Snowflake Summit 2025 Hackathon Guidebook for Participants

*This document is to help guide participants with the easy to follow step by steps to get started with the hackathon.*

# Account Provisioning

Accounts will be provisioned through DataOps.Live. We will send a link to claim your account in a separate email.

# Datasets

1. County Health Rankings:
2. FEMA National Risk Index (NRI)
3. CDC Social Vulnerability Index
4. Primary Healthcare Sites (Federally Qualified Health Centers):
5. Mapbox activity index for time period of Hurricane Ida (2021):
6. NOAA historical hurricane data
7. Power Outage Data
8. Hurricane Ida Tracking Data
9. Overture Maps
10. Homeland Infrastructure Foundation-Level Data (HIFLD)
11. Census data: American Community Survey
12. 2020 US Census: Population Block Group

## Hurricane Ida Path Data

We will provide further guidance in interpreting this data and creating a geospatial visual at the event.

### Wind Swathe

### Trajectory

### Hurricane Points

### Hurricane Lines

## Overture Maps from Carto

Participants will get these datasets directly from the Snowflake Marketplace, they are completely free.

Overture Maps is an open source foundation project: <https://overturemaps.org/open-data-sources-partners/>

* Buildings [https://app.snowflake.com/marketplace/listing/GZT0Z4CM1E9KN/carto-overture-maps-buildings](https://app.snowflake.com/marketplace/listing/GZT0Z4CM1E9KN/carto-overture-maps-buildings?search=overture)
* Places <https://app.snowflake.com/marketplace/listing/GZT0Z4CM1E9KR/carto-overture-maps-places>
* Addresses <https://app.snowflake.com/marketplace/listing/GZT0Z4CM1E9NQ/carto-overture-maps-addresses>
* Divisions <https://app.snowflake.com/marketplace/listing/GZT0Z4CM1E9M9/carto-overture-maps-divisions>
* Base: Land, Land Use, Water, etc. [https://app.snowflake.com/marketplace/listing/GZT0Z4CM1E9KV/carto-overture-maps-base](https://app.snowflake.com/marketplace/listing/GZT0Z4CM1E9KV/carto-overture-maps-base?search=overture)
* Transportation: <https://app.snowflake.com/marketplace/listing/GZT0Z4CM1E9KJ/carto-overture-maps-transportation>

### Homeland Infrastructure Foundation-Level Data (HIFLD)

Background Info and Data Dictionaries: <https://hifld-geoplatform.hub.arcgis.com/>

* The sources below have already been loaded, but you are welcome to download and load more:
  + Download in GeoJSON or WKT
  + Create Geo Data Type, and Snowflake does conversion automatically.
  + You will need a desktop application to convert the files into format 4326 before uploading to Snowflake.
  + You can use ArcGIS (Windows only), or QGIS. Export to the 4326 format, and specify a new column name to contain the “GEOMETRY” data: You can call it whatever you like, in our datasets we call it simply “GEO”
* U.S. County outlines →
  + <https://www.conservation.gov/datasets/17b89622df5643cda5339ae6649247a6_0/about> -
* Hospitals → <https://hifld-geoplatform.hub.arcgis.com/maps/9e318142490c4884bf74932af437c6c2/about>
  + **Table Name**: HOSPITAL\_LOCATIONS
* Nursing homes → <https://hifld-geoplatform.hub.arcgis.com/datasets/426ab97fbcec4f4d872340b4d8630170_0/explore?location=36.192972%2C64.379135%2C2.68>
  + **Table Name**: NURSING\_HOMES
* Dialysis Centers → <https://hifld-geoplatform.hub.arcgis.com/datasets/a600e72a409646cc9b617d104ecb5c60_0/explore?location=30.918705%2C66.347100%2C2.47>
  + **Table Name**: DIALYSIS\_CENTERS
* National Shelter System Facilities → <https://hifld-geoplatform.hub.arcgis.com/datasets/bcaf5fdb3db24c78afee52d4c8a02748_5/explore?location=32.995165%2C66.254524%2C2.56>
  + **Table Name**: NATIONAL\_SHELTER\_FACILITIES

### Census data in Esri’s Living Atlas: American Community Survey

* ACS Population Variables → <https://services.arcgis.com/P3ePLMYs2RVChkJx/arcgis/rest/services/ACS_Total_Population_Boundaries/FeatureServer>
  + **Table Name**: ACS\_POPULATION\_TRACTS
* ACS Race and Hispanic Origin Variables → <https://services.arcgis.com/P3ePLMYs2RVChkJx/arcgis/rest/services/ACS_Population_by_Race_and_Hispanic_Origin_Boundaries/FeatureServer>
  + **Table Name**: ACS\_RACE\_HISPANIC\_ORIGIN\_TRACTS
* ACS Poverty Status Variables → <https://services.arcgis.com/P3ePLMYs2RVChkJx/arcgis/rest/services/ACS_Poverty_by_Age_Boundaries/FeatureServer>
  + **Table Name**: ACS\_POVERTY\_STATUS\_TRACTS
* ACS Median Household Income Variables → <https://services.arcgis.com/P3ePLMYs2RVChkJx/arcgis/rest/services/ACS_Median_Income_by_Race_and_Age_Selp_Emp_Boundaries/FeatureServer>
  + **Table Name**: ACS\_HH\_MEDIAN\_INCOME\_TRACTS
* ACS Vehicle Availability Variables → <https://services.arcgis.com/P3ePLMYs2RVChkJx/arcgis/rest/services/ACS_Vehicle_Availability_Boundaries/FeatureServer>
  + **Table Name**: ACS\_VEHICLE\_AVAILABILITY\_TRACTS

### 2020 Census: Population Block Group

<https://services.arcgis.com/P3ePLMYs2RVChkJx/arcgis/rest/services/USA_Census_2020_DHC_Total_Population/FeatureServer/4>

* **Table Name**: POPULATION\_CENSUS\_BLOCK

### County Health Rankings:

* <https://services.arcgis.com/HRPe58bUyBqyyiCt/arcgis/rest/services/County_Health_Rankings_2024_Test_2/FeatureServer>
  + **Table Name**: COUNTY\_HEALTH\_RANKINGS\_2024

### FEMA National Risk Index (NRI)

* Counties → <https://services.arcgis.com/XG15cJAlne2vxtgt/arcgis/rest/services/National_Risk_Index_Counties/FeatureServer>
  + **Table** **Name**: NATIONAL\_RISK\_INDEX\_COUNTIES
* Census Tracks → <https://services.arcgis.com/XG15cJAlne2vxtgt/arcgis/rest/services/National_Risk_Index_Census_Tracts/FeatureServer>
  + **Table Name**: NATIONAL\_RISK\_INDEX\_TRACTS

### CDC Social Vulnerability Index

Loaded - from archive.

