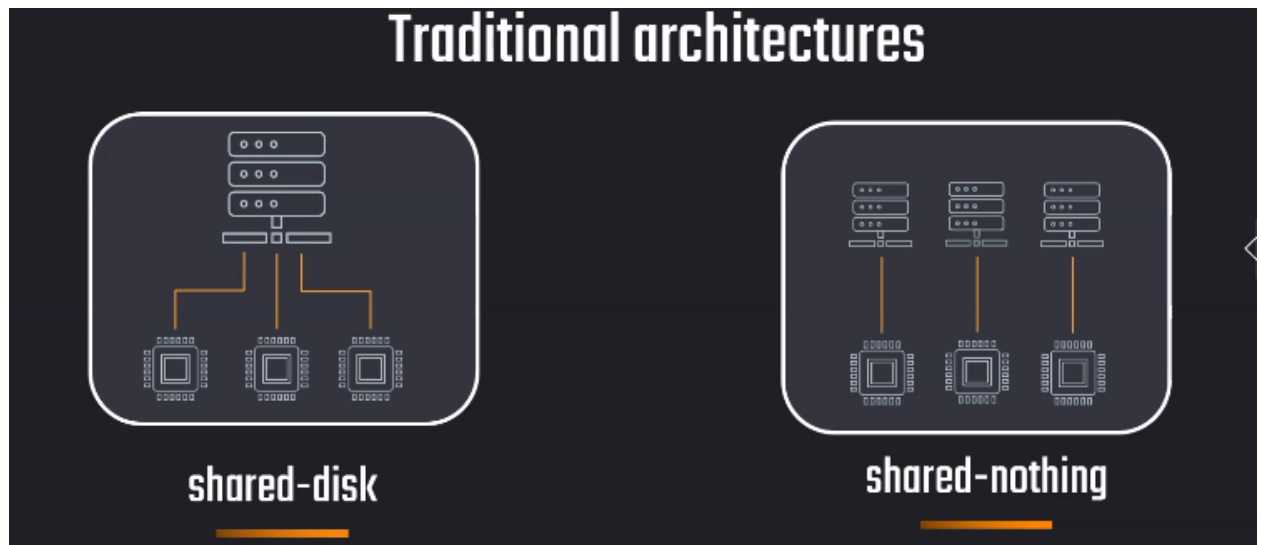
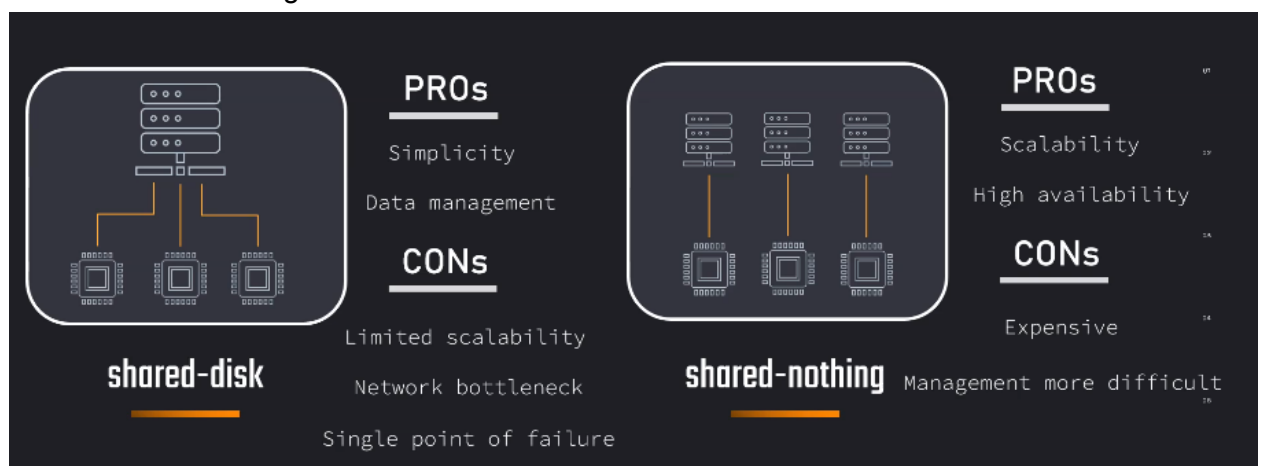


Multi cluster shared-disk

1. In traditional architectures...when there was a need to scaling they used shared disk and shared nothing architectures

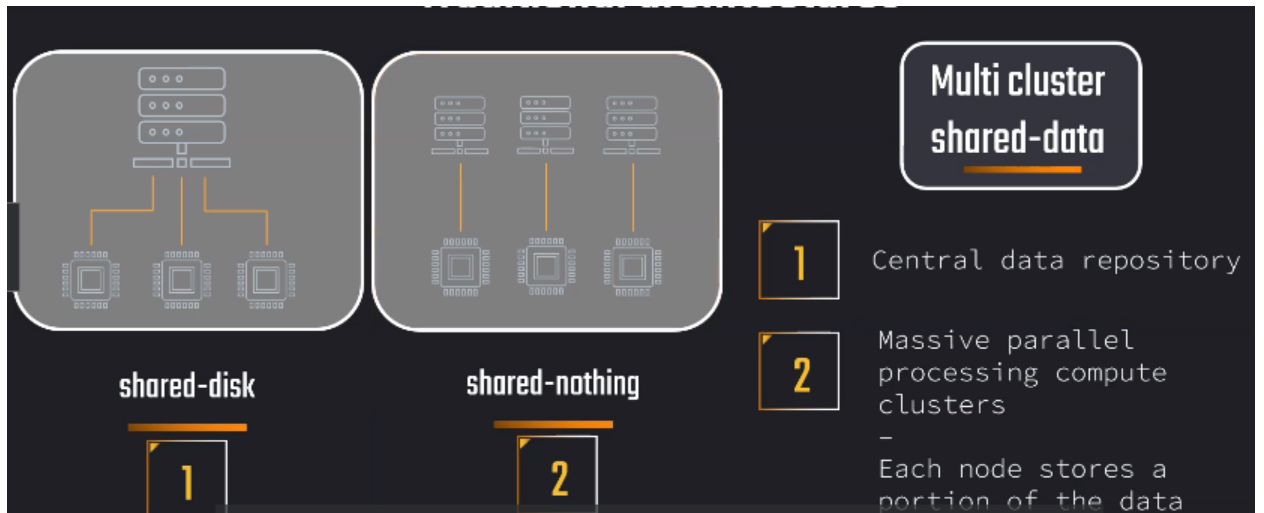


2. Pro and cons of shared disk
3. Pros...it is simple and easy to manage data....and coming to cons... scalability is possible at certain point this is a bit limited because now since all of those nodes need to communicate through this network with the central data storage, this creates a bottleneck and can really decrease the performance
4. And also availability is low...means if our database is corrupted...then it effects the entire disk
5. Pros of shared nothing...

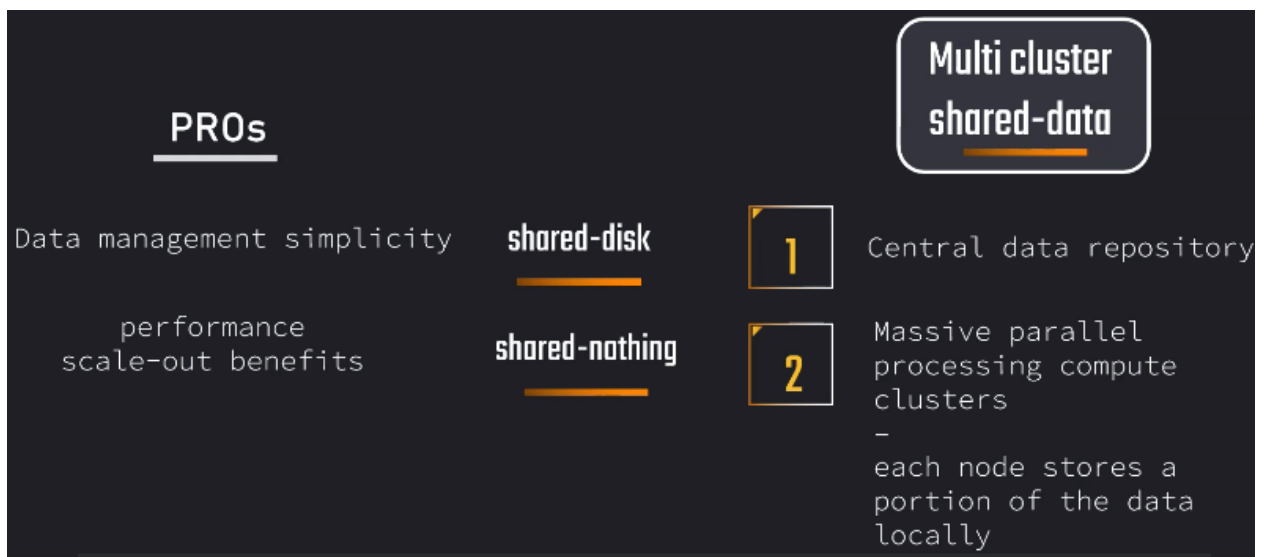


6. Therefore, the snowflake architecture takes the best from both of those two architectures and this is called multi-cluster shared data architecture

7.



