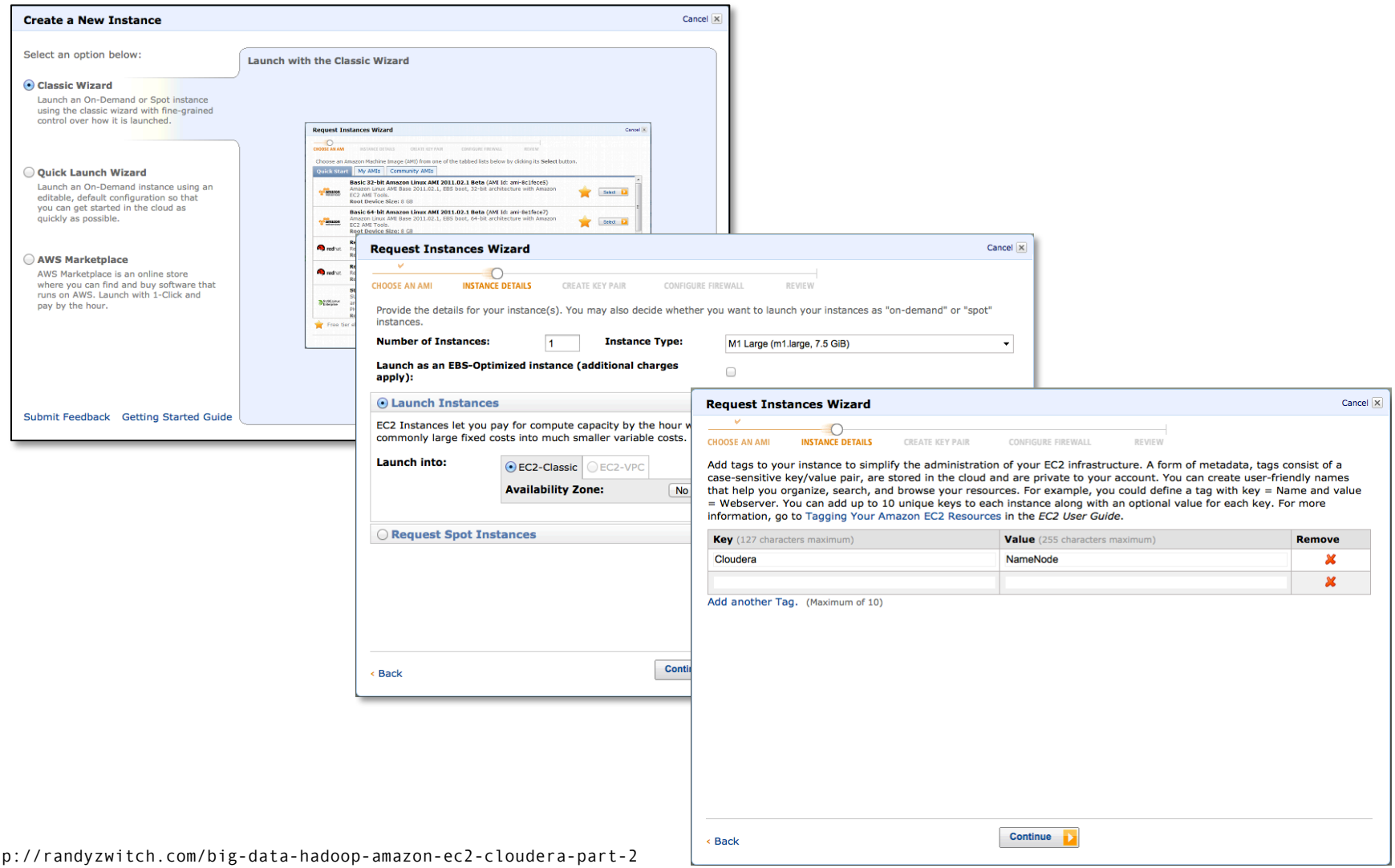
Building a Hadoop Cluster on Amazon EC2 using Cloudera

