

Sharan Patil

018327049

Prof. Keeyong Han

Assignment 10

GIT LINK- https://github.com/sharan9219790/HW10_DATAWAREHOUSE

Introduction

This report documents the implementation of a dbt (Data Build Tool) project that connects to Snowflake for building, transforming, and validating session-level data models. The purpose of this project is to demonstrate dbt model creation, dependency management, testing, and snapshotting using synthetic session data.

1. Project Setup

The project directory was located at:

`/Users/spartan/Downloads/DATA226/HW10/hw10_wau`

The Snowflake connection was configured under `~/dbt/profiles.yml` with the following details:

- **Account:** sfedu02-lvb17920
- **User:** COYOTE
- **Warehouse:** COYOTE_QUERY_WH
- **Database:** USER_DB_COYOTE
- **Schema:** ANALYTICS
- **Role:** TRAINING_ROLE

```
(.venv) spartan@MLK-SCS-J9LFH7MXXN HW10 % cd /Users/spartan/Downloads/DATA226/HW10/hw10_wau
dbt debug

00:15:49 Running with dbt=1.11.0-b3
00:15:49 dbt version: 1.11.0-b3
00:15:49 python version: 3.9.6
00:15:49 python path: /Users/spartan/Downloads/DATA226/HW10/.venv/bin/python3
00:15:49 os info: macOS-15.6.1-arm64-arm-64bit
00:15:49 /Users/spartan/Downloads/DATA226/HW10/.venv/lib/python3.9/site-packages/snowflake/connector/vendored/urllib3/__init__.py:35: NotOpenSSLWarning: urllib3 v2 only supports OpenSSL 1.1.1+, currently the 'ssl' module is compiled with 'LibreSSL 2.8.3'. See: https://github.com/urllib3/urllib3/issues/3020
warnings.warn(
00:15:50 Using profiles dir at /Users/spartan/.dbt
00:15:50 Using profiles.yml file at /Users/spartan/.dbt/profiles.yml
00:15:50 Using dbt_project.yml file at /Users/spartan/Downloads/DATA226/HW10/hw10_wau/dbt_project.yml
00:15:50 adapter type: snowflake
00:15:50 adapter version: 1.10.2
00:15:50 Configuration:
00:15:50   profiles.yml file [OK found and valid]
00:15:50   dbt_project.yml file [OK found and valid]
00:15:50 Required dependencies:
00:15:50   - git [OK found]

00:15:50 Connection:
```

2. Input Models

Two input models were developed under models/input/ to generate synthetic session and user data. Both are materialized as **views** in Snowflake under the ANALYTICS schema.

session_timestamp.sql

```
with data as (
  select
    column1::string as SESSION_ID,
    to_timestamp_ntz(column2) as EVENT_TS
  from values
    ('S000001', '2025-10-20 10:00:00'),
    ('S000001', '2025-10-20 10:15:00'),
    ('S000002', '2025-10-21 09:05:00'),
    ('S000002', '2025-10-21 09:20:00'),
    ('S000002', '2025-10-21 10:00:00'),
    ('S000003', '2025-10-22 14:30:00'),
    ('S000004', '2025-10-23 08:00:00'),
    ('S000004', '2025-10-23 08:05:00'),
    ('S000005', '2025-10-24 18:00:00'),
    ('S000005', '2025-10-24 18:30:00')
)
select SESSION_ID, EVENT_TS
from data
```

user_session_channel.sql

```
with data as (
  select
```

```

column1::string as SESSION_ID,
column2::string as USER_ID,
column3::string as CHANNEL
from values
('S000001', 'U001', 'mobile'),
('S000002', 'U002', 'web'),
('S000003', 'U001', 'web'),
('S000004', 'U003', 'mobile'),
('S000005', 'U004', 'web')
)
select SESSION_ID, USER_ID, CHANNEL
from data

