

Cloud Computing Lab 2

Phoebe Wu

EMR DASHBOARD

The screenshot shows the AWS EMR Dashboard for a cluster named "My cluster". The cluster status is "Waiting" (Cluster ready after last step completed). The dashboard includes tabs for Summary, Monitoring, Hardware, Events, Steps, Configurations, and Bootstrap actions. The Summary tab displays cluster details such as ID (j-283RSMYY9HHI8), Creation date (2018-11-07 17:44 (UTC-5)), Elapsed time (31 minutes), Auto-terminate (No), and Termination protection (Off). The Configuration details section shows Release label (emr-4.3.0), Hadoop distribution (Amazon 2.7.1), Applications (Ganglia 3.7.2, Hive 1.0.0, Hue 3.7.1, Mahout 0.11.0, Pig 0.14.0, Spark 1.6.0), and Log URI (s3://aws-logs-146747544651-us-east-1/elasticmapreduce/). The Network and hardware section shows Availability zone (us-east-1b), Subnet ID (subnet-c839b4e6), Master (Running 1 m3.2xlarge), Core (--), and Task (--). The Security and access section shows Key name (MyKeypair), EC2 instance profile (EMR_EC2_DefaultRole), EMR role (EMR_DefaultRole), and security group details for sg-011f6f5852d735e59 and sg-01f1b271a652b6c57.

console.aws.amazon.com

EMR – AWS Console

Services Resource Groups

Clusters Security configurations VPC subnets Events Help What's new

Clone Terminate AWS CLI export

Cluster: My cluster Waiting Cluster ready after last step completed.

Summary Monitoring Hardware Events Steps Configurations Bootstrap actions

Connections: Enable Web Connection – Hue, Spark History Server, Ganglia, Resource Manager ... (View All)

Master public DNS: ec2-52-207-216-37.compute-1.amazonaws.com SSH

Tags: -- View All / Edit

ID: j-283RSMYY9HHI8

Creation date: 2018-11-07 17:44 (UTC-5)

Elapsed time: 31 minutes

Auto-terminate: No

Termination protection: Off Change

Configuration details

Release label: emr-4.3.0

Hadoop distribution: Amazon 2.7.1

Applications: Ganglia 3.7.2, Hive 1.0.0, Hue 3.7.1, Mahout 0.11.0, Pig 0.14.0, Spark 1.6.0

Log URI: s3://aws-logs-146747544651-us-east-1/elasticmapreduce/

EMRFS consistent view: Disabled

Network and hardware

Availability zone: us-east-1b

Subnet ID: subnet-c839b4e6

Master: Running 1 m3.2xlarge

Core: --

Task: --

Security and access

Key name: MyKeypair

EC2 instance profile: EMR_EC2_DefaultRole

EMR role: EMR_DefaultRole

Visible to all users: All Change

Security groups for sg-011f6f5852d735e59

Master: (ElasticMapReduce-master)

Security groups for sg-01f1b271a652b6c57

Core & Task: (ElasticMapReduce-slave)

Feedback English (US)

© 2008 - 2018, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

HUE INTERFACE

The screenshot shows a web browser window titled "Hue - Quick Start" with the URL "ec2-52-207-216-37.compute-1.amazonaws.com:8888/about/". The browser has three tabs: "FoxyProxy - Downloads", "FoxyProxy options", and "Hue - Quick Start". The main content area is titled "Quick Start Wizard - Hue™ 3.7.1-amzn-5 - The Hadoop UI". It displays a navigation bar with links: "About Hue", "Quick Start" (which is underlined), "Configuration", and "Server Logs". Below the navigation bar, there are four steps: "Step 1: Check Configuration", "Step 2: Examples", "Step 3: Users", and "Step 4: Go!". The first step, "Check Configuration", is highlighted. The configuration files are located in "/etc/hue/conf.empty". A message states "All OK. Configuration check passed.". At the bottom left, there are "Back" and "Next" buttons.

