LIS NEPAL
A YÓMARÍ COMPANY

LIS Nepal Pvt. Ltd.

Setup Guide for Workshop

Prepared for:

LOCUS Workshop Participants

January 2023

Contents

Document Revision History

Version	Date	Modified By	Comments
v0.1	05/09/2022	Workshop Team	Installation Guide
v0.2	05/11/2022	Workshop Team	Data Source connection in Power BI
v0.3	01/03/2023	Workshop Team	Formatting and restructuring

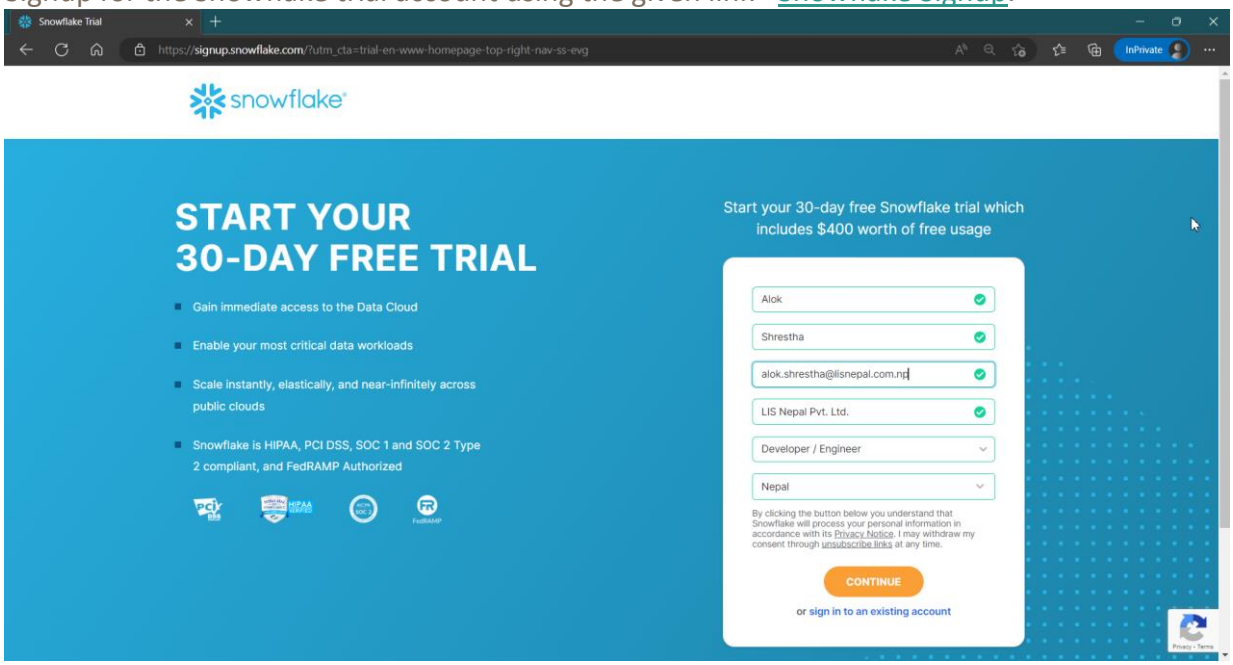
1 Setting up the Database (Snowflake)

For this workshop, we will use Snowflake, which is a cloud-based data warehousing company.

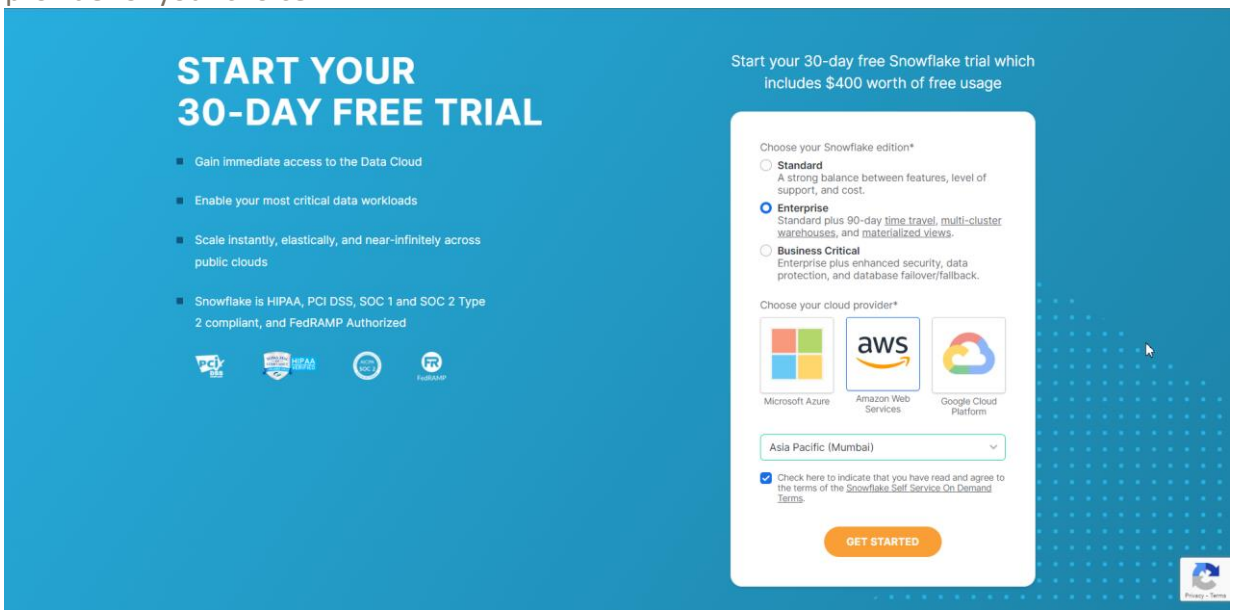
1.1 Creating a Snowflake account

Please follow the following steps to set up the Snowflake account.

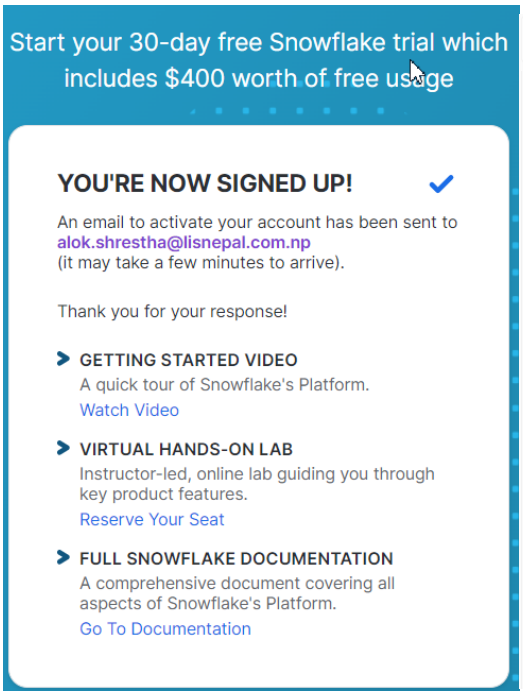
1. Signup for the Snowflake trial account using the given link - Snowflake Signup.