```

The models were executed successfully using:

```
dbt run --select 'hw10_wau.input.*'
```

```

00:44:24 Found 3 models, 1 snapshot, 2 data tests, 492 macros
00:44:24
00:44:24 Concurrency: 4 threads (target='dev')
00:44:24
/Users/spartan/Downloads/DATA226/HW10/.venv/lib/python3.9/site-packages/boto3/compat.py:84: PythonDepre
cationWarning: Boto3 will no longer support Python 3.9 starting April 29, 2026. To continue receiving s
ervice updates, bug fixes, and security updates please upgrade to Python 3.10 or later. More informatio
n can be found here: https://aws.amazon.com/blogs/developer/python-support-policy-updates-for-aws-sdks-
and-tools/
  warnings.warn(warning, PythonDeprecationWarning)
00:44:27 1 of 2 START sql view model ANALYTICS.session_timestamp ..... [RUN]
00:44:27 2 of 2 START sql view model ANALYTICS.user_session_channel ..... [RUN]
/Users/spartan/Downloads/DATA226/HW10/.venv/lib/python3.9/site-packages/boto3/compat.py:84: PythonDepre
cationWarning: Boto3 will no longer support Python 3.9 starting April 29, 2026. To continue receiving s
ervice updates, bug fixes, and security updates please upgrade to Python 3.10 or later. More informatio
n can be found here: https://aws.amazon.com/blogs/developer/python-support-policy-updates-for-aws-sdks-
and-tools/
  warnings.warn(warning, PythonDeprecationWarning)
00:44:27 1 of 2 OK created sql view model ANALYTICS.session_timestamp ..... [SUCCESS 1 i
n 0.47s]
00:44:30 2 of 2 OK created sql view model ANALYTICS.user_session_channel ..... [SUCCESS 1 i
n 2.75s]
00:44:30
00:44:30 Finished running 2 view models in 0 hours 0 minutes and 6.22 seconds (6.22s).
00:44:30
00:44:30 Completed successfully
00:44:30

```

3. Output Model – session_summary.sql

The output model aggregates session-level data and joins it with user information. It computes session start, end, duration, and total event count.

```

with ts as (
select
SESSION_ID,
min(EVENT_TS) as SESSION_START,
max(EVENT_TS) as SESSION_END,
datediff('minute', min(EVENT_TS), max(EVENT_TS)) as DURATION_MINUTES,
count(*) as EVENT_COUNT
from {{ ref('session_timestamp') }}

