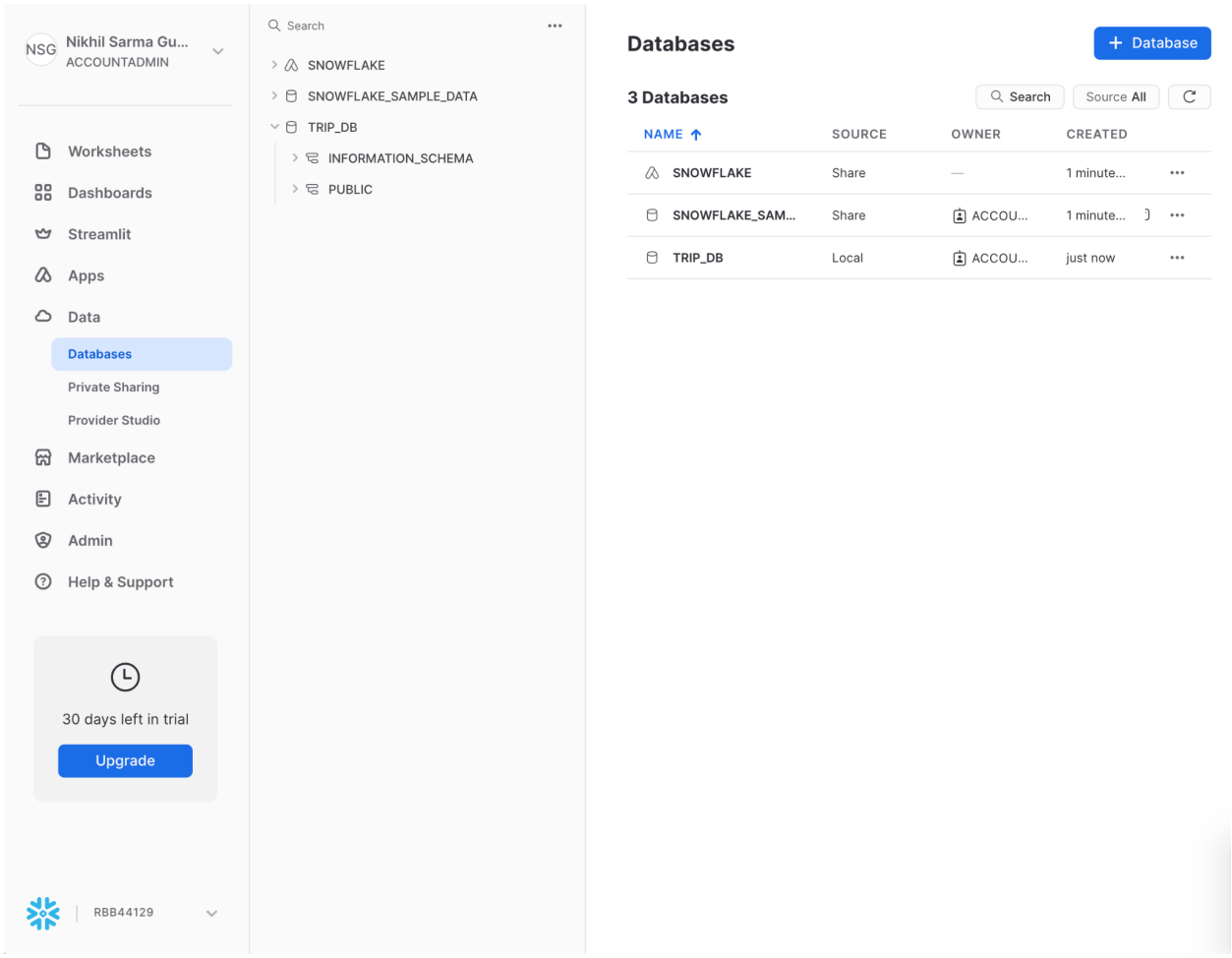
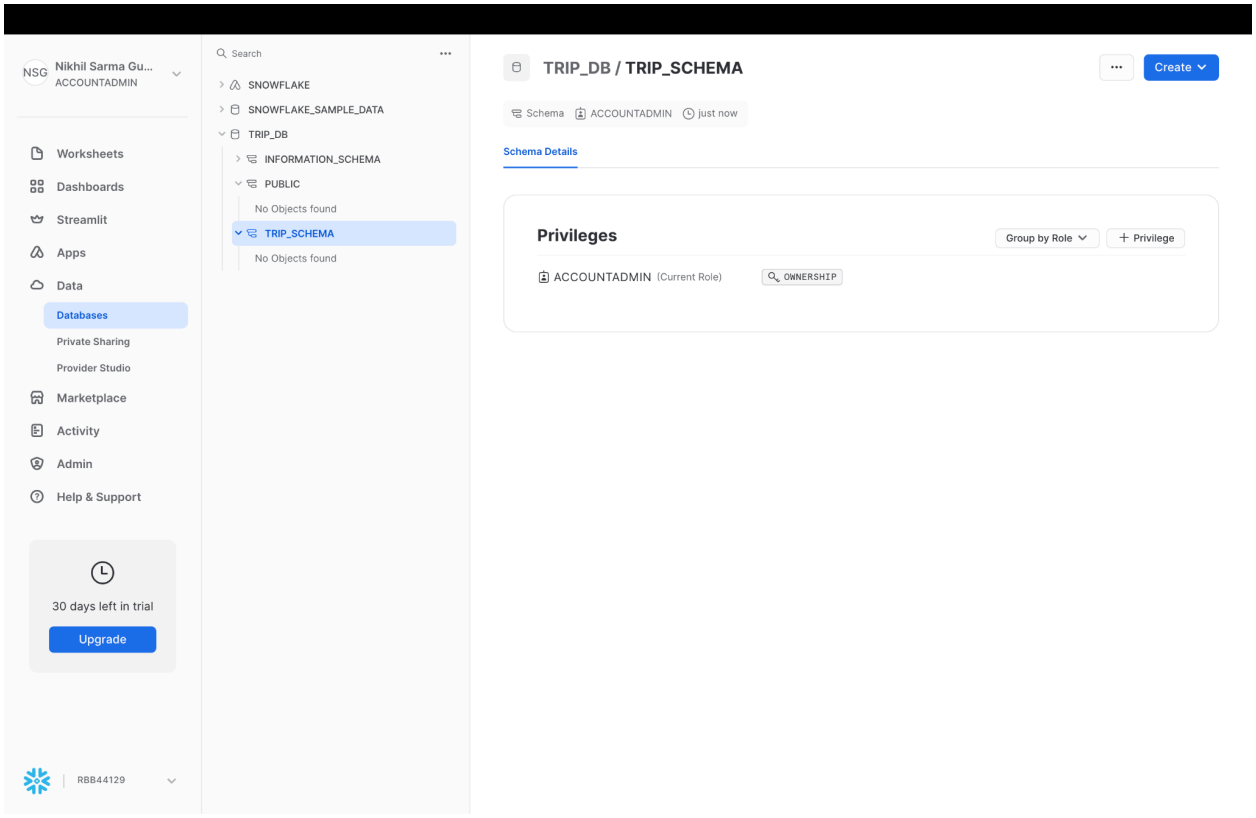
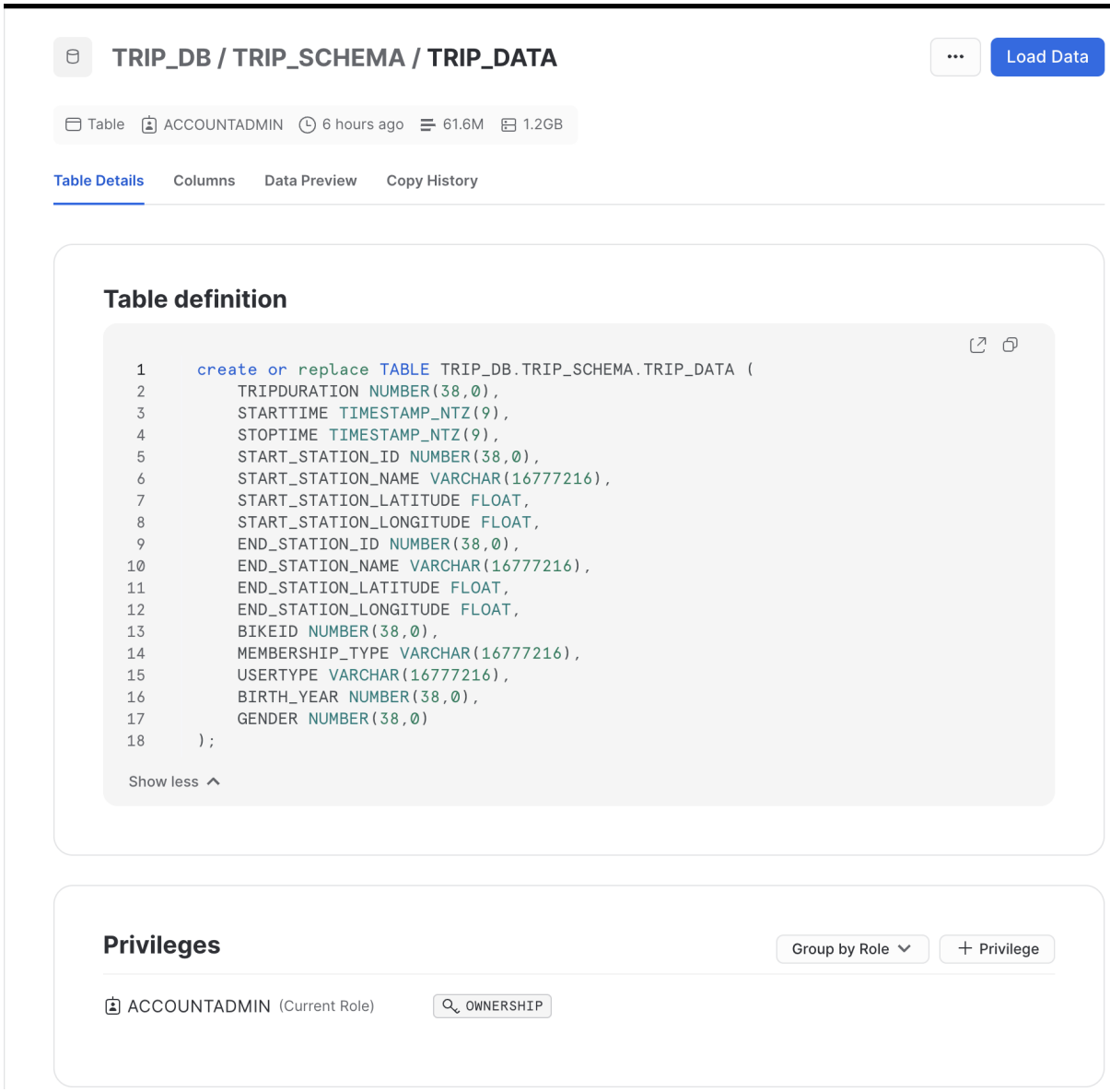
Homework 10

