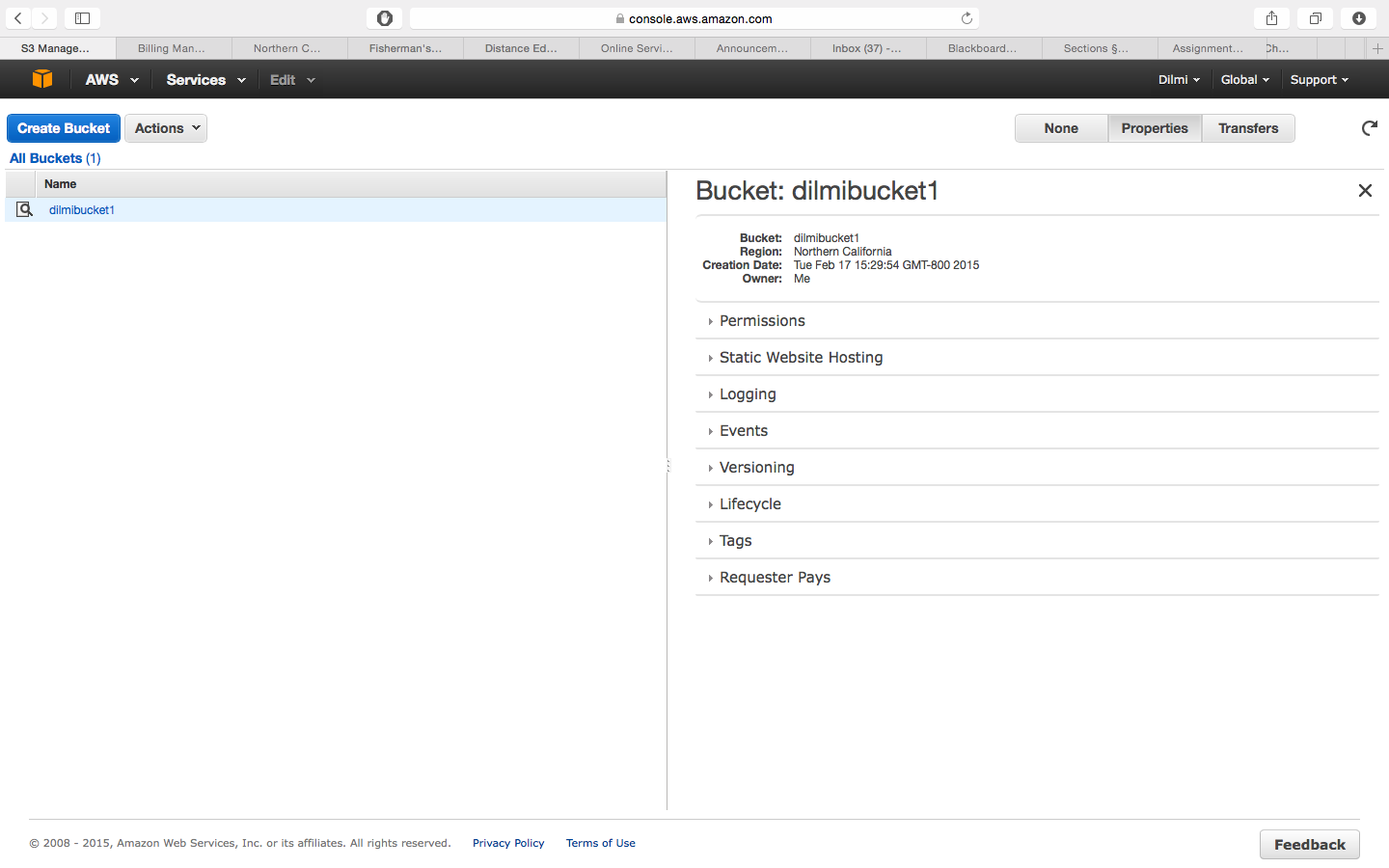
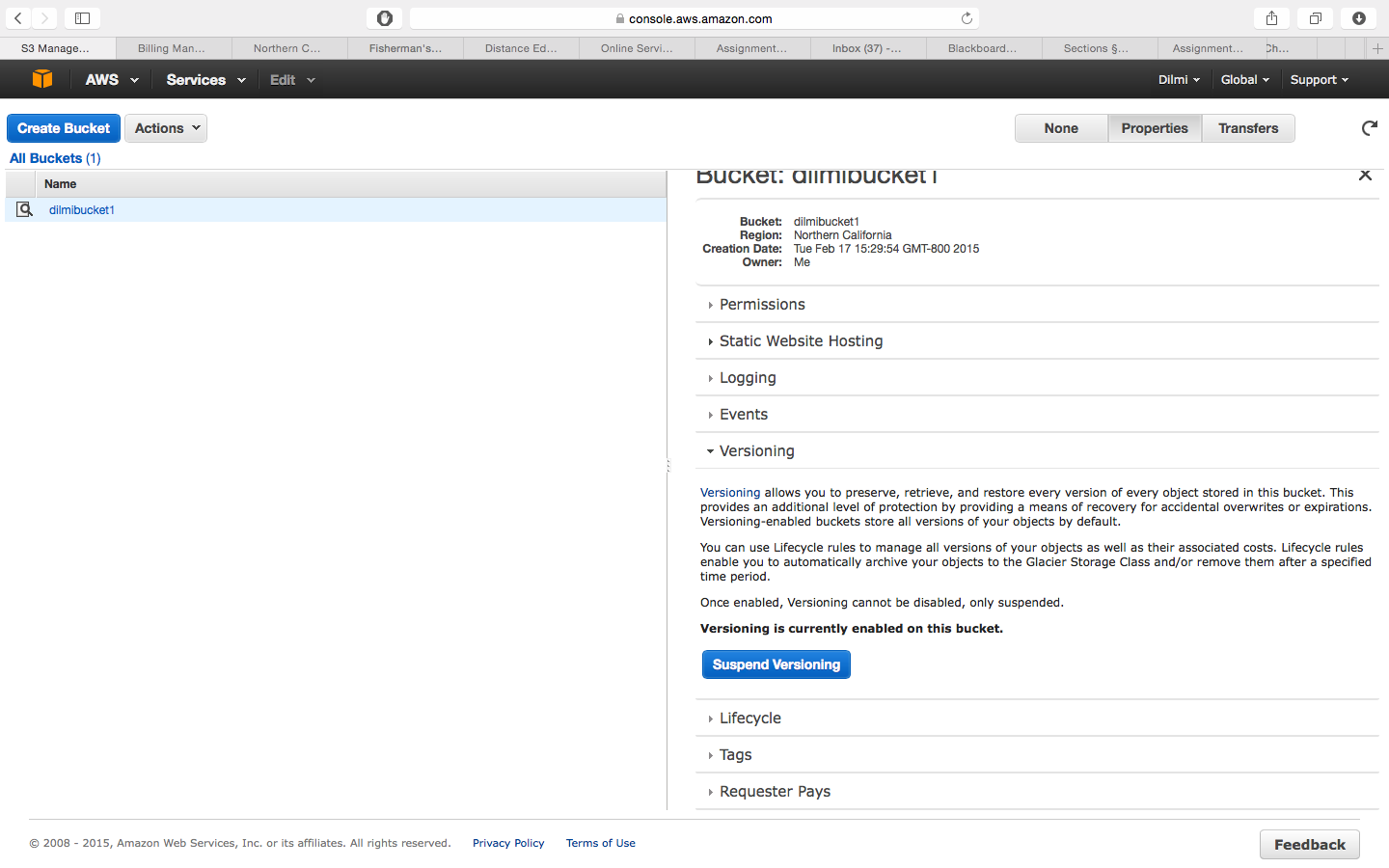
HU Extension **Assignment 03** E-63 Big Data Analytics

