

# Query Performance Tuning

---

**Sujith Nair**

Cloud Data Architect

Snowflake Snowpro Certified

# # What is query acceleration? # How can you enable query acceleration?

Query acceleration is a serverless compute service offered by snowflake which can allocate compute to queries to enhance their performance. The service must be enabled at the warehouse level for it to work.

Query acceleration is enabled at the warehouse level and can be done with the alter statement

```
ALTER WAREHOUSE COMPUTE_WH SET  
ENABLE_QUERY_ACCELERATION = true  
QUERY_ACCELERATION_MAX_SCALE_FACTOR = 1;
```

# # What is QUERY\_ACCELERATION\_MAX\_SCALE\_FACTOR?

The query acceleration scale factor determines how much compute resource will be provided to the warehouse.

If we set the value to 1 then it will consume the same number of credits as the warehouse `COMPUTE_WH`.

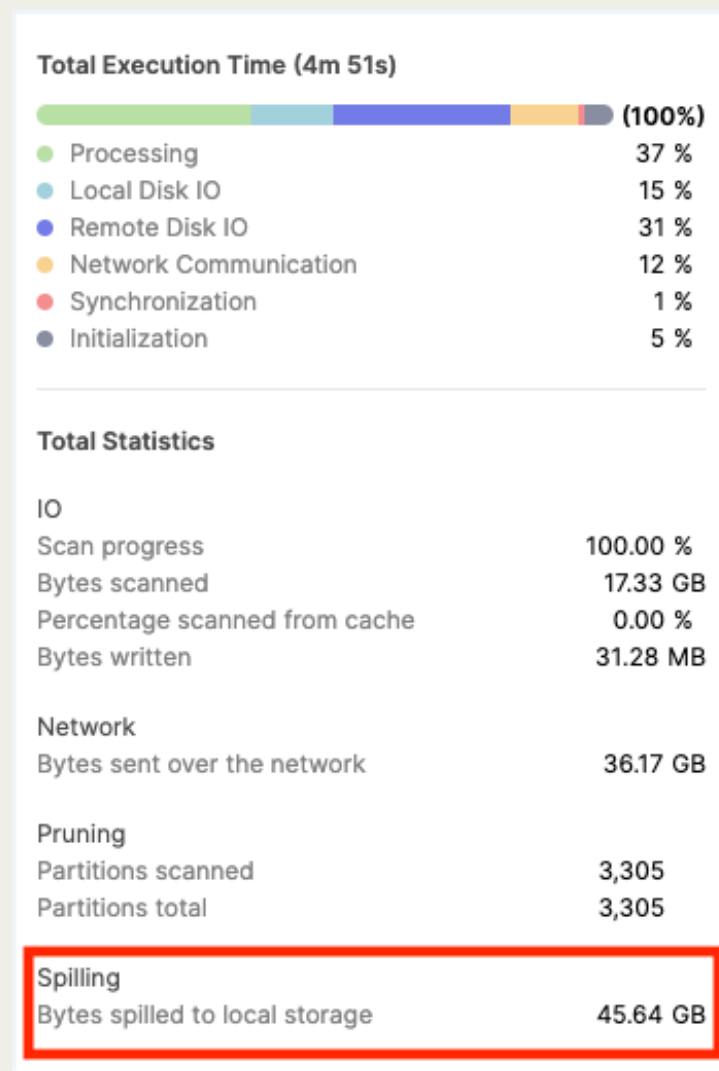
The default value is 8

```
ALTER WAREHOUSE COMPUTE_WH SET  
ENABLE_QUERY_ACCELERATION = true  
QUERY_ACCELERATION_MAX_SCALE_FACTOR = 1;
```

# # What is memory spillage ?

When a warehouse runs out of memory while executing a query it moves the data to the disk of the warehouse. This is called memory spilling. If the disk of the warehouse is not sufficient then it will spill to cloud storage.

