Setting up an Amazon Web Services Account

Diana Maclean for CS448G 2011

First of all, what do we need?

For this assignment, we'll be running MapReduce workflows on Amazon's Elastic MapReduce framework. You'll need to do a couple of things to get set up:

1. Create an AWS account
    a. Sign up for S3 (this is where you'll store your data on the cloud) & create your main and logging buckets.
        i. Sign up for S3Fox (optional, but recommended)
    b. Sign up for EC2 (the machines we'll use to run our workflows on)
    c. Sign up for EMR (the packaged framework we'll use to run our MapReduce jobs; EMR will run your workflows on EC2 for you.)
2. Create a key pair and access your account credentials
3. Redeem your (free) credit! (And check how much you have left.)
4. Familiarize yourself with S3, EMR and EC2

1: Create an AWS account

Go to http://aws.amazon.com and sign up for an account, if you do not have one already. For now, please enter the required details, including payment details if required. Amazon has generously agreed to provide each student with credit for this class; more on how to redeem this later. Once you've created an account, you will need to sign up for the services that we will be using for the Assignment.

1a: Sign up for S3

Amazon Simple Storage Service (S3) is just a storage framework – you can upload any data here and pay (very small amounts, unless you have very large data) for storage space. You can read more about S3 here:

http://aws.amazon.com/s3

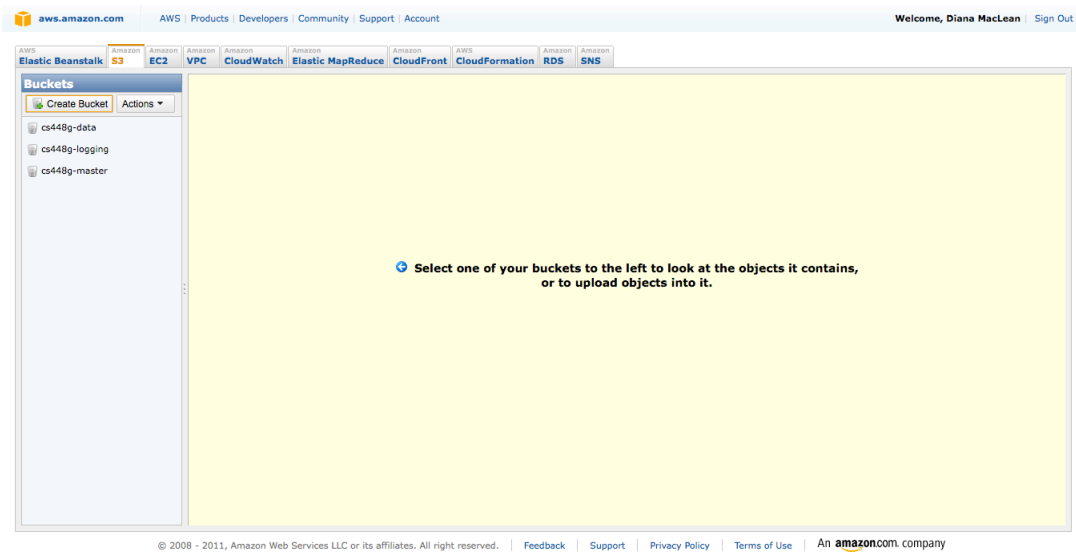
Go to the AWS Management Console, and click the "S3" tab. There should be a message informing you that you need to sign up for S3. Enter your own payment details for now – we will load credit onto the account once we're all set up.

We need S3 for two reasons: 1.) an EMR workflow requires the input data to be on S3; 2.) EMR workflow output is always *saved* to S3. Data (or objects) in S3 are stored in what we call "buckets". You can think of buckets as folders. For this class, we've put the data you'll need for Part 1 of Assignment 2 in a public bucket called:

