

Homework10_AparnaBharathiSuresh

1.Create Database 'trip_db', Schema 'trip_schema' and Table 'trip_data' in Snowflake.

Query:

Query:

```
create table trip_data (
    tripduration integer,
    starttime timestamp,
    stoptime timestamp,
    start_station_id integer,
    start_station_name string,
    start_station_latitude float,
    start_station_longitude float,
    end_station_id integer,
    end_station_name string,
    end_station_latitude float,
    end_station_longitude float,
    bikeid integer,
    membership_type string,
    usertype string,
    birth_year integer,
    gender integer
);
```

The screenshot shows the Snowflake web interface. On the left, there's a sidebar with navigation links: Search, Projects, Data (with Databases selected), Data Products, AI & ML, Monitoring, and Admin. Below the sidebar, there's a promotional banner for a \$400 trial and a blue Upgrade button. The main area is titled "Databases" and shows a list of existing databases: SNOWFLAKE and SNOWFLAKE_SAMPLE_DATA. A modal window titled "New Database" is open, prompting for a database name ("trip_db") and a comment ("Assignment10"). The "Create" button is visible at the bottom of the modal. The top of the page has several tabs and a header bar.

This screenshot shows the same Snowflake interface after the database creation. The "Databases" list now includes the newly created database "TRIP_DB". The table structure is as follows:

NAME	SOURCE	OWNER	CREAT...
SNOWFLAKE	Share	—	2 days ...
SNOWFLAKE_SAMPLE_DATA	Share	ACCOUNTADMIN	2 days ...
TRIP_DB	Local	ACCOUNTADMIN	just now

The screenshot shows the Snowflake Worksheet interface. The top navigation bar includes tabs for "Homework 10", "Getting Started with Snowflake", "Shreenithi Sivakumar's Zoom", "Activate your Snowflake account", and "2024-04-24 7:19pm - Snowflake". The main area has a title bar "Load sample data with SQL..." and the date "2024-04-24 7:19pm". A sidebar on the left lists "Databases" and "Worksheets", with "Databases" selected. Under "Databases", there are sections for "Pinned (0)", "No pinned objects", and a search bar. Below this are links to "SNOWFLAKE", "SNOWFLAKE_SAMPLE_DATA", and "TRIP_DB". The central workspace contains a single query: "select col from table where created = :daterange". The bottom right corner features an "Ask Copilot" button.

The screenshot shows the "TRIP_DB | Database" page in the Snowflake interface. The top navigation bar is identical to the previous screenshot. The main content area displays "Database Details" for "TRIP_DB". It shows the database was created by "ACCOUNTADMIN" 9 minutes ago, is local, and assigned to "Assignment10". The "Schemas" tab is selected, showing three schemas: "INFORMATION_SCHEMA", "PUBLIC", and "TRIP_SCHEMA". The "INFORMATION_SCHEMA" and "PUBLIC" schemas were created by "ACCOUNTADMIN" 9 minutes ago. The "TRIP_SCHEMA" was created by "ACCOUNTADMIN" just now. On the left sidebar, under "Data", the "Databases" option is selected. A promotional banner on the left side offers "\$400 of \$400 left" with an "Upgrade" button and a message "27 days left in trial". The bottom left corner shows the user information "AS Aparna Suresh ACCOUNTADMIN".

The screenshot shows the Snowflake SQL interface. In the top navigation bar, there are several tabs: Homework 10, OnlineZTS_LabGuide.pdf, Getting Started with Snowflake, Shreenithi Sivakumar's Z, Activate your Snowflake, and 2024-04-24 7:19pm - Snowflake. The main area displays a SQL query in the code editor:

```

1 create table trip_data (
2     tripduration integer,
3     starttime timestamp,
4     stoptime timestamp,
5     start_station_id integer,
6     start_station_name string,
7     start_station_latitude float,
8     start_station_longitude float,
9     end_station_id integer,
10    end_station_name string,
11    end_station_latitude float,
12    end_station_longitude float,
13    bikeid integer,
14    membership_type string,
15    usertype string,
16    birth_year integer,
17    gender integer
18 );

```

The results pane shows the status of the query:

```

status
1 Table TRIP_DATA successfully created.

```

On the right side, there are "Query Details" and "Table Details" sections.

The screenshot shows the Snowflake UI for the TRIP_DB database. The left sidebar shows the navigation menu with "Data" selected. The main pane displays the table definition for TRIP_DATA:

```

1 create or replace TABLE TRIP_DB.TRIP_SCHEMA.TRIP_DATA (
2     TRIPDURATION NUMBER(38, 0),
3     STARTTIME TIMESTAMP_NTZ(9),
4     STOPTIME TIMESTAMP_NTZ(9),
5     START_STATION_ID NUMBER(38, 0),
6     START_STATION_NAME VARCHAR(16777216),
7     START_STATION_LATITUDE FLOAT,
8     START_STATION_LONGITUDE FLOAT,
9     END_STATION_ID NUMBER(38, 0),
10    END_STATION_NAME VARCHAR(16777216),
11    END_STATION_LATITUDE FLOAT,
12    END_STATION_LONGITUDE FLOAT,
13    BIKEID NUMBER(38, 0),
14    MEMBERSHIP_TYPE VARCHAR(16777216),
15    USERTYPE VARCHAR(16777216),
16    BIRTH_YEAR NUMBER(38, 0),
17    GENDER NUMBER(38, 0)
18 );

```

The "Privileges" section shows "ACCOUNT... (Current Role)" and a search bar for "OWNERSHIP".

2. Create Stage 'trip_stage' in Snowflake.

Query:

list @TRIP_STAGE;

The screenshot shows the Snowflake interface with the following details:

- Left Sidebar:** Shows the user's account information (Aparna Suresh, ACCOUNTADMIN), a trial upgrade notice (\$400 of \$400 left), and a 27-day trial remaining.
- Central Panel:**
 - Database:** TRIP_DB
 - Schema:** TRIP_SCHEMA
 - Stage:** TRIP_STAGE
- Details View:** Displays the Stage Details for TRIP_STAGE, showing one file named "trips_2013_0,0,0.csv".

