

Homework 10

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

```
create or replace TABLE TRIP_DB.TRIP_SCHEMA.TRIP_DATA (  
    TRIPDURATION NUMBER(38,0),  
    STARTTIME TIMESTAMP_NTZ(9),  
    STOPTIME TIMESTAMP_NTZ(9),  
    START_STATION_ID NUMBER(38,0),  
    START_STATION_NAME VARCHAR(16777216),  
    START_STATION_LATITUDE FLOAT,  
    START_STATION_LONGITUDE FLOAT,  
    END_STATION_ID NUMBER(38,0),  
    END_STATION_NAME VARCHAR(16777216),  
    END_STATION_LATITUDE FLOAT,  
    END_STATION_LONGITUDE FLOAT,  
    BIKEID NUMBER(38,0),  
    MEMBERSHIP_TYPE VARCHAR(16777216),  
    USERTYPE VARCHAR(16777216),  
    BIRTH_YEAR NUMBER(38,0),  
    GENDER NUMBER(38,0)  
);
```

NSG

Nikhil Sarma Gu...
ACCOUNTADMIN

Worksheets

Dashboards

Streamlit

Apps

Data

Databases

Private Sharing

Provider Studio

Marketplace

Activity

Admin

Help & Support

30 days left in trial

Upgrade

RBB44129

Search

> SNOWFLAKE

> SNOWFLAKE_SAMPLE_DATA

> TRIP_DB

- > INFORMATION_SCHEMA
- > PUBLIC

Databases

+ Database

3 Databases

Search Source All

NAME ↑	SOURCE	OWNER	CREATED	
SNOWFLAKE	Share	—	1 minute...	...
SNOWFLAKE_SAM...	Share	ACCOU...	1 minute...]
TRIP_DB	Local	ACCOU...	just now	...

NSG

Nikhil Sarma Gu...
ACCOUNTADMIN

Worksheets

Dashboards

Streamlit

Apps

Data

Databases

Private Sharing

Provider Studio

Marketplace

Activity

Admin

Help & Support

30 days left in trial

Upgrade

RBB44129

Search

> SNOWFLAKE

> SNOWFLAKE_SAMPLE_DATA

> TRIP_DB

INFORMATION_SCHEMA

PUBLIC

No Objects found

TRIP_SCHEMA

No Objects found

TRIP_DB / TRIP_SCHEMA

Create

Schema

ACCOUNTADMIN

just now

Schema Details

Privileges

Group by Role

+ Privilege

ACCOUNTADMIN (Current Role)

OWNERSHIP

TRIP_DB / TRIP_SCHEMA / TRIP_DATA

...

Load Data

Table

ACCOUNTADMIN

6 hours ago

61.6M

1.2GB

Table Details

Columns

Data Preview

Copy History

Table definition

```

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 );