April 2013

http://randyzwitch.com/big-data-hadoop-amazon-ec2-cloudera-part-2

Create Name Node: m1.large

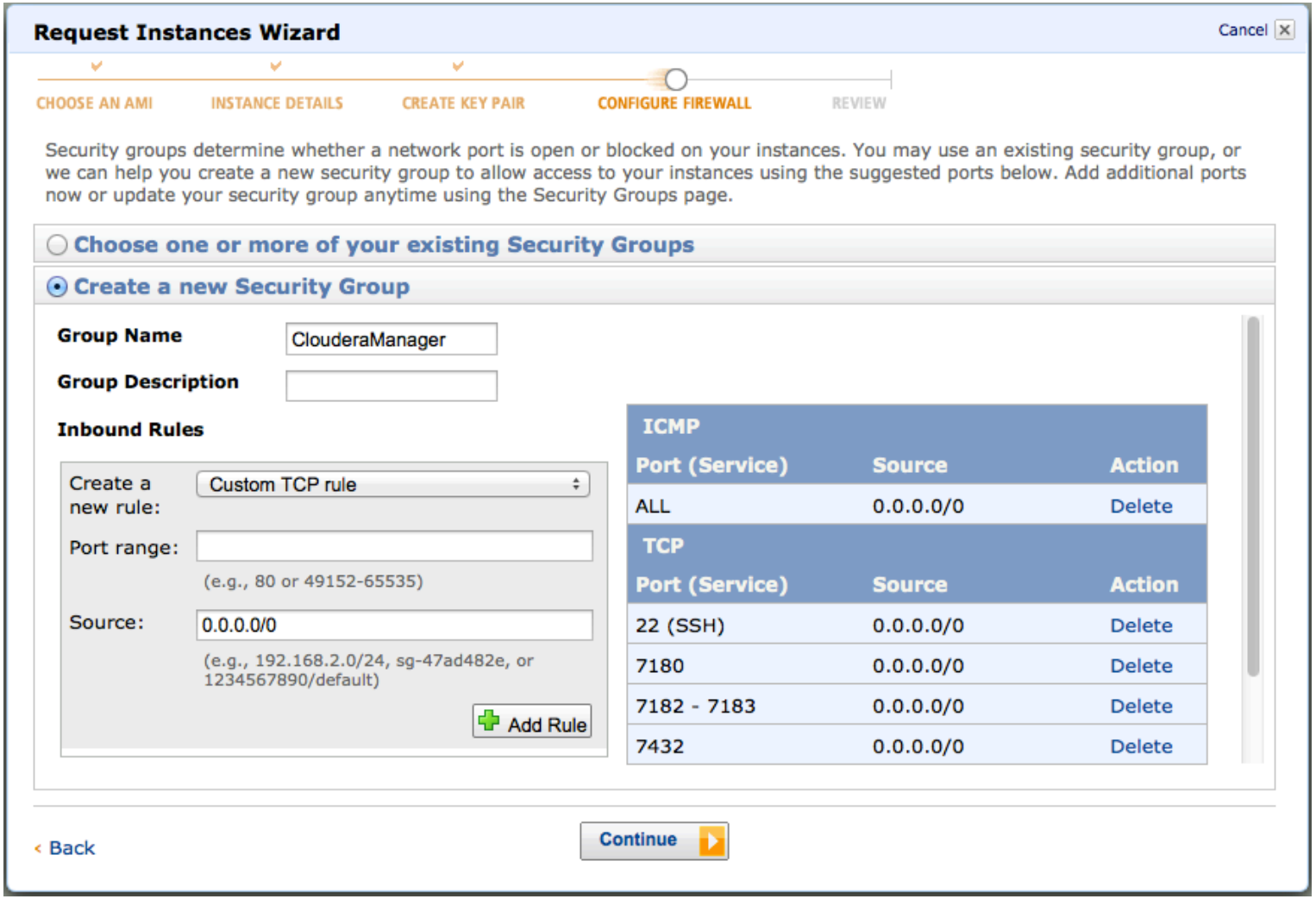
Ubuntu Server 12.04.1 LTS 64-bit



http://randyzwitch.com/big-data-hadoop-amazon-ec2-cloudera-part-2

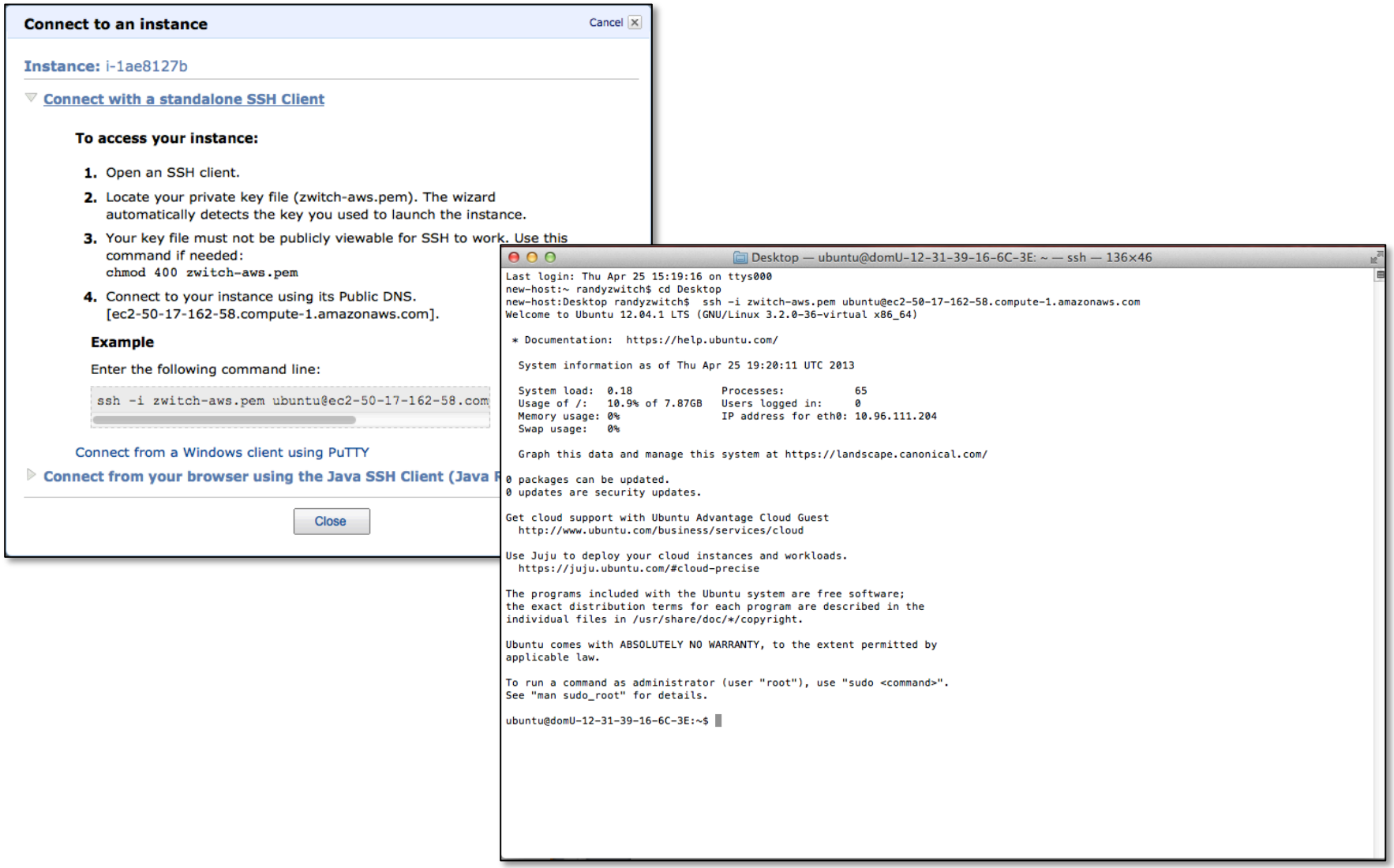
Create Name Node: m1.large

Can use defaults for most Wizard screens except Firewall



http://randyzwitch.com/big-data-hadoop-amazon-ec2-cloudera-part-2

Launch Instance, Connect Via SSH



http://randyzwitch.com/big-data-hadoop-amazon-ec2-cloudera-part-2

Download & Run Cloudera Manager

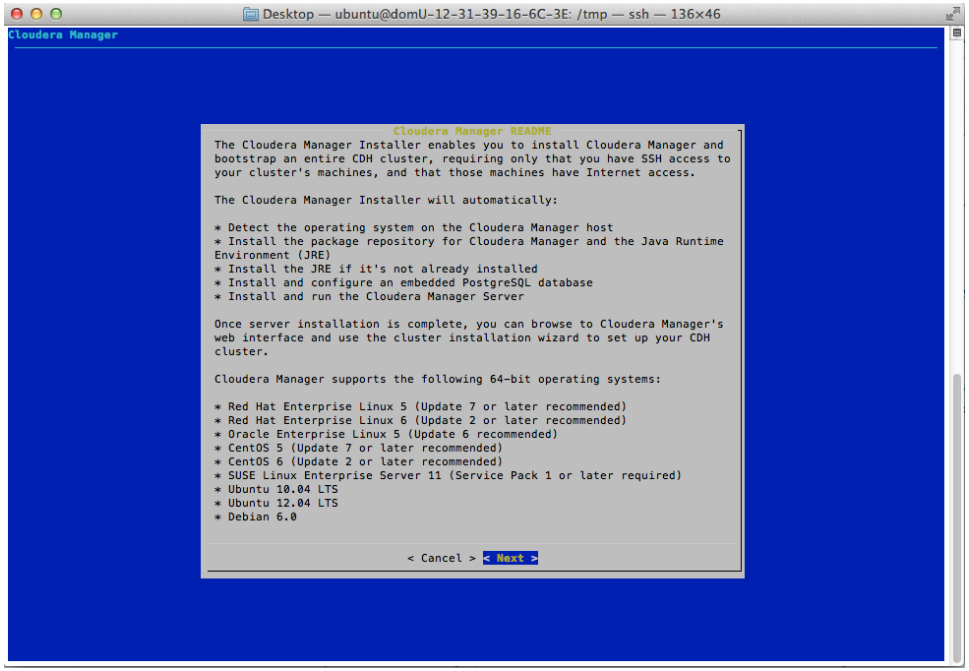
Depending on settings, might need to run as “sudo”

```
ubuntu@domU-12-31-39-16-6C-3E:~$ cd /tmp
ubuntu@domU-12-31-39-16-6C-3E:/tmp$ wget http://archive.cloudera.com/cm4/installer/latest/cloudera-manager-installer.bin
--2013-04-25 19:24:44--  http://archive.cloudera.com/cm4/installer/latest/cloudera-manager-installer.bin
Resolving archive.cloudera.com (archive.cloudera.com)... 184.73.217.71
Connecting to archive.cloudera.com (archive.cloudera.com)|184.73.217.71|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 501701 (490K) [application/octet-stream]
Saving to: `cloudera-manager-installer.bin'

100%[======================================================================================>] 501,701     --.-K/s   in 0.02s

2013-04-25 19:24:44 (19.9 MB/s) - `cloudera-manager-installer.bin' saved [501701/501701]

ubuntu@domU-12-31-39-16-6C-3E:/tmp$ chmod +x cloudera-manager-installer.bin
ubuntu@domU-12-31-39-16-6C-3E:/tmp$ ./cloudera-manager-installer.bin
```



