

CISD42_Ch08_NoSQL Databases Lab Assignment

Amazon EMR > EMR on EC2: Clusters > My cluster

My cluster

Updated 9 minutes ago

Terminate

Clone in AWS CLI

Clone

▼ Summary

Cluster info

Cluster ID
j-2UR27OJXRRV2F

Cluster configuration

Instance groups

Capacity
1 Primary 1 Core 1 Task

Applications

Amazon EMR version
emr-6.15.0

Installed applications
Hadoop 3.3.6, Hive 3.1.3, Hue 4.11.0, Pig 0.17.0, Tez 0.10.2

Cluster management

Log destination in Amazon S3
Logging not configured

Persistent application UIs
[YARN timeline server](#)
[Tez UI](#)

Primary node public DNS
[ec2-34-238-169-156.compute-1.amazonaws.com](#)
[Connect to the Primary node using SSH](#)
[Connect to the Primary node using SSM](#)

Status and time

Status
Waiting

Creation time
December 15, 2023, 16:24 (UTC-08:00)

Elapsed time
34 minutes

< Properties Bootstrap actions Instances (Hardware) Steps Applications Configurations Monitoring Events >

Operating system [Info](#)

Amazon Linux release 2.0.20231116.0

Cluster logs [Info](#)

Archive log files to Amazon S3
Turned off

Encryption for logs
Turned off

Cluster termination [Info](#)

Edit cluster termination

Termination option
Manually terminate cluster

Idle time
-

Termination protection
Turned off

Network and security [Info](#)

Network

Virtual Private Cloud (VPC)
[vpc-0ea41301bd7d11167](#)

Security configuration

Security configuration
None

Permissions

Service role for Amazon EMR
[EMR_DefaultRole](#)

Host: Username: Password: Port: Quickconnect

Status: Connecting to ec2-34-238-169-156.compute-1.amazonaws.com...

Status: Using username "hadoop".

Status: Connected to ec2-34-238-169-156.compute-1.amazonaws.com

Status: Retrieving directory listing...

