

# EVIDENCIAS

## Creación del cluster

Amazon EMR > EMR on EC2: Clusters > My cluster LULA

My cluster LULA Updated 4 minutes ago [Refresh] [Terminate] [Clone in AWS CLI] [Clone]

### Summary

<b>Cluster info</b> Cluster ID: j-18D0X460DUWPS Cluster configuration: emr-7.3.0 Instance groups: 1 Primary, 1 Core, 1 Task Capacity: 1 Primary, 1 Core, 1 Task	<b>Applications</b> Amazon EMR version: emr-7.3.0 Installed applications: Hadoop 3.5.6, HCatalog 3.1.3, Hive 3.1.3, Hue 4.11.0, JupyterEnterpriseGateway 2.6.0, JupyterHub 1.5.0, Uky 0.8.0, Spark 3.5.1, Sqoop 1.4.7, Zeppelin 0.11.1, ZooKeeper 3.9.1	<b>Cluster management</b> Log destination in Amazon S3: aws-logs-286128634116-us-east-1/elasticmapreduce Persistent application UIs: Spark History Server, YARN timeline server, Tez UI Primary node public DNS: ec2-54-224-107-30.compute-1.amazonaws.com Connect to the Primary node using SSH: Connect to the Primary node using SSM	<b>Status and time</b> Status: Waiting Creation time: October 18, 2024, 20:12 (UTC-05:00) Elapsed time: 19 minutes, 27 seconds
---	---	---	---

Properties | Bootstrap actions | Instances (Hardware) | Steps | **Applications** | Configurations | Monitoring | Events | Tags (0)

### Application user interfaces

Applications installed on your Amazon EMR cluster publish user interfaces (UI) as websites. You can use these to monitor cluster activity.

☒ On-cluster application UIs  
On-cluster UIs are available only while your cluster is running. Use the following links to get started. To access all the application UIs, set up SSH tunneling.

☐ Persistent application UIs  
Persistent UIs don't require SSH tunneling. They are hosted off of the cluster and are available for 30 days after an application ends.