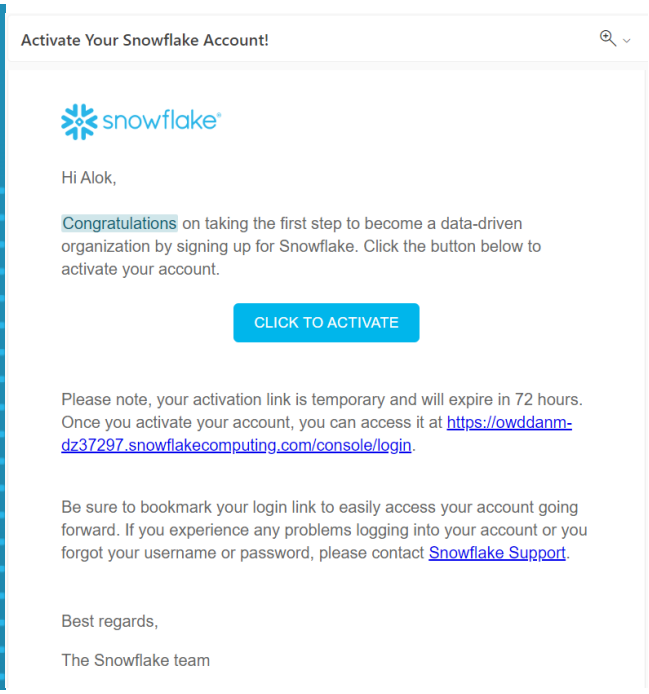


2. You can begin your free Snowflake trial by choosing any Snowflake edition and cloud provider of your choice.

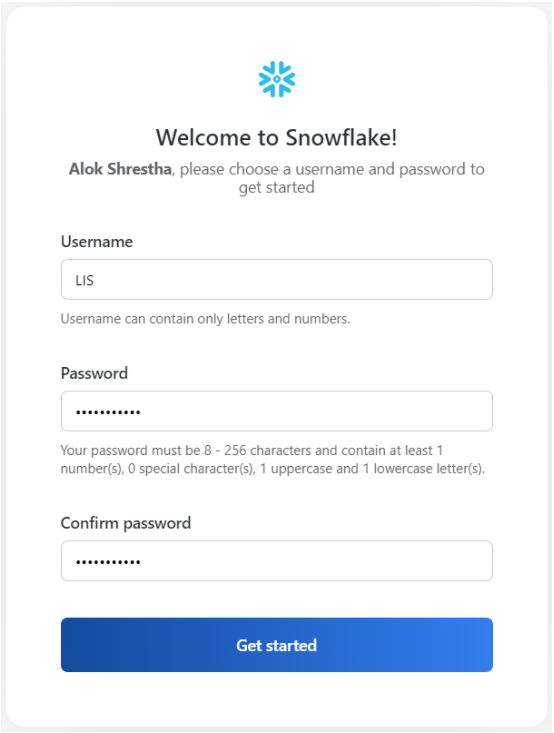


3. Click get started to receive the email containing the activation link for the Snowflake account and the link to access the cloud database as shown below.

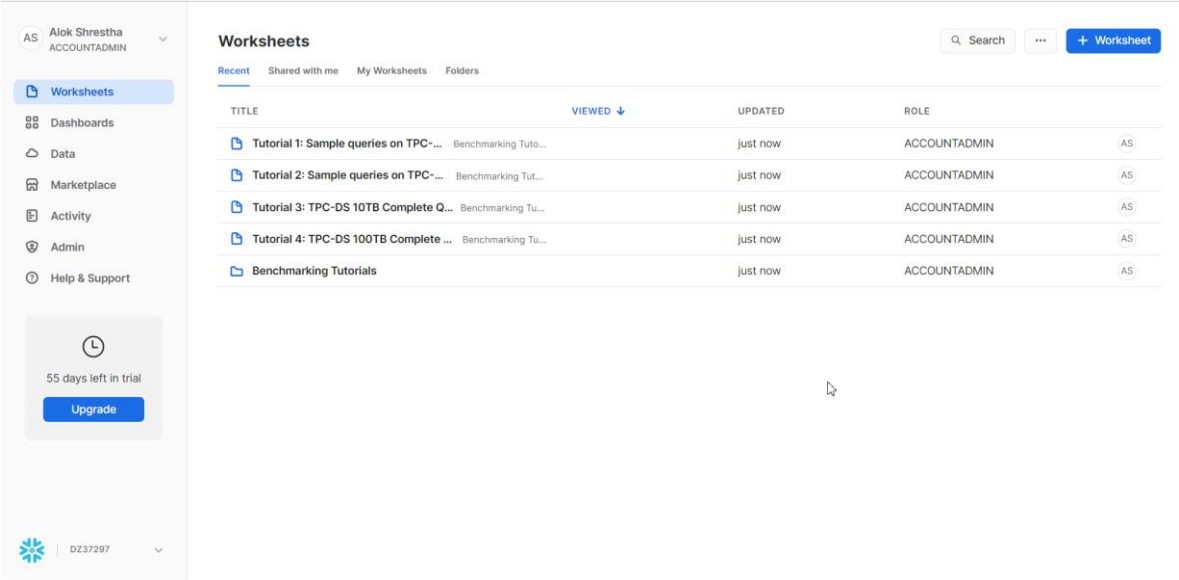




4. To activate the account, provide a username and password for the Snowflake account.

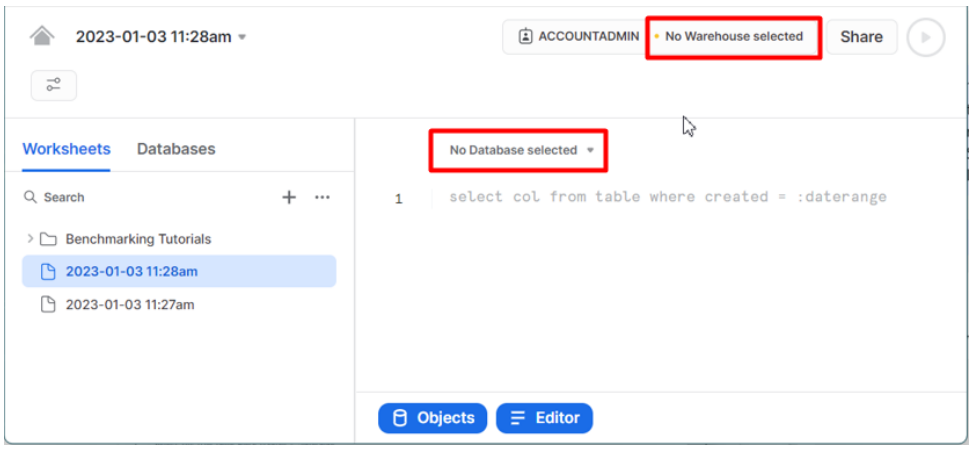


5. Once all these steps are completed, your Snowflake account is ready and you are all set to create your own database. Use the link from your email and use your credentials to open the Snowflake console.

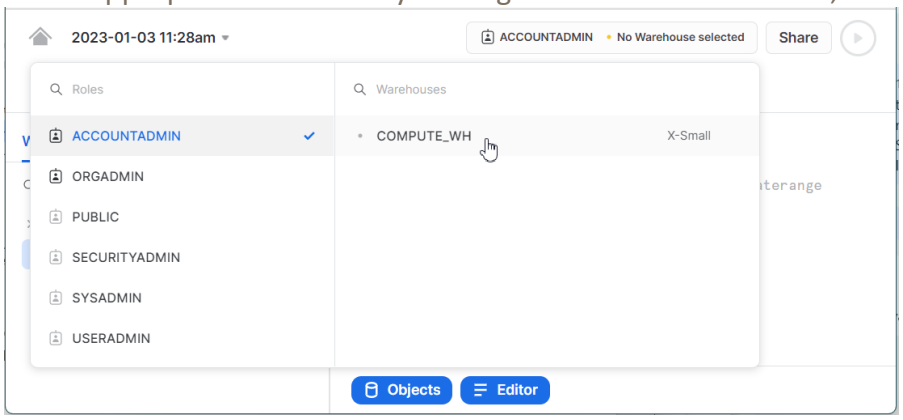


1.2 Creating database definitions

1. You can run your queries using Worksheet. A worksheet is basically a workspace where you can create and submit SQL queries. Click the "+ Worksheet" button on the top right to see the worksheet editor screen.



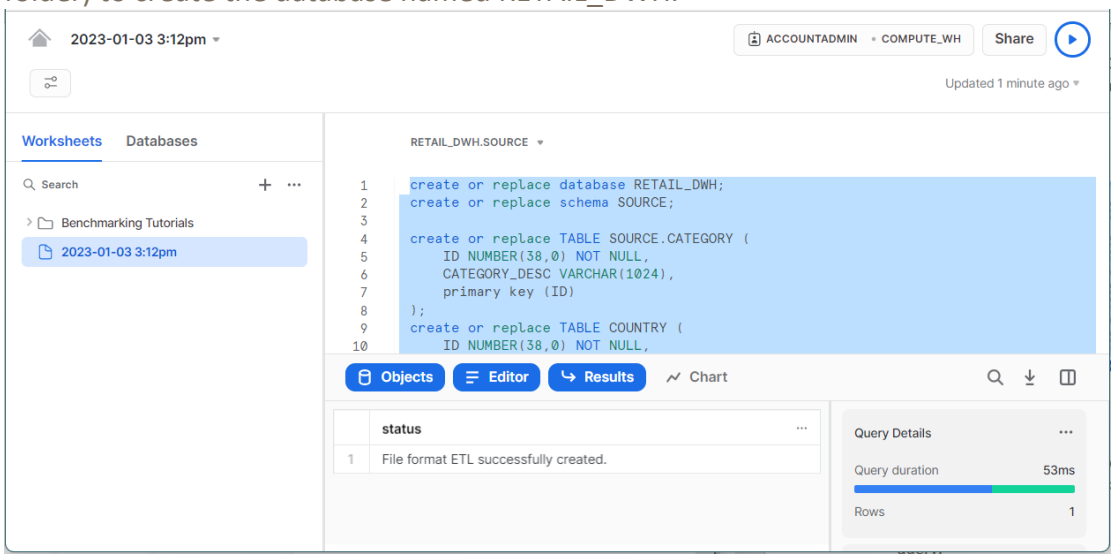
2. If the screen shows "No Warehouse selected", select default COMPUTE_WH warehouse. Select appropriate database by clicking "No Database Selected", when needed.