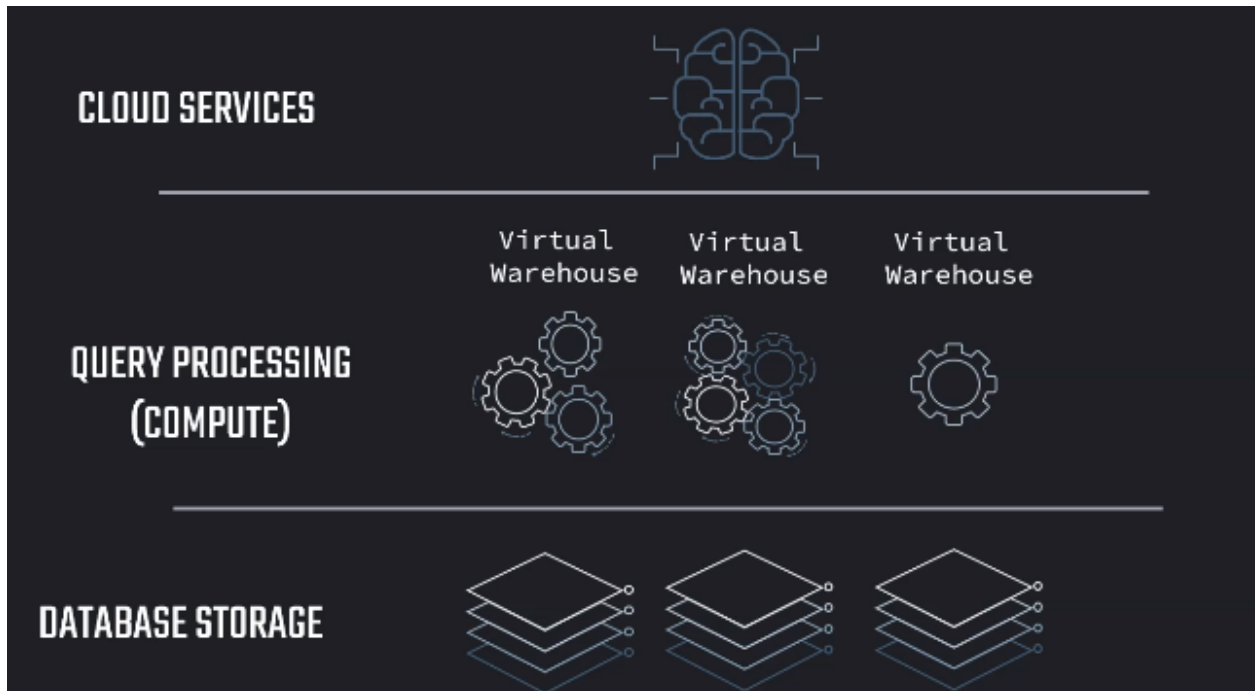
8.



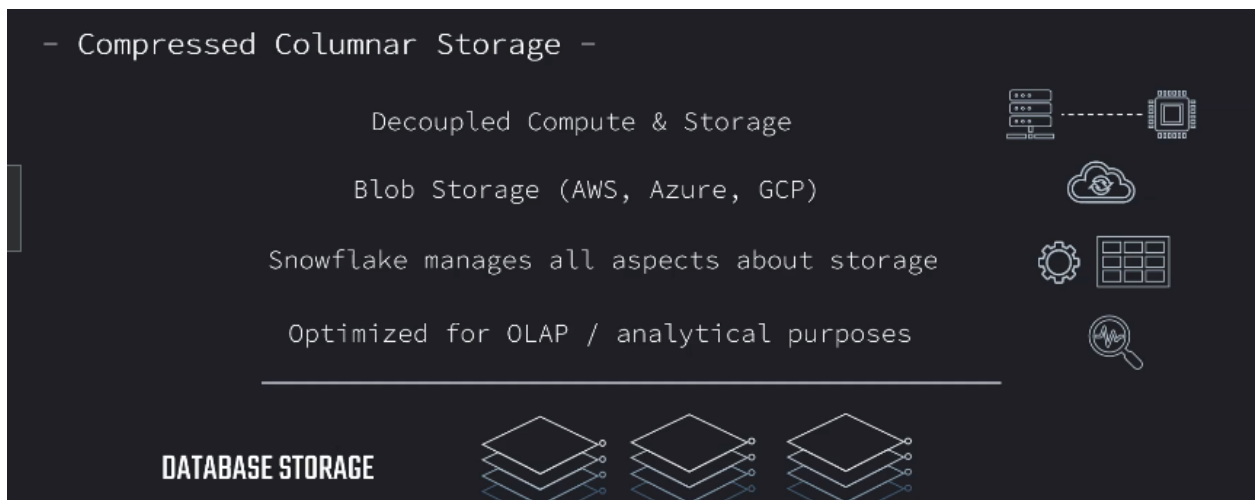
9. This is called multi-cluster shared-data

Three Distinct layers of Snowflake

1. The first layer is database storage.
2. Then we have the second independent layer, which is the compute layer or also called query processing layer. This consists of so-called virtual warehouses. They provide the compute resources to execute our queries.
3. And then lastly, we have the so-called cloud services layer.
4. This is also called the brain of the system, because in here all of the threads come together and all
5. of the different components and activities are managed.

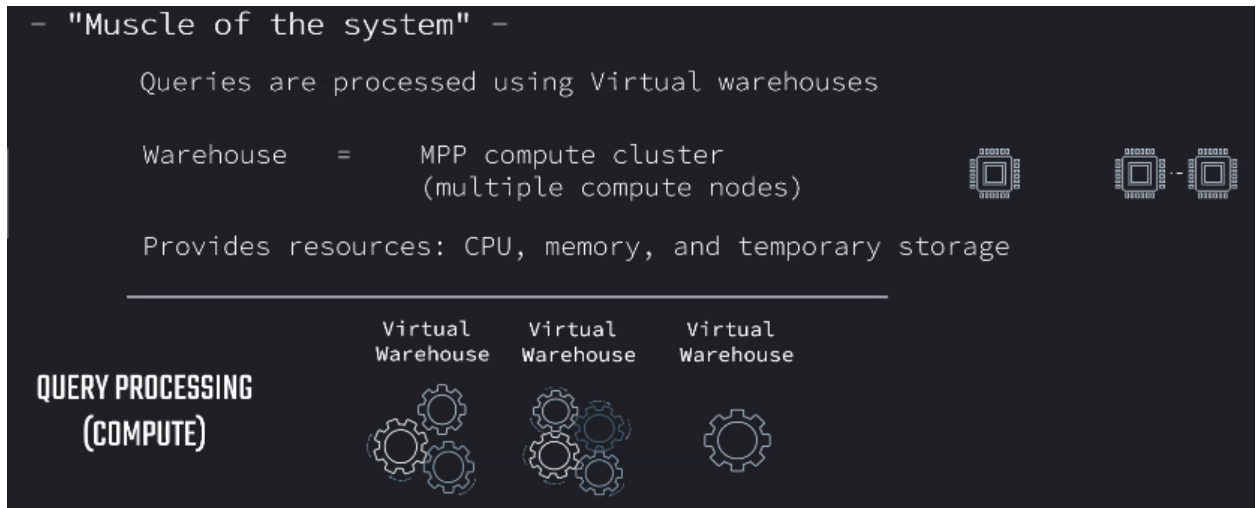


- 6.
7. Lets learn database storage first
8. The database layer is decoupled from compute.. this is a so-called hybrid columnar storage.
9. Here it compresses data into blobs..and these blobs will be stored in external cloud provider...like AWS S3, Azure container etc...we cannot see or access these blobs
10. Also this layer is optimized for making quick analysis and getting data out of database



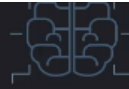
11. Coming to 2nd layer Compute
12. This is a query processing layer and also called as muscle of the system

13. All these queries will get processed in the virtual warehouse..which is a massive parallel processing(MPP) compute cluster...and this clusters can be single node or multiple nodes
14. Again, we use resources from the cloud provider.
15. So if this is AWS, we use so-called EC2 instances... for Azure, we use Azure virtual machines.



16. Coming to last layer...cloud services layer..also called as brain of the system
17. This layer is completely managed by snowflake..
18. But what we need to remember is
19. So for example, when we log in into our account, the authentication is going through this cloud service layer and then we are authenticated.
20. If we have different roles with different permissions, this will also be managed completely by this cloud service
21. This layer also stores the metadata about our queries
22. Basically the query processing is done in compute layer ...but the query parsing and query optimization is done in cloud service layer

CLOUD SERVICES



- "Brain of the system" -

- ✓ Authentication
- ✓ Access control
- ✓ Metadata management
- ✓ Query parsing and optimization
- ✓ Infrastructure management

Collection of services to coordinate & manage the components

Also run on compute instances of cloud provider

23.