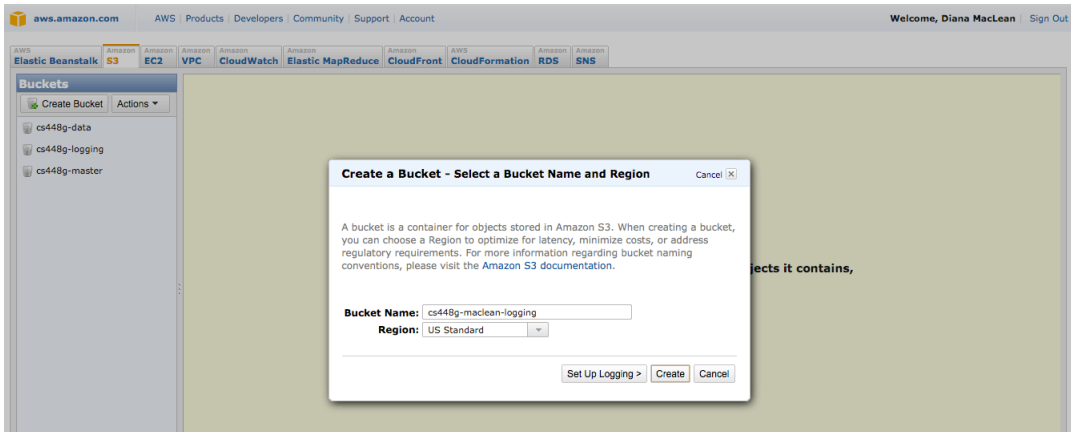
cs448g-data

You'll see how to reference this for EMR input later on. In the meanwhile, you'll need some buckets of your own to 1.) store your EMR output in, and 2.) store your log files if you want to debug your EMR runs. Once you're all signed up, you should see a message prompting you to get started by creating your own buckets. We'll create the log bucket first.

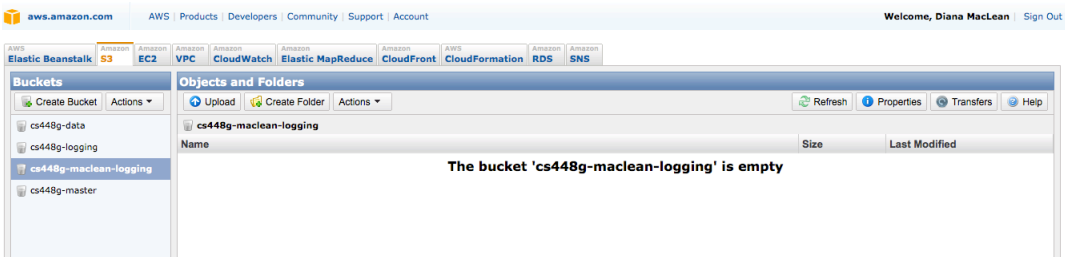
1. In the S3 console, click on "Create Bucket".



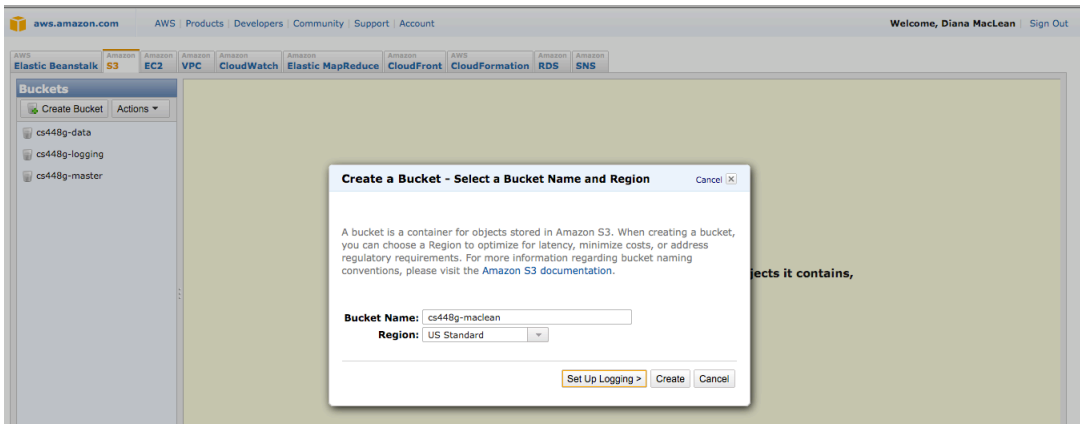
2. All S3 buckets have to have unique names, so call your logging bucket cs448g-lastname-logging. Click on "Create" (not on "Set Up Logging >>").



