

Data Warehouse Assignment 7

1. Create a dbt project with snowflake connector

```
03:31:15 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): SGEDU02-KSB65579
user (dev username): CHIPMUNK
[1] password
[2] keypair
[3] sso
Desired authentication type option (enter a number): 1
password (dev password):
role (dev role): TRAINING_ROLE
warehouse (warehouse name): CHIPMUNK_QUERY_WH
database (default database that dbt will build objects in): USER_DB_CHIPMUNK
schema (default schema that dbt will build objects in): analytics
threads (1 or more) [1]:
03:35:31 Profile build_mau written to C:\Users\aiish\.dbt\profiles.yml using target's profile_template.yml and your supplied values. Run 'dbt debug' to validate the connection.
```

2. Set up Input models

```
C:\Users\aiish\OneDrive\Desktop\build_mau\models>dbt run
16:27:34 Running with dbt=1.9.3
16:27:35 Registered adapter: snowflake=1.9.2
16:27:36 Found 2 models, 3 sources, 474 macros
16:27:36 Concurrency: 1 threads (target='dev')
16:27:36
16:27:37 1 of 2 START sql view model analytics.user_session_channel ..... [RUN]
16:27:38 1 of 2 OK created sql view model analytics.user_session_channel ..... [SUCCESS 1 in
0.83s]
16:27:38 2 of 2 START sql view model analytics.user_session_timestamp ..... [RUN]
16:27:38 2 of 2 OK created sql view model analytics.user_session_timestamp ..... [SUCCESS 1 in
0.38s]
16:27:39
16:27:39 Finished running 2 view models in 0 hours 0 minutes and 2.94 seconds (2.94s).
16:27:39 Completed successfully
16:27:39
16:27:39 Done. PASS=2 WARN=0 ERROR=0 SKIP=0 TOTAL=2
```

a. session_timestamp content

```
1  SELECT
2    |   sessionId,
3    |   ts
4  FROM {{ source('raw', 'session_timestamp') }}
5  WHERE sessionId IS NOT NULL
```

	SESSIONID	TS
1	7cdace91c487558e27ce54df7cdb299c	2019-05-01T00:13:11.783Z
2	94f192dee566b018e0acf31e1f99a2d9	2019-05-01T00:49:46.073Z
3	7ed2d3454c5eea71148b11d0c25104ff	2019-05-01T10:18:43.21Z
4	f1daf122cde863010844459363cd31db	2019-05-01T13:10:56.413Z
5	fd0efcca272f704a760c3b61dcc70fd0	2019-05-01T13:45:19.793Z
6	8804f94e16ba5b680e239a554a08f7d2	2019-05-01T14:23:07.66Z
7	c5f441cd5f43eb2f2c024e1f8b5d00cd	2019-05-01T15:03:54.65Z
8	d5fcc35c94879a4afad61cacca56192c	2019-05-01T15:13:16.14Z
9	3d191ef6e236bd1b9bdb9ff4743c47fe	2019-05-01T15:33:58.197Z
10	c17028c9b6e0c5deaad29665d582284a	2019-05-01T15:59:57.49Z
11	cd0b43eac0392accf3624b7372dec36e	2019-05-01T16:33:03.463Z
12	0a4bbceda17a6253386bc9eb45240e25	2019-05-01T17:10:29.417Z
13	c67ba7c4c5c0cd4cc3e3a7146fe5c015	2019-05-01T17:42:13.58Z
14	d8a3a3c3234392b0add43c5f9c05a246	2019-05-01T18:18:00.407Z

b. user_session_channel content

```

1  SELECT
2      userId,
3      sessionId,
4      channel
5  FROM {{ source('raw', 'user_session_channel') }}
6  WHERE sessionId IS NOT NULL

```

...	USERID	SESSIONID	CHANNEL
1	184	c41dd99a69df04044aa4e33ece9c9249	Naver
2	80	fdc0eb412a84fa549afe68373d9087e9	Organic
3	251	0a54b19a13b6712dc04d1b49215423d8	Facebook
4	264	a914ecef9c12ffdb9bede64bb703d877	Google
5	744	05ae14d7ae387b93370d142d82220f1b	Facebook
6	229	eddeb82df22554fa67c641e3f8a25566	Google
7	265	4c4ea5258ef3fb3fb1fc48fee9b4408c	Naver
8	80	8ae1da0fe37c98412768453f82490da2	Organic
9	45	60131a2a3f223dc8f4753bcc5771660c	Youtube
10	198	90f4760fcc9b69c13da7368c5c2917f3	Facebook
11	262	761efc843ff05ab74ed358713dd51c1b	Organic
12	237	233f1dd0f3f537bcb7a338ea74d63483	Organic
13	702	8c97dbeee3b0d40ced7f514b99a93d93	Instagram
14	341	e10534dd65cf727692c0f9c44ba613f8	Naver