Loading data(hands On)

1. First we have created a SQL worksheet and performed commands which are in LoadingData.txt
2. Basically we created database, renamed db...then created the table as per the txt file
3. Next we checked whether our table is empty or not by executing ..this

```
// Check that table is empty
USE DATABASE OUR_FIRST_DB;

SELECT * FROM LOAN_PAYMENT;
```

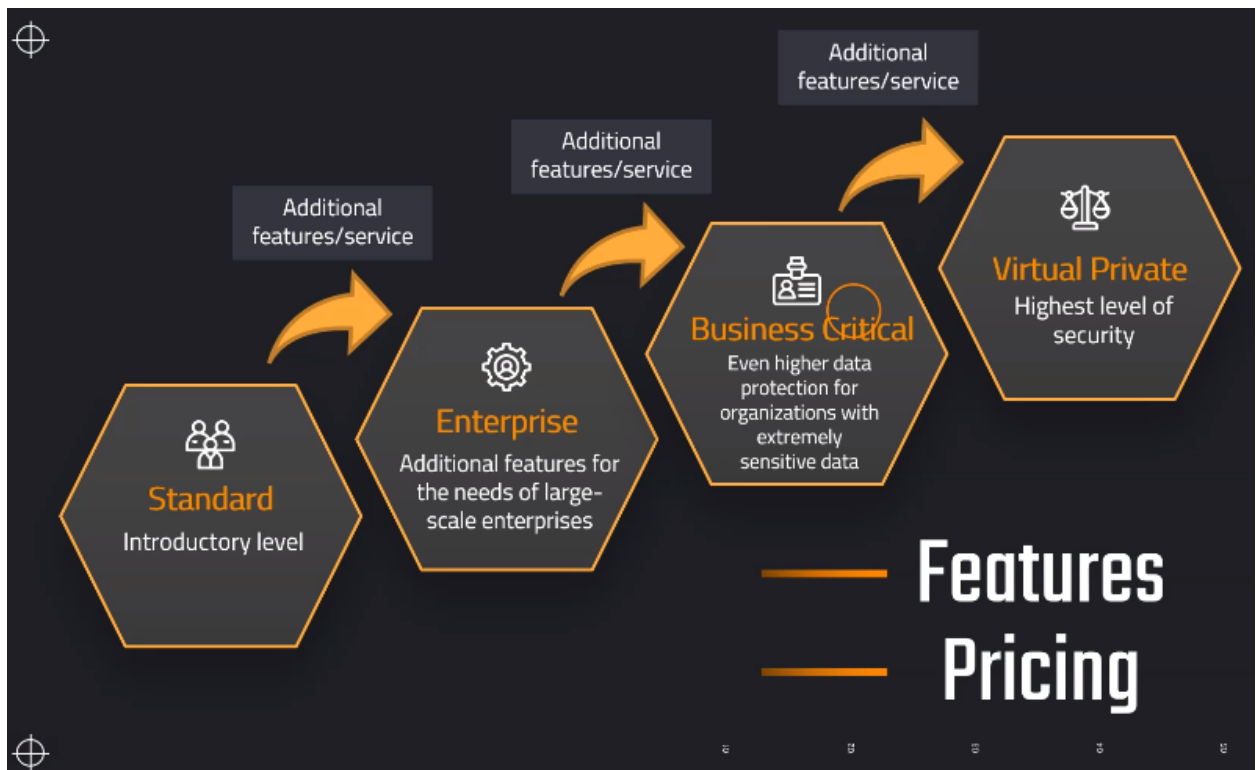
4. Later we loaded data from AWS s3 to our table and validated the data

```
// Loading the data from S3 bucket

COPY INTO LOAN_PAYMENT
FROM s3://bucketsnowflakes3/Loan_payments_data.csv
file_format = (type = csv
               field_delimiter = ','
               skip_header=1);

//Validate
SELECT * FROM LOAN_PAYMENT;
```

Snowflake Editions



- 1.
2. We cannot usually take virtual private edition...for that we have to contact snowflake

Standard	Enterprise	Business Critical	Virtual Private
<ul style="list-style-type: none"> ✓ Complete DWH ✓ Automatic data encryption ✓ Broad support for standard and special data types ✓ Time travel up to 1 day ✓ Disaster recovery for 7 days beyond time travel ✓ Network policies ✓ Secure data share ✓ Federated authentication & SSO 	<ul style="list-style-type: none"> ✓ All Standard features ✓ Multi-cluster warehouse ✓ Time travel up to 90 days ✓ Materialized views ✓ Search Optimization ✓ Column-level security ✓ 24-hour early access to weekly new releases 	<ul style="list-style-type: none"> ✓ All Enterprise features ✓ Additional security features such as customer-managed encryption ✓ Support for data specific regulation ✓ Database failover/failback (disaster recovery) 	<ul style="list-style-type: none"> ✓ All Business Critical features ✓ Dedicated virtual servers and completely separate Snowflake environment ✓ Dedicated metadata store
<ul style="list-style-type: none"> ✓ Premier support 24/7 	003-1048559	1258 003-77156.8	1760 0009-14563.7 73273

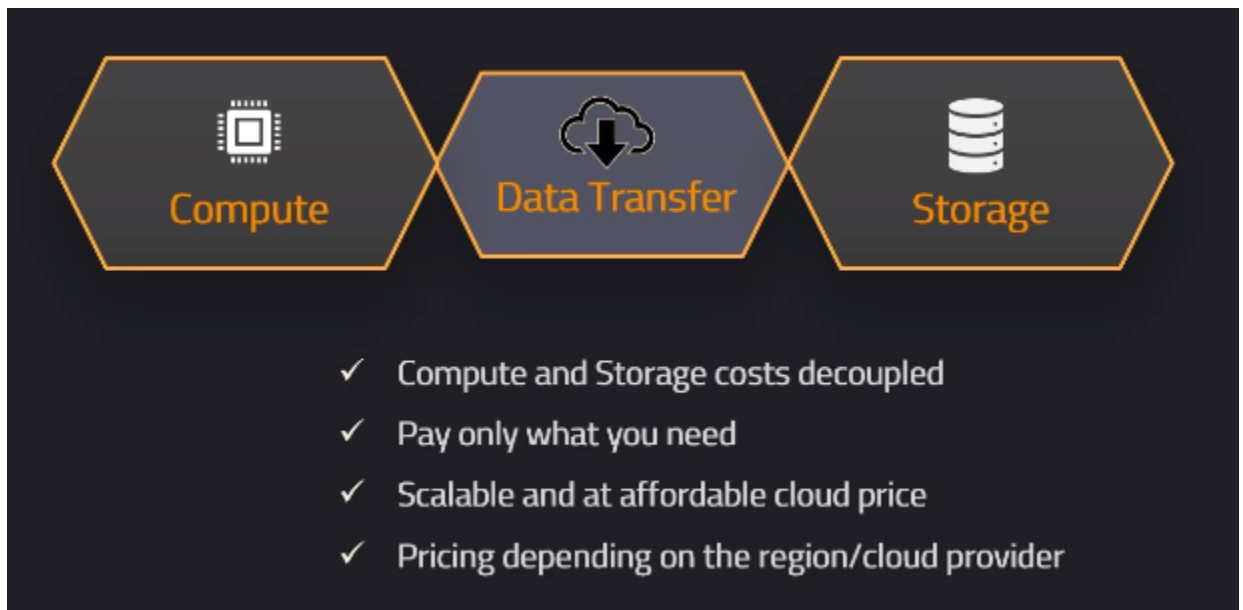
- 3.

Snowflake Time Travel works by keeping a copy of all changes made to your data, including updates, deletions, and object drops. This data is stored for a specified period of time, which is called the data retention period. The default data retention period is 1 day, but it can be extended to up to 90 days for Enterprise and Business-Critical editions.

- 4.
5. Enterprise edition has columnar security...means ..we can hide some columns ..for specific users..etc

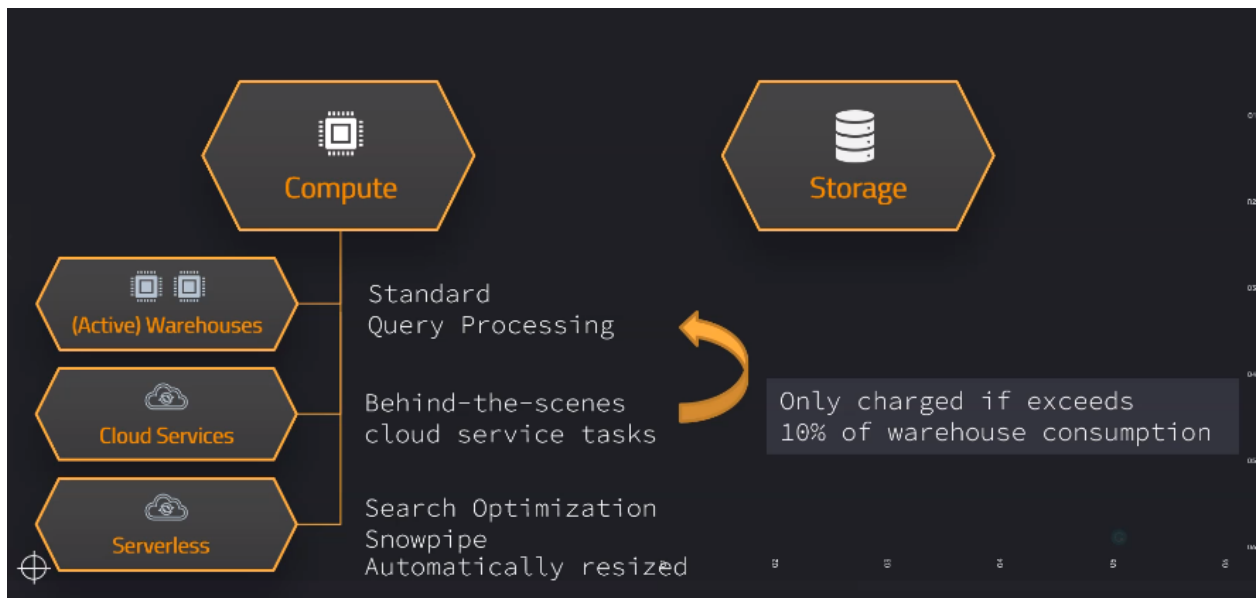
Compute Cost

1. We have two main types of cost. The first block is compute cost, and then we have also storage cost.
2. Main benefit is they are decoupled..which means they can scale independently as per the needs
3. we also have some data transfer costs in some situations.

