AWS + snowflake + tableau project.

Basics.

To create sql worksheet on snowflake.
1. Go to projects → worksplace → Add new
2. click choose database.
3. Select SNOWFLAKE_SAMPLE_DATA → TPCH-SF1
4. you can now execute queries.

snowflake architecture
① snowflake is a data warehouse that runs on cloud infrastructure.
② we can perform (eg.- data transformation, creating database schema, create tables, loading data into tables, creating virtual warehouses.
③ It separates the storage and compute layer. (helpful in cost optimization)

In snowflake we have two notions - database, virtual warehouse

when we have to perform query processing or some queries have to be executed, so it basically allocates the resources through these virtual warehouse only.
-These virtual warehouses are available in different sizes. X-Small - one node
Small - two nodes
Medium - 4
Large - 8
x-large - 16
6x-large - 512

The architecture has 3 layers.

Bottom layer - Database storage :-
- when the data is loaded into snowflake so it is basically compressed and stored in columnar format
- snowflake stores this data in cloud storage.

Compute layer - Query processing :-
- snowflake processes queries using "virtual warehouses"

Top layer - cloud service :-
- It helps in authentication and access control.
- It is a collection of services that coordinates activities across snowflake.

setting up a warehouse in snowflake
1. go to ~~manage~~ Manage → Compute → warehouse
2. First change the role.
   ↓
   1. click account icon
   2. switch role to public
   dont do now.
3. Make sure the role is account admin
4. click + warehouse
5. select size small
6. go to advance options
7. Make sure Auto-resume and Auto-suspend are ticked
8. tick Multi-cluster warehouse
9. go to projects sql worksheet, select your created warehouse

↑load file aswell.

1. create an S3 bucket with same region as of Snowflake
2. create an IAM role
   ↓ select Aws account
   2. tick option → requires external ID
   3. keep it 00000
3. click next
4. give S3 Full Access permision

Creating an integration object which will be required to establish connection between Aws and snowflake.

for the piece of code.

1. get ARN from Aws role
2. get S3 bucket name and put it in code
3. paste the code in sql query editor selecting your
4. now update trust policy in aws
5. Run the desc command
   ↓ you will be able to see ARN ID and External ID
6. Now go to Aws
7. go to your role → Trust relationships
8. Click edit trust policy.
9. In place of ARN, & Paste the IAM-ARN from snowflake
10. Replace the External ID aswell.

Loading the data from Aws to snowflake

1. run the sql queries to create database, schema and table.
2. now we have to create a stage - a stage is basically an object in snowflake that will store the location from Aws in which we have files (data) available

3. mention the s3 name and storage integration name

Now Connecting snowflake to tableau

1. We will need tableau desktop as tableau public will not work.
2. go to tableau desktop, click more.
3. download the snowflake driver
4. After downloading, go to snowflake and Copy Aws account no. server url,
5. Type snowflake username and password
6. Select warehouse.
7. Database and schema, we created.
8. table drag and drop.
9. click update automatically.

Now starting the project.

1. create an s3 bucket
2. now add files
3. now create role that will be used to establish the connection between Aws, s3 and snowflake
4. for role, select Aws account
5. tick require external id.
6. Put any random no. for now.
7. To create integration object, go to snowflake and paste. Code Provided
8. Copy arn and external Id. → edit in code
9. now go to aws Iam trust relationships
10. edit trust policies
11. copy user ARN and External_ID from the output of query (desc integration)

SQL queries

12. Now we will be creating a stage, database schema
13. stage holds the information of wherever data available on S3.
14. Run all queries.

r8x

Data profiling : data understanding

1. you can execute queries, create charts in snowsight to understand data.
2. we will create a replica table so original one doesn't get affected.

data transformation

Create a dashboard with 6 charts
first chart - month_usage_kwh by region
Second chart : monthly_usage_kwh by country
third chart - monthly_usage_kwh by energy source
fourth chart - cost_saving by region
fifth chart - cost_saving by country.
sixth chart - cost_saving by energy source.
and then dashboard.

1. change sheet name.
2. drag and drop monthly usage kwh in rows and region in columns
3. change to standard view to chart view
4. Right click on chart and format
5. select grid lines as none.
6. to not see values in vertical right click on values and untick show header.

7. press ctrl + drag Monthly usage to label for second chart.

1. swap rows and column