NOTES: For all the following steps please make sure to

a. Copy the queries in the provided file/s to a new worksheet (Files are provided separately)
b. Select the appropriate Warehouse and database from above the worksheet editor
c. Select the entire queries in the editor before executing them (Click Play button on the top right to execute)

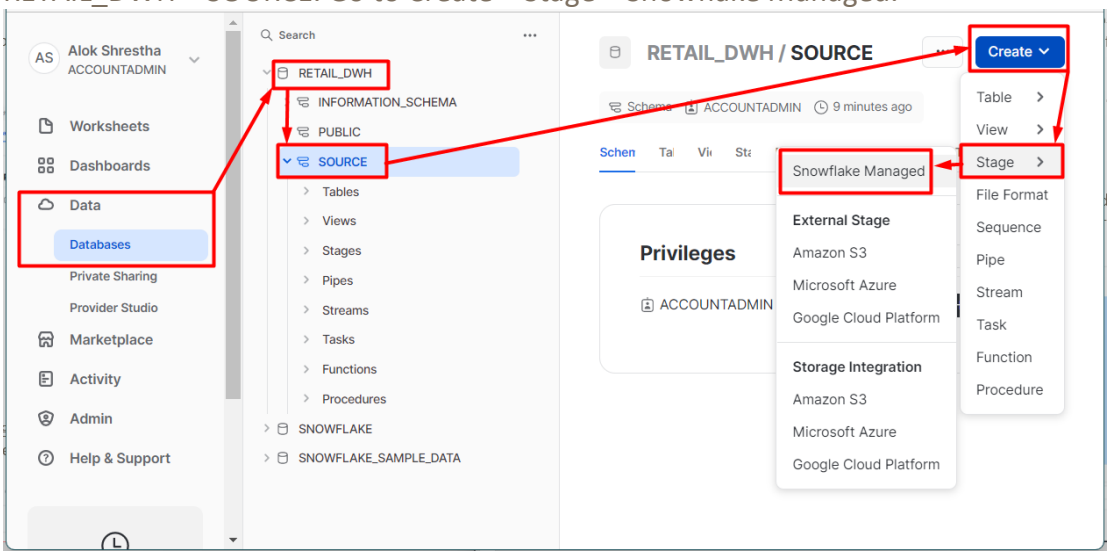
3. Create database: Using the SQL DDL (Data Definition Language), we will create our own database and tables. Execute the queries in the file (sql\tables source.sql, file inside sql folder) to create the database named RETAIL_DWH.



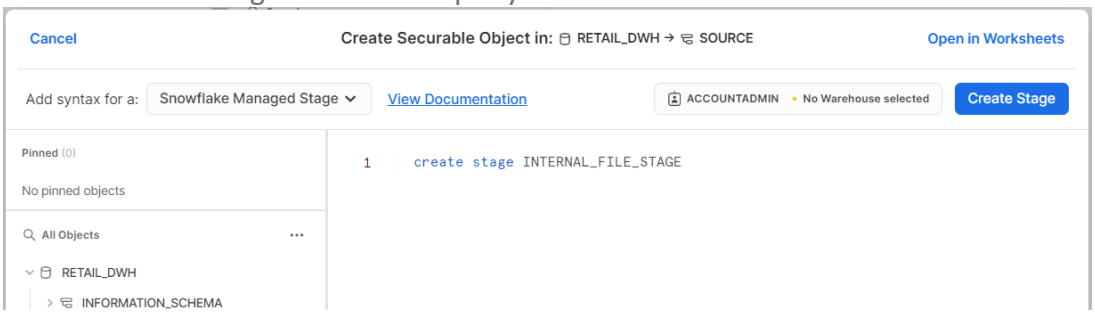
4. Insert Data to Database: Now let's insert the data into the database we just created by executing the queries in the file (sql\insert_source.sql). Select the database (RETAIL_DWH.SOURCE) before executing the query. You can verify the insertion of data by navigating to the Databases section in the leftmost panel of the Snowflake console.

6. Creating File Load Staging: We will create the stage location, to use for loading data from files into Snowflake tables and unloading data from tables into files. To do this, go to the Snowflake UI and follow the steps as below:

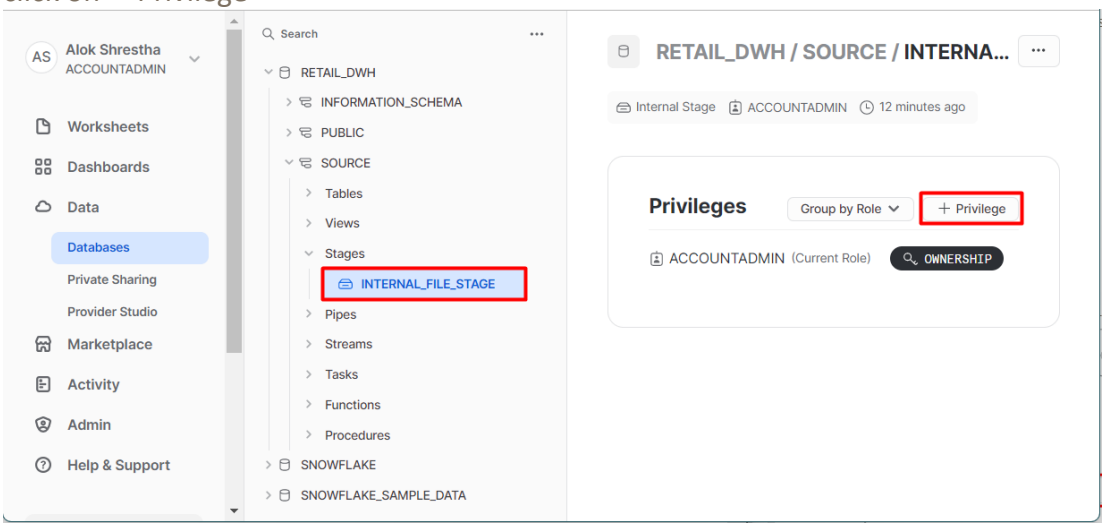
a. Go to Data > Databases from the leftmost panel on Home Screen. Click on the RETAIL_DWH > SOURCE. Go to Create > Stage > Snowflake Managed.