Memory spillage is bad for query performance because it needs to write to disk and this slows down the processing of the query



```
SELECT
    query_id,
    SUBSTR(query_text, 1, 50) partial_query_text,
    user_name,
    warehouse_name,
    bytes_spilled_to_local_storage,
    bytes_spilled_to_remote_storage
FROM snowflake.account_usage.query_history
WHERE (bytes_spilled_to_local_storage > 0
OR bytes_spilled_to_remote_storage > 0 )
AND start_time::date > dateadd('days', -45, current_date)
ORDER BY bytes_spilled_to_remote_storage,
bytes_spilled_to_local_storage DESC
LIMIT 10;
```

# # How will you fix memory spillage issues?

Memory spillage happens because memory available is not sufficient to process the data being consumed.

Reduce the amount of data being processed , this can be done by eliminating columns that are not being used , or filtering data that does not need to be processed.

Increase the size of the warehouse to increase the amount of memory available

Use snowpark optimized warehouse which has 16x memory and 10x local disk space.

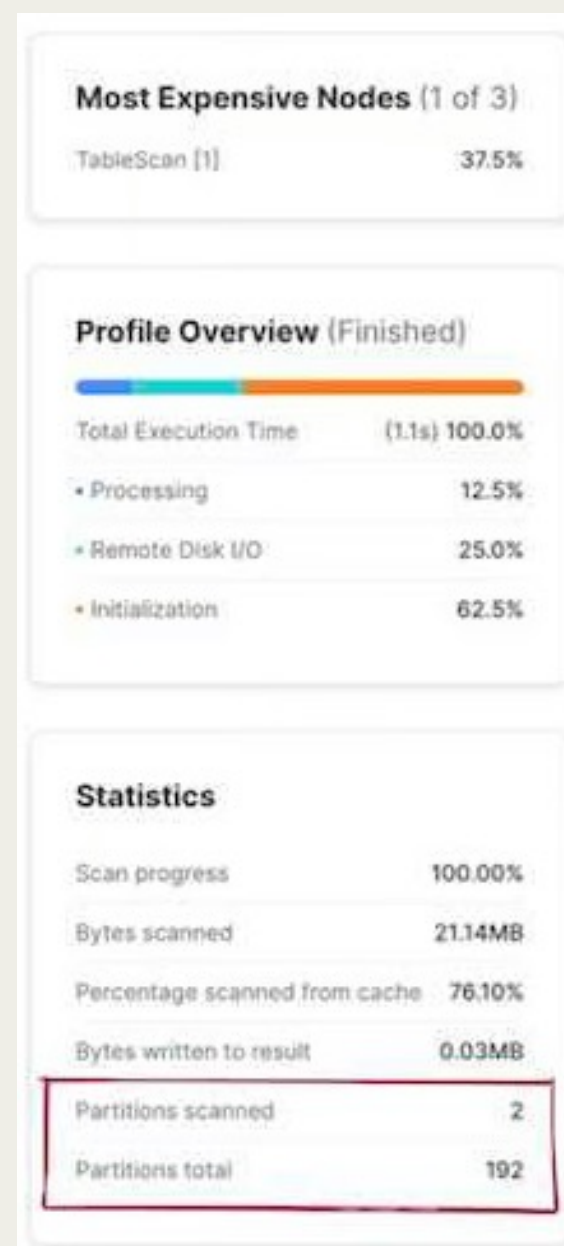
Standard Snowflake Warehouse	Snowpark Optimized Warehouse
X-small 16GB of RAM and 200GB SSD disk.	Not Available
Small 32GB of RAM and 400GB SSD disk.	Not Available
Medium 64GB of RAM and 800GB SSD disk.- 4 credits per hour	Medium 1024 GB of RAM and 8000 GB disk - 6 credits per Hour



# # What is partition pruning or query pruning in snowflake ?

The process of skipping partition when reading data from micro-partitions is called partition pruning or query pruning.