```

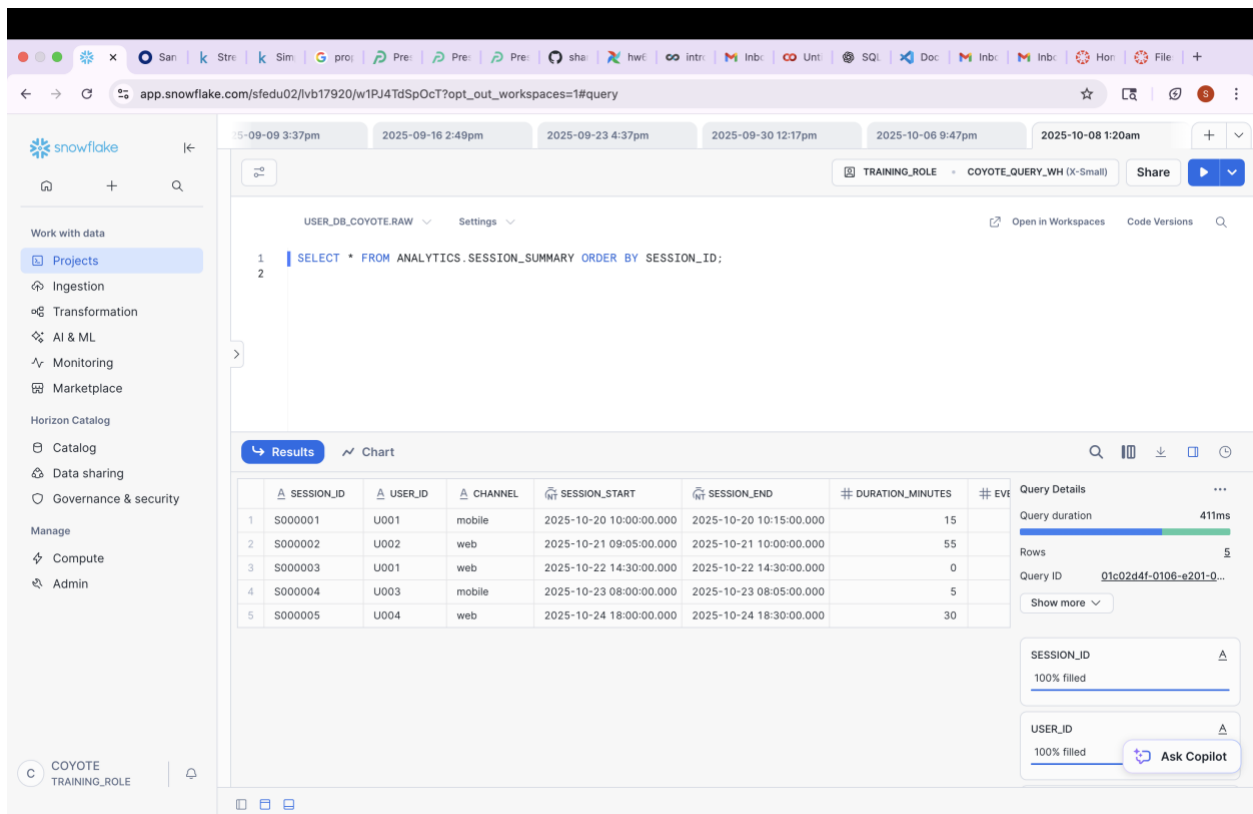
```

    group by SESSION_ID
),
map as (
    select SESSION_ID, USER_ID, CHANNEL
    from {{ ref('user_session_channel') }}
)
select
    ts.SESSION_ID,
    mp.USER_ID,
    mp.CHANNEL,
    ts.SESSION_START,
    ts.SESSION_END,
    ts.DURATION_MINUTES,
    ts.EVENT_COUNT
from ts ts
left join map mp
    on ts.SESSION_ID = mp.SESSION_ID

```

Command used to build:

```
dbt run --select 'hw10_wau.output.session_summary'
```



The screenshot displays the Snowflake web interface. The main area shows a query execution result for the query: `SELECT * FROM ANALYTICS.SESSION_SUMMARY ORDER BY SESSION_ID;`. The results are displayed in a table with 5 rows and 7 columns: SESSION_ID, USER_ID, CHANNEL, SESSION_START, SESSION_END, DURATION_MINUTES, and EVENT_COUNT. The table data is as follows:

	SESSION_ID	USER_ID	CHANNEL	SESSION_START	SESSION_END	DURATION_MINUTES	EVT
1	S000001	U001	mobile	2025-10-20 10:00:00.000	2025-10-20 10:15:00.000	15	
2	S000002	U002	web	2025-10-21 09:05:00.000	2025-10-21 10:00:00.000	55	
3	S000003	U001	web	2025-10-22 14:30:00.000	2025-10-22 14:30:00.000	0	
4	S000004	U003	mobile	2025-10-23 08:00:00.000	2025-10-23 08:05:00.000	5	
5	S000005	U004	web	2025-10-24 18:00:00.000	2025-10-24 18:30:00.000	30	

The interface also shows a sidebar with navigation options, a top bar with tabs for different queries, and a bottom right section with query details including duration (411ms) and rows (5).

