**dbt**

# Introduction to dbt

Shweta Ajay Shinde
Masters in Data Analytics, San Jose State University
Data 226: Data warehousing
Instructor: Keeyong Han
30th Oct 2024

**dbt**

**dbt**

1. **Create a dbt project with snowflake connector (1 pt).**

**dbt Snowflake connection**

```
03:01:00  Setting up your profile.
Which database would you like to use?
[1] snowflake

(Don't see the one you want? https://docs.getdbt.com/docs/available-adapters)

Enter a number: 1
account (https://<this_value>.snowflakecomputing.com): mihzyek-aeb64943
user (dev username): shweta12shinde
[1] password
[2] keypair
[3] sso
Desired authentication type option (enter a number): 1
password (dev password):
role (dev role): ACCOUNTADMIN
warehouse (warehouse name): compute_wh
database (default database that dbt will build objects in): dev
schema (default schema that dbt will build objects in): analytics
threads (1 or more) [1]: 1
03:04:13  Profile build_mau_shweta written to /Users/spartan/.dbt/profiles.yml using target's profile_template.yml and your supplied values. Run 'dbt debug' to validate the connection.
(base) spartan@MLK-SCS-FVFHW4Y5Q05N Documents %
```

2. **Set up Input models (2pt)**
   1. **Input tables should be built as CTE**
   2. **Show contents of model's/input/ user_session_channel.sql and input/session_timestamp.sql**

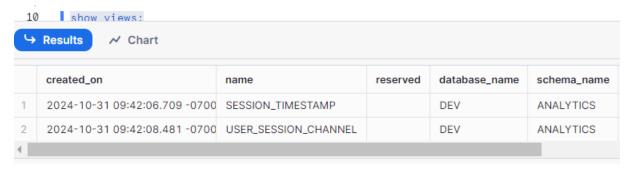   **user_session_channel.sql SQL Query**

```sql
with user_data as (
    SELECT
        userId,
        sessionId,
        channel
    FROM {{ source('raw_data', 'user_session_channel') }}
    WHERE sessionId IS NOT NULL
)

select * from user_data
```

**dbt**

**input/session_timestamp.sql SQL Query**

```sql
with session_data as (
    SELECT
        sessionId,
        ts
    FROM {{ source('raw_data', 'session_timestamp') }}
    WHERE sessionId IS NOT NULL
)


select * from session_data
```

**dbt run to create views in snowflake.**

```
(base) spartan@MLK-SCS-FVFHW4Y5Q05N build_mau_shweta % dbt run
06:00:45  Running with dbt=1.8.7
06:00:47  Registered adapter: snowflake=1.8.4
06:00:47  Found 2 models, 3 sources, 459 macros
06:00:47
06:00:49  Concurrency: 1 threads (target='dev')
06:00:49
06:00:49  1 of 2 START sql view model analytics.session_timestamp ....................... [RUN]
06:00:51  1 of 2 OK created sql view model analytics.session_timestamp ................... [SUCCESS 1 in 1.45s]
06:00:51  2 of 2 START sql view model analytics.user_session_channel .................... [RUN]
06:00:52  2 of 2 OK created sql view model analytics.user_session_channel ............... [SUCCESS 1 in 1.02s]
06:00:52
06:00:52  Finished running 2 view models in 0 hours 0 minutes and 4.33 seconds (4.33s).
06:00:52
06:00:52  Completed successfully
06:00:52
06:00:52  Done. PASS=2 WARN=0 ERROR=0 SKIP=0 TOTAL=2
(base) spartan@MLK-SCS-FVFHW4Y5Q05N build_mau_shweta %
```

**Views created in Snowflake.**

```
10    show views:
```
↳ Results    ∿ Chart

| | created_on | name | reserved | database_name | schema_name |
|---|---|---|---|---|---|
| 1 | 2024-10-31 09:42:06.709 -0700 | SESSION_TIMESTAMP | | DEV | ANALYTICS |
| 2 | 2024-10-31 09:42:08.481 -0700 | USER_SESSION_CHANNEL | | DEV | ANALYTICS |

**dbt**

3. **Set up Output models (1pt)**

   1. **Show contents of models/output/session_summary.sql**

      **session_summary.sql SQL query**

```sql
WITH u AS (
SELECT * FROM {{ ref("user_session_channel") }}
), st AS (
SELECT * FROM {{ ref("session_timestamp") }}
)
SELECT u.userId, u.sessionId, u.channel, st.ts
FROM u
JOIN st ON u.sessionId = st.sessionId
```