The screenshot shows the results of a query against the TRIP_STAGE stage:

```

4   start_time timestamp,
5   start_station_id integer,
6   start_station_name string,
7   start_station_latitude float,
8   start_station_longitude float,
9   end_station_id integer,
10  end_station_name string,
11  end_station_latitude float,
12  end_station_longitude float,
13  bikeid integer,
14  member_casual string,
15  usertype string,
16  birth_year integer,
17  gender integer
18
19
20
21  list @TRIP_STAGE
22
23

```

Results Table:

name	size	md5	last_modified
s3://snowflake-workshop-lab/citibike-trips-csv/trips_2013_0,2,0.csv.gz	2628711	ca9971f77588ec5e1204971d51380e11	Sun, 10 Jul 2022 18:19:42 GMT
s3://snowflake-workshop-lab/citibike-trips-csv/trips_2013_0,3,0.csv.gz	2331971	c0f4471b830632w85e5c50ef0b4a4e6	Sun, 10 Jul 2022 18:19:42 GMT
s3://snowflake-workshop-lab/citibike-trips-csv/trips_2013_0,4,0.csv.gz	2554446	9612401ccdf4ff4a49181762a869801e3	Sun, 10 Jul 2022 18:19:42 GMT
s3://snowflake-workshop-lab/citibike-trips-csv/trips_2013_0,5,0.csv.gz	2477099	e#79b01c545f7ec3f513b5f97ee03	Sun, 10 Jul 2022 18:19:42 GMT
s3://snowflake-workshop-lab/citibike-trips-csv/trips_2013_0,6,0.csv.gz	2569322	b178b135d9:07e859094a4051c800	Sun, 10 Jul 2022 18:19:42 GMT
s3://snowflake-workshop-lab/citibike-trips-csv/trips_2013_0,7,0.csv.gz	2572983	e#038436d9cf504491b4e0bb3e9388	Sun, 10 Jul 2022 18:19:42 GMT
s3://snowflake-workshop-lab/citibike-trips-csv/trips_2014_0,0,0.csv.gz	3039741	13a5f4587723412190n7438109x2c0	Sun, 10 Jul 2022 18:19:42 GMT
s3://snowflake-workshop-lab/citibike-trips-csv/trips_2014_0,1,0.csv.gz	2965346	0c235617958549c1bd6c16840b10c6	Sun, 10 Jul 2022 18:19:42 GMT
s3://snowflake-workshop-lab/citibike-trips-csv/trips_2014_0,2,0.csv.gz	2945099	b37250738858a1ff1283ee0f0d7	Sun, 10 Jul 2022 18:19:42 GMT
s3://snowflake-workshop-lab/citibike-trips-csv/trips_2014_0,3,0.csv.gz	2999395	69c799e5469582ce0e4441c1265062	Sun, 10 Jul 2022 18:19:42 GMT
s3://snowflake-workshop-lab/citibike-trips-csv/trips_2014_0,4,0.csv.gz	3091874	0a5332769873x9124obd3f39a83a5	Sun, 10 Jul 2022 18:19:42 GMT
s3://snowflake-workshop-lab/citibike-trips-csv/trips_2014_0,5,0.csv.gz	3088567	3ef3f3a7e9ec03b6d43f3532713213	Sun, 10 Jul 2022 18:19:42 GMT
s3://snowflake-workshop-lab/citibike-trips-csv/trips_2014_0,6,0.csv.gz	3009295	a#84841cc0a4033a252505c750e1d5	Sun, 10 Jul 2022 18:19:42 GMT
s3://snowflake-workshop-lab/citibike-trips-csv/trips_2014_0,7,0.csv.gz	3351010	85078e46deaf1da8bc44dc53959f	Sun, 10 Jul 2022 18:19:42 GMT
s3://snowflake-workshop-lab/citibike-trips-csv/trips_2014_1,0,0.csv.gz	3022861	810eae0d4408d1029ea37fe137edc	Sun, 10 Jul 2022 18:19:42 GMT
s3://snowflake-workshop-lab/citibike-trips-csv/trips_2014_1,1,0.csv.gz	3899159	e#81140215709eb0d2380b01f80c7447	Sun, 10 Jul 2022 18:19:42 GMT
s3://snowflake-workshop-lab/citibike-trips-csv/trips_2014_1,2,0.csv.gz	4287992	4#d185833d34fdfe442d3540f996	Sun, 10 Jul 2022 18:19:42 GMT

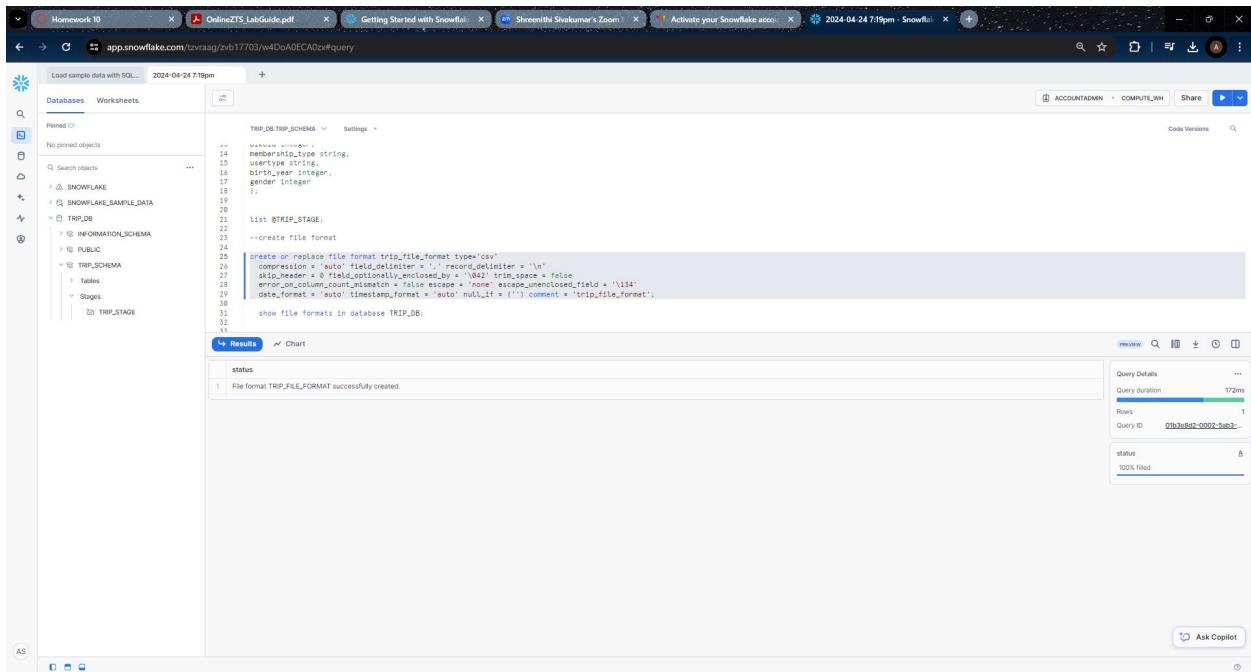
3. Create File Format 'trip_file_format' in the Snowflake.

```

create or replace file format trip_file_format type='csv'
compression = 'auto' field_delimiter = ',' record_delimiter = '\n'
skip_header = 0 field_optionally_enclosed_by = '\042' trim_space = false
error_on_column_count_mismatch = false escape = 'none' escape_unenclosed_field =
'\134'

```

```
date_format = 'auto' timestamp_format = 'auto' null_if = ('') comment = 'trip_file_format';
```