**Problem 1**

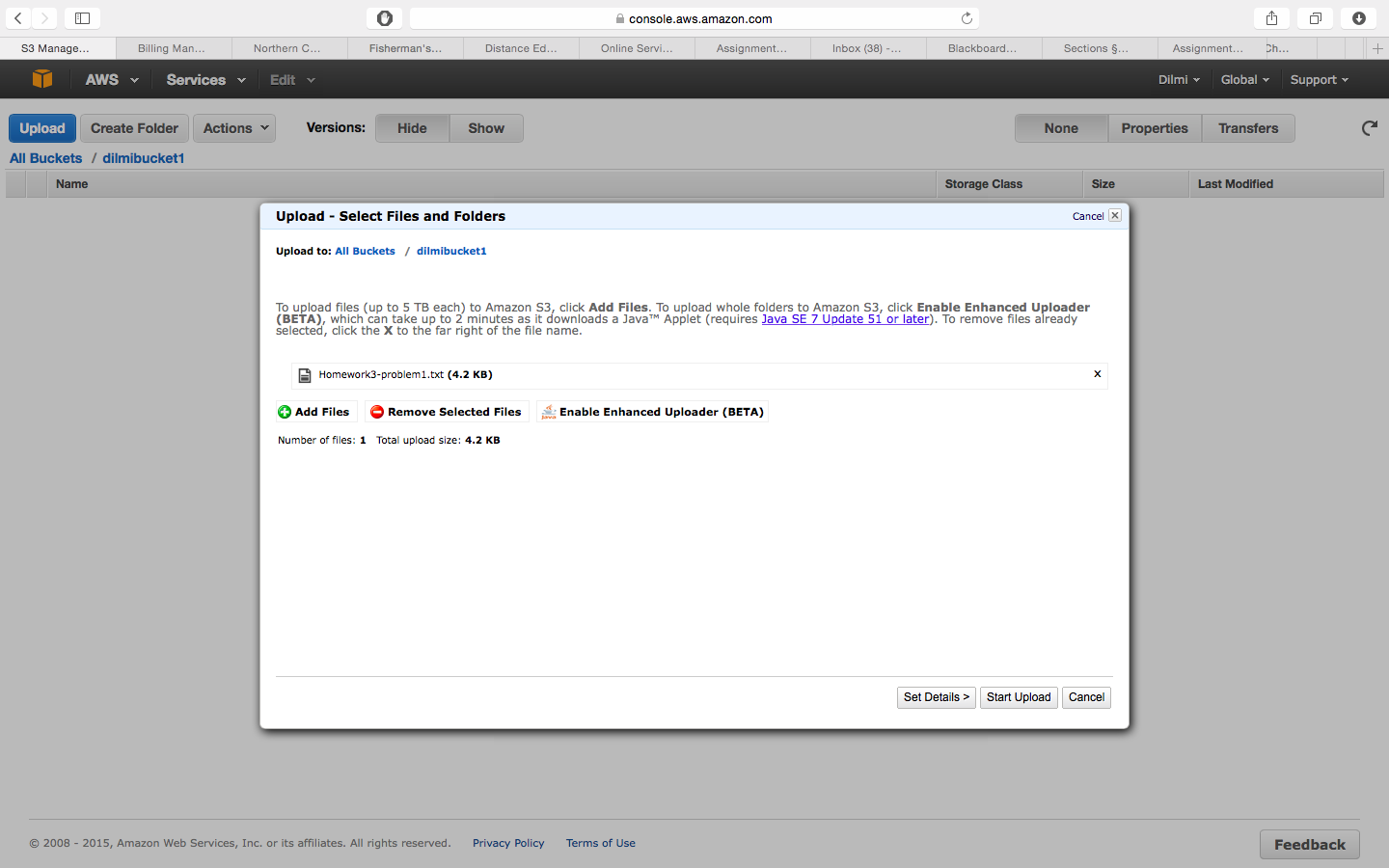
Create S3 bucket

****

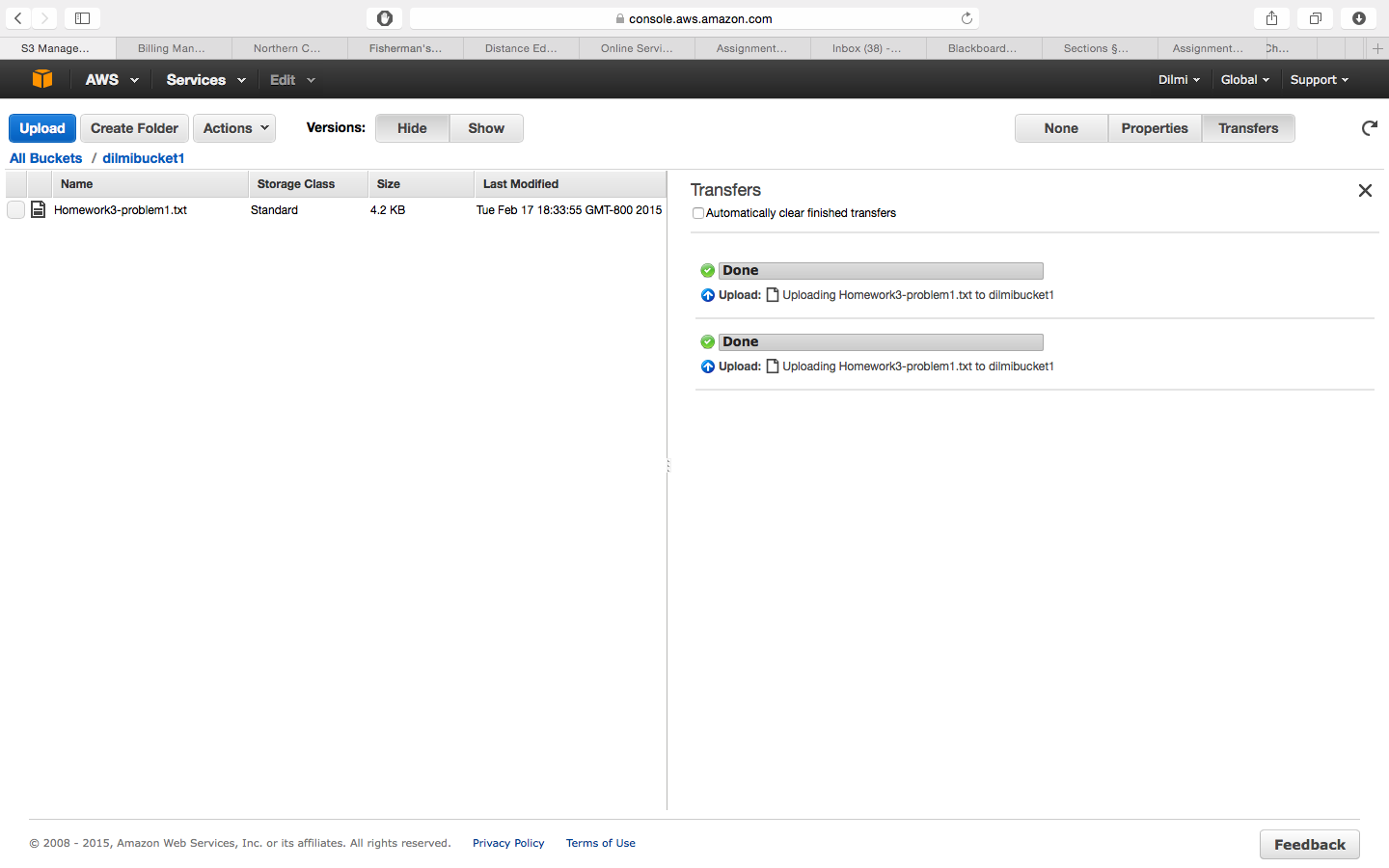
Enable versioning



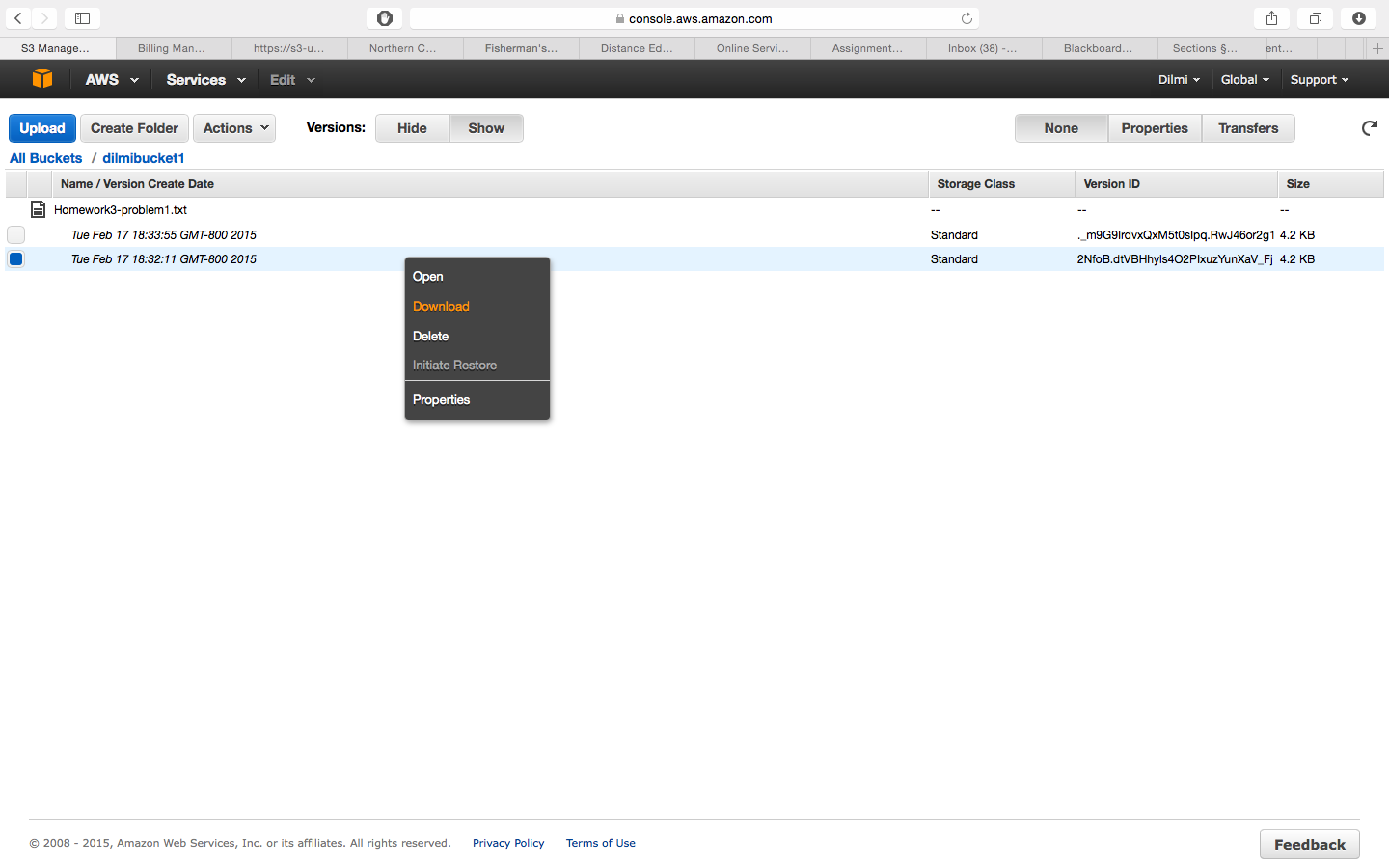
Upload file



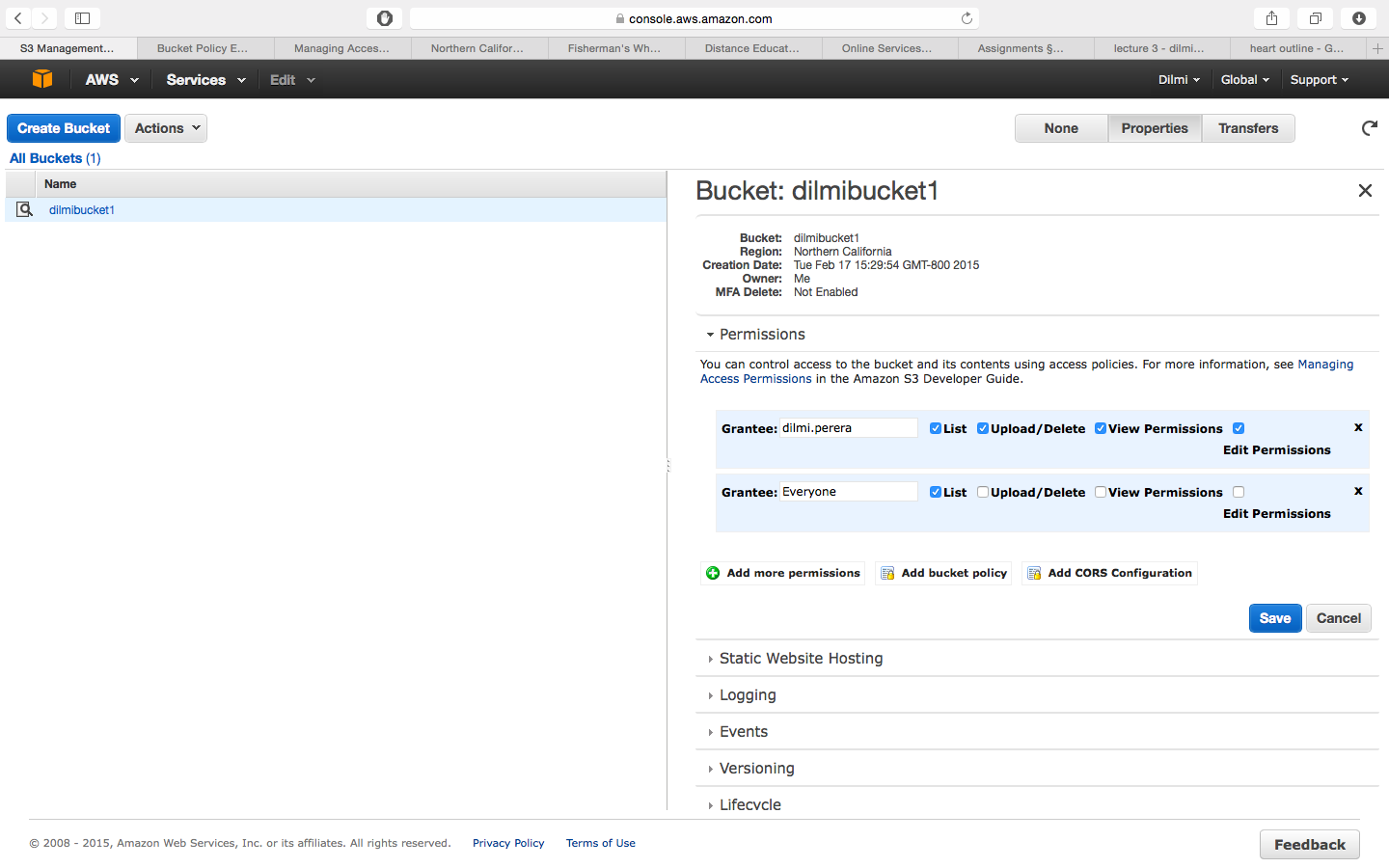
Upload modified file