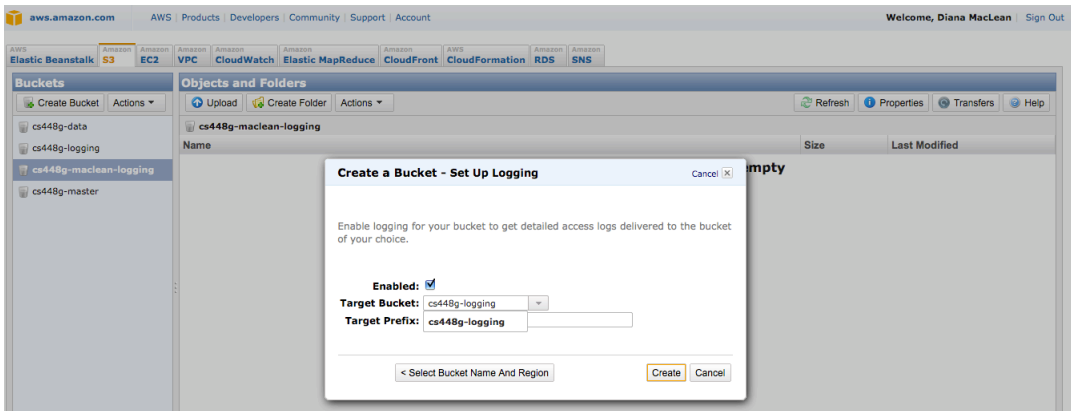
3. Your new bucket will appear in the side bar, with an informative message telling you that it is empty.



4. Now we'll create our main bucket. Again, click on "Create a Bucket". Call this one cs448g-lastname. We'll link our logging bucket to the one we're creating now, so click on "Set Up Logging >>".



5. "Enable logging", and start typing in the name of your logging bucket. It should appear in the drop down menu. Select it, and "Create".

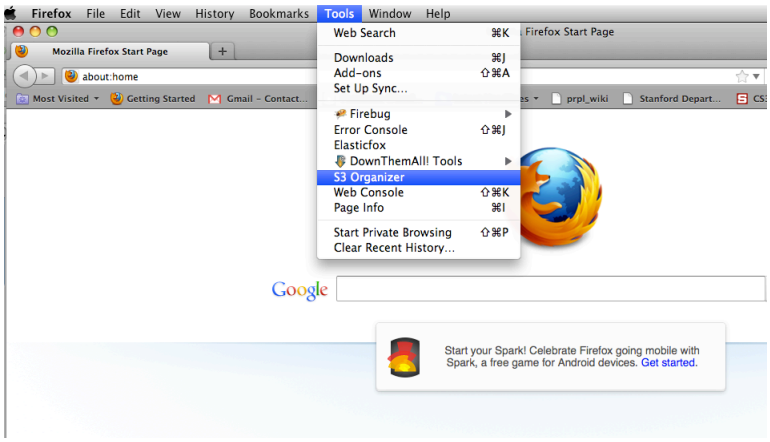


You're done creating your buckets! Awesome ☺

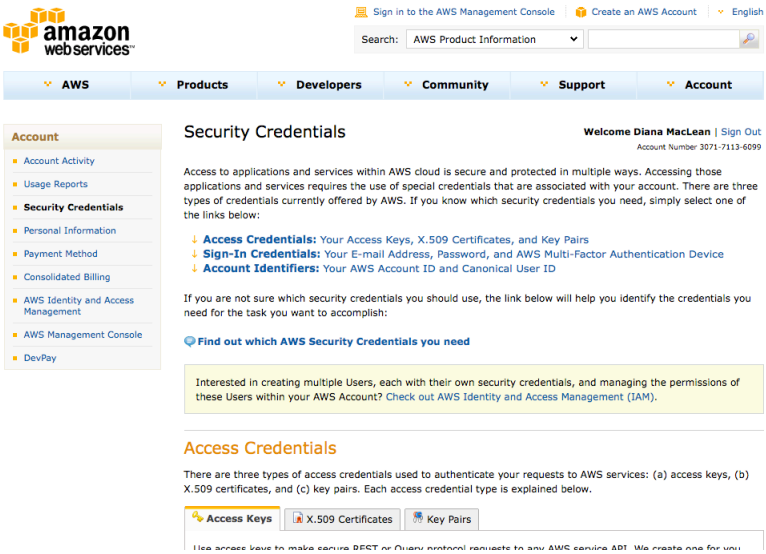
1a-i) Sign up for S3Fox (Optional, but recommended)

You can manage your buckets through the S3 console, but there is a fantastic plugin for Firefox that makes uploading and downloading between S3 and your local machine really easy. Of course, you'll need to be using Firefox.