The screenshot shows the Snowflake SQL editor interface. The left sidebar displays the database structure: TRIP_DB.TRIP_SCHEMA contains a script for creating a file format named TRIP_FILE_FORMAT. The script includes fields like membership_type, user_type, birth_year, gender, and a list of stages. The right side shows the results of the query, indicating the file format was successfully created. The status bar at the bottom shows the query duration as 172ms and the number of rows as 1.

```
TRIP_DB.TRIP_SCHEMA
...
24
25     create or replace file format trip_file_format type='csv'
26         compression = 'auto' field_delimiter = ',' record_delimiter = '\n'
27         skip_header = 0 field_optionally_enclosed_by = '\"\"'
28         error_on_column_count_mismatch = false escape = '\"'
29         date_format = 'auto' timestamp_format = 'auto' null_if = ('') comment = 'trip_file_format';
30
31     show file formats in database TRIP_DB;
32
33
```

Results

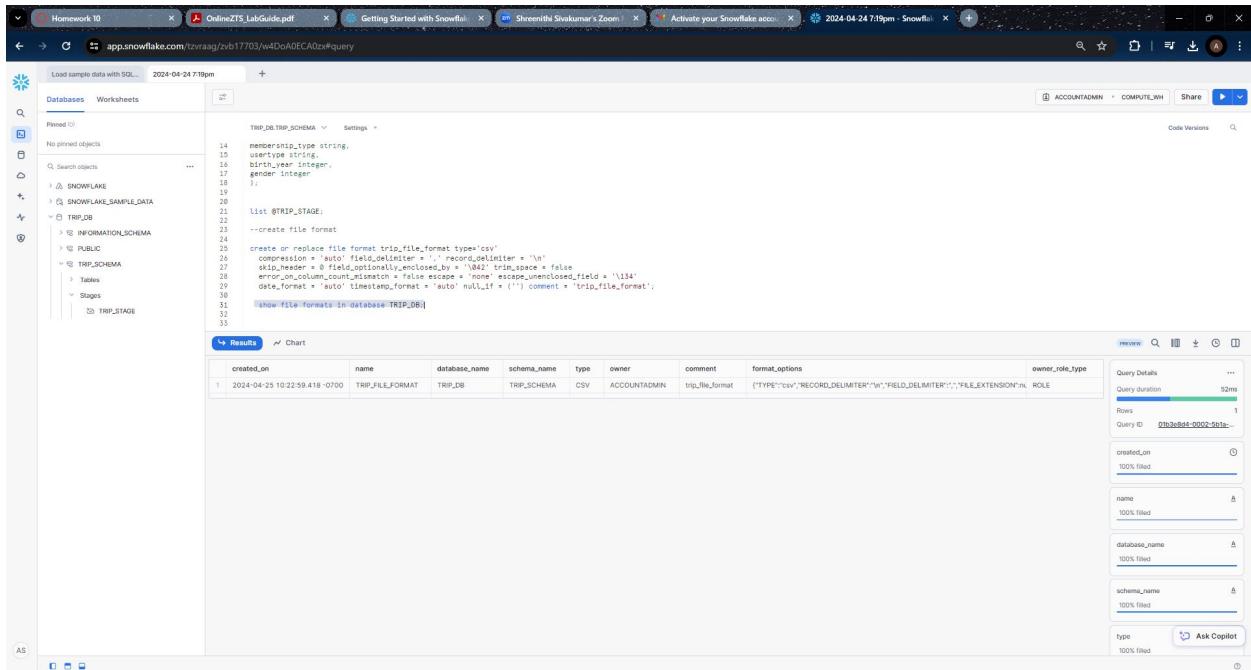
status
1 File format TRIP_FILE_FORMAT successfully created.

Query Details

- Query duration: 172ms
- Rows: 1
- Query ID: 0b3e8d2-0002-5ab2-

status

100% filled



This screenshot shows the same process as the first one, but with additional options added to the file format creation script. The options include 'skip_header = 0', 'field_optionally_enclosed_by = '\"\"', and 'error_on_column_count_mismatch = false'. The rest of the schema and table definitions remain the same.

```
TRIP_DB.TRIP_SCHEMA
...
24
25     create or replace file format trip_file_format type='csv'
26         compression = 'auto' field_delimiter = ',' record_delimiter = '\n'
27         skip_header = 0 field_optionally_enclosed_by = '\"\"'
28         error_on_column_count_mismatch = false escape = '\"'
29         date_format = 'auto' timestamp_format = 'auto' null_if = ('') comment = 'trip_file_format';
30
31     show file formats in database TRIP_DB;
32
33
```

Results

created_on	name	database_name	schema_name	type	owner	comment	format_options	owner_role_type
2024-04-24 10:22:59.418 -0700	TRIP_FILE_FORMAT	TRIP_DB	TRIP_SCHEMA	CSV	ACCOUNTADMIN	trip_file_format	{"TYPE":"csv","RECORD_DELIMITER":"\n","FIELD_DELIMITER":",","FILE_EXTENSION":in_, ROLE}	

Query Details

- Query duration: 52ms
- Rows: 1
- Query ID: 0b3e8d4-0002-5ab2-

created_on

name

database_name

schema_name

type

4. Load Data into the 'trip_data' table.

copy into trip_data from @TRIP_STAGE file_format=trip_file_format PATTERN = '*csv.*';

```

    Load sample data with SQL... 2024-04-24 7:19pm + ACCOUNTADMIN COMPUTE_WH Share ▾
    Databases Worksheets TRIP_DB.TRIP_SCHEMA Settings ▾
    No pinned objects
    Search objects ...
    > ⚑ SNOWFLAKE ...
    > ⚑ TRIP_DB ...
    > INFORMATION_SCHEMA ...
    > PUBLIC ...
    > TRIP_SCHEMA ...
        Tables
            TRIP_DATA ...
        Stages
            TRIP_STAGE ...
    Copy into trip_data from @TRIP_STAGE file_format=trip_file_format PATTERN = '*csv.*';
    16 birth_year integer;
    17 gender integer;
    18 ...
    19 ...
    20 ...
    21 ...
    22 ...
    23 ...
    24 ...
    25 ...
    26 ...
    27 ...
    28 ...
    29 ...
    30 ...
    31 ...
    32 ...
    33 ...
    34 ...
    35 ...

    Results ▾ Chart
    Query Details ...
    Query duration 1m 2s
    Rows 377
    Query ID gbae909-0002-9eab-...
    File 100% filled
    status LOADED 377
    rows_parsed 7756
    rows_loaded 27964
    Ask Copilot
  
