

SNOWFLAKE

THE CLOUD DATA

PLATFORM



Snowflake
Fundamentals



Introduction

Introduction

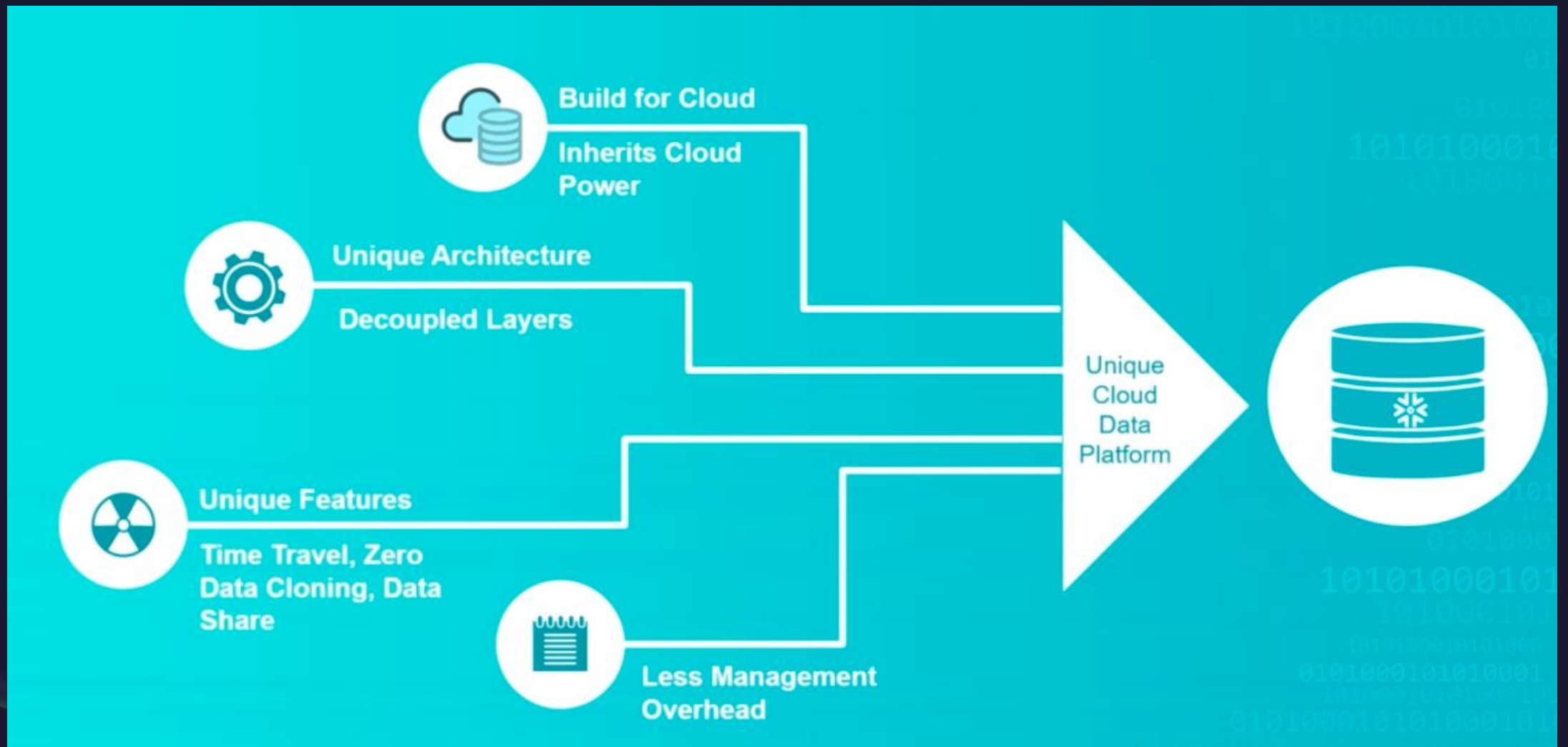
- ❖ We are going to learn basic Snowflake concepts and fundamentals
- ❖ Doing hands-on labs, no prior knowledge of Snowflake is required
- ❖ Basic SQL and database knowledge is required
- ❖ Preferred to have knowledge on Cloud
- ❖ Take your time for the hands-on

What is Snowflake ?

- Snowflake is an analytic data warehouse provided as Software-as- Services(SaaS). Snowflake provides data warehouse that is faster, easier to use and more flexible than 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.