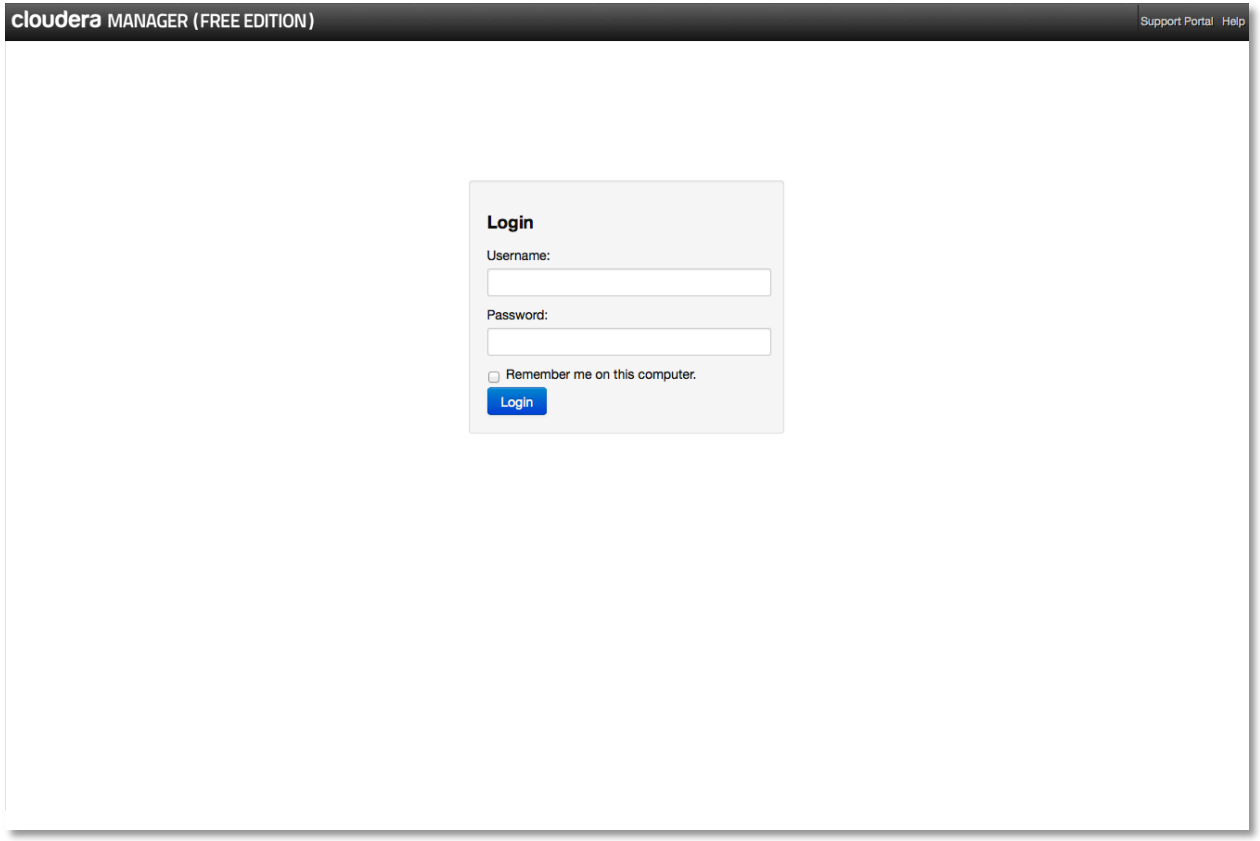
Might take 5 or so minutes to go through licensing and installation menus

http://randyzwitch.com/big-data-hadoop-amazon-ec2-cloudera-part-2

Open browser And Go To EC2 Instance Public DNS At Port 7180

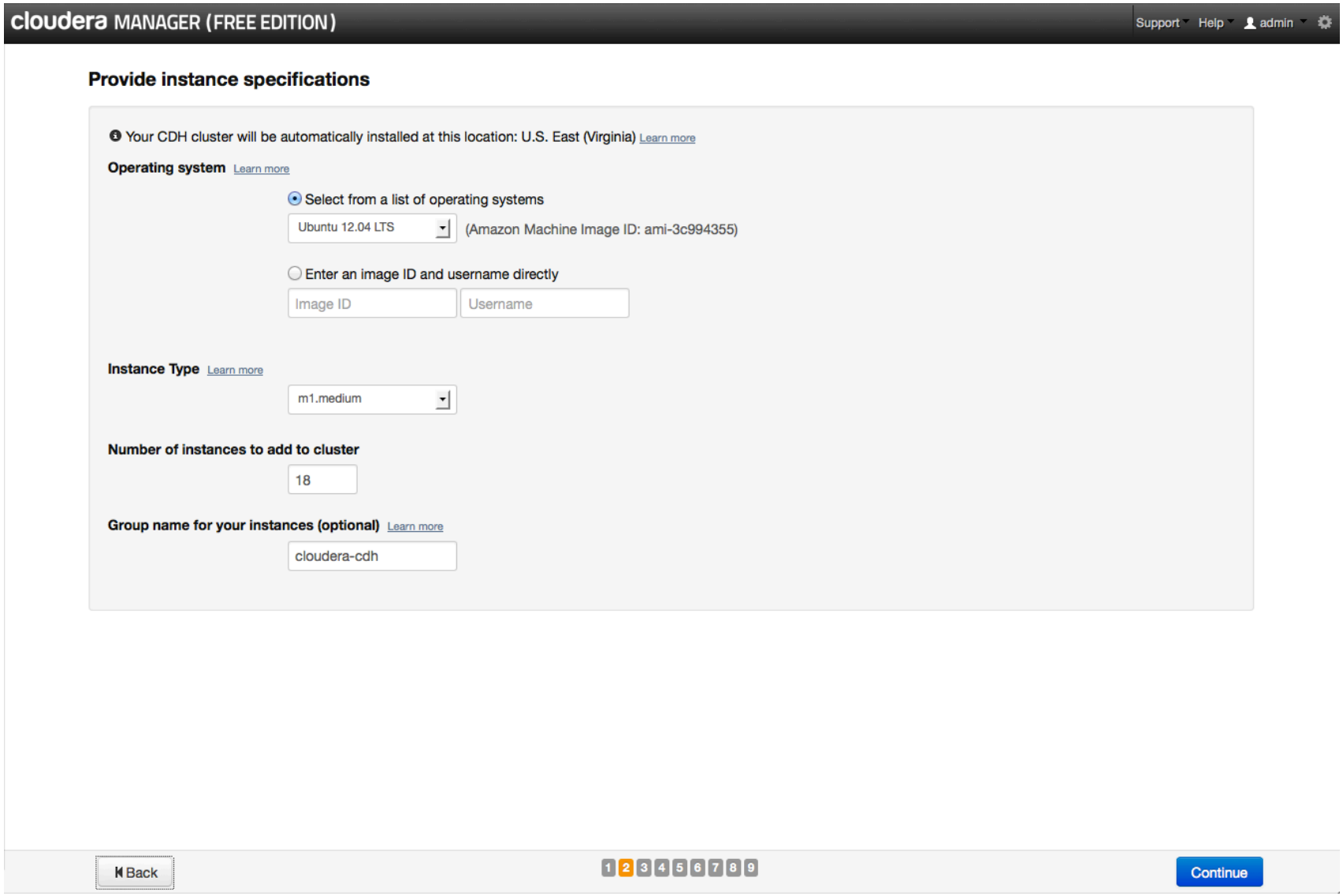
Ex: *http://ec2-50-17-162-58.compute-1.amazonaws.com:7180*

