

Snowflake: Data Load/download

Schema creation for Base table: Trips

snowsql -a kd11570.ap-southeast-2 -u SRDAS -f

/home/soumya/Documents/gitfolder/CityBikeTravel_ETL/biketrip_ddl.sql

```
(base) soumya@soumya-Inspiron-5559:~$ snowsql -a kd11570.ap-southeast-2 -u SRDAS -f /home/soumya/Documents/gitfolder/CityBikeTravel_ETL/biketrip_ddl.sql
Failed to initialize log. No logging is enabled: [Errno 13] Permission denied: '/home/snowsql_rt.log_bootstrap'
Failed to initialize log. No logging is enabled: [Errno 13] Permission denied: '/home/snowsql_rt.log'
Password:
* SnowSQL * v1.2.10
Type SQL statements or !help
+-----+
| status |
+-----+
| Statement executed successfully. |
+-----+
1 Row(s) produced. Time Elapsed: 0.410s
+-----+
| status |
+-----+
| Statement executed successfully. |
+-----+
1 Row(s) produced. Time Elapsed: 0.406s
+-----+
| status |
+-----+
| Statement executed successfully. |
+-----+
1 Row(s) produced. Time Elapsed: 0.411s
+-----+
| status |
+-----+
| Table TRIPS successfully created. |
+-----+
1 Row(s) produced. Time Elapsed: 0.495s
Goodbye!
```

Stage table Creation and Bulk Data load:

snowsql -a kd11570.ap-southeast-2 -u SRDAS -f

/home/soumya/Documents/gitfolder/CityBikeTravel_ETL/staging.sql

```
+-----+
| status |
+-----+
| Table STAGING successfully created. |
+-----+
1 Row(s) produced. Time Elapsed: 0.494s
2016.csv_c.gz(4.94MB): [#####] 100.00% Done (11.807s, 0.42MB/s).
+-----+
| source | target | source_size | target_size | source_compression | target_compression | status | message |
+-----+
| 2016.csv | 2016.csv.gz | 40281861 | 5184079 | NONE | GZIP | UPLOADED | |
+-----+
1 Row(s) produced. Time Elapsed: 22.509s
+-----+
| name | size | md5 | last_modified |
+-----+
| 2016.csv.gz | 5184080 | 7ccdc3b3f65f9a0843dfbe833bd419f2 | Sat, 12 Dec 2020 19:42:01 GMT |
+-----+
1 Row(s) produced. Time Elapsed: 0.418s
+-----+
| file | status | rows_parsed | rows_loaded | error_limit | errors_seen | first_error | first_error_line |
+-----+
| 2016.csv.gz | LOADED | 408160 | 408160 | 1 | 0 | NULL | NULL |
+-----+
1 Row(s) produced. Time Elapsed: 3.261s
Goodbye!
```

Row	DEPARTURE_TIME	RETURN_TIME	DEPARTURE_ST/	DEPARTURE_ST/	RETURN_STATIO	RETURN_STATIO	COVERED_DISTA	DURATION	YEAR_MONTH	YEAR_VALUE
1	2016-09-30T23:58:00	2016-10-01T00:02:00	A06	Vliskulma	A03	Kapteeninpui...	561	236	2016-09	2016
2	2016-09-30T23:56:00	2016-09-30T23:59:00	A18	Ritarikatu	A12	Unioninkatu	641	165	2016-09	2016
3	2016-09-30T23:54:00	2016-10-01T00:15:00	A03	Kapteeninpui...	B07	Brahen kenttä	4392	1236	2016-09	2016
4	2016-09-30T23:54:00	2016-09-30T23:59:00	A31	Eläinmuseo	A23	Rautatientori ...	673	273	2016-09	2016
5	2016-09-30T23:53:00	2016-10-01	A12	Unioninkatu	A03	Kapteeninpui...	1257	397	2016-09	2016

Data view snowflake: staging table

Data Integration: Moving data from staging to => final table (incremental load every year)

snowsql -a kd11570.ap-southeast-2 -u SRDAS -f

/home/soumya/Documents/gitfolder/CityBikeTravel_ETL/data_integration.sql

```
(base) soumya@soumya-Inspiron-5559:~$ snowsql -a kd11570.ap-southeast-2 -u SRDAS -f /home/soumya/Documents/gitf
older/CityBikeTravel_ETL/data_integration.sql
Failed to initialize log. No logging is enabled: [Errno 13] Permission denied: '/home/snowsql_rt.log_bootstrap'
Failed to initialize log. No logging is enabled: [Errno 13] Permission denied: '/home/snowsql_rt.log'
Password:
* SnowSQL * v1.2.10
Type SQL statements or !help
+-----+
| status |
+-----+
| Statement executed successfully. |
+-----+
1 Row(s) produced. Time Elapsed: 0.409s
+-----+
| status |
+-----+
| Statement executed successfully. |
+-----+
1 Row(s) produced. Time Elapsed: 0.431s
+-----+
| status |
+-----+
| Statement executed successfully. |
+-----+
1 Row(s) produced. Time Elapsed: 0.383s
+-----+
| number of rows inserted |
+-----+
| 408160 |
+-----+
408160 Row(s) produced. Time Elapsed: 2.985s
Goodbye!
```