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

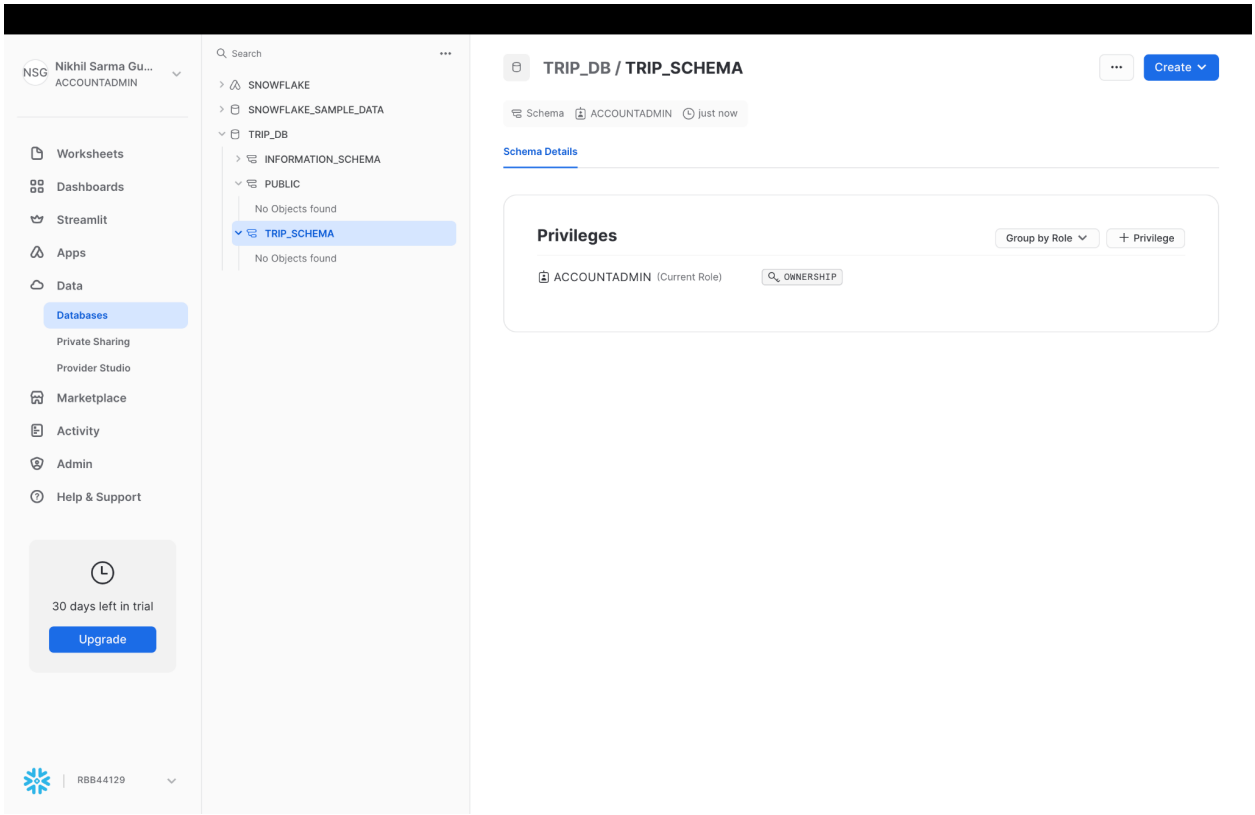

NSG
Nikhil Sarma Gu...
ACCOUNTADMIN
Worksheets
Dashboards
Streamlit
Apps
Data
Databases
Private Sharing
Provider Studio
Marketplace
Activity
Admin
Help & Support
30 days left in trial
Upgrade
RBB44129
Search
SNOWFLAKE
SNOWFLAKE_SAMPLE_DATA
TRIP_DB
INFORMATION_SCHEMA
PUBLIC
No Objects found
TRIP_SCHEMA
No Objects found
TRIP_DB / TRIP_SCHEMA
Create
Schema
ACCOUNTADMIN
just now
Schema Details
Privileges
Group by Role
Privilege
ACCOUNTADMIN (Current Role)
OWNERSHIP