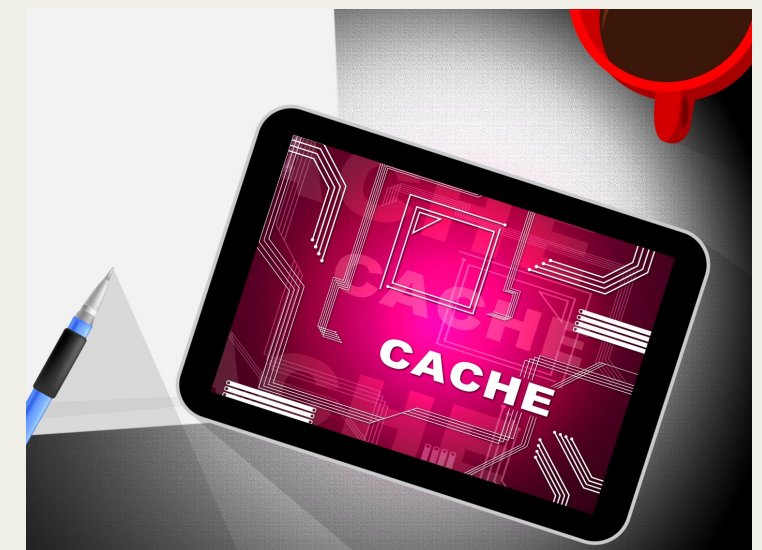
The metadata stored in partitions allow snowflake to skip partitions that do not contain data included in the where clause or join condition



# # What is metadata cache ?

Snowflake stores certain information about each table like row count, min max values of a column in its metadata in the cloud services layer so that certain queries about the table can be easily answered without reading the table.

- ROW COUNT of a table.
- MIN and MAX values of a column.
- Number of DISTINCT values in a column.





## # What is the difference between automatic clustering and manually clustering tables ?

Automatic clustering as the name suggests is automatically done by snowflake and is a process of partitioning data into micro-partitions based on snowflakes best guess based on its analysis of the data.

In manual clustering which is only recommended for tables that are over 1 TB in size, we pick the cluster keys based on our understanding of the data and the conditions that are used in filter conditions and joins for that table. We use the `CLUSTER BY` keyword to cluster the table.

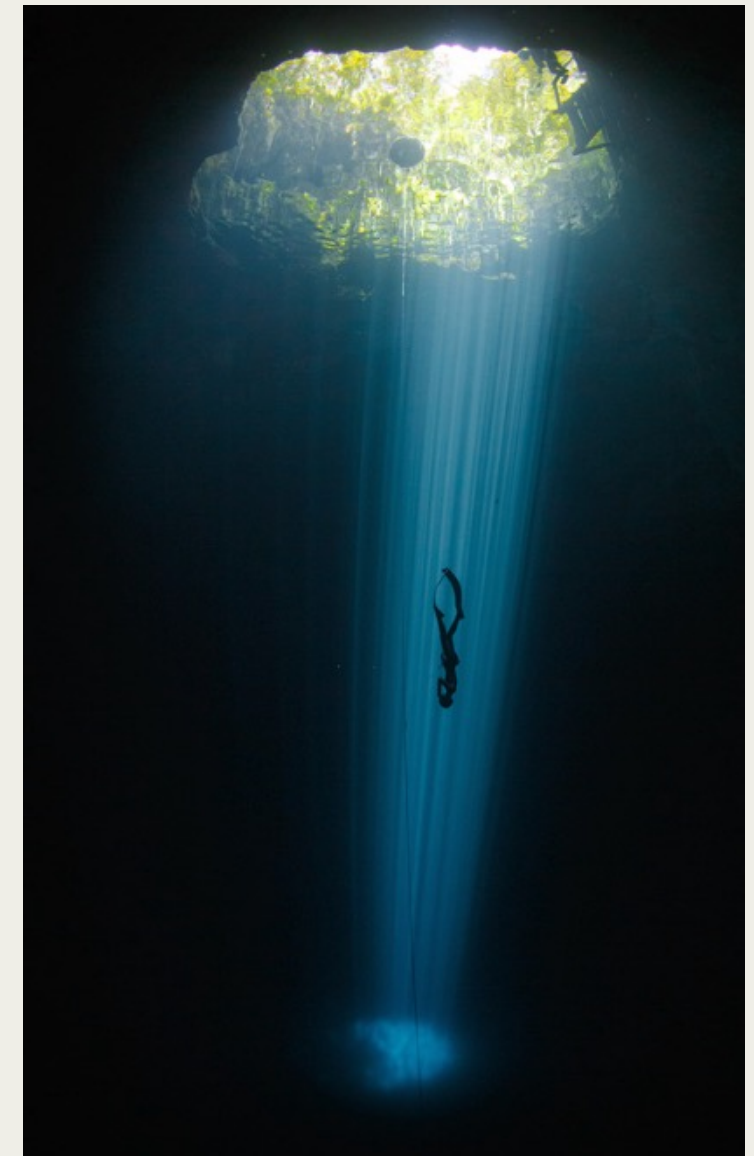
The table can be clustered either during creation or latter with the alter table statement.