3. Set up Output models

```

C:\Users\aiish\OneDrive\Desktop\build_mau\models>dbt run
16:48:33 Running with dbt=1.9.3
16:48:34 Registered adapter: snowflake=1.9.2
16:48:35 Found 4 models, 3 sources, 474 macros
16:48:35
16:48:35 Concurrency: 1 threads (target='dev')
16:48:35
16:48:36 1 of 1 START sql table model analytics.session_summary ..... [RUN]
16:48:38 1 of 1 OK created sql table model analytics.session_summary ..... [SUCCESS 1 in
2.55s]
16:48:39
16:48:39 Finished running 1 table model in 0 hours 0 minutes and 3.96 seconds (3.96s).
16:48:39
16:48:39 Completed successfully
16:48:39
16:48:39 Done. PASS=1 WARN=0 ERROR=0 SKIP=0 TOTAL=1

```

a. session_summary content

```

1 with __dbt__cte__user_session_channel as (
2   SELECT
3     userId,
4     sessionId,
5     channel
6   FROM USER_DB_CHIPMUNK.raw.user_session_channel
7   WHERE sessionId IS NOT NULL
8 ), __dbt__cte__session_timestamp as (
9   SELECT
10    sessionId,
11    ts
12  FROM USER_DB_CHIPMUNK.raw.session_timestamp
13  WHERE sessionId IS NOT NULL
14 ) SELECT u.userId, u.sessionId, u.channel, st.ts
15 FROM __dbt__cte__user_session_channel u
16 JOIN __dbt__cte__session_timestamp st ON u.sessionId = st.sessionId

```

	USERID	SESSIONID	CHANNEL	TS
1	779	7cdace91c487558e27ce54df7cdb299c	Instagram	2019-05-01T00:13:11.783Z
2	230	94f192dee566b018e0acf31e1f99a2d9	Naver	2019-05-01T00:49:46.073Z
3	369	7ed2d3454c5eea71148b11d0c25104ff	Youtube	2019-05-01T10:18:43.21Z
4	248	f1daf122cde86301084459363cd31db	Naver	2019-05-01T13:10:56.413Z
5	676	fd0efcca272f704a760c3b61dcc70fd0	Instagram	2019-05-01T13:45:19.793Z
6	40	8804f94e16ba5b680e239a554a08f7d2	Youtube	2019-05-01T14:23:07.66Z
7	468	c5f441cd5f43eb2f2c024e1f8b5d00cd	Instagram	2019-05-01T15:03:54.65Z
8	69	d5fcc35c94879a4afad61cacca56192c	Facebook	2019-05-01T15:13:16.14Z
9	420	3d191ef6e236bd1b9bdb9ff4743c47fe	Youtube	2019-05-01T15:33:58.197Z
10	572	c17028c9b8e0c5deaad29665d582284a	Organic	2019-05-01T15:59:57.49Z
11	253	cd0b43eac0392accf3624b7372dec36e	Facebook	2019-05-01T16:33:03.463Z
12	87	0a4bbceda17a6253386bc9eb45240e25	Youtube	2019-05-01T17:10:29.417Z
13	277	c67ba7c4c5c0cd4cc3e3a7146fe5c015	Naver	2019-05-01T17:42:13.58Z
14	97	d8a3a3c3234392b0add43c5f9c05a246	Organic	2019-05-01T18:18:00.407Z

4. Add Snapshot to the output table (session_summary)

```
C:\Users\aiish\OneDrive\Desktop\build_mau>dbt snapshot
17:05:53 Running with dbt=1.9.3
17:05:54 Registered adapter: snowflake=1.9.2
17:05:57 Found 3 models, 1 snapshot, 3 sources, 474 macros
17:05:57
17:05:57 Concurrency: 1 threads (target='dev')
17:05:57
17:06:00 1 of 1 START snapshot snapshot.snapshot_session_summary .....
..... [RUN]
17:06:03 1 of 1 OK snapshotted snapshot.snapshot_session_summary .....
..... [SUCCESS 1 in 2.84s]
17:06:03
17:06:03 Finished running 1 snapshot in 0 hours 0 minutes and 6.36 seconds (6.36s).
17:06:03
17:06:03 Completed successfully
17:06:03
17:06:03 Done. PASS=1 WARN=0 ERROR=0 SKIP=0 TOTAL=1
```

a. snapshot_session_summary content

```
1  {% snapshot snapshot_session_summary %}
2  {{
3    config(
4      target_schema='snapshot',
5      unique_key='sessionId',
6      strategy='timestamp',
7      updated_at='ts',
8      invalidate_hard_deletes=True
9    )
10 }}
11 SELECT * FROM {{ ref('session_summary') }}
12 {% endsnapshot %}
```

```

61
62 SELECT *
63 FROM snapshot.snapshot_session_summary
64 LIMIT 10;
65
66