```

TRIPID	STARTTIME	STOPTIME	START_STATION_ID	START_STATION_NAME	START_STATION_LATITUDE	START_STATION_LONGITUDE	END_STATION_ID	END_STATION_NAME
695	2013-06-01 00:00:01	2013-06-01 00:11:36	444	Broadway & W 24 St	40.7423543	-73.98915076	434	9 Ave
693	2013-06-01 00:00:08	2013-06-01 00:11:41	444	Broadway & W 24 St	40.7423543	-73.98915076	434	9 Ave
2059	2013-06-01 00:00:44	2013-06-01 00:35:03	406	Hicks St & Montague St	40.69512645	-73.99595065	408	Hicks
123	2013-06-01 00:01:04	2013-06-01 00:03:07	475	E 15 St & Living Pl	40.73524278	-73.98758581	282	Wash
1521	2013-06-01 00:01:22	2013-06-01 00:28:43	2005	Little West St & 1 Pl	40.70589254	-74.01677685	310	State
2028	2013-06-01 00:01:47	2013-06-01 00:35:35	485	W 37 St & 5 Ave	40.75038009	-73.9838988	406	Hicks
2057	2013-06-01 00:02:23	2013-06-01 00:36:50	285	Broadway & E 14 St	40.73454587	-73.99074142	532	S 5 Pl
139929	2013-06-01 00:02:39	2013-06-01 02:54:48	432	E 7 St & Avenue A	40.72621788	-73.98379855	402	Broad
369	2013-06-01 00:03:29	2013-06-01 00:09:38	509	9 Ave & W 22 St	40.7454973	-74.00197139	521	8 Ave
512	2013-06-01 00:03:36	2013-06-01 01:12:08	309	Murray St & West St	40.7149787	-74.01301212	300	Shev
504	2013-06-01 00:03:45	2013-06-01 01:12:10	309	Murray St & West St	40.7149787	-74.01301212	347	Green
1829	2013-06-01 00:03:47	2013-06-01 01:36:16	265	Stanton St & Chrystie St	40.72229946	-73.99147135	436	Hanc
829	2013-06-01 00:04:22	2013-06-01 01:08:11	404	9 Ave & W 14 St	40.7405926	-74.00550867	303	Merx
1316	2013-06-01 00:04:22	2013-06-01 00:26:24	423	W 54 St & 9 Ave	40.76584491	-73.98909036	314	Cadm
1456	2013-06-01 00:04:41	2013-06-01 02:08:57	501	Henry St & Grand St	40.71421915	-73.981346	532	S 5 Pl
388	2013-06-01 00:05:13	2013-06-01 01:11:39	241	Dekabt Ave & Portland Ave	40.68981035	-73.97493121	385	Fulton
924	2013-06-01 00:05:21	2013-06-01 02:20:45	486	Broadway & W 29 St	40.7482009	-73.9885723	521	8 Ave
1233	2013-06-01 00:06:44	2013-06-01 02:27:17	527	E 33 St & 2 Ave	40.7440223	-73.9786058	298	Divisi
833	2013-06-01 00:07:29	2013-06-01 02:21:22	503	E 20 St & Park Ave	40.73827428	-73.98751688	503	E 20 St
1818	2013-06-01 00:08:10	2013-06-01 03:08:28	257	Lipendorf St & Broadway	40.71939226	-74.00247214	500	Broad
553	2013-06-01 00:08:47	2013-06-01 01:08:00	489	10 Ave & W 28 St	40.75066386	-74.00176802	448	W 37
682	2013-06-01 00:08:53	2013-06-01 01:08:15	486	Broadway & W 29 St	40.7482009	-73.9885723	521	8 Ave
899	2013-06-01 00:09:25	2013-06-01 02:24:24	494	W 20 St & 8 Ave	40.74734825	-73.99723551	494	W 26
626	2013-06-01 00:09:52	2013-06-01 03:02:18	380	W 4 St & 7 Ave S	40.73401143	-74.00209387	382	Unive
219	2013-06-01 00:10:28	2013-06-01 01:04:07	309	E 58 St & 3 Ave	40.76095796	-73.96724467	338	Christ
339	2013-06-01 00:10:45	2013-06-01 01:16:24	480	W 53 St & 10 Ave	40.7666971	-73.99061728	432	E 7 St
511	2013-06-01 00:10:50	2013-06-01 01:19:21	223	W 13 St & 7 Ave	40.73781509	-73.999494661	345	W 13
97	2013-06-01 00:11:00	2013-06-01 01:12:37	453	W 22 St & 6 Ave	40.74475148	-73.9915362	453	W 22
547	2013-06-01 00:11:04	2013-06-01 00:20:11	432	E 7 St & Avenue A	40.72621788	-73.98379855	null	null
198	2013-06-01 00:11:10	2013-06-01 04:28	476	E 31 St & 3 Ave	40.74384314	-73.97966069	461	E 20 St

5. Analyze the performance of the node upon changing configurations from x-small to x-large in Snowflake.

The performance was little slow when the warehouse size was X-small and comparatively higher when the size was changed to X-large.

Query	Time taken in X-small warehouse	Time Taken in X-large warehouse
Copy Query to load the data	1 min 2 seconds	7.6 seconds
Select Query	3.4 seconds	1.8 seconds

For larger data loading operations or more complicated queries, a larger warehouse can be used.

Screenshots of X-Small:

The screenshot shows the Snowflake web interface with the URL <https://app.snowflake.com/tzvraag/cyb17703/4/compute/warehouses?whotype=all&status=all&size=all&columns=name%2Cstatus%2Csize%2Ctype%2Cclusters%2Crunning%2Cqueued%2Cowner%2ClastResumedTime>. The left sidebar has 'Warehouses' selected under 'Admin'. The main area displays a table titled 'Warehouses' with one row:

NAME	SIZE	STATUS	RUNNING	QUEUED	OWNER	RESUMED
COMPUTE_WH	(S)	Started	0	0	ACCOUNTADMIN	10 minutes ago

Below the table, there's a message: '\$400 of \$400 left: ⏪ ...' with a 'Upgrade' button, and '27 days left in trial'.

Query - 01b3e909-0002-5ab4-005d-9b070001b032

Details

Status	Success	Duration	1m 2s	Driver Status	Supported
Start Time	4/25/2024, 11:17:58 AM	Query ID	01b3e909-0002-5ab4-005d-9b070001b032	Client Driver	Go 11.5
End Time	4/25/2024, 11:19:00 AM	Query Tag	—	Session ID	26347627201310734
Warehouse Size	X-Small				

SQL Text

```
copy into trip_data from @TRIP_STAGE file_format=trip_file_format PATTERN = '*.csv';
```

Results 377 of 377 Rows × Cluster 1

file	status	rows_parsed	rows_loaded	error_limit	errors_seen	first_error	first_error_line	first_error_character	first_error_column_name
1 s3://snowflake-workshop-lab/cobike-trips/csv/trips_2013_0_0.csv.gz	LOADED	103,700	103,700	1	0	null	null	null	null
2 s3://snowflake-workshop-lab/cobike-trips/csv/trips_2013_1_5_0.csv.gz	LOADED	93,833	93,833	1	0	null	null	null	null
3 s3://snowflake-workshop-lab/cobike-trips/csv/trips_2013_1_4_0.csv.gz	LOADED	79,328	79,328	1	0	null	null	null	null
4 s3://snowflake-workshop-lab/cobike-trips/csv/trips_2013_1_3_0.csv.gz	LOADED	114,248	114,248	1	0	null	null	null	null
5 s3://snowflake-workshop-lab/cobike-trips/csv/trips_2013_1_3_0.csv.gz	LOADED	111,697	111,697	1	0	null	null	null	null
6 s3://snowflake-workshop-lab/cobike-trips/csv/trips_2013_5_5_0.csv.gz	LOADED	108,605	108,605	1	0	null	null	null	null
7 s3://snowflake-workshop-lab/cobike-trips/csv/trips_2013_5_5_0.csv.gz	LOADED	82,969	82,969	1	0	null	null	null	null
8 s3://snowflake-workshop-lab/cobike-trips/csv/trips_2014_1_5_0.csv.gz	LOADED	103,229	103,229	1	0	null	null	null	null
9 s3://snowflake-workshop-lab/cobike-trips/csv/trips_2014_1_5_0.csv.gz	LOADED	135,516	135,516	1	0	null	null	null	null

