Parquet dataset load into Snowflake

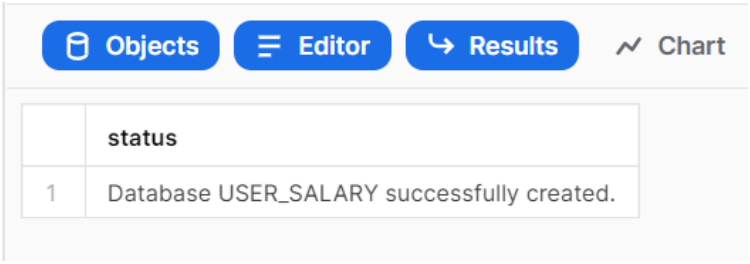
Neelesh Palaparthi

Login into Snowflake account:

#Login into snowflake account and create a new warehouse and perform the below steps.

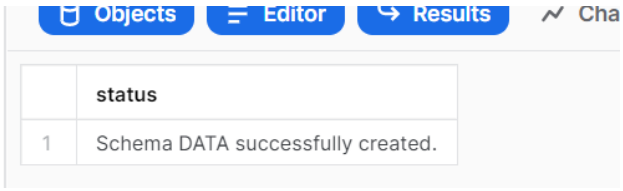
Creation of Database:

#Create a new database with a appropriate name.
create or replace database user_salary;
use database user_salary;



Creation of Schema:

#Create a schema so that the tables are store into it.
create or replace schema data
use schema data



Grant Permissions for Roles:

#before creating integration layer grant the permissions for different roles.
grant usage on database user_salary to role sysadmin;

Integration Layer:

#integrating AWS with snowflake. For that you need to create S3 bucket and IAM role. Refer this link to create S3 bucket and IAM role. Then create a new integration with the following commands.

```
CREATE or replace STORAGE INTEGRATION usersal_integ
TYPE = EXTERNAL_STAGE
STORAGE_PROVIDER = S3
ENABLED = TRUE
STORAGE_AWS_ROLE_ARN ='arn:aws:iam::627540493785:role/snowflake_connectivity'
STORAGE_ALLOWED_LOCATIONS=('s3://testbucketmt/parquet_data/userdata1.parquet
');
```

Desribe the integration details:

```
#To show the integration details.
desc integration usersal_integ
```

Objects | Editor | Results | Chart

	property	property_type	property_value	property_defa
1	ENABLED	Boolean	true	false
2	STORAGE_PROVIDER	String	S3	
3	STORAGE_ALLOWED_LOCATIONS	List	s3://testb[redacted]serdata1.parquet	[]
4	STORAGE_BLOCKED_LOCATIONS	List		[]
5	STORAGE_AWS_IAM_USER_ARN	String	arn:a[redacted]s	
6	STORAGE_AWS_ROLE_ARN	String	arn:aws:[redacted]_connectivity	
7	STORAGE_AWS_EXTERNAL_ID	String	QM[redacted]1g=	
8	COMMENT	String		

File format:

#here we have to give the details of file format like csv, excel, parquet.

```
--file format object
CREATE OR REPLACE FILE FORMAT sfparquet_format
  TYPE = parquet;
```

Table creation:

you have to create table and write all the columns so that when loading data from aws the columns has to match with staging otherwise you will get errors.

```
create or replace table usersalary(
registration_dttm varchar,
id int,
first_name varchar,
last_name varchar,
email varchar,
gender varchar,
```

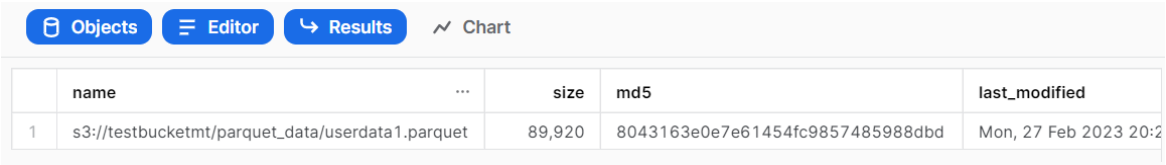
```
ip_address varchar,
cc varchar,
country varchar,
birthdate varchar,
salary float,
title varchar,
comments varchar
);
```

Stage creation:

```
#Creation of stage using integration. Type these commands and run them.
create or replace stage sfparquet_stage
url = 's3://testbucketmt/parquet_data/userdata1.parquet'
storage_integration = usersal_integ
file_format = sfparquet_format;
```

