

Guide to Working with AWS CLI

To Start AWS CLI:

1. Navigate to the folder, Kingas-MacBook-Pro:AmazonEMRandRHadoop Kinga, then enter Conda environment, amazonEnv:

```
Kingas...Kinga$ source activate amazonEnv
```

2. Activate the virtual environment:

```
(amazonEnv) Kingas...Hadoop Kinga$ source ~/cli-ve/bin/activate
```

To Exit AWS CLI:

1. To exit Virtualenv:

```
(cli-ve) (amazonEnv) Kingas...Hadoop Kinga$ deactivate
```

2. To exit the Conda environment, amazonEnv:

```
(amazonEnv) Kingas...RHadoop Kinga$ source deactivate
```

Using AWS CLI

Checking the installation:

```
(cli-ve) (amazonEnv) Kingas...RHadoop Kinga$ aws --version
```

To configure AWS CLI:

```
(amazonEnv) Kingas...RHadoop Kinga$ aws configure
```

The AWS CLI will prompt you for four pieces of information

```
AWS Access Key ID [None]: see IAMaccessKeys.csv
```

```
Secret Access Key [None]: see IAMaccessKeys.csv
```

```
Default region name [None]: us-west-2
```

```
Default output format [None]: json
```

To update any of the settings, simply run **aws configure** again and enter new values as appropriate,

The CLI stores credentials specified with **aws configure** in a local file named **credentials** in a folder named **.aws** in your home directory. To look at the file:

```
(cli-ve) (amazonEnv) Kingas...RHadoop Kinga$ ls ~/.aws
```

```
$ aws ec2 create-key-pair --key-name theKey --query  
'KeyMaterial' --output text > theKey.pem
```

```
$ aws emr create-cluster --applications Name=Hadoop Name=Spark  
Name=Hive Name=Pig Name=Tez Name=Ganglia \  
--release-label emr-5.2.0 --name "EMR 5.2.0 RStudio +  
sparklyr" --service-role EMR_DefaultRole \  
--instance-groups  
InstanceGroupType=MASTER,InstanceCount=1,InstanceType=m  
3.2xlarge \  
InstanceGroupType=CORE,InstanceCount=5,InstanceType=m3.  
2xlarge --bootstrap-actions \  
Path=s3://aws-bigdata-blog/artifacts/aws-blog-emr-  
rstudio-sparklyr/rstudio_sparklyr_emr5.sh,\  
Args=["--rstudio","--sparkr","--rexamples","--  
plyrmr","--rhdfs","--sparklyr"],\  
Name="Install RStudio" --ec2-attributes  
InstanceProfile=EMR_EC2_DefaultRole,KeyName=theKey \  
--configurations  
'[{"Classification":"spark","Properties":  
{"maximizeResourceAllocation":"true"}}]' \  
--region us-west-2
```

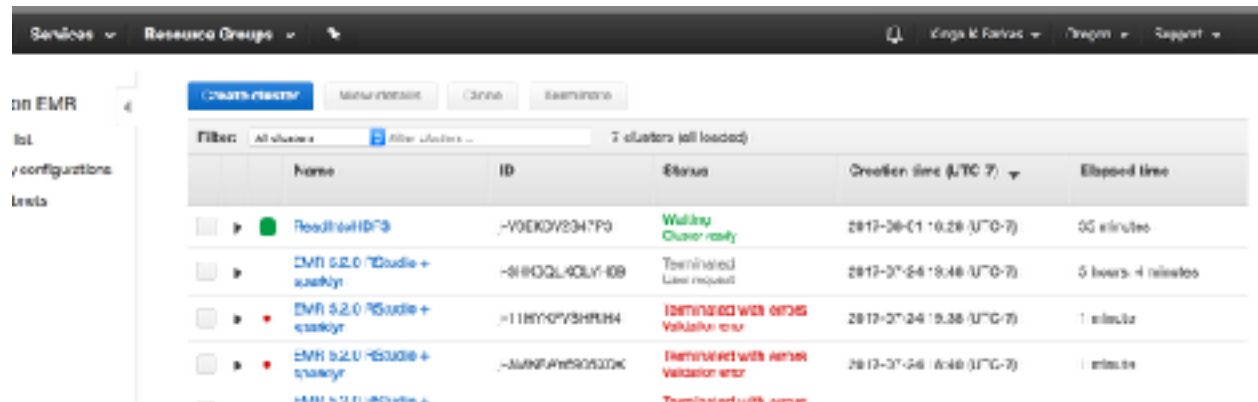
```
aws emr socks --cluster-id j-3HK6QLKOLYH09 --key-pair-file  
theKey.pem
```

To connect to R-studio go to the link:

```
http://ec2-54-187-74-95.us-west-2.compute.amazonaws.com:8787
```

