Tasks

Sujith Nair
Cloud Data Architect
Snowflake Snowpro Certified

#What is a task in snowflake ?

A task is a scheduling tool that can be used to run SQL statements or stored procedures in snowflake.

What can tasks do for us ?

- Run Tasks on a schedule
- Run Tasks at regular intervals
- Run Task to consume data from a stream
- Create dependency between Tasks

What are its Limitations ?

- Cannot create dependency with external systems

How do you monitor snowflake tasks?

If we are monitoring tasks then we can use the information_schema.task_history.

```
select * from table (information_schema.task_history
 (scheduled_time_range_start=>dateadd('hour',-4,current_timestamp()),
    scheduled_time_range_end=>current_timestamp(),
    task_name =>'<task_name>')
```

- Data in information schema appears immediately.
- Data is stored for 7 days.
- If you need data more than 7 days old then go to account_usage.task_history.

Create a view so that you don't have to write the query

```
CREATE OR REPLACE VIEW CURRENT_TASK_HISTORY_VW
AS
select * from table (information_schema.task_history
    (scheduled_time_range_start=>dateadd('hour',-24*7,current_timestamp()),
    scheduled_time_range_end=>current_timestamp()))
```

When a task created is it active or inactive by default?

The task is inactive by default and needs to be activated before it can start executing on schedule.

We can do that by using the
Alter task <taskname> resume;

What would happen if we created a task without a warehouse?

The task will run as per the schedule provided.
Snowflake will manage the compute resources for the task.

Before sept 2021, it was required to provide a warehouse for tasks.
To make the platform easier to use snowflake introduced the ability to create and run tasks without assigning a warehouse to it.

You can provide the optional parameter
USER_TASK_MANAGED_INITIAL_WAREHOUSE_SIZE
which will be the starting size for the task when
it begins execution as a serverless task.

If this parameter is not provided then the task is run on a medium task

What is the maximum number of tasks that a DAG can have?

A DAG(Directed Acyclic Graph) is a series of tasks composed of a single root task and additional child tasks organized by their dependencies.

A DAG is limited to 1000 tasks in total, including the root task

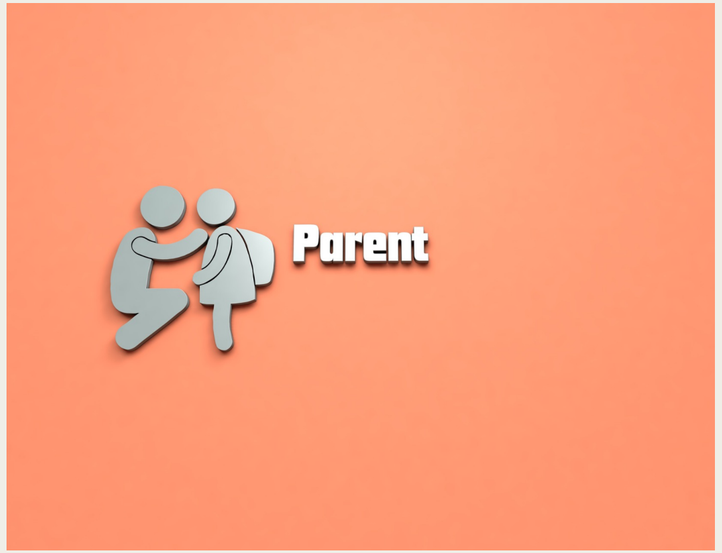
A single task can have a maximum of 100 predecessor tasks and 100 child tasks.

How can we create dependencies between tasks?

We can create dependencies between tasks by using the AFTER keyword ?

```
CREATE OR REPLACE TASK SNOWFLAKE_PARENT_ROOT
WAREHOUSE = COMPUTE_WH
SCHEDULE = '1 MINUTE'
AS SELECT CURRENT_TIMESTAMP();

CREATE TASK SNOWFLAKE_CHILD
WAREHOUSE = COMPUTE_WH
AFTER  SNOWFLAKE_PARENT_ROOT
AS SELECT CURRENT_TIMESTAMP();
```



You changed a table column causing a task to fail, how will you handle it ?

Suspend the task, if the task is part of a tree of tasks, remove the task from the tree. Add the task back once the table maintenance is complete and you made changes if needed to the task.
If you don't want child tasks to run suspend the child tasks

Alter task <taskname> suspend;

Alter task <taskname> remove after <parent_task>

Can you explain a scenarios in your project where you have used tasks ?

Use Case 1:
Copy data from ACCOUNT_USAGE.TABLE_STORAGE_METRICS for data growth analysis for 3 years to Data Governance schema.

Use Case 2:
We use tasks with Streams in our project to update several tables in near real time. We use the stream condition
SYSTEM$STREAM_HAS_DATA('<STREAMNAME>')
to run the task when there is data in the incoming stream.

Use Case 3:
Refresh reporting table/Data Mart from base table every 60 mins.

How do you find dependency information between tasks?

- Tasks can have parent tasks
- Tasks can have child task
- Task can have more than 1 child task
- Task can have more than 1 parent task

We can find dependency between tasks by using the information_schema.task_dependents.

```
select * from table (information_schema.task_dependents(
task_name => '<>' , recursive=> true));
```

task_dependents can only get us child task information.
Recursive means that it will give child, grandchild tasks.

Thank you!

Learn2CloudData Solutions