On this first login, you set your username and password credentials



http://randyzwitch.com/big-data-hadoop-amazon-ec2-cloudera-part-2

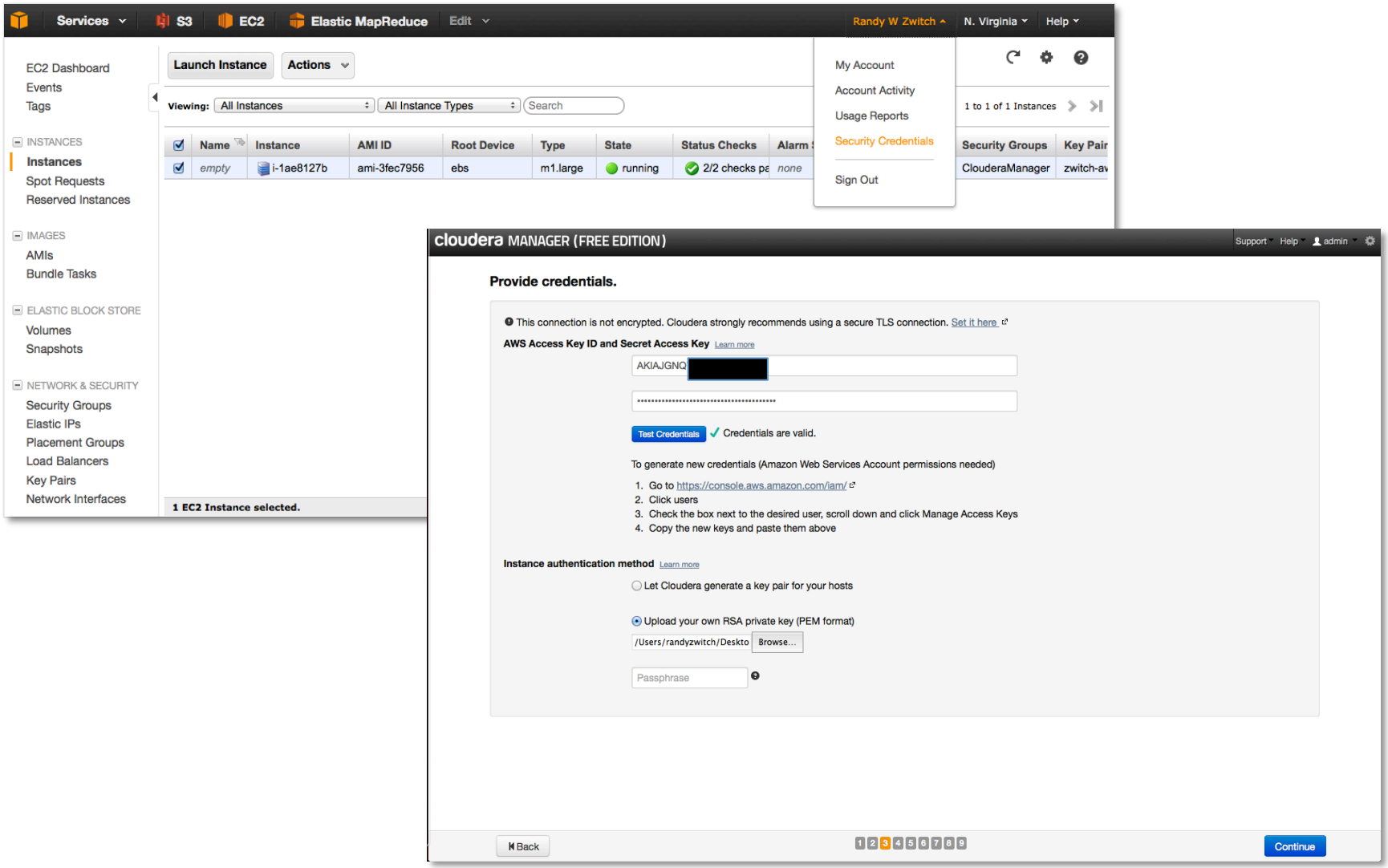
Use Defaults,‘18’ for instances



http://randyzwitch.com/big-data-hadoop-amazon-ec2-cloudera-part-2

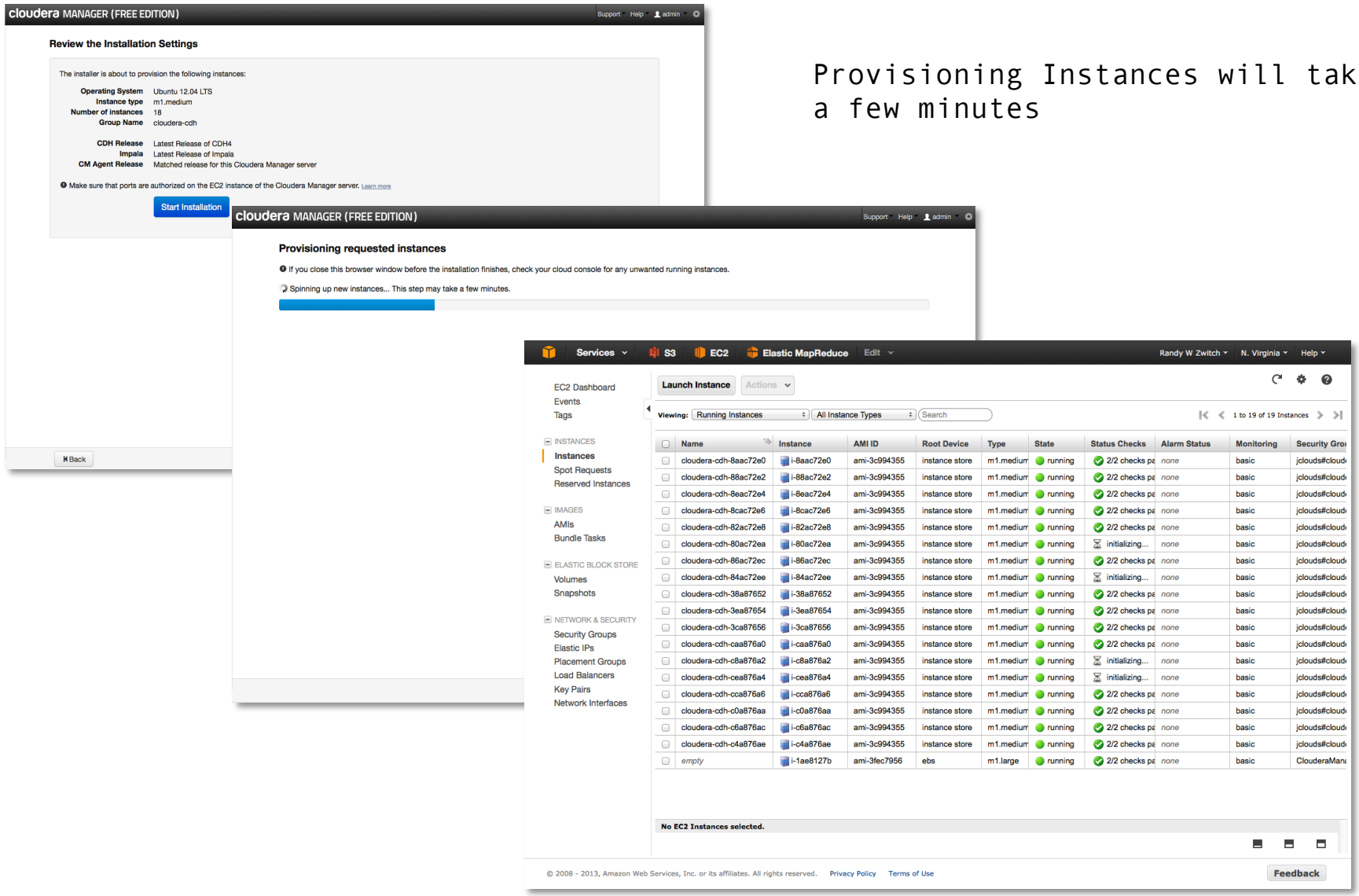
Type In AWS Access Key ID & Secret Access Key

