

UNIVERSIDAD EAFIT
INGENIERÍA DE SISTEMAS
ST0263 TÓPICOS ESPECIALES EN TELEMÁTICA, 2022-1

UNIDAD 3 BIG DATA

DOCUMENTACIÓN y EVIDENCIAS DE LAB 5

ESTE DOCUMENTO DEBE SER ENTREGADO POR BUZON DE ENTREGAS DE LA MATERIA
Además entregar por buzón la URL del repo github donde tiene todo de los LABs 5 y 6

NOMBRE DEL ALUMNO: Juan Felipe Lopez Gutierrez

EMAIL: jflopez@eafit.edu.co

GITHUB DEL ALUMNO PARA LOS LABS <https://github.com/JuanL-Code/st0263jflopezg>

Realizar las actividades relacionadas con el lab5 en: <https://github.com/st0263eafit/st0263-2261/tree/main/bigdata>

LAB5-EVIDENCIAS DE INSTALACIÓN DE EMR EN AWS

párrafos descriptivos, Screenshots, códigos fuente, extractos de código, scripts, urls, etc.



The screenshot shows the AWS Management Console interface for Amazon EMR. The top navigation bar includes the AWS logo, 'Services', a search bar, and user information. The left sidebar shows the navigation menu with 'Amazon EMR' selected. The main content area displays the 'Create cluster' button and a table of existing clusters. The table has columns for Name, ID, Status, Creation time (UTC-5), Elapsed time, and Normalized instance hours. One cluster is listed: 'My cluster 6.3.1.jflopezg' with ID 'j-2MJM7YNRH2KWD', status 'Waiting' and 'Cluster ready', creation time '2022-05-22 11:04 (UTC-5)', elapsed time '1 hour, 21 minutes', and normalized instance hours '0'.

Name	ID	Status	Creation time (UTC-5)	Elapsed time	Normalized instance hours
My cluster 6.3.1.jflopezg	j-2MJM7YNRH2KWD	Waiting Cluster ready	2022-05-22 11:04 (UTC-5)	1 hour, 21 minutes	0

aws

Services

Search for services, features, blogs, docs, and more

[Option+S]

N. Virginia

voclabs/user1907286-jflopezg@eafit.edu.co @ 2539-3939-5255

Amazon EMR

EMR Studio

EMR on EC2

Clusters

Notebooks

Git repositories

Security configurations

Block public access

VPC subnets

Events

EMR on EKS

Virtual clusters

Help

What's new

Clone

Terminate

AWS CLI export

Cluster: My cluster 6.3.1 jflopezg

Waiting

Cluster ready after last step completed.

Summary

Application user interfaces

Monitoring

Hardware

Configurations

Events

Steps

Bootstrap actions

Summary

Configuration details

Application user interfaces

Network and hardware

Security and access

ID: j-2MJM7YNRH2KWD

Creation date: 2022-05-22 11:04 (UTC-5)

Elapsed time: 1 hour, 22 minutes

After last step completes: Cluster waits

Termination protection: On [Change](#)

Tags: -- [View All](#) / [Edit](#)

Master public DNS: ec2-34-239-156-97.compute-1.amazonaws.com [Connect to the Master Node Using SSH](#)

Release label: emr-6.3.1

Hadoop distribution: Amazon 3.2.1

Applications: Hive 3.1.2, Hue 4.9.0, JupyterHub 1.2.0, Sqoop 1.4.7, Zeppelin 0.9.0, Tez 0.9.2, JupyterEnterpriseGateway 2.1.0, Spark 3.1.1, Livy 0.7.0, HCatalog 3.1.2

Log URI: s3://aws-logs-253939395255-us-east-1/elasticmapreduce/ [View](#)

EMRFS consistent view: Disabled

Custom AMI ID: --

Persistent user interfaces [View](#): [Spark history server](#), [YARN timeline server](#), [Tez UI](#)

On-cluster user interfaces [View](#): Not Enabled [Enable an SSH Connection](#)

Availability zone: us-east-1f

Subnet ID: [subnet-064e88531bbc08fe5](#) [View](#)

Master: **Running** 1 m4.xlarge Spot (max on-demand)

Core: **Running** 2 m4.xlarge Spot (max on-demand)

