

Getting Started with Amazon Redshift and Tableau

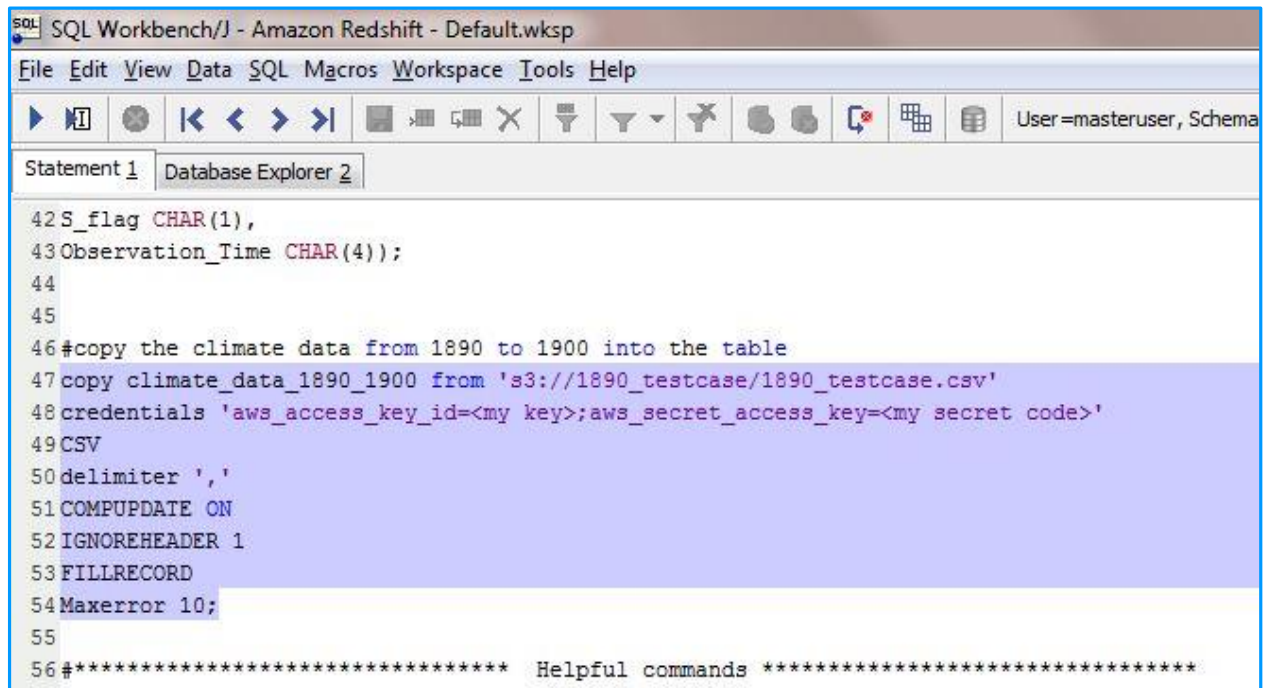
Ken Black 10/30/13

1. Set up an amazon AWS account

- Retrieve your access code and your secret code. Keep this information private!

2. Setup the SQL workbench to interact with Amazon Servers via SQL Commands

- This took me longer than learning some of the Amazon nomenclature, principally because of the steps needed to install the SQL workbench program were not well documented. If you are an SQL guru, this most likely won't be a problem for you. However, for me, it took a while to figure out that the postgresql-9.2-1003.jdbc4.jar file had to be separately downloaded from the installation package. Once I got that downloaded and the workbench configured for it, the workbench worked fine.
- Figure 1 shows the workbench program with one SQL command highlighted for processing and Figure 2 shows the menu choice needed to run the command.



```
SQL Workbench/J - Amazon Redshift - Default.wksp
File Edit View Data SQL Macros Workspace Tools Help
User=masteruser, Schema
Statement 1 Database Explorer 2
42 S_flag CHAR(1),
43 Observation_Time CHAR(4));
44
45
46 #copy the climate data from 1890 to 1900 into the table
47 copy climate_data_1890_1900 from 's3://1890_testcase/1890_testcase.csv'
48 credentials 'aws_access_key_id=<my key>;aws_secret_access_key=<my secret code>'
49 CSV
50 delimiter ','
51 COMPUPDATE ON
52 IGNOREHEADER 1
53 FILLRECORD
54 Maxerror 10;
55
56 #***** Helpful commands *****
```

Figure 1- The SQL workbench showing an SQL command for copying data into an SQL table. My Amazon codes have been removed from the example.