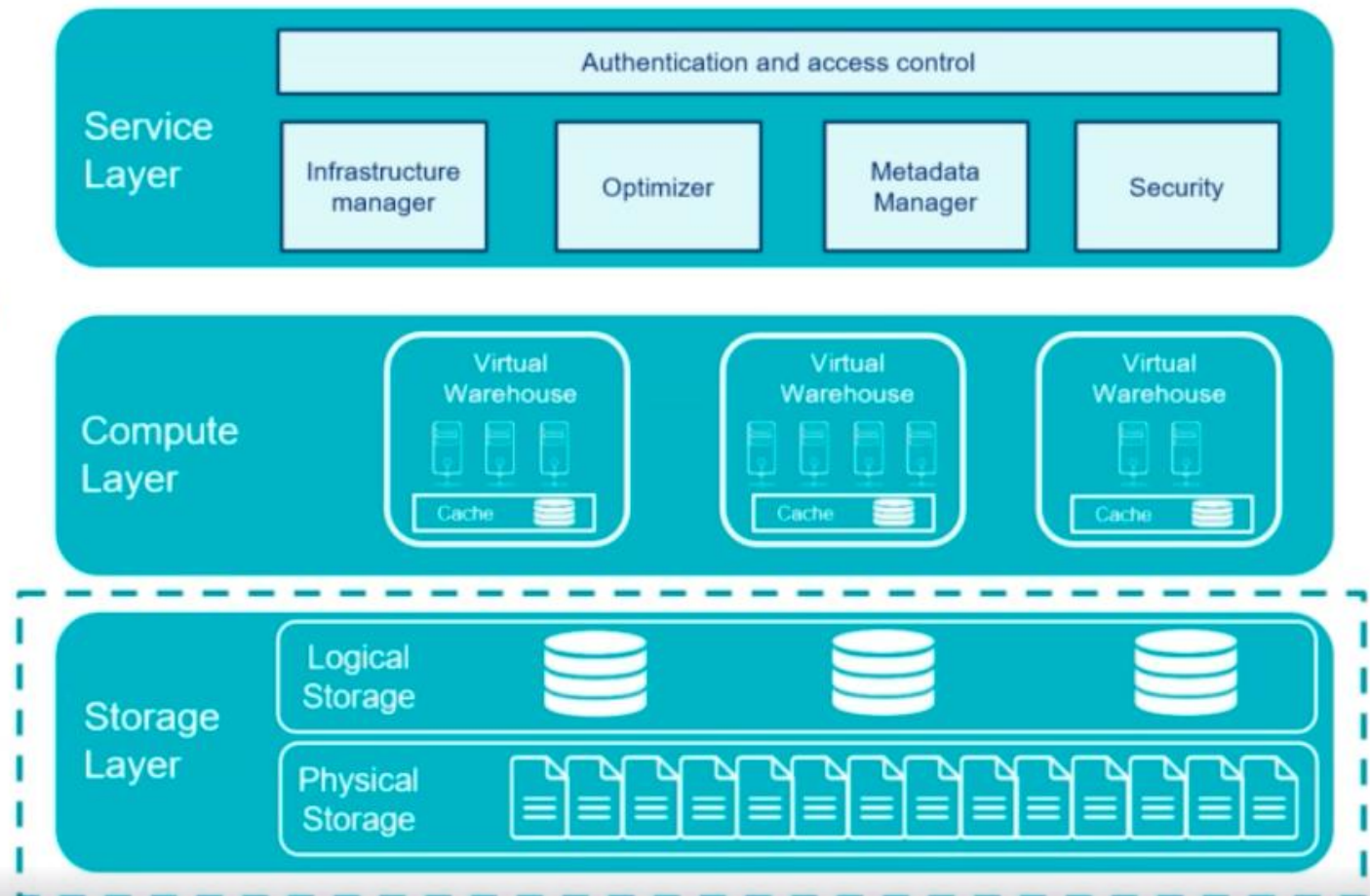
Snowflake completely runs on cloud infrastructure. All the component of the snowflake service runs on public cloud infrastructure.

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 premises)

Snowflake Architecture

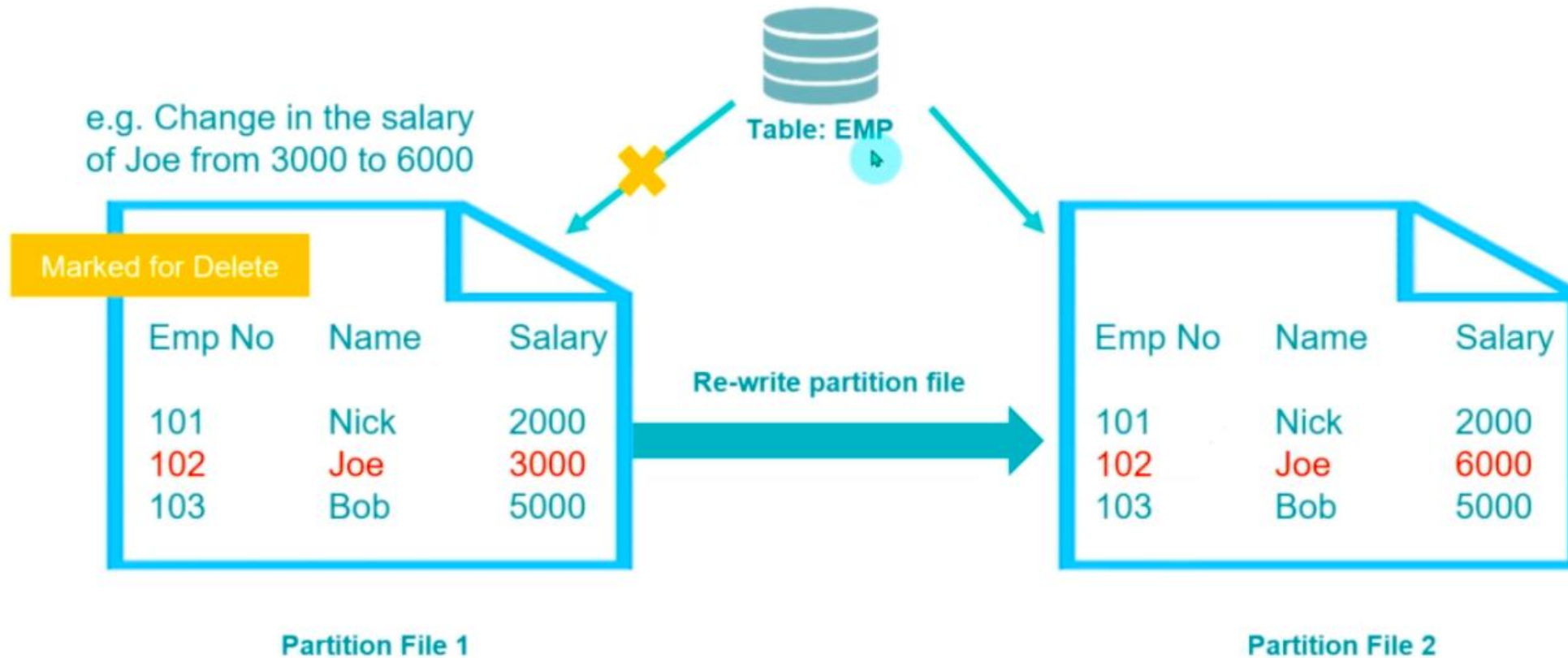
Storage Layer

- 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 proprietary to Snowflake
- Files are immutable
- Table definitions are logical and stored in its metadata layer