Task: --

Cluster scaling: Not enabled

Auto-termination: Terminate if idle for 1 hour

Key name: st0263-jflopezg

EC2 Instance profile: EMR_EC2_DefaultRole

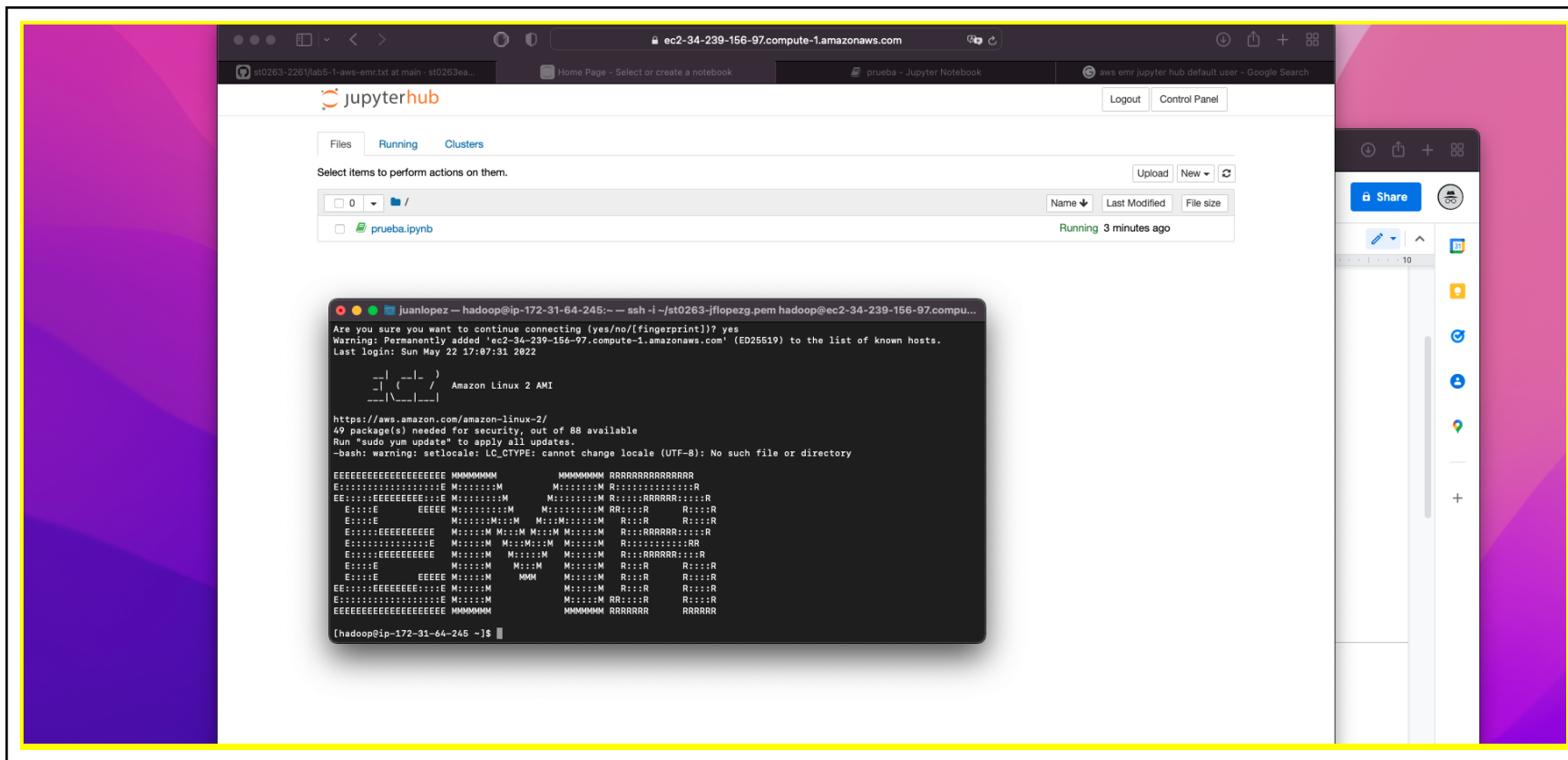
EMR role: EMR_DefaultRole

Auto Scaling role: EMR_AutoScaling_DefaultRole

Visible to all users: All [Change](#)

Security groups for Master: [sg-091f859ec5245004a](#) [View](#) (ElasticMapReduce-master)