Query - 01b3e937-0002-5b18-005d-9b070001a112

Details

Status	Success	Duration	3.4s	Driver Status	Supported
Start Time	4/25/2024, 12:03:28 PM	Query ID	01b3e937-0002-5b18-005d-9b070001a112	Client Driver	Go 11.5
End Time	4/25/2024, 12:03:32 PM	Query Tag	—	Session ID	26347627201310734
Warehouse Size	X-Small				

SQL Text

```
select date_trunc('hour', starttime) as "date", count(*) as "num trips", avg(tripduration)/60 as "avg duration (mins)", avg(haversine(start_station.latitude, start_station.longitude, end_station.latitude, end_station.longitude)) as "avg distance (km)" from trip_data group by 1 order by 1;
```

Results 10,000 of 44,295 Rows × Cluster 1

date	num trips	avg duration (mins)	avg distance (km)
2013-06-01 00:00:00	304	56.058429833333	—
2013-06-01 01:00:00	204	26.525163400000	—
2013-06-01 02:00:00	134	36.119005000000	—
2013-06-01 03:00:00	82	44.485365850000	—
2013-06-01 04:00:00	32	23.278125000000	—
2013-06-01 05:00:00	26	34.584615383333	—
2013-06-01 06:00:00	80	22.397500000000	—
2013-06-01 07:00:00	186	66.397849466667	—

Screenshot of the Snowflake SQL interface showing a query editor and results panel.

Query Editor:

```

-- Load sample data with SQL
-- 2024-04-24 7:19pm
-- 2024-04-25 11:32am

TRIP_DB.TRIP_SCHEMA
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37

```

Results Panel:

date	num trips	avg duration (mins)	avg distance (km)
2013-06-01 00:00:00.000	304	56.05844298333	2.127917476
2013-06-01 01:00:00.000	204	28.52518340000	2.087906273
2013-06-01 02:00:00.000	134	36.11980050000	2.31744827
2013-06-01 03:00:00.000	82	44.88538585000	2.349126632
2013-06-01 04:00:00.000	32	23.27812500000	1.840026007
2013-06-01 05:00:00.000	26	34.58461538333	3.337844489
2013-06-01 06:00:00.000	80	22.39750000000	2.832927896
2013-06-01 07:00:00.000	186	66.39784946667	2.691774983
2013-06-01 08:00:00.000	354	18.53022983333	2.244399992
2013-06-01 09:00:00.000	560	40.26101190000	2.292639558
2013-06-01 10:00:00.000	750	28.87346666667	2.225768226
2013-06-01 11:00:00.000	946	35.83620085000	2.26058226
2013-06-01 12:00:00.000	1208	33.87635210000	2.177910979
2013-06-01 13:00:00.000	1276	48.62209300000	2.270395192
2013-06-01 14:00:00.000	1376	34.74004800000	2.271991408
2013-06-01 15:00:00.000	1368	47.81639668333	2.221532454
2013-06-01 16:00:00.000	1386	51.48276383333	2.265854068

Role Selection: ACCOUNTADMIN (Default) → COMPUTE_WH (X-Small)

Query Details:

- Query duration: 3.4s
- Rows: 44.3K
- Query ID: qb3d37-0002-3b39-
- Date range: 2013-06-01 to 2019-06-30
- num trips distribution: 1 to 8810
- avg duration (mins) distribution: 2.211116667 to 66.55662275
- avg distance (km) distribution: 0 to 66.55662275

Change Datawarehouse size to X-Large:

Screenshot of the Snowflake Admin interface showing the Warehouses page.

Warehouses:

NAME	SL...	STATUS	RUNNING	QUEUED	OWNER	RESUMED
COMPUTE_WH	...	Started	1	0	ACCOUNTADMIN	10 minutes ago

Cost Management:

- \$400 of \$400 left: Upgrade
- 27 days left in trial

User & Roles:

- Aparna Suresh (ACCOUNTADMIN)

Query - 01b3e922-0002-5b88-005d-9b07000190be

Query Details

Details

Status	Success	Duration	7.6s
Start Time	4/25/2024, 11:42:58 AM	Query ID	01b3e922-0002-5b88-005d-9b07000190be
End Time	4/25/2024, 11:43:06 AM	Query Tag	—
Warehouse Size	X-Large	Session ID	26347627201314870

SQL Text

```
copy into trip_data from @TRIP_STAGE file_format=trip_file_format PATTERN = '*.csv' ;
```

Results 376 of 376 Rows × Cluster 1

file	status	rows_parsed	rows_loaded	error_limit	errors_seen	first_error	first_error_line	first_error_character	first_error_column_name
1 s3://snowflake-workshop-lab/citibike-trips/csv/trips_2013_3_6_0.csv.gz	LOADED	85,049	85,049	1	0	null	null	null	null
2 s3://snowflake-workshop-lab/citibike-trips/csv/trips_2014_6_3_0.csv.gz	LOADED	153,370	153,370	1	0	null	null	null	null
3 s3://snowflake-workshop-lab/citibike-trips/csv/trips_2018_4_7_0.csv.gz	LOADED	128,382	128,382	1	0	null	null	null	null
4 s3://snowflake-workshop-lab/citibike-trips/csv/trips_2019_1_4_0.csv.gz	LOADED	169,136	168,136	1	0	null	null	null	null
5 s3://snowflake-workshop-lab/citibike-trips/csv/trips_2017_1_4_0.csv.gz	LOADED	234,783	234,783	1	0	null	null	null	null
6 s3://snowflake-workshop-lab/citibike-trips/csv/trips_2014_3_4_0.csv.gz	LOADED	153,739	153,739	1	0	null	null	null	null
7 s3://snowflake-workshop-lab/citibike-trips/csv/trips_2016_2_10_0.csv.gz	LOADED	220,372	220,372	1	0	null	null	null	null
8 s3://snowflake-workshop-lab/citibike-trips/csv/trips_2014_6_4_0.csv.gz	LOADED	143,919	143,919	1	0	null	null	null	null
9 s3://snowflake-workshop-lab/citibike-trips_csv/trips_2017_0_4_0.csv.gz	LOADED	269,455	269,455	1	0	null	null	null	null

Query - 01b3e939-0002-5ab2-005d-9b070001c112

Query Details

Details

Status	Success	Duration	1.8s
Start Time	4/25/2024, 12:05:45 PM	Query ID	01b3e939-0002-5ab2-005d-9b070001c112
End Time	4/25/2024, 12:05:47 PM	Query Tag	—
Warehouse Size	X-Large	Session ID	26347627201314870

SQL Text

```
select date_trunc('hour', starttime) as "date", count(*) as "num trips", avg(tripduration)/60 as "avg duration (mins)", avg(haversine(start_station.latitude, start_station.longitude, end_station.latitude, end_station.longitude)) as "avg distance (km)" from trip_data group by 1 order by 1;
```

Results 10,000 of 44,295 Rows × Cluster 1

date	num trips	avg duration (mins)	avg distance (km)
2013-06-01 00:00:00	152	56.05842983333	—
2013-06-01 01:00:00	102	26.525163400000	—
2013-06-01 02:00:00	67	38.119005000000	—
2013-06-01 03:00:00	41	44.485365850000	—
2013-06-01 04:00:00	16	23.278125000000	—
2013-06-01 05:00:00	13	34.584615383333	—
2013-06-01 06:00:00	40	22.397500000000	—
2013-06-01 07:00:00	93	66.397849466667	—

6. Create a copy of the Spreadsheet shared in the announcement, add a few more records, create a named range and load data into Snowflake with Fivetran.