* No longer available from the traditional source in ESRI Living Atlas.
* Esri archive of 2022 data → <https://services2.arcgis.com/FiaPA4ga0iQKduv3/arcgis/rest/services/CDC_SVI_2022_(Archive)/FeatureServer>
  + **Table Name**: SOCIAL\_VULNERABILITY\_INDEX

### Primary Healthcare Sites (Federally Qualified Health Centers):

* <https://services1.arcgis.com/ZGrptGlLV2IILABw/arcgis/rest/services/FQHC_LAL_AllSites_2023/FeatureServer>
  + **Table Name**: FQHC\_SITES
* Original source: US Government data: HRSA, HHS.
  + <https://www.hrsa.gov/>
  + <https://www.hrsa.gov/foia/electronic-reading.html>

### Mapbox activity index for time period of Hurricane Ida (2021):

* <https://docs.mapbox.com/data/movement/guides/activity-index/>

### NOAA historical hurricane data

Ida was extracted from other hurricanes:

* Hurricane tracks: <https://services2.arcgis.com/FiaPA4ga0iQKduv3/arcgis/rest/services/IBTrACS_ALL_list_v04r00_lines_1/FeatureServer>
* Point data: <https://services2.arcgis.com/FiaPA4ga0iQKduv3/arcgis/rest/services/IBTrACS_ALL_list_v04r00_points_1/FeatureServer>

### Power Outage Data

* PowerOutage.US - Posted at county level, and city level in US. CrisisReady has an API Key - a state GEO can be queried when it updates, every 15 minutes.

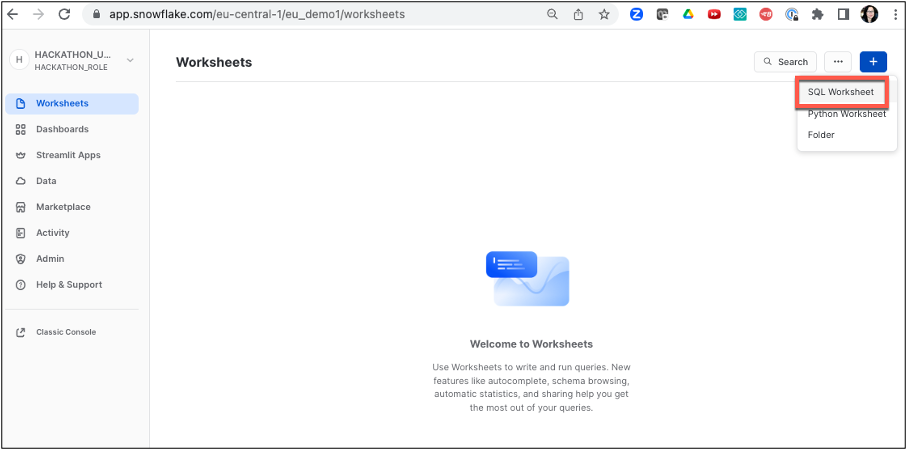
# Data Exploration App in Streamlit

# Uploading Your Own Data

If you wish to upload your own data, you can follow these example steps:

These are from another event using World Bank data, but you can follow along with whatever data you choose.

1- Let’s create your first worksheet by clicking on the + sign and choose SQL Worksheet



A worksheet with a date and time as name will open. Double click on the tab name to rename it to “YOURNAME WORKSHEET” .

2- Run the following commands to create the necessary database, schema & warehouse.

-- create database

CREATE OR REPLACE DATABASE WORLD\_BANK\_DATA

-- create schema

CREATE OR REPLACE SCHEMA RAW\_DATA

-- create warehouses

CREATE OR REPLACE WAREHOUSE HACKATHON\_WH

WAREHOUSE\_SIZE = medium

WAREHOUSE\_TYPE = 'standard'

AUTO\_SUSPEND = 60

AUTO\_RESUME = TRUE

INITIALLY\_SUSPENDED = TRUE;

3- Create a ‘file format’ and the ‘stage’ to upload the csv files. This is an internal stage we are creating within Snowflake which allows us to upload files to snowflake from local machine.

CREATE OR REPLACE FILE FORMAT mycsvformat

TYPE = 'CSV'

FIELD\_DELIMITER = ','

FIELD\_OPTIONALLY\_ENCLOSED\_BY = '"'

SKIP\_HEADER = 1;

CREATE OR REPLACE STAGE my\_csv\_stage

FILE\_FORMAT = mycsvformat;

4- Install SnowSQL. Download the binary and perform installation as per your local laptop’s OS. This is needed to upload files to the internal stage created in the previous step.

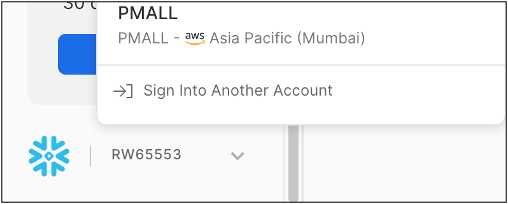
[Hint: Files can also be uploaded to internal stage from drag and drop but the limitation is of 50 MB]

<https://docs.snowflake.com/en/user-guide/snowsql-install-config>

5- Update the SnowSQL configuration file using the following instructions.

<https://docs.snowflake.com/en/user-guide/snowsql-config#modify-the-snowsql-configuration-file>

Get the necessary details at the bottom left of Snowsight UI. You can hover on the account details to get details needed for the config file.