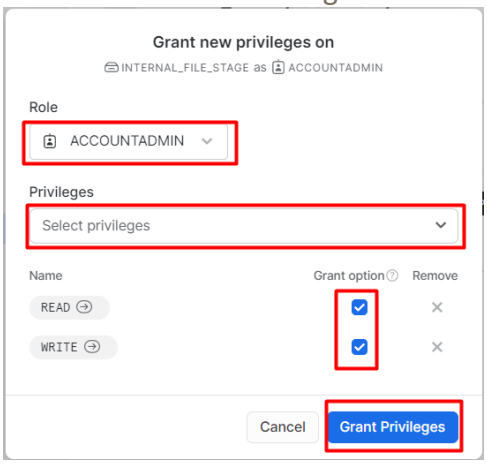
b. This will open a new worksheet with a create stage command in it. Rename the name of the stage to “INTERNAL_FILE_STAGE” so that the query is: create stage INTERNAL_FILE_STAGE Make sure the name is same as mentioned above, since this is used in config. Execute this query.



c. Next, we need to grant privileges of read and write to the current role. To do this, click on “+Privilege”

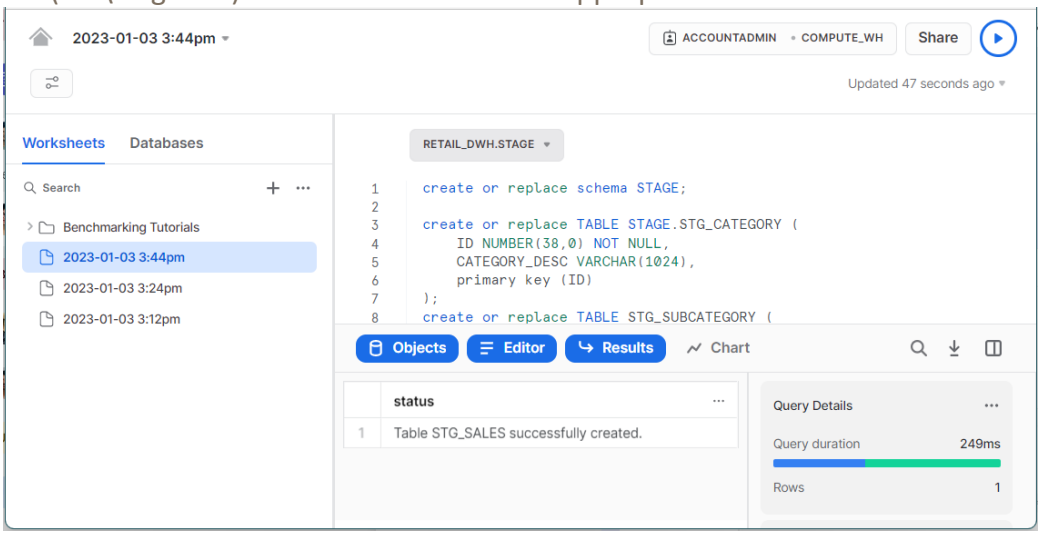


d. Select the used role and grant both read and write privileges to the role.

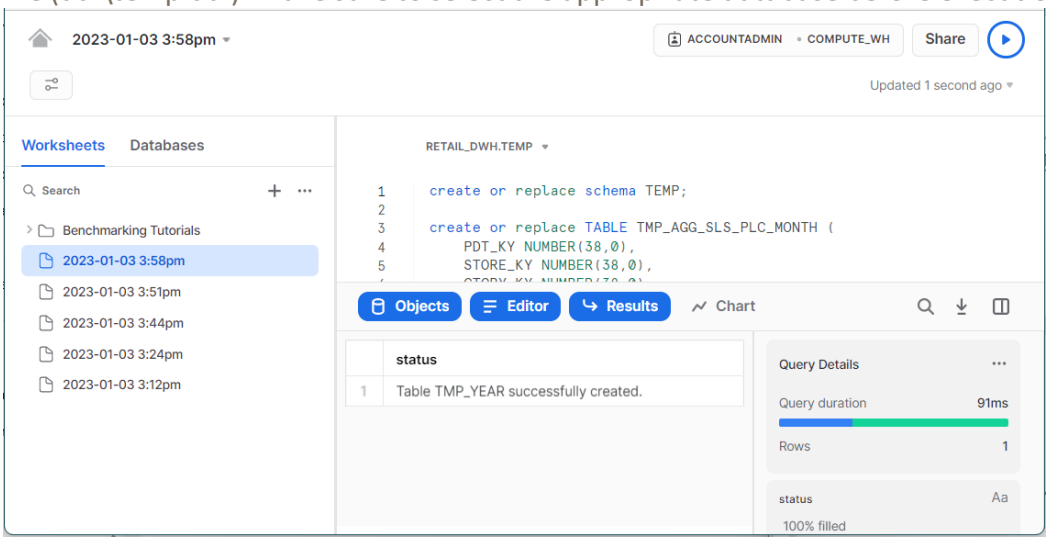


7. Creating Schema: Now we will create the schema definition for our workshop. We will create 3 schemas (stage, temp, and target).

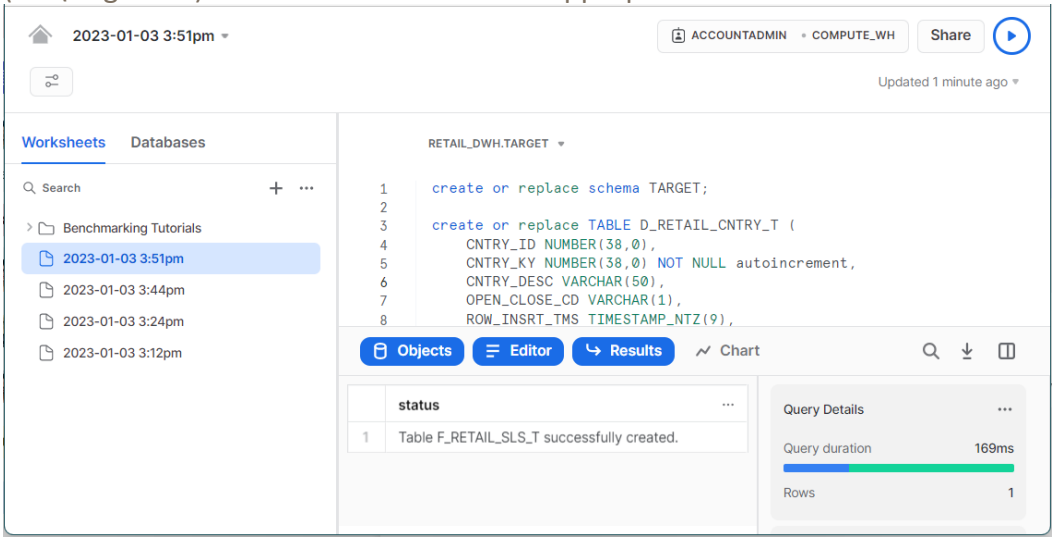
a. First, we will create the stage schema. For that, we will execute the queries in the file (ddl\stage.ddl). Make sure to select the appropriate database before execution.



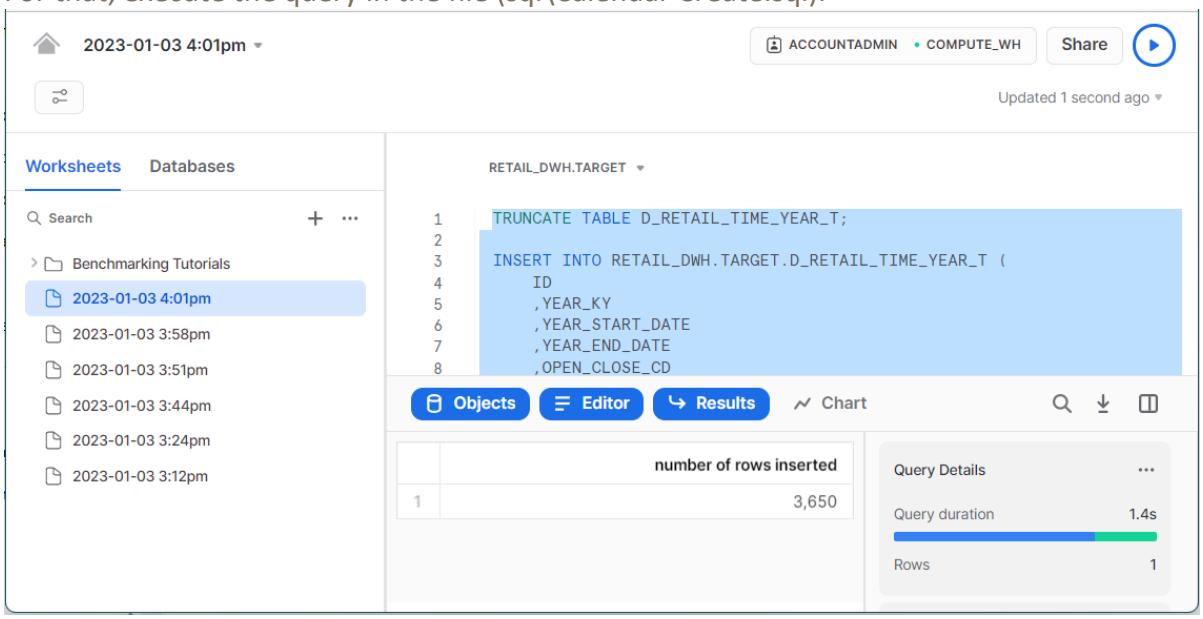
b. Now we will create the temp schema. For that, we will execute the queries in the file (ddl\temp.ddl). Make sure to select the appropriate database before execution.



c. Finally, we will create the target schema by executing the queries in the file (ddl\target.ddl). Make sure to select the appropriate database before execution.



