**Virtual Warehouse overview**

To run anything we need compute resource which is nothing but a Virtual Warehouse.

It would be the first thing we have to set up once we set up snowflake account

**Virtual warehouse = cluster of compute nodes (resources)/machines**

Cluster => group of nodes/machines connected together

Compute Resources => CPU + Memory+ Temporary storage

Machine/computer/node

**Snowflake Warehouse comes in various T-Shirt sizes**

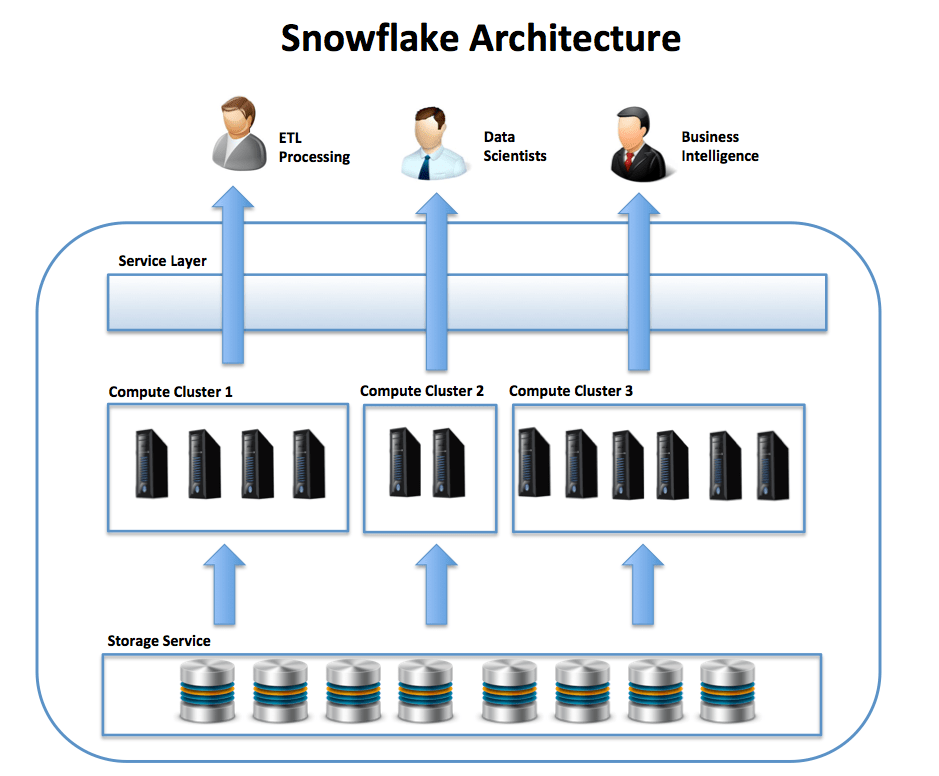


XS(1) will have one compute nodes/machines, S(2) will have 2 compute nodes, M(4), L(8), XL(16), 2XL(32), 3XL(64), 4XL(128) -- generally available

5XL(256) and 6 XL(512) are in preview state

MAXIMUM CLUSTER AS 10 => 128x10 =1280

If we set 6XL warehouse to Multicluster Warehouse=> 512x10 =5120 computes nodes



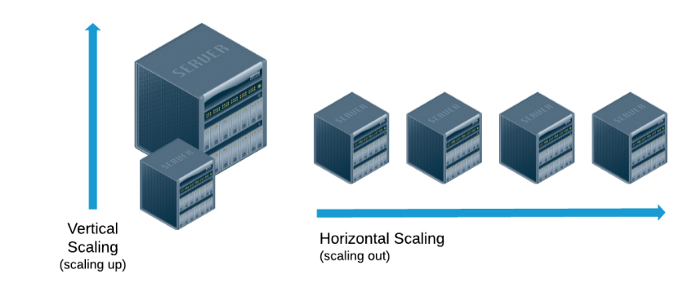
You can start & stop the warehouse anytime as Snowflake storage & computing are de coupled (separate and independent)

What are Auto-suspension and Auto-resumption?