Update Record

Update Record

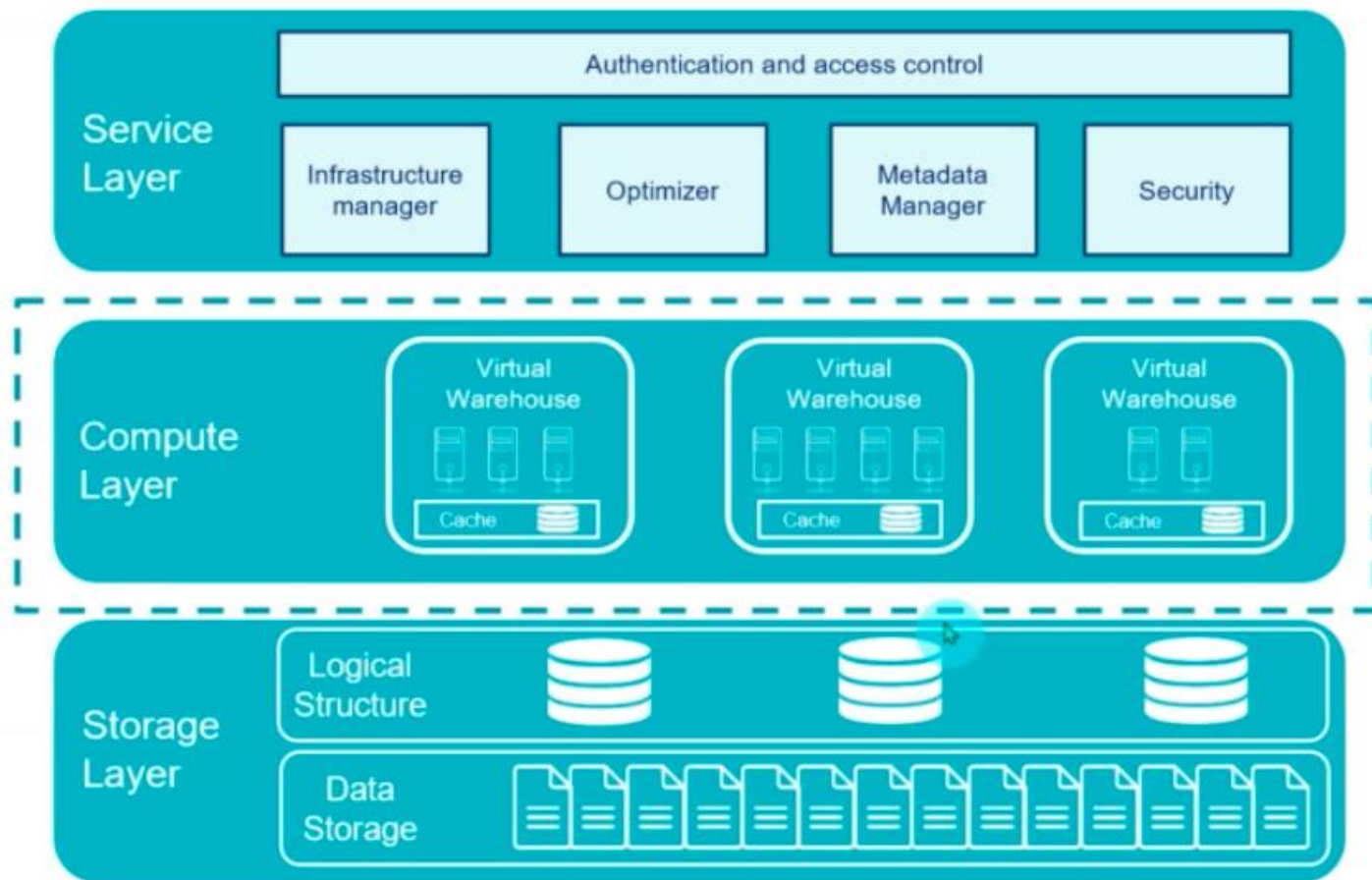


After record update new partition file 2 will be created and partition file 1 will be marked for delete

