

15319/15619: Cloud Computing, Recitation 3, 27/01/2015

Hands-On Exercise Details

Welcome! Glad you can join us! Today we will try to perform a hands-on exercise in the latter half of the recitation. To save some time, please get the following done **while we discuss the recitation slides**. If you choose to participate in this exercise, the expense will come from your allocated budget for Project 1.2. The budget allocated for Project 1.2 is \$15, and this exercise should not cost more a few cents for the hour, assuming you terminate the instance right after the demo.

Steps:

1. Sign into your AWS account at <http://aws.amazon.com>
2. Click on “**EC2**”
3. Click on “**Launch Instance**”
4. Click on **Community AMIs** and enter the AMI ID: **ami-b6094dde** in the search box
5. Select **t2.micro** as the instance type and click on “**Next: Configure Instance Details**”
6. Configure this instance as an On-Demand instance
7. Continue through the steps until you see Tag Instance, tag the instance with **Key=Project, Value=1.2**
8. And finally, finish the launch, if you choose an existing key-pair, make sure you have it ready with you to connect to the instance, otherwise, create a new keypair to connect.
9. Once your instance has launched, note down the Public DNS name of the instance and connect to it over SSH

Once you are logged in through SSH, navigate to **r3demo/python** folder. We will try running a sample wordcount flow in the t2.micro using the given map and reduce python example programs. Similar instructions apply for running the java version in **r3demo/java**.

- 1) The input file is **input.txt**
- 2) The input file will be piped into the mapper program (**wordcount_mapper.py**)
- 3) The output of the mapper program will be piped into the reducer program (**wordcount_reducer.py**)
- 4) The output of the reducer program will be sent to output file (output.txt)

To Run the Python Flow:

```
cat input.txt | ./wordcount_mapper.py | sort | ./wordcount_reducer.py > output.txt
```

To Run the Java Flow:

```
cat input.txt | java wordcountMapper | sort | java wordcountReducer > output.txt
```

Please Note:

This sheet only contains the instructions to run a small-scale test version of streaming mapreduce by emulating the operation on unix pipes. This method cannot be used to process the entire dataset. Instead, we will need to configure a streaming MapReduce job to use these mappers and reducer to run in Elastic MapReduce (EMR).