Both versions can be retrieved



Permission for everyone to view the bucket



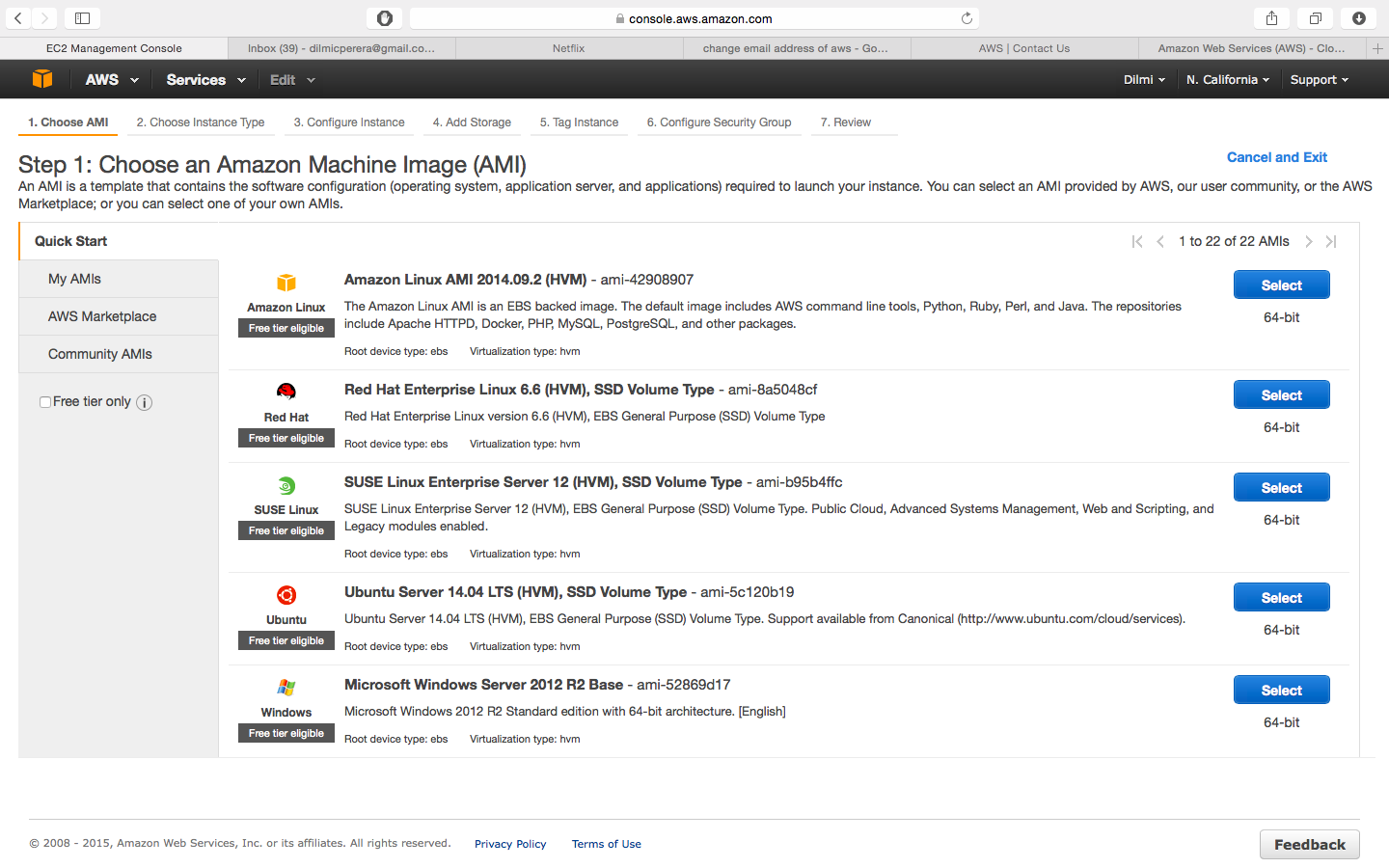
Link to download image

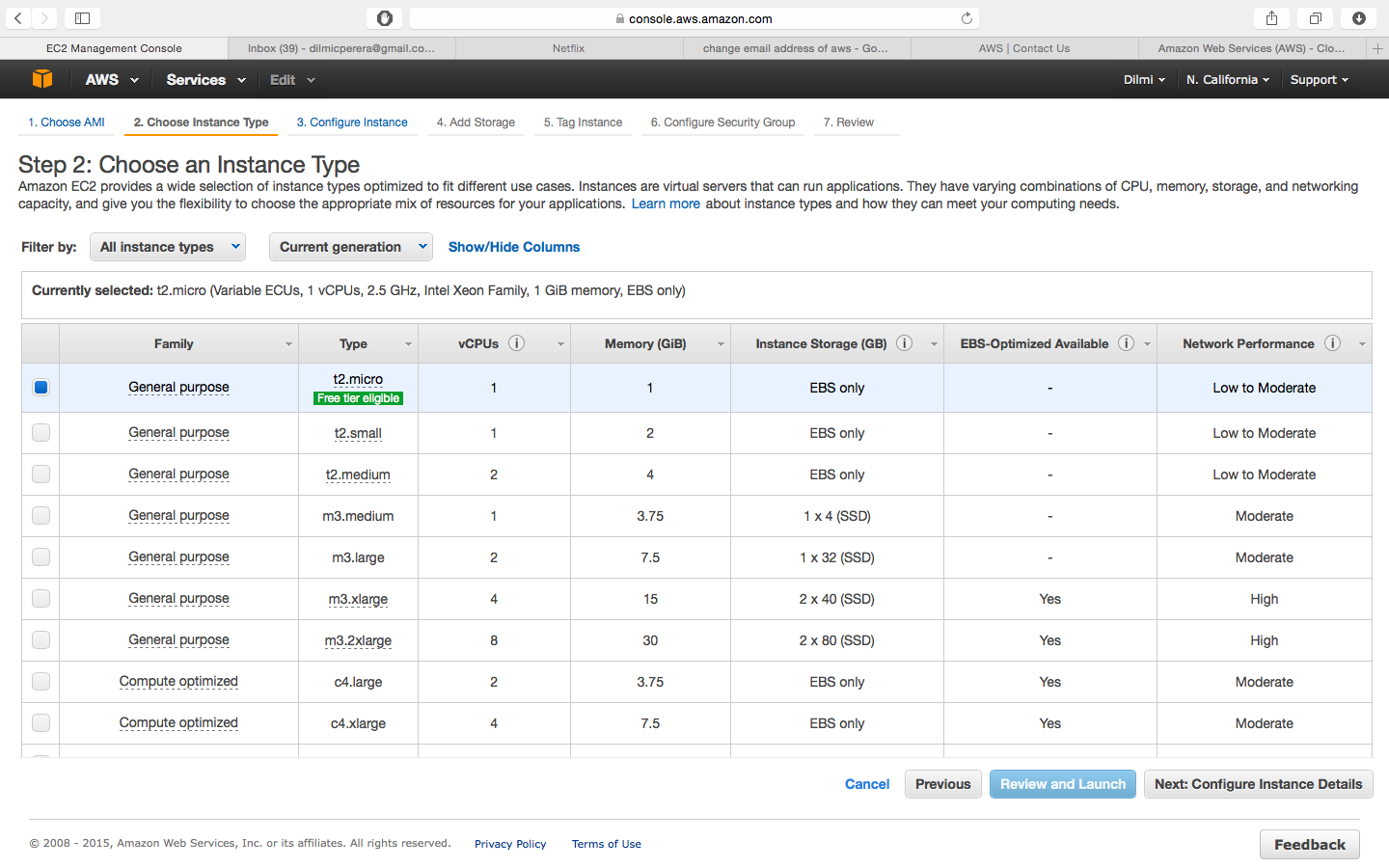
<https://s3-us-west-1.amazonaws.com/dilmibucket1/heart.jpg>