## # What does clustering depth of a table tell us ?

It provides us with information on number of overlapping micro partitions. In general, a table with a large clustering depth will perform poorly in queries.

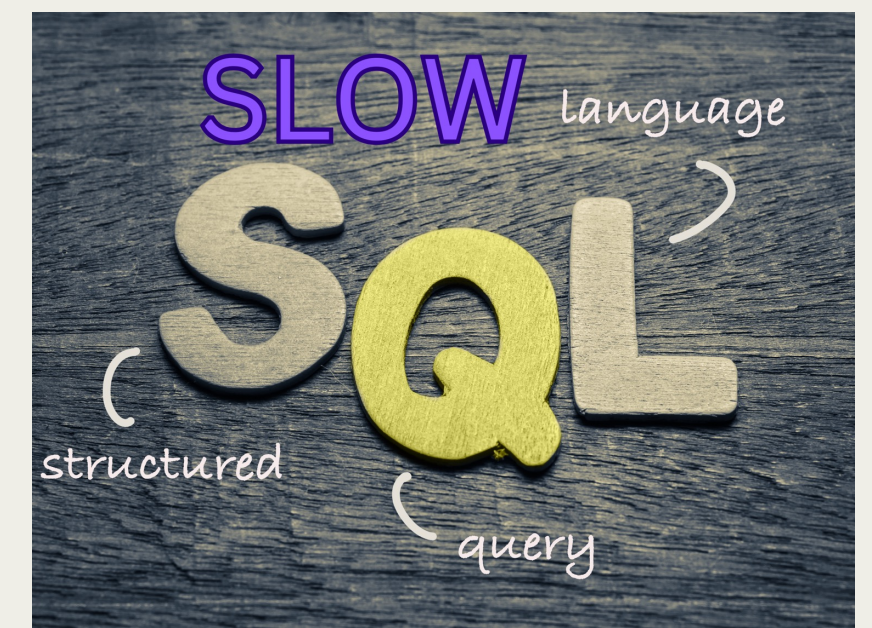
The function is called `CLUSTERING_DEPTH('<TableName>')`



**# You suspect that queries are waiting to get picked up by a warehouse to be processed, how will you confirm this ?**

If simple queries are taking more time than it should then the cause of the problem could be multiple queries running concurrently.

So how can we determine if there is a queuing problem.



```

SELECT
    DATE_TRUNC('hour', start_time) start_time_truncated_at_hour,
    HOUR(start_time)                start_time_hour,
    AVG(avg_running)                avg_nbr_of_query_running_per_hour,
    AVG(avg_queued_load)            avg_nbr_of_queued_query_due2overloaded_WH_per_hour,
    AVG(avg_queued_provisioning)    avg_nbr_of_queued_query_due2provisioning_WH_per_hour,
    AVG(avg_blocked)                avg_blocked
FROM snowflake.account_usage.warehouse_load_history
WHERE DATE_TRUNC('DAY', start_time) = '2021-13-19'
    AND warehouse_name = 'WAREHOUSE_NAME'
GROUP BY start_time_truncated_at_hour, start_time_hour
ORDER BY start_time_truncated_at_hour;

```

If we observe that there is queuing, then we can change the cluster to multi-warehouse or increase the maximum number of warehouses in the multi-cluster warehouse.

# Thank you!

---