Quick Start Wizard - Hue™ 3.7.1-amzn-5 - The Hadoop UI

Step 1: Check Configuration Step 2: Examples Step 3: Users Step 4: Go!

Checking current configuration

Configuration files located in `/etc/hue/conf.empty`

All OK. Configuration check passed.

Back Next

LOGIN EMR

```
[Phoebebede-MacBook-Air:Downloads phoebewu$ ssh -i MyKeyPair.pem.txt hadoop@ec2-52-207-216-37.compute-1.amazonaws.com
Last login: Wed Nov  7 23:46:15 2018

      _|_ _|_
      | ( /   Amazon Linux AMI
      ___\_\_|_|
https://aws.amazon.com/amazon-linux-ami/2015.09-release-notes/
22 package(s) needed for security, out of 34 available
Run "sudo yum update" to apply all updates.
Amazon Linux version 2018.03 is available.

EEEEEEEEEEEEEEEEEE MMMMMMM          MMMMMMM RRRRRRRRRRRRRRR
E:::::::EE:::::E M:::::::M          M:::::::M R:::::::R
EE:::::EE:::::E M:::::::M          M:::::::M R:::::RRRRRR:::::R
  E:::E     EEEEE M:::::::M          M:::::::M RR:::::R      R:::::R
  E:::E     M::::::M:::M          M:::M:::::M R:::::R      R:::::R
  E:::::EEEEEEEEEEE M:::::M M:::M M:::M M:::::M R:::::RRRRRR:::::R
  E:::::::EE:::::E M:::::M M:::M:::M M:::M M:::::M R:::::::::::RR
  E:::::EEEEEEEEEEE M:::::M M:::::M M:::::M R:::::RRRRRR:::::R
  E:::E     M:::::M M:::M          M:::::M R:::::R      R:::::R
  E:::E     EEEEE M:::::M          MMM  M:::::M R:::::R      R:::::R
EE:::::EE:::::E M:::::M          M:::::M R:::::R      R:::::R
E:::::::EE:::::E M:::::M          M:::::M RR:::::R      R:::::R
EEEEEEEEEEEEEEEEEE MMMMMMM          MMMMMMM RRRRRRRR      RRRRRR

[hadoop@ip-172-31-94-49 ~]$
```

I can not run 'sudo apt-get install python-pip' so I use yum instead.

