE TABLES

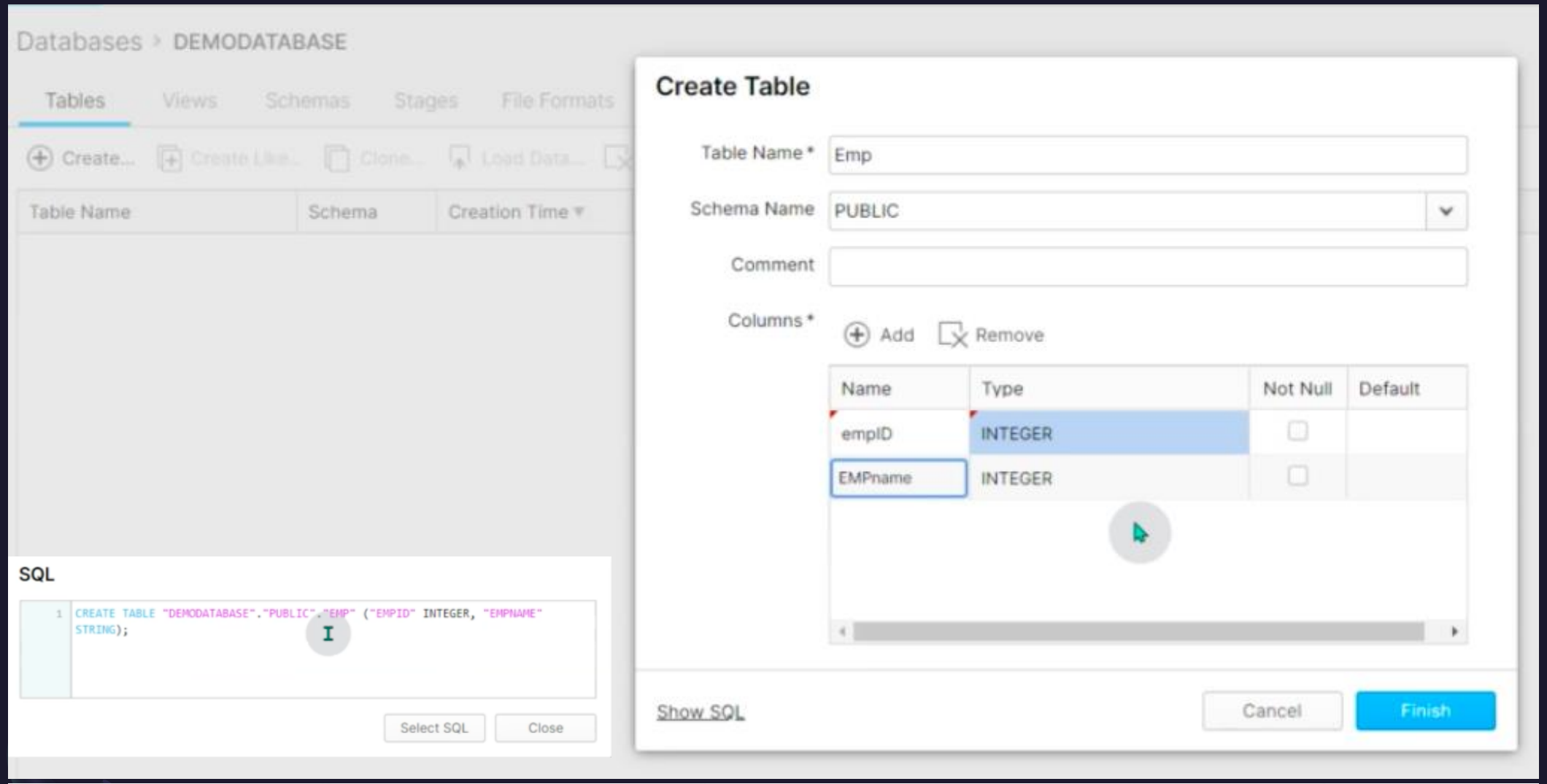
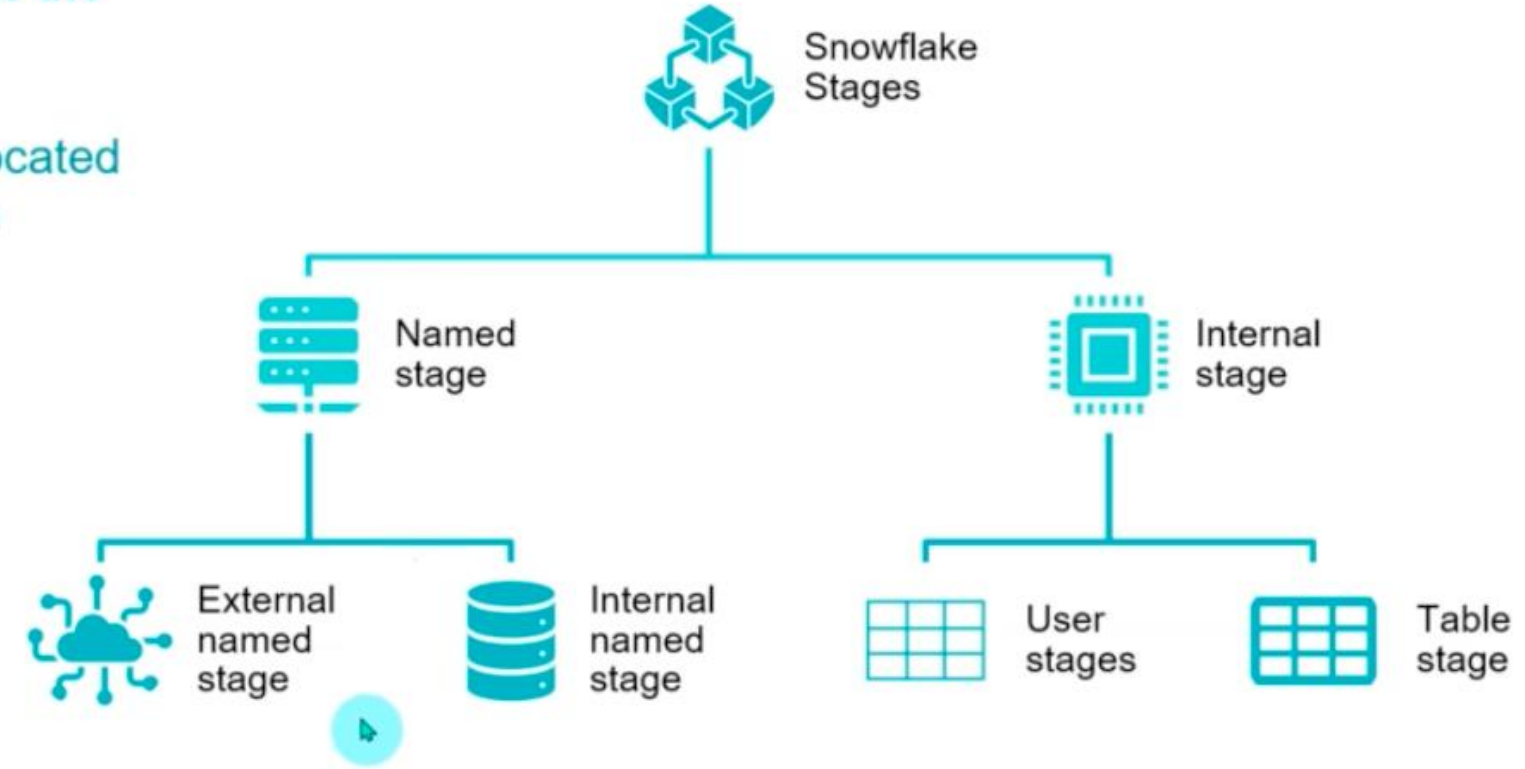