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





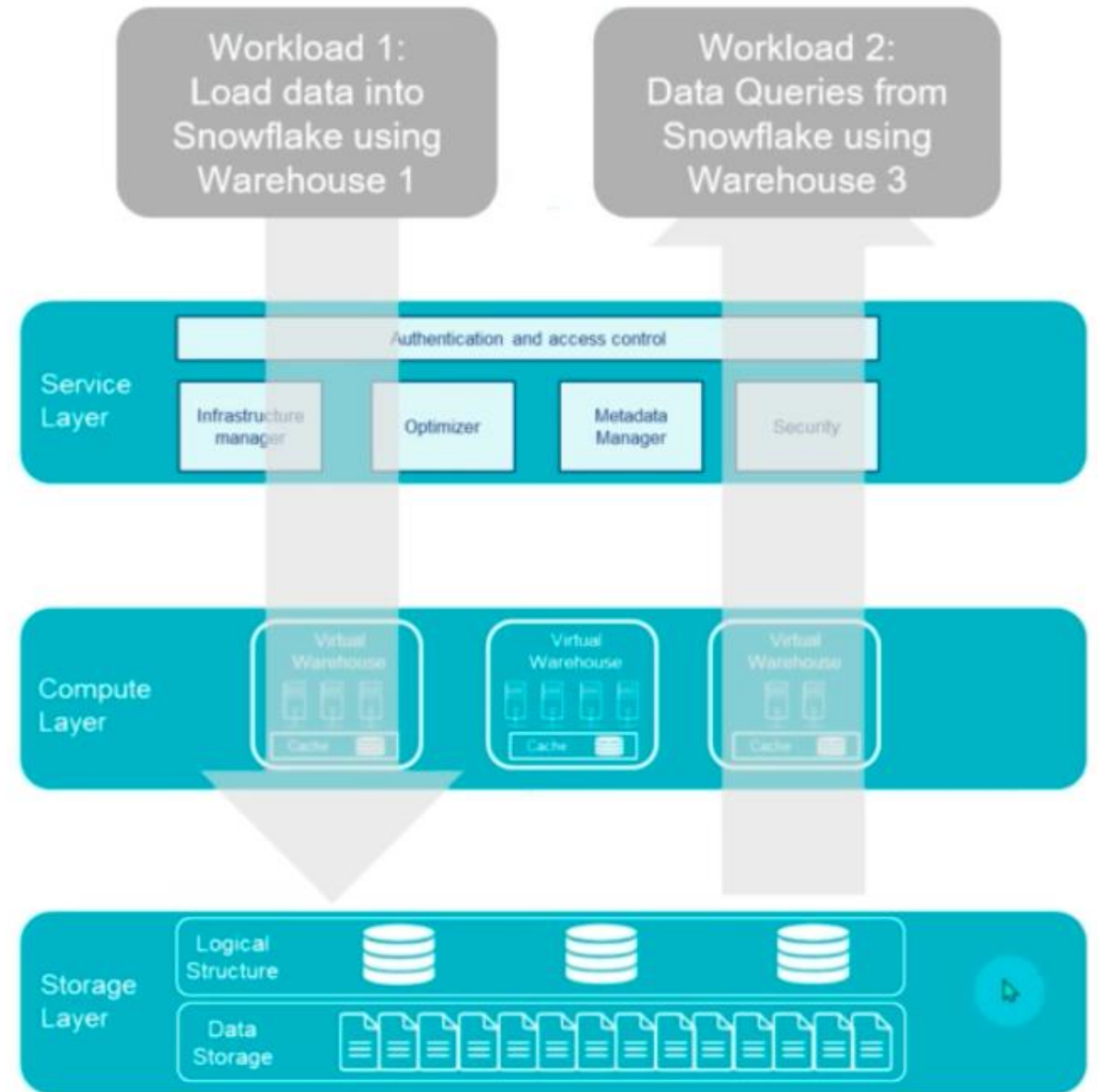
Warehouse Size & Pricing

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

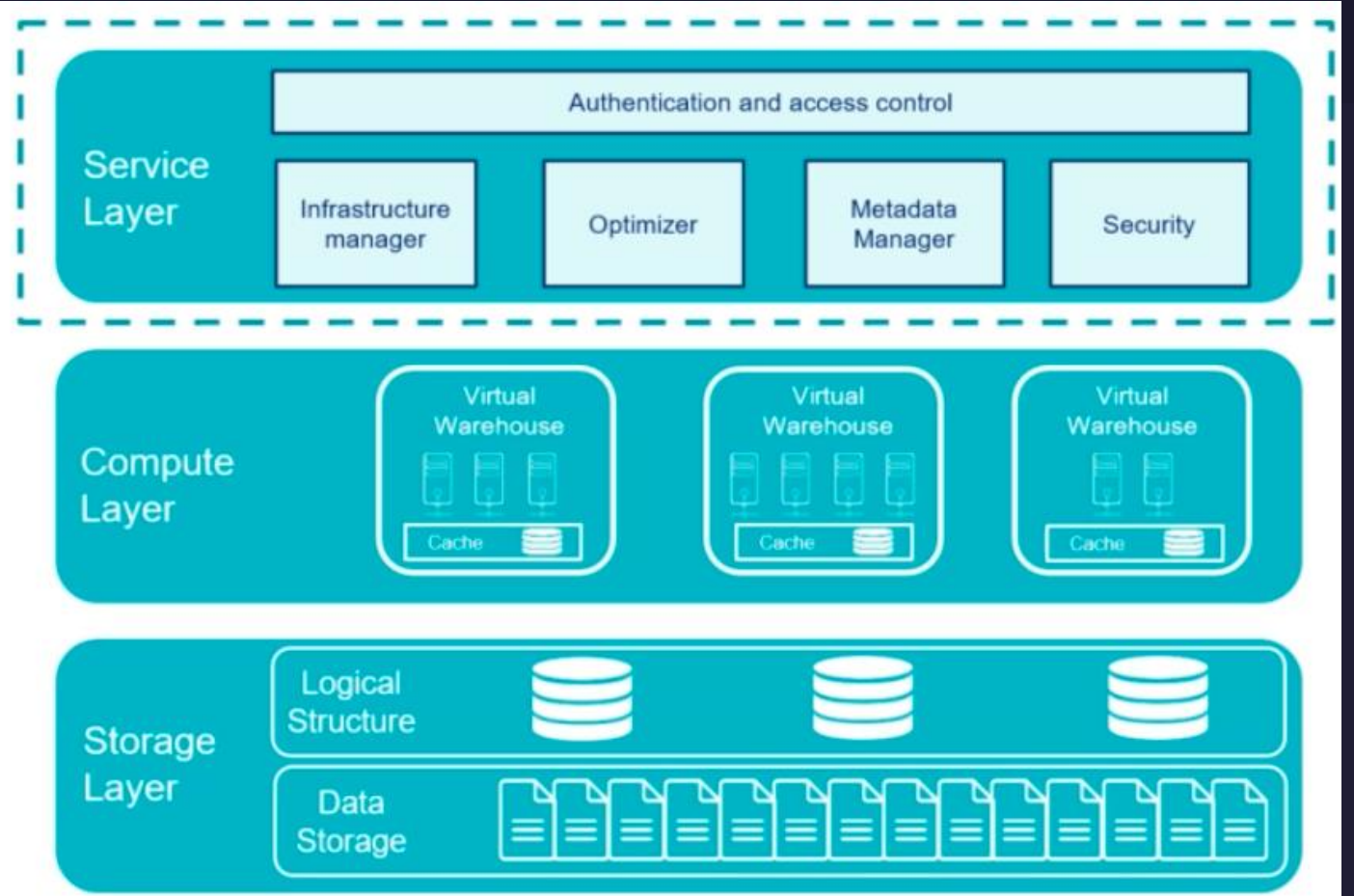


Table and Partition File Mapping

Query 1: Select * from EMP where Emp=103

Partition scan

Emp No	Name	Salary
101	Nick	2000
102	Joe	3000
103	Bob	5000

Partition File 1

Query 2: Select * from EMP where Emp=104

Partition scan

Emp No	Name	Salary
104	David	2000
105	Nicole	6000
106	Jack	5000

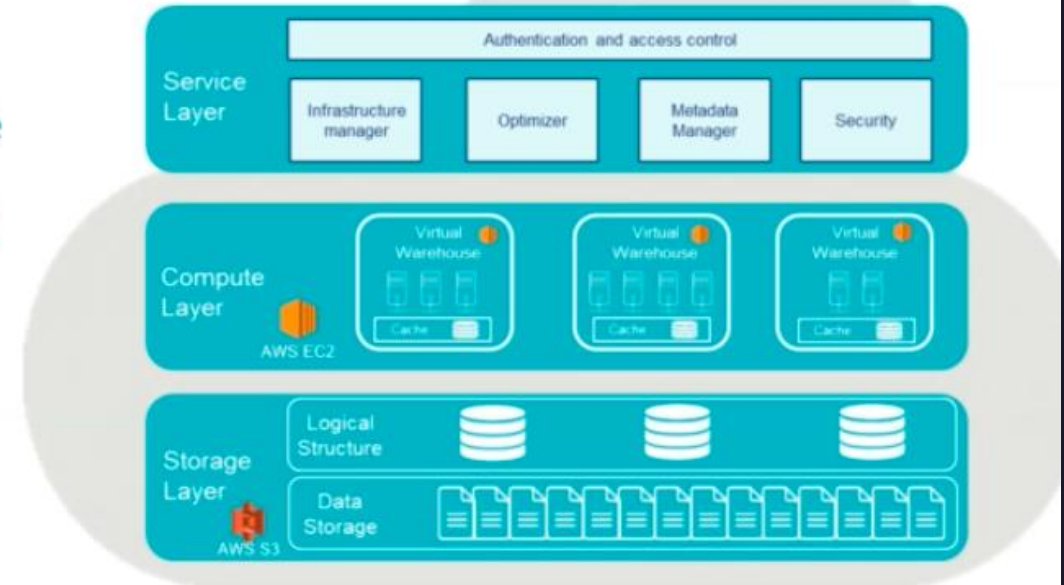
Partition File 2

Query 3: Select * from EMP where Emp in (103,105)

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




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.

Cloud provider

 Amazon Web Services

 Google Cloud Platform

 Microsoft Azure

Asia Pacific (Mumbai)

IN+91 In...

☒ Check here to indicate that you have read and agree to the terms of the [Snowflake Self Service On Demand Terms](#).

GET STARTED

Start Your 30-Day Free Trial
Receive \$400 worth of free usage with sign-up.

First Name

Last Name

Email

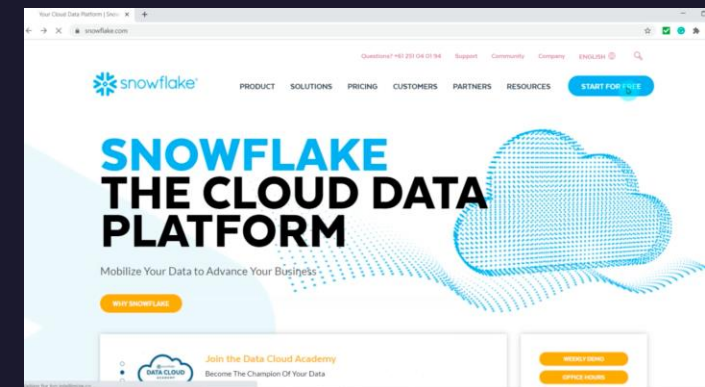
Title

Company


Country

By clicking the button below you understand that Snowflake will process your personal information in accordance with its [Privacy Notice](#).

CONTINUE



GET STARTED



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



Welcome to Snowflake, Nishant Jain!

Please choose your user name and password to get started.

User Name

Password

Confirm Password

[Get Started](#)

SNOWFLAKE CONSOLE

The screenshot displays the Snowflake Console interface in a web browser. The browser's address bar shows the URL: `ni35845.ap-south-1.aws.snowflakecomputing.com/console#/internal/worksheet`. The console features a top navigation bar with icons for Databases, Shares, Data Marketplace, Warehouses, Worksheets (active), and History. On the right of this bar are links for Preview App, Partner Connect, Help, and a user profile for NISHANTDEMO (SYSADMIN). Below the navigation bar, a banner reads: "Enjoy your free trial! Visit our documentation to learn more about using Snowflake or contact our support team with any