Update the following sections in the “.snowsql/config” file

[connections.example]

#Can be used in SnowSql as #connect example

accountname = <locator>.ap-southeast-1

username = <Snowflake username>

password = <Snowflake password>

6- Upload the files to the internal stage. Follow the instructions given here.

<https://docs.snowflake.com/en/user-guide/data-load-internal-tutorial-stage-data-files>

Connect to snowsql by default connection ‘example’ and then change the database and schema to ‘WORLD\_BANK\_DATA’ and ‘RAW\_DATA’.

Run the following commands one at a time.

> ~ % snowsql -c example

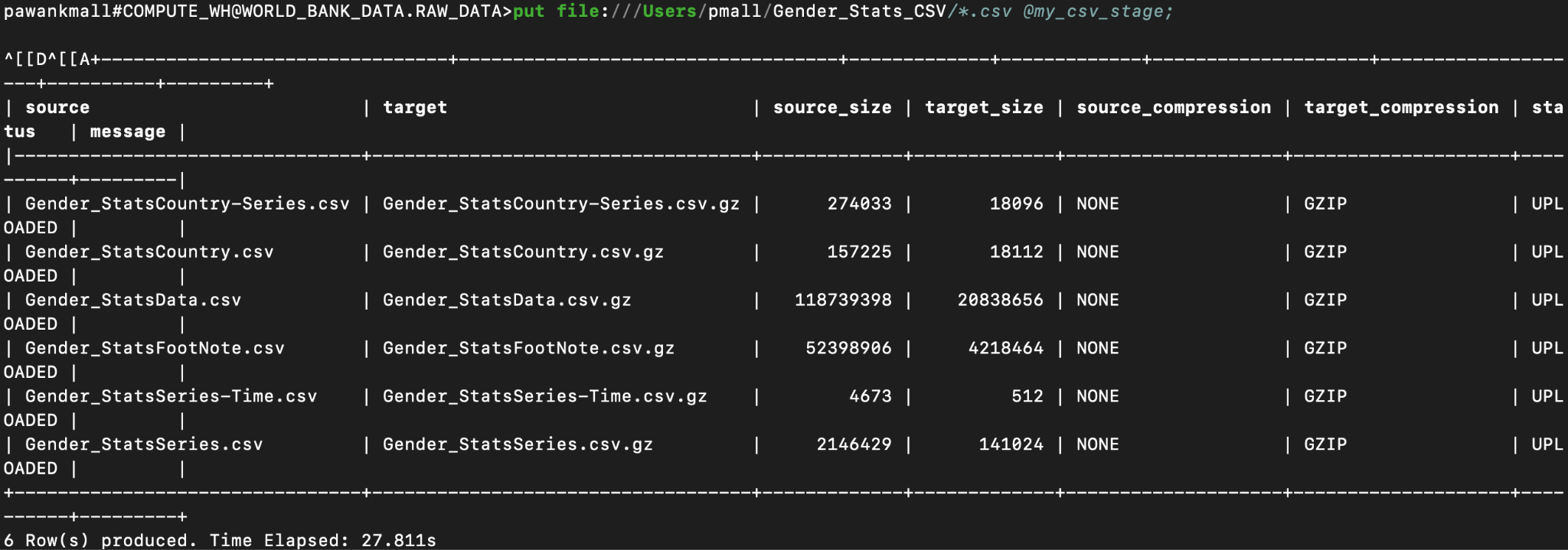
**>use** **database** WORLD\_BANK\_DATA;

**>use** **SCHEMA** RAW\_DATA;

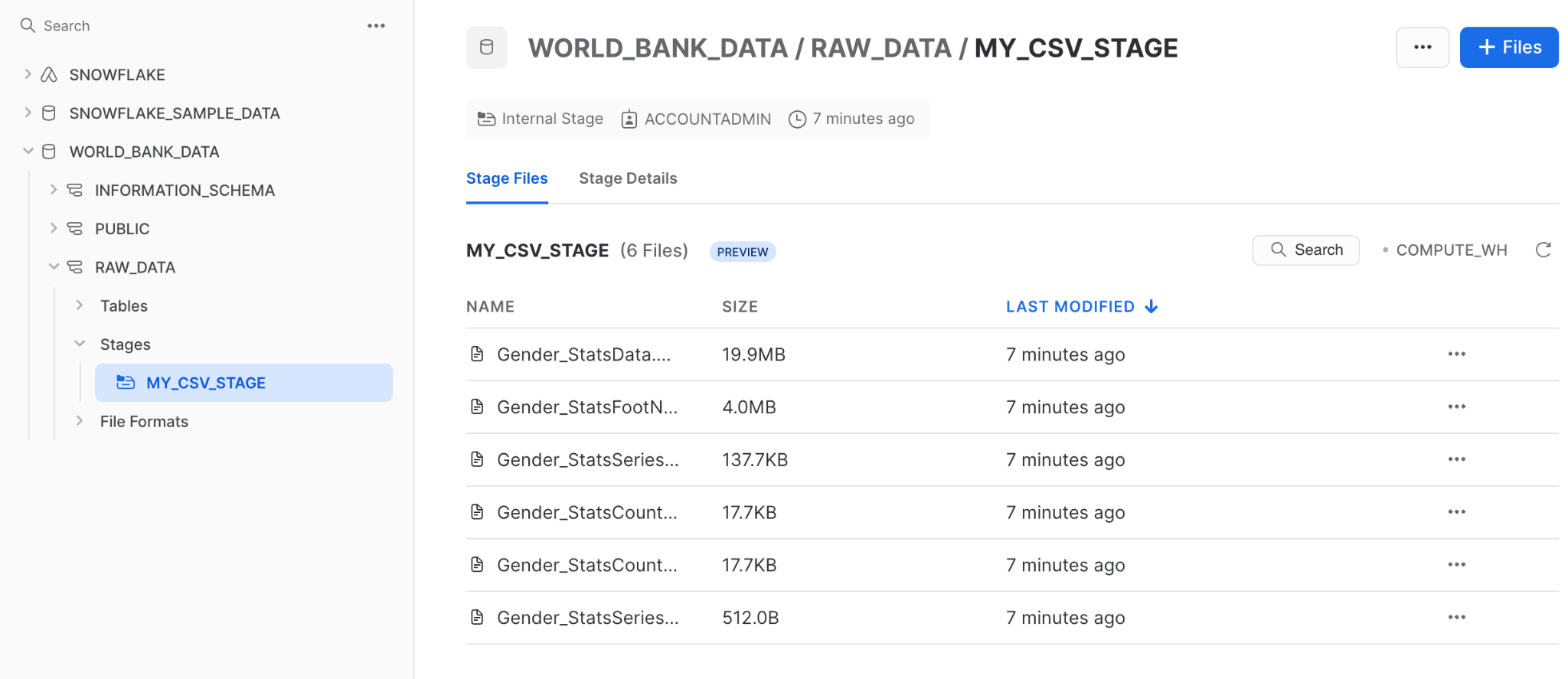
Run the following command pointing to the folder location of the downloaded files in your laptop. This will upload the files from your local machine to Snowflake’s internal stage.