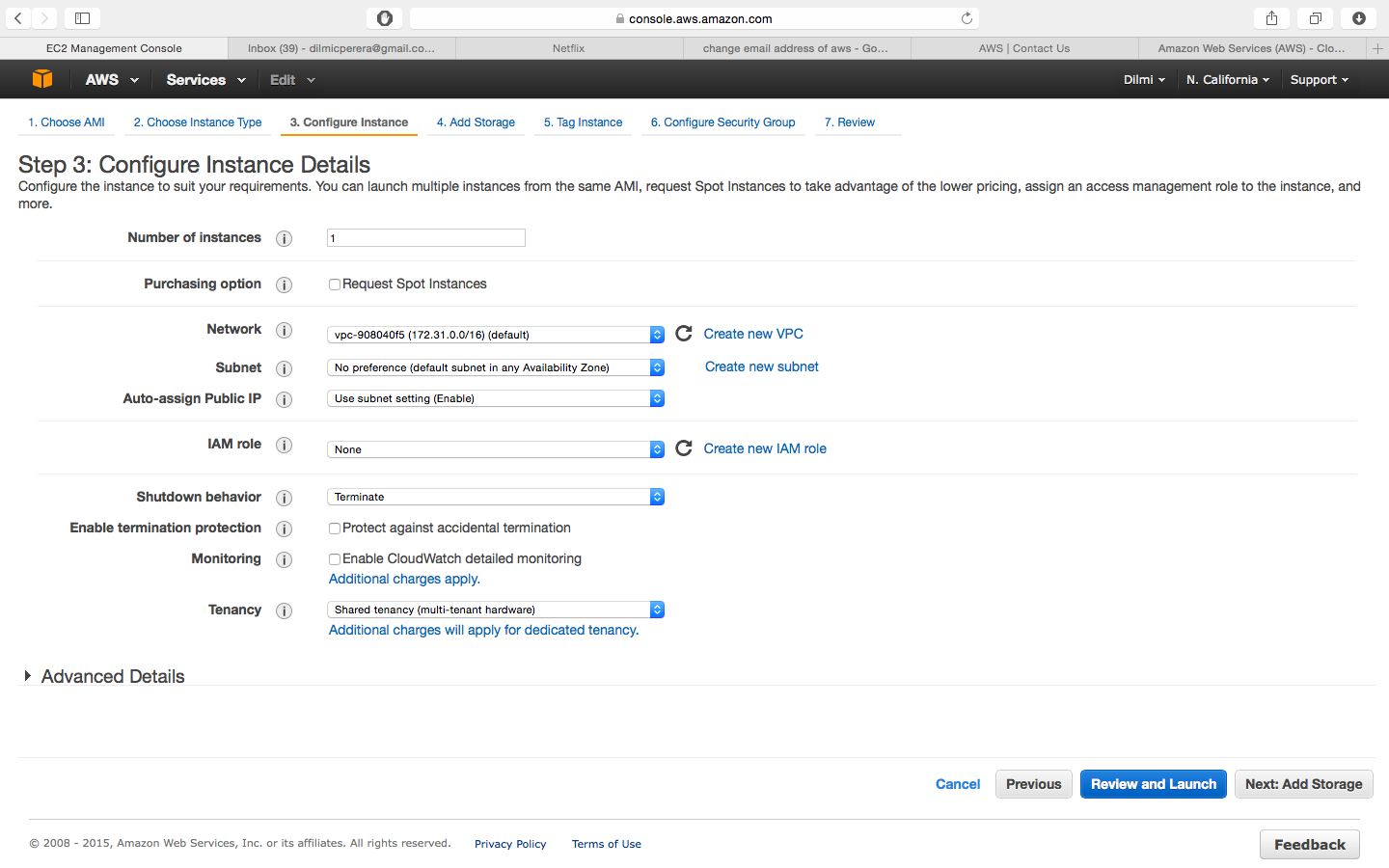
**Problem 2**

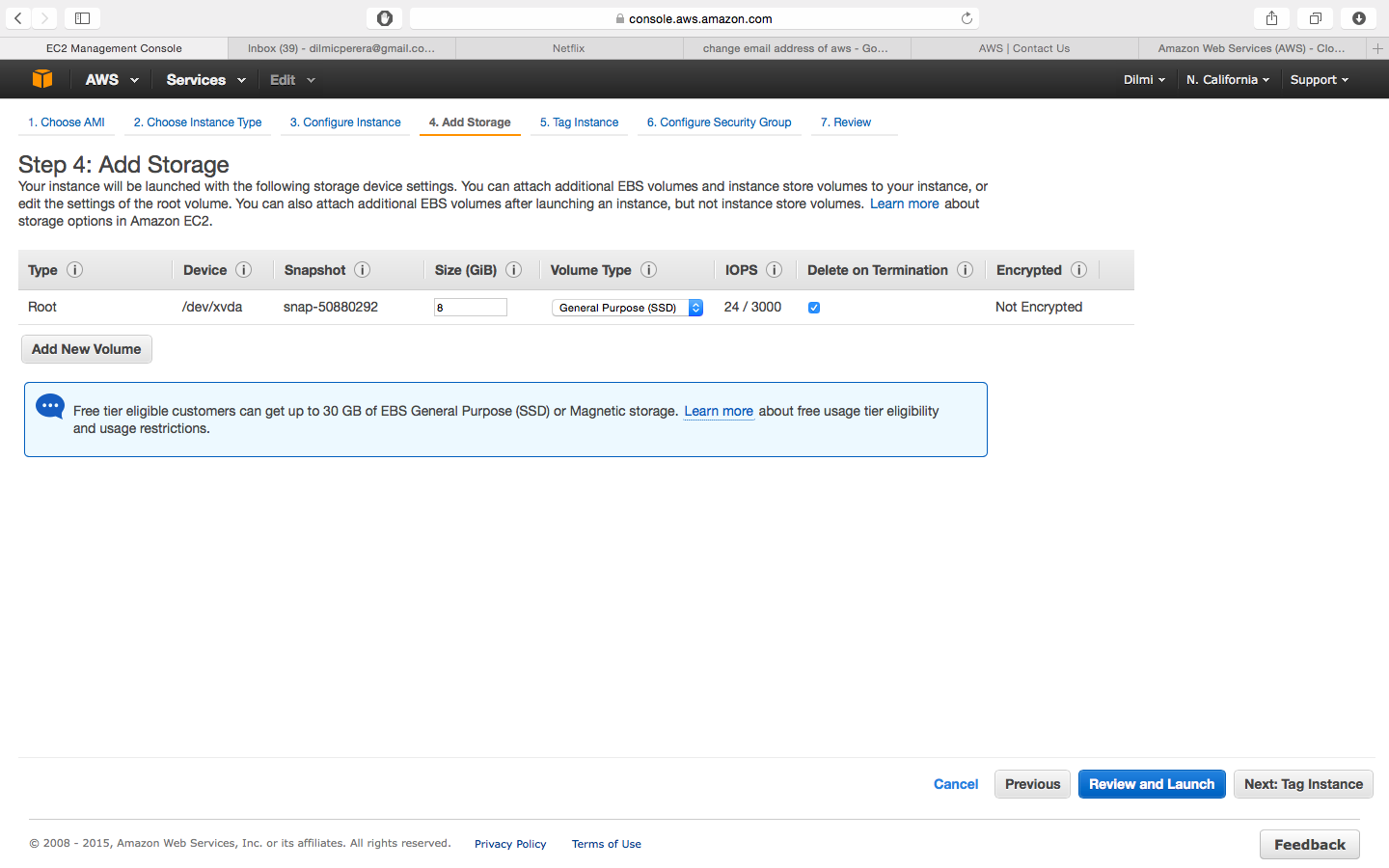
**Create EC2**

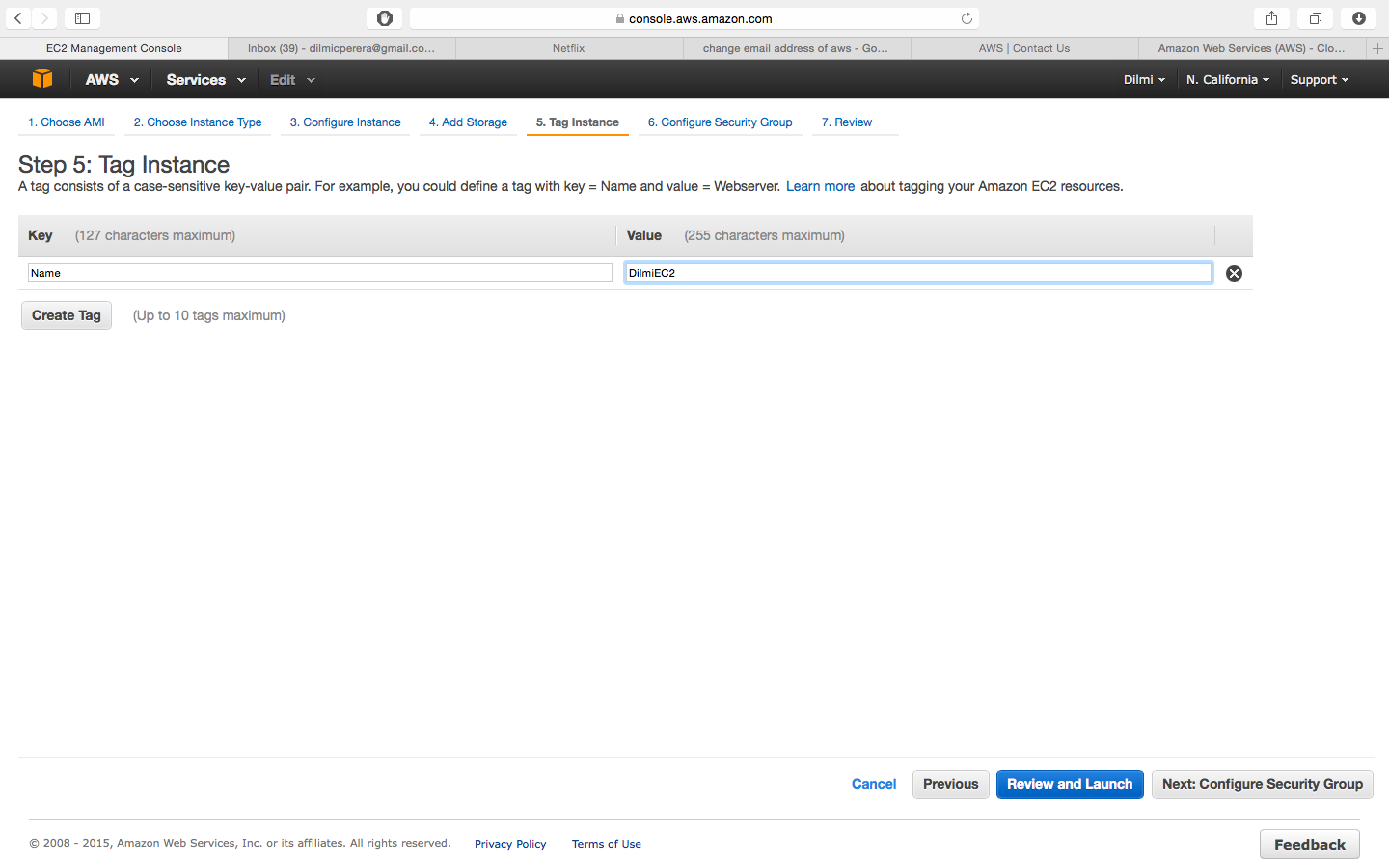