Status: Listing directory /home/hadoop

Status: Directory listing of "/home/hadoop" successful

Local site: /Users/macpro/ Remote site: /home/hadoop

> /

> support

> Volumes

> bin

> cores

> dev

> etc

> home

> opt

> private

>/sbin

> tmp

> usr

> var

Filename

Filesize

Filetype

Last modified

Permissions

Owner/Group

sda-0.log.0

10,224

0-file

06/25

requirements.txt

20,188

txt-file

10/25/

package-lock.json

7,988

JSON document

02/01/

.zshrc

541

File

06/06

.zshenv

81

File

08/28

.zsh_history

89,036

File

12/11/

.zprofile.pysave

743

pysave-file

10/01/

.zprofile.macports-saved_2019-11-15_at_00:19:19

164

macports-saved_2

10/26/

.zprofile

909

File

10/25/

.xonshrc

671

File

06/06

.viminfo

1,547

File

09/03

.tcshrc

308

File

06/06

.python_history

7

File

09/14/

.mysql_history

314

File

07/01/

.mime.types

764

types-file

08/23

.mailcap

1,459

File

08/23

.lessstat

20

File

04/12/

.gitconfig

55

File

11/07/

.condarc

23

File

09/11/

.bash_profile

524

File

06/06

.bash_history

5,427

File

08/24

23 files and 66 directories. Total size: 186,060 bytes

2 files and 3 directories. Total size: 712 bytes

Server/Local file

Direction

Remote file

Size

Priority

Status

171

Madelyn Hensley

10075 Thierer Plaza

New York

New York

81377

1976

M

TRUE

FALSE

FALSE

FALSE

172

Lonny Foster

23901 Park Meadow Dr

Austin Texas

13498 1981

F

TRUE

FALSE

FALSE

TRUE

173

Karina Livingston

95 Anderson Park

Chattanooga

Tennessee

94518

1977

F

TRUE

TRUE

TRUE

TRUE

174

Avery McCormick

09992 Sunfield Parkway

Chicago Illinois

38300 1992

F

TRUE

FALSE

FALSE

TRUE

175

Peter King

0486 Dryden Road

Chicago Illinois

21012 1991

M

TRUE

FALSE

TRUE

TRUE

176

Bret Ibarra

14 Transport Place

San Diego

California

12865

1984

M

TRUE

FALSE

FALSE

FALSE

177

Leonardo Wheeler

806 Corry Crossing

New York

New York

44464

1972

F

FALSE

FALSE

FALSE

TRUE

178

Bennett Noble

8 South Terrace

Hixson

Tennessee

52890 1989

F

FALSE

FALSE

TRUE

TRUE

179

Marcia Mathews

0380 Knutson Road

Dallas Texas

80477 1967

F

TRUE

FALSE

TRUE

TRUE

180

Avis Kramer

49624 Hanover Junction

New York

New York

62542

1965

M

TRUE

FALSE

TRUE

TRUE

181

Lynnette Tate

30741 Paget Court

New York

New York

34886

1987

M

TRUE

FALSE

TRUE

TRUE

182

Lakisha Estrada

50 Dahle Crossing

Dallas Texas

16042 1991

F

FALSE

TRUE

FALSE

TRUE

183

Bill Silva

4 McBride Crossing

Detroit Michigan

15871

1968

M

TRUE

TRUE

TRUE

TRUE

```
-- V -->
A newer version of Amazon Linux is available!
Amazon Linux 2023. GA and supported until 2026-03-15.
https://aws.amazon.com/linux/amazon-linux-2023/

19 package(s) needed for security, out of 35 available
Run "sudo yum update" to apply all updates.

EEEEEEEEEEEEEEEEEEEE MMMMMMMM MMMMMMMM RRRRRRRRRRRRRRRR
E::::::::::::::::::::E M::::::::M M::::::::M R:::::::::::::R
EE::::::::EEEEEEEEEE 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::::M R::R R::::R
E::::EEEEEEEEEE M::::M M::M M::M M::::M R::RRRRRR:::::R
E::::::::::::::::::::E M::::M M::M:M M::::M R::::::::::::RR
E::::EEEEEEEEEE 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::::::::EEEEEEEEEE M::::M M::::M R::R R::::R
E::::::::::::::::::::E M::::M M::::M RR::::R R::::R
EEEEEEEEEEEEEEEEEEEE MMMMMMMM MMMMMMMM RRRRRRR RRRRRR

[hadoop@ip-172-31-40-176 ~]$
```