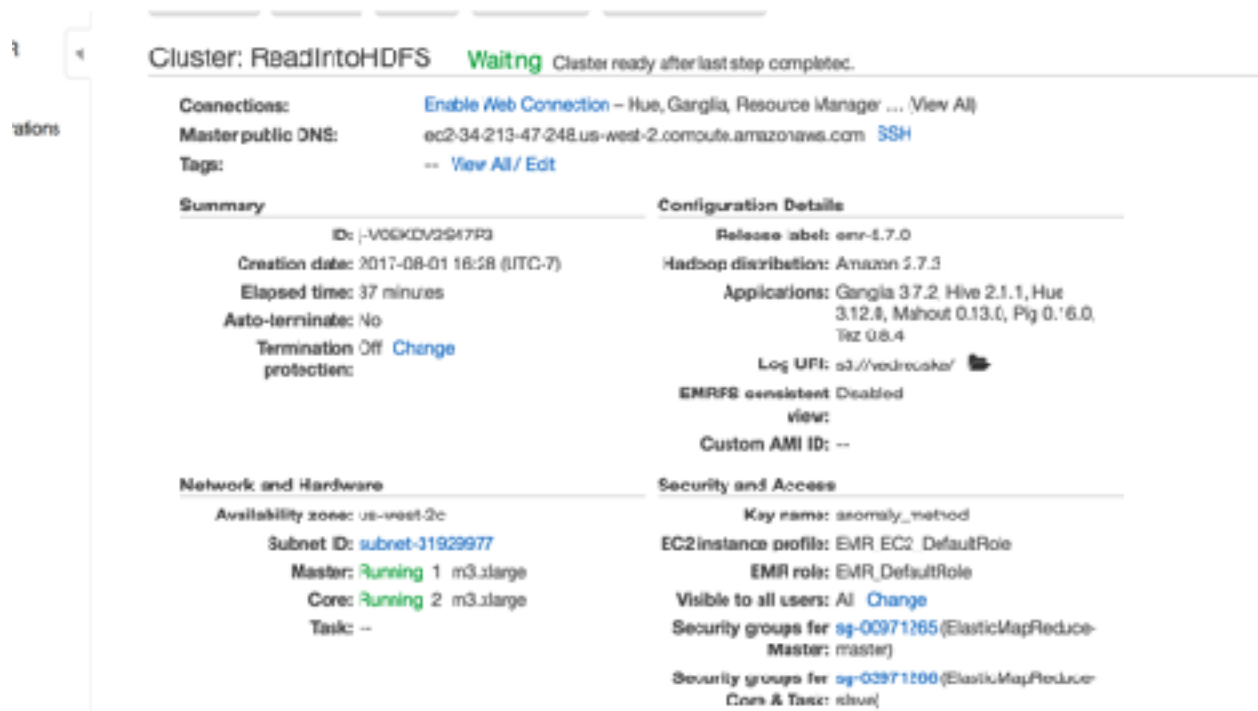
where <http://ec2-54-187-74-95.us-west-2.compute.amazonaws.com> is the Master Public DNS of your cluster. It can be found using AWS EMR console:

1. Go to the Cluster List



| | Name | ID | Status | Creation time (UTC-7) | Elapsed time |
|--------------------------|--|-----------------|--|--------------------------|--------------------|
| <input type="checkbox"/> | ReadIntoHDFS | j-V06KDV2S47P3 | Waiting Cluster ready | 2017-08-01 16:28 (UTC-7) | 0G minutes |
| <input type="checkbox"/> | EMR 5.2.0 Hadoop + auxsh | j-4W4OQL4OLY40B | Terminated Last request | 2017-07-24 19:48 (UTC-7) | 5 hours, 4 minutes |
| <input type="checkbox"/> | EMR 5.2.0 Hadoop + auxsh | j-11BY0PVS4PH4 | Terminated with errors Validation error | 2017-07-24 19:38 (UTC-7) | 1 minute |
| <input type="checkbox"/> | EMR 5.2.0 Hadoop + auxsh | j-MJNFAH620530K | Terminated with errors Validation error | 2017-07-24 19:48 (UTC-7) | 1 minute |

2. Select the cluster you are working on to get:



Cluster: ReadIntoHDFS Waiting Cluster ready after last step completed.

Connections: [Enable Web Connection](#) - Hue, Ganglia, Resource Manager ... [View All](#)

Master public DNS: ec2-54-213-47-248.us-west-2.compute.amazonaws.com [SSH](#)

Tags: -- [View All/ Edit](#)

Summary

ID: j-V06KDV2S47P3

Creation date: 2017-08-01 16:28 (UTC-7)

Elapsed time: 37 minutes

Auto-terminate: No

Termination protection: Off [Change](#)

Configuration Details

Release label: emr-5.7.0

Hadoop distribution: Amazon 2.7.3

Applications: Ganglia 3.7.2, Hive 2.1.1, Hue 3.12.8, Mahout 0.13.0, Pig 0.16.0, Tez 0.8.4

Log URI: s3://youreuser/ [View](#)

EMRFS consistent view: Disabled

Custom AMI ID: --

Network and Hardware

Availability zone: us-west-2c

Subnet ID: [subnet-31929977](#)

Master: Running 1 m3.large

Core: Running 2 m3.large

Task: --

Security and Access

Key name: anomaly_method

EC2 instance profile: EMR_EC2_DefaultRole

EMR role: EMR_DefaultRole

Visible to all users: All [Change](#)

Security groups for [sg-00971265](#) (ElasticMapReduce-Master): master

Security groups for [sg-03971200](#) (ElasticMapReduce-Core & Task): shue

3. The Master Public DNS is the second line in the table.

Important:

The login and password are both **hadoop**, unless you specified something different.