```

Show less ^

Privileges

Group by Role ▾

+ Privilege

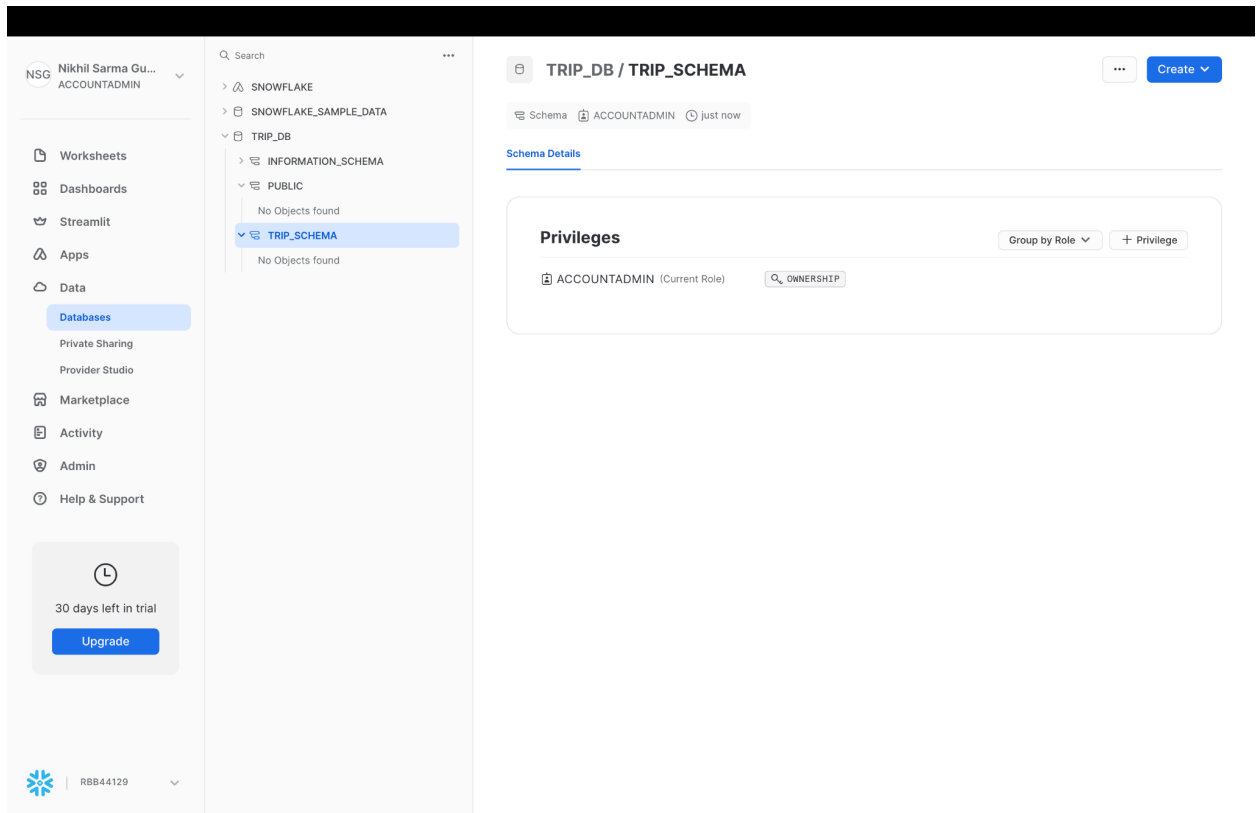
ACCOUNTADMIN (Current Role)

OWNERSHIP

2. Create Stage 'trip_stage' in Snowflake.

```
CREATE STAGE trip_stage
```

```
URL = 's3://snowflake-workshop-lab/citibike-trips';
```



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

```
CREATE 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 = 'citi_file_format';
```

The screenshot displays the Snowflake web interface. At the top, the header shows 'Worksheets' and the timestamp '2023-11-20 1:36pm'. The left sidebar contains a 'Databases' tab and a tree view with 'SNOWFLAKE', 'SNOWFLAKE_SAMPLE_DATA', and 'TRIP_DB'. The main editor area shows a SQL query for creating a file format:

```
1 CREATE FILE FORMAT trip_file_format
2 TYPE=CSV COMPRESSION = 'AUTO'
3 FIELD_DELIMITER = ','
4 RECORD_DELIMITER = '\n'
5 SKIP_HEADER = 0
6 FIELD_OPTIONALLY_ENCLOSED_BY = '\042'
7 TRIM_SPACE = FALSE
8 ERROR_ON_COLUMN_COUNT_MISMATCH = FALSE ESCAPE = 'NONE' ESCAPE_UNENCLOSED_FIELD = '\134'
9 DATE_FORMAT = 'AUTO' TIMESTAMP_FORMAT = 'AUTO' NULL_IF = ( '') COMMENT = 'citi_file_format';
```

 Below the editor, the 'Results' tab is active, showing a single row with the status: 'File format TRIP_FILE_FORMAT successfully created.' To the right of the results, a 'Query Details' panel provides additional information: 'Query duration' is 55ms, 'Rows' is 1, 'Query ID' is 01b078b8-0001-e5c4-..., and the 'status' is 100% filled.

4. Load Data into the 'trip_data' table.

```
COPY INTO TRIP_DATA FROM @TRIP_STAGE
FILE_FORMAT = trip_file_format
ON_ERROR = 'CONTINUE';
```

The screenshot displays the Snowflake SQL Editor interface. At the top, the 'Worksheets' tab is active, showing a timestamp of 2023-11-20 1:36pm. The left sidebar contains a 'Databases' section with a search bar and a list of objects: SNOWFLAKE, SNOWFLAKE_SAMPLE_DATA, and TRIP_DB. The main editor area shows a SQL query in the 'TRIP_DB.TRIP_SCHEMA' database. The query consists of two parts: a 'COPY INTO' statement and a 'SELECT' statement. The 'COPY INTO' statement copies data from '@TRIP_STAGE' into 'TRIP_DATA', specifying 'FILE_FORMAT = trip_file_format' and 'ON_ERROR = 'CONTINUE''. The 'SELECT' statement follows, querying all data from 'trip_data'. Below the query editor, the 'Results' tab is selected, displaying a table with columns 'file', 'status', and 'rows_parsed'. The table contains 8 rows of data, all with a 'status' of 'LOADED'. To the right of the results table, a 'Query Details' panel provides additional information: 'Query duration' is 13m 8s, 'Rows' is 12.1K, and 'Query ID' is 01b076b9-0001-e68872-... The panel also shows a progress bar for 'file' at 100% and a 'status' section.