Credentials can be found under “Security Credentials” in EC2 dashboard



http://randyzwitch.com/big-data-hadoop-amazon-ec2-cloudera-part-2

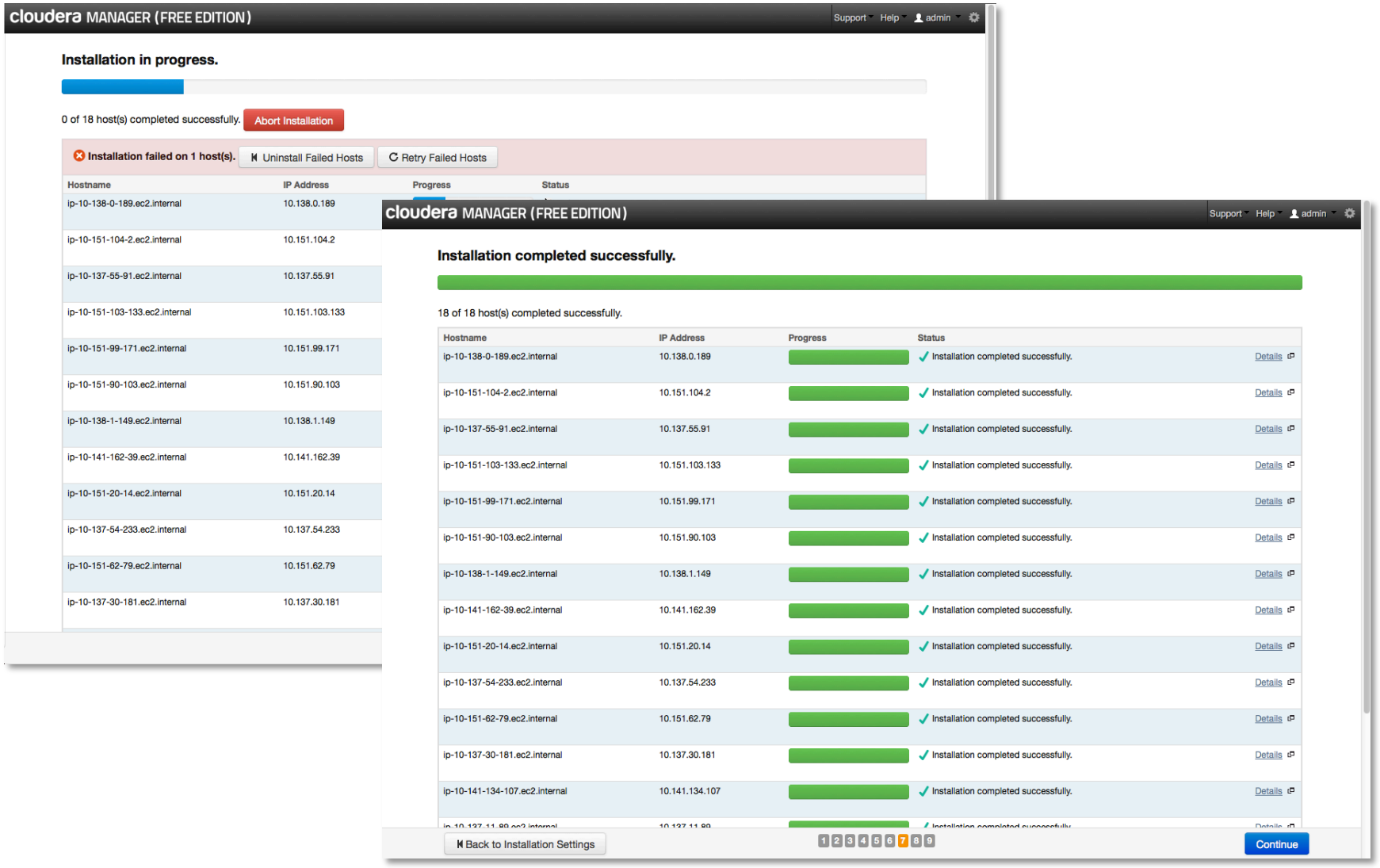
Review Settings Then Install!



Provisioning Instances will take a few minutes

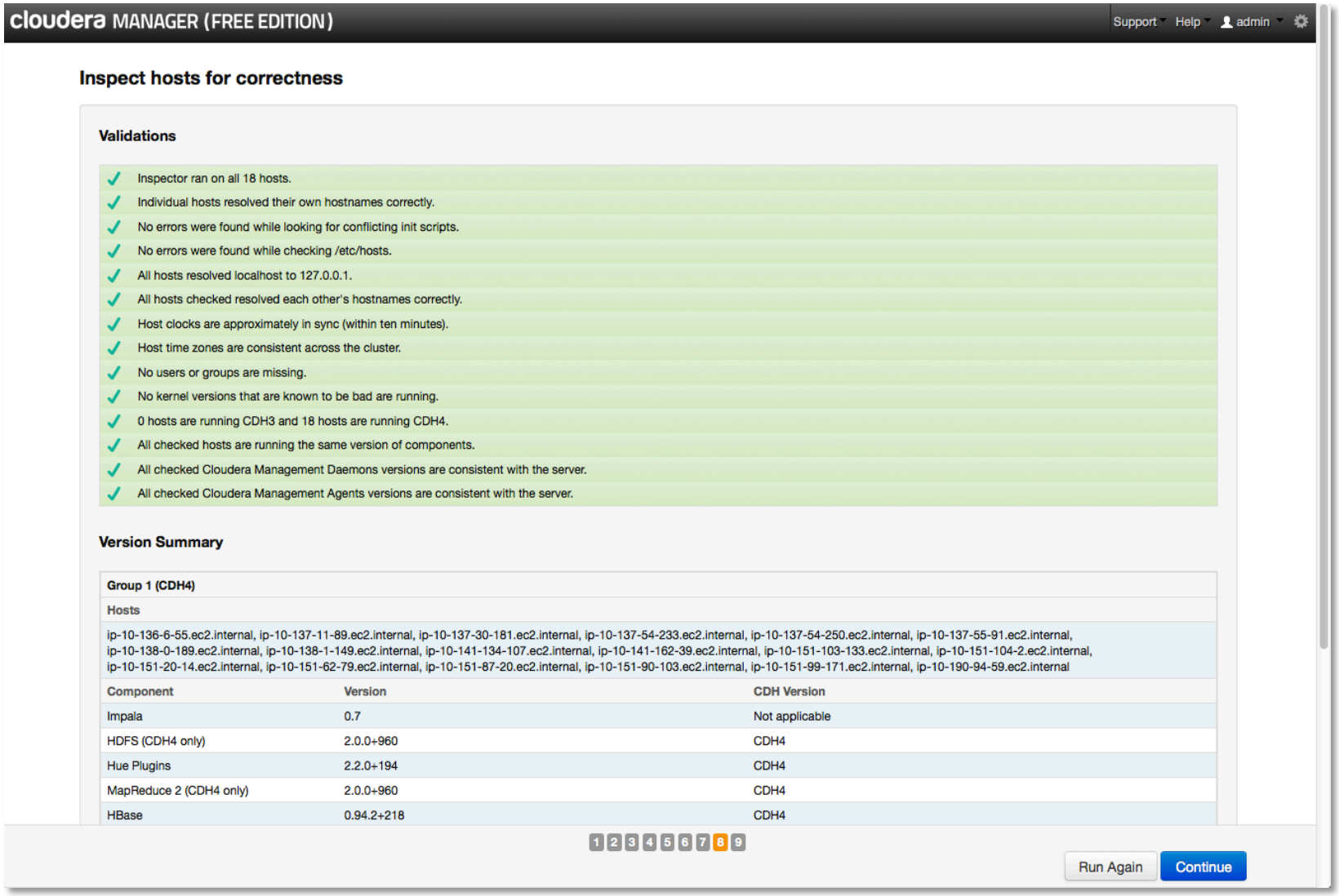
http://randyzwitch.com/big-data-hadoop-amazon-ec2-cloudera-part-2