Security groups for Core & Task: [sg-08d1d420718254551](#) [View](#) (ElasticMapReduce-slave)



Inbound rules (22)



Manage tags

Edit inbound rules

< 1 >

<input type="checkbox"/>	Name	Security group rule...	IP version	Type	Protocol	Port range	Source	Description
<input type="checkbox"/>	-	sgr-0f766d872bae9ac58	-	All TCP	TCP	0 - 65535	sg-091f859ec5245004...	-
<input type="checkbox"/>	-	sgr-0aa7db36690578e...	IPv4	Custom TCP	TCP	8443	72.21.217.0/24	-
<input type="checkbox"/>	-	sgr-013de029127953c7f	IPv4	Custom TCP	TCP	8888	0.0.0.0/0	-
<input type="checkbox"/>	-	sgr-04fe3ead33634efc4	IPv4	Custom TCP	TCP	8443	72.21.198.64/29	-
<input type="checkbox"/>	-	sgr-0e91c58d7abdddb64	IPv4	Custom TCP	TCP	8443	54.240.217.64/28	-
<input type="checkbox"/>	-	sgr-01c6c8ff914fea694	IPv4	Custom TCP	TCP	8443	54.240.217.8/29	-
<input type="checkbox"/>	-	sgr-0e72a29e360e946...	IPv4	SSH	TCP	22	0.0.0.0/0	-
<input type="checkbox"/>	-	sgr-03546ce64da51b9...	IPv4	Custom TCP	TCP	8443	72.21.196.64/29	-
<input type="checkbox"/>	-	sgr-054f8f95c78241c00	IPv4	Custom TCP	TCP	8443	54.240.217.16/29	-
<input type="checkbox"/>	-	sgr-0d527d39cd8e5c752	-	All UDP	UDP	0 - 65535	sg-091f859ec5245004...	-
<input type="checkbox"/>	-	sgr-071141ad0b4090f...	IPv4	Custom TCP	TCP	8443	54.239.98.0/24	-
<input type="checkbox"/>	-	sgr-0766d8ead4c9670...	IPv4	Custom TCP	TCP	8443	207.171.167.101/32	-
<input type="checkbox"/>	-	sgr-0fe07ce3f3ee67210	-	All UDP	UDP	0 - 65535	sg-08d1d4207182545...	-
<input type="checkbox"/>	-	sgr-0a1c730fc787f6834	IPv4	Custom TCP	TCP	8890	0.0.0.0/0	-
<input type="checkbox"/>	-	sgr-0b4e14e70dcdaa3ad	IPv4	Custom TCP	TCP	8443	54.240.217.80/29	-
<input type="checkbox"/>	-	sgr-02d0088807c125...	-	All ICMP - IPv4	ICMP	All	sg-08d1d4207182545...	-
<input type="checkbox"/>	-	sgr-016f644249d1db73c	IPv4	Custom TCP	TCP	8443	207.171.167.25/32	-
<input type="checkbox"/>	-	sgr-0246b48cdb740c4...	-	All ICMP - IPv4	ICMP	All	sg-091f859ec5245004...	-
<input type="checkbox"/>	-	sgr-088f2f4f7defda847	IPv4	Custom TCP	TCP	8443	207.171.167.26/32	-
<input type="checkbox"/>	-	sgr-0dffb2f552367980e	-	All TCP	TCP	0 - 65535	sg-08d1d4207182545...	-
<input type="checkbox"/>	-	sgr-07a207c9252653b...	IPv4	Custom TCP	TCP	9443	0.0.0.0/0	-
<input type="checkbox"/>	-	sgr-0eb4c5418511d93...	IPv4	Custom TCP	TCP	8443	207.171.172.6/32	-

LAB5-EVIDENCIAS DE COPIADO DE ARCHIVOS EN HDFS EN EL CLUSTER EMR

párrafos descriptivos, Screenshots, códigos fuente, extractos de código, scripts, urls, etc.

juanlopez — hadoop@ip-172-31-64-245:~ — ssh -i ~/st0263-jflopezg.pe...