A warehouse can be set to automatically resume or suspend, based on activity:

* By default, auto-suspend is enabled. Snowflake automatically suspends the warehouse if it is IDLE for the specified period of time 10 minutes
* By default, auto-resume is enabled. Snowflake automatically resumes the warehouse when any statement that requires a warehouse is submitted
* Auto-suspend and auto-resume apply only to the entire warehouse and not to the individual clusters in the warehouse.

**Horizontal and Vertical scaling in Snowflake**



**Vertical Scaling (Resizing the warehouse)**

Scale up => **increasing the warehouse size** example from small => large, M to XL

**Purpose** is to **increase performance** of complex queries

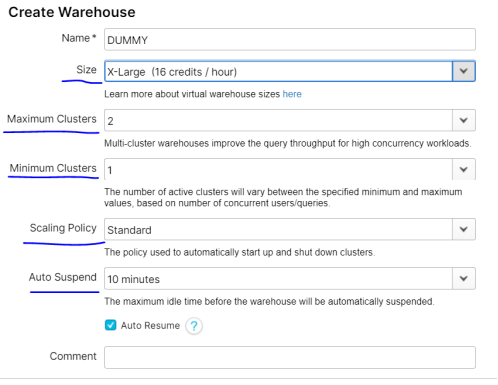
An increase in T-Shirt size (XS⇾XL) => increase in CPU, Memory, Temporary Storage.

Scale down => **reducing the warehouse size** example from large => small

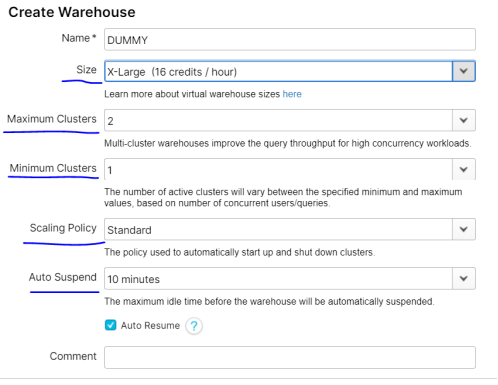
Scale up and down (Re-sizing warehouse) is manual process

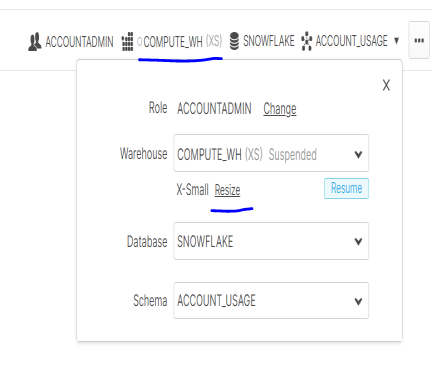
We can modify warehouse-size on the fly even if it is in a **running state**, provided the new size change will only be applicable for newly queued queries and all existing queries will still use old warehouse.

At the time of creation, you can provide the size, multi-cluster attribute (Enterprise & above) & Scaling policy



You can modify warehouse size from the Context menu and also modify the min & max cluster under warehouse>config.





**How to decide on optimum Snowflake Warehouse Size**

Creating multiple warehouses in Snowflake is free, you can create as much as you can, and you will be charged for what you use only.

For each department we can create a warehouse with different size based on workloads

1. Development Team
2. ETL Team
3. Test Team
4. Reporting Team



**MULTI CLUSTER WAREHOUSE**

To create a multi-cluster virtual warehouse, you must have the Enterprise or higher edition of Snowflake.

Lower editions will through error stating 'MULTI\_CLUSTER\_WAREHOUSES' not enabled.

Multi Cluster Warehouse=> Max cluster count>1

User should have the SYSADMIN role or higher

Since the SYSADMIN (or the even higher ACCOUNTADMIN) role can create a virtual warehouse.