****

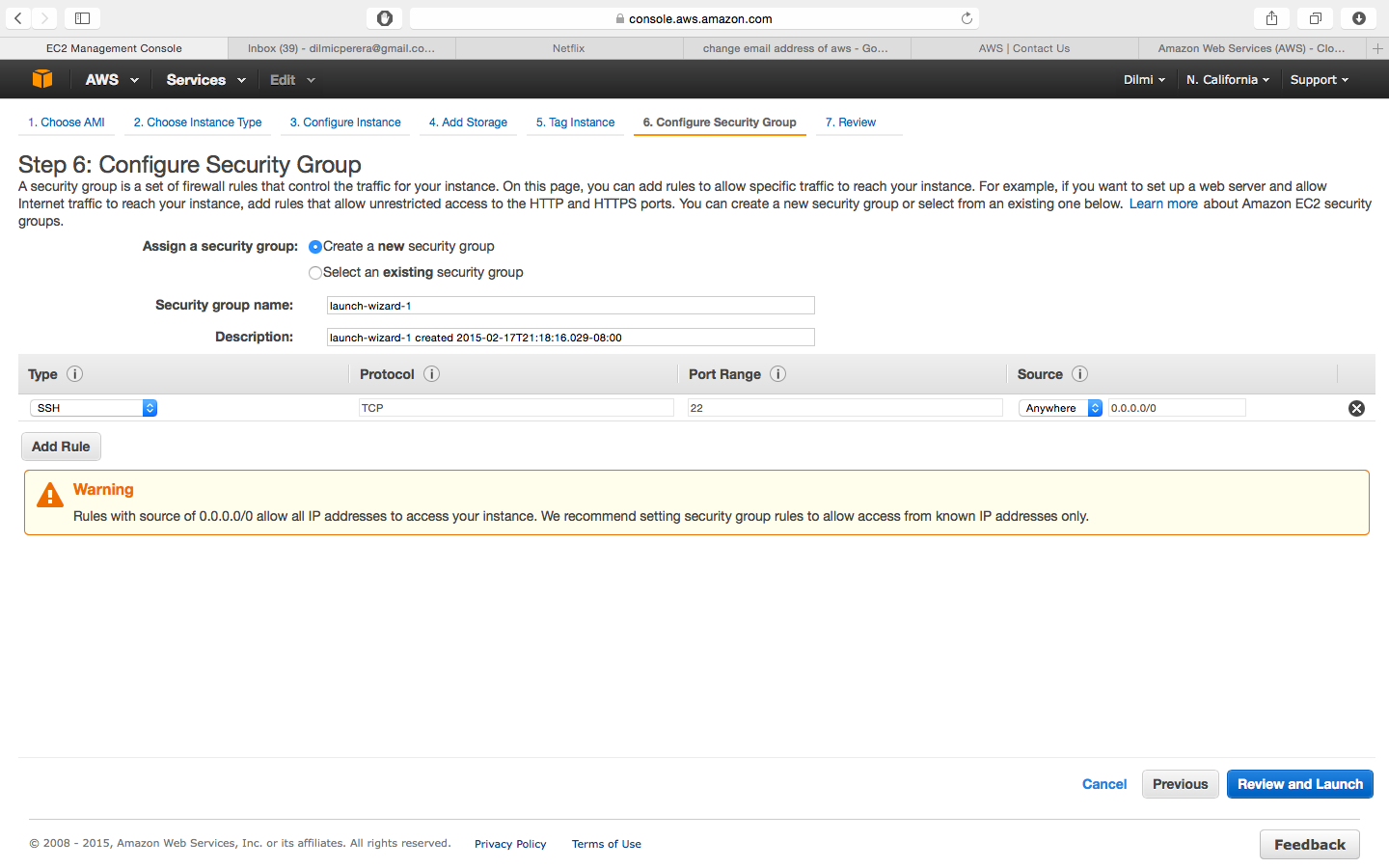
****

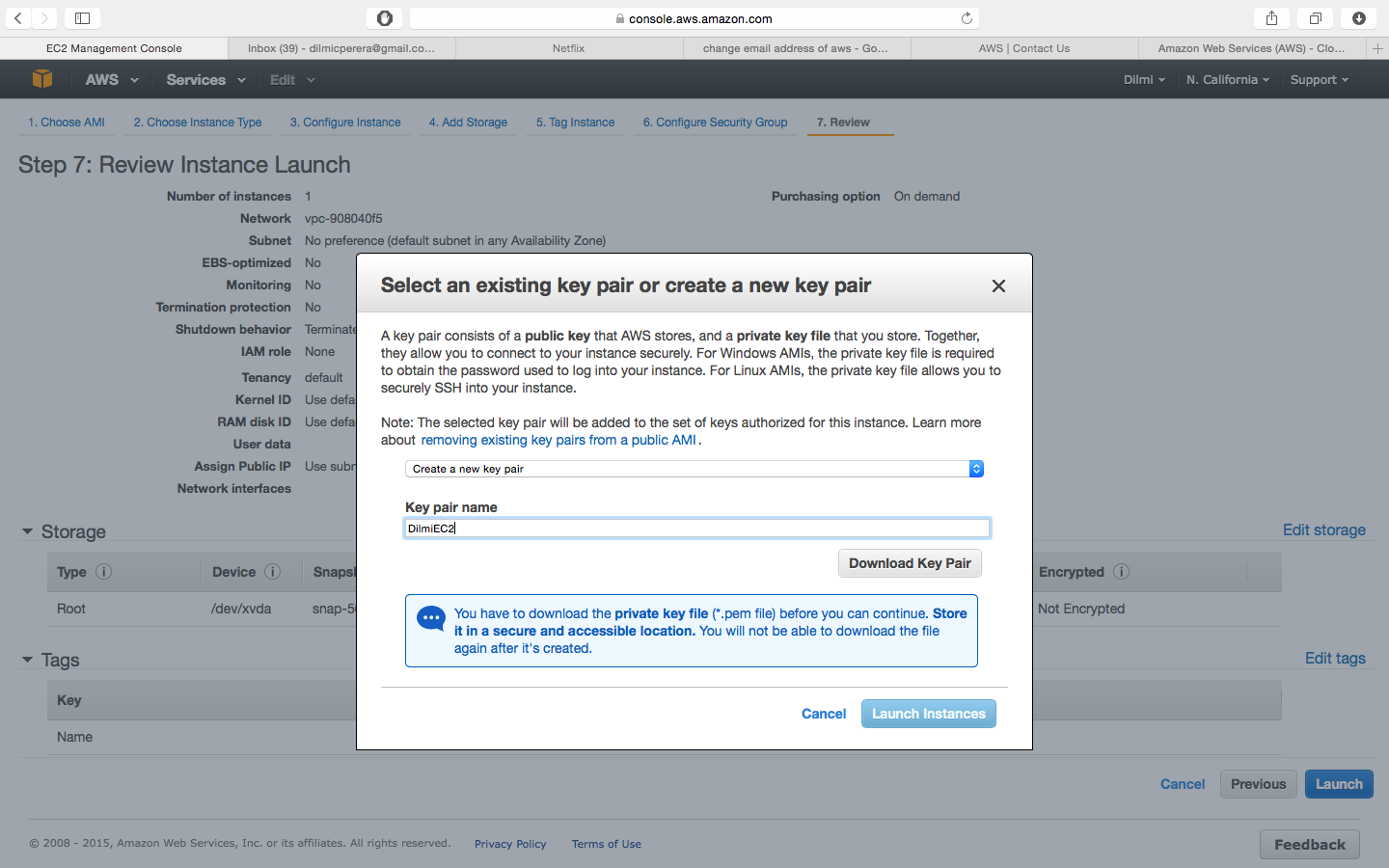
****

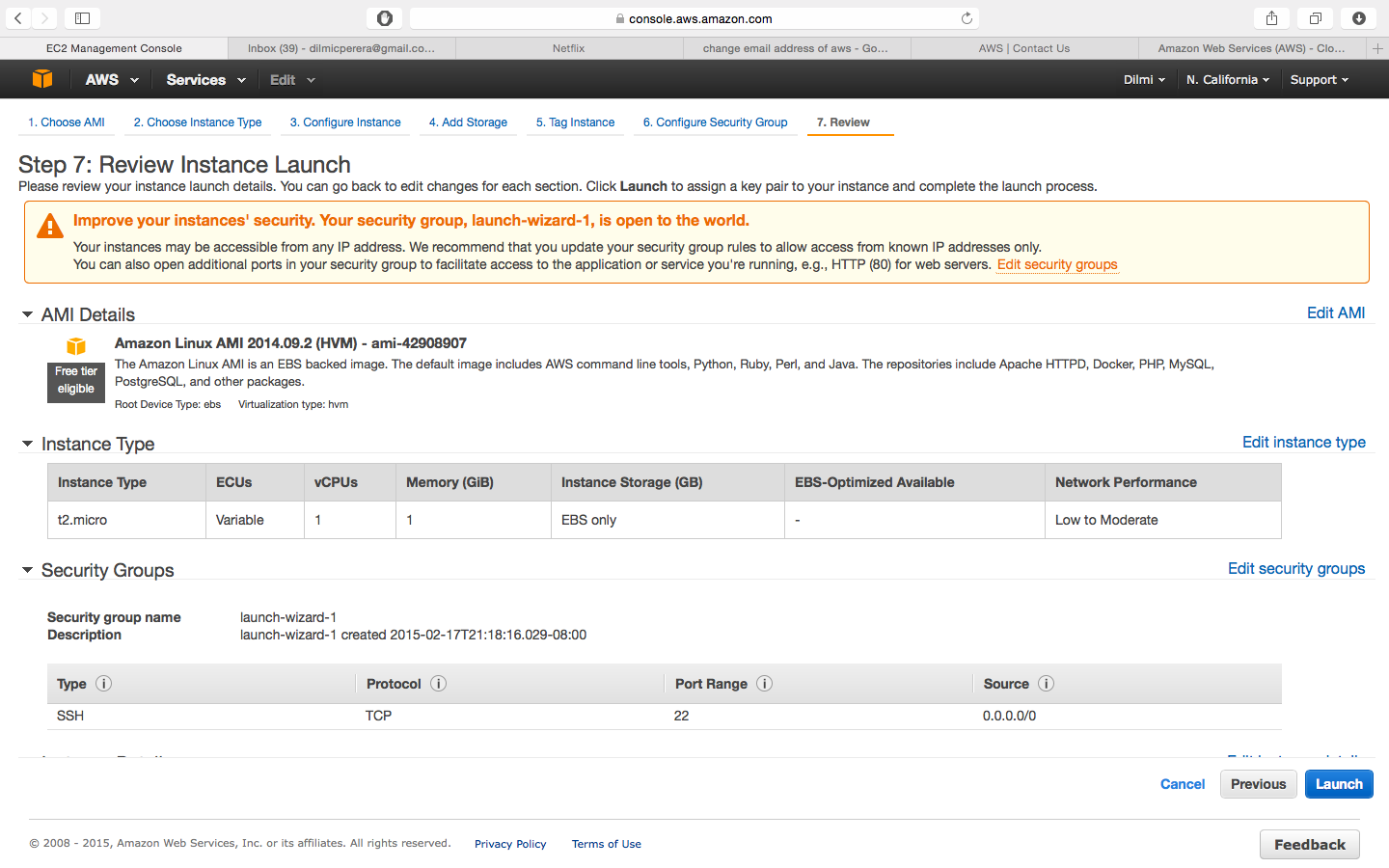
****

****