-bash: warning: setlocale: LC_CTYPE: cannot change locale (UTF-8): No such file or directory

```
EEEEEEEEEEEEEEEEEEEE MMMMMMMM MMMMMMMM RRRRRRRRRRRRRRRR
E::::::::::::::::::::E M::::::::M M::::::::M R:::::::::R
EE:::::EEEEEEEE::::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:::::EEEEEEEE M::::M M::M M::M M::::M R::RRRRRR::::R
  E:::::E M::::M M::M:M::M M::::M R:::::::::RR
  E:::::EEEEEEEE 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:::::EEEEEEEE::::E 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-172-31-64-245 ~]\$ hdfs dfs -ls /

Found 4 items

drwxr-xr-x	-	hdfs	hdfsadmingroup	0	2022-05-22	16:17	/apps
drwxrwxrwt	-	hdfs	hdfsadmingroup	0	2022-05-22	16:21	/tmp
drwxr-xr-x	-	hdfs	hdfsadmingroup	0	2022-05-22	16:17	/user
drwxr-xr-x	-	hdfs	hdfsadmingroup	0	2022-05-22	16:17	/var

[hadoop@ip-172-31-64-245 ~]\$

Not Secure | ec2-34-239-156-97.compute-1.amazonaws.com:8888/hue/filebrowser/view=%2Fuser%2Fhadoop

Search data and saved documents...

File Browser

Search for file name

Actions

Delete forever

Upload

New

Home / user / hadoop

	Name	Size	User	Group	Permissions	Date
<input type="checkbox"/>	<div></div>		hdfs	hdfsadmingroup	drwxr-xr-x	May 22, 2022 09:17 AM
<input type="checkbox"/>	.		hadoop	hdfsadmingroup	drwxrwxrwx	May 22, 2022 10:50 AM
<input type="checkbox"/>	datasets		hadoop	hdfsadmingroup	drwxr-xr-x	May 22, 2022 10:50 AM

Show 45 of 1 items

Page 1 of 1

juanlopez — hadoop@ip-172-31-64-245:~ — ssh -i ~/st0263-jflopezg.pe...

Last login: Sun May 22 17:40:07 2022 from 181.51.33.9

Amazon Linux 2 AMI

https://aws.amazon.com/amazon-linux-2/
49 package(s) needed for security, out of 88 available
Run "sudo yum update" to apply all updates.
-bash: warning: setlocale: LC_CTYPE: cannot change locale (UTF-8): No such file or directory

EEEEEEEEEEEEEEEEEEEE MMMMMMMM MMMMMMMM RRRRRRRRRRRRRRRR
E:EEEEEEEEEEEEEEEEEEEE M:EEEEEEEEEEEEEEEEEEEE M:EEEEEEEEEEEEEEEEEEEE R:EEEEEEEEEEEEEEEEEEEE
EE:EEEEEEEEEEEEEEEEEEEE M:EEEEEEEEEEEEEEEEEEEE M:EEEEEEEEEEEEEEEEEEEE R:RRRRRRRRRRRRRRRRRRRR
E:EEEEEEEEEEEEEEEEEEEE M:EEEEEEEEEEEEEEEEEEEE M:EEEEEEEEEEEEEEEEEEEE RR:EEEEEEEEEEEEEEEEEEEE R:EEEEEEEEEEEEEEEEEEEE
E:EEEEEEEEEEEEEEEEEEEE M:EEEEEEEEEEEEEEEEEEEE M:EEEEEEEEEEEEEEEEEEEE R:EEEEEEEEEEEEEEEEEEEE R:EEEEEEEEEEEEEEEEEEEE
E:EEEEEEEEEEEEEEEEEEEE M:EEEEEEEEEEEEEEEEEEEE M:EEEEEEEEEEEEEEEEEEEE R:EEEEEEEEEEEEEEEEEEEE R:EEEEEEEEEEEEEEEEEEEE
E:EEEEEEEEEEEEEEEEEEEE M:EEEEEEEEEEEEEEEEEEEE M:EEEEEEEEEEEEEEEEEEEE R:EEEEEEEEEEEEEEEEEEEE R:EEEEEEEEEEEEEEEEEEEE
E:EEEEEEEEEEEEEEEEEEEE M:EEEEEEEEEEEEEEEEEEEE M:EEEEEEEEEEEEEEEEEEEE R:EEEEEEEEEEEEEEEEEEEE R:EEEEEEEEEEEEEEEEEEEE
EE:EEEEEEEEEEEEEEEEEEEE M:EEEEEEEEEEEEEEEEEEEE M:EEEEEEEEEEEEEEEEEEEE R:EEEEEEEEEEEEEEEEEEEE R:EEEEEEEEEEEEEEEEEEEE
E:EEEEEEEEEEEEEEEEEEEE M:EEEEEEEEEEEEEEEEEEEE M:EEEEEEEEEEEEEEEEEEEE R:EEEEEEEEEEEEEEEEEEEE R:EEEEEEEEEEEEEEEEEEEE
EEEEEEEEEEEEEEEEEEEE MMMMMMMM MMMMMMMM RRRRRRRRRRRRRRRR RRRRRRRRRRRRRRRR

[hadoop@ip-172-31-64-245 ~]\$ hdfs dfs -ls /
Found 4 items
drwxr-xr-x - hdfs hdfsadmingroup 0 2022-05-22 16:17 /apps
drwxrwxrwt - hdfs hdfsadmingroup 0 2022-05-22 16:21 /tmp
drwxr-xr-x - hdfs hdfsadmingroup 0 2022-05-22 16:17 /user
drwxr-xr-x - hdfs hdfsadmingroup 0 2022-05-22 16:17 /var
[hadoop@ip-172-31-64-245 ~]\$ hdfs dfs -ls /users
ls: '/users': No such file or directory
[hadoop@ip-172-31-64-245 ~]\$ hdfs dfs -mkdir /user/hadoop/datasets
[hadoop@ip-172-31-64-245 ~]\$