YouTube - Sheet1

	A	B	C	D	E	F	G
1	name	brand_channel	subscribers_in_millions	primary_language	category	country	
2	T-Series	Yes	249	Hindi	Music	India	
3	MrBeast	No	184	English	Entertainment	United States	
4	Cocomelon	Yes	165	English	Education	United States	
5	Sony Entertainment Television India	Yes	162	Hindi	Entertainment	India	
6	Kids Diana Show	Yes	113	English	Entertainment	Ukraine- United States	
7	PewDiePie	No	111	English	Entertainment	Sweden	
8	Like Nastya	No	107	English	Entertainment	Russia- United States	
9	Vlad and Niki	No	101	English	Entertainment	Russia- United States	
10	Zee Music Company	Yes	99.5	Hindi	Music	India	
11	WWE	Yes	97.1	English	Sports	United States	
12	Blackpink	No	91.2	Korean	Music	South Korea	
13	Goldmines	Yes	89.5	Hindi	Film	India	
14	Dude Perfect	Yes	64.6	English	Sports	United States	
15	5-Minute Crafts	Yes	64.1	English	DIY & Crafts	United States	
16	Ariana Grande	Yes	63.4	English	Music	United States	
17	Troom Troom	Yes	61.8	English	DIY & Crafts	United States	
18	BTS	Yes	61.6	Korean	Music	South Korea	
19	El Reino Infantil	Yes	61.3	Spanish	Kids	Argentina	
20	The Jazzy Chicks	Yes	60.0	English	Music	United States	

Screenshot of the Fivetran configuration interface for connecting Google Sheets to Snowflake.

Destination: google_sheets_YouTube

Authentication Method: Authorize Service Account

Sheet URL: https://docs.google.com/spreadsheets/d/1jmpT2QJT8LsSubpNM6sOeQ36sxDBg-II2Vv

Setup Guide:

How Fivetran connects to Google Sheets depends on which authentication method you choose. Fivetran supports two authentication methods:

- User OAuth:** The User OAuth authentication method allows Fivetran read-only access to any Google Sheet your user has access to. This is for organizations that do not allow users from other domains to be added to a Google Sheet.
- Service Account:** The Service Account authentication method provides fine-grained control over access to your data. You add a unique Fivetran service account as a read-only user to specific sheets that you want to sync.

Follow these instructions to sync your Google Sheets to your destination using Fivetran.

Prerequisites:

To connect Google Sheets to Fivetran, you must have Read / Write access to a Google Sheet.

User OAuth setup instructions:

STEP 1 Select a range

Open your Google Sheet and select the range that you want added to your destination. You can change your selected range later if needed.

To select a range, you can do either of the following:

- manually select the range as shown below,
- select just the columns (for example, Sheet1!A:D). If you select just the columns, Fivetran only creates rows for up to the final row that has values in your sheet (for example, Sheet1!A:06).

Screenshot of the Fivetran configuration interface for connecting Google Sheets to Snowflake, showing a different configuration.

Destination: google_sheets_YouTube_table

Authentication Method: Authorize Service Account

Sheet URL: https://docs.google.com/spreadsheets/d/1jmpT2QJT8LsSubpNM6sOeQ36sxDBg-II2Vv

Named Range: YouTubeDataRange

Setup Guide:

How Fivetran connects to Google Sheets depends on which authentication method you choose. Fivetran supports two authentication methods:

- User OAuth:** The User OAuth authentication method allows Fivetran read-only access to any Google Sheet your user has access to. This is for organizations that do not allow users from other domains to be added to a Google Sheet.
- Service Account:** The Service Account authentication method provides fine-grained control over access to your data. You add a unique Fivetran service account as a read-only user to specific sheets that you want to sync.

Follow these instructions to sync your Google Sheets to your destination using Fivetran.

Prerequisites:

To connect Google Sheets to Fivetran, you must have Read / Write access to a Google Sheet.

User OAuth setup instructions:

STEP 1 Select a range

Open your Google Sheet and select the range that you want added to your destination. You can change your selected range later if needed.

To select a range, you can do either of the following:

- manually select the range as shown below,
- select just the columns (for example, Sheet1!A:D). If you select just the columns, Fivetran only creates rows for up to the final row that has values in your sheet (for example, Sheet1!A:06).

File YouTube - Google Sheets Getting Started with Shreenithi Sivakumar Activate your Snowflake Partner Connect Dashboard | Fivetran Data | Cloud | MongoDB YouTube - Google Sheets

Connect destination Connect source Complete initial sync

Your 14-day trial starts once the initial sync is complete

Sjsu_244849894843949...

Connectors Transformations Destinations

Trial usage estimator

Alerts

Account Settings +

Resources & Support +

Aparna +

Google Sheets

Connection tests:

- Finding specified sheet
- Validating named range

All connection tests passed!

Back Continue

How was connecting to this source?

File YouTube - Google Sheets Getting Started with Shreenithi Sivakumar Activate your Snowflake GOOGLE SHEETS_YOUTUBE Dashboard | Fivetran Data | Cloud | MongoDB YouTube - Google Sheets

app.snowflake.com/tzvraag/zvb17703/#/data/databases/PC_FIVETRAN_DB/schemas/GOOGLE_SHEETS_YOUTUBE/table/GOOGLE_SHEETS_YOUTUBE_TABLE/data-preview