If Any Installations Fail, Retry Until Success



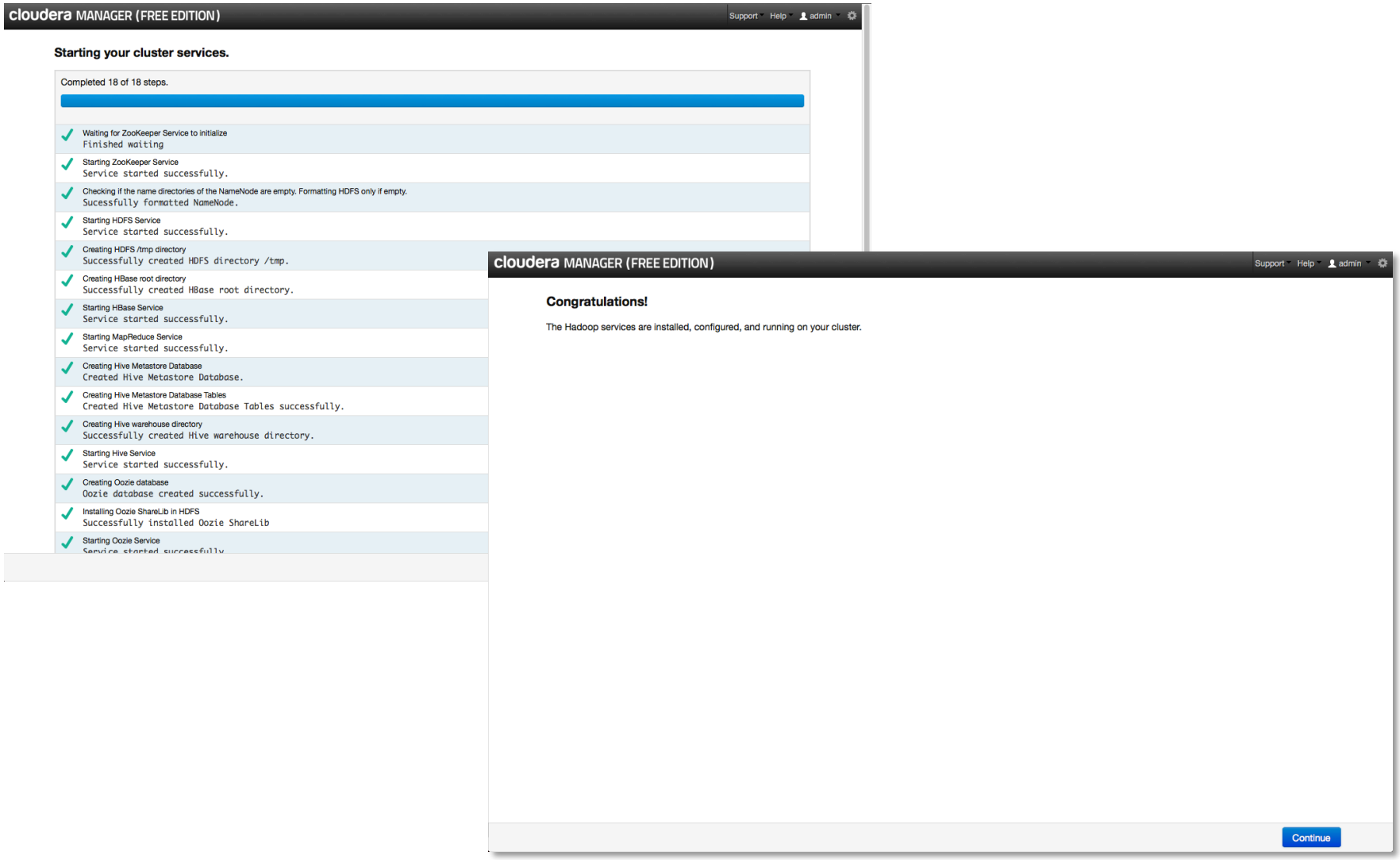
http://randyzwitch.com/big-data-hadoop-amazon-ec2-cloudera-part-2

Make Sure Consistency Check Passes



http://randyzwitch.com/big-data-hadoop-amazon-ec2-cloudera-part-2

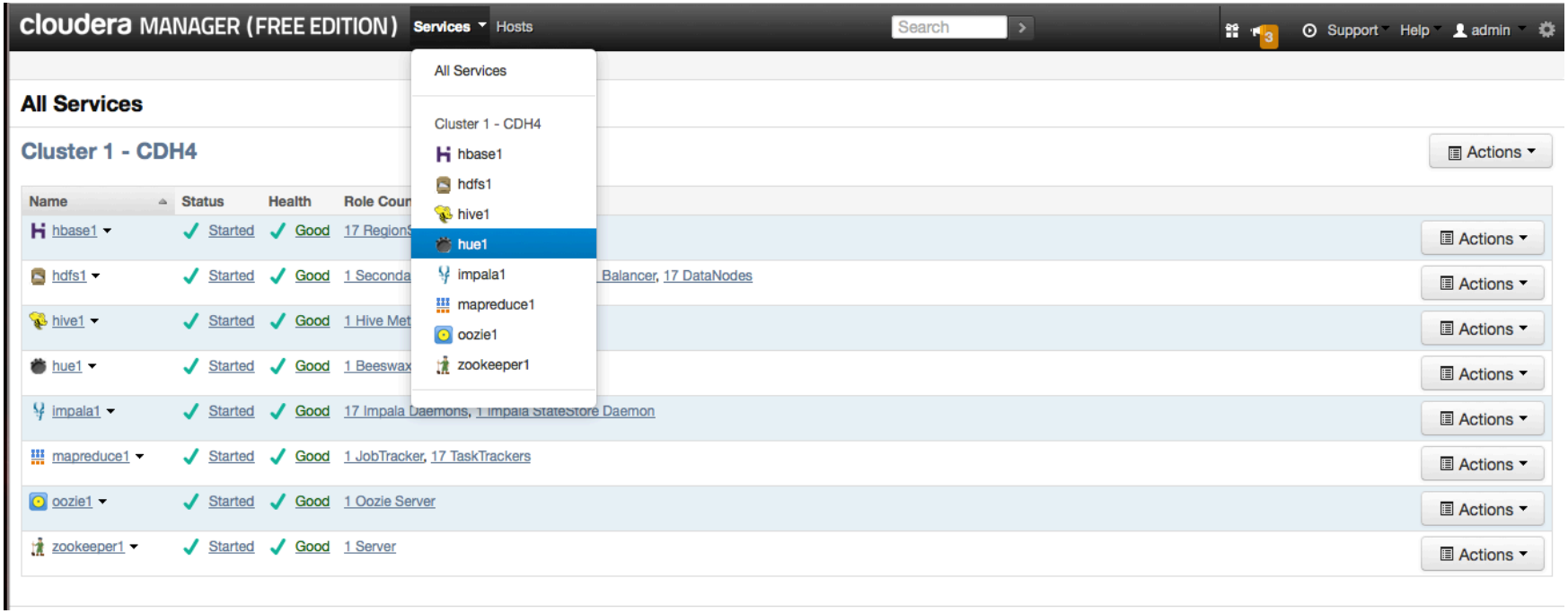
Cluster Services Will Start, Then Success!



