

Introduction to Docker

Shweta Ajay Shinde

Masters in Data Analytics, San Jose State University

Data 226: Data warehousing

Instructor: Keeyong Han

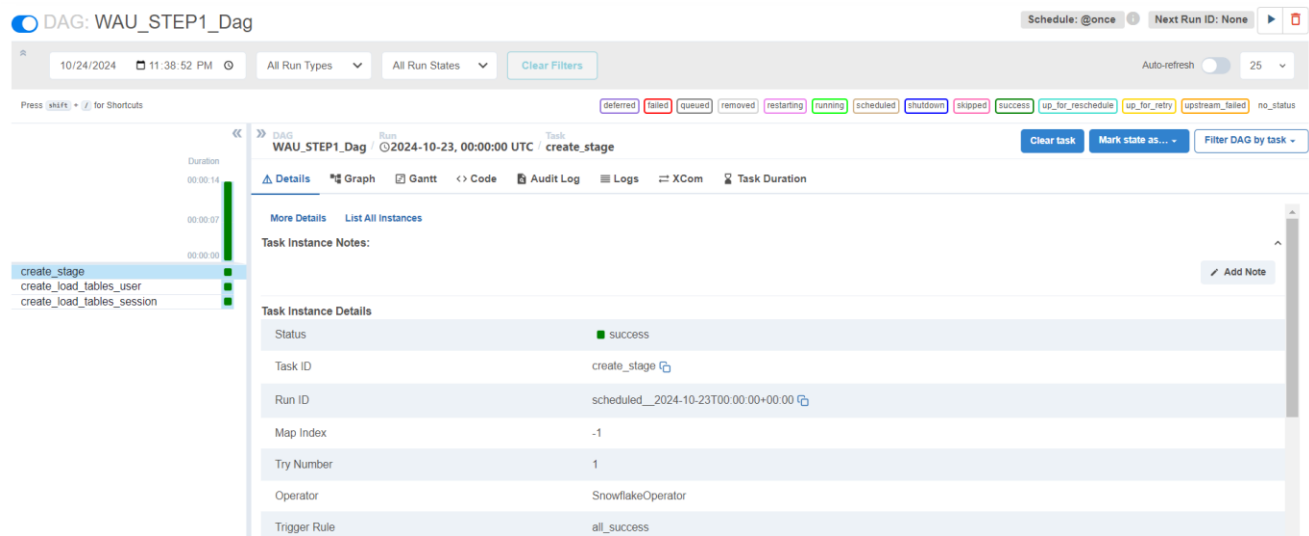
25th Oct 2024

Docker

Docker

1. Import two tables in your SnowflakeLinks to an external site. as an ETL DAG in your Airflow (+3pt)
 1. user_session_channel and session_timestamp (under raw_data schema or equivalent)
 2. Capture the screenshot of this DAG's detailed page from the Web UI (#1)

Screenshot for Step 1 Dag execution:-

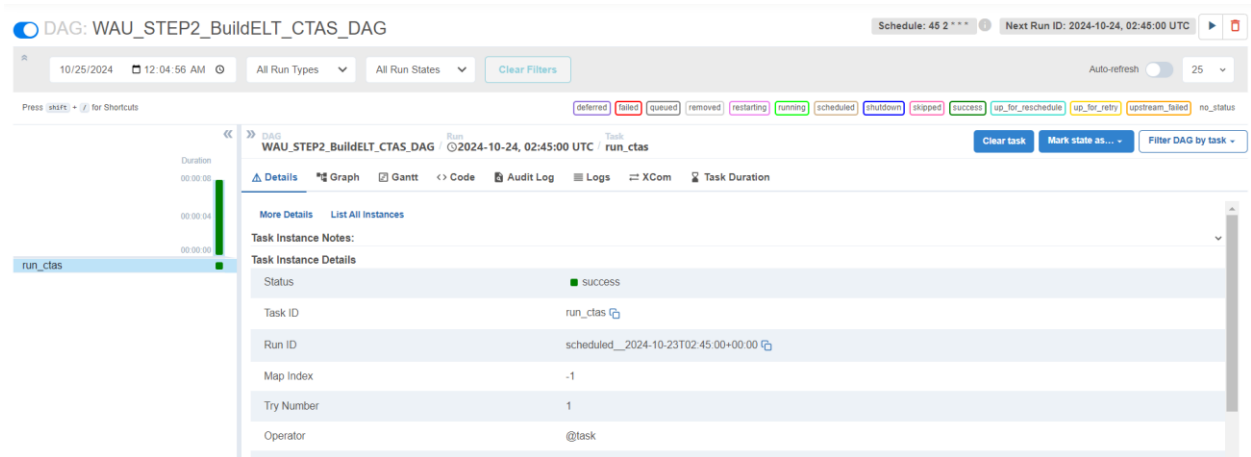


1. Create a ELT DAG in your Airflow to create a JOINED table of the two (+3pt)
 1. session_summary (under analytics)
 2. Extra point: add one more condition to check duplicate records (+1pt)

```
table = "dev.analytics.session_summary"
select_sql = """SELECT u.*, s.ts
FROM dev.raw_data.user_session_channel u
JOIN dev.raw_data.session_timestamp s ON u.sessionId=s.sessionId
GROUP BY u.userid,u.sessionId, u.channel, s.ts
HAVING COUNT(u.sessionId) = 1
"""
```

Docker

3. Capture the screenshot of this DAG's detailed page from the Web UI (#2) Screenshot for Step 2 Dag execution:-



Github Link for code click [here](#)

1. Set up your Preset account or [Links to an external site.](#) (+2pt)

1. This includes setting up Snowflake connection and import session_summary from step 2

Edit database

SNOWFLAKE

Shweta

BASIC

ADVANCED

...

DATABASE NAME *

dev

Copy the name of the database you are trying to connect to.

USERNAME

shweta12shinde

PASSWORD

.....

DISPLAY NAME *

Shweta

Pick a nickname for how the database will display in Superset.

ACCOUNT *

MIHZYEK-AEB64943

CLOSE

FINISH

Docker

2. Capture the screenshot of your Datasets (#3)

The screenshot shows the Snowflake Data Catalog interface. On the left, there are filters for DATABASE (snowflake), SCHEMA (analytics), and TABLE (session_summary). The main area displays the 'session_summary' dataset with its columns: userid (DECIMAL), sessionid (VARCHAR), channel (VARCHAR), and ts (TIMESTAMP...). At the bottom, there is a 'Datasets' section with a table listing the dataset.

NAME	TYPE	DATABASE	SCHEMA	OWNER	CERTIFIED
session	Select or type a value	Select or type a value	Select or type a value	Select or type a value	Select or type a value

Name	Type	Database	Schema	Owners	Last modified	Actions
session_summary	Physical	Shweta	analytics	SS	13 minutes ago	

2. Create your WAU chart (+2pt)

1. Make sure you rename the metrics field to WAU
2. Capture the screenshot of the chart (#4)