Bulk loading => Staging for next year(2017)

```
+-----+
| status |
+-----+
| Table STAGING successfully created. |
+-----+
1 Row(s) produced. Time Elapsed: 0.556s
2017.csv_c.gz(30.31MB): [#####] 100.00% Done (29.928s, 1.01MB/s).
+-----+
| source | target | source_size | target_size | source_compression | target_compression | status | mes
sage |
+-----+
| 2017.csv | 2017.csv.gz | 231216074 | 31783615 | NONE | GZIP | UPLOADED |
+-----+
1 Row(s) produced. Time Elapsed: 72.284s
+-----+
| name | size | md5 | last_modified |
+-----+
| 2017.csv.gz | 31783616 | 11b090783efc7382846fcefcabeba77a | Sat, 12 Dec 2020 22:21:59 GMT |
+-----+
1 Row(s) produced. Time Elapsed: 0.454s
+-----+
| file | status | rows_parsed | rows_loaded | error_limit | errors_seen | first_error | first_error_line |
| first_error_character | first_error_column_name |
+-----+
| 2017.csv.gz | LOADED | 2201552 | 2201552 | 1 | 0 | NULL | NULL |
| NULL | NULL |
+-----+
1 Row(s) produced. Time Elapsed: 9.248s
Goodbye!
```

Integrating 2017 data with 2016 in TRIPS table:

```
(base) soumya@soumya-Inspiron-5559:~$ snowsql -a kd11570.ap-southeast-2 -u SRDAS -f /home/soumya/Documents/gitf
older/CityBikeTravel_ETL/data_integration.sql
Failed to initialize log. No logging is enabled: [Errno 13] Permission denied: '/home/snowsql_rt.log_bootstrap'
Failed to initialize log. No logging is enabled: [Errno 13] Permission denied: '/home/snowsql_rt.log'
Password:
* SnowSQL * v1.2.10
Type SQL statements or !help
+-----+
| status |
+-----+
| Statement executed successfully. |
+-----+
1 Row(s) produced. Time Elapsed: 0.412s
+-----+
| status |
+-----+
| Statement executed successfully. |
+-----+
1 Row(s) produced. Time Elapsed: 0.407s
+-----+
| status |
+-----+
| Statement executed successfully. |
+-----+
1 Row(s) produced. Time Elapsed: 0.377s
+-----+
| number of rows inserted |
+-----+
| 2201552 |
+-----+
2201552 Row(s) produced. Time Elapsed: 3.820s
Goodbye!
```

```
3 select DISTINCT(YEAR_VALUE) from "CITYBIKES"."PUBLIC"."TRIPS";
```

Results Data Preview

✓ Query ID SQL 352ms 2 rows

Filter result...



Copy

Row	YEAR_VALUE
1	2016
2	2017

```
3 select MAX(YEAR_VALUE) from "CITYBIKES"."PUBLIC"."TRIPS";
```

Results Data Preview

✓ Query ID SQL 510ms 1 rows

Filter result...



Copy

Row	MAX(YEAR_VALUE)
1	2017

Download LATEST year from the database as .csv file:

```
(base) soumya@soumya-Inspiron-5559:~$ snowsql -a kd11570.ap-southeast-2 -u SRDAS -f /home/soumya/Documents/gitf
older/CityBikeTravel_ETL/yearvalue_download.sql
Failed to initialize log. No logging is enabled: [Errno 13] Permission denied: '/home/snowsql_rt.log_bootstrap'
Failed to initialize log. No logging is enabled: [Errno 13] Permission denied: '/home/snowsql_rt.log'
Password:
* SnowSQL * v1.2.10
Type SQL statements or !help
+-----+
| status |
+-----+
| Statement executed successfully. |
+-----+
1 Row(s) produced. Time Elapsed: 0.415s
+-----+
| status |
+-----+
| Statement executed successfully. |
+-----+
1 Row(s) produced. Time Elapsed: 0.453s
+-----+
| status |
+-----+
| Statement executed successfully. |
+-----+
1 Row(s) produced. Time Elapsed: 0.681s
+-----+
| rows_unloaded | input_bytes | output_bytes |
+-----+
| 1 | 5 | 25 |
+-----+
1 Row(s) produced. Time Elapsed: 0.754s
data_0_0_0.csv.gz(0.00MB): [#####] 100.00% Done (0.818s, 0.00MB/s).
+-----+
| file | size | status | message |
+-----+
| data_0_0_0.csv.gz | 25 | DOWNLOADED | |
+-----+
1 Row(s) produced. Time Elapsed: 6.731s
Goodbye!
```

This value is used by the scrapper for downloading the latest zip file from the web.