2. Create Stage 'trip_stage' in Snowflake.

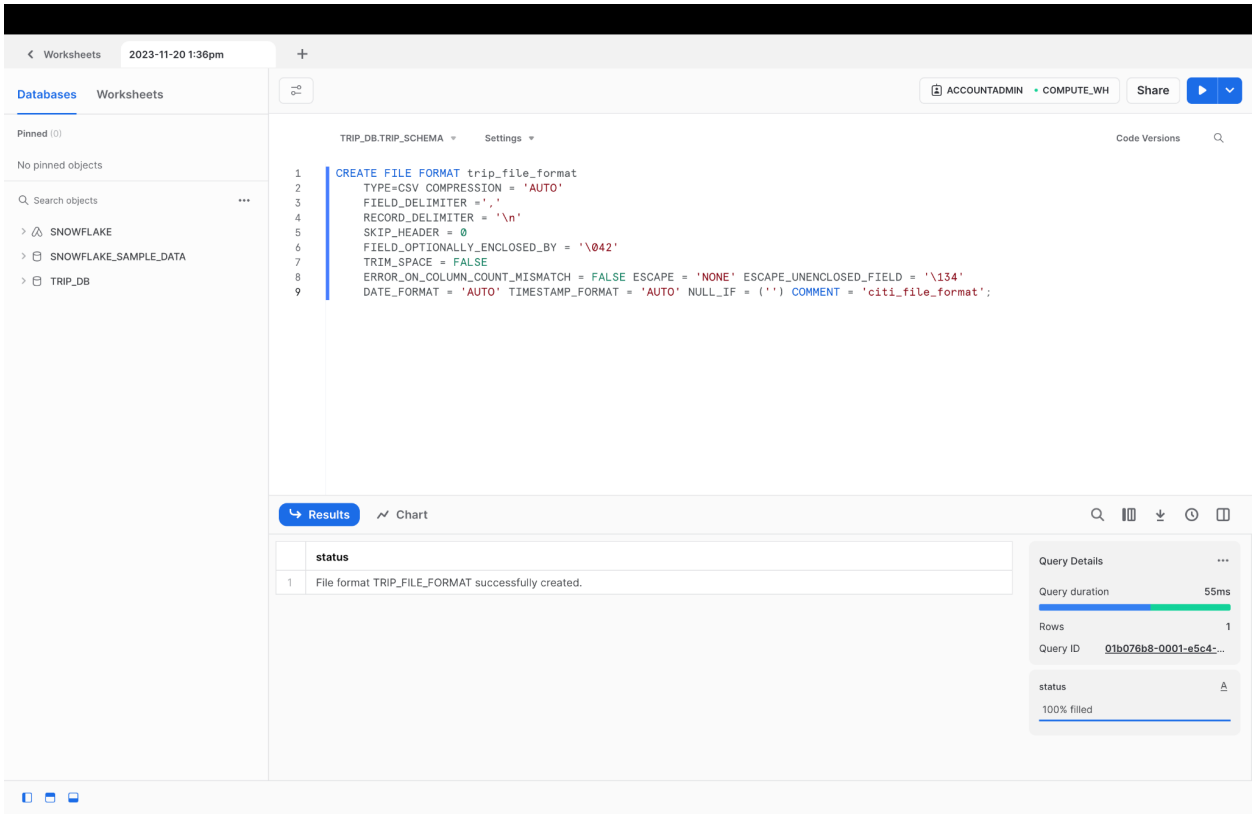
CREATE STAGE trip_stage

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



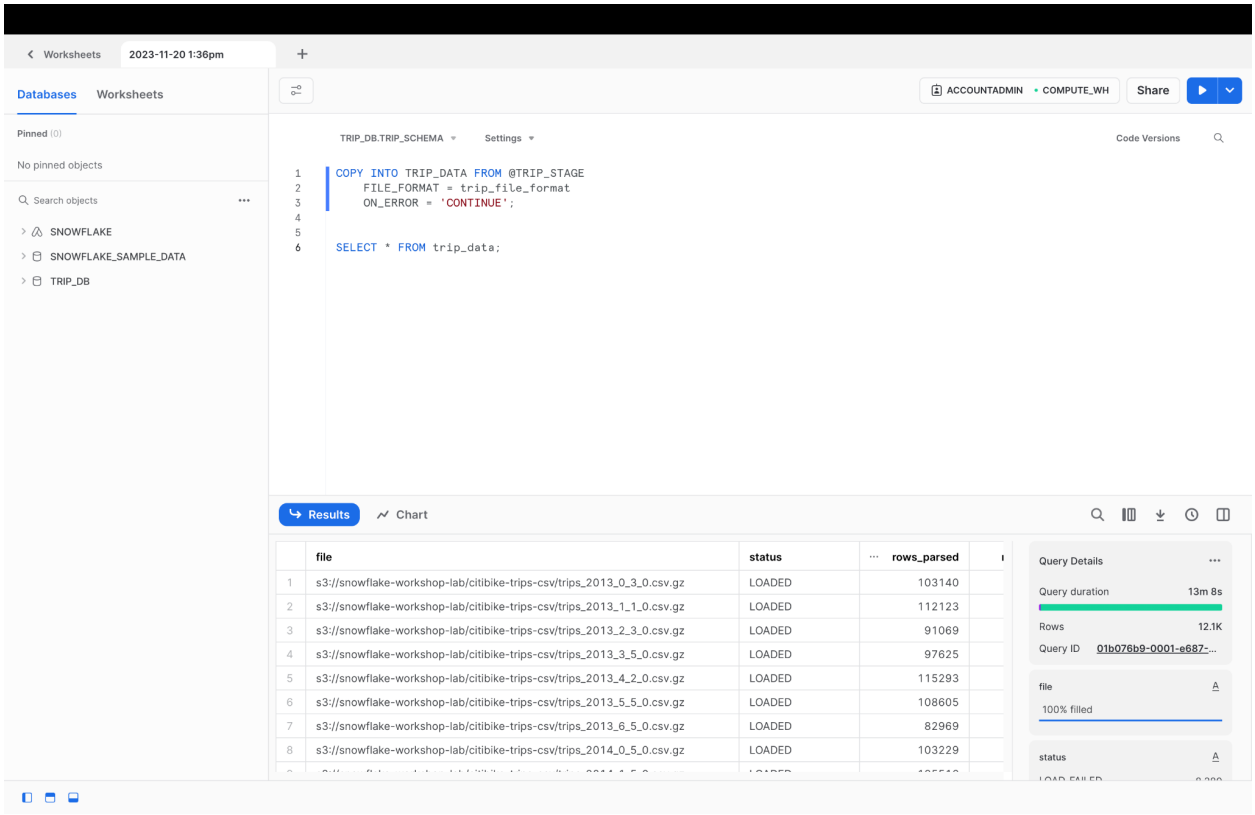
3. Create File Format 'trip_file_format' in the Snowflake.

```
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';
```