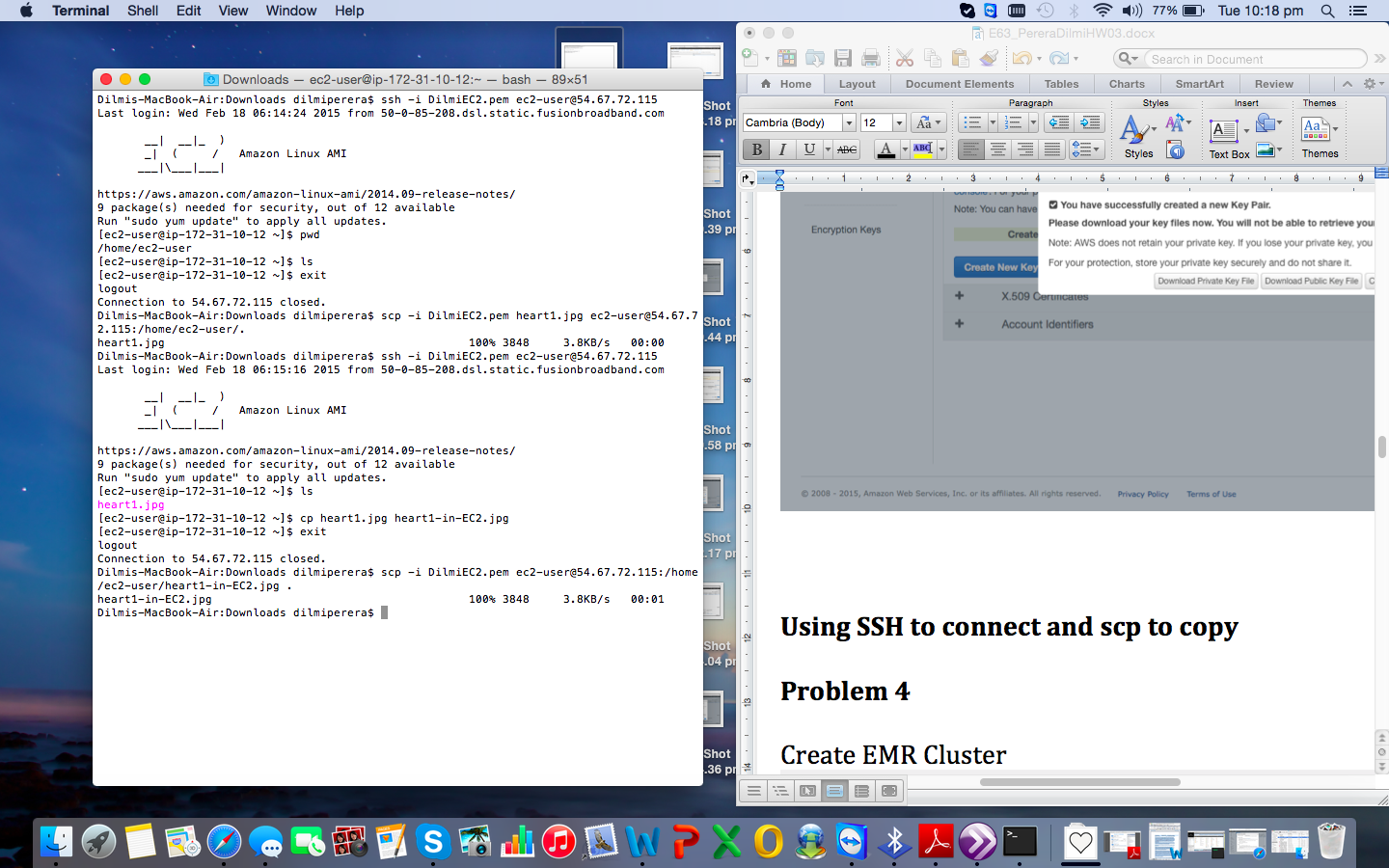
****

****

****

****

**Using SSH to connect and SCP to copy to and from AMI**

****

**Problem 3**

Java code: MapReduce.java.