Q Search

Projects

Data

Databases

PC_FIVETRAN_DB

GOOGLE_SHEETS_YOUTUBE

Tables

INFORMATION_SCHEMA

PUBLIC

SNOW-LAKE

SNOW-LAKE_SAMPLE_DATA

TRP_DB

TRP_DB1

\$399 of \$400 left: 27 days left in trial

Upgrade

AS Aparna Suresh ACCOUNTADMIN

Table Details Columns Data Preview Copy History

COMPUTE_WH | 19 Rows • Updated just now

_ROW	_FIVETRAN_SYNCED	COUNTRY	NAME	PRIMARY_LANGUAGE	BRAND_CHANNEL	CATEGORY	SUBSCRIBERS_IN_MILLIONS
1	11	2024-04-25 21:29:29	South Korea	Korean	No	Music	91.2
2	6	2024-04-25 21:29:29	Sweden	PerDiePie	No	Entertainment	111
3	2	2024-04-25 21:29:29	United States	MrBeast	No	Entertainment	184
4	7	2024-04-25 21:29:29	Russia - United States	Lika Nastya	No	Entertainment	107
5	18	2024-04-25 21:29:29	Argentina	Ei Reino Infantil	Spanish	Kids	61.3
6	15	2024-04-25 21:29:29	United States	Ariana Grande	English	Music	63.4
7	5	2024-04-25 21:29:29	Ukraine - United States	Kids Diana Show	English	Entertainment	113
8	16	2024-04-25 21:29:29	United States	Trom Trom	English	DIY & Crafts	61.8
9	10	2024-04-25 21:29:29	United States	WWE	English	Sports	97.1
10	8	2024-04-25 21:29:29	Russia - United States	Vlad and Niki	English	Entertainment	101
11	13	2024-04-25 21:29:29	United States	Dude Perfect	English	Sports	64.6
12	1	2024-04-25 21:29:29	India	T-Series	Hindi	Music	249
13	12	2024-04-25 21:29:29	India	Goldmies	Hindi	Film	89.5
14	17	2024-04-25 21:29:29	South Korea	BTS	Korean	Music	61.6
15	9	2024-04-25 21:29:29	India	Zee Music Company	Hindi	Music	99.5
16	4	2024-04-25 21:29:29	India	Sony Entertainment Television India	Hindi	Entertainment	162
17	19	2024-04-25 21:29:29	United States	Taylor Swift	English	Music	60.8
18	3	2024-04-25 21:29:29	United States	Cocomelon	English	Education	165
19	14	2024-04-25 21:29:29	United States	5-Minute Crafts	English	DIY & Crafts	64.1

https://app.snowflake.com/tzvraag/zvb17703/#/data/databases/PC_FIVETRAN_DB/schemas/GOOGLE_SHEETS_YOUTUBE/table/GOOGLE_SHEETS_YOUTUBE_TABLE/data-preview

7. Create the connector 'load_sam_data' and use it for the above-mentioned operation.

The screenshot shows the Fivetran connector configuration interface. At the top, there's a header with tabs for 'Status', 'Schema' (which is selected), 'Usage', and 'Setup'. A status message says 'First time syncing data for this connector'. Below the header, the connector name 'load_sam_data' is shown along with its source ('Google Sheets') and destination ('SnowflakePC'). The 'Schema' tab displays the table structure: 'load_sam_data_table' with columns '_row', 'brand_channel', 'category', 'country', 'name', 'primary_language', and 'subscribers_in_millions'. On the left sidebar, there are sections for 'Connectors', 'Transformations', and 'Destinations'. A trial usage estimator is also present. On the right side, there are buttons for 'Documentation', 'Release Notes', and a 'SYNCING' toggle switch which is set to 'ENABLED'.

The screenshot shows the Fivetran connector list page. It lists three active connectors: 'load_sam_data.load_sam_data_table' (Source: Google Sheets, Status: Active, Last synced: a minute ago) and 'google_sheets.youtube.google_sheets.youtube_table' (Source: Google Sheets, Status: Active, Last synced: 19 minutes ago). There are also sections for 'New Automatic PII Detection' and 'My Bookmarks'. A 'Select destinations' button is visible. The left sidebar includes sections for 'Alerts', 'Account Settings', 'Resources & Support', and a user profile for 'Aparna'. A 'Need help?' button is located at the bottom right.

PC_FIVETRAN_DB / LOAD_SAM_DATA

NAME	TYPE	OWNER	ROWS	BY...	CREAT...
LOAD_SAM_DATA_TABLE	Table	PC_FIVETRAN_ROLE	19	40...	1 day ago

PC_FIVETRAN_DB / LOAD_SAM_DATA / LOAD_SAM_DATA_TABLE

_ROW	_FIVETRAN_SYNCED	COUNTRY	NAME	PRIMARY_LANGUAGE	BRAND_CHANNEL	CATEGORY	SUBSCRIBERS_IN_MILLIONS	
1	2	2024-04-25 21:46:49	United States	MrBeast	English	No	Entertainment	184
2	15	2024-04-25 21:46:49	United States	Ariana Grande	English	Yes	Music	63.4
3	7	2024-04-25 21:46:49	Russia-United States	Like Nasty	English	No	Entertainment	107
4	10	2024-04-25 21:46:49	United States	WWE	English	Yes	Sports	97.1
5	18	2024-04-25 21:46:49	Argentina	El Reino Infantil	Spanish	Yes	Kids	61.3
6	19	2024-04-25 21:46:49	United States	Taylor Swift	English	Yes	Music	60.8
7	6	2024-04-25 21:46:49	Sweden	PewDiePie	English	No	Entertainment	111
8	5	2024-04-25 21:46:49	Ukraine-United States	Kids Diana Show	English	Yes	Entertainment	113
9	11	2024-04-25 21:46:49	South Korea	Blackpink	Korean	No	Music	91.2
10	3	2024-04-25 21:46:49	United States	Cocomelon	English	Yes	Education	185
11	16	2024-04-25 21:46:49	United States	Trotm Trotm	English	Yes	DIY & Crafts	61.8
12	8	2024-04-25 21:46:49	Russia-United States	Vlad and Niki	English	No	Entertainment	101
13	14	2024-04-25 21:46:49	United States	5-Minute Crafts	English	Yes	DIY & Crafts	64.1
14	4	2024-04-25 21:46:49	India	Sony Entertainment Television India	Hindi	Yes	Entertainment	162
15	13	2024-04-25 21:46:49	United States	Dude Perfect	English	Yes	Sports	64.8
16	9	2024-04-25 21:46:49	India	Zee Music Company	Hindi	Yes	Music	99.5
17	17	2024-04-25 21:46:49	South Korea	BTS	Korean	Yes	Music	61.8
18	1	2024-04-25 21:46:49	India	T-Series	Hindi	Yes	Music	249
19	12	2024-04-25 21:46:49	India	Goldnames	Hindi	Yes	Film	89.5