Showing the stage creation:

```
list @awsf_stage;
```

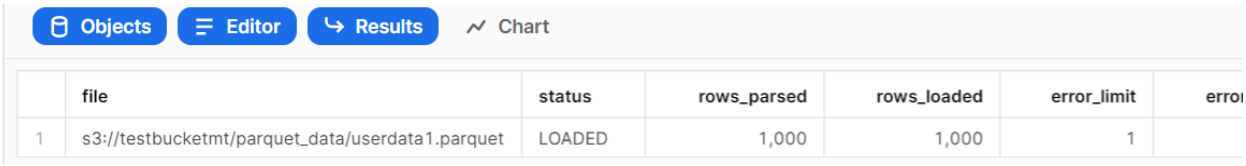


	name	size	md5	last_modified
1	s3://testbucketmt/parquet_data/userdata1.parquet	89,920	8043163e0e7e61454fc9857485988dbd	Mon, 27 Feb 2023 20:2

Copying table from stage:

#Copying the data into target table which we have created In the above steps.

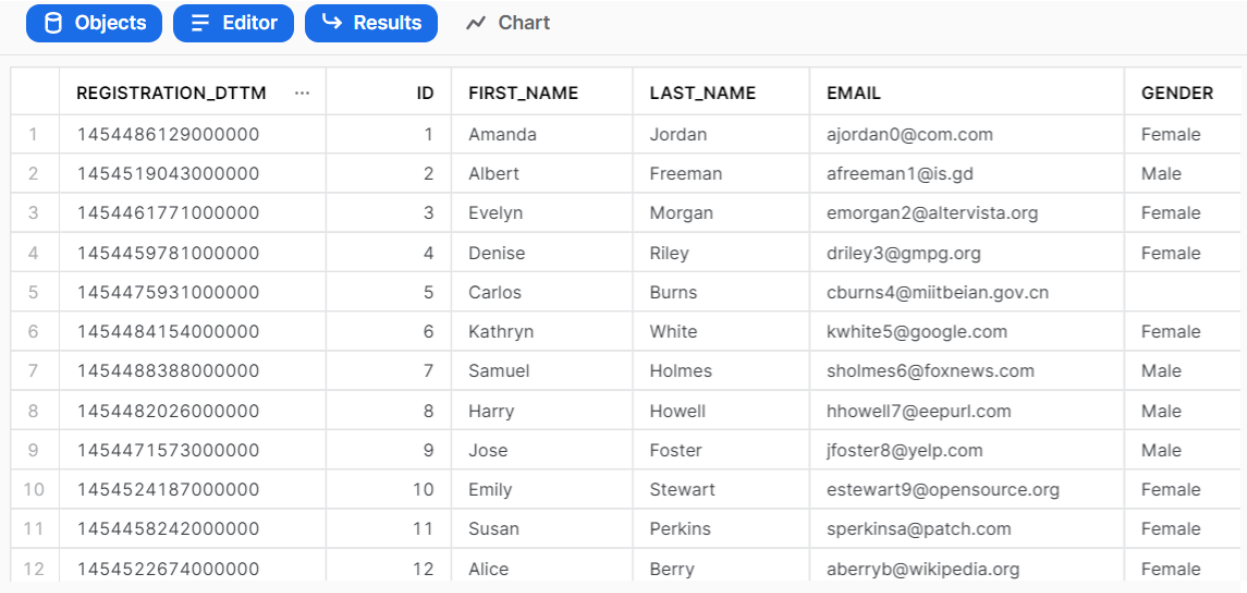
```
COPY INTO usersalary FROM @sfparquet_stage
ON_ERROR = 'skip_file'
match_by_column_name = case_insensitive
```



	file	status	rows_parsed	rows_loaded	error_limit	erro
1	s3://testbucketmt/parquet_data/userdata1.parquet	LOADED	1,000	1,000	1	

Display the table:

SELECT * FROM usersalary;



	REGISTRATION_DTTM	ID	FIRST_NAME	LAST_NAME	EMAIL	GENDER
1	1454486129000000	1	Amanda	Jordan	ajordan0@com.com	Female
2	1454519043000000	2	Albert	Freeman	afreeman1@is.gd	Male
3	1454461771000000	3	Evelyn	Morgan	emorgan2@altervista.org	Female
4	1454459781000000	4	Denise	Riley	driley3@gmpg.org	Female
5	1454475931000000	5	Carlos	Burns	cburns4@miitbeian.gov.cn	
6	1454484154000000	6	Kathryn	White	kwhite5@google.com	Female
7	1454488388000000	7	Samuel	Holmes	sholmes6@foxnews.com	Male
8	1454482026000000	8	Harry	Howell	hhowell7@eepurl.com	Male
9	1454471573000000	9	Jose	Foster	jfoster8@yelp.com	Male
10	1454524187000000	10	Emily	Stewart	estewart9@opensource.org	Female
11	1454458242000000	11	Susan	Perkins	sperkinsa@patch.com	Female
12	1454522674000000	12	Alice	Berry	aberryb@wikipedia.org	Female