```

ip-172-31-64-245: No such file or directory: hdfs://ip-172-31-64-245.ec2.internal:8020/user/datasets
(hadoopip-172-31-64-245 -18 hadoop distcp s3://notebooksjflopezg/raw/datasets/airlines.csv /tmp/
2022-05-22 17:56:39.466 INFO tools.DistCp: Input Options: DistCpOptions(atomicCommit=false, syncFolder=false, deleteMissing=false, ignoreFailures=false, overwrite=false, append=false, useDiff=false, useRdiff=false, fromSnapshot=null, toSnapshot=null, skipCRC=
e, blocking=true, numListStatusThreads=0, maxMaps=20, mapBandwidth=0.0, copyStrategy='uniformsize', preserveStatus=[BLOCKSIZE], atomicWorkPath=null, logPath=null, sourceFileListing=null, sourcePaths=[s3://notebooksjflopezg/raw/datasets/airlines.csv], targetPa
tmp, filtersFile='null', blocksPerChunk=0, copyBufferSize=8192, verboseLog=false, directWrite=false), sourcePaths=[s3://notebooksjflopezg/raw/datasets/airlines.csv], targetPathExists=true, preserveRawXattrsfalse
2022-05-22 17:56:39.765 INFO client.RMPProxy: Connecting to ResourceManager at ip-172-31-64-245.ec2.internal/172.31.64.245:8032
2022-05-22 17:56:39.928 INFO client.AHSProxy: Connecting to Application History server at ip-172-31-64-245.ec2.internal/172.31.64.245:10200
2022-05-22 17:56:42.959 INFO tools.SimpleCopyListing: Paths (files+dirs) cnt = 1; dirCnt = 0
2022-05-22 17:56:42.959 INFO tools.SimpleCopyListing: Build file listing completed.
2022-05-22 17:56:42.961 INFO Configuration.deprecation: io.sort.mb is deprecated. Instead, use mapreduce.task.io.sort.mb
2022-05-22 17:56:42.961 INFO Configuration.deprecation: io.sort.factor is deprecated. Instead, use mapreduce.task.io.sort.factor
2022-05-22 17:56:43.040 INFO tools.DistCp: Number of paths in the copy list: 1
2022-05-22 17:56:43.074 INFO tools.DistCp: Number of paths in the copy list: 1
2022-05-22 17:56:43.093 INFO client.RMPProxy: Connecting to ResourceManager at ip-172-31-64-245.ec2.internal/172.31.64.245:8032
2022-05-22 17:56:43.093 INFO client.AHSProxy: Connecting to Application History server at ip-172-31-64-245.ec2.internal/172.31.64.245:10200
2022-05-22 17:56:43.179 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1653236289364_0004
2022-05-22 17:56:43.387 INFO mapreduce.JobSubmitter: number of splits:1
2022-05-22 17:56:43.585 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1653236289364_0004
2022-05-22 17:56:43.586 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-05-22 17:56:43.736 INFO conf.Configuration: resource-types.xml not found
2022-05-22 17:56:43.736 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2022-05-22 17:56:43.817 INFO impl.YarnClientImpl: Submitted application application_1653236289364_0004
2022-05-22 17:56:43.868 INFO mapreduce.Job: The url to track the job: http://ip-172-31-64-245.ec2.internal:20888/proxy/application_1653236289364_0004/
2022-05-22 17:56:43.869 INFO tools.DistCp: DistCp job-id: job_1653236289364_0004
2022-05-22 17:56:43.869 INFO mapreduce.Job: Running job: job_1653236289364_0004
2022-05-22 17:56:50.954 INFO mapreduce.Job: Job job_1653236289364_0004 running in uber mode : false
2022-05-22 17:56:50.955 INFO mapreduce.Job: map 0% reduce 0%
2022-05-22 17:56:58.017 INFO mapreduce.Job: map 100% reduce 0%
2022-05-22 17:56:59.020 INFO mapreduce.Job: Job job_1653236289364_0004 completed successfully
2022-05-22 17:56:59.133 INFO mapreduce.Job: Counters: 42
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=242745
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=397
  HDFS: Number of bytes written=780058
  HDFS: Number of read operations=14
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=5
  HDFS: Number of bytes read erasure-coded=0
  S3: Number of bytes read=780058