**put** **file**:///**Users**/<path>/Gender\_Stats\_CSV*/\*.csv @my\_csv\_stage;*

You will see the files zipped and uploaded to the Snowflake’s internal stage we created earlier



You can verify the files uploaded on Snowsight



7- Once the files are uploaded. Go ahead and create the tables based on the column names of the files. You can create the table by using simple DDL statements from the worksheet.

create or replace TABLE WORLD\_BANK\_DATA.RAW\_DATA.GENDER\_STATS\_COUNTRY (

COUNTRY\_CODE VARCHAR(16777216),

SHORT\_NAME VARCHAR(16777216),

TABLE\_NAME VARCHAR(16777216),

LONG\_NAME VARCHAR(16777216),

"2\_ALPHA\_CODE" VARCHAR(16777216),

CURRENCY\_UNIT VARCHAR(16777216),

SPECIAL\_NOTES VARCHAR(16777216),

REGION VARCHAR(16777216),

INCOME\_GROUP VARCHAR(16777216),

WB\_2\_CODE VARCHAR(16777216),

NATIONAL\_ACCOUNTS\_BASE\_YEAR VARCHAR(16777216),

NATIONAL\_ACCOUNTS\_REFERENCE\_YEAR VARCHAR(16777216),

SNA\_PRICE\_VALUATION VARCHAR(16777216),

LENDING\_CATEGORY VARCHAR(16777216),

OTHER\_GROUPS VARCHAR(16777216),

SYSTEM\_OF\_NATIONAL\_ACCOUNTS VARCHAR(16777216),

ALTERNATIVE\_CONVERSION\_FACTOR VARCHAR(16777216),

PPP\_SURVEY\_YEAR VARCHAR(16777216),

BALANCE\_OF\_PAYMENTS\_MANUAL\_IN\_USE VARCHAR(16777216),

EXTERNAL\_DEBT\_REPORTING\_STATUS VARCHAR(16777216),

SYSTEM\_OF\_TRADE VARCHAR(16777216),

GOVERNMENT\_ACCOUNT\_CONCEPT VARCHAR(16777216),

IMF\_DATA\_DISSEMINATION\_STANDARD VARCHAR(16777216),

LATEST\_POPULATION\_CENSUS VARCHAR(16777216),

LATEST\_HOUSEHOLD\_SURVEY VARCHAR(16777216),

SOURCE\_OF\_MOST\_RECENT\_INCOME\_AND\_EXPENDITURE VARCHAR(16777216),

VITAL\_REGISTRATION\_COMPLETE VARCHAR(16777216),

LATEST\_AGRICULTURAL\_CENSUS VARCHAR(16777216),

LATEST\_INDUSTRIAL\_DATA VARCHAR(16777216),

LATEST\_TRADE\_DATA VARCHAR(16777216)

);

create or replace TABLE WORLD\_BANK\_DATA.RAW\_DATA.GENDER\_STATS\_COUNTRY\_SERIES (

COUNTRY\_CODE VARCHAR(16777216),

SERIES\_CODE VARCHAR(16777216),

DESCRIPTION VARCHAR(16777216)

);

8- If you have the tables created from the CSV files using DDLs. Please go ahead and run copy into command to load the table from CSV staged in the internal stage.<https://docs.snowflake.com/en/user-guide/data-load-internal-tutorial-copy-into>

Use the following template to load the files from Internal Stage to the tables created.

COPY INTO "WORLD\_BANK\_DATA"."RAW\_DATA"."GENDER\_STATS\_FOOT\_NOTE"

FROM '@"WORLD\_BANK\_DATA"."RAW\_DATA"."MY\_CSV\_STAGE"'

FILES = ('Gender\_StatsFootNote.csv.gz')

FILE\_FORMAT = (

TYPE=CSV,

SKIP\_HEADER=1,

FIELD\_DELIMITER=',',

TRIM\_SPACE=TRUE,

FIELD\_OPTIONALLY\_ENCLOSED\_BY='"',

DATE\_FORMAT=AUTO,

TIME\_FORMAT=AUTO,

TIMESTAMP\_FORMAT=AUTO

ERROR\_ON\_COLUMN\_COUNT\_MISMATCH = FALSE

)