questions." The main workspace is titled "New Worksheet" and includes a "Find database objects" search bar. A sidebar on the left lists available databases: DEMO_DB, SNOWFLAKE_SAMPLE_DATA, and UTIL_DB. The central area contains a query editor with a "Run" button, a checkbox for "All Queries", and a timestamp "Saved 11 minutes ago". Below the editor is a "Results" pane with a "Data Preview" tab. The results area is currently empty, displaying the text "Query results will appear here." A blue circle with a mouse cursor is positioned over the "Results" tab. At the bottom right of the results pane, there is a link to "Open History".

CREATING WAREHOUSE

Enjoy your free trial! Visit our documentation to learn more about using Snowflake or contact our support team with any questions.

Databases Shares Data Marketplace **Warehouses** Worksheets History Preview App Partner Connect Help

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

[Create...](#) [Configure...](#) [Suspend...](#) [Resume...](#) [Drop...](#) [Transfer Ownership](#)

Status	Create a new warehouse	Size	Clusters	Scaling Pol...	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 ▼
[Learn more about virtual warehouse sizes here](#)

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

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

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

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

☒ Auto Resume ?

Comment

[Show SQL](#) Cancel Finish

CREATE DATABASE AND TABLES

Enjoy your free trial! Visit our documentation to learn more about using Snowflake or contact our support team w

Databases

Shares

Data Marketplace

Warehouses

Worksheets

History

Databases

+

Create...

Clone...

Drop...

Transfer Ownership

Search Databases

3 databases

Database	Origin	Creation Time	Owner	Comment
SNOWFLAKE_SAMPLE_DATA	SFC_SAMPLES.SAMPLE_DATA	12:39 PM	ACCOUNTADMIN	TPC-H, OpenWeatherMap, etc
DEMO_DB		12:39 PM	SYSADMIN	demo database
UTIL_DB		12:38 PM	SYSADMIN	utility database

Databases > SNOWFLAKE_SAMPLE_DATA

Tables

Views

Schemas

Stages

File Formats

Sequences

Pipes

+

Create...

Create Like...

Clone...

Load Data...

Drop...