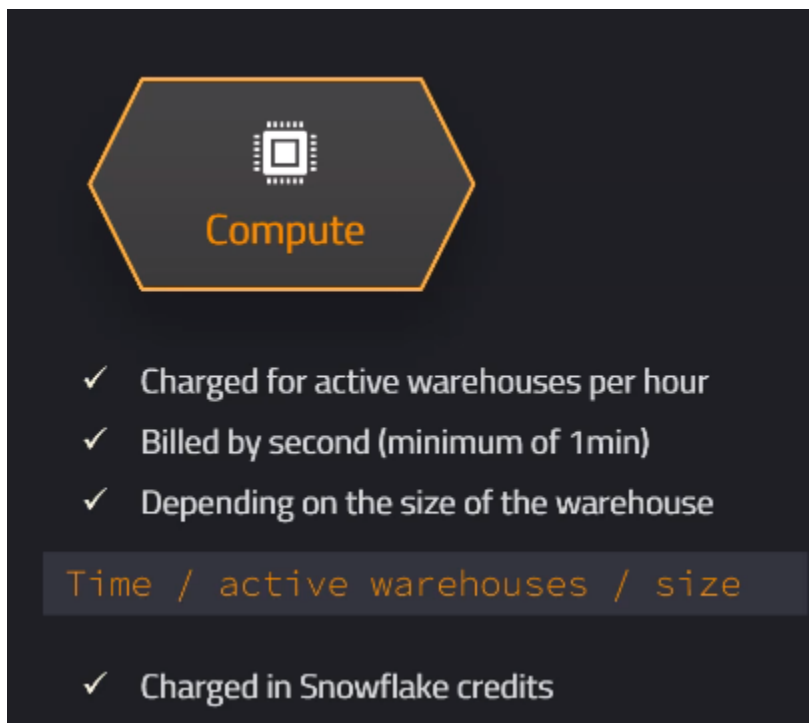


4. Lets check compute costs first
5. So we have again here three different aspects of compute costs and the main one is active warehouses.. we are only paying for warehouses that are active.
6. If we suspend them...then we stop paying for them
7. Second is cloud services..
8. The last is serverless...These are also some warehouses that are needed for some specific features like automatic reclustering,search optimization and

snowpipe.

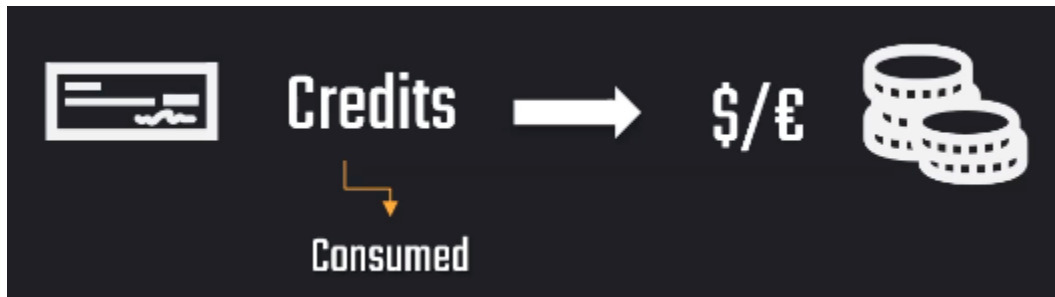


9. Lets deep dive into active warehouses



10.

11. We don't get charged directly in ...we get charged on snowflake credits...from these credits...USD will get charged



12. So basically if we have xs(xtra small) which has only 1 node...it consumes 1 credit

13. Like was for M(medium) 4 nodes ..then 4 credits consumed



The number of Snowflake credits consumed by a particular activity depends on the type of activity and the resources used. For example, running a Large virtual warehouse will consume more credits than running a Small virtual warehouse.

Snowflake charges for credits in one-second increments, with a minimum charge of 60 seconds per virtual warehouse start. This means that you only pay for the time that you are actually using Snowflake resources.

Example:

Suppose you have a Small virtual warehouse that you use to run queries and load data. You start the virtual warehouse at 10:00 AM and run it for 2 hours. At 12:00 PM, you stop the virtual warehouse.

During the 2 hours that the virtual warehouse was running, you consumed 2 Snowflake credits. This is because a Small virtual warehouse consumes 1 Snowflake credit per hour.

Snowflake will bill you for the 2 Snowflake credits that you consumed. The billing will occur at the end of the month.

14. For different regions we have diff prices per credit

Service Level	Price per Credit
Standard	✓ \$2 / Credit
Enterprise	✓ \$3 / Credit
Business Critical	✓ \$4 / Credit
Virtual Private	✓ Contact Snowflake

Region: US East (Ohio)
Platform: AWS

15.

Service Level	Price per Credit
Standard	✓ \$2.70 / Credit
Enterprise	✓ \$4 / Credit
Business Critical	✓ \$5.40 / Credit
Virtual Private	✓ Contact Snowflake

Region: EU (Frankfurt)
Platform: AWS

Storage and Data Transfer costs

1. Basically the storage costs depends upon cloud provider and region

\$40
per TB / per month

Region: US East (Ohio)
Platform: AWS

Storage

- ✓ Monthly storage fees
- ✓ Based on average storage used per month
- ✓ Cloud Providers
- ✓ Cost calculated after compression

2. We have 2 scenarios...on demand and capacity storage
3. In on-demand..we pay as we use....but in capacity..we have to pay upfront

On Demand Storage

✓ We think we need 1 TB of storage

❖ Scenario 1: 100GB of storage used
 $0.1 \text{ TB} \times \$40 = \4

❖ Scenario 2: 800GB of storage used
 $0.8 \text{ TB} \times \$40 = \32

Capacity Storage

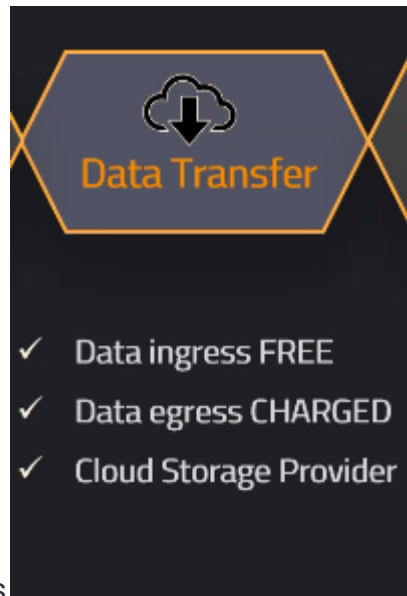
❖ Scenario 1: 100GB of storage used
 $1 \text{ TB} \times \$23 = \23

❖ Scenario 2: 800GB of storage used
 $0.8 \text{ TB} \times \$40 = \23

Region: US East (Northern Virginia)
Platform: AWS

003-1040559 1258 003-77 56.8 1760 0808-14563.7 73273

- 4.



