

Big Data on AWS v3.0 - Lab 5: Running Pig Scripts in Hue on Amazon EMR

==================================================================================================================

Using this command reference.

==================================================================================================================

1. Locate the section you need. Each section in this file matches a section in the lab instructions.

2. Replace items in angle brackets - < > - with appropriate values. For example, in this command you would replace the value - <JobFlowID> - (including the angle brackets) with the parameter indicated in the lab instructions:

elastic-mapreduce --list <JobFlowID>. You can also use find and replace to change bracketed parameters in bulk.

3. Do NOT enable the Word Wrap feature in Windows Notepad or the text editor you use to view this file.

==================================================================================================================

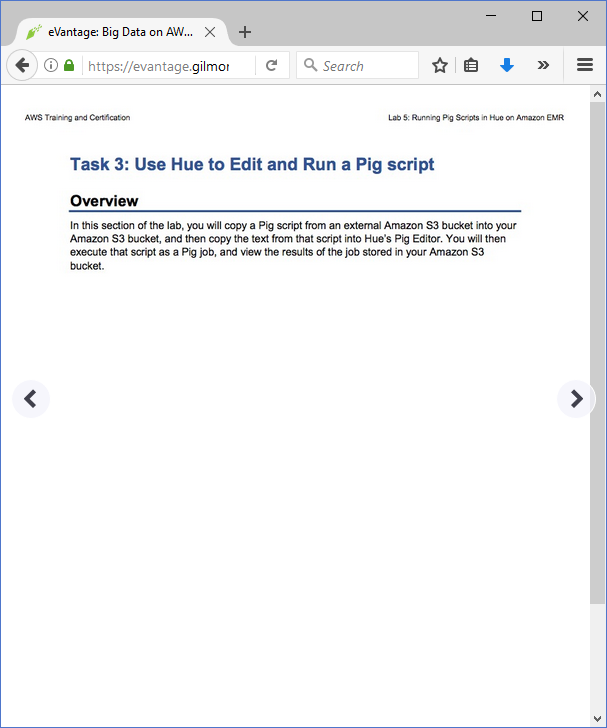
Task 1.3: Setting up your connection to Hue (Linux or MacOS)

==================================================================================================================

+++ Step 1.3.2

chmod 400 <key.pem>

ssh -i <key.pem> -L 8000:localhost:80 -L 8888:localhost:8888 hadoop@<master node hostname>



=================================================================================================================

Task 3.2: Use Hue to edit and run a Pig script

==================================================================================================================

+++ Step 3.2.3

s3://us-west-2-aws-training/awsu-ilt/AWS-200-BIG/v3.0/lab-5-pig-hue/scripts/piglabscript.pig

+++ Step 3.3.9, Input field:

s3://us-west-2-aws-training/awsu-ilt/AWS-200-BIG/v3.0/data/lab-5-pig-hue/data/

+++ Step 3.3.9, Output field:

s3://pig-bucket<yourinitials>/output/