TRIP_DB.TRIP_SCHEMA Settings

```

1 COPY INTO TRIP_DATA FROM @TRIP_STAGE
2   FILE_FORMAT = trip_file_format
3   ON_ERROR = 'CONTINUE';
4
5
6 SELECT * FROM trip_data;

```

Results Chart

	file	status	rows_parsed
1	s3://snowflake-workshop-lab/citibike-trips-csv/trips_2013_0_3_0.csv.gz	LOADED	103140
2	s3://snowflake-workshop-lab/citibike-trips-csv/trips_2013_1_1_0.csv.gz	LOADED	112123
3	s3://snowflake-workshop-lab/citibike-trips-csv/trips_2013_2_3_0.csv.gz	LOADED	91069
4	s3://snowflake-workshop-lab/citibike-trips-csv/trips_2013_3_5_0.csv.gz	LOADED	97625
5	s3://snowflake-workshop-lab/citibike-trips-csv/trips_2013_4_2_0.csv.gz	LOADED	115293
6	s3://snowflake-workshop-lab/citibike-trips-csv/trips_2013_5_5_0.csv.gz	LOADED	108605
7	s3://snowflake-workshop-lab/citibike-trips-csv/trips_2013_6_5_0.csv.gz	LOADED	82969
8	s3://snowflake-workshop-lab/citibike-trips-csv/trips_2014_0_5_0.csv.gz	LOADED	103229

Query Details

Query duration 13m 8s

Rows 12.1K

Query ID 01b076b9-0001-e68872-...

file 100% filled

status

SELECT * FROM trip_data;

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

x-large performance is better than that of x-small

Worksheets2023-11-20 1:36pm

DatabasesWorksheets

Pinned (0)

No pinned objects

Search objects

PC_FIVETRAN_DB

DB_ASSIGNMENT

Tables

LOAD_SAM_DATA

INFORMATION_SCHEMA

PUBLIC

SNOWFLAKE

SNOWFLAKE_SAMPLE_DATA

TRIP_DB

TRIP_DB.TRIP_SCHEMA

Settings

Code Versions

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

ON_ERROR = 'CONTINUE';

SELECT * FROM trip_data;

SELECT tripduration ,

starttime ,

stoptime ,

start_station_id ,

start_station_name ,

start_station_latitude ,

start_station_longitude ,

membership_type ,

usertype ,birth_year ,gender, end_station_id ,

end_station_name ,

end_station_latitude ,

end_station_longitude ,

bikeid FROM trip_data;

ResultsChart

Search

Download

Refresh

Fullscreen

LOAD_SAM_DATA

9 Rows

Columns

More

#

_ROW

NUMBER(38,0)

△

POSITION

VARCHAR(256)

△

FIRST_NAME

VARCHAR(256)

△

TEAM

VARCHAR(256)

△

LAST_NAME

VARCHAR(256)

🕒

_FIVETRAN_SYNCED

TIMESTAMP_TZ(9)

Snowflake executing...

23s

Start Time

ID

Warehouse

Produced Rows

Nov 20, 2:31 PM

01b076e7-0001-e5ff-0000-0043e8072159

COMPUTE_WH(X-Small)

8.3M (8,292,352)

Cancel

🔗

📄

📄

NSG

Nikhil Sarma Gu...

ACCOUNTADMIN

Worksheets

Dashboards

Streamlit

Apps

Data

Marketplace

Activity

Query History

Copy History

Task History

Dynamic Tables

Admin

Help & Support

🕒

30 days left in trial

Upgrade

❄️

RBB44129

Query - 01b076e7-0001-e5ff-0000-0043e8072159

COMPUTE_WH

NIKHILGUDUR

Query Details

Query Profile

Details

Status

Success

Duration

1m 13s

Driver Status

Supported

Start Time

11/20/2023, 2:31 PM

Query ID

01b076e7-0001-e5ff-0000-0043e8072159

Client Driver

Go 1.1.5

End Time

11/20/2023, 2:32 PM

Query Tag

—

Session ID

291655598101

Warehouse Size

X-Small

SQL Text

SELECT tripduration ,

starttime ,

stoptime ,

start_station_id ,

start_station_name ,

start_station_latitude ,

start_station_longitude ,

membership_type ,

usertype ,birth_year ,gender, end_station_id ,

end_station_name ,

end_station_latitude ,

Worksheets2023-11-20 1:36pm

Databases

Worksheets

Pinned (0)

No pinned objects

Search objects

...