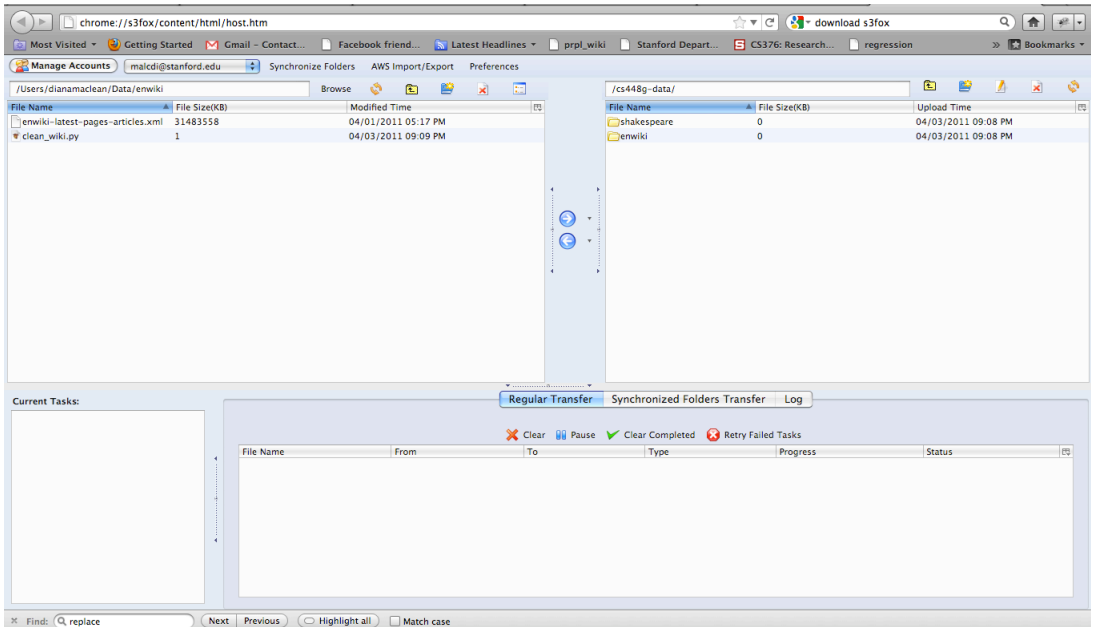
From http://www.s3fox.net, click on "Download". You should get a popup window asking you if you want to install the plugin (example here: http://www.s3fox.net/DownloadPage.aspx). Once you've installed, restart Firefox. You should be able to access "S3 Organizer" by clicking on the "Tools" tab in the Firefox taskbar, as shown below.



When you first open S3Fox, you'll need to enter your Access Key and your Secret Key. You can find these by clicking on "Account" in the AWS Management Console (http://aws.amazon.com/account), and then on the "Security Credentials" link. This will take you to a page where you can view your Access Keys.

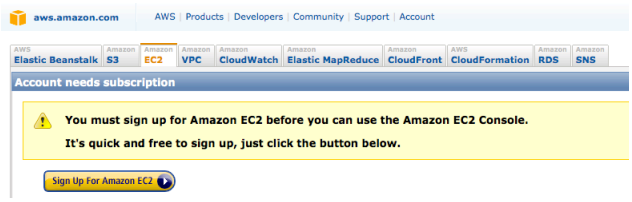


Once you've correctly entered your credentials in S3Fox, you should see a console that looks something like the following (only yours won't have files in it yet). S3Fox will make it much easier to upload and download your scripts and output files once we start running our EMR jobs.

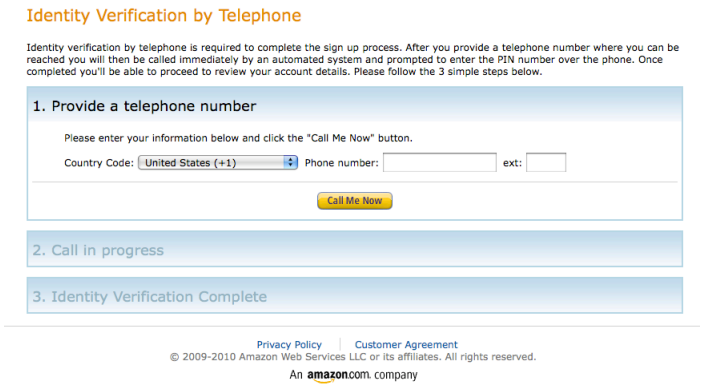


1b: Sign up for EC2

In the AWS console, click on the "EC2" tab, and sign up for EC2.