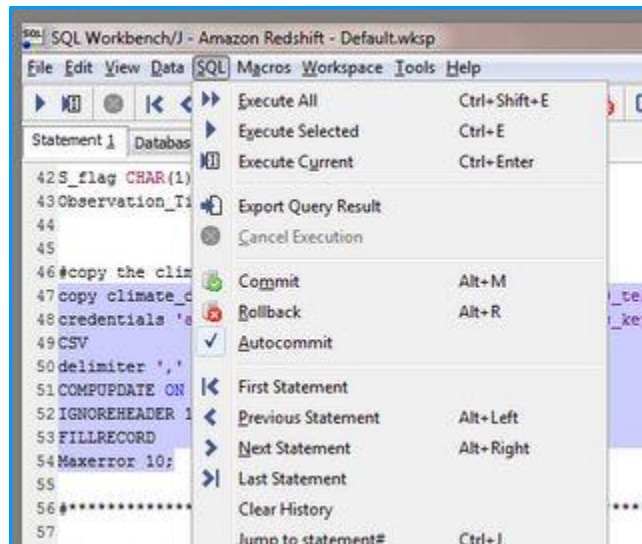


Figure 2- Shows how to find the "Execute Selected" menu choice for running the SQL command that is selected.

3. Learn to use the S3 service for uploading your data

- Simple to use - create a bucket, upload files into the bucket.

4. Learn to use the Redshift service for accessing your data via Tableau. Getting the data ready for Tableau takes four steps.

- 1) Create the data.
- 2) Upload the data to Amazon servers using the S3 service.
- 3) Create SQL data tables to hold your data. This is completed using the SQL workbench as shown in the examples below.
- 4) Populate your tables with a copy command and the data stored in S3, as shown in the two examples shown below.

5. Connect Tableau to the Redshift databases you have created. This connection menu is shown in Figure 3.

- The biggest issues here were realizing that the endpoint link (not the jdbc link) provided by Amazon is used in the "Step 1: Enter a server name:" field for Tableau. Secondly, you need to use the correct port number (5439) that Amazon provides, not the 8912 standard port given by Tableau. This information is found on your cluster page, once that is created. Also, you will find your data by selecting the public schema in "Step 5: Select a schema on the server:" of Figure 3.

