UNIVERSIDAD EAFIT

Tópicos Especiales en Telemática

Laboratorio 3-3

Juan Manuel Gómez - Ingeniería de Sistemas

(jmgomezp@eafit.edu.co)

Profesor: Edwin Nelson Montoya

Medellín, 7 de noviembre de 2024

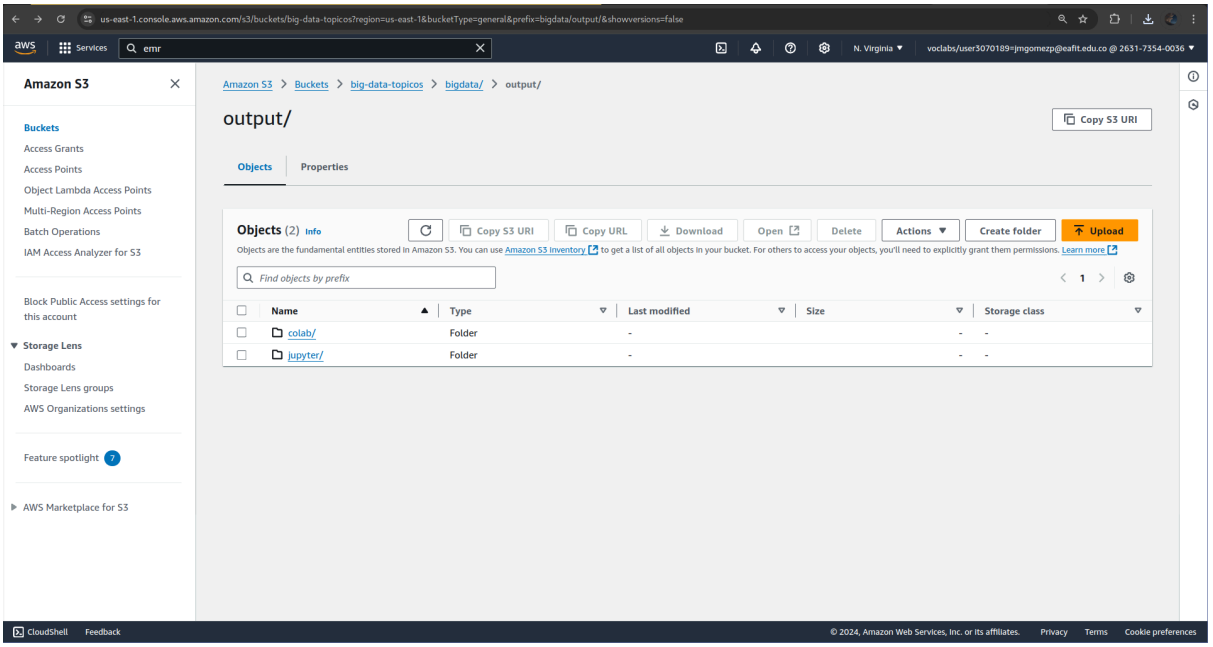
Procesamiento del dataset Casos_positivos_de_COVID-19_en_Colombia-100K.csv usando PySpark en Jupyter Hub

