

Optimizing the warehouse cache

This topic discusses how a warehouse owner or administrator can optimize a warehouse's cache in order to improve the performance of queries running on the warehouse.

A running warehouse maintains a cache of table data that can be accessed by queries running on the same warehouse. This can improve the performance of subsequent queries if they are able to read from the cache instead of from tables.

See also [Using Persisted Query Results](#), which explains how the results of specific queries may be cached and reused.

Note

You must have [access to the shared SNOWFLAKE database](#) to execute the diagnostic queries provided in this topic. By default, only the ACCOUNTADMIN role has the privileges needed to execute the queries.

Finding data scanned from cache

The following query provides the percentage of data scanned from cache, aggregated across all queries and broken out by warehouse.

If you have queries that can benefit from scanning data from the cache (e.g. frequent, similar queries) and the percentage of data scanned from cache is low, you might see a performance boost by optimizing the cache.

```
SELECT warehouse_name
, COUNT(*) AS query_count
, SUM(bytes_scanned) AS bytes_scanned
, SUM(bytes_scanned*percentage_scanned_from_cache) AS
bytes_scanned_from_cache
, SUM(bytes_scanned*percentage_scanned_from_cache) /
SUM(bytes_scanned) AS percent_scanned_from_cache
FROM snowflake.account_usage.query_history
WHERE start_time >= dateadd(month,-1,current_timestamp())
AND bytes_scanned > 0
GROUP BY 1
ORDER BY 5;
```

About the cache and auto-suspension

The auto-suspension setting of the warehouse can have a direct impact on query performance because the cache is dropped when the warehouse is suspended. If a warehouse is running frequent and similar queries, it might not make sense to suspend the warehouse in between queries because the cache might be dropped before the next query is executed.

You can use the following general guidelines when setting that auto-suspension time limit:

- For [tasks](#), Snowflake recommends immediate suspension.
- For DevOps, DataOps and Data Science use cases, Snowflake recommends setting auto-suspension to approximately 5 minutes, as the cache is not as important for ad-hoc and unique queries.
- For query warehouses, for example BI and SELECT use cases, Snowflake recommends setting auto-suspension to at least 10 minutes to maintain the cache for users.

Cost considerations

Keep in mind that a running warehouse consumes credits even if it is not processing queries. Be sure that your auto-suspension setting matches your workload. For example, if a warehouse executes a query every 30 minutes, it does not make sense to set the auto-suspension setting to 10 minutes. The warehouse will consume credits while sitting idle without gaining the benefits of a cache because it will be dropped before the next query executes.

How to configure auto-suspension

To change how much time must elapse before a warehouse is suspended and its cache dropped:

Snowsight

1. Sign in to Snowsight.
2. Navigate to **Admin » Warehouses**.
3. Find the warehouse, and select ... » **Edit**.
4. Ensure that **Auto Suspend** is turned on.
5. In the **Suspend After (min)** field, enter the number of minutes that must elapse before the warehouse is suspended.
6. Select **Save Warehouse**.

SQL

Use the [ALTER WAREHOUSE](#) command to change the auto-suspend time limit, which is specified in seconds, not minutes. For example:

```
ALTER WAREHOUSE my_wh SET AUTO_SUSPEND = 600;
```

Was this page helpful?