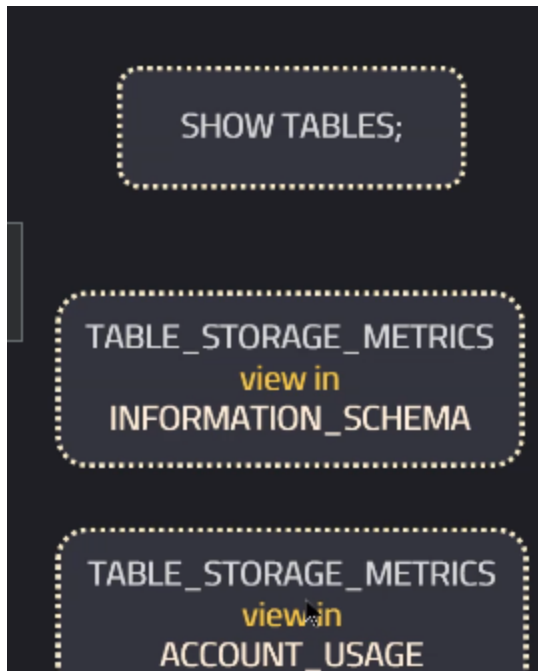
5. Data transfer charges

Cloud	Region where data is stored	Transfer to same region, same cloud (per TB)	Transfer to different region, same cloud (per TB)	Transfer to different cloud (per TB)
AWS	US East (Northern Virginia)	\$0.00	\$20.00	\$90.00
AWS	US East (Ohio)	\$0.00	\$20.00	\$90.00
AWS	US West (Oregon)	\$0.00	\$20.00	\$90.00
AWS	Canada (Central)	\$0.00	\$20.00	\$90.00
AWS	EU (Ireland)	\$0.00	\$20.00	\$90.00
AWS	EU (London)	\$0.00	\$20.00	\$90.00
AWS	EU (Frankfurt)	\$0.00	\$20.00	\$90.00
AWS	EU (Paris)	\$0.00	\$20.00	\$90.00
AWS	EU (Stockholm)	\$0.00	\$20.00	\$90.00
AWS	Asia Pacific (Sydney)	\$0.00	\$140.00	\$140.00
AWS	Asia Pacific (Singapore)	\$0.00	\$90.00	\$120.00
AWS	Asia Pacific (Osaka)	\$0.00	\$90.00	\$114.00

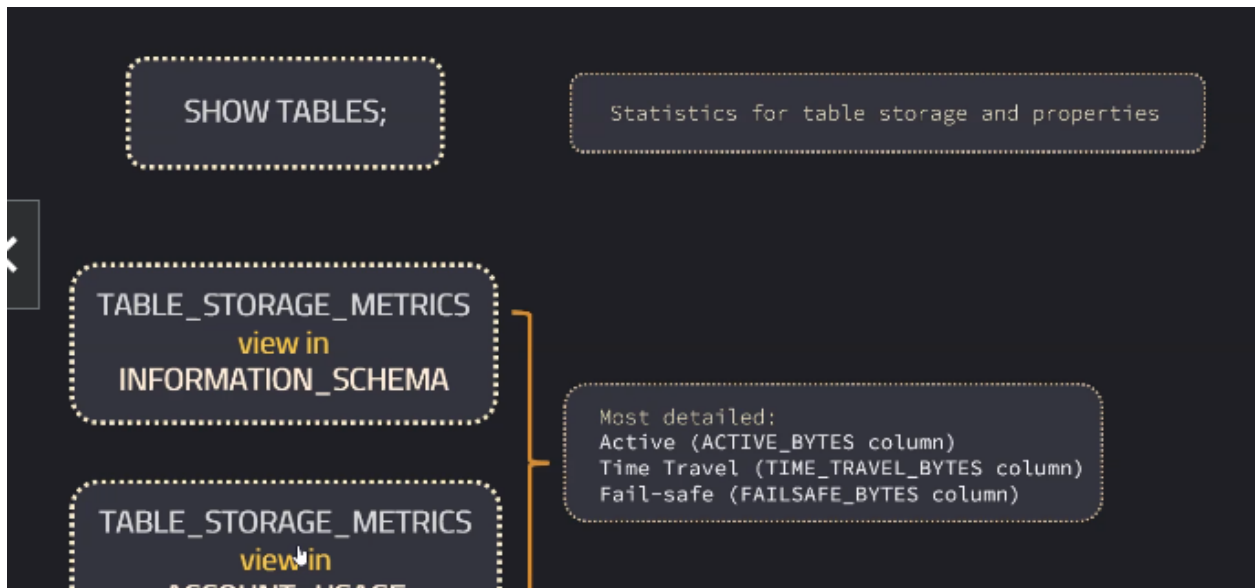
6.

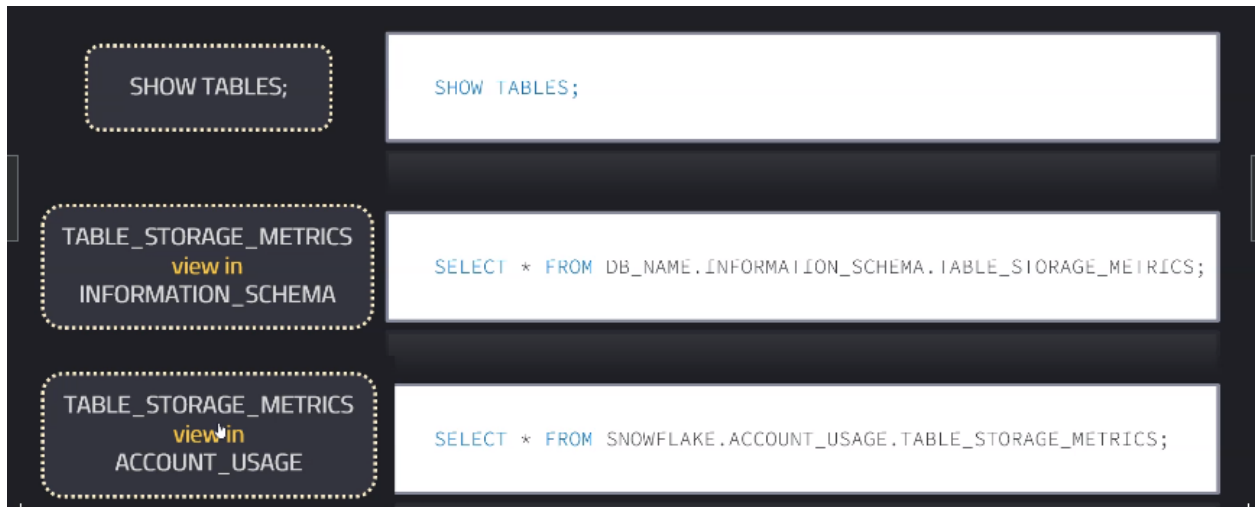
Storage Monitoring

1. Individual table storage...just checking how many tables are in our account



- 2.
3. Show table give statistics





- 4.
5. So we need to use account admin role for querying table storage metrics..

```
USE ROLE ACCOUNTADMIN;
-- Most detailed information:
SELECT * FROM SNOWFLAKE.ACCOUNT_USAGE.TABLE_STORAGE_METRICS;
```

6. And now if we run the show Tables command, we see information about the table properties like clustering search optimization and then also about the amount of bytes used. `SHOW TABLES;`



7. We can also check usage in

Resource Monitors

1. Those are objects that can be used to control and monitor the credit usage of both warehouses and our entire account.

2. In the resource monitor we can set credit quota

Resource Monitors

Control and monitor credit usage of warehouses and account

Standard Edition

Credit Quota **Set Credit Limits**

Schedule Start Monitoring **Immediately** End Monitoring **Never** Resets **Monthly** **Customize** **In defined cycle**

3. We can also set the resource monitor for a group of data warehouses

Virtual warehouse

Account

Multiple virtual warehouses

Monitor Type

Schedule

Warehouse

Schedule ✓
 ✓

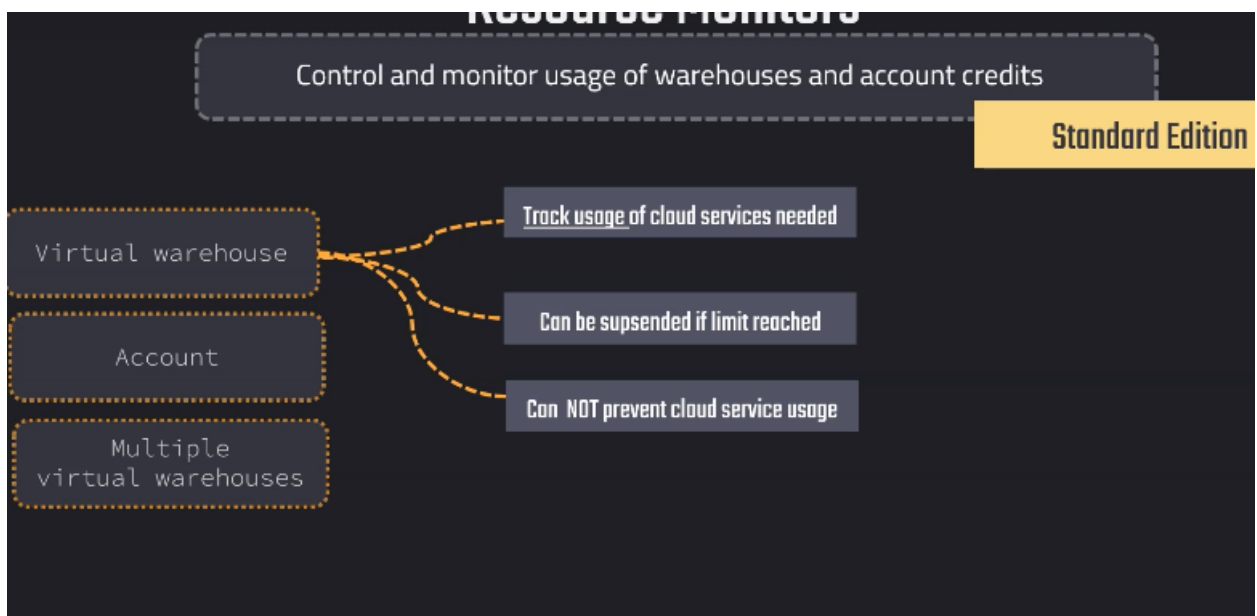
4. And then based on the quota that we have set, we can define specific actions....