Table Name	Schema	Creation Time	Owner	Rows	Size	Comment
EMP	PUBLIC	5:35:01 PM	SYSADMIN			

Snowflake Stages

- Snowflake requires the staging area where data gets loaded before ingesting into the Snowflake tables
- Stage can be internal or externally located on the cloud or your local file system



DATA LOAD

Bulk Loading

- Using the COPY command from stage
- Relies on the user-provisioned COMPUTE resources i.e. Data Warehouse
- Supports simple transformations

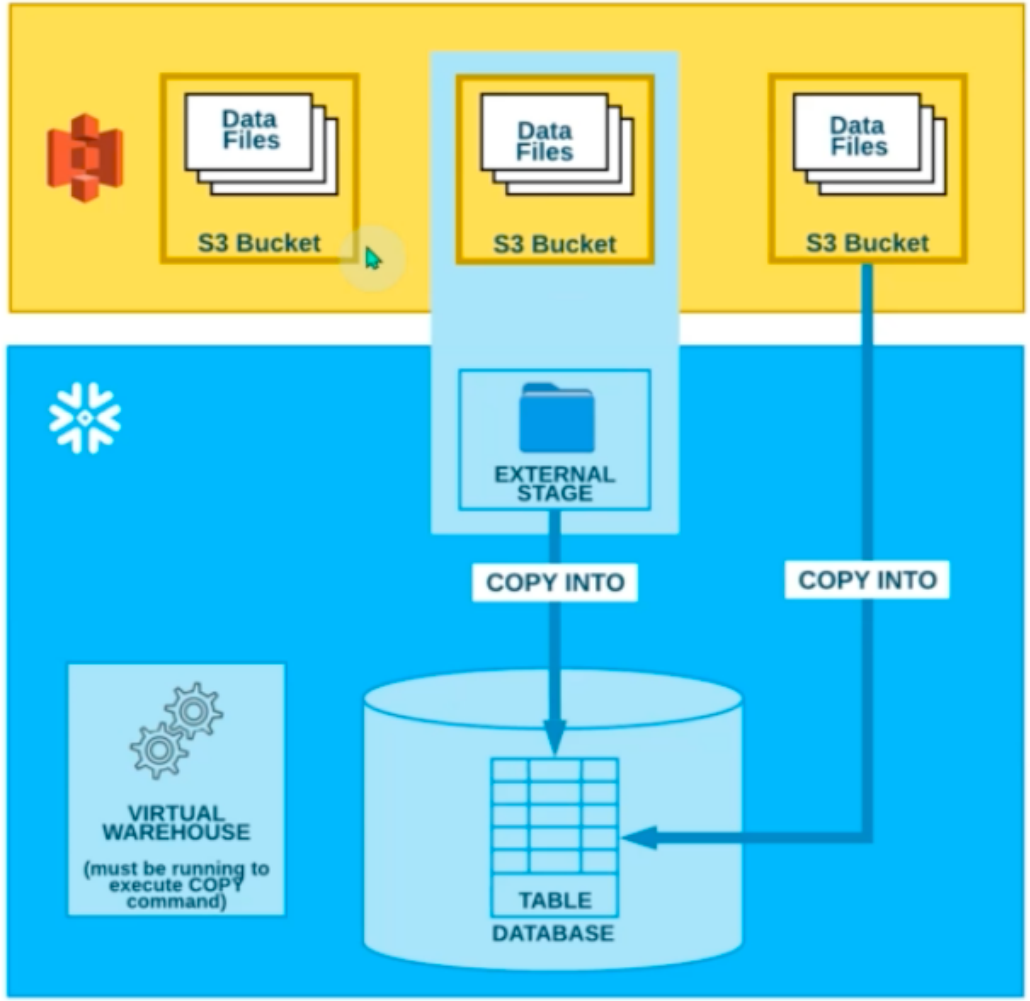
Continuous Loading

- Using Snowpipe
- Doesn't rely on the user-provisioned COMPUTE resource but provisioned inter nally by Snowflake
- Supports simple and complex transformations

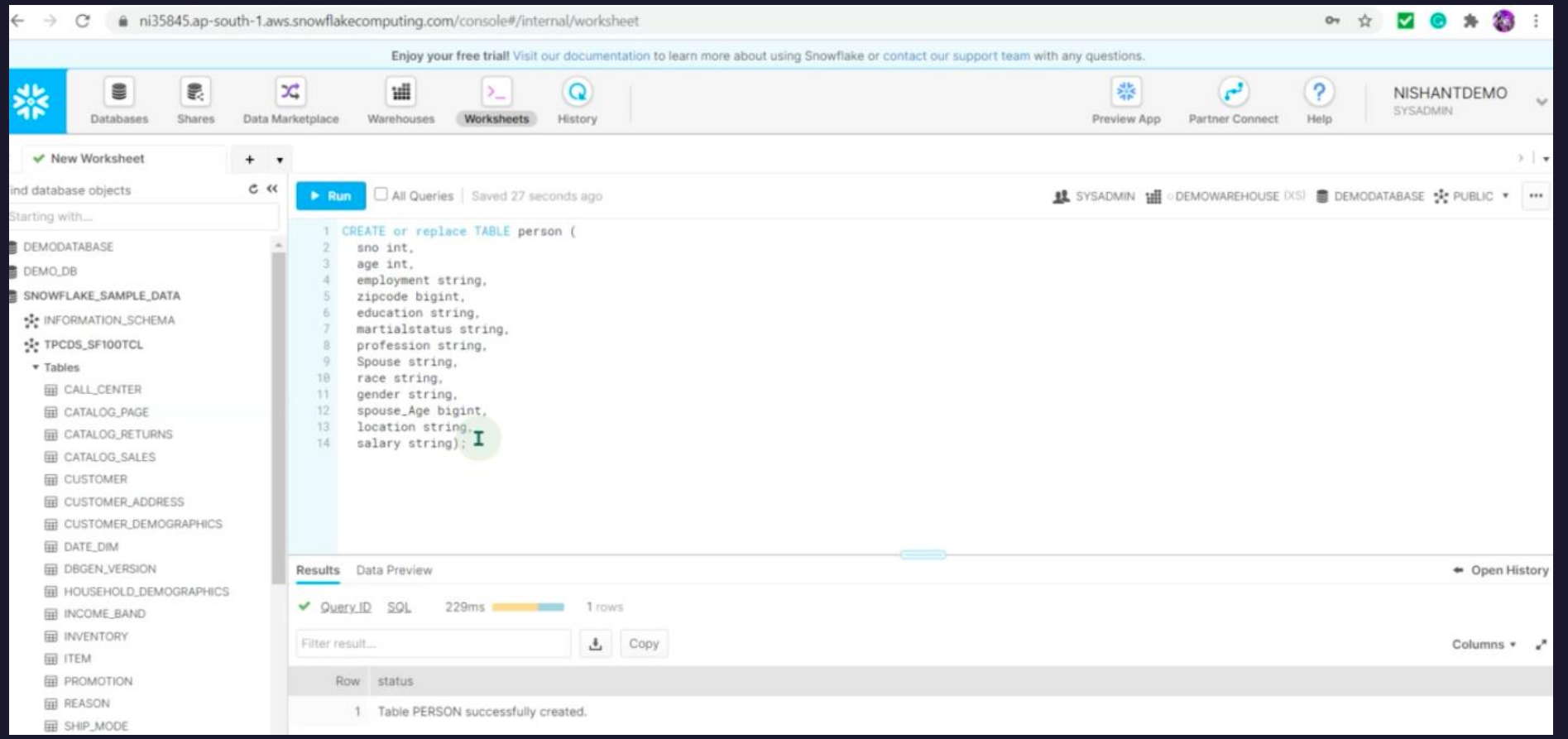
DATA LOAD CONTINUES..