PC_FIVETRAN_DB

DB_ASSIGNMENT

Tables

LOAD_SAM_DATA

INFORMATION_SCHEMA

PUBLIC

SNOWFLAKE

SNOWFLAKE_SAMPLE_DATA

TRIP_DB

TRIP_DB.TRIP_SCHEMA

Settings

Code Versions

Q

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

SELECT * FROM trip_data;

SELECT tripduration ,

stoptime ,

start_station_id ,

start_station_latitude ,

end_station_longitude ,

start_station_longitude ,

membership_type ,

usertype ,

birth_year ,

gender ,

starttime ,

end_station_id ,

end_station_name ,

start_station_name ,

end_station_latitude ,

bikeid FROM trip_data;

Results

Chart

Q

Download

Refresh

Fullscreen

LOAD_SAM_DATA

9 Rows

...

#_ROW

NUMBER(38,0)

POSITION

VARCHAR(256)

FIRST_NAME

VARCHAR(256)

TEAM

VARCHAR(256)

LAST_NAME

VARCHAR(256)

_FIVETRAN_SYNCED

TIMESTAMP_TZ(9)

Snowflake executing...

13s

Start Time

Nov 20, 2:37 PM

ID

01b076ed-0001-e5c4-0000-0043e8077135

Warehouse

COMPUTE_WH(X-Large)

Produced Rows

3.1M (3,088,384)

Cancel

NSG Nikhil Sarma Gu... ACCOUNTADMIN

Worksheets

Dashboards

Streamlit

Apps

Data

Marketplace

Activity

Query History

Copy History

Task History

Dynamic Tables

Admin

Help & Support

30 days left in trial

Upgrade

RBB44129

Query - 01b076ed-0001-e5c4-0000-0043e8077135

COMPUTE_WH NIKHILGUDUR

Query Details

Query Profile

Details

Status

Success

Duration

12s

Driver Status

Supported

Start Time

11/20/2023, 2:37 PM

Query ID

01b076ed-0001-e5c4-0000-0043e8077135

Client Driver

Go 1.1.5

End Time

11/20/2023, 2:37 PM

Query Tag

—

Session ID

291655598101

Warehouse Size

X-Large

SQL Text

SELECT tripduration ,

stoptime ,

start_station_id ,

start_station_latitude ,

end_station_longitude ,

start_station_longitude ,

membership_type ,

usertype ,

birth_year ,

gender ,

starttime ,

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.

The screenshot shows a Google Sheet titled "NBA Players" with a menu bar (File, Edit, View, Insert, Format, Data, Tools, Extensions, Help) and a toolbar with various icons. The spreadsheet contains a table with 4 columns: First Name, Last Name, Position, and Team. The data is as follows:

1	First Name	Last Name	Position	Team
2	LeBron	James	Forward	Lakers
3	Jimmy	Butler	Forward	Heat
4	Stepher	Curry	Guard	Warriors
5	Kyrie	Irving	Guard	Nets
6	Joel	Embiid	Center	76ers
7	Monica	Geller	Forward	Lakers
8	Chandler	Bing	Centre	Warriors
9	Joey	Tribbiani	Guard	Nets
10	Ross	Geller	Forward	Heat

The sidebar on the right shows "Named ranges" with two entries: "DB_Assignment" (Sheet1!A1:D10) and "PlayerList" (Sheet1!A1:D6). The "PlayerList" entry has an edit icon.

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

1 Connect destination

2 Connect source

3 Complete initial sync

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

Sjsu_260832062316349...

Connectors

Transformations

Destinations

Trial usage estimator

Alerts

Account Settings

Resources & Support

Nikhil Sarma

Give Feedback

Google Sheets

Connection tests:

Finding specified sheet

Validating named range

All connection tests passed!