**Live Application UIs**  
These on-cluster application UIs are available without SSH tunneling.


Application UIs

## Comando put

```
hadoop@ip-172-31-38-45 ~$ hadoop distcp s3://at0263datasets/datasets/airlines.csv /tmp/
2024-10-19 01:56:57,866 INFO tools.DistCp: Input Options: DistCpOptions(atomicCommit=false, syncFolder=false, deleteMissing=false, ignoreFailures=false, overwrite=false, append=false, useDiff=false,
f=false, fromSnapshot=null, toSnapshot=null, skipCRC=false, blocking=true, numListStatusThreads=0, maxMaps=20, mapBandwidth=0.0, copyStrategy='uniformize', preserveStatus=[], atomicWorkPath=null,
null, sourceFileListing=null, sourcePaths=[s3://at0263datasets/datasets/airlines.csv], targetPath=/tmp, filterFile='null', blocksPerChunk=0, copyBufferSize=8192, verboseLog=false, directWrite=false,
erator=false), sourcePaths=[s3://at0263datasets/datasets/airlines.csv], targetPath=/tmp, filterFile='null', blocksPerChunk=0, copyBufferSize=8192, verboseLog=false, directWrite=false
2024-10-19 01:56:58,066 INFO client.DefaultHadoopMFollowerProxyProvider: Connecting to ResourceManager at ip-172-31-38-45.ec2.internal/172.31.38.45:8032
2024-10-19 01:56:58,197 INFO client.AMRProxy: Connecting to Application History server at ip-172-31-38-45.ec2.internal/172.31.38.45:10200
2024-10-19 01:57:00,269 ERROR tools.DistCp: Exception encountered
java.io.IOException: com.amazonaws.emr.hadoop.fs.shaded.com.amazonaws.services.s3.model.AmazonS3Exception: The specified bucket does not exist (Service: Amazon S3; Status Code: 404; Error Code: NoBu
cket; Request ID: G0E32Y9QF65W9AJ7; S3 Extended Request ID: d+FBWDT6a5EM0yohG/hAbtriN3hryygrQ43cNBy63myq0E+a5yDNQ32nod6V4sryuBulw*QK-; Proxy: null), S3 Extended Request ID: d+FBWDT6a5EM0yohG/hAbtriN
3hryygrQ43cNBy63myq0E+a5yDNQ32nod6V4sryuBulw*QK-
at com.amazonaws.emr.hadoop.fs.s3n.Jets3tNativeFileSystemStore.list(Jets3tNativeFileSystemStore.java:429)
at com.amazonaws.emr.hadoop.fs.s3n.Jets3tNativeFileSystemStore.isFileUsingFolderObject(Jets3tNativeFileSystemStore.java:255)
at com.amazonaws.emr.hadoop.fs.s3n.Jets3tNativeFileSystemStore.isFileUsingFolderObject(Jets3tNativeFileSystemStore.java:218)
at com.amazonaws.emr.hadoop.fs.s3n.S3NativeFileSystem.getFileStatus(S3NativeFileSystem.java:554)
at org.apache.hadoop.fs.Globber.getFileStatus(Globber.java:115)
at org.apache.hadoop.fs.Globber.doGlob(Globber.java:349)
at org.apache.hadoop.fs.Globber.glob(Globber.java:202)
at org.apache.hadoop.fs.FileSystem.globStatus(FileSystem.java:2287)
at com.amazonaws.emr.hadoop.fs.FileSystem.globStatus(FileSystem.java:461)
at org.apache.hadoop.tools.GlobberCopyListing.doBuildListing(GlobberCopyListing.java:77)
at org.apache.hadoop.tools.CopyListing.buildListing(CopyListing.java:86)
at org.apache.hadoop.tools.DistCp.createInputFileListing(DistCp.java:370)
at org.apache.hadoop.tools.DistCp.prepareFileListing(DistCp.java:346)
at org.apache.hadoop.tools.DistCp.createAndSubmitJob(DistCp.java:205)
at org.apache.hadoop.tools.DistCp.execute(DistCp.java:182)
at org.apache.hadoop.tools.DistCp.run(DistCp.java:153)
at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:82)
at org.apache.hadoop.tools.DistCp.main(DistCp.java:443)
Caused by: com.amazonaws.emr.hadoop.fs.shaded.com.amazonaws.services.s3.model.AmazonS3Exception: The specified bucket does not exist (Service: Amazon S3; Status Code: 404; Error Code: NoSuchBucket;
Request ID: G0E32Y9QF65W9AJ7; S3 Extended Request ID: d+FBWDT6a5EM0yohG/hAbtriN3hryygrQ43cNBy63myq0E+a5yDNQ32nod6V4sryuBulw*QK-; Proxy: null), S3 Extended Request ID: d+FBWDT6a5EM0yohG/hAbtriN3hryygrQ43
cNBy63myq0E+a5yDNQ32nod6V4sryuBulw*QK-
at com.amazonaws.emr.hadoop.fs.shaded.com.amazonaws.http.AmazonHttpClient$RequestExecutor.handleErrorResponse(AmazonHttpClient.java:1879)
```

```
[hadoop@ip-172-31-38-45 ~]$ hdfs dfs -copyFromLocal /home/ec2-home/st0263-242/bigdata/datasets/* /user/hadoop/datasets/
copyFromLocal: '/home/ec2-home/st0263-242/bigdata/datasets/*': No such file or directory
[hadoop@ip-172-31-38-45 ~]$ hdfs dfs -copyFromLocal /home/hadoop/st0263-242/bigdata/datasets/* /user/hadoop/datasets/