4. Load Data into the 'trip_data' table.

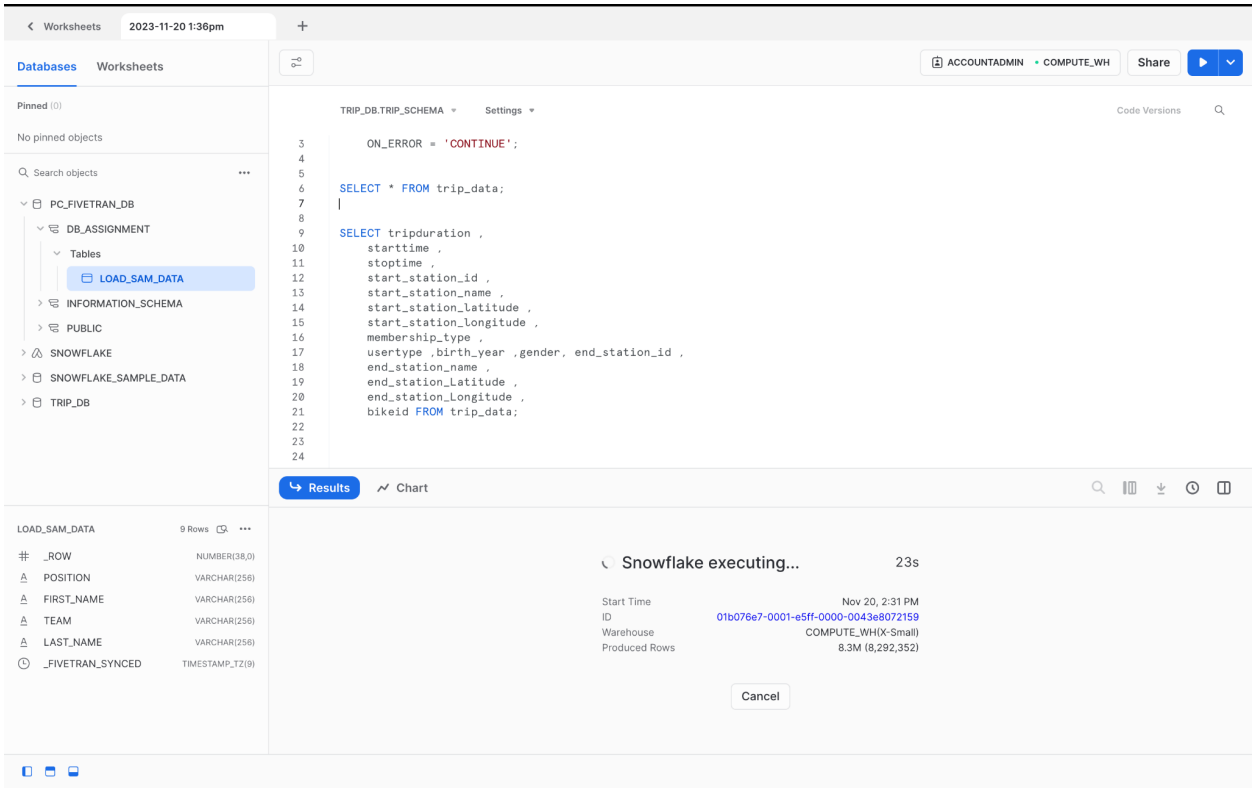
```
COPY INTO TRIP_DATA FROM @TRIP_STAGE
  FILE_FORMAT = trip_file_format
  ON_ERROR = 'CONTINUE';
```