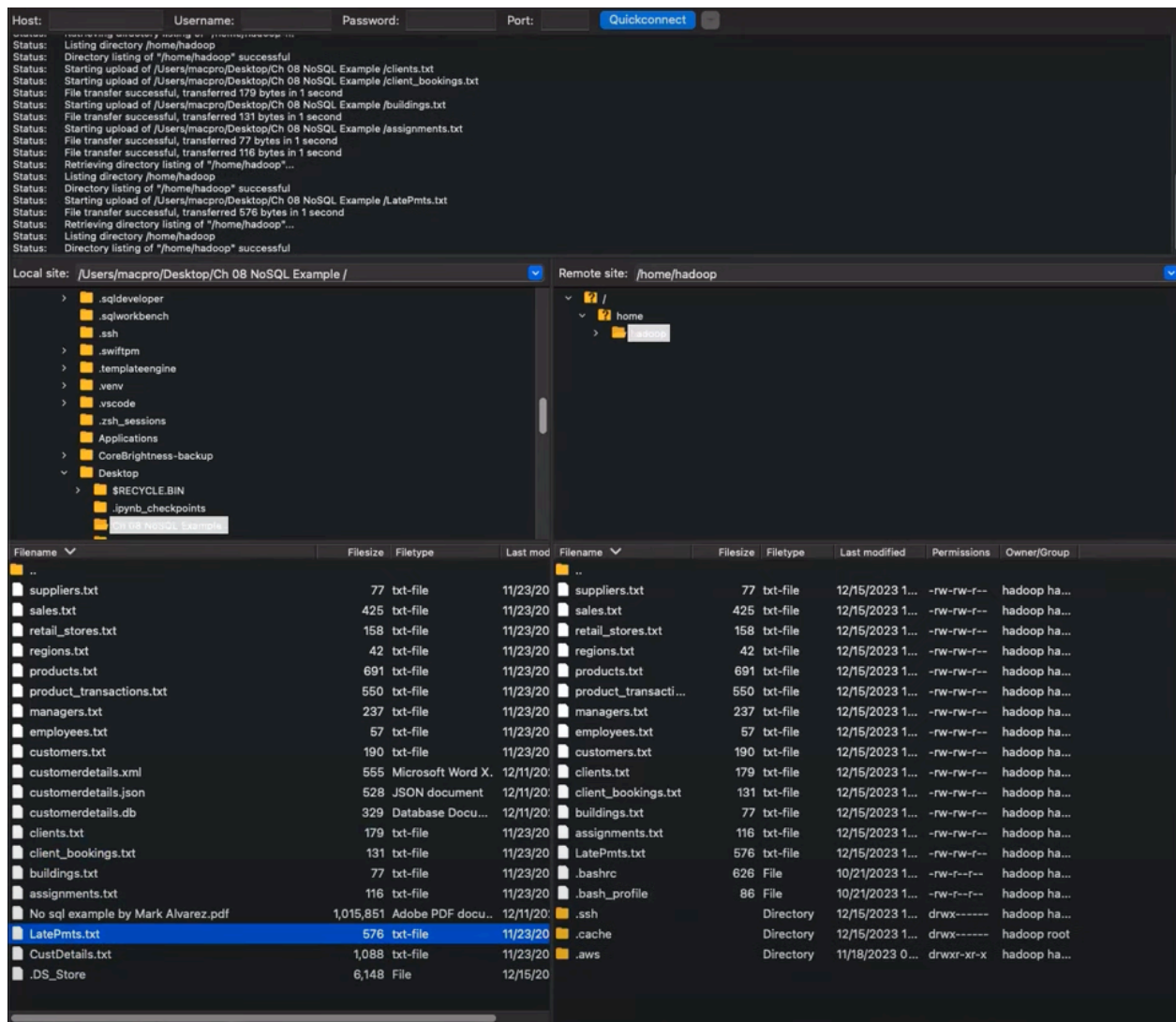
```
19 package(s) needed for security, out of 20 available
Run "sudo yum update" to apply all updates.

EEEEEEEEEEEEEEEEEEEE MMMMMMMM MMMMMMMM RRRRRRRRRRRRRRRR
E::::::::::::::::::::E M::::::::M M::::::::M R:::::::::::::R
EE::::::::EEEEEEEEEE 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::::M R::R R::::R
E::::EEEEEEEEEE M::::M M::M M::M M::::M R::RRRRRR:::::R
E::::::::::::::::::::E M::::M M::M:M M::::M R::::::::::::RR
E::::EEEEEEEEEE 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::::::::EEEEEEEEEE M::::M M::::M R::R R::::R
E::::::::::::::::::::E M::::M M::::M RR::::R R::::R
EEEEEEEEEEEEEEEEEEEE MMMMMMMM MMMMMMMM RRRRRRR RRRRRR

[hadoop@ip-172-31-39-134 ~]$
```

```
E::::E EEEEE M::::::::M M::::::::M RR::::R R::::R
E::::E M::::M:M::M M::M::::M R::R R::::R
E::::EEEEEEEEEE M::::M M::M M::M M::::M R::RRRRRR:::::R
E::::::::::::::::::::E M::::M M::M:M M::::M R::::::::::::RR
E::::EEEEEEEEEE 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::::::::EEEEEEEEEE M::::M M::::M R::R R::::R
E::::::::::::::::::::E M::::M M::::M RR::::R R::::R
EEEEEEEEEEEEEEEEEEEE MMMMMMMM MMMMMMMM RRRRRRR RRRRRR