Text file used: temp.txt

Output: WordCounts.txt

**First 5 words from program output(**WordCounts.txt**) and output from linux egrep+wc**

many 1

periodically 1

value 2

workers 4

key 5

egrep + wc

Dilmis-MacBook-Air:~ dilmiperera$ egrep -o '\<many\>' /Users/dilmiperera/Downloads/temp.txt | wc -w

1

Dilmis-MacBook-Air:~ dilmiperera$ egrep -o '\<Periodically\>' /Users/dilmiperera/Downloads/temp.txt | wc -w

1

Dilmis-MacBook-Air:~ dilmiperera$ egrep -o '\<value\>' /Users/dilmiperera/Downloads/temp.txt | wc -w

2

Dilmis-MacBook-Air:~ dilmiperera$ egrep -o '\<workers\>' /Users/dilmiperera/Downloads/temp.txt | wc -w

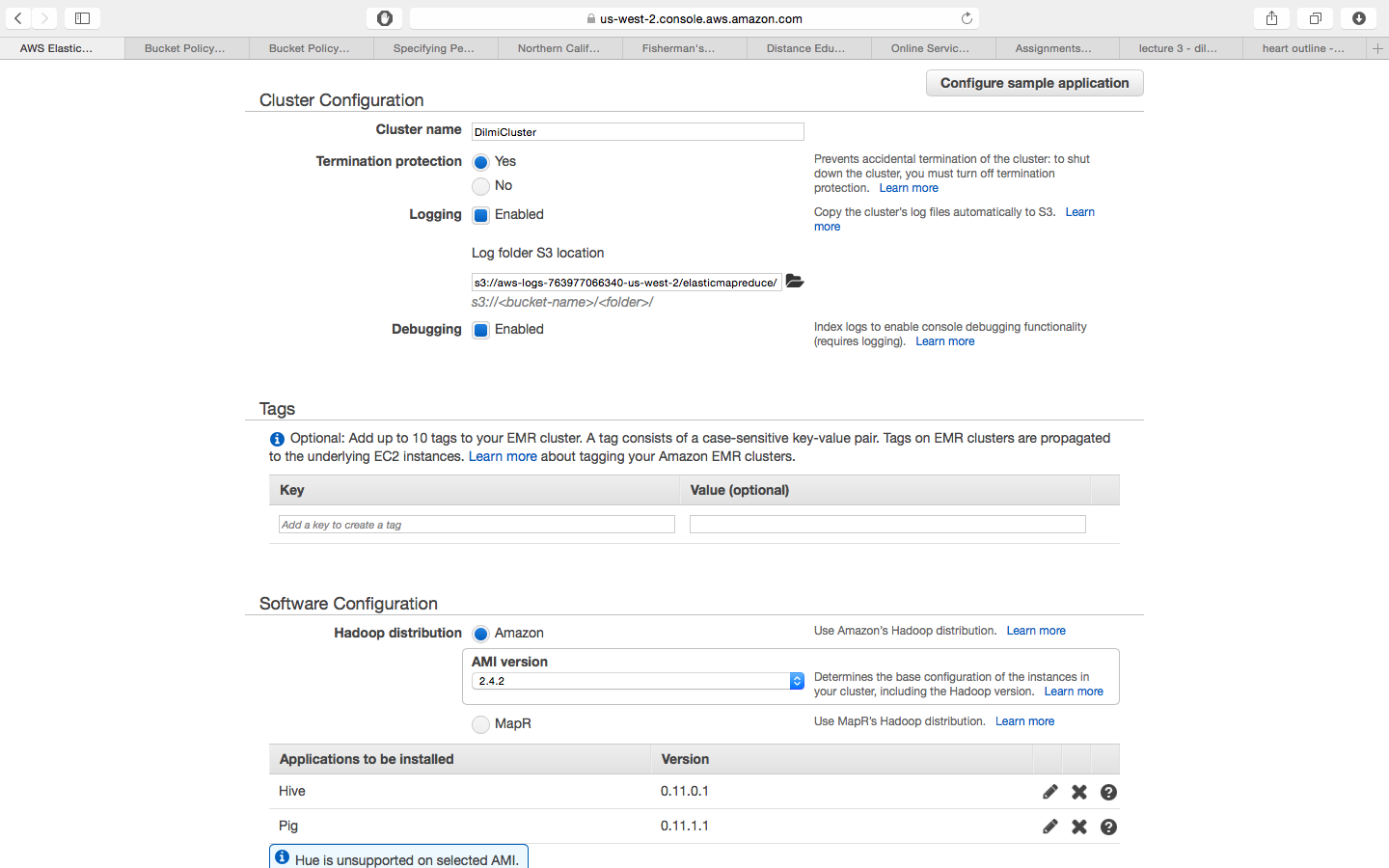
4

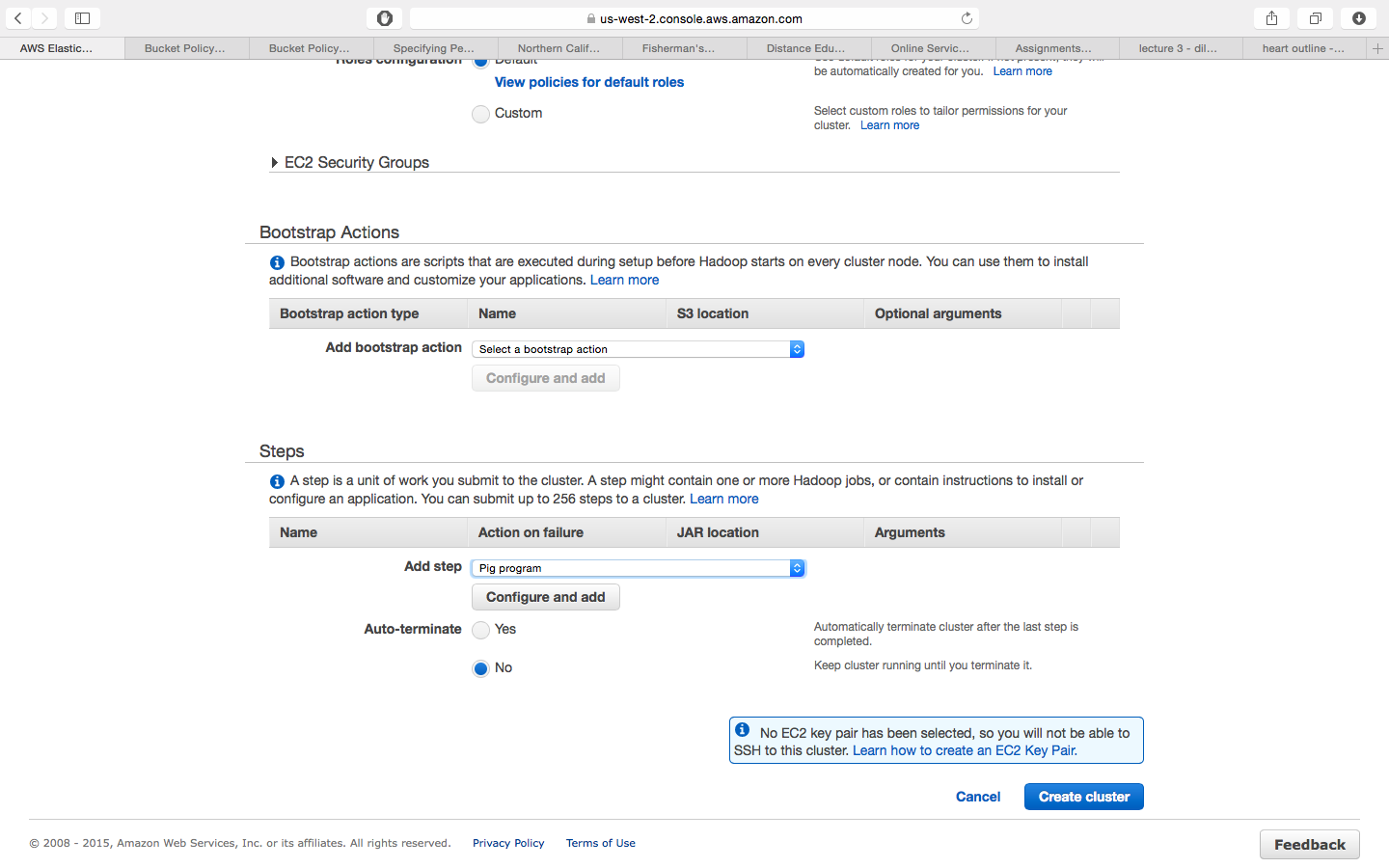
Dilmis-MacBook-Air:~ dilmiperera$ egrep -o '\<key\>' /Users/dilmiperera/Downloads/temp.txt | wc -w