RUN MAPREDUCE JOB: WORDCOUNT.PY

```
Removing HDFS temp directory hdfs:///user/hadoop/tmp/mrjob/wordcount.hadoop.20181108.005234.389385...
Removing temp directory /tmp/wordcount.hadoop.20181108.005234.389385...
[hadoop@ip-172-31-94-49 ~]$ hdfs dfs -cat /out/*
"87" 1
"90," 1
"Frisco," 2
"How" 1
"I" 20
"If" 1
"It" 3
"Lost," 1
"Pon" 1
"Remarkable" 1
"Singular" 1
"Tell" 1
"and" 7
"band," 1
"e's" 1
"his" 1
"the" 3
"un" 1
"(available" 1
"(if" 1
"(trademark/copyright)" 1
"***" 6
"1.A." 1
"1.E.1" 3
"1.E.3." 1
"1.E.8" 2
"1.F.1." 1
"[10s." 1
"11:15.]" 2
"1500" 1
"16A," 1
"1884--there" 1
"2011" 1
"22nd" 1
"29," 1
"40" 1
"4:35" 1
"4th," 1
"5:15" 1
"64-6221541." 1
"6:30" 1
"7" 1
"99712.." 1
"A," 2
"ACTUAL," 1
"AGREE" 2
"AK," 1
"ANYTHING" 1
"ARTHUR" 1
"Above" 2
"Accustomed" 1
"Adler?"" 2
"Again" 3
"Albert" 3
"Alexander" 1
"Allegro." 1
"American," 1
"And" 83
"Any" 2
"Archie," 1
"Evidently" 1
"Eyford.\\" 1
"FITNESS" 1
"Five" 1
"Farrington" 1
"Fashionable" 1
"Fordham" 1
"Frank" 14
"Freemason," 1
"Fritz!'" 1
"GUTENBERG-tm" 1
"Gang." 1
"Getting" 1
"Good-night.\\" 1
"Greenwich." 1
"Gutenberg-tm's" 1
"Gutenberg-tm." 1
"Hafiz" 1
"Hampshire" 3
"Hardy," 2
"Hart," 1
"He'd" 1
"Hence," 1
"Here," 1
" Hill." 1
"His" 50
"Holmes," 127
"Holmes--a" 1
"Horsham" 2
"Hum!" 2
"Hyde" 1
"INCLUDING" 1
"INDEMNITY" 1
"IS" 1
"Information" 4
"Insensibly" 1
"Irene" 13
"Isa" 3
"Isa." 1
"Its" 8
"James's," 1
"January," 2
"Jersey" 1
"Jump" 1
"K.';" 1
"King;" 1
"Klan." 1
"LEAGUE:" 1
"LICENSE" 2
"Lane." 1
"Leadenhall" 5
"League" 7
"Lebanon," 1
"Lee," 2
"Local" 1
"Lothman" 1
"MAN" 1
"Major-General" 1
"Malay" 1
"Mary," 3
"Mauritius." 1
"McCarthy.\\" 2
"McCarthys" 3
"To-day" 1
"Toller" 10
"Tollers" 1
"Treppoff" 1
"Trincomalee," 1
"Tuesday," 1
"Turner's" 2
"Turner," 4
"U.S.A." 1
"Use" 2
"Valley" 4
"Victoria!" 1
"Victoria" 2
"Victoria,\\" 1
"Vincent" 5
"Volunteers" 1
"WARRANTIES" 2
"WARRANTY" 1
"WARRANTY," 1
"Wednesday," 2
"Westphail," 1
"Wharf," 1
"Whatever" 1
"Whitney" 1
"Wilson." 4
"Wilson?\\" 2
"Wimpole" 1
"XI." 2
"You're" 1
"Young" 2
"\'A" 1
"\'At" 1
"\'But," 1
"\'Certainly,\\" 1
"\'Dear" 2
"\'Dearest" 1
"\'How" 2
"\'Lord" 1
"\'No," 9
"\'No," 1
"\'One" 2
"\'Pon" 1
"\'Precisely.\\" 1
"\'Some" 2
"\'Why," 1
"\'After" 1
"\'Alas!\\" 1
"\'Away" 2
"\'BALLARAT.\\" 1
"\'Because" 5
"\'Bring" 1
"\'But" 72
"\'DEAR" 1
"\'Do" 15
"\'Done" 1
"\'Entirely.\\" 2
"\'Fancy" 1
"\'Farewell," 1
"\'Fine" 1
"\'G\\" 1
"\'Good-evening," 1
"\'Good-night," 1
"\'Has" 1
"perpetrators." 1
"perplexity," 1
"person" 21
"person.\\" 1
"perspired" 1
"pestered" 1
"pestering" 1
"photograph;" 1
"phrase" 4
"pieces" 1
"pierced" 3
"pinnacles" 1
"pitch," 1
"pittance," 1
"pity," 2
"place--within" 1
"places." 2
"planking" 1
"planning," 1
"plantation," 1
"plaster." 1
"platform." 2
"playing" 4
"pleading" 1
"pleased," 2
"pleasure.\\" 1
"plentiful" 1
"plot," 1
"pockets" 4
"points." 2
"poison" 5
"poker" 2
"police," 1
"police-court" 1
"police-court,\\" 1
"police?\\" 1
"politicians" 1
"poor" 20
"positive" 5
"possession," 1
"possible" 33
"possible;" 1
"post" 4
"post." 1
"pounds," 4
"powers" 5
"practice," 2
"premises," 2
"presence--in" 1
"presently," 1
"presently:" 1
"press" 3
"pressing" 6
"pressure.\\" 1
"pretence" 1
"prevent?" 1
"preventing" 2
"prices." 1
"pride." 1
"print," 1
"prisoner's" 1
"prize," 1
"probably" 14
"vehement" 1
"veil," 4
"veins" 2
"ventilators" 1
"ventured" 2
"verify" 1
"vestry." 1
"view," 5
"viewed," 2
"views." 1
"vile" 3
"villain" 3
"villains" 1
"villes" 1
"violates" 1
"violence," 3
"violent," 2
"violin-player," 1
"visiting" 1
"visitor," 6
"visits?" 1
"vital" 3
"vizard" 1
"volume," 1
"vulnerable" 1
"waddling" 1
"wages," 1
"waggled" 1
"waistcoat." 1
"waited" 13
"walked" 27
"walks," 1
"wall," 3
"want," 1
"want?\\" 1
"warned" 1
"was?" 1
"wash" 2
"wasted.\\" 1
"watch" 5
"water." 4
"way," 16
"way:" 3
"ways." 2
"we" 413
"weakening" 1
"weapon" 4
"weapon.\\" 1
"wears" 1
"weather" 3
"weave," 1
"weaver" 1
"web" 5
"wedding?\\" 3
"weedy" 1
"week," 7
"week.\\" 1
"weigh" 1
"weighed" 2
"welcomed" 1
"well?\\" 1
"were,\\" 1
"western" 1
```