```

	USER	SESSIONID	CHANN	TS	DBT_SCD_ID	DBT_UPDATED_AT	DBT_VALID_FROM
1	779	7cdace91c487558e27ce54df7cdb299c	Instagram	2019-05-01 00:13:11.783	968a20fba603a2f1669abf0e39255815	2019-05-01 00:13:11.783	2019-05-01 00:13:11
2	230	94f192dee566b018e0ac31e1f99a2d9	Naver	2019-05-01 00:49:46.073	d1484d42374543038d705da14d6259ab	2019-05-01 00:49:46.073	2019-05-01 00:49:46
3	369	7ed2d3454c5eea71148b11d0c25104ff	Youtube	2019-05-01 10:18:43.210	8fdec2c89c86d082fe65756e64c7b83	2019-05-01 10:18:43.210	2019-05-01 10:18:43
4	248	f1daf122cde863010844459363cd31db	Naver	2019-05-01 13:10:56.413	3140d8abcb37b1cf782c3fd72685fe0	2019-05-01 13:10:56.413	2019-05-01 13:10:56
5	676	fd0efcca272f704a760c3b61dcc70fd0	Instagram	2019-05-01 13:45:19.793	5d6a0aab72232c7f67f4420e2a870bd7	2019-05-01 13:45:19.793	2019-05-01 13:45:19
6	40	8804f94e16ba5b680e239a554a08f7d2	Youtube	2019-05-01 14:23:07.660	c068e54e00defd8eddb661f6ce9e8035	2019-05-01 14:23:07.660	2019-05-01 14:23:07
7	468	c5f441cd5f43eb2f2c024e1f8b5d00cd	Instagram	2019-05-01 15:03:54.650	d73ce4e551a457d41304cbda7fc0b4c3	2019-05-01 15:03:54.650	2019-05-01 15:03:54
8	69	d5fcc35c94879a4afad61cacca56192c	Facebook	2019-05-01 15:13:16.140	1ea183e4b8ebab807df02bbdbc93a3b	2019-05-01 15:13:16.140	2019-05-01 15:13:16
9	420	3d191fe6e236bd1b9bdb9ff4743c47fe	Youtube	2019-05-01 15:33:58.197	5536b40dc468a35cb851a422e8ae6a14	2019-05-01 15:33:58.197	2019-05-01 15:33:58
10	572	c17028c9b6e0c5deaad29e65d582284a	Organic	2019-05-01 15:59:57.490	439009581c657cd1ac0345f80e8c5e90	2019-05-01 15:59:57.490	2019-05-01 15:59:57

Query Details

Query duration 326ms

Rows 10

Query ID 07bb5602-0104-d352-0...

Show more

USERID

40 779

SESSIONID

Ask Copilot

5. Add at least 2 tests to sessionId field of session_summary

```

C:\Users\aiish\OneDrive\Desktop\build_mau>dbt test
17:18:56 Running with dbt=1.9.3
17:18:56 Registered adapter: snowflake=1.9.2
17:18:57 Found 3 models, 1 snapshot, 3 data tests, 3 sources, 474 macros
17:18:57
17:18:57 Concurrency: 1 threads (target='dev')
17:18:57
17:18:58 1 of 3 START test not_null_session_summary_sessionId .....
..... [RUN]
17:18:59 1 of 3 PASS not_null_session_summary_sessionId .....
..... [PASS in 1.02s]
17:18:59 2 of 3 START test not_null_session_summary_userId .....
..... [RUN]
17:18:59 2 of 3 PASS not_null_session_summary_userId .....
..... [PASS in 0.44s]
17:18:59 3 of 3 START test unique_session_summary_sessionId .....
..... [RUN]
17:19:00 3 of 3 PASS unique_session_summary_sessionId .....
..... [PASS in 0.81s]
17:19:00
17:19:00 Finished running 3 data tests in 0 hours 0 minutes and 3.38 seconds (3.38s).
17:19:00
17:19:00 Completed successfully
17:19:00
17:19:00 Done. PASS=3 WARN=0 ERROR=0 SKIP=0 TOTAL=3

```

```
1  version: 2
2
3  models:
4    - name: session_summary
5      description: "Combines user session channel data with timestamps"
6      columns:
7        - name: userId
8          description: "The user identifier"
9          tests:
10           - not_null
11
12        - name: sessionId
13          description: "Unique identifier for each session"
14          tests:
15           - unique
16           - not_null
```

Lineage Graph



[GitHub Link](#)