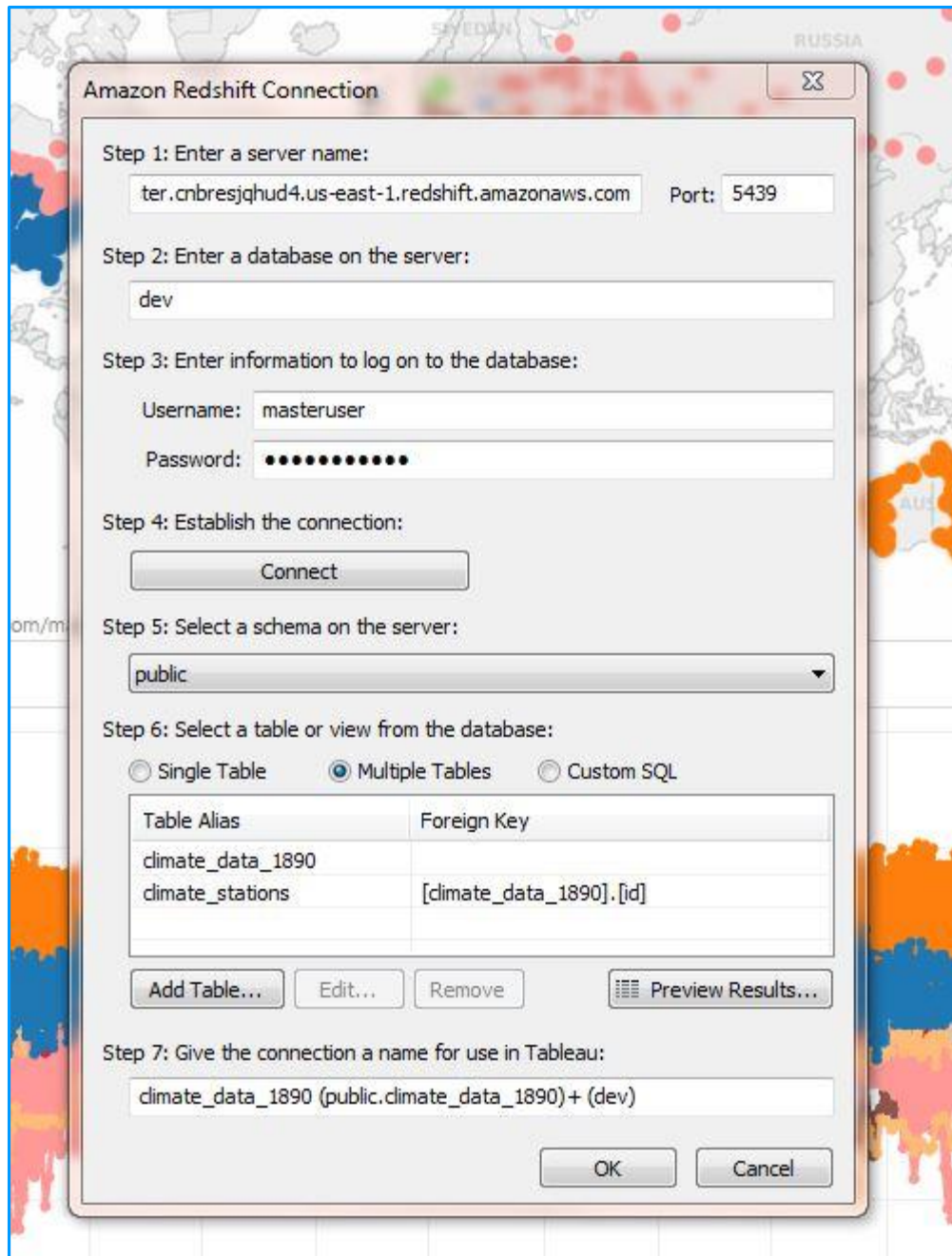


Figure 3- Tableau - Redshift connection menu

Example 1: Feeding the 1890 Climate data into an SQL Table

***** Step 1: Here is the 1890 Climate data - 1.1 million rows - csv format *****

```
ID,Date,Measure Type,Data Value,M-FLAG,Q-FLAG,S-FLAG,Observation Time
ASN00070200,01/01/1890,PRCP,0,,,a,
SF000782720,01/01/1890,PRCP,0,,,l,
CA005022790,01/01/1890,TMAX,-222,,,G,
```

```
CA005022790,01/01/1890,TMIN,-261,,,G,
CA005022790,01/01/1890,PRCP,0,,,G,
...
```

***** Step 2: Creating a table for to hold 1890 Climate data *****
#SQL command and options to create the table to hold this climate data. Command is executed in SQL workbench.

```
create table climate_data_1890(
  Id VARCHAR(11),
  Measurement_Date CHAR(10),
  Measurement_Type CHAR(4),
  Measurement_Value REAL,
  M_flag CHAR(1),
  Q_flag CHAR(1),
  S_flag CHAR(1),
  Observation_Time CHAR(4));
```

***** Step 3: Copy the 1890 Climate data into the table *****
#SQL command to copy the climate data from 1890 into the table created in step 2. Command is executed in SQL workbench.

```
copy climate_data_1890 from 's3://1890_testcase/1890_only_testcase.csv'
credentials 'aws_access_key_id=<your access key>;aws_secret_access_key=<your secret key>'
CSV
delimiter ','
COMPUPDATE ON
IGNOREHEADER 1
FILLRECORD
Maxerror 10;
```

Notes

1. Your <your access key> and <your secret key> are long alphanumeric key strings provided to you by Amazon. They are downloaded and you must keep them private. You don't include the brackets < and > in this command.
2. It is imperative to use the CSV option to be able to properly read the csv files stored on the S3 system. Otherwise, you will get errors and it might take you a couple of hours to figure out what to do about them!

Example 2: Feeding the Climate monitoring station data into an SQL Table

***** Step 1: Here is the Climate Monitoring Stations - 90,000 rows - csv format *****

```
Id,Latitude,Longitude,Elevation (M),State,Name,Gsnflag,Hcnflag,Wmoid,Country Code,Country Name
ACW00011604,17.1167,-61.7833,10.1,,ST JOHNS COOLIDGE FLD,,,,AC,Antigua and Barbuda
ACW00011647,17.1333,-61.7833,19.2,,ST JOHNS,,,,AC,Antigua and Barbuda
AE000041196,25.333,55.517,34,,SHARJAH INTER. AIRP,GSN,,41196,AE,United Arab Emirates
AF000040930,35.317,69.017,3366,,NORTH-SALANG,GSN,,40930,AF,Afghanistan
AG000060390,36.7167,3.25,24,,ALGER-DAR EL BEIDA,GSN,,60390,AG,Algeria
AG000060590,30.5667,2.8667,397,,EL-GOLEA,GSN,,60590,AG,Algeria
...
```

***** Step 2: Creating a table to hold Climate monitoring station data *****

#SQL commands to create the table to hold this climate monitoring station data.

```
create table climate_stations(  
    Id CHAR(11),  
    Latitude DOUBLE PRECISION,  
    Longitude DOUBLE PRECISION,  
    Elevation DOUBLE PRECISION,  
    State CHAR(2),  
    Name VARCHAR(40),  
    Gsnflag CHAR(3),  
    Hcnflag CHAR(3),  
    Wmoid BIGINT,  
    Country_Code CHAR(2),  
    Country_Name VARCHAR(60));
```

Notes

1. Char variables are fixed length, whereas varchar variables have variable length.

***** Step 3: Copy the Climate monitoring station data into the table *****

#SQL commands to copy the climate data from 1890 into the table created in step 2.

```
copy climate_stations from 's3://1890_testcase/Climate_Stations.csv'  
credentials 'aws_access_key_id=<your access key>;aws_secret_access_key=<your secret key>'  
CSV  
delimiter ','  
COMPUPDATE ON  
IGNOREHEADER 1;
```