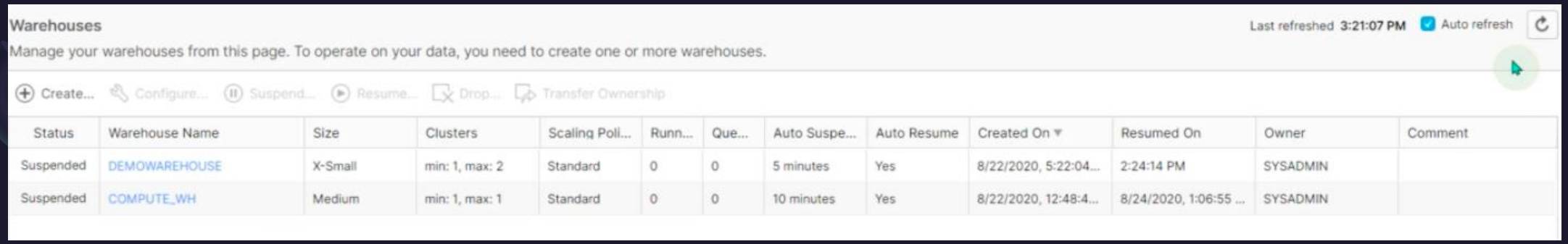
Bulk Loading

- Without named stage
- Using External stage
- With simple transformations

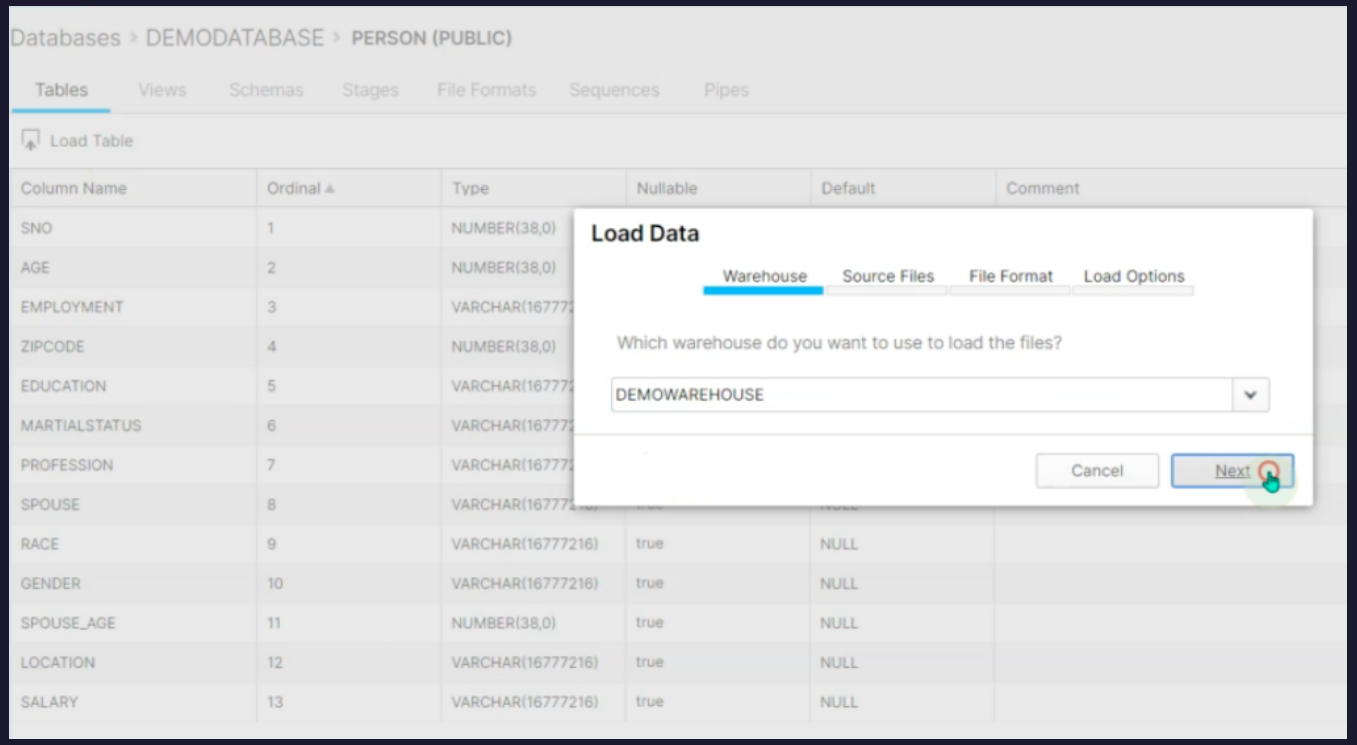