Transfer Ownership

Table Name	Schema	Creation Time	Owner	Rows	Size	Comment
STORE_SALES	TPCDS_SF1...	4/23/2020, 9:15:30 ...		288.0G	10.9TB	
CATALOG_SALES	TPCDS_SF1...	4/23/2020, 9:15:30 ...		144.0G	9.3TB	Size: 10.9TB
WEB_SALES	TPCDS_SF1...	4/23/2020, 9:15:29 ...		72.0G	4.8TB	
WEB_RETURNS	TPCDS_SF1...	4/23/2020, 9:15:29 ...		7.2G	356.4GB	
CATALOG_RETURNS	TPCDS_SF1...	4/23/2020, 9:15:29 ...		14.4G	771.8GB	
INVENTORY	TPCDS_SF1...	4/23/2020, 9:15:28 ...		2.0G	7.5GB	
CATALOG_PAGE	TPCDS_SF1...	4/23/2020, 9:15:28 ...		50K	2.6MB	
PROMOTION	TPCDS_SF1...	4/23/2020, 9:15:27 ...		2.5K	94KB	
WEB_PAGE	TPCDS_SF1...	4/23/2020, 9:15:27 ...		5.0K	62.5KB	
HOUSEHOLD_DEMOGRAPHICS	TPCDS_SF1...	4/23/2020, 9:15:27 ...		7.2K	26KB	
STORE_RETURNS	TPCDS_SF1...	4/23/2020, 9:15:26 ...		28.8G	976.9GB	
WEB_SITE	TPCDS_SF1...	4/23/2020, 9:15:26 ...		96	15KB	
CUSTOMER	TPCDS_SF1...	4/23/2020, 9:15:26 ...		100M	4.5GB	
CALL_CENTER	TPCDS_SF1...	4/23/2020, 9:15:25 ...		60	11.5KB	
STORE	TPCDS_SF1...	4/23/2020, 9:15:25 ...		1.9K	138.5KB	

CREATE SCHEMAS

Databases > DEMODATABASE

Tables Views Schemas Stages File Formats Sequences Pipes

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

Schema	Creation Time ▾	Owner	Managed Access	Comment
INFORMATION_SCHEMA	5:33:45 PM			View
PUBLIC	5:33:39 PM	SYSADMIN		

Create Schema

Name*

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

Schema	Creation Time ▾	Owner	Managed Access	Comment
INFORMATION_SCHEMA	5:32:23 PM			V
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*

Comment

☐ Managed Access

Show SQL

Cancel

Finish

CREATE DATABASE

The screenshot displays the Snowflake web interface for managing databases. The top navigation bar includes icons for Databases, Shares, Data Marketplace, Warehouses, Worksheets, and History. The 'Databases' section is active, showing a list of existing databases: SNOWFLAKE_SAMPLE_DATA, DEMO_DB, and UTIL_DB. A modal dialog titled 'Create Database' is open, allowing the user to enter a new database name and comment. The 'Name' field is pre-filled with 'demoDatabase'. Below the dialog, a small SQL editor window shows the generated SQL command: 'CREATE DATABASE demoDatabase;'. The background interface also shows action buttons like '+ Create...', 'Clone...', 'Drop...', and 'Transfer Ownership', along with a search bar and a table of database details.

Databases

+ Create... Clone... Drop... Transfer Ownership

Search Databases 3 databases

Database	Origin	Creation Time	Owner	Comment
SNOWFLAKE_SAMPLE_DATA	SFC_SAMPLES.SAMPLE_DATA	12:39 PM	ACCOUNTADMIN	TDC-H GreenWeatherMap, etc
DEMO_DB				
UTIL_DB				

Create Database

Name * demoDatabase

Comment

Show SQL Cancel Finish

SQL

```
1 CREATE DATABASE demoDatabase;
```

Select SQL Close

CREATE TABLES

Databases > DEMODATABASE

Tables Views Schemas Stages File Formats

+ Create... + Create Like... Clone... Load Data...

Table Name	Schema	Creation Time
------------	--------	---------------

SQL

1 CREATE TABLE "DEMODATABASE"."PUBLIC"."EMP" ("EMPID" INTEGER, "EMPNAME" STRING);

Select SQL Close

Create Table

Table Name * Emp

Schema Name PUBLIC

Comment

Columns *

+ Add - Remove

Name	Type	Not Null	Default
empID	INTEGER	<input type="checkbox"/>	
EMPname	INTEGER	<input type="checkbox"/>	