RUNNING PIG ON HADOOP MASTER NODE

```
grunt> a = LOAD '/home/hadoop/book.txt' as (lines:chararray);
18/11/08 01:35:23 INFO Configuration.deprecation: fs.default.name is deprecated. Instead, use fs.defaultFS
18/11/08 01:35:23 INFO Configuration.deprecation: io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
grunt>
grunt>
grunt> b = FOREACH a GENERATE FLATTEN(TOKENIZE(lines)) as word;
grunt> c = GROUP b by word;
grunt> d = FOREACH c GENERATE group, COUNT(b);
grunt> store d into '/home/hadoop/pig_wordcount';
```

```
18/11/08 01:35:29 INFO mapReduceLayer.MapReduceLauncher: 100% complete
18/11/08 01:35:29 INFO mapreduce.SimplePigStats: Script Statistics:
```

HadoopVersion	PigVersion	UserId	StartedAt	FinishedAt	Features
2.7.1-amzn-0	0.14.0-amzn-0	root	2018-11-08 01:35:24	2018-11-08 01:35:29	GROUP_BY

Success!

Job Stats (time in seconds):

JobId	Maps	Reduces	MaxMapTime	MinMapTime	AvgMapTime	MedianMapTime	MaxReduceTime	MinReduceTime	AvgReduceTime	MedianReducetime	AliaFeature	Outputs
job_local291507561_0003	1	1	n/a	n/a	n/a	n/a	n/a	n/a	a,b,c,d	GROUP_BY,COMBINER	/home/hadoop/pig_wordcount,	

Input(s):

Successfully read 13052 records from: "/home/hadoop/book.txt"

Output(s):

Successfully stored 12272 records in: "/home/hadoop/pig_wordcount"

Counters:

Total records written : 12272
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:

job_local291507561_0003

```
18/11/08 01:35:29 INFO jvm.JvmMetrics: Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
18/11/08 01:35:29 INFO jvm.JvmMetrics: Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
18/11/08 01:35:29 INFO jvm.JvmMetrics: Cannot initialize JVM Metrics with processName=JobTracker, sessionId= - already initialized
18/11/08 01:35:29 INFO mapReduceLayer.MapReduceLauncher: Success!
```

grunt>