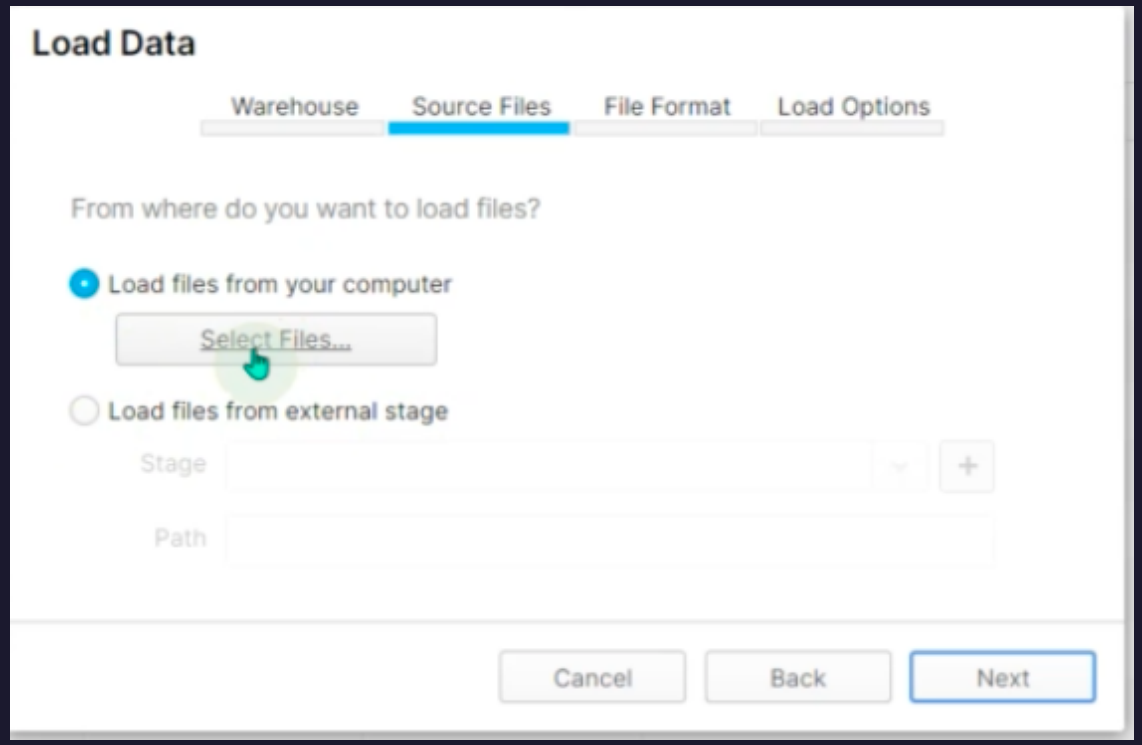
LAB : DATA LOAD



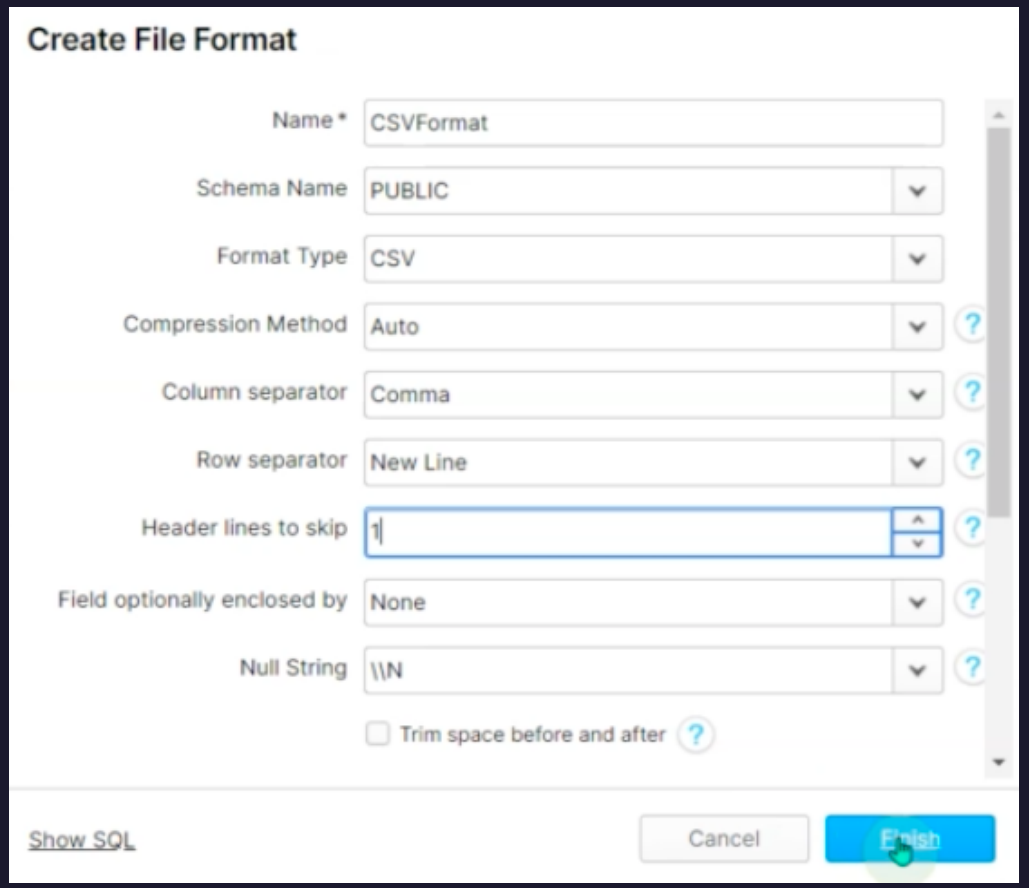


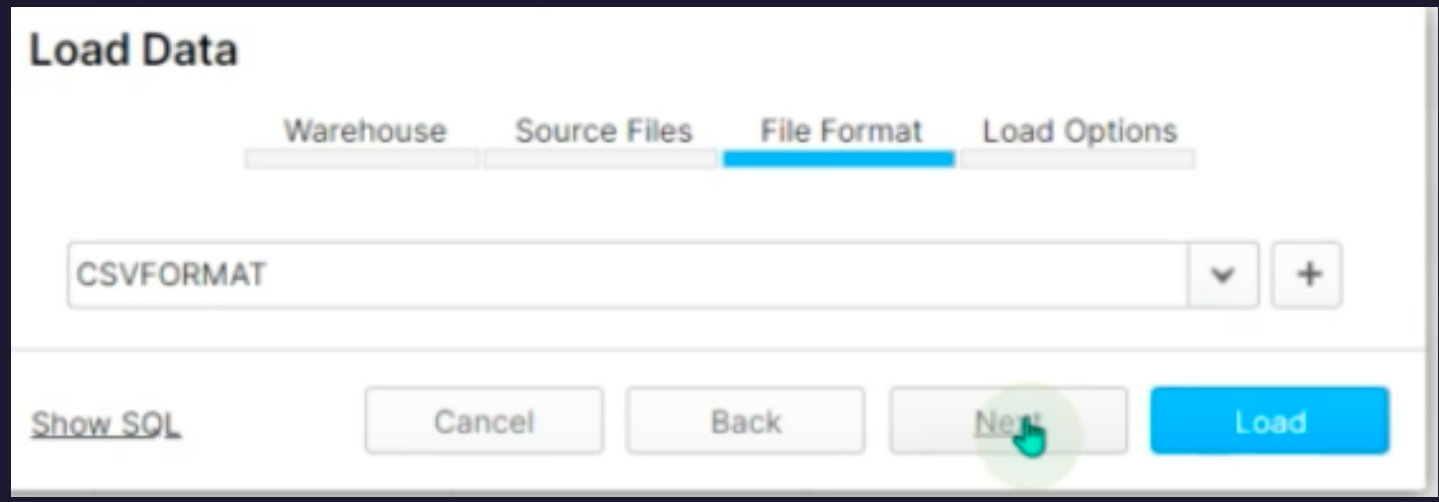
LAB : DATA LOAD

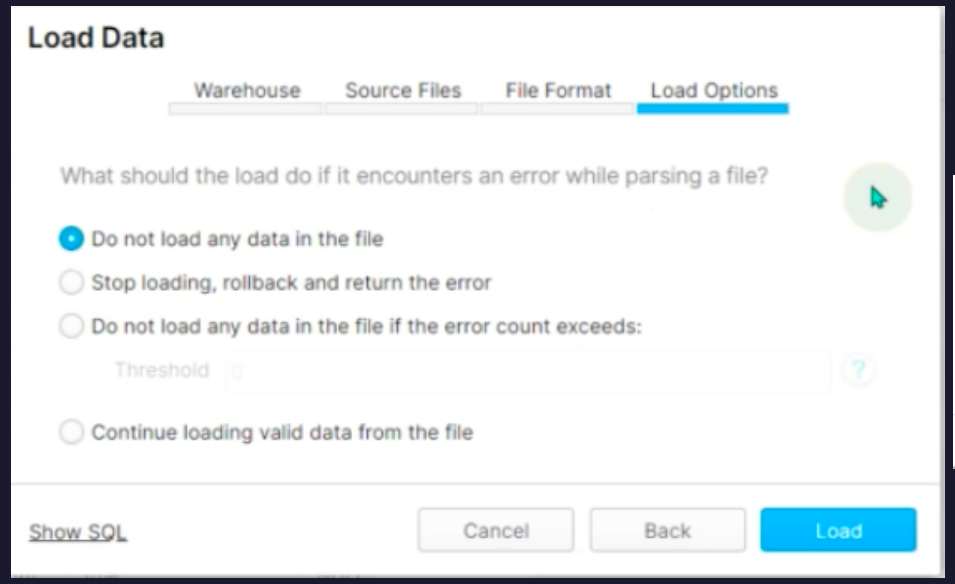