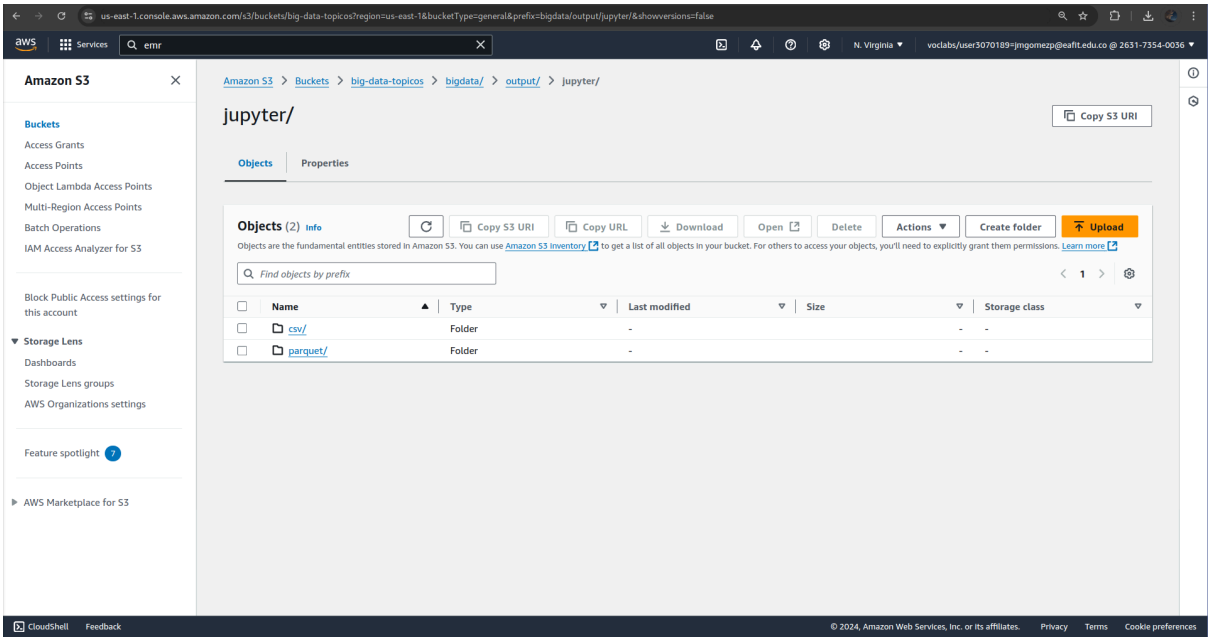
En esta ocasión se hizo una exploración de los datos referentes a los casos de COVID-19 en Colombia con los datos proporcionados por el Ministerio de Salud. Luego de haber hecho todo el procesamiento se subieron los archivos al Bucket de S3.

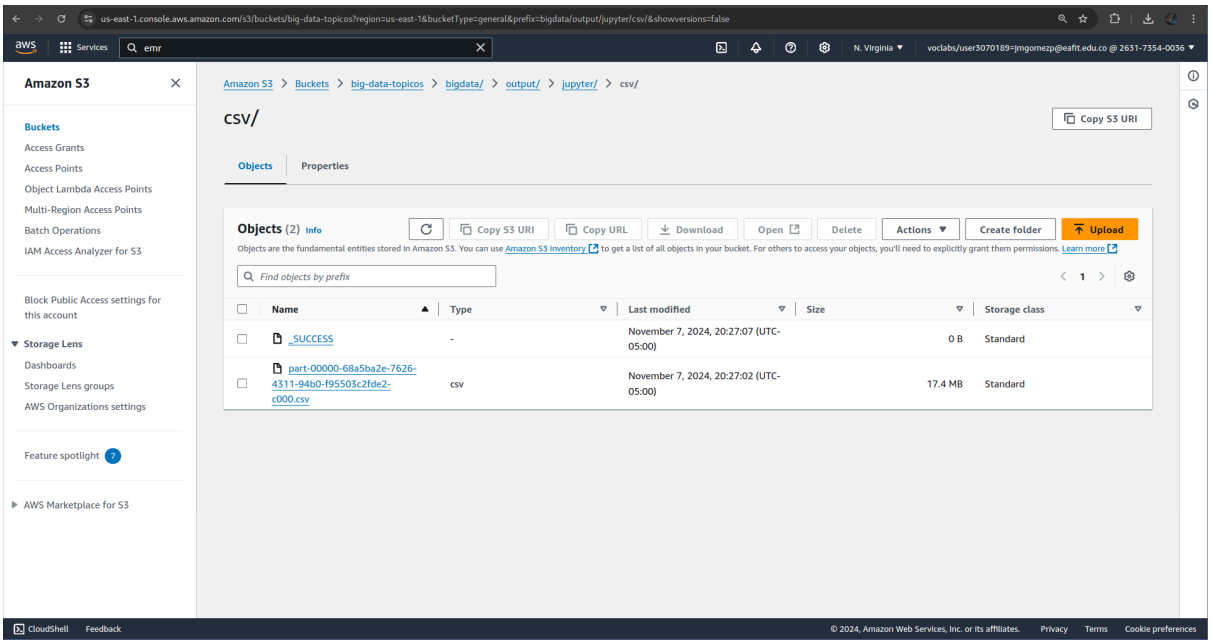
En este caso se desarrollaron los siguientes puntos:

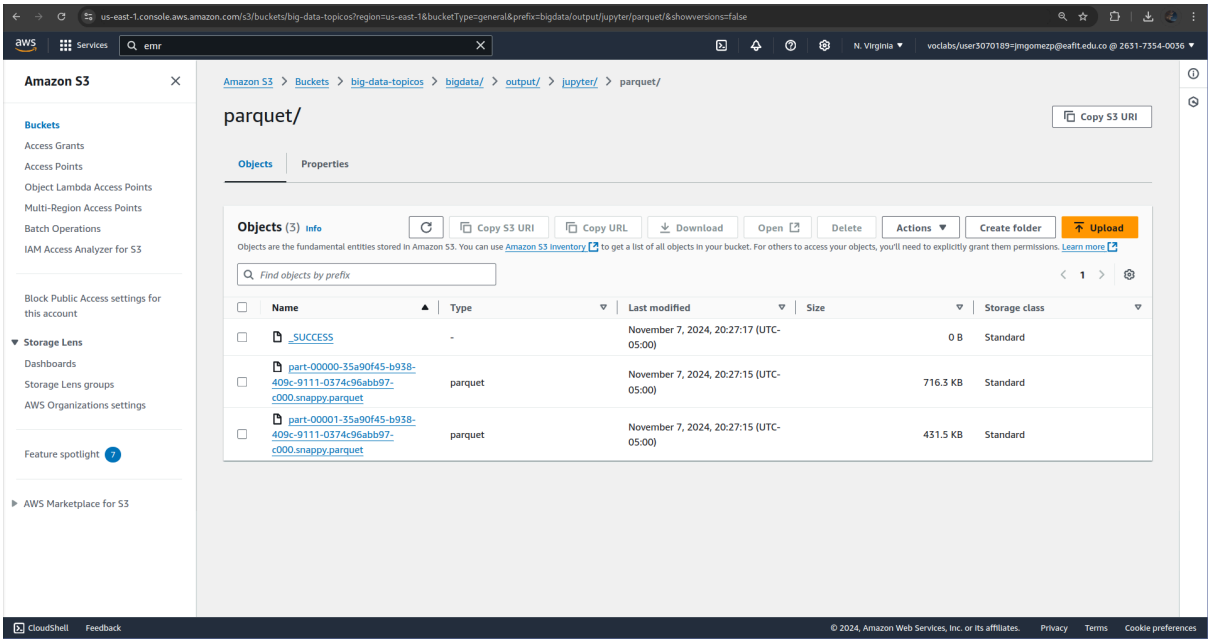
- Columnas.
- Tipos de datos.
- Seleccionar algunas columnas.
- Renombrar columnas.
- Agregar columnas.
- Borrar columnas.
- Filtrar datos.
- Ejecutar alguna función UDF sobre alguna columna creando una nueva.
- Ejecutar una función lambda sobre alguna columna creando una nueva.