8. Insert Seed Data to Calendar: Next, we perform calendar seeding in the Target schema. For that, execute the query in the file (sql\Calendar Create.sql).

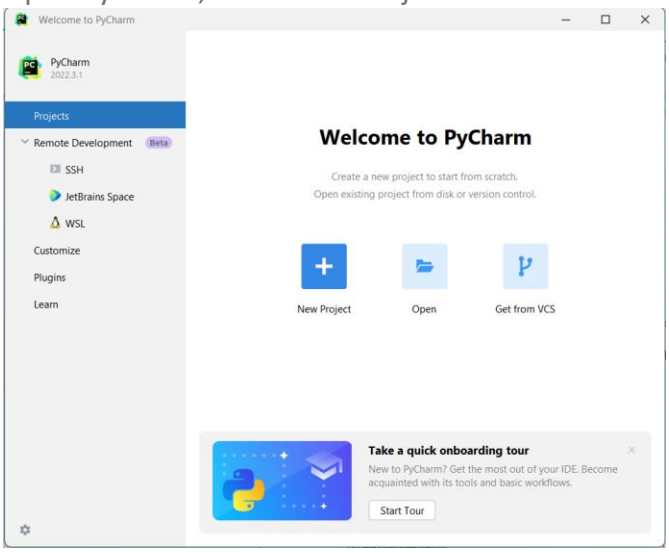


2 Setting up the environment for Python

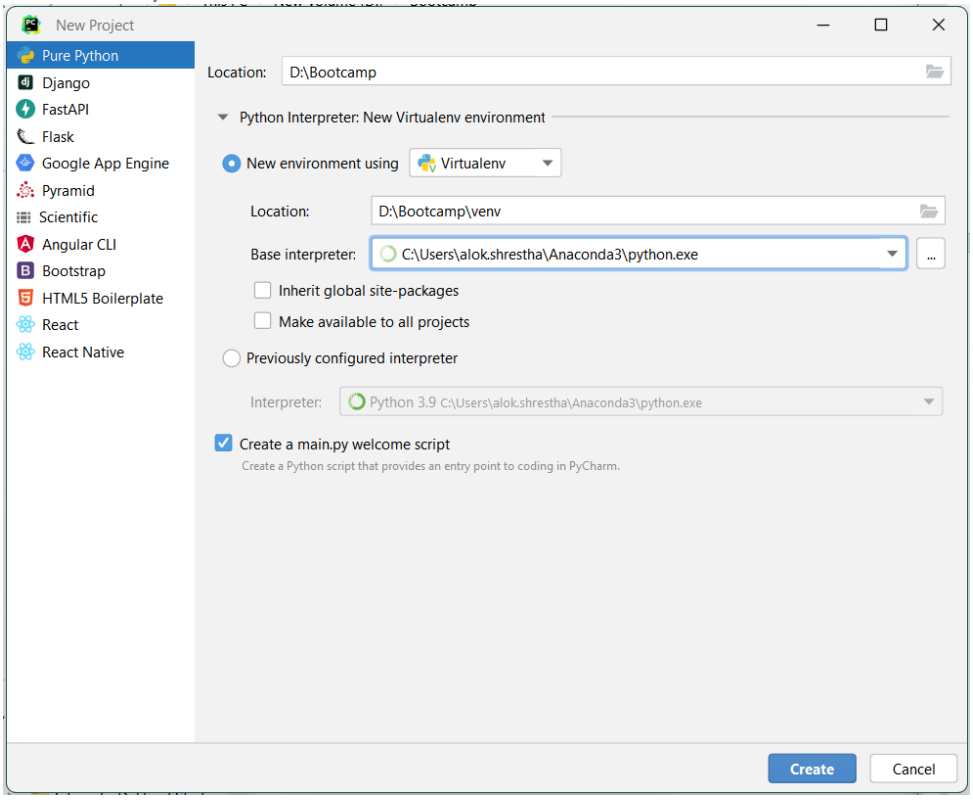
We will use the latest version of Python and PyCharm (Community Edition) as IDE/code editor.

1. Installing Python and PyCharm
    a. Please download and install the latest version of the Python using the following link - Python Download. Make sure to add the Python to PATH variables in Windows.
    b. Download and install PyCharm, using the following link - PyCharm Download. While installing, make sure to set the checkbox option “Add bin folder to PATH”

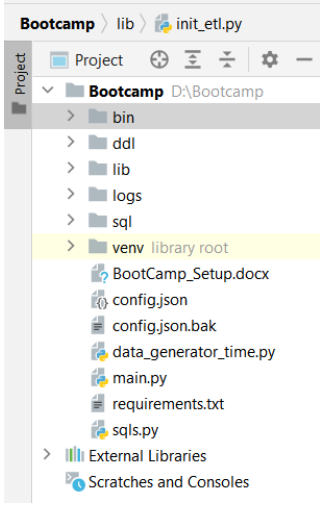
2. Setting up PyCharm: Once the above are installed, we will setup project for our workshop.
    a. Open PyCharm, click “New Project” on the welcome screen to create a project.



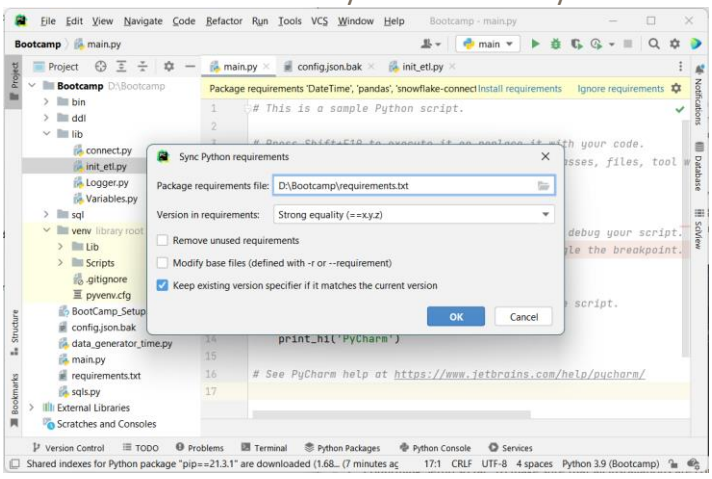
    b. PyCharm allows us to create a new environment in multiple ways. We will create a one for our project using Virtualenv. Make sure to update Location field to the location of your choice. Create the folder beforehand if needed.



c. Once virtualenv is created, please extract the zip file you received into the folder you entered in Location field above. Once extracted, you will see the folder structure as shown below in your PyCharm IDE. (If you do not see it, Right Click any file/folder > Reload from disk to refresh the folder structure.)



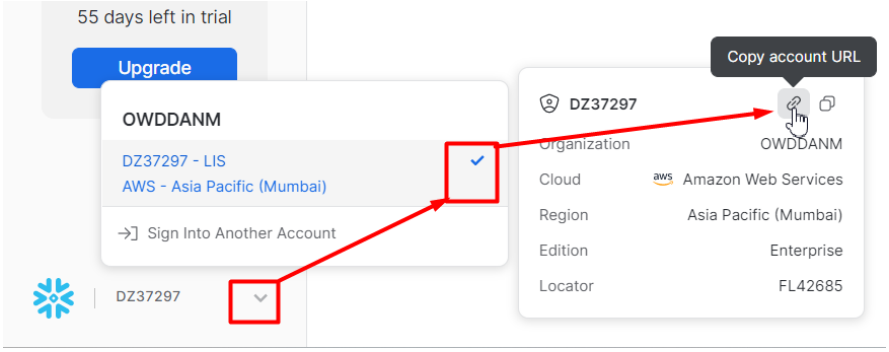
3. Installing Requirements: Now we will install the required packages for our project.
   a. We will import the packages using the Tool > Sync Python Requirements to sync the requirements for our project.
   b. Choose the requirements.txt file provided to sync the requirements for our project, which is also in the same PyCharm directory. Make sure to provide the full location.