5. We have 3 types of actions..

The screenshot shows a configuration window titled 'Actions' with the subtitle 'Specify an action to perform when the quota is reached.' and an 'Add' button. It lists three actions, each with a percentage input field and a description:

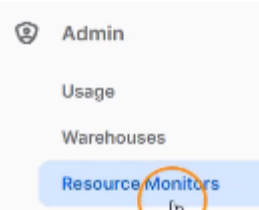
Action	Percentage	Description
Suspend immediately and notify	%	Suspend immediately and notify when this % of credit is used. ?
Suspend and notify	%	Suspend and notify when this % of credit is used. ?
Notify	%	Notify when this % of credit is used. ?

6. In suspend and notify...it completes the task that is started ..and if it exceeds the limit...it stop after completing this task
7. Where as suspend immediately will stop ASAP if it reaches the specified limit..and it will stop..ongoing tasks
8. This can only be created by users with the account admin role.
9. Also account admin can grant privileges of monitor and modify some resource monitors



Hands On resource monitoring

1. To Setup a resource monitor we use account admin role...as it has full privileges



2. Then go to resource monitor section under admin
3. Then press on resource monitor

A screenshot of a web application showing the 'Resource Monitors' section. A modal window titled 'New Resource Monitor' is open, showing a form to create a new monitor. The form includes fields for 'Name', 'Credit Quota', 'Monitor Type', 'Schedule', and 'Actions'. The 'Name' field contains the text 'Name of resource monitor'. The 'Credit Quota' field contains the text 'Number of credits allowed per cycle'. The 'Monitor Type' field is a dropdown menu with the text 'Select Monitor Type'. The 'Schedule' field has three radio buttons: 'Start Monitoring Immediately', 'End Monitoring Never', and 'Resets Monthly'. The 'Actions' field has a text input with the placeholder 'Specify an action to perform when the quota is reached.' and an 'Add' button. Below the 'Actions' field, there are two rows of input fields for percentages, each with a placeholder text: 'Suspend immediately and notify when this % of credit is used.' and 'Suspend and notify when this % of credit is'. The background shows a list of resource monitors with filters for 'Warehouses All' and 'Frequency All'.

4. Then we choose the name and set credit quota..etc..

5. and also we can choose monitor type..if we choose warehouses...then this monitoring will be applied to that warehouse..here we have chosen account

New Resource Monitor
Creating as ACCOUNTADMIN

Name: Monthly_Account_Budget

Credit Quota: 50

Monitor Type: Account

Schedule: Start Monitoring Immediately, End Monitoring Never, Resets Monthly

Actions: Specify an action to perform when the quota is reached.

Buttons: Cancel, Create Resource Monitor

Monitor Type: Account

Schedule: Start Monitoring Immediately, End Monitoring Never, Resets Monthly

Actions: Specify an action to perform when the quota is reached.

Buttons: Cancel, Create Resource Monitor

Selected Action: 105 Suspend immediately and notify when this % of credit is used.

6. then we have to select at what % our services

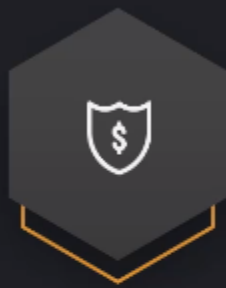
Warehouses and multi clustering

What is a virtual warehouse?



Provides compute
resources to
execute queries
and operations

1. What is a virtual warehouse?
2. We have 2 diff types of warehouse



Standard

Most suitable in
most use cases



Snowpark-optimized

Recommended for
memory-intensive
workloads such as
ML training

3. The snowpark optimized warehouse sizes starts from M(medium)

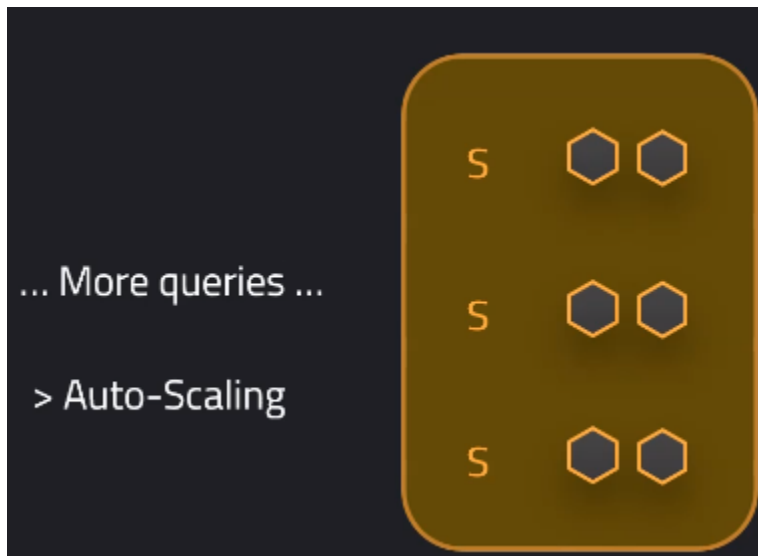
M  **6**

L  **12**

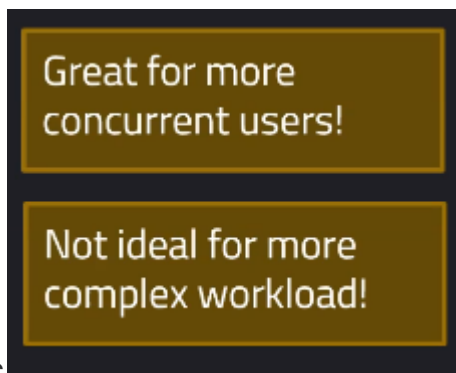
XL  **24**

6XL **768**

4. We have multicluster warehouse...
5. if we have more queries to execute in one machine..then it will become too slow and we need to wait for results
6. To avoid this we use multi cluster warehouse..snowflake autoscales clusters in the warehouse..it reaches a limit