[hadoop@ip-172-31-39-134 ~]$ hadoop fs -mkdir NoSQL
[hadoop@ip-172-31-39-134 ~]$ hadoop fs -mkdir NoSQL/Exercise1
[hadoop@ip-172-31-39-134 ~]$ hadoop fs -mkdir NoSQL/Exercise1/input
[hadoop@ip-172-31-39-134 ~]$ hadoop fs -ls
Found 1 items
drwxr-xr-x - hadoop hdfsadmin group 0 2023-12-16 01:02 NoSQL
[hadoop@ip-172-31-39-134 ~]$
```



```

E:::::EEEEEEEEEE M::::M M::::M M::::M R::RRRRR:::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:::::EEEEEEEE:::E M::::M M::::M R::R R:::R
E:::::EEEEEEEE:::E M::::M M::::M RR:::R R:::R
EEEEEEEEEEEEEEEEEEEE MMMMMM MMMMMM RRRRRR RRRRRR

[hadoop@ip-172-31-39-134 ~]$ hadoop fs -mkdir NoSQL
[hadoop@ip-172-31-39-134 ~]$ hadoop fs -mkdir NoSQL/Exercise1
[hadoop@ip-172-31-39-134 ~]$ hadoop fs -mkdir NoSQL/Exercise1/input
[hadoop@ip-172-31-39-134 ~]$ hadoop fs -ls
Found 1 items
drwxr-xr-x - hadoop hdfsadmingroup 0 2023-12-16 01:02 NoSQL
[hadoop@ip-172-31-39-134 ~]$ hadoop fs -copyFromLocal CustDetails.txt NoSQL/Exercise1/input
copyFromLocal: 'CustDetails.txt': No such file or directory
[hadoop@ip-172-31-39-134 ~]$ hadoop fs -copyFromLocal CustDetails.txt NoSQL/Exercise1/input
[hadoop@ip-172-31-39-134 ~]$ hive