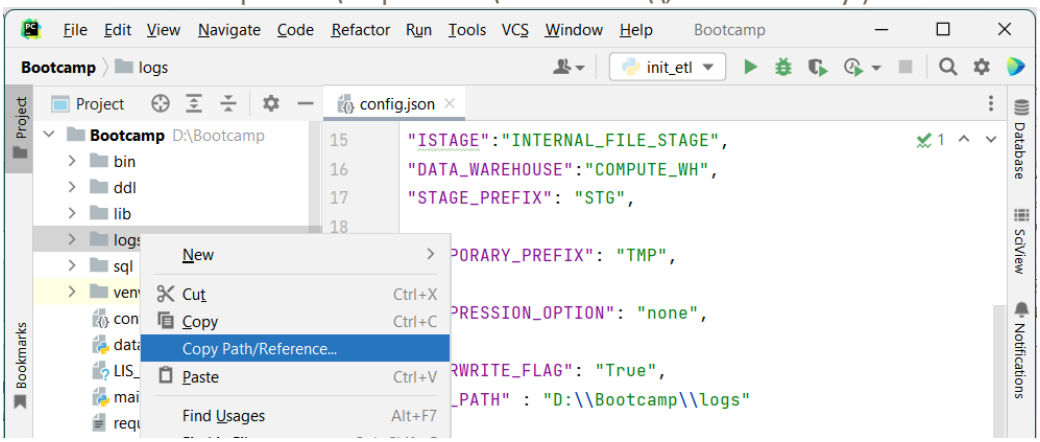
   c. You will see the prompt asking for the installation of the requirements. Accept the prompt to install of the packages required for our Workshop.

4. Updating Configuration: Now we will update the configuration file for this Workshop. Rename the config.json.bak file in your PyCharm directory to config.json. This file contains the configuration for the Workshop. Please open the config.json file and update the following with appropriate values.
   a. USER = “Your Snowflake Username”,
   b. PASSWORD = “Your Snowflake Password”,

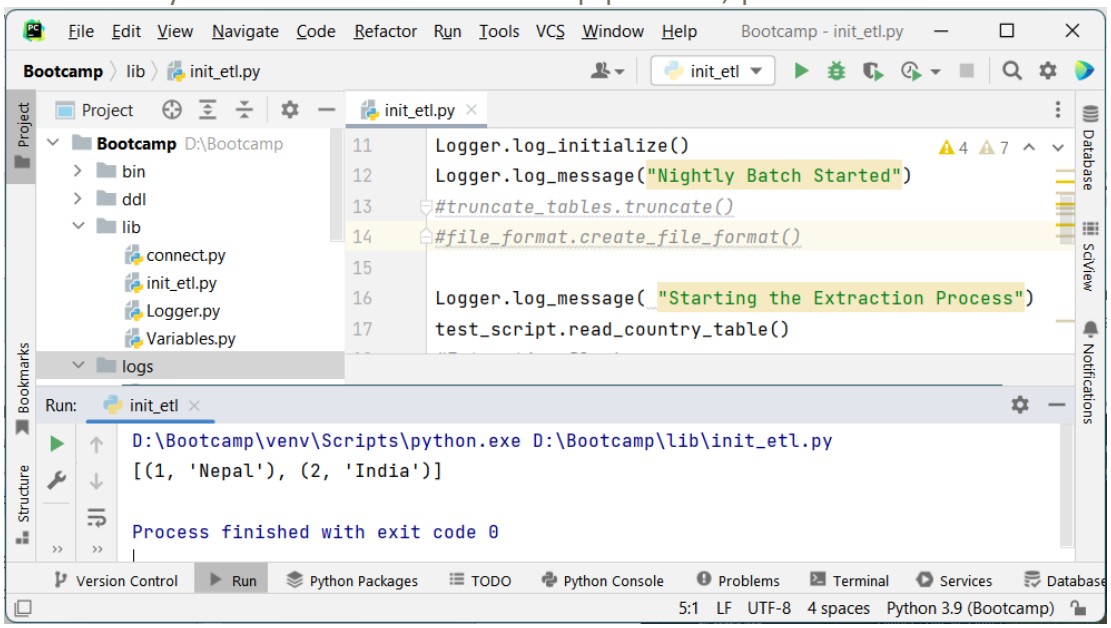
c. ACCOUNT = (If your Snowflake account URL is 'https://uk28978.ap-south-1.aws.snowflakecomputing.com' then your account is 'uk28978.ap-south-1.aws')



d. LOG_PATH = "Path to the log folder" (For log path, please create a folder in the project folder then we can navigate to the log folder and copy its absolute path. Make sure to escape the \ in path i.e \ should be \\, if not already.)



5. Confirming Setup so far: To make sure that all installations are correct, we have provided you with a small code snippet that will fetch the data from the database. Please execute the lib\init_etl.py file (Right Click > Run 'init_etl') from your PyCharm directory. This should result displaying the name of the country as shown below. If you get the output as shown below then you are all done with the setup part else, please recheck the above steps.

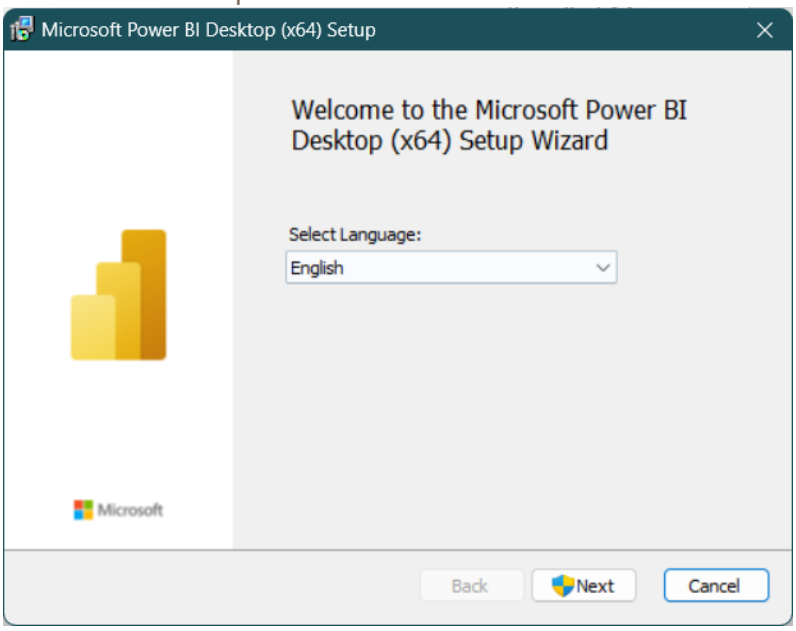


3 Setting up Microsoft Power BI

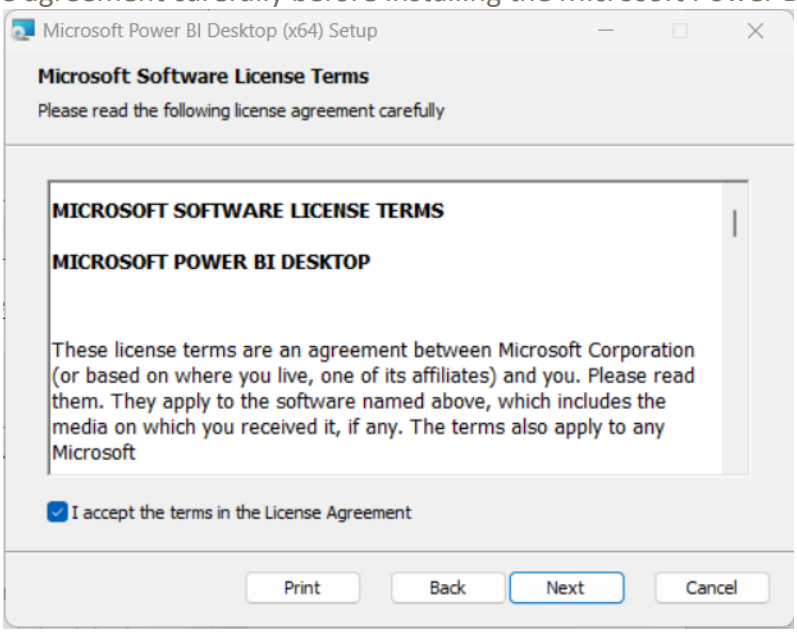
Now we will install Microsoft Power BI, which is an interactive visualization software.