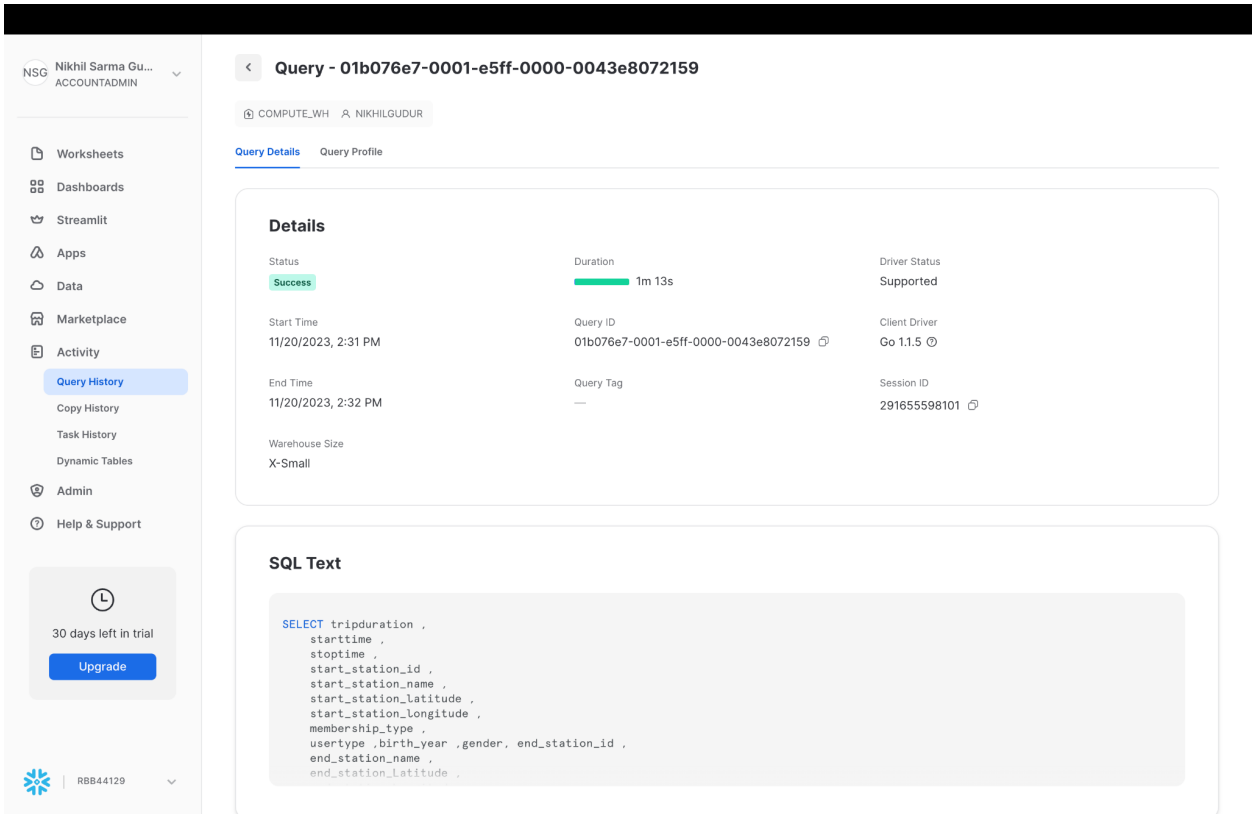


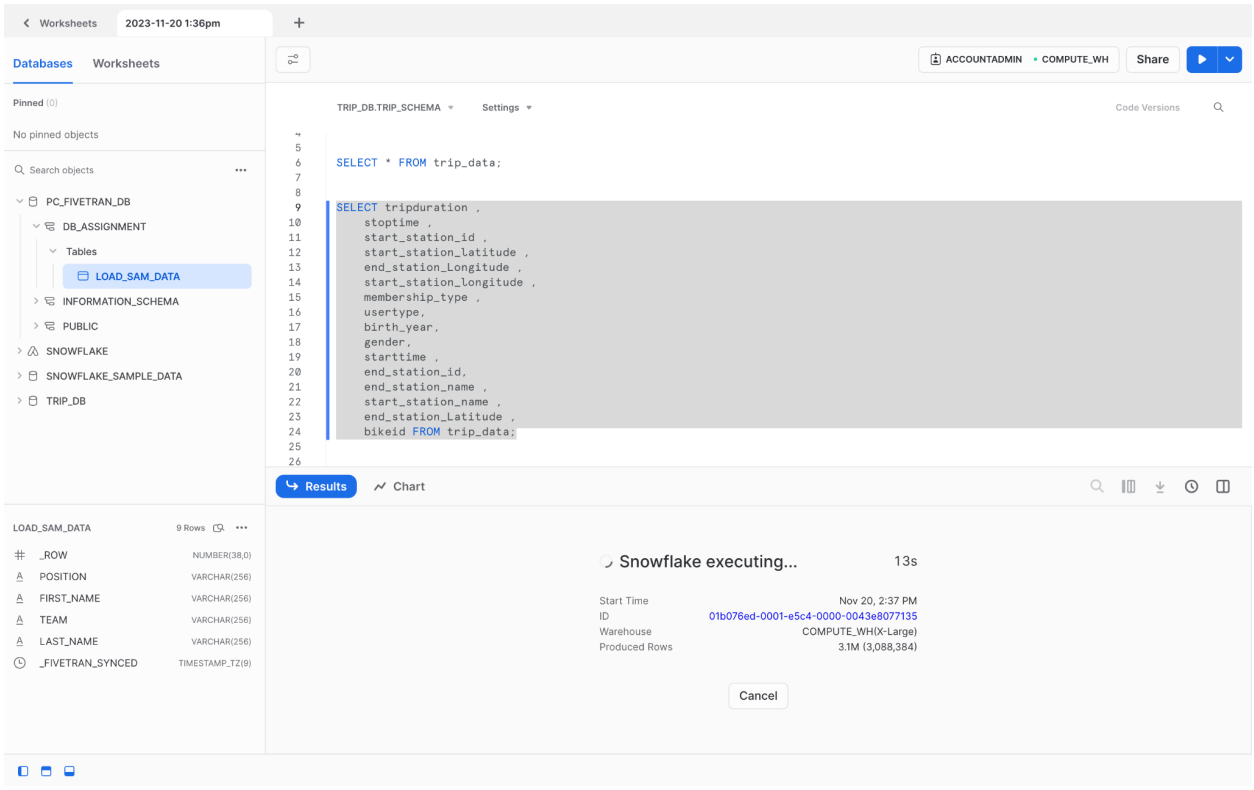
SELECT * FROM trip_data;

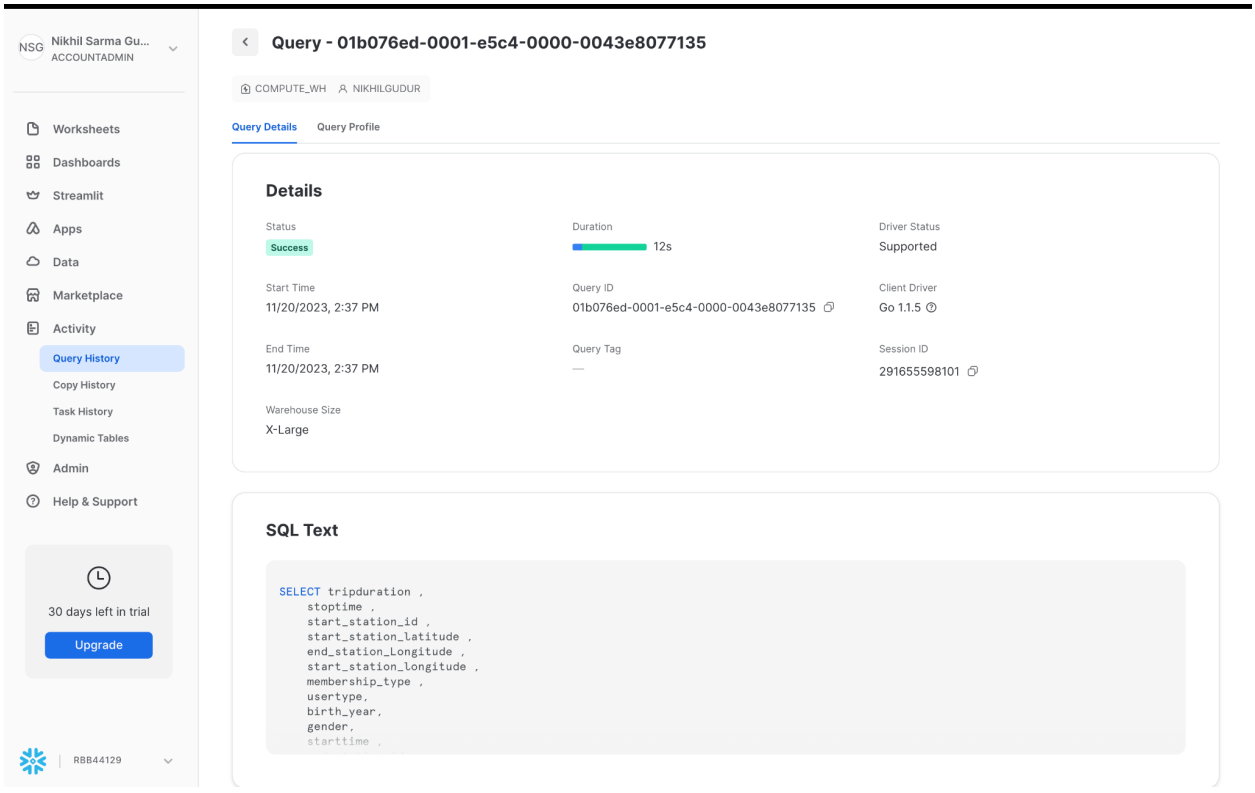
5. Analyze the performance of the node upon changing configurations from x-small to x-large in Snowflake.