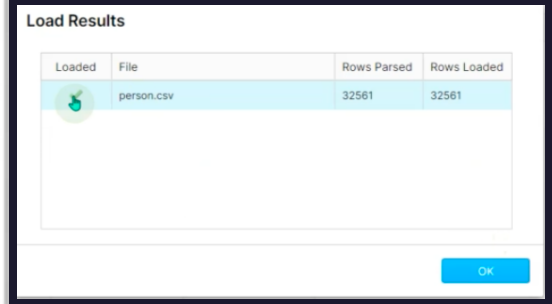


LAB : DATA LOAD









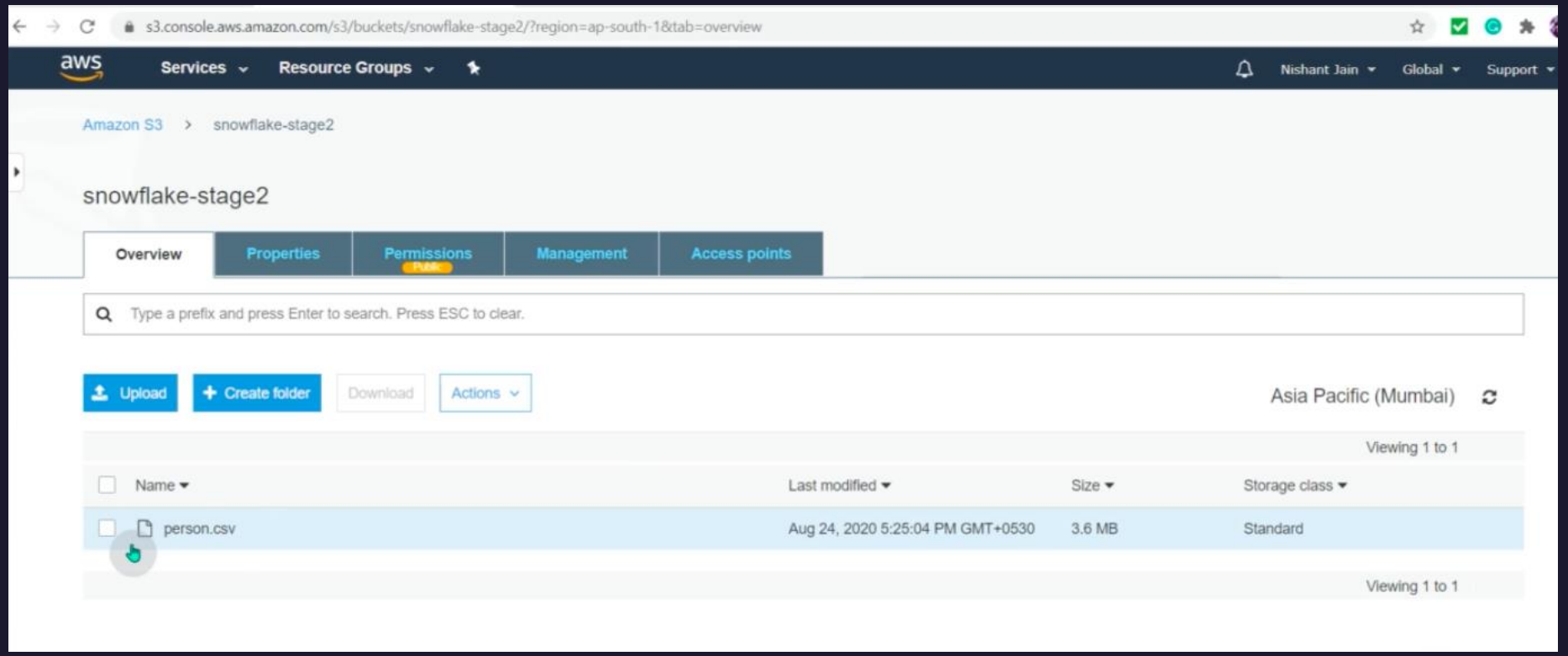
*Untitled - Notepad

File Edit Format View Help

```
PUT file://<file_path>/person.csv @PERSON/ui1598522962697

COPY INTO "DEMODATABASE"."PUBLIC"."PERSON" FROM @/ui1598522962697 FILE_FORMAT =
'"DEMODATABASE"."PUBLIC"."CSVFORMAT"' ON_ERROR = 'CONTINUE' PURGE = TRUE;
```