3.1 Installing Microsoft Power BI

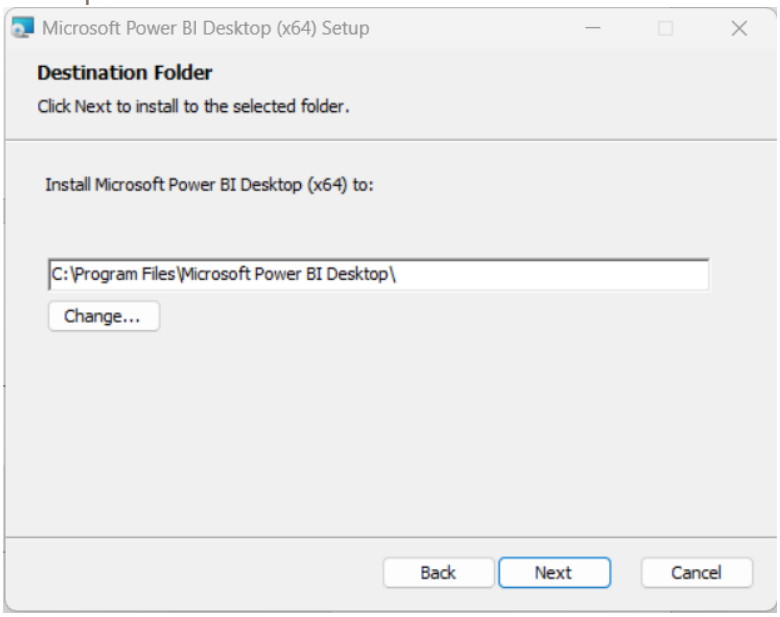
1. Using the following link - [Power BI Downloads](), please download and install Power BI Desktop version using the executable setup.



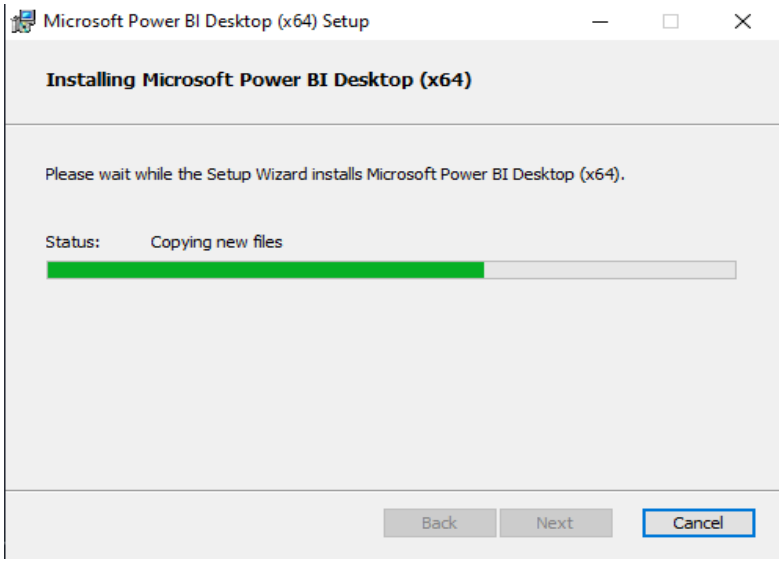
2. Read the license agreement carefully before installing the Microsoft Power BI.



3. Navigate the folder path where you want to install Microsoft Power BI, alternatively you can leave it to default path.



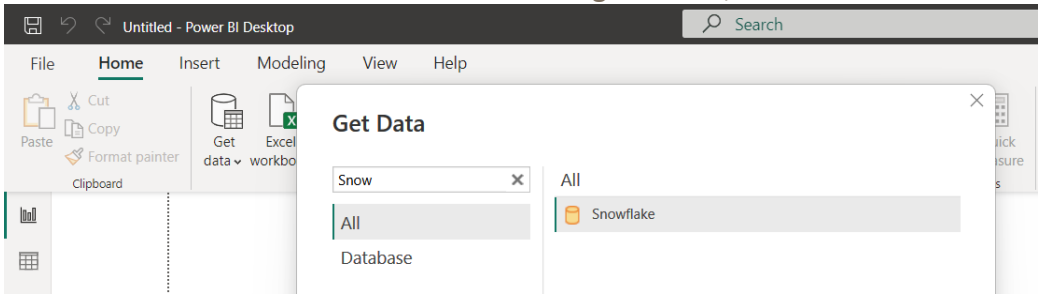
4. Clicking Next will result in installation of Microsoft Power BI which usually takes couple of minutes to complete the installation.



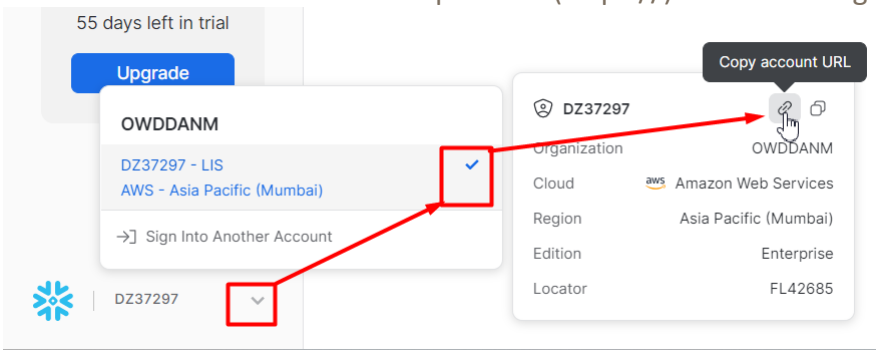
3.2 Connecting Power BI to Data Source

Once you have successfully installed the Power BI, we need to connect and import data from the source. These data will be used to build reports in Power BI.

1. Click on Home > Get Data > More.. in the navigation bar, then search and select Snowflake.

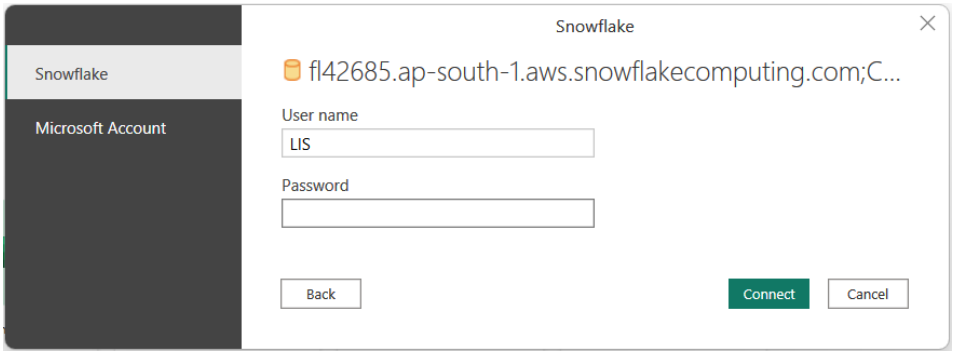


2. Enter your Snowflake server details and warehouse name in the prompt.
    a. Server: This is URL of the Snowflake trial account you created above and sent to your email. OR you can retrieve this from the console by following the screenshot below. Make sure to remove the protocol (https://) while entering in this field.