4. Snapshot – snapshot_session_summary.sql

Snapshots track historical changes to the session_summary model. This ensures any updates to session data can be versioned over time.

```
{% snapshot snapshot_session_summary %}
{{
  config(
    target_schema='ANALYTICS',
    target_database='USER_DB_COYOTE',
    unique_key='SESSION_ID',
    strategy='timestamp',
    updated_at='SESSION_END'
  )
}}

select * from {{ ref('session_summary') }}
{% endsnapshot %}
```

Command to run:

dbt snapshot

```
{% snapshot snapshot_session_summary %}
{{
  config(
    target_schema='ANALYTICS',
    target_database='USER_DB_COYOTE',
    unique_key='SESSION_ID',
    strategy='timestamp',
    updated_at='SESSION_END'
  )
}}

select * from {{ ref('session_summary') }}
{% endsnapshot %}
```

5. Data Quality Tests – schema.yml

Data validation was configured to ensure SESSION_ID values are unique and non-null, and all records contain valid USER_ID values.

version: 2

models:

- name: session_summary
 - description: "Aggregated session data by user and channel"
 - columns:
 - name: SESSION_ID
 - tests:
 - unique
 - not_null
 - name: USER_ID
 - tests:
 - not_null

Command used:

dbt test

Result: All tests passed successfully.

```
(.venv) spartan@MLK-SCS-J9LFH7MXXN hw10_wau % dbt test

01:18:36 Running with dbt=1.11.0-b3
/Users/spartan/Downloads/DATA226/HW10/.venv/lib/python3.9/site-packages/snowflake/connector/vendored/urllib3/__init__.py:35: NotOpenSSLWarning: urllib3 v2 only supports OpenSSL 1.1.1+, currently the 'ssl' module is compiled with 'LibreSSL 2.8.3'. See: https://github.com/urllib3/urllib3/issues/3020
  warnings.warn(
01:18:36 Registered adapter: snowflake=1.10.2
01:18:37 Found 1 snapshot, 3 models, 2 data tests, 492 macros
01:18:37
01:18:37 Concurrency: 4 threads (target='dev')
01:18:37
/Users/spartan/Downloads/DATA226/HW10/.venv/lib/python3.9/site-packages/boto3/compat.py:84: PythonDeprecationWarning: Boto3 will no longer support Python 3.9 starting April 29, 2026. To continue receiving service updates, bug fixes, and security updates please upgrade to Python 3.10 or later. More information can be found here: https://aws.amazon.com/blogs/developer/python-support-policy-updates-for-aws-sdks-and-tools/
  warnings.warn(warning, PythonDeprecationWarning)
01:18:39 1 of 2 START test not_null_session_summary_session_id ..... [RUN]
01:18:39 2 of 2 START test unique_session_summary_session_id ..... [RUN]
/Users/spartan/Downloads/DATA226/HW10/.venv/lib/python3.9/site-packages/boto3/compat.py:84: PythonDeprecationWarning: Boto3 will no longer support Python 3.9 starting April 29, 2026. To continue receiving service updates, bug fixes, and security updates please upgrade to Python 3.10 or later. More information can be found here: https://aws.amazon.com/blogs/developer/python-support-policy-updates-for-aws-sdks-and-tools/
  warnings.warn(warning, PythonDeprecationWarning)
01:18:40 2 of 2 PASS unique_session_summary_session_id ..... [PASS in 0.84s]
01:18:40 1 of 2 PASS not_null_session_summary_session_id ..... [PASS in 0.84s]
01:18:41
01:18:41 Finished running 2 data tests in 0 hours 0 minutes and 4.04 seconds (4.04s).
01:18:41
01:18:41 Completed successfully
01:18:41
01:18:41 Done. PASS=2 WARN=0 ERROR=0 SKIP=0 NO-OP=0 TOTAL=2
(.venv) spartan@MLK-SCS-J9LFH7MXXN hw10_wau %
```

6. Conclusion

The dbt project was successfully implemented with Snowflake as the data warehouse.

Key achievements include:

- Successful model creation and execution (input, output, snapshot).
- Automated data testing and validation using dbt's native test framework.
- Modular model design ensuring easy reusability and scalability.

This project demonstrates the complete lifecycle of ELT (Extract, Load, Transform) using dbt and Snowflake integration.