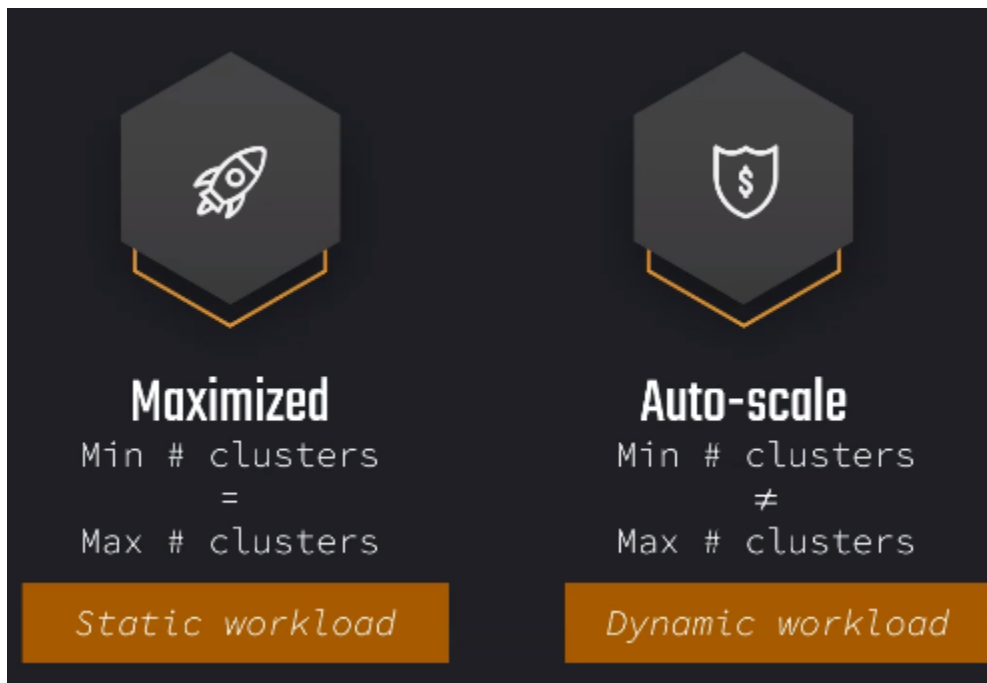


7. It is great for concurrent users...but not good for complex workload like Machine

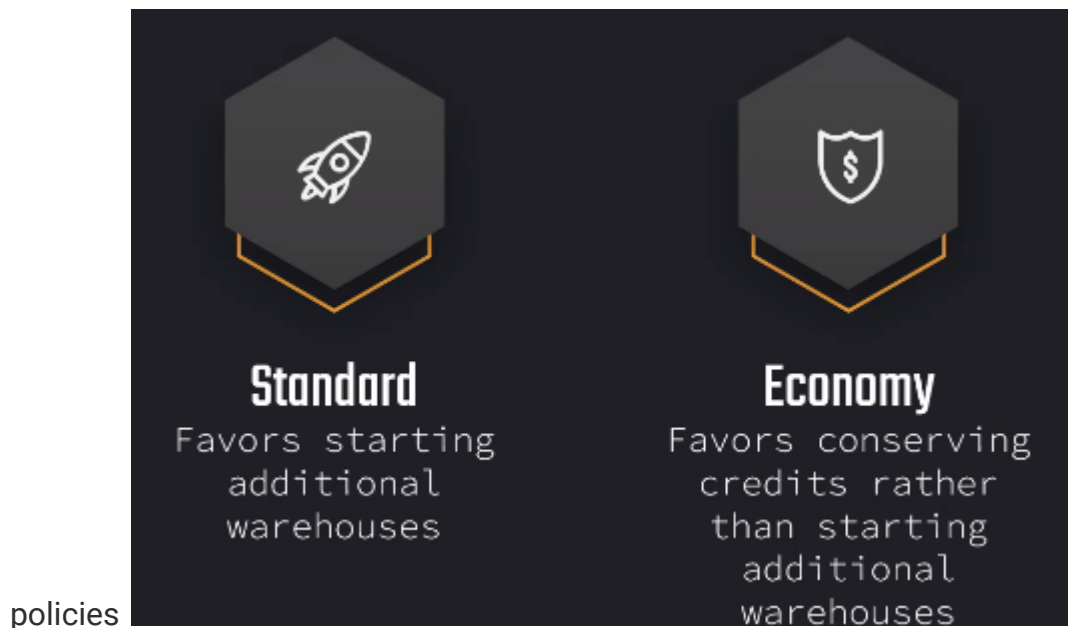


learning works etc

8. In multi cluster warehouse...we have 2 modes



9. When does auto scaling starts?..it depends upon the scaling policy..so we have 2



policies

10. Detailed look at scaling policies

Policy	Description	Cluster Starts...	Cluster Shuts Down...
Standard (default)	Prevents/minimizes queuing by <u>favoring starting additional clusters</u> over conserving credits.	<u>Immediately</u> when either a query is <u>queued</u> or the system detects that there are more queries than can be executed by the currently available clusters.	<u>After 2 to 3 consecutive successful checks</u> (performed at 1 minute intervals), which determine whether the load on the least-loaded cluster could be redistributed to the other clusters
Economy	Conserves credits by favoring keeping running clusters fully-loaded rather than starting additional clusters. <i><u>Result</u></i> May result in queries being queued and <u>taking longer to complete.</u>	Only if the system estimates there's enough query load to keep the cluster <u>busy for at least 6 minutes.</u>	<u>After 5 to 6 consecutive successful checks</u> -

Warehouses Hands on

1. To create a warehouse ..go to admin section and select warehouses...
2. Next click +warehouse..then give name,specify size of the warehouse

The screenshot shows the 'Warehouses' management page with a 'New Warehouse' modal open. The modal is titled 'New Warehouse' and shows 'Creating as ACCOUNTADMIN'. It has two input fields: 'Name' with the value 'FIRST_WH' and 'Size' with a dropdown menu showing 'X-Small 1 credit/hour'. Below these is a 'Comment (optional)' text area containing 'T'. There are two toggle switches: 'Query Acceleration' (off) and 'Multi-cluster Warehouse' (off). At the bottom, there is a section for 'Advanced Warehouse Options' with a dropdown arrow, and two buttons: 'Cancel' and 'Create Warehouse'.

NAME ↑	COMPUTE
1 Warehouse	

Warehouses

+ Warehouse

QAS ... OWN... CREAT...

Disab... S... 4 days ...

New Warehouse

Creating as ACCOUNTADMIN

Name: FIRST_WH

Size: X-Small 1 credit/hour

Comment (optional): T

Query Acceleration: ☐

Multi-cluster Warehouse: ☐

Advanced Warehouse Options

Cancel Create Warehouse

3. Why we use query acceleration? In our case we dont need it

There are a few reasons why we use query acceleration in Snowflake warehouse:

- To improve the performance of queries that scan large tables. Query acceleration can significantly improve the performance of queries that scan large tables, especially if the query is also filtering or aggregating the data.
- To reduce the impact of outlier queries on overall warehouse performance. Query acceleration can help to reduce the impact of outlier queries on overall warehouse performance by offloading these queries to dedicated compute resources.
- To improve the cost-effectiveness of running a mixed workload. Query acceleration can be more cost-effective than scaling up to a larger warehouse for workloads that include a mix of queries, some of which require more compute resources than others.

Here are some specific examples of when you might want to use query acceleration:

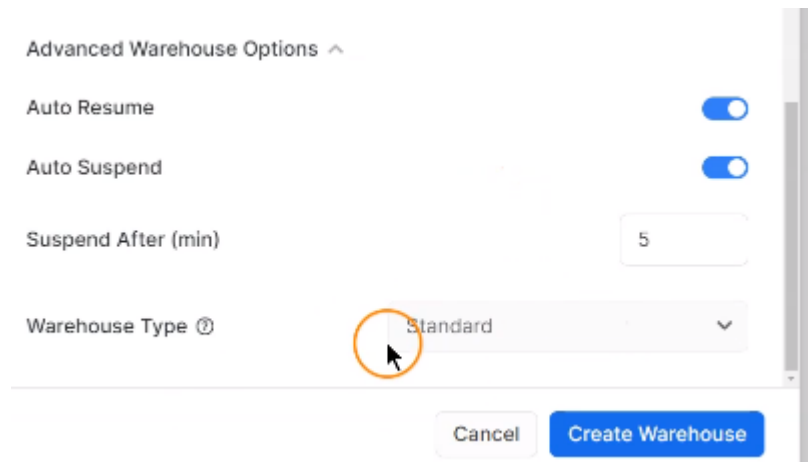
- You have a query that scans a large table and takes a long time to run.
- You have a workload that includes a mix of queries, some of which are more resource-intensive than others.
- You are experiencing performance problems due to outlier queries.
- You want to reduce the cost of running your workload.

4. Next we can setup multi-scaling...while creating it..we also need to select scaling

The image displays two screenshots of the Snowflake warehouse configuration interface. The top screenshot shows the 'Multi-cluster Warehouse' settings. The 'Mode' is set to 'Auto-scale', 'Min Clusters' is set to '1', and 'Max Clusters' is set to '1'. The 'Scaling Policy' is set to 'Standard'. The 'Create Warehouse' button is highlighted. The bottom screenshot shows the 'Scaling Policy' dropdown menu with 'Standard' selected. The 'Advanced Warehouse Options' section is expanded, showing 'Auto Resume' and 'Auto Suspend' options. The 'Create Warehouse' button is also highlighted in this screenshot.

5. We have advanced options like..auto resume and auto suspend..

6. we can choose to suspend our warehouse..If we stay idle for like 5 min..and auto



resume again

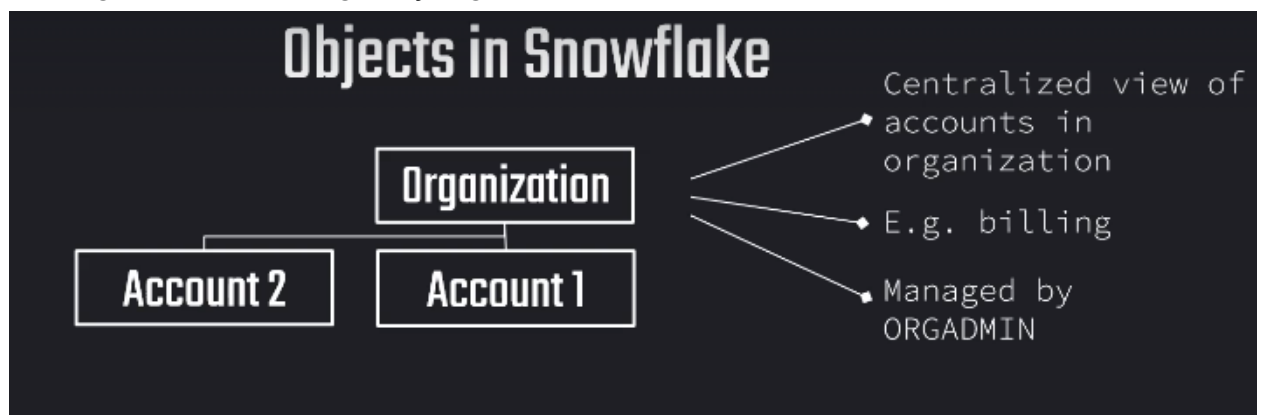
7. We can also create warehouse using SQL commands

```
CREATE WAREHOUSE SECOND_WH
WITH
WAREHOUSE_SIZE = XSMALL
MIN_CLUSTER_COUNT = 1
MAX_CLUSTER_COUNT = 2
AUTO_RESUME = TRUE
AUTO_SUSPEND = 300
COMMENT = 'This is our second warehouse'
```

8. We can drop warehouse using drop command | `DROP WAREHOUSE SECOND_WH;`

Snowflake objects

1. We have a new feature in snowflake..that is “setting up multiple account under one org” it will be managed by orgadmin user



- To create an account..we go to admin section and click on Accounts

The screenshot shows the Snowflake 'Accounts' management interface. At the top, there's a '+ Account' button. Below it, a table lists existing accounts. The 'Create New Account' modal is open, showing options for Cloud (Amazon Web Services), Region (US East (Ohio)), and Edition (Standard - \$2 Credit, \$40 TB). The modal also includes a 'Cancel' button and a 'Next' button.

ACCOUNT	EDITION	CLOUD	REGION	CREATED	LOCATOR
OY93219	Enterprise	AWS	US East (Ohio)	1 day ago	OH34131

Create New Account

Each account in your organization will have its own set of users, roles, databases, and warehouses.