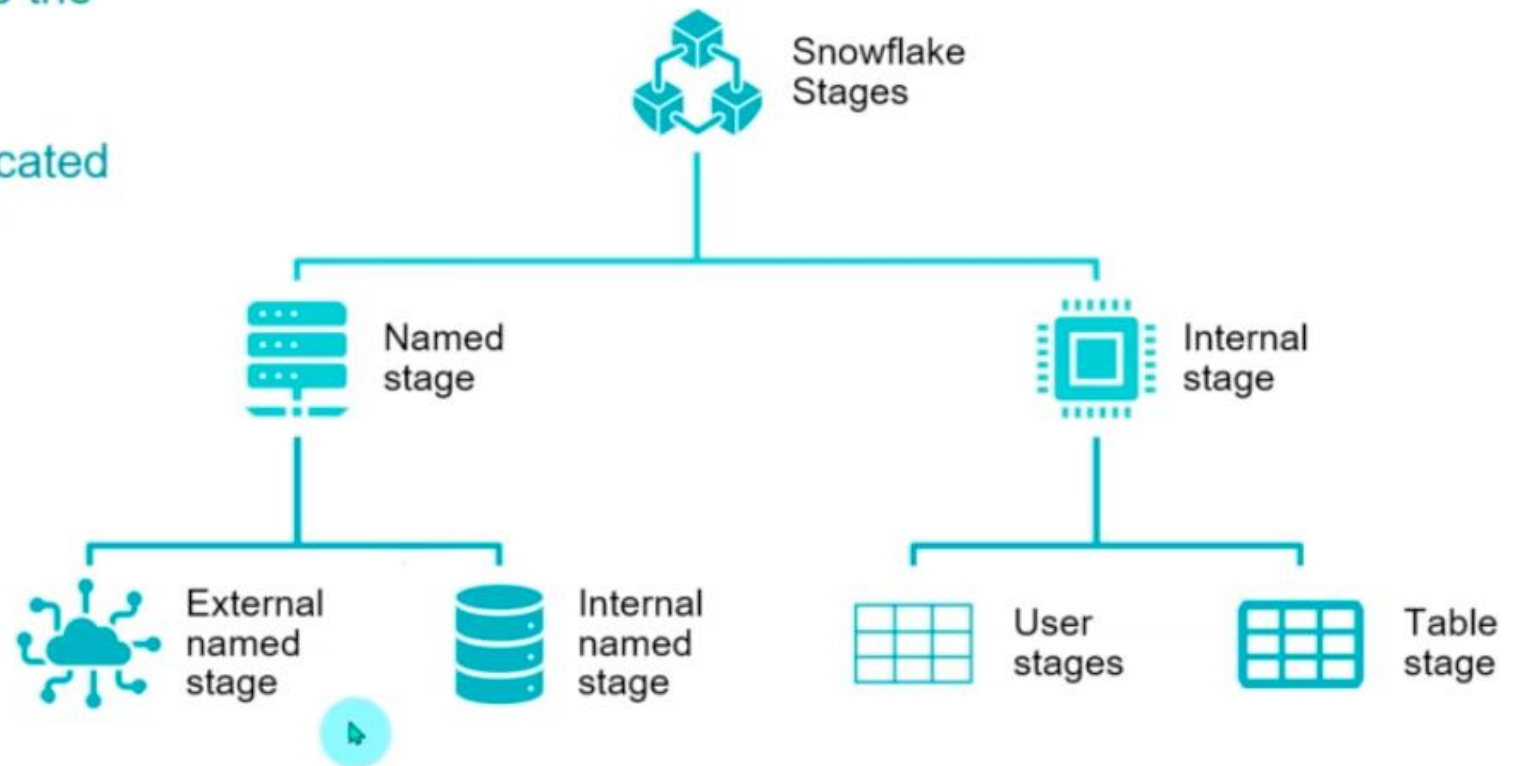
Show SQL

Cancel Finish

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 internally by Snowflake
- ❖ Supports simple and complex transformations

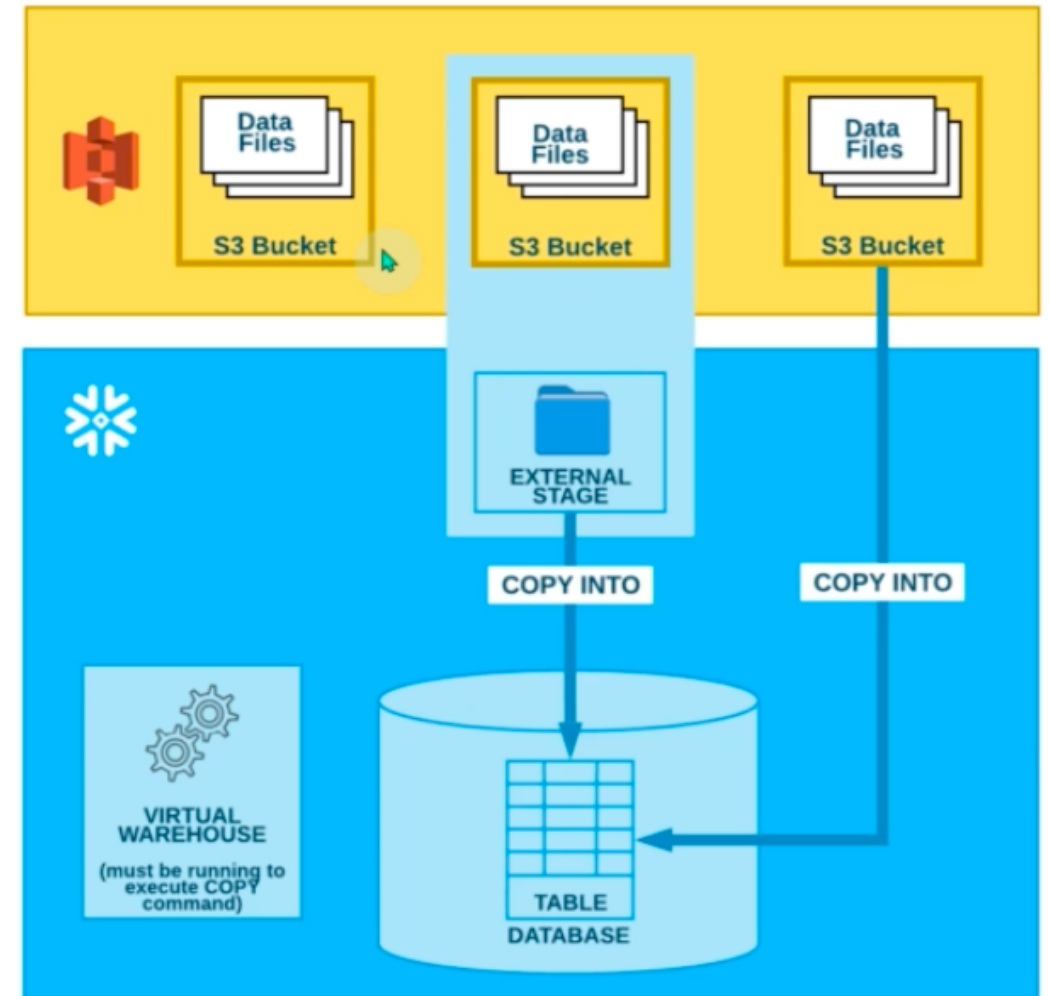
0011100101011001
0101001110101000
1010100010110101
1011011010001010
1110001010100010
1000101110101100
0100110100110100
1000010101001111
0111011011011010



DATA LOAD CONTINUES..

Bulk Loading

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



LAB : DATA LOAD

The screenshot shows the Snowflake web interface. The top navigation bar includes icons for Databases, Shares, Data Marketplace, Warehouses, Worksheets, and History. The 'Worksheets' tab is active. On the left, a sidebar shows the database hierarchy: DEMODATABASE, DEMO_DB, SNOWFLAKE_SAMPLE_DATA, INFORMATION_SCHEMA, and TPCDS_SF100TCL. The main area displays a SQL query: `CREATE or replace TABLE person (sno int, age int, employment string, zipcode bigint, education string, martialstatus string, profession string, Spouse string, race string, gender string, spouse_Age bigint, location string, salary string);`. Below the query, the 'Results' tab shows a single row with the status 'Table PERSON successfully created.'.