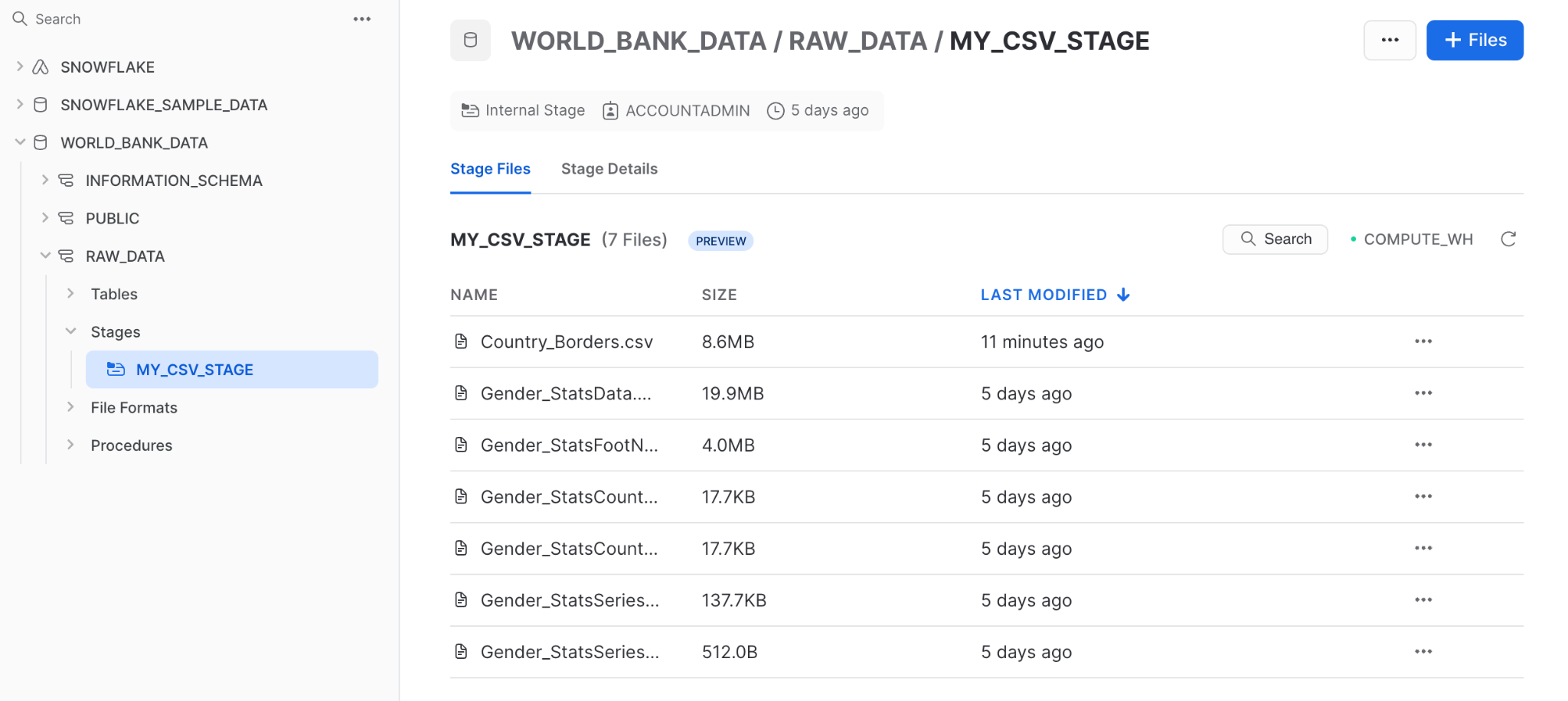
ON\_ERROR=ABORT\_STATEMENT

[ Please note we are using ERROR\_ON\_COLUMN\_COUNT\_MISMATCH=FALSE as there is additional blank column in the source csv files]

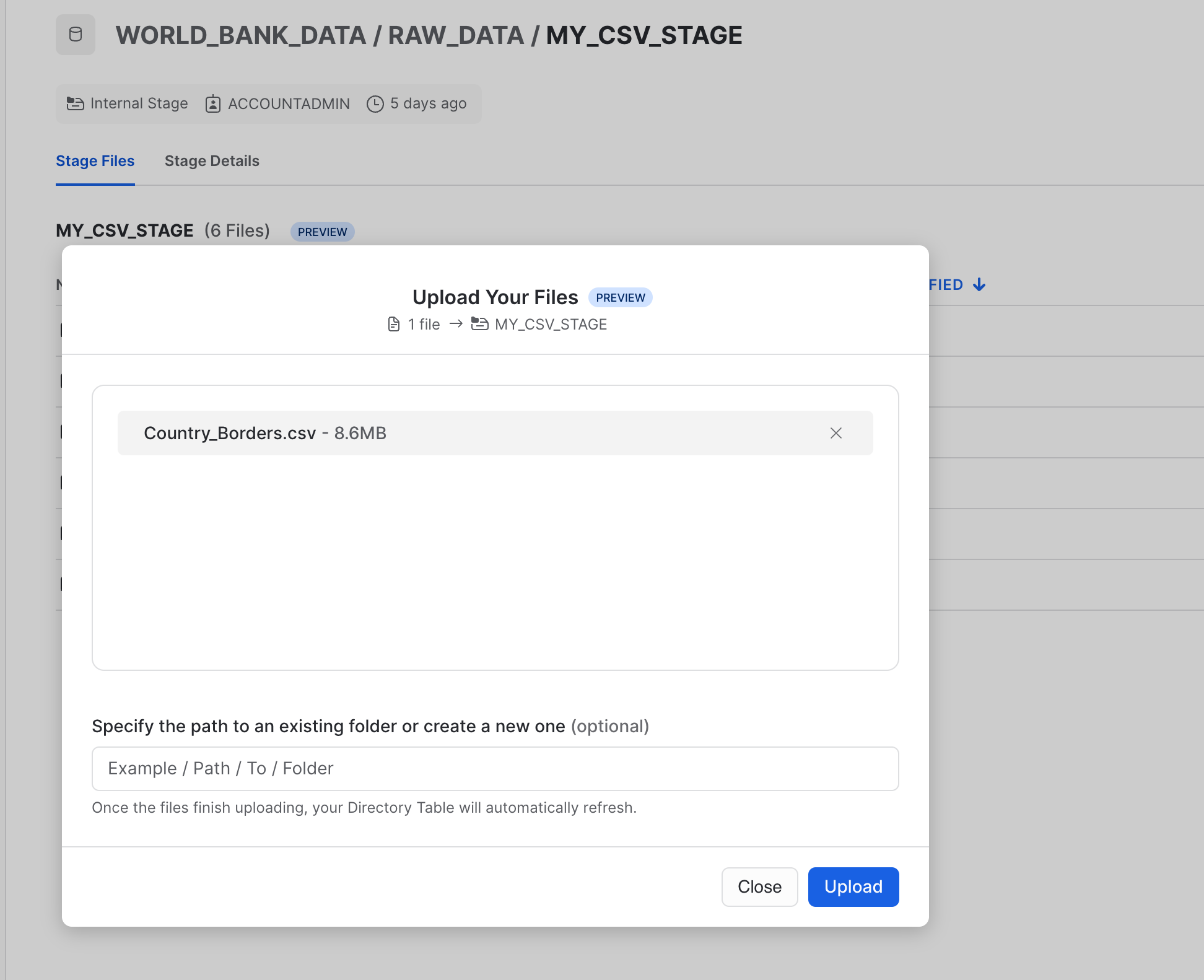
GENDER\_STATS\_DATA will take approximately 3 mins when running on medium warehouse due to the large number of columns.