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

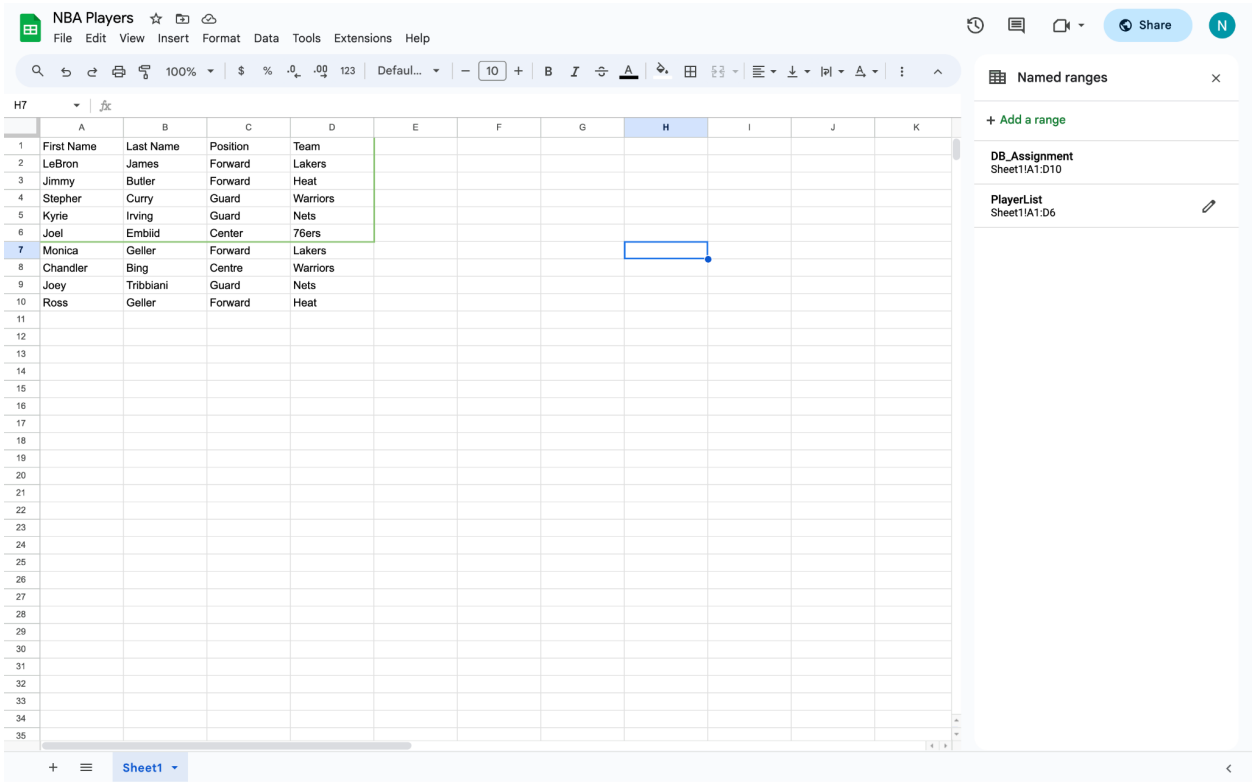

Worksheets
2023-11-20 1:36pm
Databases
Worksheets
ACCOUNTADMIN
COMPUTE_WH
Share
Pinned (0)
No pinned objects
Search objects
PC_FIVETRAN_DB
DB_ASSIGNMENT
Tables
LOAD_SAM_DATA
INFORMATION_SCHEMA
PUBLIC
SNOWFLAKE
SNOWFLAKE_SAMPLE_DATA
TRIP_DB
TRIP_DB.TRIP_SCHEMA
Settings
Code Versions
ON_ERROR = 'CONTINUE';
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;
Results
Chart
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...
23s
Start Time Nov 20, 2:31 PM
ID 01b076e7-0001-e5ff-0000-0043e8072159
Warehouse COMPUTE_WH(X-Small)
Produced Rows 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 ,


Worksheets
2023-11-20 1:36pm
Databases
Worksheets
ACCOUNTADMIN
COMPUTE_WH
Share
Pinned (0)
No pinned objects
TRIP_DB.TRIP_SCHEMA
Settings
Code Versions
Search objects
PC_FIVETRAN_DB
DB_ASSIGNMENT
Tables
LOAD_SAM_DATA
INFORMATION_SCHEMA
PUBLIC
SNOWFLAKE
SNOWFLAKE_SAMPLE_DATA
TRIP_DB
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
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

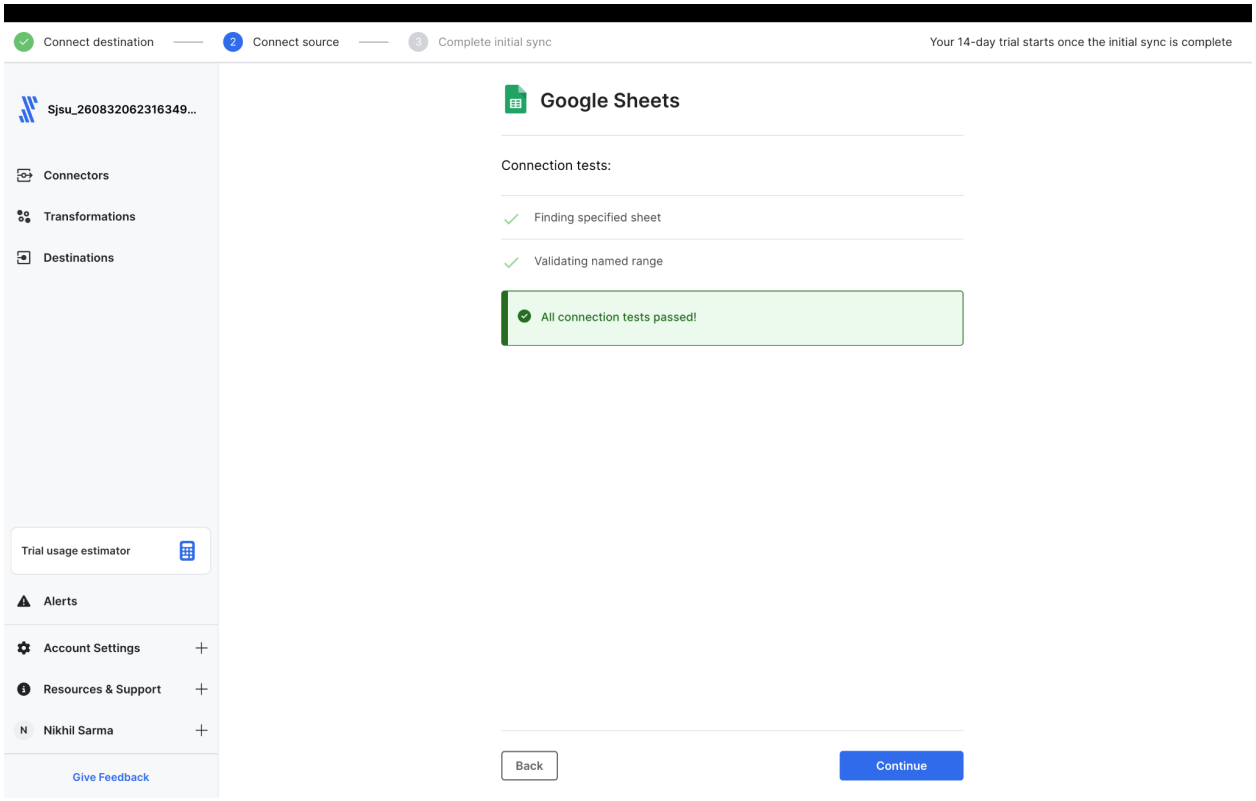

NSG
Nikhil Sarma Gu...
ACCOUNTADMIN
Worksheets
Dashboards
Streamlit
Apps
Data
Marketplace
Activity
Query History
Copy History
Task History
Dynamic Tables
Admin
Help & Support
30 days left in trial
Upgrade
RBB44129
Query - 01b076ed-0001-e5c4-0000-0043e8077135
COMPUTE_WH
NIKHILGUDUR
Query Details
Query 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 ,