Proceed as before; this time you will probably be asked to verify your identity via phone. This should not take more than a couple of minutes to complete.



Once you've verified your identity, you should get an email verification stating that you're ready to use EC2. Awesome!

You can read more about EC2 here: http://aws.amazon.com/ec2.

EC2 is essentially a big mass of machines that you can "rent" to run your own workloads on. EC2 is great because you can customize your EC2 instances according to your compute requirements (e.g. specify RAM, softward etc.). We won't launch instances directly, but when we run MapReduce jobs through EMR, EMR will use EC2 instances.

1c: Sign up for EMR

Repeat step 1b, except under the "Elastic MapReduce" tab on the AWS Management Console.

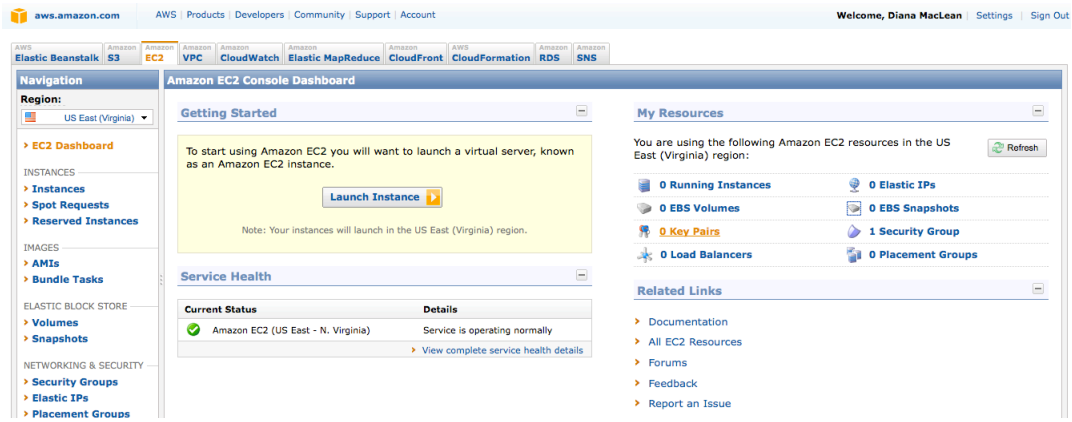
You can read more about EMR here: http://aws.amazon.com/elasticmapreduce

In a nutshell, EMR provides a framework for running MapReduce jobs using data stored on S3, and machines from the EC2 service. We'll be working mostly with EMR from here on.

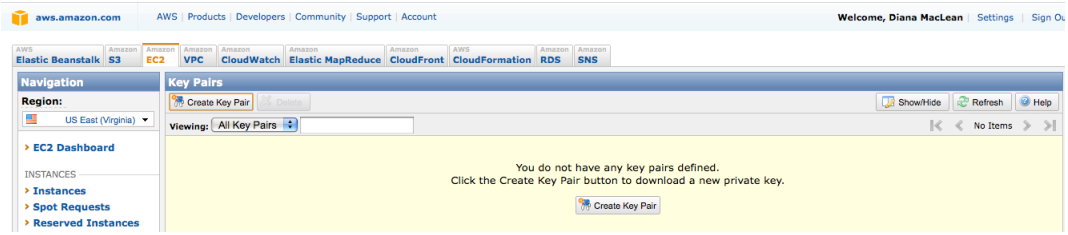
2: Access your account credentials, and create a key pair

If you signed up for S3Fox, then you already know how to access your account credentials. Take a look at section *1a-i* of this tutorial, if not.