You can also use direct upload functionality in Snowsight to upload a file.

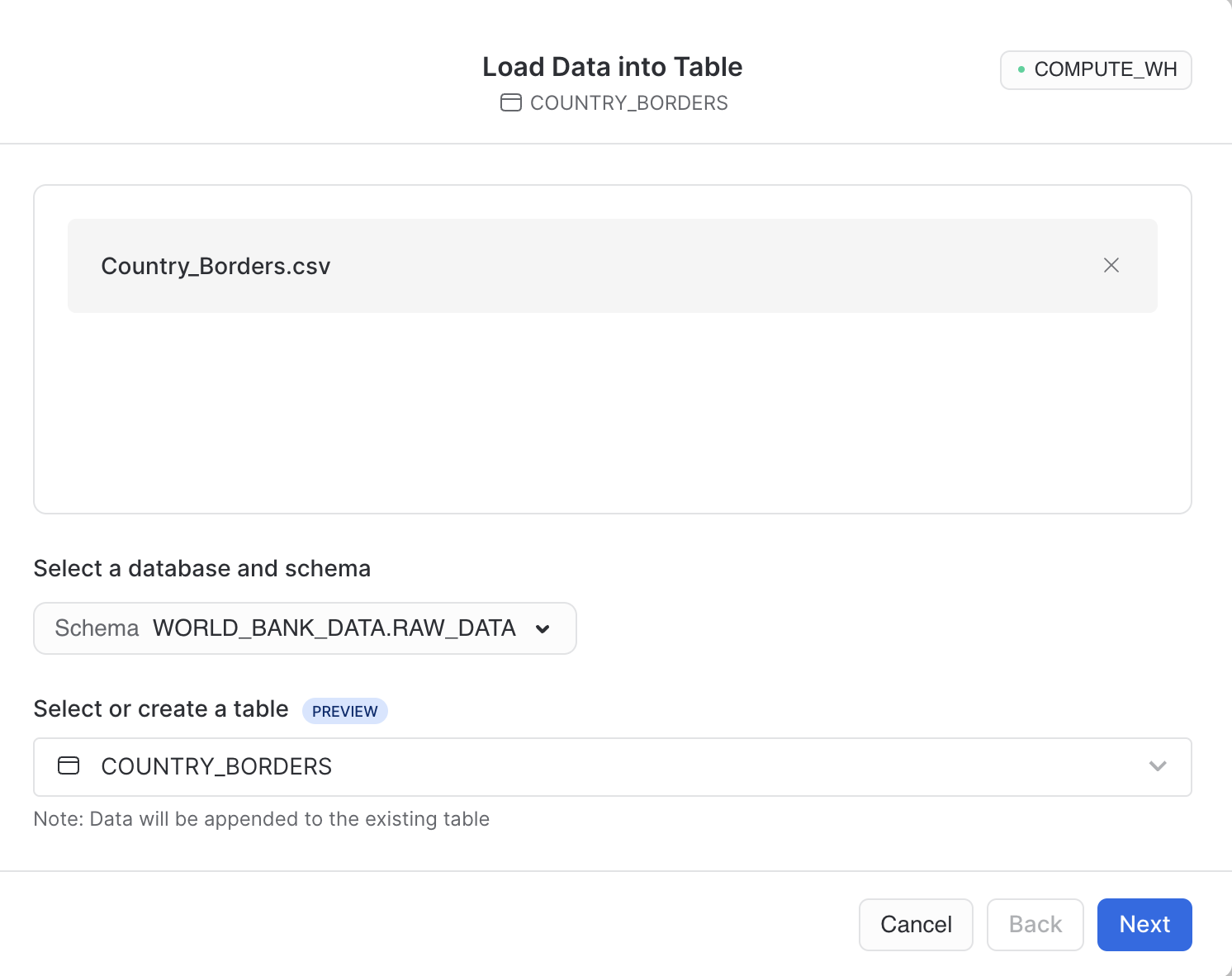
* Navigate to the internal stage created during earlier exercise (my\_csv\_stage).
* Click on ‘+Files’