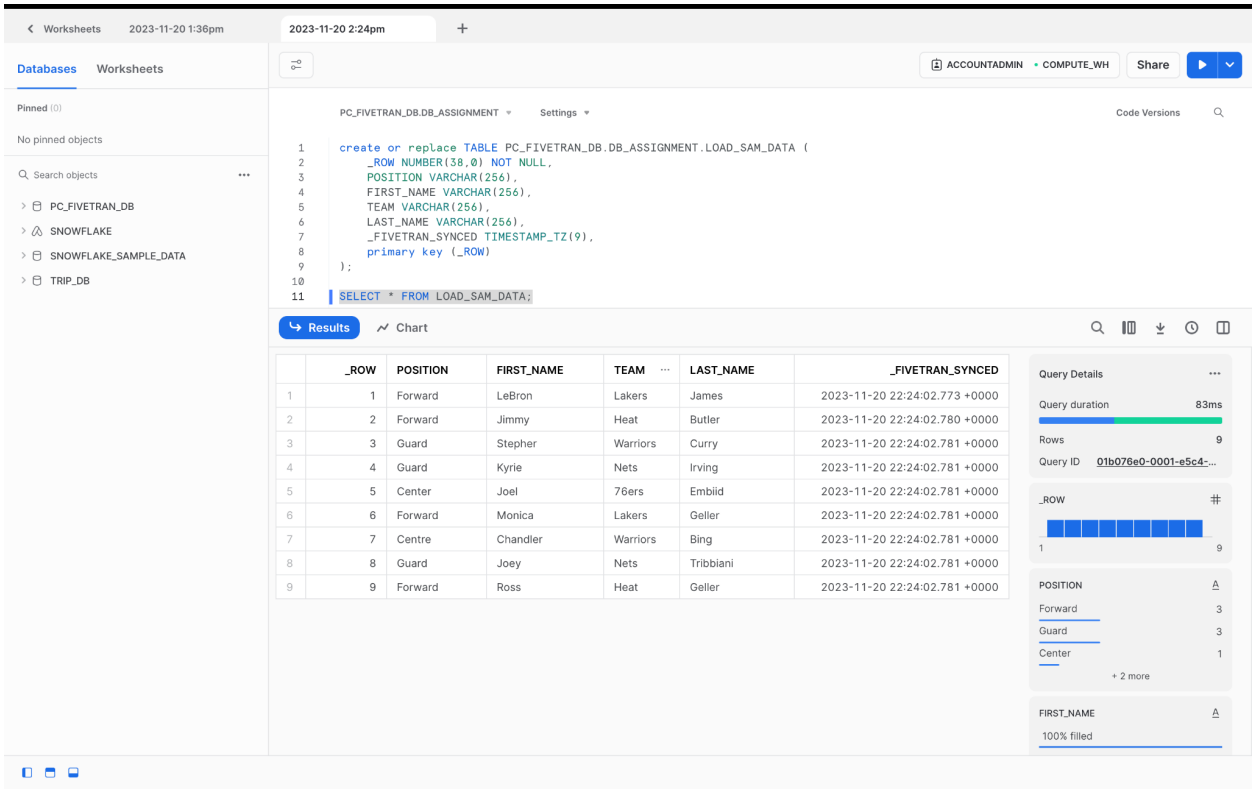
6. Create a copy of the Spreadsheet shared in the announcement, add a few more records, create a named range and load data into Snowflake with Fivetran.



First Name	Last Name	Position	Team
LeBron	James	Forward	Lakers
Jimmy	Butler	Forward	Heat
Stepher	Curry	Guard	Warriors
Kyrie	Irving	Guard	Nets
Joel	Embiid	Center	76ers
Monica	Geller	Forward	Lakers
Chandler	Bing	Centre	Warriors
Joey	Tribbiani	Guard	Nets
Ross	Geller	Forward	Heat

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

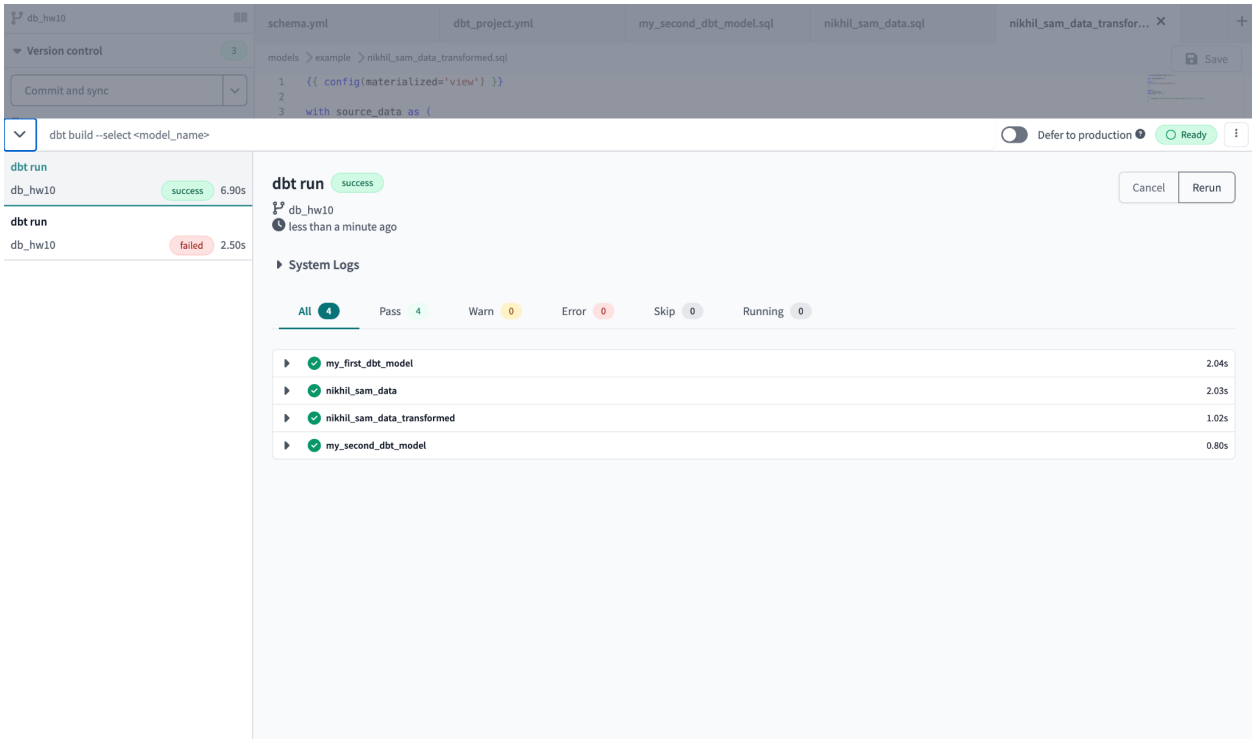

Connect destination
2 Connect source
3 Complete initial sync
Your 14-day trial starts once the initial sync is complete
Sjsu_260832062316349...
Connectors
Transformations
Destinations
Trial usage estimator
Alerts
Account Settings
Resources & Support
Nikhil Sarma
Give Feedback
Google Sheets
Connection tests:
Finding specified sheet
Validating named range
All connection tests passed!
Back
Continue