copyFromLocal: '/user/hadoop/datasets/gutenberg-small/AbrahamLincoln_LincolnLetters.txt': File exists
copyFromLocal: '/user/hadoop/datasets/gutenberg-small/AbrahamLincoln_LincolnsFirstInauguralAddress.txt': File exists
copyFromLocal: '/user/hadoop/datasets/gutenberg-small/AbrahamLincoln_LincolnsGettysburgAddressGivenNovember-19-1863.txt': File exists
copyFromLocal: '/user/hadoop/datasets/gutenberg-small/AbrahamLincoln_LincolnsInauguralsAddressesandLettersSelections.txt': File exists
copyFromLocal: '/user/hadoop/datasets/gutenberg-small/AbrahamLincoln_LincolnsSecondInauguralAddress.txt': File exists
copyFromLocal: '/user/hadoop/datasets/gutenberg-small/AbrahamLincoln_SpeechesandLettersofAbrahamLincoln1832-1865.txt': File exists
copyFromLocal: '/user/hadoop/datasets/gutenberg-small/AbrahamLincoln_StateoftheUnionAddresses.txt': File exists
copyFromLocal: '/user/hadoop/datasets/gutenberg-small/AbrahamLincoln_TheEmancipationProclamation.txt': File exists
copyFromLocal: '/user/hadoop/datasets/gutenberg-small/AbrahamLincoln_TheLifeandPublicServiceofGeneralZacharyTaylorAnAddress.txt': File exists
copyFromLocal: '/user/hadoop/datasets/gutenberg-small/AbrahamLincoln_TheWritingsofAbrahamLincolnVolume1.txt': File exists
copyFromLocal: '/user/hadoop/datasets/gutenberg-small/AbrahamLincoln_TheWritingsofAbrahamLincolnVolume2.txt': File exists
copyFromLocal: '/user/hadoop/datasets/gutenberg-small/AbrahamLincoln_TheWritingsofAbrahamLincolnVolume3.txt': File exists
copyFromLocal: '/user/hadoop/datasets/gutenberg-small/AbrahamLincoln_TheWritingsofAbrahamLincolnVolume4.txt': File exists
copyFromLocal: '/user/hadoop/datasets/gutenberg-small/AbrahamLincoln_TheWritingsofAbrahamLincolnVolume5.txt': File exists
copyFromLocal: '/user/hadoop/datasets/gutenberg-small/AbrahamLincoln_TheWritingsofAbrahamLincolnVolume6.txt': File exists
copyFromLocal: '/user/hadoop/datasets/gutenberg-small/AbrahamLincoln_TheWritingsofAbrahamLincolnVolume7.txt': File exists
[hadoop@ip-172-31-38-45 ~]$ hdfs dfs -ls /user/hadoop/datasets
Found 11 items
drwxr-xr-x 1 hadoop hdfsadmingroup 780058 2024-10-19 02:00 /user/hadoop/datasets/airlines.csv
drwxr-xr-x 1 hadoop hdfsadmingroup 0 2024-10-19 02:00 /user/hadoop/datasets/all-news
drwxr-xr-x 1 hadoop hdfsadmingroup 0 2024-10-19 02:00 /user/hadoop/datasets/covid19
drwxr-xr-x 1 hadoop hdfsadmingroup 0 2024-10-19 02:00 /user/hadoop/datasets/flights
drwxr-xr-x 1 hadoop hdfsadmingroup 0 2024-10-19 02:00 /user/hadoop/datasets/gutenberg
drwxr-xr-x 1 root hdfsadmingroup 0 2024-10-19 01:53 /user/hadoop/datasets/gutenberg-small
drwxr-xr-x 1 hadoop hdfsadmingroup 0 2024-10-19 02:00 /user/hadoop/datasets/omn
drwxr-xr-x 1 hadoop hdfsadmingroup 0 2024-10-19 02:00 /user/hadoop/datasets/otrs
drwxr-xr-x 1 hadoop hdfsadmingroup 0 2024-10-19 02:00 /user/hadoop/datasets/retail_logs
-rw-r--r-- 1 hadoop hdfsadmingroup 534 2024-10-19 02:00 /user/hadoop/datasets/sample_data.csv
drwxr-xr-x 1 hadoop hdfsadmingroup 0 2024-10-19 02:00 /user/hadoop/datasets/spark
[hadoop@ip-172-31-38-45 ~]$ hdfs dfs -ls /user/hadoop/datasets/gutenberg-small/
Found 17 items
-rw-r--r-- 1 hadoop hdfsadmingroup 5878 2024-10-19 01:53 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_LincolnLetters.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 21586 2024-10-19 01:53 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_LincolnsFirstInauguralAddress.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 1653 2024-10-19 01:53 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_LincolnsGettysburgAddressGivenNovember-19-1863.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 262083 2024-10-19 01:53 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_LincolnsInauguralsAddressesandLettersSelections.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 4093 2024-10-19 01:53 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_LincolnsSecondInauguralAddress.txt
-rw-r--r-- 1 hadoop hdfsadmingroup 516298 2024-10-19 01:53 /user/hadoop/datasets/gutenberg-small/AbrahamLincoln_SpeechesandLettersofAbrahamLincoln1832-1865.txt
```

