**The Easy Button for Your AWS Environment**

Some of you have had difficulty getting the AWS environments set up correctly.  As we move into the second half of the course, you need to be at parity with you classmates.  To make that simpler, these instructions will get your environment set up correctly without much work on your side.

If you have been struggling with the AWS setups, follow these instructions **exactly**.

**Note: THIS WILL DELETE ALL DATA ON YOUR EBS VOLUME.**

**Preliminaries**

1.   If you have not already, create an EBS volume.  Note the region in which the volume was created.

2.   Start an instance of ucbw205_complete_plus_postgres (AMI ID: ami-bff58fda) in the **same region** as your EBS volume.

3.   Attach the volume to the instance.

4.   Login to the instance via ssh.  At login, you will be the root user.

**Setup and install**

1.   Determine which device is your EBS drive by running fdisk –l

2.   The last entry is typically your EBS volume.  For example /dev/xvdf

3.   Copy the device path (/dev/xvdf) to your clipboard

4.   Download the setup script like this:

wget https://s3.amazonaws.com/ucbdatasciencew205/setup_ucb_complete_plus_postgres.sh

5.   Run the script like this:

bash setup_ucb_complete_plus_postres.sh <paste your device path>

6.   Follow the onscreen instructions

**After the script runs**

1.   Hadoop will be installed and started

2.   Postgres will be installed and started

3. Hive and SparkSQL will use Postgres as a metastore

4. A w205 user will exist.  This is the user you should work as

5. A script called setup_zeppelin.sh has been created

a. If you want to setup zeppelin, type: ./setup_zeppelin.sh
b. You can start zeppelin by typing /data/zeppelin/bin/zeppelin.sh

You only need to go through this process **ONCE**.  After the install, you should interact with your instance like this:

**When you shut the instance down (every time):**

1. Make sure you're root, run whoami

a. If you're not root, type exit and check again

2. Run: /data/stop_postgres.sh

3. Run: /root/stop_hadoop.sh

4. Run: umount /data

5. Terminate the instance

**When you start the instance again (every time):**
1. Find your EBS volume by running: fdisk –l
2. Mount your data volume by running: mount –t ext4 /dev/<your EBS device> /data
3. Run: /root/start_hadoop.sh
4. Run: /data/start_postgres.sh
5. Log into your named user, run: su – w205
6. If you are using Hive, PySpark, Postgres, or SparkSQL you can access those programs using the appropriate commands.
7. If you are going to use zeppelin, run: /data/zeppelin/bin/zeppelin.sh
a. Open <the URL of your EC2 instance>:8080 in your web browser
b. You will **NOT** be able to access zeppelin unless you have configured your instance security group to have port 8080 open.