Back

Continue

Worksheets2023-11-20 1:36pm2023-11-20 2:24pm

DatabasesWorksheets

PC_FIVETRAN_DB

PC_FIVETRAN_DB.DB_ASSIGNMENT

create or replace TABLE PC_FIVETRAN_DB.DB_ASSIGNMENT.LOAD_SAM_DATA (_ROW NUMBER(38,0) NOT NULL, POSITION VARCHAR(256), FIRST_NAME VARCHAR(256), TEAM VARCHAR(256), LAST_NAME VARCHAR(256), _FIVETRAN_SYNCED TIMESTAMP_TZ(9), primary key (_ROW)); SELECT * FROM LOAD_SAM_DATA;

	_ROW	POSITION	FIRST_NAME	TEAM	LAST_NAME	_FIVETRAN_SYNCED
1	1	Forward	LeBron	Lakers	James	2023-11-20 22:24:02.773 +0000
2	2	Forward	Jimmy	Heat	Butler	2023-11-20 22:24:02.780 +0000
3	3	Guard	Stepher	Warriors	Curry	2023-11-20 22:24:02.781 +0000
4	4	Guard	Kyrie	Nets	Irving	2023-11-20 22:24:02.781 +0000
5	5	Center	Joel	76ers	Embiid	2023-11-20 22:24:02.781 +0000
6	6	Forward	Monica	Lakers	Geller	2023-11-20 22:24:02.781 +0000
7	7	Centre	Chandler	Warriors	Bing	2023-11-20 22:24:02.781 +0000
8	8	Guard	Joey	Nets	Tribbiani	2023-11-20 22:24:02.781 +0000
9	9	Forward	Ross	Heat	Geller	2023-11-20 22:24:02.781 +0000

Query Details

Query duration83ms

Rows9

Query ID01b076e0-0001-e5c4-

POSITION

Forward3

Guard3

Center1

+ 2 more

FIRST_NAME

100% filled

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

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

```
SELECT 1 as id, 'David' as name, NULL as age
```

```
UNION ALL
```

```
SELECT 2 as id, 'Rachel' as name, 49 as age
```

```
UNION ALL
```

```
SELECT 3 as id, 'Chandler' as name, NULL as age
```

```
UNION ALL
```

```
SELECT 4 as id, 'Phoebe' as name, 19 as age
```

```
UNION ALL
```

```
SELECT 5 as id, 'Ross' as name, 20 as age
```

The screenshot displays the dbt CLI interface. At the top, there's a file explorer showing the project structure with files like schema.yml, dbt_project.yml, my_second_dbt_model.sql, nikhil_sam_data.sql, and nikhil_sam_data_transfor... The main area shows the command 'dbt run --select <model_name>' being executed. Below this, a table lists the results of the run:

Model Name	Status	Duration
my_first_dbt_model	Success	2.04s
nikhil_sam_data	Success	2.03s
nikhil_sam_data_transformed	Success	1.02s
my_second_dbt_model	Success	0.80s

The interface also includes a 'System Logs' section and a 'Defer to production' toggle set to 'Ready'.

Worksheets 2023-11-20 1:36pm 2023-11-20 9:43pm +

Databases Worksheets

Pinned (0)

No pinned objects

Search objects ...

- PC_DBT_DB
- PC_FIVETRAN_DB
- SNOWFLAKE
- SNOWFLAKE_SAMPLE_DATA
- TRIP_DB

PC_DBT_DB.DBT_NGUDUR Settings

Code Versions

1 `SELECT * FROM nikhil_sam_data;`

Results Chart

	ID	NAME	AGE
1	1	David	null
2	2	Rachel	49
3	3	Chandler	null
4	4	Phoebe	19
5	5	Ross	20

Query Details

Query duration 535ms

Rows 5

Query ID 01b0789a-0001-e687-...

ID

NAME 100% filled

AGE

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

```
{{ config(materialized='view') }}
```

```
with source_data as (
```

```
select *
```

```
from {{ ref('nikhil_sam_data') }}
```

```
)
```

```
select *
```

```
from source_data
```

```
where age is not null
```

Worksheets2023-11-20 1:36pm2023-11-20 9:43pm+

DatabasesWorksheets

Pinned (0)

No pinned objects

Search objects

PC_DBT_DB

PC_FIVETRAN_DB

SNOWFLAKE

SNOWFLAKE_SAMPLE_DATA

TRIP_DB

PC_DBT_DB.DBT_NGUDURSettings

1SELECT * FROM nikhil_sam_data_transformed;

ResultsChart

	ID	NAME	AGE
1	2	Rachel	49
2	4	Phoebe	19
3	5	Ross	20

Query Details

Query duration335ms

Rows3

Query ID01b0789b-0001-e5ff-0...

ID

25

NAME

100% filled

AGE

1949