http://randyzwitch.com/big-data-hadoop-amazon-ec2-cloudera-part-2

Finding Hue Public DNS

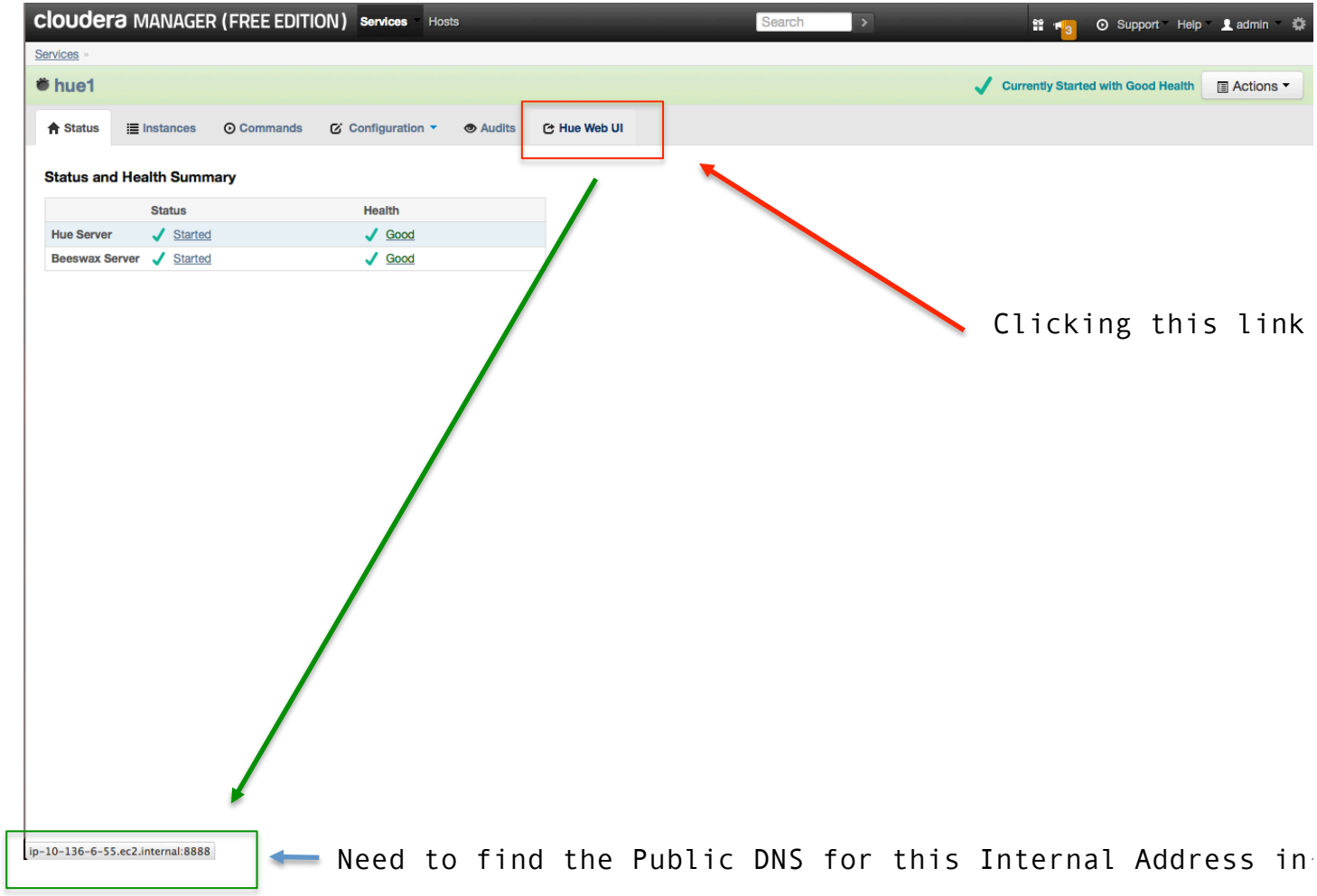
Hue (Hadoop User Experience) is the more user-friendly way to interact with Hadoop



http://randyzwitch.com/big-data-hadoop-amazon-ec2-cloudera-part-2

Finding Hue Public DNS

Clicking on the “Hue Web UI” button doesn’t work, because it references the Internal Address for Amazon EC2



Clicking this link button won’t work!

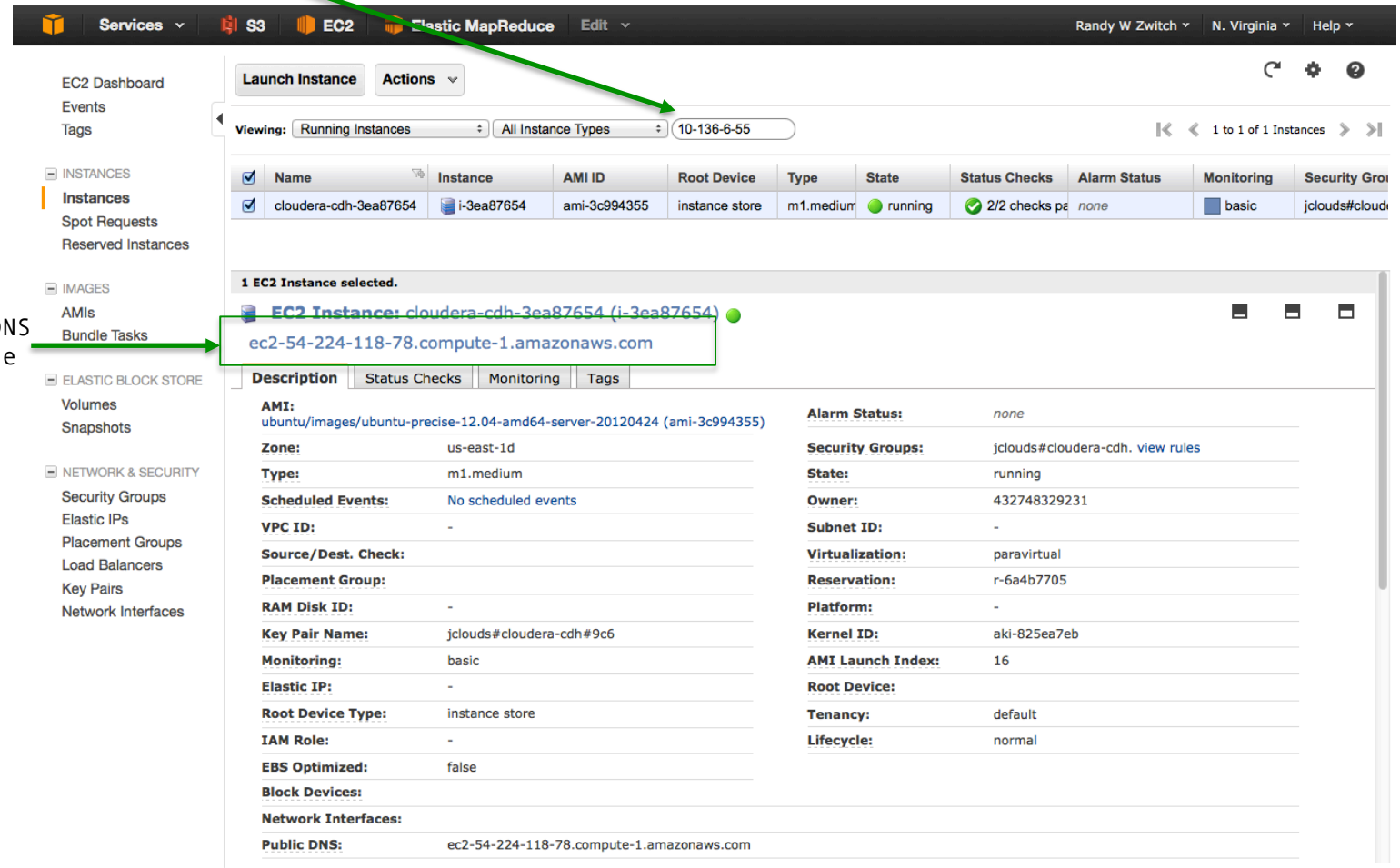
Need to find the Public DNS for this Internal Address in Amazon Dashboard