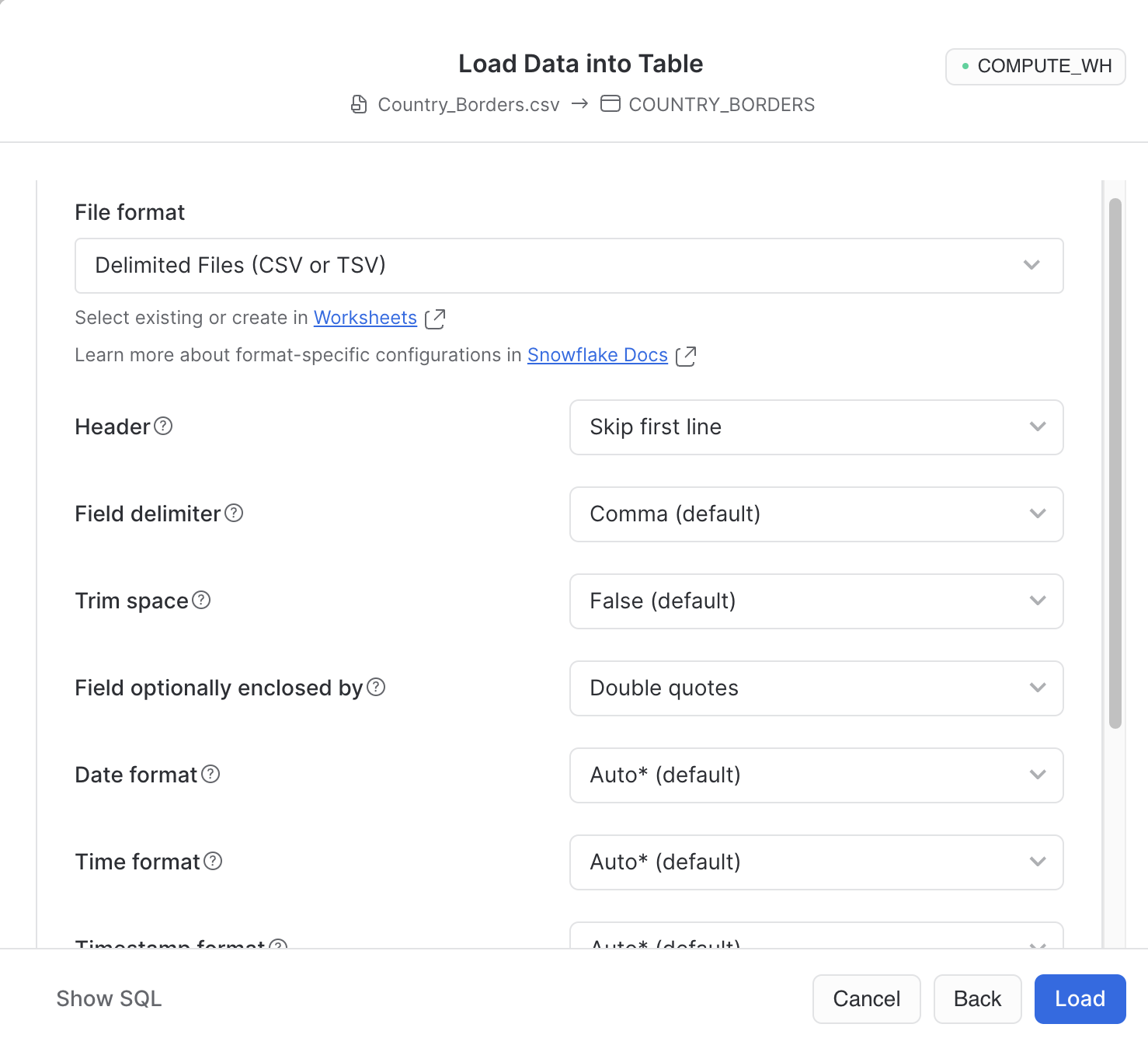
* Upload the file from the location you saved earlier on your laptop.



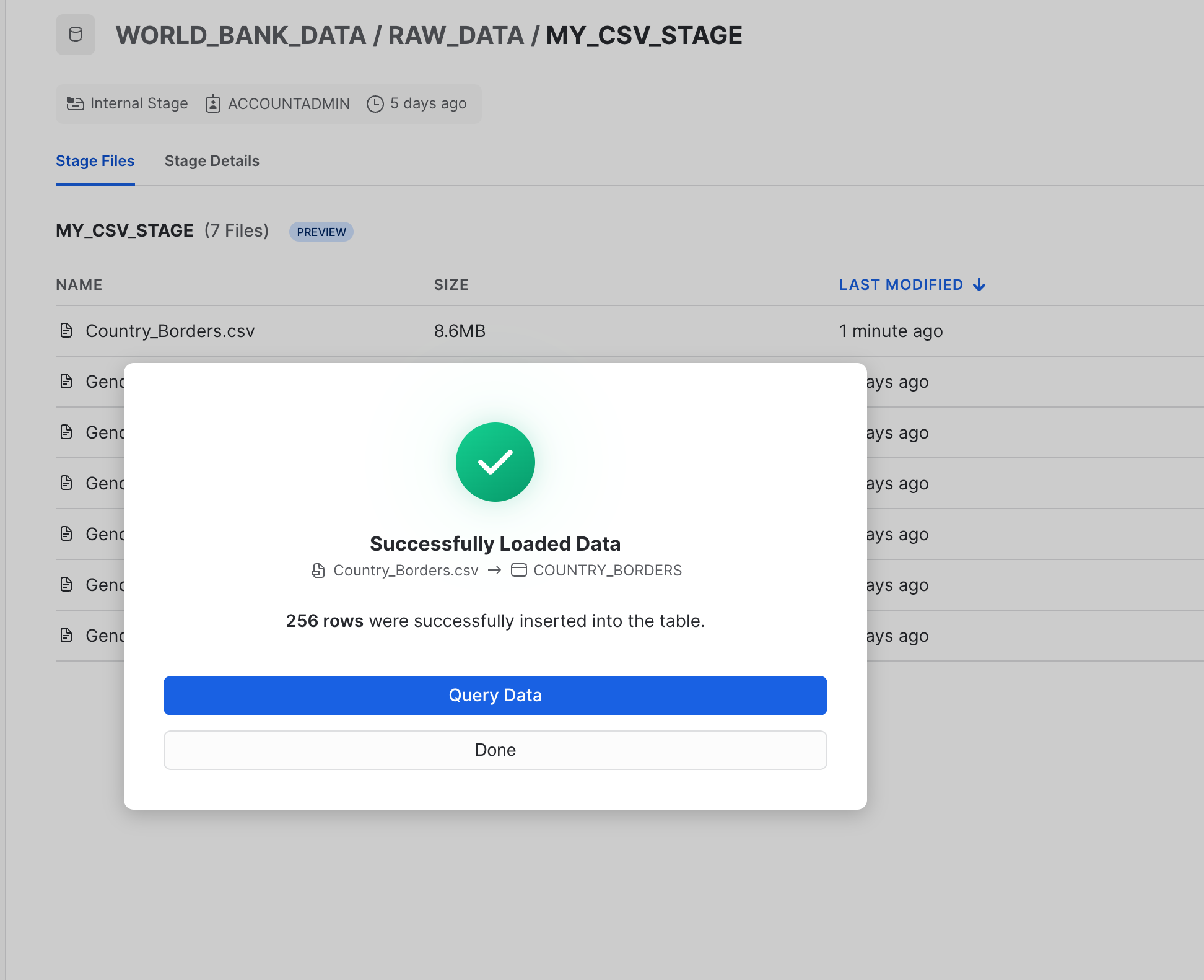
* Select the Schema & table as shown below.