  S3: Number of bytes written=0
  S3: Number of read operations=0
  S3: Number of large read operations=0
  S3: Number of write operations=0
Job Counters
  Launched map tasks=1
  Other local map tasks=1
  Total time spent by all maps in occupied slots (ms)=241872
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=5039
  Total vcore-milliseconds taken by all map tasks=5039
  Total megabyte-milliseconds taken by all map tasks=7739904
Map-Reduce Framework
  Map input records=1
  Map output records=0

```

LAB5-EVIDENCIAS DE COPIADO DE ARCHIVOS EN AWS S3

←

→

↻

s3.console.aws.amazon.com/s3/buckets?region=us-east-1®ion=us-east-1

🔖

☆

⚙️

🖨️

👤

⋮

aws

Services

🔍 Search for services, features, blogs, docs, and more

[Option+S]

📧

🔔

❓

Global ▾

voclabs/user1907286=jflopezg@eafit.edu.co @ 2539-3939-5255 ▾

Amazon S3

×

📘 Replicate your data to any storage class to save on storage costs.

Get started

ⓘ

Amazon S3 > Buckets

▼ Account snapshot

Last updated: May 22, 2022 by Storage Lens. Metrics are generated every 24 hours. [Learn more](#)

Total storage

2.1 MB

Object count

249

Avg. object size

8.8 KB

You can enable advanced metrics in the "default-account-dashboard" configuration.

[View Storage Lens dashboard](#)

Buckets (4) [Info](#)

↻

📄 Copy ARN

Empty

Delete

Create bucket

🔍 Find buckets by name

< 1 > ⚙️

	Name ▲	AWS Region ▼	Access ▼	Creation date ▼
<input type="radio"/>	aws-logs-253939395255-us-east-1	US East (N. Virginia) us-east-1	Objects can be public	May 21, 2022, 23:02:20 (UTC-05:00)
<input type="radio"/>	jflopezgdatalake	US East (N. Virginia) us-east-1	Objects can be public	May 21, 2022, 22:17:03 (UTC-05:00)
<input type="radio"/>	notebooksjflopezg	US East (N. Virginia) us-east-1	Objects can be public	May 21, 2022, 23:27:59 (UTC-05:00)
<input type="radio"/>	st0263jflopezg	US East (N. Virginia) us-east-1	Objects can be public	May 21, 2022, 22:30:47 (UTC-05:00)

Buckets

Access Points

Object Lambda Access Points

Multi-Region Access Points

Batch Operations

Access analyzer for S3

Block Public Access settings for this account

▼ Storage Lens

Dashboards

AWS Organizations settings

Feature spotlight 3

► AWS Marketplace for S3

ChromeFileEditViewHistoryBookmarksProfilesTabWindowHelp

Identity and access management xLearner Lab - Foundational Ser xS3 Management Console x

s3.console.aws.amazon.com/s3/upload/jflopezgdatalake?region=us-east-1&prefix=raw/

awsServicesSearch for services, features, blogs, docs, and more[Option+S]

Globalvoclabs/user1907286=jflopezg@eafit.edu.co @ 2539-3939-5255

Upload succeededView details below.

Upload: status

Close

The information below will no longer be available after you navigate away from this page.

Summary

Destination

s3://jflopezgdatalake/raw/

Succeeded

36 files, 18.7 MB (100.00%)

Failed

0 files, 0 B (0%)

Files and folders

Configuration

Files and folders (36 Total, 18.7 MB)

Find by name

Name	Folder	Type	Size	Status	Error
hdi-data.csv	onu/	text/csv	9.0 KB	Succeeded	-
hdi-metadata.txt	onu/	text/plain	260.0 B	Succeeded	-
movie_reviews.csv	spark/	text/csv	367.1 KB	Succeeded	-
online-retail-dataset.csv.zip	spark/	application/zip	7.2 MB	Succeeded	-
sample_data.csv	spark/	text/csv	534.0 B	Succeeded	-
url.txt	all-news/	text/plain	46.0 B	Succeeded	-

párrafos descriptivos, Screenshots, códigos fuente, extractos de código, scripts, urls, etc.

OJO: Adjuntar la URL del bucket público donde tendrá todos los datos de entrada y salida de LAB5 y LAB6

arn:aws:s3:::jflopezgdatalake

LAB5-EVIDENCIAS DEL DESARROLLO DE LOS PROGRAMAS EN MRJOB

párrafos descriptivos, Screenshots, códigos fuente, extractos de código, scripts, urls, etc.

Punto 1