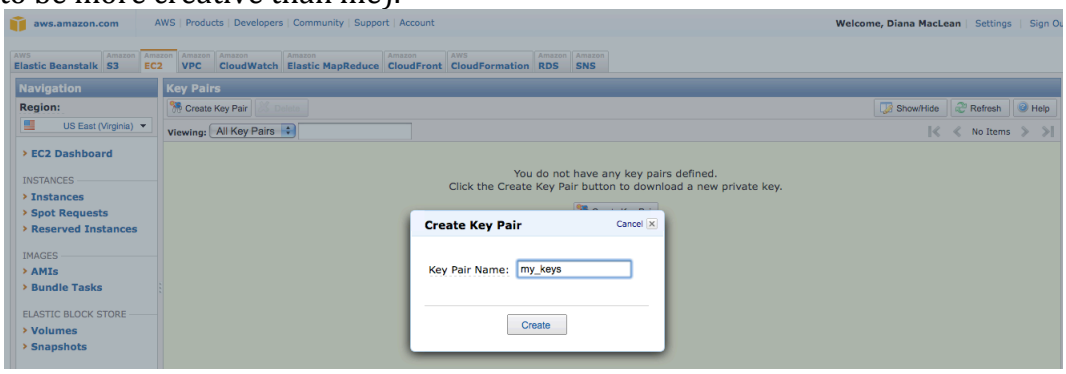
When we run jobs on EMR, you'll need to have a valid public/private key pair. You can also use these keys to ssh into EC2 instance machines that you launched, but we won't go through that here. (Come and find me if you want to do this.) To create your first key pair, click on the "EC2" tab in the AWS Management Console.



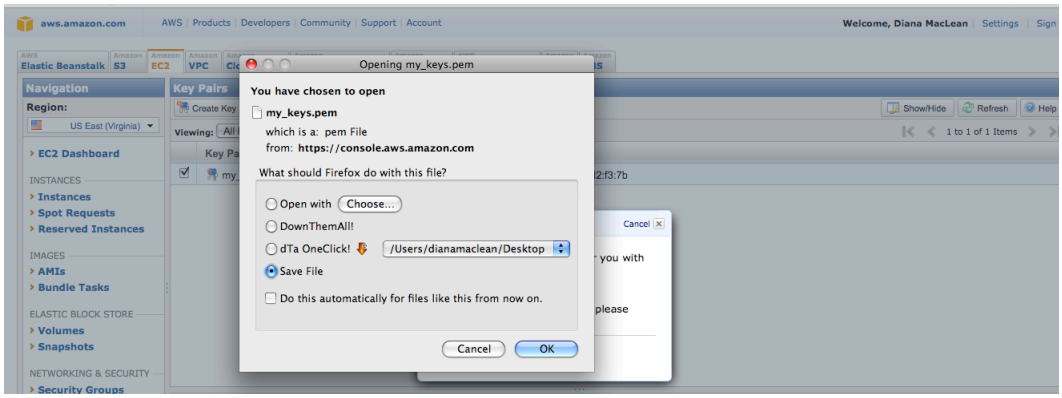
On the right, you should see a link stating "0 Key Pairs". Click on this.



You'll be given an option to "Create Key Pair". Name your key pair as you wish (try to be more creative than me).



Upon providing a name and clicking on "Create", your private key (a .pem file) will automatically download. Save it in a safe place where you'll be able to find it again.



If you need to access your public key, you'll be able to find it in the same place where you found your account credentials. Amazon keeps no record of your private key, however, so if you lose it, you'll need to generate a new set. (Note: You won't really need to access your private key if you use the AWS Management Console, but you will be asked to name your key pair each time you run an EMR job.)

3: Add Credit to your Account

Finally, the part you've been holding your breath for! Adding (free) credit to your account! In order to be credited, you'll need your unique Credit Coupon Code. If you have not received this yet, e-mail me and I will issue you one.

Once you have your code, go to your account page (http://aws.amazon.com/account) and click on "Payment Method". You'll see your billing details; if you scroll down, you will also see an option to redeem a Credit Coupon.

Redeem an AWS Credit Coupon

If you were issued an AWS Credit Coupon, you may add it by clicking the link below. We will charge your AWS credit balance before charging your credit card.

› Redeem/View AWS Credits

Click on Redeem/View AWS Credits, and enter your code to get your credit. *Please email the CS448G course staff* ***immediately*** *if this does not work.*

Unfortunately, we can only give you so much free credit, so don't go too wild! You can check on how much credit you have left by clicking on the "Account Activity" link from your account page. Sometimes this can take a while to update, so don't be surprised if recent changes aren't immediately apparent.

4: Familiarize yourself with S3, EMR and EC2

Now you should be all set up to start with Assignment 2. The more familiar you are with AWS, however, the easier it will be. Browse around the AWS site – there are plenty of helpful tutorials and resources there. Party on ☺