Enjoy your free trial! Visit our documentation to learn more about using Snowflake or contact our support team with any questions.

NEW Worksheet

Find database objects

Starting with...

DEMODATABASE

DEMO_DB

SNOWFLAKE_SAMPLE_DATA

INFORMATION_SCHEMA

TPCDS_SF100TCL

Tables

CALL_CENTER

CATALOG_PAGE

CATALOG_RETURNS

CATALOG_SALES

CUSTOMER

CUSTOMER_ADDRESS

CUSTOMER_DEMOGRAPHICS

DATE_DIM

DBGEN_VERSION

HOUSEHOLD_DEMOGRAPHICS

INCOME_BAND

INVENTORY

ITEM

PROMOTION

REASON

SHIP_MODE

Run

All Queries | Saved 27 seconds ago

SYSADMIN DEMOWAREHOUSE (XS) DEMODATABASE PUBLIC

```
1 CREATE or replace TABLE person (  
2   sno int,  
3   age int,  
4   employment string,  
5   zipcode bigint,  
6   education string,  
7   martialstatus string,  
8   profession string,  
9   Spouse string,  
10  race string,  
11  gender string,  
12  spouse_Age bigint,  
13  location string,  
14  salary string);
```

Results Data Preview

Open History

Query ID SQL 229ms 1 rows

Filter result...

Copy

Row	status
1	Table PERSON successfully created.

Warehouses Last refreshed 3:21:07 PM Auto refresh

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

Create... Configure... Suspend... Resume... Drop... Transfer Ownership

Status	Warehouse Name	Size	Clusters	Scaling Poll...	Runn...	Que...	Auto Suspe...	Auto Resume	Created On	Resumed On	Owner	Comment
Suspended	DEMOWAREHOUSE	X-Small	min: 1, max: 2	Standard	0	0	5 minutes	Yes	8/22/2020, 5:22:04...	2:24:14 PM	SYSADMIN	
Suspended	COMPUTE_WH	Medium	min: 1, max: 1	Standard	0	0	10 minutes	Yes	8/22/2020, 12:48:4...	8/24/2020, 1:06:55 ...	SYSADMIN	

LAB : DATA LOAD

Databases > DEMODATABASE > PERSON (PUBLIC)

Tables Views Schemas Stages File Formats Sequences Pipes

Load Table

Column Name	Ordinal ▲	Type	Nullable	Default	Comment
SNO	1	NUMBER(38,0)			
AGE	2	NUMBER(38,0)			
EMPLOYMENT	3	VARCHAR(16777216)			
ZIPCODE	4	NUMBER(38,0)			
EDUCATION	5	VARCHAR(16777216)			
MARTIALSTATUS	6	VARCHAR(16777216)			
PROFESSION	7	VARCHAR(16777216)			
SPOUSE	8	VARCHAR(16777216)			
RACE	9	VARCHAR(16777216)	true	NULL	
GENDER	10	VARCHAR(16777216)	true	NULL	
SPOUSE_AGE	11	NUMBER(38,0)	true	NULL	
LOCATION	12	VARCHAR(16777216)	true	NULL	
SALARY	13	VARCHAR(16777216)	true	NULL	

Load Data

Warehouse Source Files File Format Load Options

Which warehouse do you want to use to load the files?

DEMOWAREHOUSE

Cancel Next

Load Data

Warehouse Source Files File Format Load Options

From where do you want to load files?

☒ Load files from your computer

Select Files...

☐ Load files from external stage

Stage

Path

Cancel Back Next

LAB : DATA LOAD

Create File Format

Name * CSVFormat

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

Show SQL Cancel Finish

Load Data

Warehouse Source Files File Format Load Options

CSVFORMAT

Show SQL Cancel Back Next Load

Load Data

Warehouse Source Files File Format Load Options

What should the load do if it encounters an error while parsing a file?

☒ Do not load any data in the file

☐ Stop loading, rollback and return the error


☐ Do not load any data in the file if the error count exceeds:

Threshold 0

☐ Continue loading valid data from the file

Show SQL Cancel Back Load

Load Results

Loaded	File	Rows Parsed	Rows Loaded
	person.csv	32561	32561

OK

SQL

```
1 PUT file://<file_path>/person.csv @PERSON/ui1598522962697
2
3 COPY INTO "DEMODATABASE"."PUBLIC"."PERSON" FROM @/ui1598522962697 FILE_FORMAT
  = "DEMODATABASE"."PUBLIC"."CSVFORMAT" ON_ERROR = 'CONTINUE' PURGE = TRUE;
```

Select SQL Close

Staging Files...

Encrypting Files

```
*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

s3.console.aws.amazon.com/s3/buckets/snowflake-stage2/?region=ap-south-1&tab=overview

aws Services Resource Groups

Amazon S3 > snowflake-stage2

snowflake-stage2

Overview Properties Permissions Management Access points

Search: Type a prefix and press Enter to search. Press ESC to clear.

Upload Create folder Download Actions

Asia Pacific (Mumbai)

Name	Last modified	Size	Storage class
person.csv	Aug 24, 2020 5:25:04 PM GMT+0530	3.6 MB	Standard

Viewing 1 to 1

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 > 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

The screenshot displays the Snowflake web console interface. The top navigation bar includes icons for Databases, Shares, Data Marketplace, Warehouses, Worksheets, and History. The user is logged in as NISHANTDEMO, SYSADMIN. The main workspace shows a query editor with the following SQL code:

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

The query has been executed successfully, as indicated by the green checkmark and the message "Query_ID SQL 630ms 32,562 rows". The results are displayed in a table with 7 rows and 3 columns: Row, \$1, \$2, and \$3.

Row	\$1	\$2	\$3
1	sno	Age	Employment
2	1	39	State-gov
3	2	50	Self-emp-not-inc
4	3	38	Private
5	4	53	Private
6	5	28	Private
7	6	37	Private



THANK
YOU...!!!