1. Switch role to the SYSADMIN role (or higher):

USE ROLE SYSADMIN;

1. Create a virtual warehouse that can auto-scale between one and three clusters, depending on the demand:

CREATE WAREHOUSE ETL\_WH

WAREHOUSE\_SIZE = XSMALL

MAX\_CLUSTER\_COUNT = 3

MIN\_CLUSTER\_COUNT = 1

SCALING\_POLICY = ECONOMY

AUTO\_SUSPEND = 300 -- suspend after 5 minutes (300 seconds) of inactivity

AUTO\_RESUME = TRUE

INITIALLY\_SUSPENDED = TRUE

COMMENT = 'Virtual warehouse for ETL workloads. Auto scales between 1 and on the workload'

**How it works…**

Setting MIN\_CLUSTER\_COUNT and MAX\_CLUSTER\_COUNT to different values ensures that the multi cluster virtual warehouse will start with a number of clusters equal to MIN\_CLUSTER\_COUNT initially.

However, it will scale out to MAX\_CLUSTER\_COUNT if the number of concurrent queries exceeds the server's capacity and queries start to queue.

Setting SCALING\_POLICY to ECONOMY (as opposed to STANDARD) ensures that the cluster is only scaled up if there are enough queries to keep the additional cluster busy for at least 6 minutes.

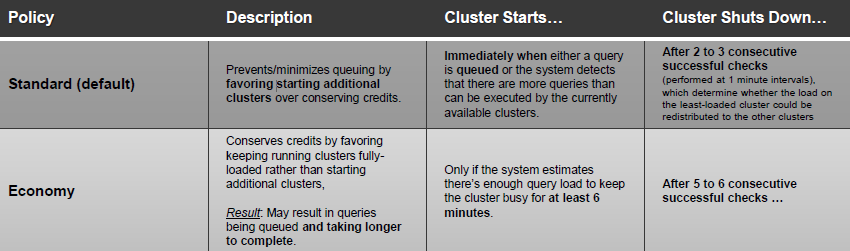
AUTO\_SUSPEND setting ensures that the cluster is suspended automatically after 300 seconds (or 5 minutes) of inactivity.

**Scaling Policy**

In Snowflake, we have predefined 2 types of scaling policy.

1. Standard (default policy) – performance (starts immediately when a query is queued)
2. Economy -- conserves/saves credits(cluster starts only if it estimates there is enough query load to keep the cluster busy for at least 6 minutes)





**Horizontal Scaling (Adding clusters of same size)**

**Scale out => adding clusters to a warehouse (multi cluster warehouse)**

Will help resolve concurrency issues. Multiple users accessing your application at the same time, and you don’t have resources, and then auto-scaling will help resolve that issue

Concurrency is executing two or more queries at the same time (in parallel) by single/multiple users.

**Scale in => removing clusters in a warehouses (Multi cluster warehouse)** concurrency =>queue (add clusters to existing warehouse)

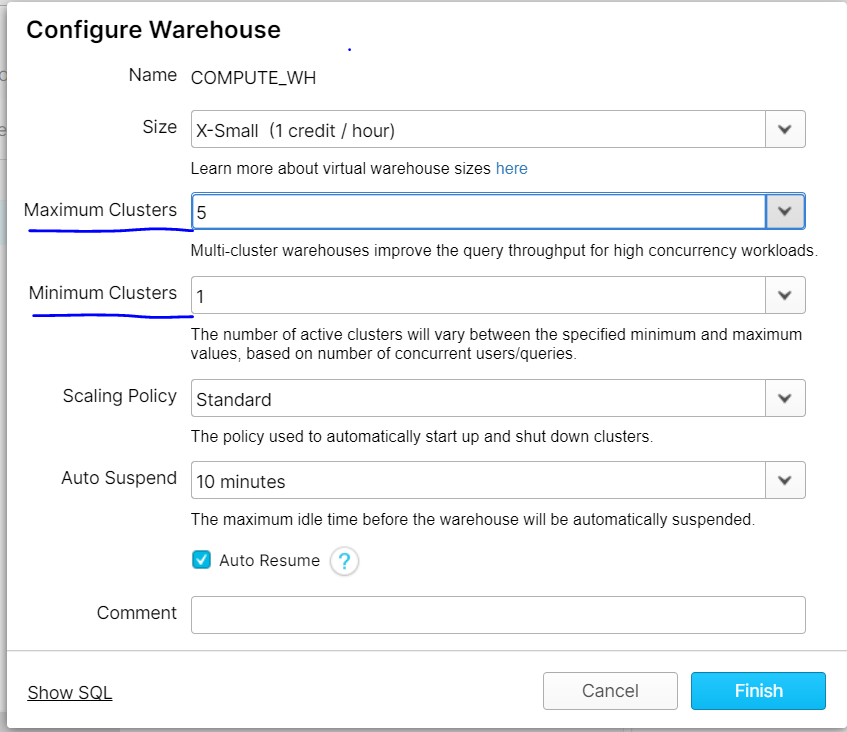
**Maximized and Auto-scale mode**

We can provision warehouses in 2 different modes.