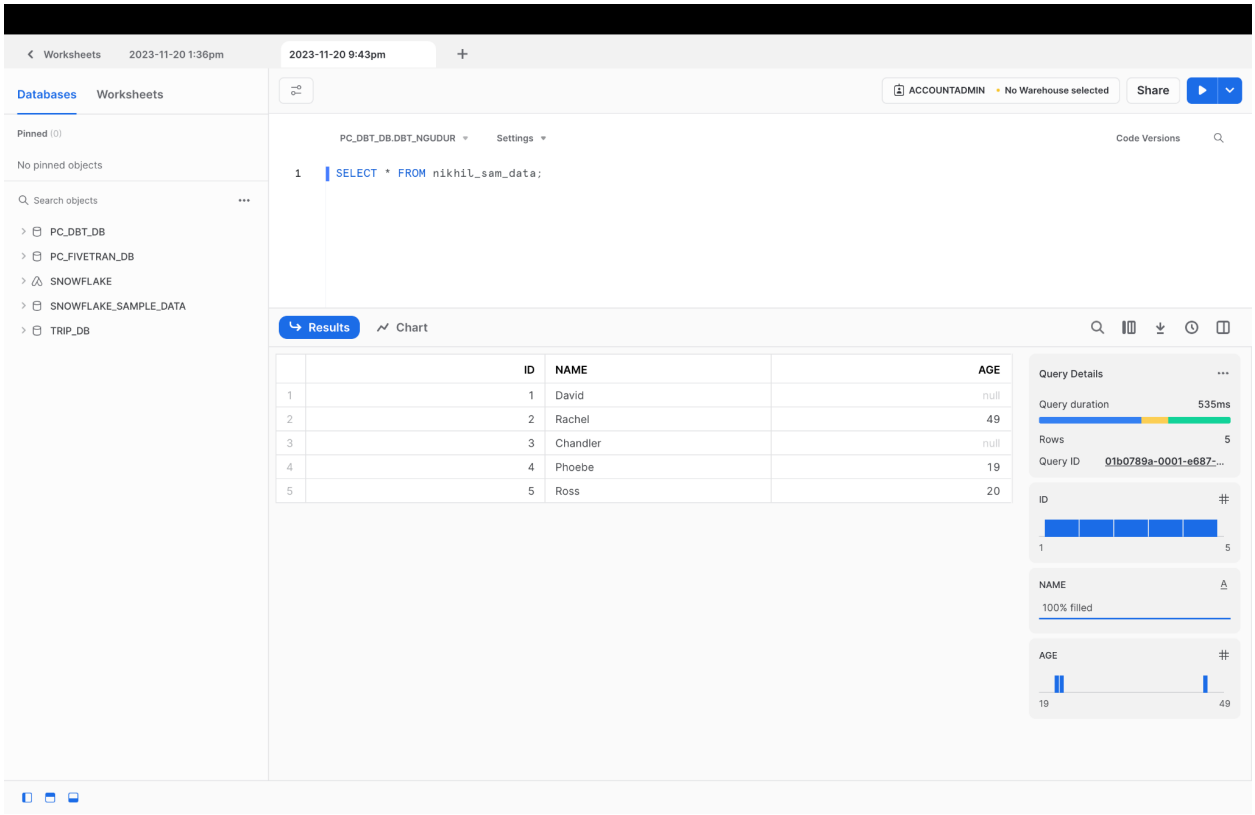

Worksheets
2023-11-20 1:36pm
2023-11-20 2:24pm
Databases
Worksheets
Pinned (0)
No pinned objects
Search objects
PC_FIVETRAN_DB
SNOWFLAKE
SNOWFLAKE_SAMPLE_DATA
TRIP_DB
ACCOUNTADMIN
COMPUTE_WH
Share
PC_FIVETRAN_DB.DB_ASSIGNMENT
Settings
Code Versions
create or replace TABLE PC_FIVETRAN_DB.DB_ASSIGNMENT.LOAD_SAM_DATA (
_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;
Results
Chart
_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 duration 83ms
Rows 9
Query ID 01b076e0-0001-e5c4-...
_ROW
1
9
POSITION
Forward 3
Guard 3
Center 1
+ 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') }}
```

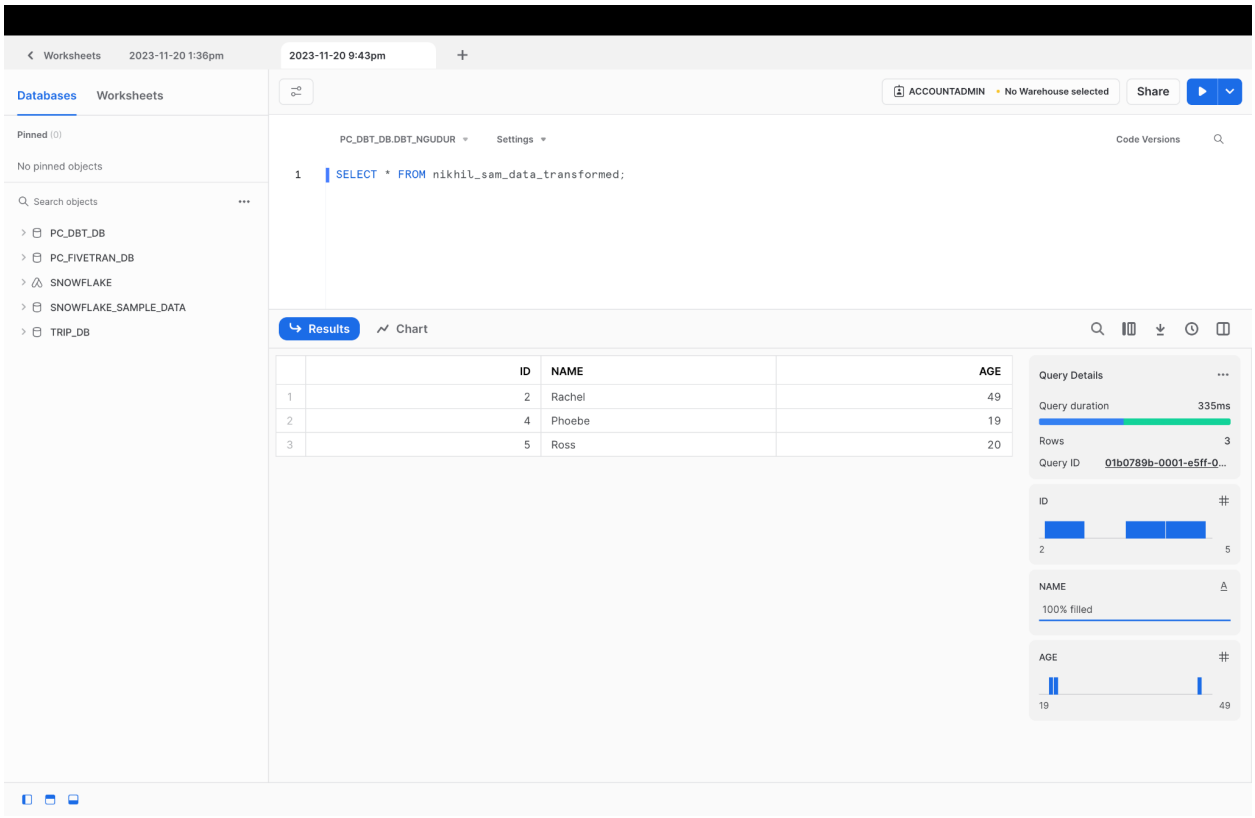
```
SELECT 1 as id, 'David' as name, NULL as age

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
```





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
```


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
ACCOUNTADMIN
No Warehouse selected
Share
PC_DBT_DB.DBT_NGUDUR
Settings
Code Versions
1 SELECT * FROM nikhil_sam_data_transformed;
Results
Chart
ID NAME AGE
1 2 Rachel 49
2 4 Phoebe 19
3 5 Ross 20
Query Details
Query duration 335ms
Rows 3
Query ID 01b0789b-0001-e5ff-0...
ID
2 5
NAME
100% filled
AGE
19 49