## HUE



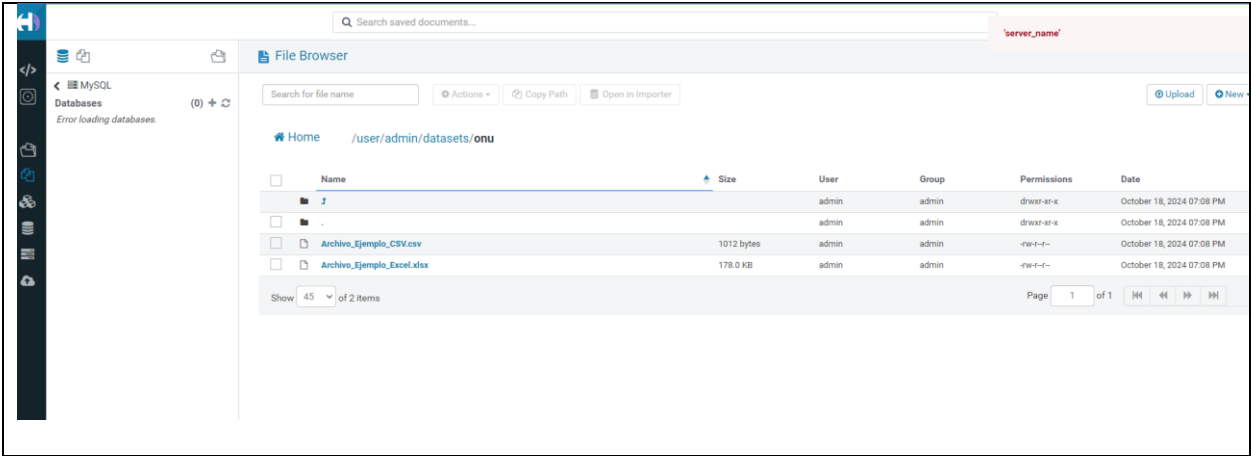
Query. Explore. Repeat.

Since this is your first time logging in, pick any username and password. Be sure to remember these, as they will become your Hue superuser credentials.

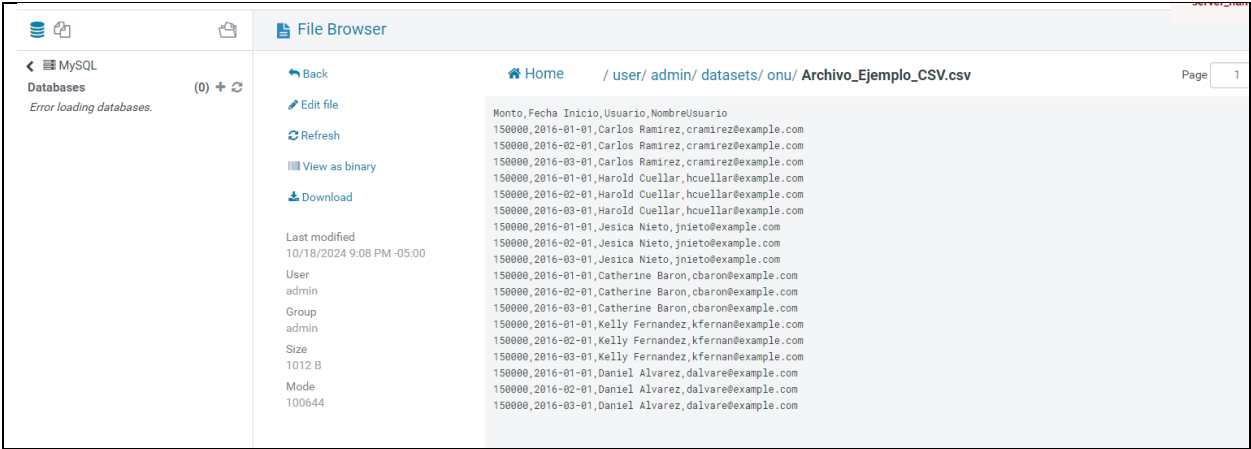
The password must be at least 8 characters long, and must contain both uppercase and lowercase letters, at least one number, and at least one special character.

Create Account

Creación de directorios




Upload file



Bucket creation

lulanotebooks Info

- Objects
- Properties
- Permissions
- Metrics
- Management
- Access Points

Bucket overview		
AWS Region US East (N. Virginia) us-east-1	Amazon Resource Name (ARN)  am:aws:s3::lulanotebooks	Creation date October 18, 2024, 20:10:13 (UTC-05:00)

**Bucket Versioning**

Edit

Versioning is a means of keeping multiple variants of an object in the same bucket. You can use versioning to preserve, retrieve, and restore every version of every object stored in your Amazon S3 bucket. With versioning, you can easily recover from both unintended user actions and application failures. [Learn more](#)

Bucket Versioning

Disabled

Multi-factor authentication (MFA) delete

An additional layer of security that requires multi-factor authentication for changing Bucket Versioning settings and permanently deleting object versions. To modify MFA delete settings, use the AWS CLI, AWS SDK, or the Amazon S3 REST API. [Learn more](#)

Disabled

**Tags (0)**

Edit

You can use bucket tags to track storage costs and organize buckets. [Learn more](#)

Key	Value
No tags associated with this resource.	