* 1. **Auto-Scale mode:** At the time of warehouse creation if we set

**Max cluster count(5) > Minimum cluster count (1)**

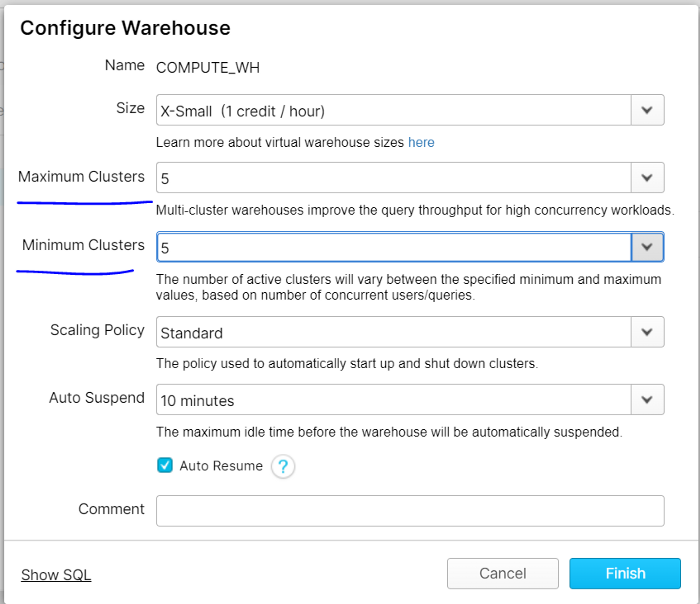
As the query load or concurrent user increases or decreases, the warehouse automatically scale-out/in respectively



* 1. **Maximized mode:**

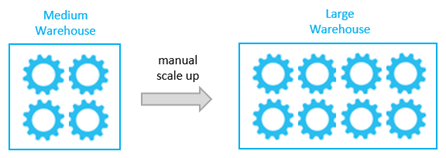
Max cluster count = Minimum cluster count

Snowflake starts all the clusters so that maximum resources are available while the warehouse is running.

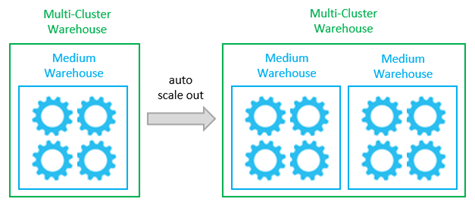


# Actual Behavior of Virtual Warehouses in Snowflake

1. Vertical scaling is manual process ( scaling up or down) in Snowflake (can’t be automated)

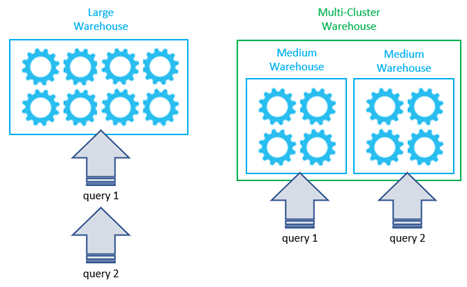


Snowflake auto scales horizontally (i.e scale in or scale out)

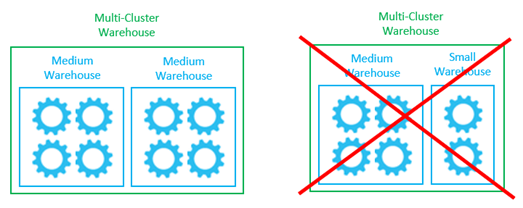


1. Multi-cluster warehouses improve concurrency.

They are designed specifically for handling queuing and performance issues related to large numbers of concurrent users and/or queries.



# Can we have warehouses of different sizes in a multiple-cluster



# A query can be processed by more than one cluster at the same time?

