

For the full-launch of Data Analytics, we’ve set up an Amazon relational database server to use as a teaching tool throughout the course, as part of our initiative to make the course more akin to what an analyst would do, day to day.

Please read below for more context and instructions on how to use the service:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**What is AWS?**

Taken from aws.amazon.com:

“Amazon Relational Database Service (Amazon RDS) makes it easy to set up, operate, and scale a relational database in [the cloud](http://aws.amazon.com/what-is-cloud-computing/). It provides cost-efficient and resizable capacity while managing time-consuming database management tasks, freeing you up to focus on your applications and business.”

Much like a hosted website, AWS allows you to host your database using their service. Read more [here](http://aws.amazon.com/rds/).

**Why did we set this up?**

In our pilot version of the course, the SQL unit focused on downloading and installing PostgreSQL locally, creating tables and databases, and copying data to your local database before querying it. While this approach gave students a really unique understanding of databases and PostgreSQL, it became clear that most analysts are not asked to do this.

In talking with both internal and external stakeholders, we found that hosting our data on this service and providing students with the proper authentication credentials would closely mimic an analyst’s day to day practices.

**What will it be used for?**

Since unit one is focused on analytics in Excel, for the in-class exercises in units two and three - as well as the unit projects for both units - this AWS server will contain all the data you’ll need for instructional teams and students. Find the [details below](#id.6kp59v7vebr0) of the datasets that correspond with each unit and project.

**Who will be using it? How will they use it?**

With a third-party client - we recommend [PGAdmin](http://www.pgadmin.org/) - you can connect to and read from each database - no one will have write access except the admin from HQ. For instructions on how to connect with PGAdmin, click [here.](http://suite.opengeo.org/opengeo-docs/dataadmin/pgGettingStarted/pgadmin.html) Unfortunately, this article doesn’t list steps for how to get to the query terminal: simply click “Tools” >> “Query Tool”, once you’ve connected.

Students and instructors should use the following credentials to connect (please use the correct host for your region):

* **host**:
  + West-coast (includes Austin and Chicago)
    - [analyticsga.cuwj8wuu6wbh.us-west-2.rds.amazonaws.com](http://analyticsga.cuwj8wuu6wbh.us-west-2.rds.amazonaws.com)
  + East-coast
    - [analyticsga-east2.c20gkj5cvu3l.us-east-1.rds.amazonaws.com](http://analyticsga-east2.c20gkj5cvu3l.us-east-1.rds.amazonaws.com/)
  + APAC and Europe
    - [analyticsga-apac2.csuojbfcexhv.ap-southeast-2.rds.amazonaws.com](http://analyticsga-apac2.csuojbfcexhv.ap-southeast-2.rds.amazonaws.com/)
* **port**: 5432
* **username**: analytics\_student
* **password**: analyticsga

Any issues connecting to, reading from, or questions/suggestions about the data, should be documented on [the help desk](http://ga.co/helpdesk).

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Dataset Details**

* Unit 2
  + Iowa Liquor Sales
    - This dataset contains the spirits purchase information of Iowa Class "E" liquor licensees by product and date of purchase from January 1, 2014 to current. The dataset can be used to analyze total spirits sales in Iowa of individual products at the store level.
    - More documentation [here](http://flowingdata.com/2015/06/05/iowa-liquor-sales-data-3m-rows/)
* Unit 2 Project
  + Firefox User Study
    - This is a dataset on firefox browser user activity including surveys, users, and browser events.
    - More documentation [here](http://datahub.io/dataset/a-week-in-the-life-of-a-browser-version-2/resource/ec631c92-c975-4549-b646-74bb2ec2d778?inner_span=True)
* Unit 3
  + Airline Flight Delays
    - This dataset provides information about airlines and flight delays. Interesting, large dataset that has a lot of [great examples](https://www.google.com/webhp?sourceid=chrome-instant&ion=1&espv=2&ie=UTF-8#q=tableau%20airline%20delays%20data) of being explored in Tableau
* Final Project
  + NYC Taxi Trips
    - This dataset has 12 different csv files for information collect around trips and fares
    - More documentation, stories, and visualizations [here](http://www.andresmh.com/nyctaxitrips/)
  + Reddit Survey Results
    - This was a survey of over 30,000 Reddit users; could be used to create user segments, demographics
    - You must [download](http://static.reddit.com/RedditShortDemoSurvey-1.csv) the dataset or query it to get the schema