Cloud: Amazon Web Services

Region: US East (Ohio)

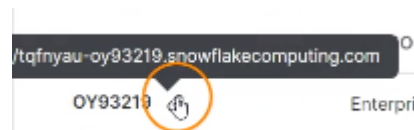
Edition: Standard - \$2 Credit, \$40 TB

You will be billed for usage in all of your accounts on a single bill.

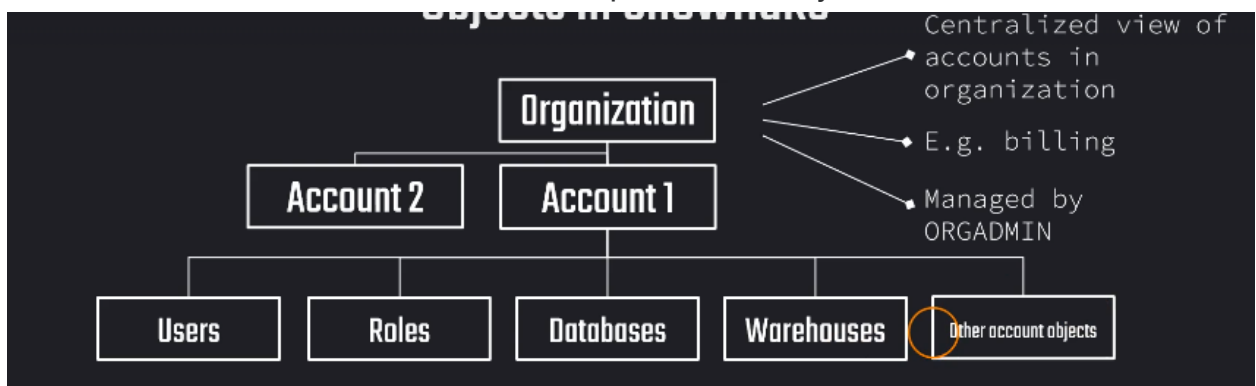
The current rates displayed are standard list prices and may not reflect any discounts you may have. Learn more about [Snowflake editions](#).

Cancel Next

- The pricing will be based on cloud provider and region



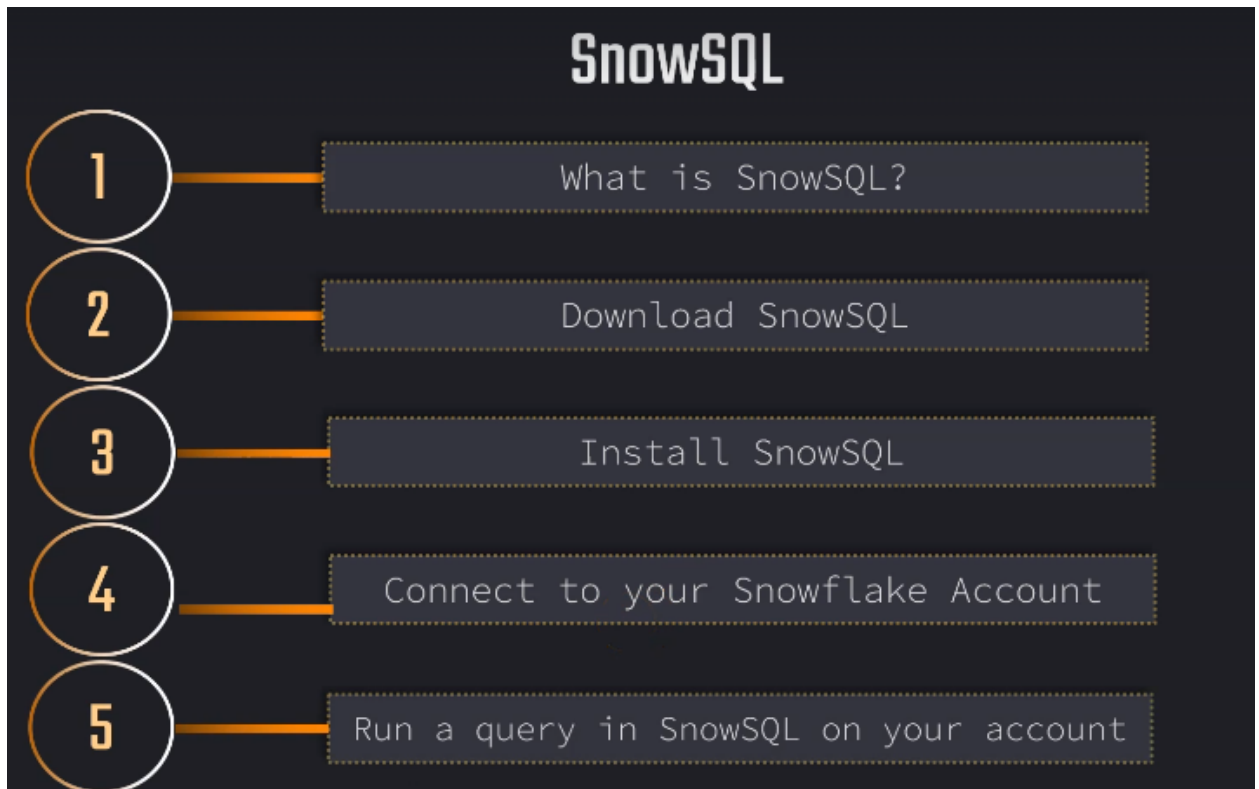
- To access this account we use this link
- And now under one account, we have multiple account objects.

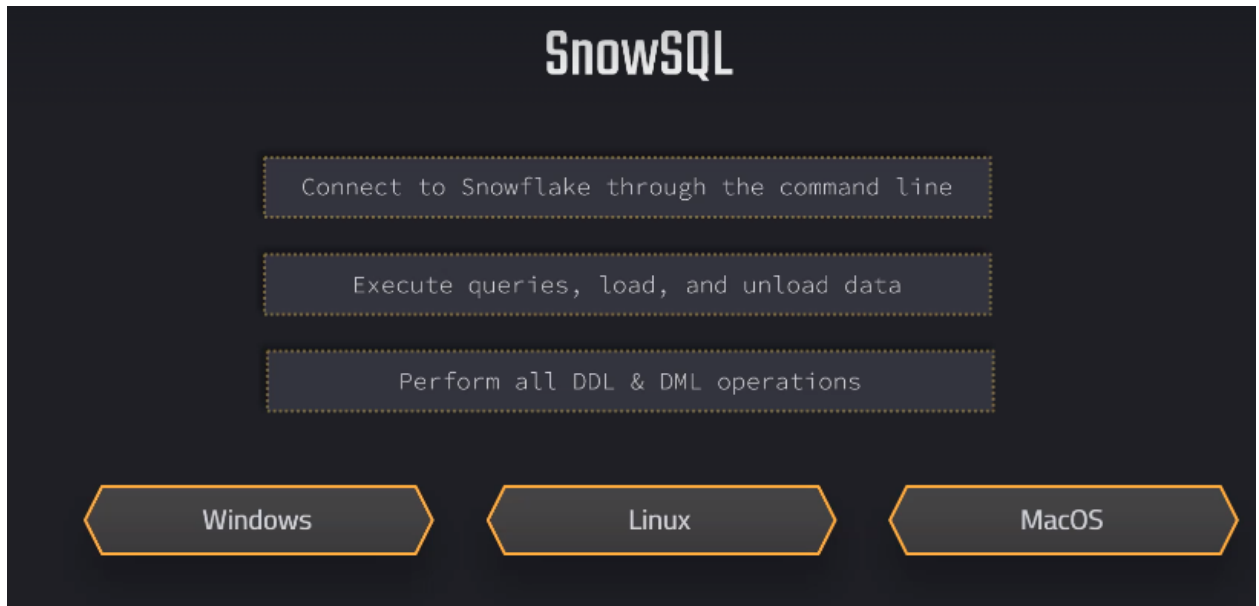


-

7. And this databases have schemas to organize a database..these schemas
..contains table,views etc
8. So these are all snowflake objects

SnowSQL





- 2.
3. First we have to install SNOWsql for windows
4. Then just follow instructions

```
C:\Users\iamka>snowsql -a al49335.us-east-2.aws -u KAUSHI
Password:
* SnowSQL * v1.2.28
Type SQL statements or !help
KAUSHI#COMPUTE_WH@(no database).(no schema)>
```

- 5.

```
User: NIKOLAI
Password:
* SnowSQL * v1.2.24
Type SQL statements or !help
NIKOLAI#COMPUTE_WH@(no database).(no schema)>USE DATABASE SNOWFLAKE_SAMPLE_DATA
+-----+
| status |
+-----+
| Statement executed successfully. |
+-----+
1 Row(s) produced. Time Elapsed: 0.198s
NIKOLAI#COMPUTE_WH@SNOWFLAKE_SAMPLE_DATA.(no schema)>USE SCHEMA TPCH_SF1
+-----+
| status |
+-----+
| Statement executed successfully. |
+-----+
1 Row(s) produced. Time Elapsed: 0.743s
NIKOLAI#COMPUTE_WH@SNOWFLAKE_SAMPLE_DATA.TPCH_SF1>
```

6.