BULK DATA LOAD : PART II



BULK DATA LOAD : PART II

Run | All Queries | Saved 2 seconds ago

```
1  create or replace stage mystage url='s3://snowflake-stage2/';
2
3  copy into person from @mystage file_format=CSVFORMAT;
4
5  copy into persondata from s3://snowflake-stage2;
```

Run | All Queries

```
1  show stages;
2
3  select t.$1,$2, | from @mystage t;
```

Databases | Shares | Data Marketplace | Warehouses | Worksheets | History

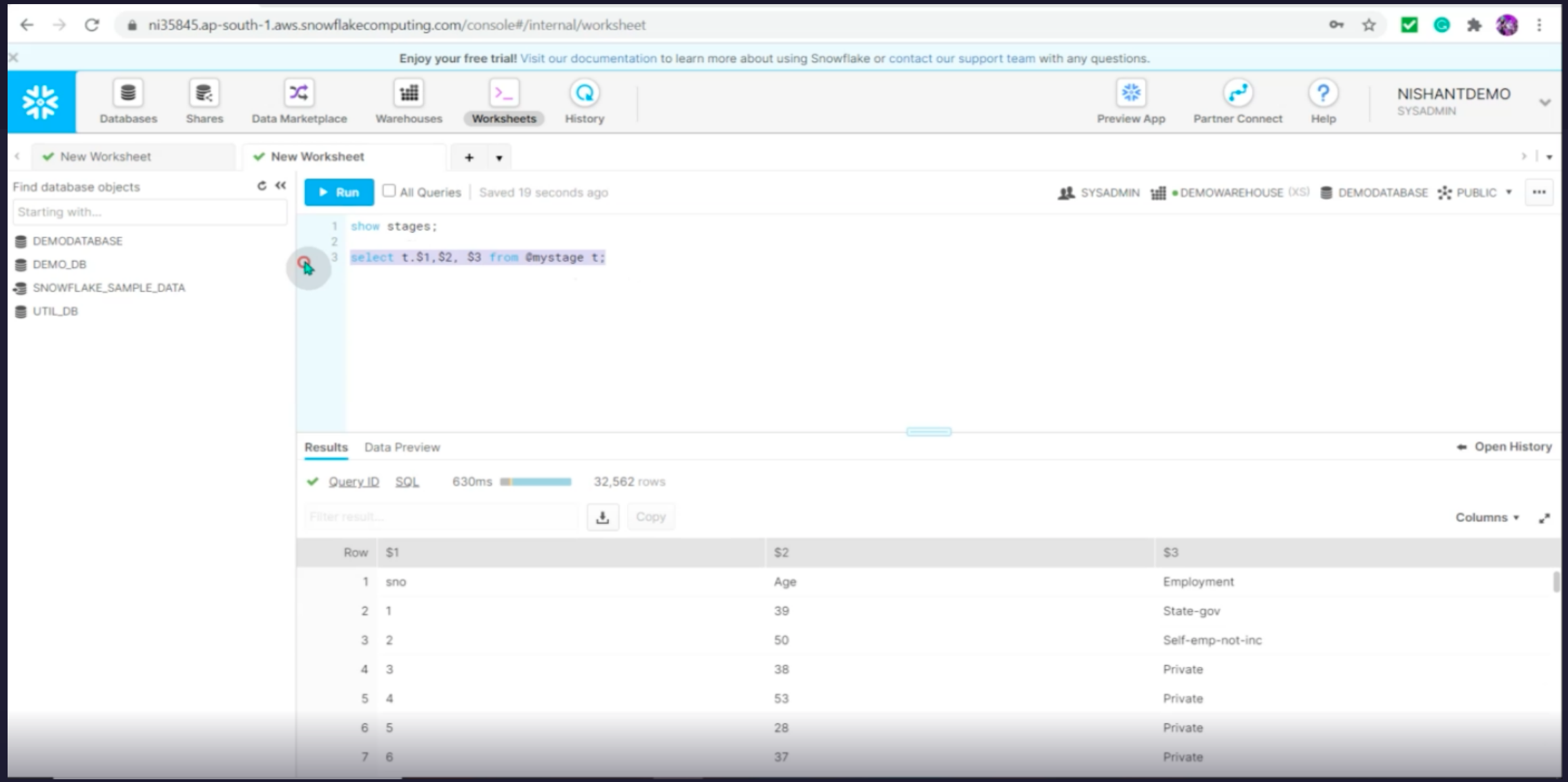
Databases › DEMODATABASE

Tables | Views | Schemas | Stages | File Formats | Sequences | Pipes

Create... | Clone... | Edit... | Drop... | Transfer Ownership

Stage	Schema	Location	Creation Time	Owner	Comment
MYSTAGE	PUBLIC	s3://snowflake-stage2/	4:01:28 PM	SYSADMIN	

BULK DATA LOAD : PART II



THANK YOU...!!!