Step by step: How to run the Run the sample word-count application from AWS

By Elaheh Raisi (elaheh@vt.edu) and B. Aditya Prakash

After creating an AWS account, storage buckets on S3 and a key pair:

1. Click on the Elastic MapReduce link in the Analytics Section of the AWS management console. This will take you to the EMR Cluster List page.

2. Click on the Create cluster link, you will be taken to ‘cluster configuration’ page.

3. Click on ‘Configure Sample Application’ on top right corner.

4. On the Configure Sample Application page select ‘Word count’ and in output location select the name of the bucket you will like the output to be. Enable logging and select the folder from the logging bucket you will like the logs to be.

5. After clicking ‘ok’ change the EC2 key pair in ‘Security and Access’. Choose the key pair created by you. For the number of EC2 instance, the default number is enough for word count.
6. Then click on Create Cluster.

7. If you choose there will be no Termination protection to Terminate the job:

And also you choose to Auto-terminate the step the job will finish by itself.

8. The following will be the Life Cycle of a Cluster:
9. Go to the EMR Cluster List Page, the job will be finished when you see ‘Terminated All steps completed’. It will take about 7 minutes. (The time taken to terminate the cluster totally depend on the number of cores you have selected. Here the number of cores is 10. The more the number of cores, the less time taken to terminate the cluster)

10. After this, go to the S3 console and click on the bucket you specify will contain the output in Step 4.

11. Go to the folder in the bucket that will contain the output. There should be files with the following format part-xxxxx. These files contain the output from each of the reducers in the job.
12. Select the file you will like to inspect and right click on it to open or download it. The file will contain a segment of the words from the example and the number of counts each one of them have.

```
A 14716
a 52
okar 3
sargau 3
obad 3
abandoned 46
abandonment 6
abate 9
abuj 3
abbassid 4
abbes 3
abbi 3
abbreviated 27
abc 6
obda 3
obdal 3
abdallah 6
abdelaziz 18
abdellah 9
abdellahamed 6
abderrahmane 3
abderrazzak 3
abdication 6
abdikashid 3
abdominal 1
obdo 3
obdoulaye 33
obducted 3
abdualim 3
abduljabbar 3
```