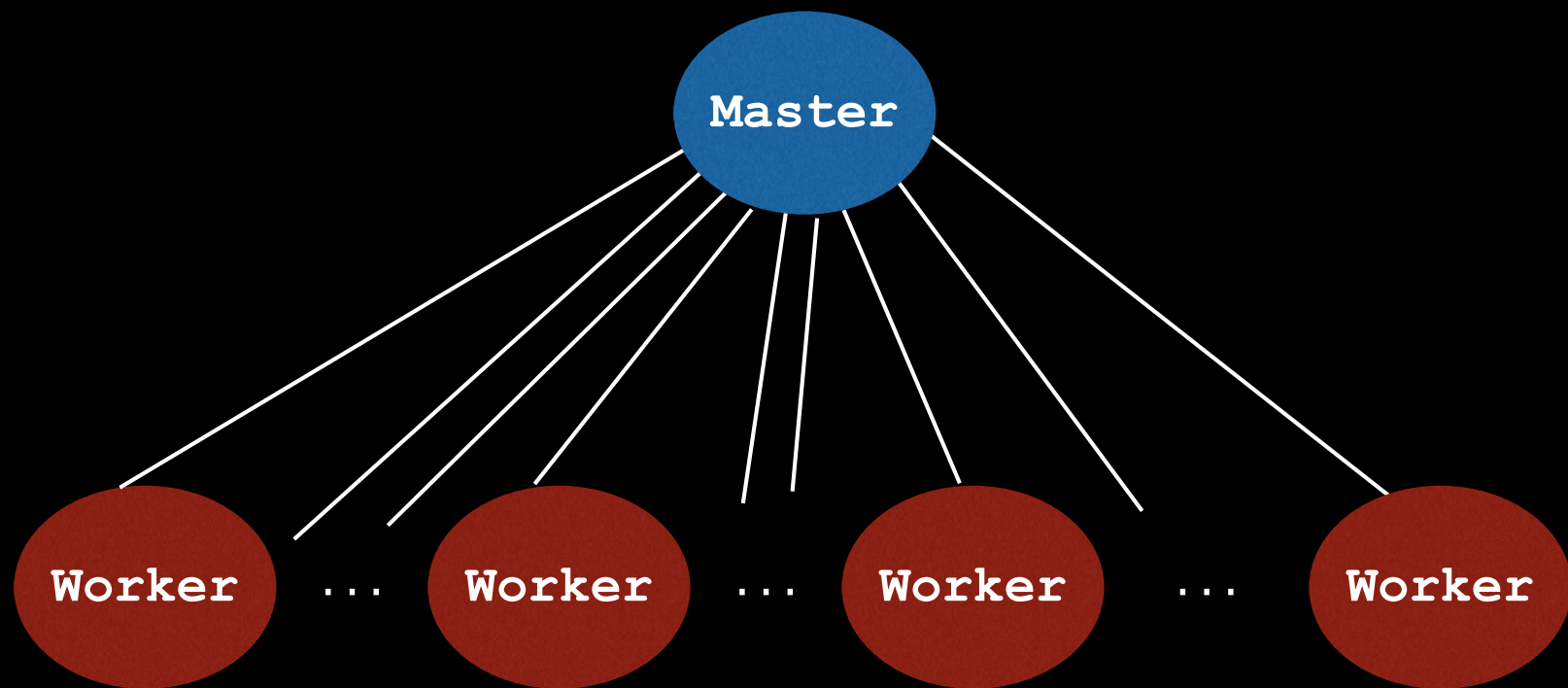


Spark on AWS

EC2 Spark Cluster



Command Line Tool

`spark-ec2`

- A tool to start an AWS cluster on command line
- Using Boto behind the scenes

Launch Cluster

```
./spark-ec2 -k keyname \
```

```
-i ~/.ssh/keyname.pem \
```

```
--copy-aws-credentials \
```

**Copy Access / Secret Key
to AWS**

```
-instance-type=m1.large \
```

Worker instance type

```
-m m1.large \
```

Master instance type

```
-s 19
```

19 workers

```
launch spark
```

Launch a Cluster

- Install Spark on the Master and Workers
- Connecting Master to Workers

Sign in to Cluster

```
./spark-ec2 -k keyname -i ~/.ssh/  
keyname.pem login spark
```

- Equivalent to SSH

Stop/ Terminate/ Restart Cluster

- You can stop a cluster without terminating
- Quicker to restart the cluster
- Instead of re-launching
- Storage costs from AWS still applies

Memory Allocation

- `IPYTHON_OPTS="notebook -ip=0.0.0.0"\
/root/spark/bin/pyspark
--executor-memory 4G
--driver-memory 4G`



Spark Master at spark://ec2-52-5-18-219.

URL: spark://ec2-52-5-18-219.compute-1.amazonaws.com:7077

Workers: 19

Cores: 38 Total, 38 Used

Memory: 119.3 GB Total, 76.0 GB Used

Applications: 1 Running, 5 Completed

Drivers: 0 Running, 0 Completed

Status: ALIVE

**Important to use
all/ most memory**

Workers

Id	Address
worker-20150605031608-ip-172-31-54-213.ec2.internal-53771	ip-172-31
worker-20150605031608-ip-172-31-54-214.ec2.internal-38195	ip-172-31
worker-20150605031608-ip-172-31-54-215.ec2.internal-39474	ip-172-31
worker-20150605031608-ip-172-31-54-216.ec2.internal-42892	ip-172-31
worker-20150605031608-ip-172-31-54-217.ec2.internal-57467	ip-172-31
worker-20150605031608-ip-172-31-54-218.ec2.internal-61000	ip-172-31

Running Applications

ID	Name	Cores	Memory per Node
app-20150605072059-0005	PySparkShell	38	4.0 GB

Completed Applications

ID	Name	Cores	Memory per Node
app-20150605061056-0004	PySparkShell	18	4.0 GB
app-20150605055312-0003	PySparkShell	0	4.0 GB
app-20150605035717-0002	PySparkShell	38	5.0 GB

**Make sure only
1 Application is running**

WIKIPEDIA

English

The Free Encyclopedia

4 675 000+ articles

Русский

Свободная энциклопедия

1 172 000+ статей

Español

La enciclopedia libre

1 145 000+ artículos

Deutsch

Die freie Enzyklopädie

1 789 000+ Artikel

中文

自由的百科全書

801 000+ 條目

Português

A enciclopédia livre

857 000+ artigos

日本語

フリー百科事典

938 000+ 記事

Français

L'encyclopédie libre

1 573 000+ articles

Italiano

L'enciclopedia libera

1 164 000+ voci

Polski

Wolna encyklopedia

1 082 000+ haseł



[s3n://jyt109/wiki_articles](https://jyt109/wiki_articles)