5

**Problem 4**

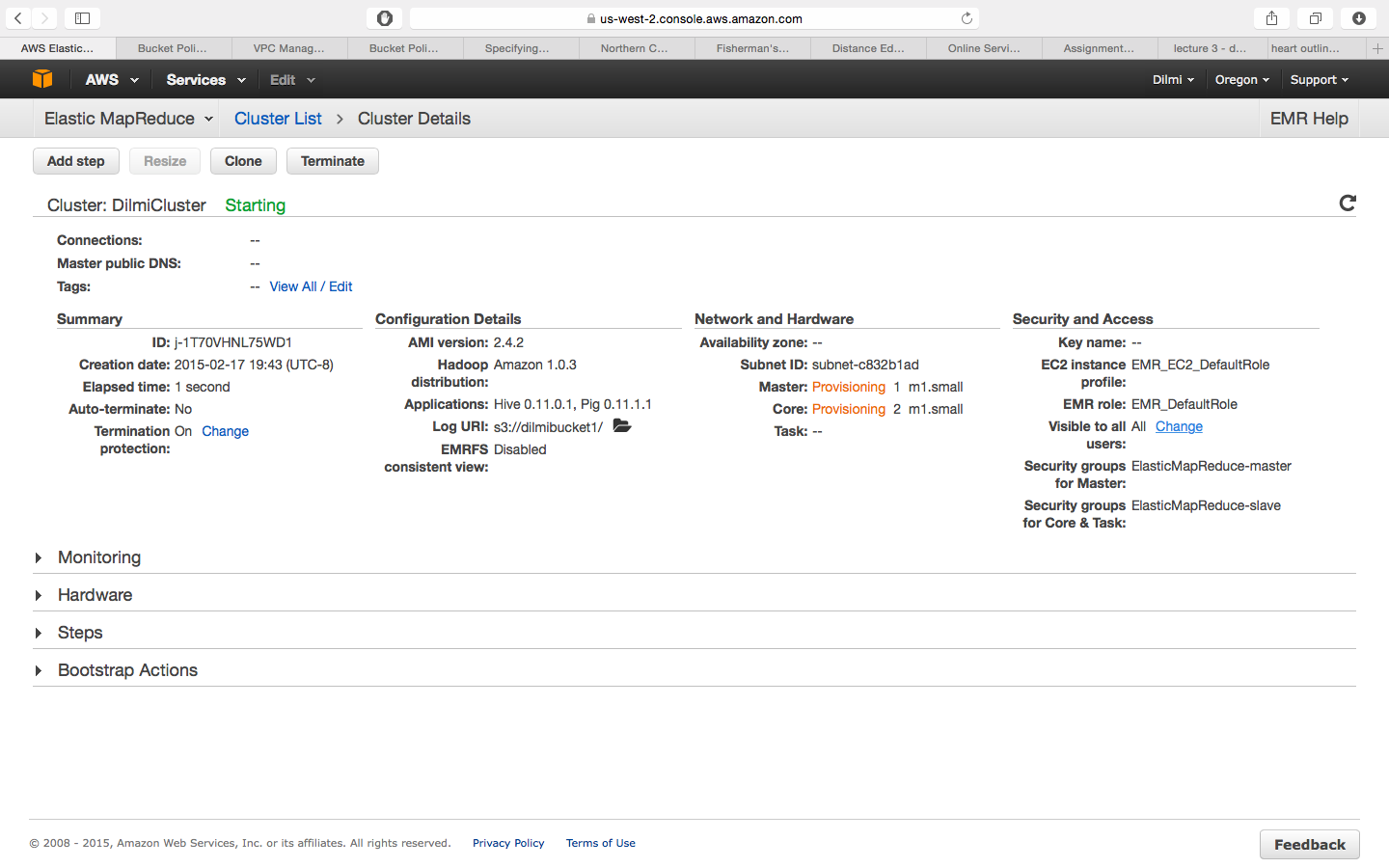
Create EMR Cluster

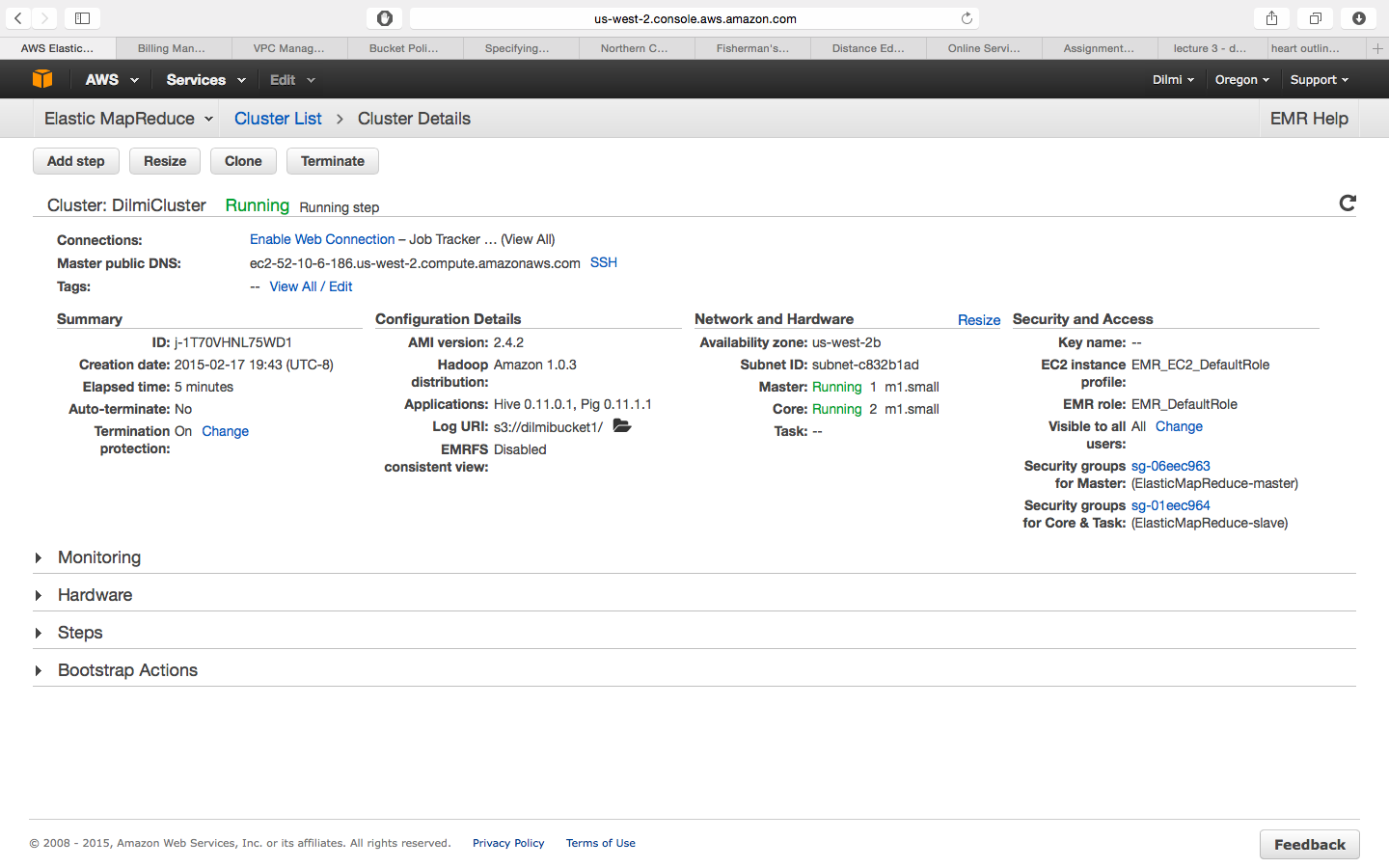


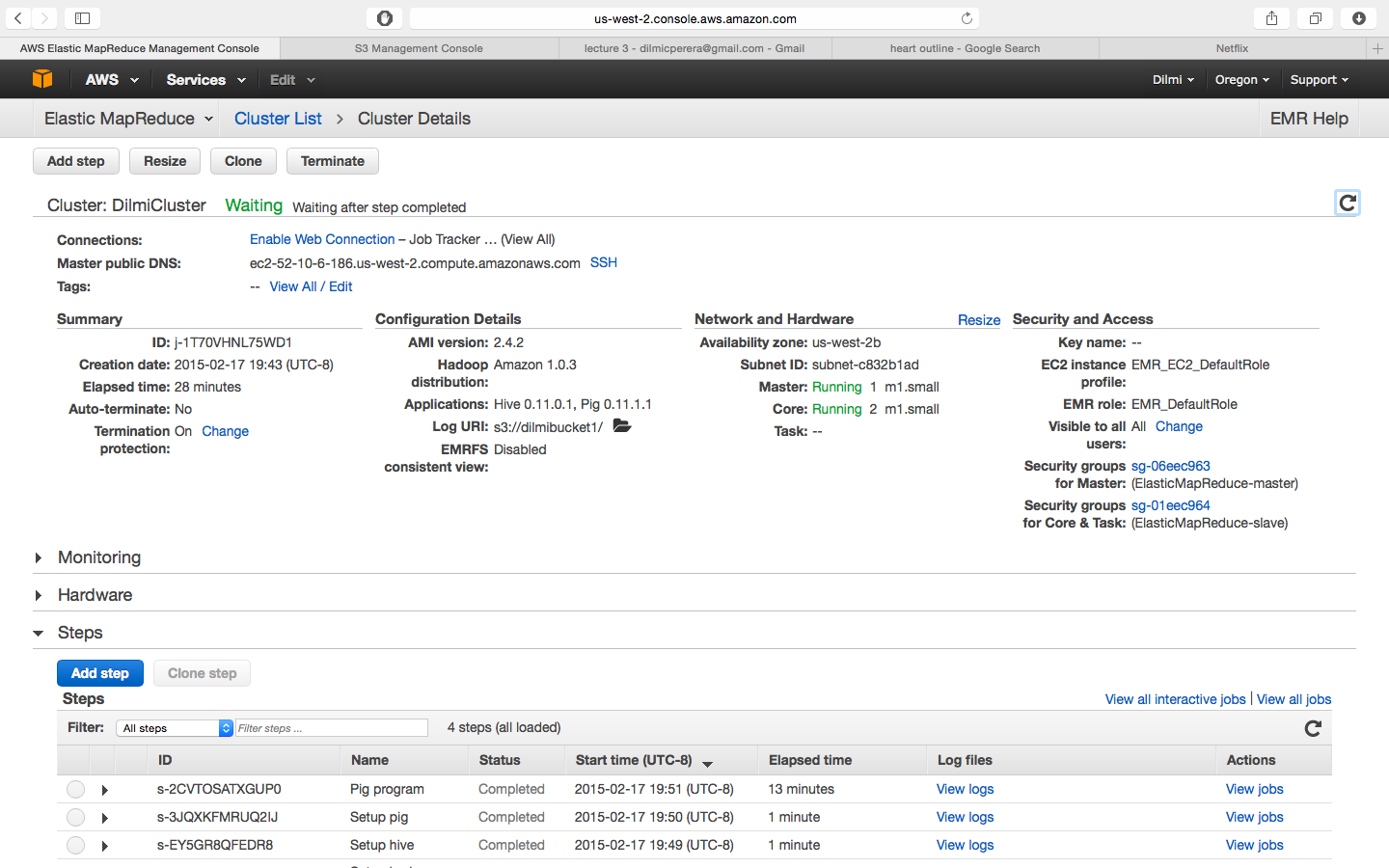




Starting and running cluster







Results