http://randyzwitch.com/big-data-hadoop-amazon-ec2-cloudera-part-2

Finding Hue Public DNS

Type in Internal Address in Search Box to find the Instance having Hue

ip-10-136-6-55.ec2.internal:8888



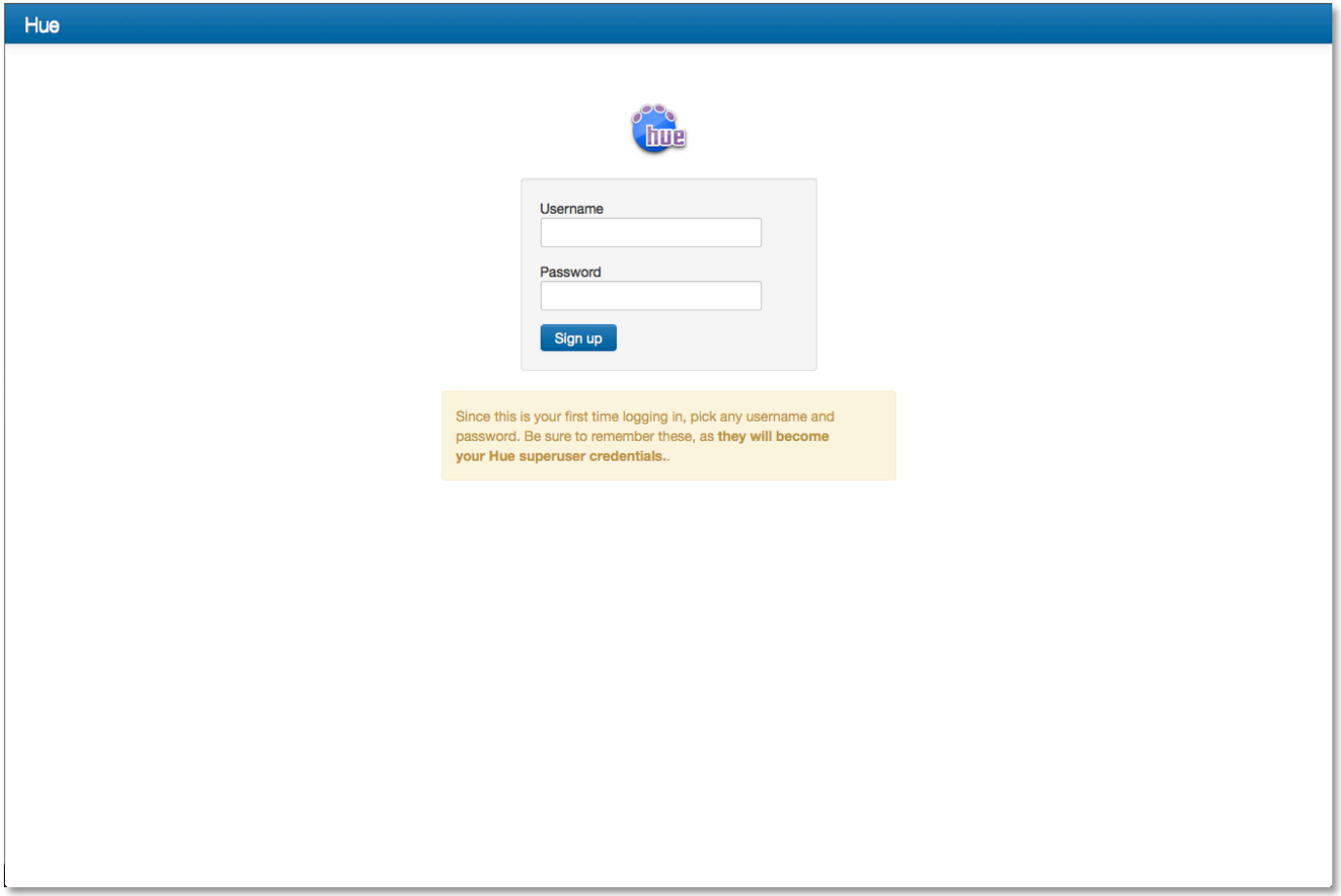
This is the public DNS Address to access Hue

http://randyzwitch.com/big-data-hadoop-amazon-ec2-cloudera-part-2

Finding Hue Public DNS

Hue is accessed via Port 8888
Ex: *http://ec2-54-224-118-78.compute-1.amazonaws.com:8888*

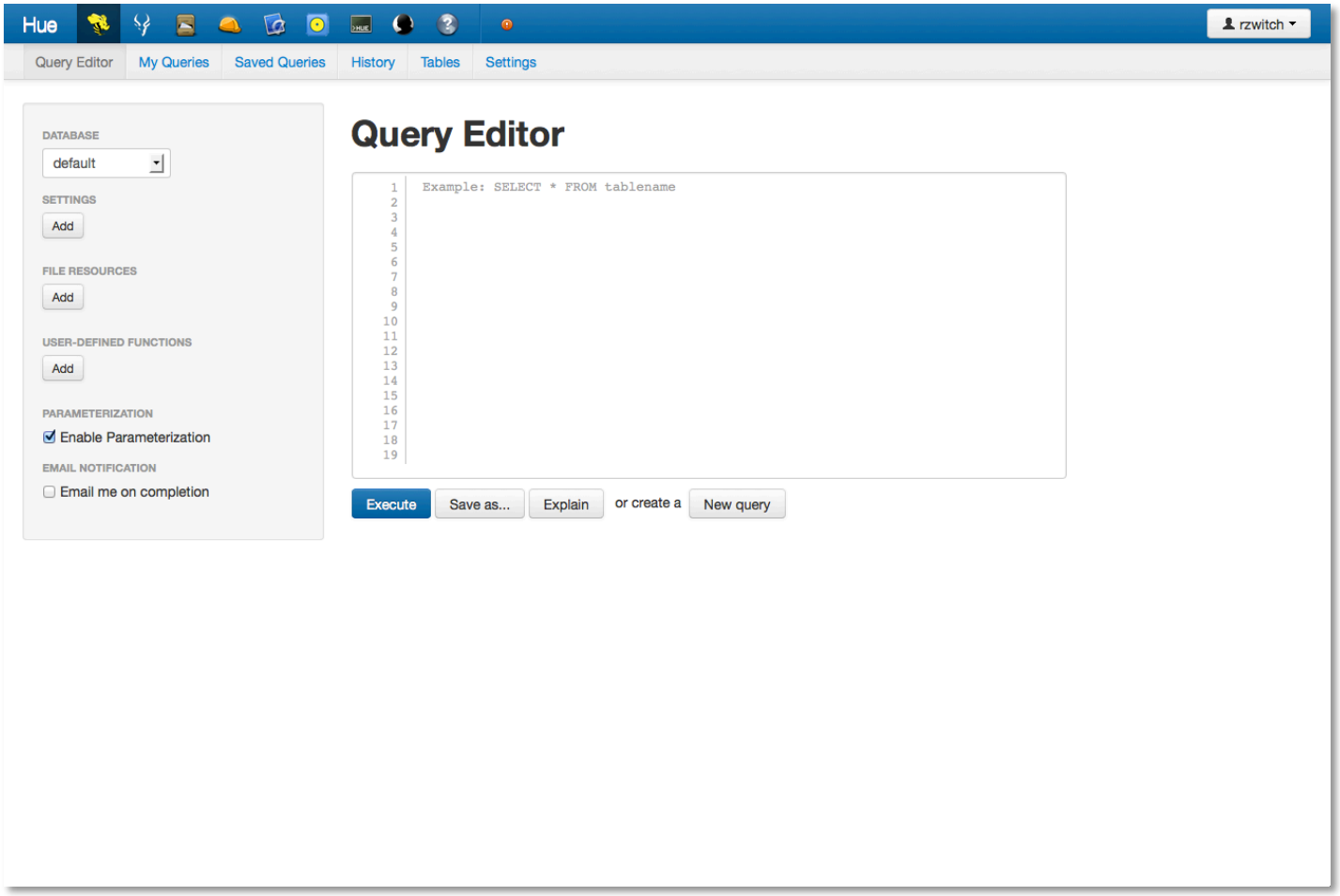
Pick your username/password carefully, this is the superuser



http://randyzwitch.com/big-data-hadoop-amazon-ec2-cloudera-part-2

If You See Hue, You're Ready For Analysis!

This is the Hive editor, which allows for SQL-Like Syntax to create MapReduce jobs



http://randyzwitch.com/big-data-hadoop-amazon-ec2-cloudera-part-2

Reference

http://blog.cloudera.com/blog/2013/03/how-to-create-a-cdh-cluster-on-amazon-ec2-via-cloudera-manager/