En este caso se utilizó un Notebook de Jupyter Hub (cluster EMR) con el kernel de PySpark. Dentro del repositorio de GitHub se encuentra el Notebook (covid_data_processing_using_pyspark_jupyterhub.ipynb) con los resultados de este laboratorio.




us-east-1.console.aws.amazon.com/s3/buckets/big-data-topicos?region=us-east-1&bucketType=general&prefix=bigdata/output/jupyter/&showversions=false
Services
emr
N. Virginia
voclabs/user3070189=jmgomezp@eafit.edu.co @ 2631-7354-0036
Amazon S3
Buckets
Access Grants
Access Points
Object Lambda Access Points
Multi-Region Access Points
Batch Operations
IAM Access Analyzer for S3
Block Public Access settings for this account
Storage Lens
Dashboards
Storage Lens groups
AWS Organizations settings
Feature spotlight 7
AWS Marketplace for S3
Amazon S3 > Buckets > big-data-topicos > bigdata/ > output/ > jupyter/
jupyter/
Copy S3 URI
Objects
Properties
Objects (2) Info
Copy S3 URI
Copy URL
Download
Open
Delete
Actions
Create folder
Upload
Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 Inventory to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. Learn more
Find objects by prefix
Name
Type
Last modified
Size
Storage class
csv/
Folder
parquet/
Folder
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences


us-east-1.console.aws.amazon.com/s3/buckets/big-data-topicos?region=us-east-1&bucketType=general&prefix=bigdata/output/jupyter/csv/&showversions=false
Services
emr
N. Virginia
voclabs/user3070189=jmgomezp@eafit.edu.co @ 2631-7354-0036
Amazon S3
Buckets
Access Grants
Access Points
Object Lambda Access Points
Multi-Region Access Points
Batch Operations
IAM Access Analyzer for S3
Block Public Access settings for this account
Storage Lens
Dashboards
Storage Lens groups
AWS Organizations settings
Feature spotlight 7
AWS Marketplace for S3
Amazon S3 > Buckets > big-data-topicos > bigdata/ > output/ > jupyter/ > csv/
csv/
Copy S3 URI
Objects
Properties
Objects (2) Info
Copy S3 URI
Copy URL
Download
Open
Delete
Actions
Create folder
Upload
Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 Inventory to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. Learn more
Find objects by prefix
Name
Type
Last modified
Size
Storage class
_SUCCESS
-
November 7, 2024, 20:27:07 (UTC-05:00)
0 B
Standard
part-00000-68a5ba2e-7626-4311-94b0-f95503c2fde2-c000.csv
csv
November 7, 2024, 20:27:02 (UTC-05:00)
17.4 MB
Standard
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences


us-east-1.console.aws.amazon.com/s3/buckets/big-data-topicos?region=us-east-1&bucketType=general&prefix=bigdata/output/jupyter/parquet/&showversions=false
aws
Services
emr
N. Virginia
voclabs/user3070189=jmgomezp@eafit.edu.co @ 2631-7354-0036
Amazon S3
Buckets
Access Grants
Access Points
Object Lambda Access Points
Multi-Region Access Points
Batch Operations
IAM Access Analyzer for S3
Block Public Access settings for this account
Storage Lens
Dashboards
Storage Lens groups
AWS Organizations settings
Feature spotlight
7
AWS Marketplace for S3
Amazon S3 > Buckets > big-data-topicos > bigdata/ > output/ > jupyter/ > parquet/
parquet/
Copy S3 URI
Objects
Properties
Objects (3) Info
Copy S3 URI
Copy URL
Download
Open
Delete
Actions
Create folder
Upload
Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 inventory to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. Learn more
Find objects by prefix
1
Name
Type
Last modified
Size
Storage class
_SUCCESS
-
November 7, 2024, 20:27:17 (UTC-05:00)
0 B
Standard
part-00000-35a90f45-b938-409c-9111-0374c96abb97-c000.snappy.parquet
parquet
November 7, 2024, 20:27:15 (UTC-05:00)
716.3 KB
Standard
part-00001-35a90f45-b938-409c-9111-0374c96abb97-c000.snappy.parquet
parquet
November 7, 2024, 20:27:15 (UTC-05:00)
431.5 KB
Standard
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences