

EMR Report for K-Means cluster (Project 3 - EECS 219)

Abhishek Madav Student # 86378148

The applications has been tested on EMR cluster with the settings of a regular cluster and the default HBase version available over the AWS. The cluster had been created, with custom jar settings as follows:-

Cluster: K-Means Cluster **Starting** Provisioning Amazon EC2 capacity

Connections: --
Master public DNS: --
Tags: -- [View All / Edit](#)

Summary	Configuration Details	Network and Hardware	Security and Access
ID: j-201SSNQ5AJD9M Creation date: 2015-06-03 11:20 (UTC-7) Elapsed time: 3 minutes Auto-terminate: No Termination On Change protection: Change	AMI version: 3.7.0 Hadoop distribution: Amazon 2.4.0 Applications: HBase 0.94.18 Log URI: s3://aws-logs-224362316408-us-west-2/elasticmapreduce/ EMRFS consistent view: Disabled	Availability zone: us-west-2b Subnet ID: subnet-17f6072 Master: Provisioning 1 m3.xlarge Core: Provisioning 2 m3.xlarge Task: --	Key name: Cluster-key EC2 instance profile: EMR_EC2_DefaultRole EMR role: EMR_DefaultRole Visible to all users: All Change Security groups for sg-4a435e2f (ElasticMapReduce-Master: master) Security groups for sg-55435e30 (ElasticMapReduce-Core & Task: slave)

Monitoring
Hardware
Steps

[Add step](#) [Clone step](#)

Steps [View all interactive jobs](#) [View all jobs](#)

Filter: All steps [Filter steps...](#) 3 steps (all loaded)

ID	Name	Status	Start time (UTC-7)	Elapsed time	Log files	Actions
s-3B189GFSJA6U7	K-Means	Pending			View logs	View jobs
s-1O48QSY1KGZAF	Setup hadoop debugging	Pending			View logs	View jobs
s-1EN6US9309NMT	Start HBase	Pending			View logs	View jobs

Bootstrap Actions

Bootstrap actions are scripts that are executed during setup before Hadoop starts on every cluster node. You can use them to install

Add Step

Step type: Custom JAR

Name*

JAR location* JAR location maybe a path into S3 or a fully qualified java class in the classpath.

Arguments These are passed to the main function in the JAR. If the JAR does not specify a main class in its manifest file you can specify another class name as the first argument.

Action on failure: What to do if the step fails.

[Cancel](#) [Save](#)

[Configure and add](#)

Auto-terminate ☐ Yes ☒ No

Automatically terminate cluster after the last step is completed.
Keep cluster running until you terminate it.

The application input four parameters as required-

- 1) S3 location of the dataset.txt file.
- 2) Table name for the input dataset.
- 3) Table name for the center table.
- 4) K- the number of clusters.

After the application has been successfully completed running, an SSH connection to the AWS master was established.

```
hadoop@ip-172-31-30-201:~$ ssh hadoop@ec2-52-26-88-15.us-west-2.compute.amazonaws.com -i /home/anadav/Desktop/cl
user-key.pem
Last login: Wed Jun 3 18:26:31 2015 from ip72-211-200-213.oc.oc.cox.net

 _ _ | _ _ | _ _ |
 _ | ( _ _ | _ _ |
 _ | \ _ _ | _ _ |

Amazon Linux AMI

https://aws.amazon.com/amazon-linux-ami/2015.03-release-notes/
26 package(s) needed for security, out of 39 available
Run "sudo yum update" to apply all updates.

-----

Welcome to Amazon Elastic MapReduce running Hadoop and Amazon Linux.

Hadoop is installed in /home/hadoop. Log files are in /mnt/var/log/hadoop. Check
/mnt/var/log/hadoop/steps for diagnosing step failures.

The Hadoop UI can be accessed via the following commands:

ResourceManager      lynx http://ip-172-31-30-201:9026/
NameNode              lynx http://ip-172-31-30-201:9101/

-----

[hadoop@ip-172-31-30-201 ~]$
```

The master provides the setting for hbase shell prompt and the tables could be scanned for the values. Scan 'center' results into –

ROW	COLUMN+CELL
center0	column=Area:X1, timestamp=1433368639022, value=0.8900027054976737
center0	column=Area:X5, timestamp=1433368639022, value=569.623389874067
center0	column=Area:X6, timestamp=1433368639022, value=306.24973874932493
center0	column=Area:Y1, timestamp=1433368639022, value=131.68682556237104
center0	column=Area:Y2, timestamp=1433368639022, value=6.9999992115170775
center0	column=Property:X2, timestamp=1433368639022, value=3.499925657324421
center0	column=Property:X3, timestamp=1433368639022, value=0.23436730994482746
center0	column=Property:X4, timestamp=1433368639022, value=2.8124619369255712
center0	column=Property:X7, timestamp=1433368639022, value=28.34576414701536
center0	column=Property:X8, timestamp=1433368639022, value=30.49437543059147
center1	column=Area:X1, timestamp=1433368491087, value=0.6402566409891939
center1	column=Area:X5, timestamp=1433368491087, value=783.7587053127913
center1	column=Area:X6, timestamp=1433368491087, value=342.8292448071897
center1	column=Area:Y1, timestamp=1433368491087, value=220.64673025280085
center1	column=Area:Y2, timestamp=1433368491087, value=3.50140423678298
center1	column=Property:X2, timestamp=1433368491087, value=3.4956210738685867
center1	column=Property:X3, timestamp=1433368491087, value=0.23431918875234997
center1	column=Property:X4, timestamp=1433368491087, value=2.8119737694300517
center1	column=Property:X7, timestamp=1433368491087, value=14.569918168579095
center1	column=Property:X8, timestamp=1433368491087, value=17.111189704986096
center2	column=Area:X1, timestamp=1433368490962, value=0.7132145237290664
center2	column=Area:X5, timestamp=1433368490962, value=710.6244708361493
center2	column=Area:X6, timestamp=1433368490962, value=269.6655778328452
center2	column=Area:Y1, timestamp=1433368490962, value=220.47944650165203
center2	column=Area:Y2, timestamp=1433368490962, value=3.50083774498761
center2	column=Property:X2, timestamp=1433368490962, value=3.5053221446271685
center2	column=Property:X3, timestamp=1433368490962, value=0.2343316167774272
center2	column=Property:X4, timestamp=1433368490962, value=2.8119737694300517
center2	column=Property:X7, timestamp=1433368490962, value=12.112868604861262
center2	column=Property:X8, timestamp=1433368490962, value=15.037270790507428
center3	column=Area:X1, timestamp=1433368620112, value=0.7750120376770102
center3	column=Area:X5, timestamp=1433368620112, value=649.2406887995188
center3	column=Area:X6, timestamp=1433368620112, value=379.7365803608536
center3	column=Area:Y1, timestamp=1433368620112, value=134.75205421933256
center3	column=Area:Y2, timestamp=1433368620112, value=6.999999604723442
center3	column=Property:X2, timestamp=1433368620112, value=3.5002371659348404
center3	column=Property:X3, timestamp=1433368620112, value=0.23440519559978132
center3	column=Property:X4, timestamp=1433368620112, value=2.8126549060597026
center3	column=Property:X7, timestamp=1433368620112, value=37.13442943644058
center3	column=Property:X8, timestamp=1433368620112, value=38.3221744149107

4 row(s) in 0.5580 seconds

The application has thus been tested over the EMR cluster for correct centroid readings.