**Snowflake Architecture**

What’s unique with snowflake architecture?

How snowflake architecture is different from traditional data bases(RDBMS like SQL Server, Oracle, My SQL) ?

What’s your understanding of Snowflake architecture?

Tell me about snowflake architecture?

It's a hybrid of shared-disk and shared-nothing architecture

Basically its Multi cluster shared data architecture. Storage and compute are separate and can be scaled independently to any extent and also it charges separately for storage and compute

(i.e. storage and compute are decoupled)

It’s 3 layer architecture

Storage layer => that stores data in hybrid columnar format and saved in cloud

Query Processing => Virtual Warehouse to run queries/for compute resources (muscle)

Cloud Services => brain of snowflake architecture coordinates between storage and query processing layers

/\*Data is stored in the cloud storage and works as a shared-disk model

(i.e. centralized data store like RDBMS)

Compute it will take advantage of performance and scale-out benefits of shared nothing architecture

(i.e. compute is distributed like hadoop) \*/

Why to go with Snowflake instead of other Cloud Data warehouses like AWS Redshift, Azure Synapse, Google Big query?

Snowflake is available on all cloud platforms like AWS, Azure and GCP (Multi cloud). Other cloud Data warehouses were restricted to that one cloud example: Redshift will be available in only AWS, Big Query will be available in GCP.

Unique features like Data Market place, Zero Copy Clone, Time travel, Data sharing, Streams for CDC

Separate Storage and compute costs, Zero Administration/Maintenance (partitioning, data backups, index maintenance, performance tuning will be taken care by snowflake itself).No software or hardware required to be installed or upgraded as snowflake is SaaS (Software as a service). Synapse and BigQuery are PaaS (Platform as a service)

How Snowflake charges for Storage and compute?

Snowflake Charges separately for Storage and separately for compute

**Virtual Warehouse(Query Processing)**

How is compute done in Snowflake?

Using virtual warehouses

What is Virtual Warehouse?

Cluster of compute nodes/machines

Cluster => group of nodes/machines connected together

Compute Resources => CPU + Memory+ Temporary storage

Different sizes of warehouses?

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

What is the default Auto suspend time in a Warehouse?

10 minutes (By default, auto-suspend is enabled. Snowflake automatically suspends the warehouse if it is IDLE for the specified period of time)

What is Auto resume in Virtual Warehouse?

If the warehouse is in suspended state and if we set Auto Resume as true then as soon as a query is ran using this warehouse it will automatically get resumed

(i.e. from suspended state it becomes active)

How many Virtual Warehouses have you used in your project?

ETL team ETL\_S\_WH

ETL\_L\_WH

Reporting\_WH

Datascience\_WH

Testing\_WH

We were using Snowflake\_WH warehouse

Maximum Warehouse size that we used in our project is XL or 2XL

ETL tools => Informatica, Microsoft Onpremises ETL tool SSIS, ADF (azure data factory), Datastage

Reporting tools => Power BI (Microsoft’s reporting tool), Tableau

Reporting\_WH

Testing\_WH

Datascience\_WH

ML\_WH

Can we increase or decrease the warehouse size any time?

Yes on fly (instantly) it will get resized

Can we increase or decrease the warehouse size while warehouse is in suspended?

Yes we can increase/decrease warehouse size irrespective whether your warehouse is in running state or suspended state

Can we start or stop the warehouse any time?

Yes, suspend will stop it and resume will start it again

We were using Snowflake\_WH warehouse for our development

What’s the Maximum Warehouse size that you have used in your project XL?

Can we increase or decrease the warehouse size any time?

Yes on fly (instantly) it will get resized

What will happen to already running queries if we resize the warehouse?

Already running queries will still use the old warehouse size but subsequent queries in queue will run with the resized warehouse

Can we start or stop the warehouse any time?

Yes, suspend will stop it and resume will start it again

What is difference between Scale up and scale out?

Increasing the warehouse size is Scale up and it’s to increase performance of complex queries and for large volumes of data (when remote disk spillage happens) (from S to X)

Scale down => decreasing the warehouse size (for example from L to S)

Adding clusters to a warehouse is scale out and it is to handle concurrency (parallel queries) issues

When to go with scale up, when to with scale out?

to handle long running/complex queries, we will increase the warehouse size i.e scale up

to handle concurrency issues(parallel users), we add clusters to a warehouse i.e scale out

how to set a warehouse to auto scale mode?

While creating a warehouse If you set max\_cluster\_count (10)> min\_cluster\_count(1) then the Virtual warehouse will be in auto scale mode

it will start with minimum number of clusters and scale out to maximum clusters based on workloads on demand

**Maximized Mode**=> max\_cluster\_count(5)= min\_cluster\_count (5)

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

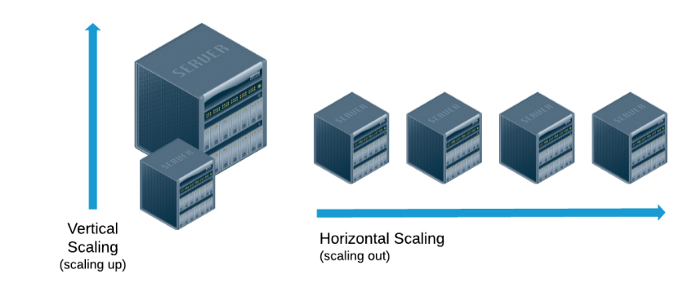
what is horizontal scaling and vertical scaling?

Scale up/scale down=> Horizontal scaling

Scale out/scale in=> Vertical scaling

What is the maximum clusters that can be added in case of scale out?

Maximum 10 clusters can be added



Can we automate scale up? No, we have to manual resize(increase/decrease) the warehouse

Scaling out => automated process taken care by snowflake based on your work load