* Select ‘Skip first line’ for Header and “Double quotes” for Field optionally enclosed by.
* Click on Load.



* You should see 256 records loaded. Click on ‘Query Data’ and verify if the data has been loaded successfully.

****

Once all the files are loaded proceed to the next step.

# Working with Snowsight

## Setting your context

Open a new worksheet and run the following commands to set the context.

use role SYSADMIN;

use warehouse hackathon\_wh;

use database world\_bank\_data;

use schema raw\_data;

If you have not used Snowsight before, here are some tips;

* ⌘+Enter (mac) or [CTRL]+[ENTER] (windows) will run the statement(s) that are currently selected, or, if there is no selection, it will execute the statement that the cursor is currently on.
* If you can't see any or some of the panels, they can be shown/hidden using the three small buttons on the bottom left.



* Snowsight can do some cool data profiling for you in the results pane. Remember to take a look at the bottom right hand side of the window and click around the dataset when you run a query that returns data.
* If you need to comment out lines of code press ⌘+/ (Mac) or [CTRL]+/ (Windows) to comment or uncomment multiple lines of code. Alternatively, if your keyboard does not support this, you can add two dashes (--) at the beginning of each line or /\* and \*/ around blocks of code that should be commented out.

# Intro to Streamlit

It's at this stage that Snowsight's simple charting tools might not be enough. We want to more quickly explore this data to find other metrics - like maybe poverty or stability metrics - in the dataset that might help us identify relationships in the data. For this, we can use Streamlit!

Streamlit is an open-source Python library that makes it easy to build data apps, including the ability to display and style data, draw charts and maps, add interactive widgets, customize app layouts, cache computation, and define themes.

# Working with Streamlit

## Creating Your First Streamlit App

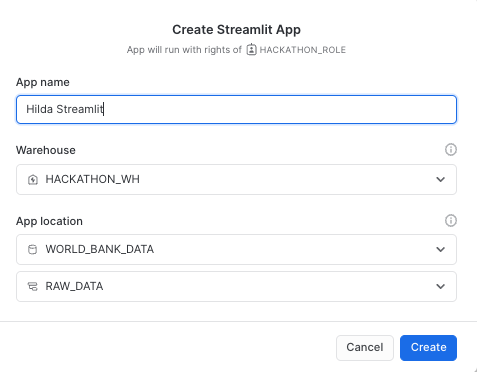
Navigate to the “Home” icon



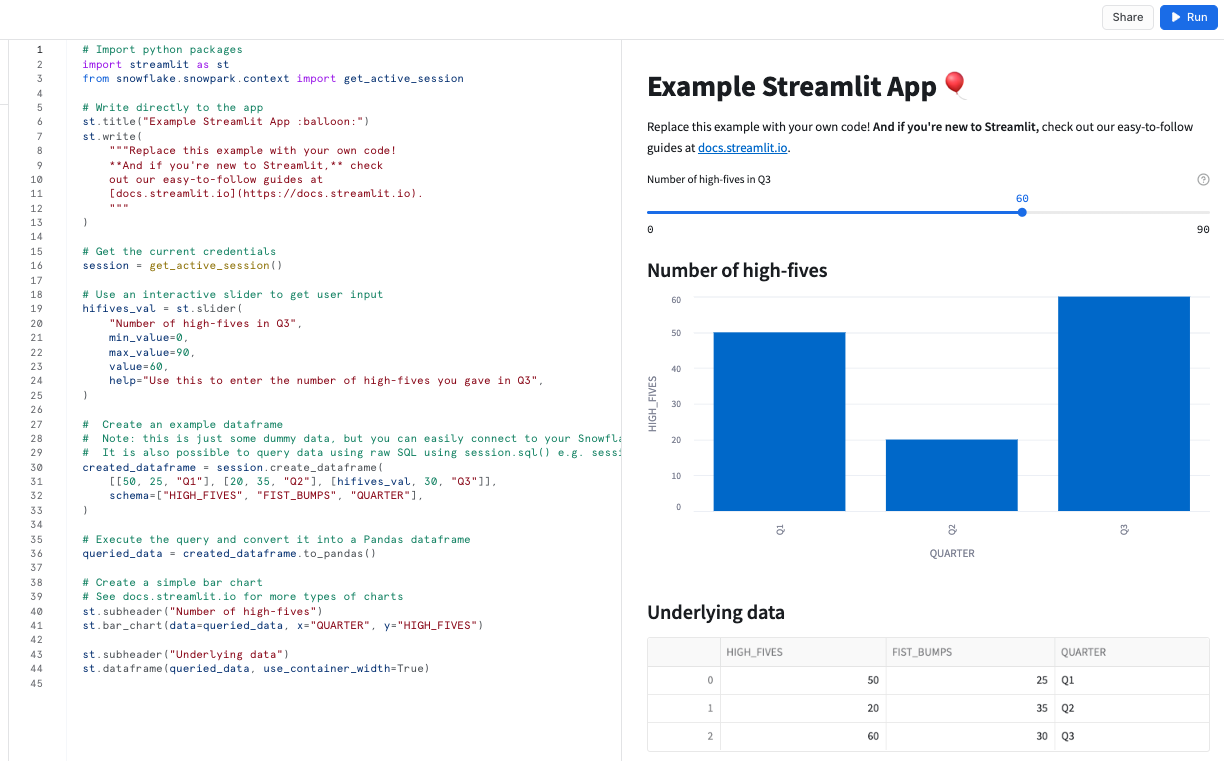
From the home page, navigate to the “Streamlit Apps” , and then click on the “+ Streamlit App” button



It will prompt you to fill in the relevant details. Call your App: “YOURNAME STREAMLIT” and click “Create”



By default, a sample Streamlit App will be created for you.



# Appendix

* To get a Snowflake trial account, sign up [here](https://signup.snowflake.com/) (30 days and $400 in credit)
* If you would like to repeat this on your own, find the World Bank Gender Stats Data [here](https://databank.worldbank.org/data/download/Gender_Stats_CSV.zip)
* To find out more about Snowflake Education and Training, go to this link [here](https://www.snowflake.com/en/resources/learn/training/)