```



```
[hadoop@ip-172-31-39-134 ~]$ hadoop fs -ls
Found 1 items
drwxr-xr-x - hadoop hdfsadmin group 0 2023-12-16 01:02 NoSQL
[hadoop@ip-172-31-39-134 ~]$ hadoop fs -copyFromLocal CustDetails.txt NoSQL/Exercise1/input
copyFromLocal: `CustDetails.txt': No such file or directory
[hadoop@ip-172-31-39-134 ~]$ hadoop fs -copyFromLocal CustDetails.txt NoSQL/Exercise1/input
[hadoop@ip-172-31-39-134 ~]$ hive
Hive Session ID = 9b252c42-1c1a-40de-81a6-58e59eec8581

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j2.properties Async: false
hive> create database exercise1;
OK
Time taken: 0.856 seconds
hive> create table exercise1.customers_NoSQL(record map<int,string>)
> row format delimited
> map keys terminated by '\t';
█
```

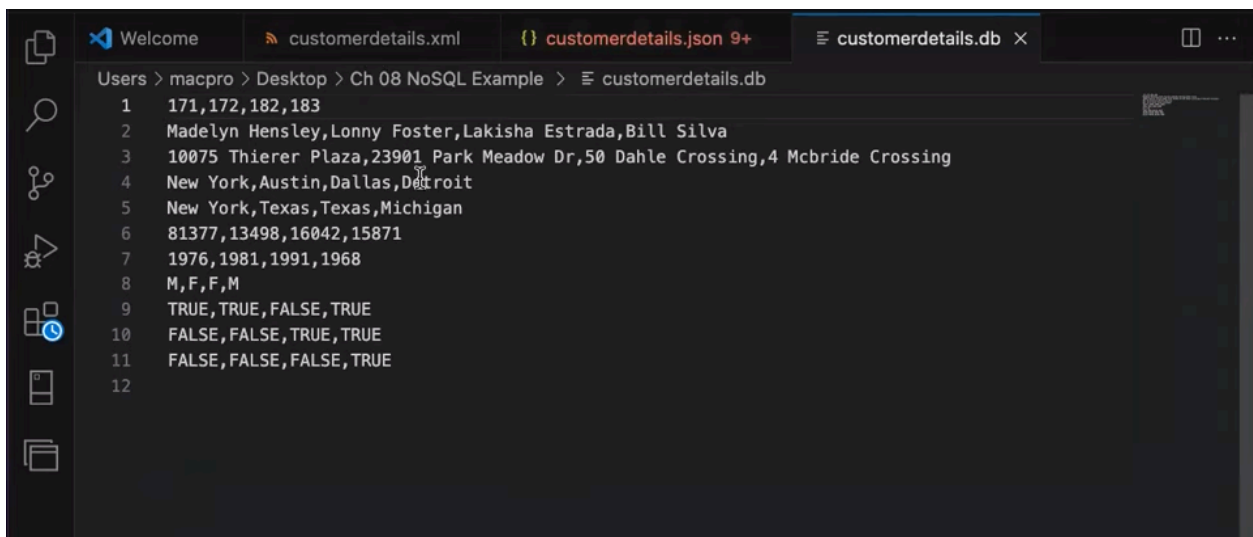
```
drwxr-xr-x - hadoop hdfsadmin group 0 2023-12-16 01:02 NoSQL
[hadoop@ip-172-31-39-134 ~]$ hadoop fs -copyFromLocal CustDetails.txt NoSQL/Exercise1/input
copyFromLocal: `CustDetails.txt': No such file or directory
[hadoop@ip-172-31-39-134 ~]$ hadoop fs -copyFromLocal CustDetails.txt NoSQL/Exercise1/input
[hadoop@ip-172-31-39-134 ~]$ hive
Hive Session ID = 9b252c42-1c1a-40de-81a6-58e59eec8581

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j2.properties Async: false
hive> create database exercise1;
OK
Time taken: 0.856 seconds
hive> create table exercise1.customers_NoSQL(record map<int,string>)
> row format delimited
> map keys terminated by '\t';
OK
Time taken: 0.396 seconds
hive> █
```

```
hive> use exercise1;
OK
Time taken: 0.032 seconds
hive> LOAD DATA INPATH 'NoSQL/Exercise1/input/CustDetails.txt' OVERWRITE INTO TABLE customers_NoSQL;
Loading data to table exercise1.customers_nosql
OK
Time taken: 0.751 seconds
hive> █
```

```
se
hive> create database exercise1;
OK
Time taken: 0.856 seconds
hive> create table exercise1.customers_NoSQL(record map<int,string>)
> row format delimited
> map keys terminated by '\t';
OK
Time taken: 0.396 seconds
hive> use exercise1;
OK
Time taken: 0.027 seconds
hive> LOAD DATA INPATH 'NoSQL/Exercise1/input/CustDetails.txt' OVERWRITE INTO TABLE customers_NoSQL;
Loading data to table exercise1.customers_nosql
OK
Time taken: 0.701 seconds
hive> select * from customers_nosql;
█
```

```
hive> select * from customers_nosql;
OK
{171:"Madelyn Hensley\t10075 Thierer Plaza\tNew York\tNew York\t81377\t1976\tM\tTRUE\tFALSE\tFALSE"}
{172:"Lonny Foster\t23901 Park Meadow Dr\tAustin\tTexas\t13498\t1981\tF\tTRUE\tFALSE\tFALSE"}
{173:"Karina Livingston\t95 Anderson Park\tChattanooga\tTennessee\t94518\t1977\tF\tTRUE\tTRUE\tTRUE"}
{174:"Avery McCormick\t09992 Sunfield Parkway\tChicago\tIllinois\t38300\t1992\tF\tTRUE\tFALSE\tFALSE"}
}
{175:"Peter King\t0486 Dryden Road\tChicago\tIllinois\t21012\t1991\tM\tTRUE\tFALSE\tTRUE"}
{176:"Bret Ibarra\t14 Transport Place\tSan Diego\tCalifornia\t12865\t1984\tM\tTRUE\tFALSE\tFALSE"}
{177:"Leonardo Wheeler\t806 Corry Crossing\tNew York\tNew York\t44464\t1972\tF\tFALSE\tFALSE\tTRUE"}
{178:"Bennett Noble\t8 South Terrace\tHixson\tTennessee\t52890\t1989\tF\tFALSE\tFALSE\tTRUE"}
{179:"Marcia Mathews\t0380 Knutson Road\tDallas\tTexas\t80477\t1967\tF\tTRUE\tFALSE\tTRUE"}
{180:"Avis Kramer\t49624 Hanover Junction\tNew York\tNew York\t62542\t1965\tM\tTRUE\tFALSE\tTRUE"}
{181:"Lynnette Tate\t30741 Paget Court\tNew York\tNew York\t34886\t1987\tM\tTRUE\tFALSE\tTRUE"}
{182:"Lakisha Estrada\t50 Dahle Crossing\tDallas\tTexas\t16042\t1991\tF\tFALSE\tTRUE\tFALSE"}
{183:"Bill Silva\t4 McBride Crossing\tDetroit\tMichigan\t15871\t1968\tM\tTRUE\tTRUE\tTRUE"}
Time taken: 2.758 seconds, Fetched: 13 row(s)
hive>
```



Users > macpro > Desktop > Ch 08 NoSQL Example > customerdetails.db

```
1 171,172,182,183
2 Madelyn Hensley,Lonny Foster,Lakisha Estrada,Bill Silva
3 10075 Thierer Plaza,23901 Park Meadow Dr,50 Dahle Crossing,4 McBride Crossing
4 New York,Austin,Dallas,Detroit
5 New York,Texas,Texas,Michigan
6 81377,13498,16042,15871
7 1976,1981,1991,1968
8 M,F,F,M
9 TRUE,TRUE,FALSE,TRUE
10 FALSE,FALSE,TRUE,TRUE
11 FALSE,FALSE,FALSE,TRUE
12
```

```
customerdetails.xml
1 <customer_details>
2   <customer>
3     <id>182</id>
4     <name>Lakisha Estrada</name>
5     <street>50 Dahle Crossing</street>
6     <city>Dallas</city>
7     <state>Texas</state>
8     <zip>16042</zip>
9     <yob>1991</yob>
10    <gender>F</gender>
11    <credit_card>FALSE</credit_card>
12    <internet>TRUE</internet>
13    <mobile>FALSE</mobile>
14  </customer>
15  <customer>
16    <id>183</id>
17    <name>Bill Silva</name>
18    <street>4 McBride Crossing</street>
19    <city>Detroit</city>
20    <state>Michigan</state>
21    <zip>15871</zip>
22    <yob>1968</yob>
23    <gender>M</gender>
24    <credit_card>TRUE</credit_card>
25    <internet>TRUE</internet>
26    <mobile>TRUE</mobile>
27  </customer>
```

```
customerdetails.json 9+
1 {
2   "id" : 171,
3   "name": Madelyn Hensley,
4   "street": 10075 Thierer Plaza,
5   "city": New York,
6   "state": New York,
7   "zip": 81377,
8   "yob": 1976,
9   "gender": M,
10  "credit_card": TRUE,
11  "internet": FALSE,
12  "mobile": FALSE
13 },
14 {
15   "id" : 172,
16   "name": Lonny Foster,
17   "street": 23901 Park Meadow Dr,
18   "city": Austin,
19   "state": Texas,
20   "zip": 13498,
21   "yob": 1981,
22   "gender": F,
23   "credit_card": TRUE,
24   "internet": FALSE,
25   "mobile": FALSE
26 }
27 }
```

OK

```
{171:"Madelyn Hensley\t10075 Thierer Plaza\tNew York\tNew York\t81377\t1976\tM\tTRUE\tFALSE\tFALSE"}
{172:"Lonny Foster\t23901 Park Meadow Dr\tAustin\tTexas\t13498\t1981\tF\tTRUE\tFALSE\tFALSE"}
{173:"Karina Livingston\t95 Anderson Park\tChattanooga\tTennessee\t94518\t1977\tF\tTRUE\tTRUE\tTRUE"}
{174:"Avery McCormick\t09992 Sunfield Parkway\tChicago\tIllinois\t38300\t1992\tF\tTRUE\tFALSE\tFALSE"}
}
{175:"Peter King\t0486 Dryden Road\tChicago\tIllinois\t21012\t1991\tM\tTRUE\tFALSE\tTRUE"}
{176:"Bret Ibarra\t14 Transport Place\tSan Diego\tCalifornia\t12865\t1984\tM\tTRUE\tFALSE\tFALSE"}
{177:"Leonardo Wheeler\t806 Corry Crossing\tNew York\tNew York\t44464\t1972\tF\tFALSE\tFALSE\tTRUE"}
{178:"Bennett Noble\t8 South Terrace\tHixson\tTennessee\t52890\t1989\tF\tFALSE\tFALSE\tTRUE"}
{179:"Marcia Mathews\t0380 Knutson Road\tDallas\tTexas\t80477\t1967\tF\tTRUE\tFALSE\tTRUE"}
{180:"Avis Kramer\t49624 Hanover Junction\tNew York\tNew York\t62542\t1965\tM\tTRUE\tFALSE\tTRUE"}
{181:"Lynnette Tate\t30741 Paget Court\tNew York\tNew York\t34886\t1987\tM\tTRUE\tFALSE\tTRUE"}
{182:"Lakisha Estrada\t50 Dahle Crossing\tDallas\tTexas\t16042\t1991\tF\tFALSE\tTRUE\tFALSE"}
{183:"Bill Silva\t4 McBride Crossing\tDetroit\tMichigan\t15871\t1968\tM\tTRUE\tTRUE\tTRUE"}
Time taken: 2.758 seconds, Fetched: 13 row(s)
hive> exit;
hadoop@ip-172-31-39-134 ~]$
```


Services

Amazon EMR

My cluster

Updated 11 minutes ago

Terminate

View in AWS console

Clone

Summary

Cluster info

Cluster ID
j-2UR27OJXRRV2F

Cluster configuration
Instance groups
m4.xlarge
1 Primary 1 Core 1 Task

Applications

Amazon EMR server
emr-6.15.0

Installed application
Hadoop 3.4.6, Hive 3.13, Hue
4 11.0, Pig 0.17.0, Tez 0.10.2

Cluster management

Log definitions in Amazon S3
Logging not configured

Parquet application (HLS)
VARS resolve cluster
EMRFS

Primary mode public IP
ec2-34-238-169-158.compute-1.amazonaws.com

Status and time

Status
Waiting

Creation time
December 15, 2023, 16:24 (UTC-08:00)

Elapsed time
51 minutes, 17 seconds

Terminate cluster

Are you sure you want to terminate Cluster j-2UR27OJXRRV2F (My cluster)? Any pending work or data residing on this cluster will be lost, such as data stored in HDFS. This action is irreversible.

Cancel

Terminate

Operating system

Amazon Linux (Release)
2.0.20231116.0

Cluster logs

Amazon log files in Amazon S3
Turned off

Encryption for logs
Turned off

Cluster termination

EMR cluster termination

Termination option
Manually terminate cluster

Wait time

Termination protection
Turned off

Network and security

Network

Virtual Private Cloud (VPC)
vpc-00047401bd7d171388

Security configuration

Security configuration
None

Permissions

Security role for Amazon EMR
EMR_DefaultRole

aws

Services

Search

[Option+S]

N. Virginia

voclabs/user1265657=Alvarez_Mark @ 6894-2263-3867

Amazon EMR > EMR on EC2: Clusters > My cluster

My cluster

Updated less than a minute ago

Terminate

▼ Summary

Cluster info

Cluster ID
j-2UR27OJXRRV2F

Cluster configuration

Instance groups

Capacity
1 Primary 1 Core 1 Task

Applications

Amazon EMR version
emr-6.15.0

Installed applications
Hadoop 3.3.6, Hive 3.1.3, Hue 4.11.0, Pig 0.17.0, Tez 0.10.2

Cluster management

Log destination in Amazon S3
Logging not configured

Persistent application UIs
[YARN timeline server](#)
[Tez UI](#)

Primary node public DNS
[ec2-34-238-169-156.compute-1.amazonaws.com](#)
[Connect to the Primary node using SSH](#)
[Connect to the Primary node using SSM](#)

Account ID: 6894-2263-3867

Federated user: voclabs/user1265657=Alvarez_Mark

Account

Organization

Service Quotas

Billing and Cost Management

Switch role

Sign out

Elapsed time
51 minutes, 21 seconds

< Properties Bootstrap actions Instances (Hardware) Steps Applications Configurations Monitoring Events >

Operating system Info

Amazon Linux release 2.0.20231116.0

Cluster logs Info

Archive log files to Amazon S3
Turned off

Encryption for logs
Turned off

Cluster termination Info

Edit cluster termination

Termination option
Manually terminate cluster

Idle time
-

Termination protection
Turned off

Network and security Info

Network

Virtual Private Cloud (VPC)
[vpc-0ea41301bd7d11167](#)

Security configuration

Security configuration
None

Permissions

Service role for Amazon EMR
[EMR_DefaultRole](#)

© 2023 Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy | Terms | Contact us