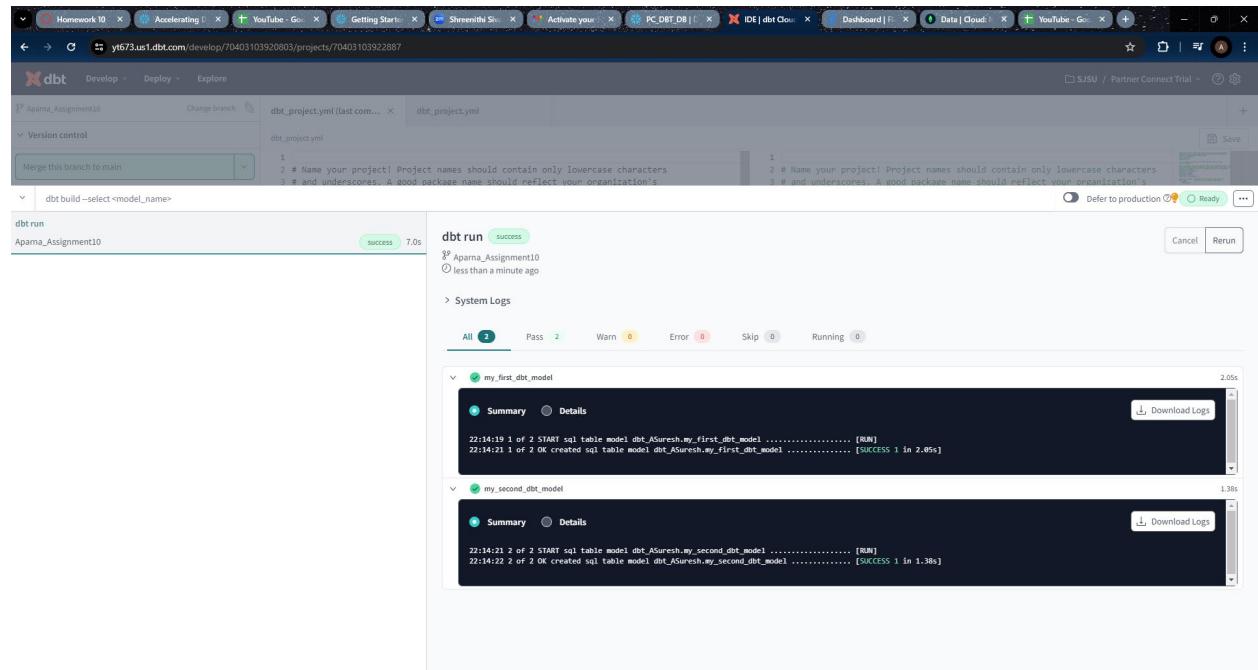
8. Initialize a DBT Repository, create a table sam_data with a few values (missing values too)

```
{% config(materialized='table') %}
```

```
with sam_data as (
```

```
    select 1 as student_id,'Aparna' as student_name, null as age union all
    select 2 as student_id,'Abc' as student_name, 20 as age union all
    select 3 as student_id,'Def' as student_name, 21 as age union all
    select 4 as student_id,'Xyz' as student_name, null as age union all
    select 5 as student_id,'Lmn' as student_name, null as age
)
```

```
select * from sam_data
```



Activate your Snowflake account | Partner Connect | IDE dbt Cloud | Homework 10 | Shreenithi Sivakumar's Zoom | +

dbt Develop Deploy Explore

Aparna_Assignment10 sam_data.sql

Version control Commit and sync Changes

```

models/example: sam_data.sql
1 {{ config(materialized='table') }}
2
3 with sam_data as (
4
5     select 1 as student_id,'Aparna' as student_name, null as age union all
6     select 2 as student_id,'Abc' as student_name, 20 as age union all
7     select 3 as student_id,'Def' as student_name, 21 as age union all
8     select 4 as student_id,'Xyz' as student_name, null as age union all
9     select 5 as student_id,'Lmn' as student_name, null as age
10 )
11
12 select * from sam_data
13

```

File explorer

- dbt cloud 70403103920803/partner-connect-trial-repo
 - analyses
 - dbt_packages
 - macros
 - gitskeep
 - generate_schema_name.sql
 - query.tag.sql
 - models
 - my_first_dbt_model.sql
 - my_second_dbt_model.sql
 - sam_data.sql
 - schema.yml
 - marts
 - staging
 - seeds
 - gitskeep
 - snapshots
 - target
 - tests
 - .gitignore
 - README.md
 - dbt_project.yml

Preview Build Format Results Code quality Compiled code Lineage

2+sam_data+2 Update Graph

Defer to production Ready

Activate your Snowflake account | DBT ASURESH| Schema | IDE dbt Cloud | Homework 10 | Shreenithi Sivakumar's Zoom | +

dbt Develop Deploy Explore

Aparna_Assignment10 sam_data.sql

Version control Commit and sync Changes

```

models/example: sam_data.sql
1 {{ config(materialized='table') }}
2
3 with sam_data as (

```

dbt run

Aparna_Assignment10 success 5.6s

dbt run success

Aparna_Assignment10 less than a minute ago

System Logs

All	3
Pass	3
Warn	0
Error	0
Skip	0
Running	0

> my_first_dbt_model 2.25s
> sam_data 2.24s
> my_second_dbt_model 1.18s

dbt run succeeded

The screenshot shows the Snowflake Data Preview interface. On the left, the sidebar navigation includes Database, Data, Projects, and Admin. A message indicates '\$369 of \$400 left' and '23 days left in trial'. The main content area displays the database structure under 'PC_DBT_DB / DBT_ASURESH / SAM_DATA'. It shows a table named 'SAM_DATA' with 5 rows. The columns are 'STUDENT_ID', 'STUDENT_NAME', and 'AGE'. The data is as follows:

STUDENT_ID	STUDENT_NAME	AGE
1	Aparna	null
2	Aloc	20
3	Def	21
4	Xyz	null
5	Lmn	null

9. Apply transformation and remove null values and load data into Snowflake.

```
select
  student_id,student_name,age
from {{ ref('sam_data') }} Where age IS NOT NULL
```

The screenshot shows the dbt Cloud interface with the following details:

- File:** sam_data_transformation.sql
- Code:**

```
1 select
2   student_id,student_name,age
3   from {{ ref('sam_data') }} Where age IS NOT NULL
```
- Lineage Graph:** A diagram showing the flow from 'sam_data' to 'sam_data_transformation'. The 'sam_data' node is connected to the 'sam_data_transformation' node. The 'sam_data_transformation' node has a tooltip indicating it contains 2+ rows.
- Buttons:** Preview, Compile, Build, Format, Results, Code quality, Compiled code, Lineage.
- Bottom Bar:** dbt build --select <model_name>

The screenshot shows the dbt Cloud interface with the following details:

- File:** sam_data_transformation.sql
- Status:** success (5.9s)
- Logs:**

```
dbt run [SUCCESS]
89 Aparna_Assignment10
⌚ less than a minute ago
```
- System Logs:** A table showing the execution status of each model:

Model	Duration
my_first_dbt_model	2.21s
sam_data	2.26s
my_second_dbt_model	1.35s
sam_data_transformation	1.19s
- Buttons:** Cancel, Rerun.

The screenshot shows the Snowflake Data Preview interface. On the left, the navigation sidebar is visible with options like Search, Projects, Data (selected), Databases (highlighted in blue), Data Products, AI & ML, Monitoring, and Admin. A promotional banner for an upgrade is present, stating '\$351 of \$400 left' and 'Upgrade'. The main area displays a table titled 'PC_DBT_DB / DBT_ASURESH / SAM_DATA_TRANSFORMATION'. The table has columns 'STUDENT_ID', 'STUDENT_NAME', and 'AGE'. The data preview shows three rows:

	STUDENT_ID	STUDENT_NAME	AGE
1	2	Abc	20
2	3	Def	21

At the bottom left of the interface, the user's account information is displayed: AS Aparna Suresh ACCOUNTADMIN.