      **dbt run to create session_summary view**

```
(base) spartan@MLK-SCS-FVFHW4Y5Q05N build_mau_shweta % dbt run
06:05:51  Running with dbt=1.8.7
06:05:53  Registered adapter: snowflake=1.8.4
06:05:54  Found 3 models, 3 sources, 459 macros
06:05:54
06:05:55  Concurrency: 1 threads (target='dev')
06:05:55
06:05:55  1 of 1 START sql table model analytics.session_summary ........................ [RUN]
06:05:59  1 of 1 OK created sql table model analytics.session_summary ................... [SUCCESS 1 in 3.92s]
06:05:59
06:05:59  Finished running 1 table model in 0 hours 0 minutes and 5.42 seconds (5.42s).
06:05:59
06:05:59  Completed successfully
06:05:59
06:05:59  Done. PASS=1 WARN=0 ERROR=0 SKIP=0 TOTAL=1
(base) spartan@MLK-SCS-FVFHW4Y5Q05N build_mau_shweta %
```

      **View created in snowflake.**

```
10    | show views;
```

      ↳ Results    ∿ Chart

| | created_on | name | reserved | database_name | schema_name |
|---|---|---|---|---|---|
| 1 | 2024-10-31 09:59:30.414 -0700 | SESSION_SUMMARY | | DEV | ANALYTICS |

**dbt**

**4.Add Snapshot to the output table (session_summary) (2pt)**
             **1.Show contents of snapshots/snapshot_session_summary.sql**

       **snapshot_session_summary.sql**

```
{% snapshot snapshot_session_summary %}
{{
 config(
   target_schema='snapshot',
   unique_key='sessionId',
   strategy='timestamp',
   updated_at='ts',
   invalidate_hard_deletes=True
 )
}}

SELECT * FROM {{ ref('session_summary') }}
{% endsnapshot %}
```

       **Screenshot of dbt snapshot**

```
(base) spartan@MLK-SCS-FVFHW4Y5Q05N build_mau_shweta % dbt snapshot
06:07:32  Running with dbt=1.8.7
06:07:34  Registered adapter: snowflake=1.8.4
06:07:35  Found 3 models, 1 snapshot, 3 sources, 459 macros
06:07:35
06:07:37  Concurrency: 1 threads (target='dev')
06:07:37
06:07:37  1 of 1 START snapshot snapshot.snapshot_session_summary ....................... [RUN]
06:07:39  1 of 1 OK snapshotted snapshot.snapshot_session_summary ....................... [SUCCESS 1 in 2.48s]
06:07:39
06:07:39  Finished running 1 snapshot in 0 hours 0 minutes and 4.85 seconds (4.85s).
06:07:40
06:07:40  Completed successfully
06:07:40
06:07:40  Done. PASS=1 WARN=0 ERROR=0 SKIP=0 TOTAL=1
(base) spartan@MLK-SCS-FVFHW4Y5Q05N build_mau_shweta %
```

**dbt**

**5.Add at least 2 tests to sessionId field of session_summary (2pt)**
        **1.Show contents of models/schema.yml**

        **Schema.yml file**

```yaml
version: 2
models:
  - name: session_summary
    description: "Analytics model for session data"
    columns:
      - name: sessionId
        description: "Unique identifier for each stock price record"
        data_tests:
          - unique
          - not_null
```

        **Screenshot of debt test**

```
[(base) spartan@MLK-SCS-FVFHW4Y5Q05N build_mau_shweta % dbt test
06:11:17  Running with dbt=1.8.7
06:11:19  Registered adapter: snowflake=1.8.4
06:11:19  Found 3 models, 1 snapshot, 2 data tests, 3 sources, 459 macros
06:11:19
06:11:21  Concurrency: 1 threads (target='dev')
06:11:21
06:11:21  1 of 2 START test not_null_session_summary_sessionId .......................... [RUN]
06:11:22  1 of 2 PASS not_null_session_summary_sessionId ................................ [PASS in 0.53s]
06:11:22  2 of 2 START test unique_session_summary_sessionId ........................... [RUN]
06:11:23  2 of 2 PASS unique_session_summary_sessionId ................................. [PASS in 0.99s]
06:11:23
06:11:23  Finished running 2 data tests in 0 hours 0 minutes and 3.23 seconds (3.23s).
06:11:23
06:11:23  Completed successfully
06:11:23
06:11:23  Done. PASS=2 WARN=0 ERROR=0 SKIP=0 TOTAL=2
(base) spartan@MLK-SCS-FVFHW4Y5Q05N build_mau_shweta % 
```

**For Git Hub link click here**