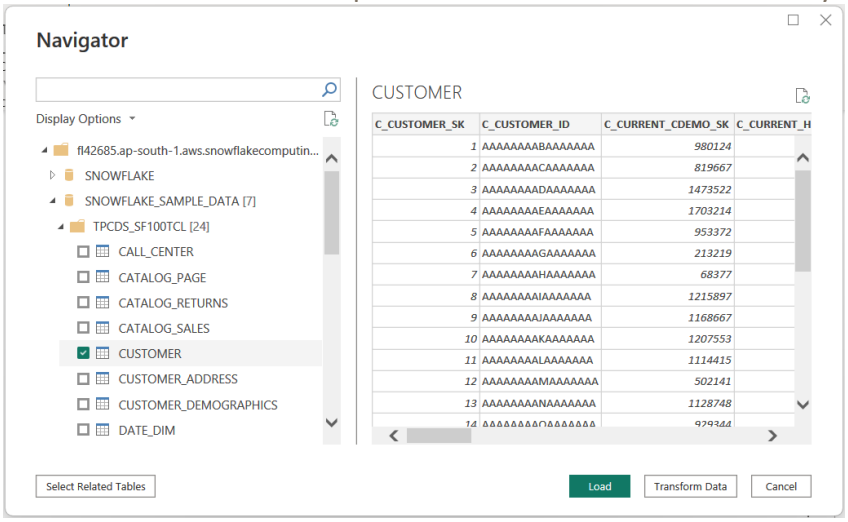


    b. Warehouse: This is the warehouse you used. Default is COMPUTE_WH

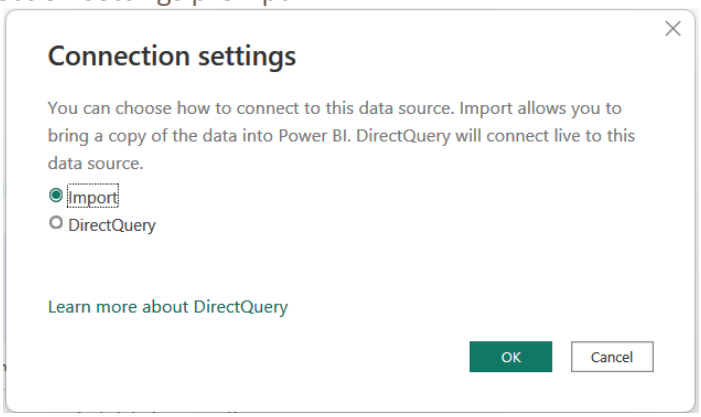
3. Click on 'OK' button of the prompt above and you'll be asked for the Snowflake credentials. Please enter your Snowflake username and password you created during setup. Click on 'Connect' button.



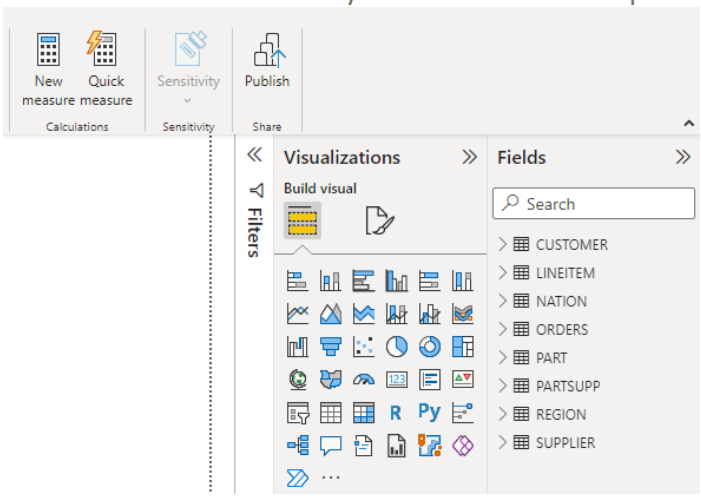
4. You will see the databases present in the warehouse. Select any dataset present.



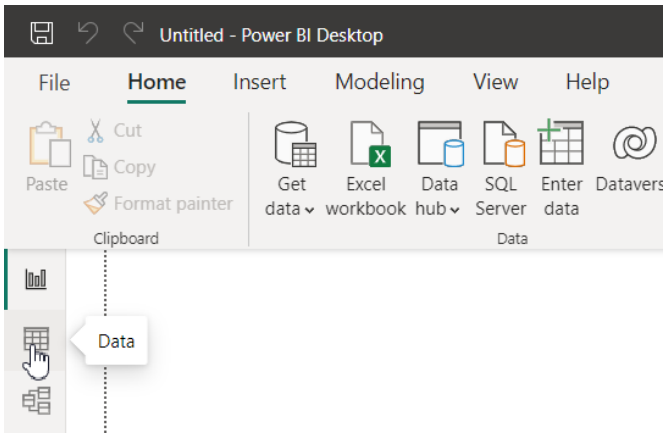
5. After selecting the tables, click Load button to load these data in Power BI. Select 'Import' option in the connection settings prompt.



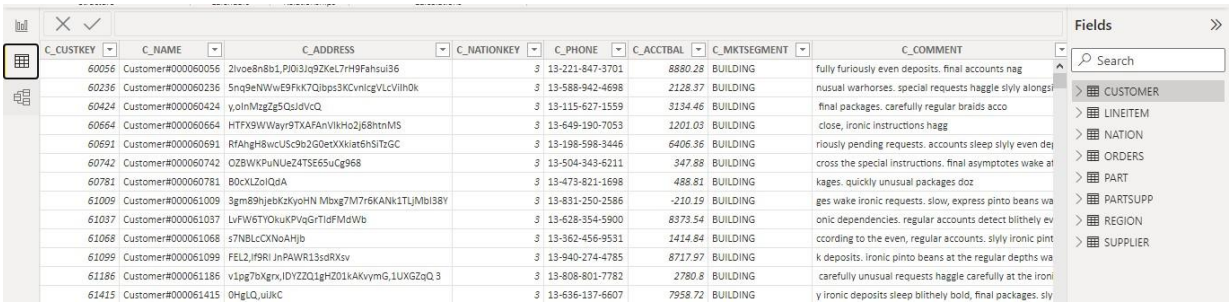
6. Data import might take few minutes to complete depending upon the data size and after completion you will be able to see the tables you selected in Fields pane of Power BI.



7. Go to the Data tab by selecting the second option as shown in the screenshot below.



8. With the correct execution of above steps, you should be able to see the data in Data tab of Power BI as shown below.



C_CUSTKEY	C_NAME	C_ADDRESS	C_NATIONKEY	C_PHONE	C_ACCTBAL	C_MKTSEGMENT	C_COMMENT
60056	Customer#000060056	2Ivoe8n8b1,PJ0i3Jq9ZKeL7rH9Fahsui36	3	13-221-847-3701	8880.28	BUILDING	fully furiously even deposits. final accounts nag
60236	Customer#000060236	5nq9eNWwE9FkK7Qjbps3KCvnlcgVLcVilh0k	3	13-588-942-4698	2128.37	BUILDING	nusual warhorses. special requests haggle slyly alongsi
60424	Customer#000060424	y,oInMzgZg5QsJdVcQ	3	13-115-627-1559	3134.46	BUILDING	final packages. carefully regular braids acco
60664	Customer#000060664	HTFX9WWayr9TXAFAnVIkHo2j68htnMS	3	13-649-190-7053	1201.03	BUILDING	close, ironic instructions hagg
60691	Customer#000060691	RfAhgH8wcUSc9b2G0etXXkiat6hSiTzGC	3	13-198-598-3446	6406.36	BUILDING	riously pending requests. accounts sleep slyly even dep
60742	Customer#000060742	OZBWKPuNUeZ4TSE65uCg968	3	13-504-343-6211	347.88	BUILDING	cross the special instructions. final asymptotes wake at
60781	Customer#000060781	B0cXLZoIQdA	3	13-473-821-1698	488.81	BUILDING	kages. quickly unusual packages doz
61009	Customer#000061009	3gm89hjebKzKyoHN Mbxg7M7r6KANk1TLjMbI38Y	3	13-831-250-2586	-210.19	BUILDING	ges wake ironic requests. slow, express pinto beans wa
61037	Customer#000061037	LvFW6TYOkuKPVqGrTIdFMdWb	3	13-628-354-5900	8373.54	BUILDING	onic dependencies. regular accounts detect blithely ev
61068	Customer#000061068	s7NBLcCXNoAHJb	3	13-362-456-9531	1414.84	BUILDING	ccording to the even, regular accounts. slyly ironic pint
61099	Customer#000061099	FEL2,Jf9RI JnPAWR13sdRXsv	3	13-940-274-4785	8717.97	BUILDING	k deposits. ironic pinto beans at the regular depths wa
61186	Customer#000061186	v1pg7bXgrx,IDYZZQ1gHZ01kAKvymG,1UXGZqQ,3	3	13-808-801-7782	2780.8	BUILDING	carefully unusual requests haggle carefully at the ironi
61415	Customer#000061415	0HgLQ,uiJkC	3	13-636-137-6607	7958.72	BUILDING	y ironic deposits sleep blithely bold, final packages. sly

Fields
Search
CUSTOMER
LINEITEM
NATION
ORDERS
PART
PARTSUPP
REGION
SUPPLIER

9. With this, we have completed the Power BI setup.