```
from mrjob.job import MRJob

class MRWordFrequencyCount(MRJob):

    def mapper(self, _, line):
        idemp, sector, salary, year = line.split(',')
        try:
            int(salary)
        except:
            pass
        else:
            yield idemp, sector

    def reducer(self, idemp, values):
        l = list(values)
        num = len(l)
        yield idemp, num

if __name__ == '__main__':
    MRWordFrequencyCount.run()
```

P2 Empresas

```

from datetime import date
from mrjob.job import MRJob

class MRWordFrequencyCount(MRJob):

    def mapper(self, _, line):
        company, price, date = line.split(',')
        try:
            float(price)
        except:
            pass
        else:
            yield company, price

    def reducer(self, company, values):
        l = list(values)
        prev_val = l[0]
        grow = True
        for i in l[1:]:
            if (prev_val > i):
                grow = False
                break
            else:
                prev_val = i
        if (grow):
            yield company, l

if __name__ == '__main__':
    MRWordFrequencyCount.run()

```

```

from datetime import date
from mrjob.job import MRJob

class MRWordFrequencyCount(MRJob):

    def mapper(self, _, line):
        company, price, date = line.split(',')
        try:
            float(price)
        except:
            pass
        else:
            yield company, (float(price), date)

    def reducer(self, company, values):
        l = list(values)
        menor = min(l, key = lambda x: x[0])
        mayor = max(l, key = lambda x: x[0])
        yield company, (menor, mayor)

if __name__ == '__main__':
    MRWordFrequencyCount.run()

```

```
from audioop import avg
from datetime import date
from mrjob.job import MRJob

class MRWordFrequencyCount(MRJob):

    def mapper(self, _, line):
        user, movie, rating, genre, date = line.split(',')
        try:
            float(rating)
        except:
            pass
        else:
            yield user, (movie, float(rating))

    def reducer(self, user, values):
        l = list(values)
        num = len(l)
        avg = sum([x[1] for x in l]) / len(l)
        yield user, (num, avg)

if __name__ == '__main__':
    MRWordFrequencyCount.run()
```

```
from audiop import avg
from datetime import date
from mrjob.job import MRJob

class MRWordFrequencyCount(MRJob):

    def mapper(self, _, line):
        user, movie, rating, genre, date = line.split(',')
        try:
            float(rating)
        except:
            pass
        else:
            yield movie, (user, float(rating))

    def reducer(self, movie, values):
        l = list(values)
        num = len(l)
        avg = sum([x[1] for x in l]) / len(l)
        yield movie, (num, avg)

if __name__ == '__main__':
    MRWordFrequencyCount.run()
```