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

Google Colab Setup PySpark AWS

En este apartado se creó conexión de un Notebook de Google Colab a un Bucket de S3 donde hay un dataset (archivo *csv*) de ejemplo para verificar que la conexión ha sido exitosa y se pueden acceder a los recursos que hay dentro de este Bucket.

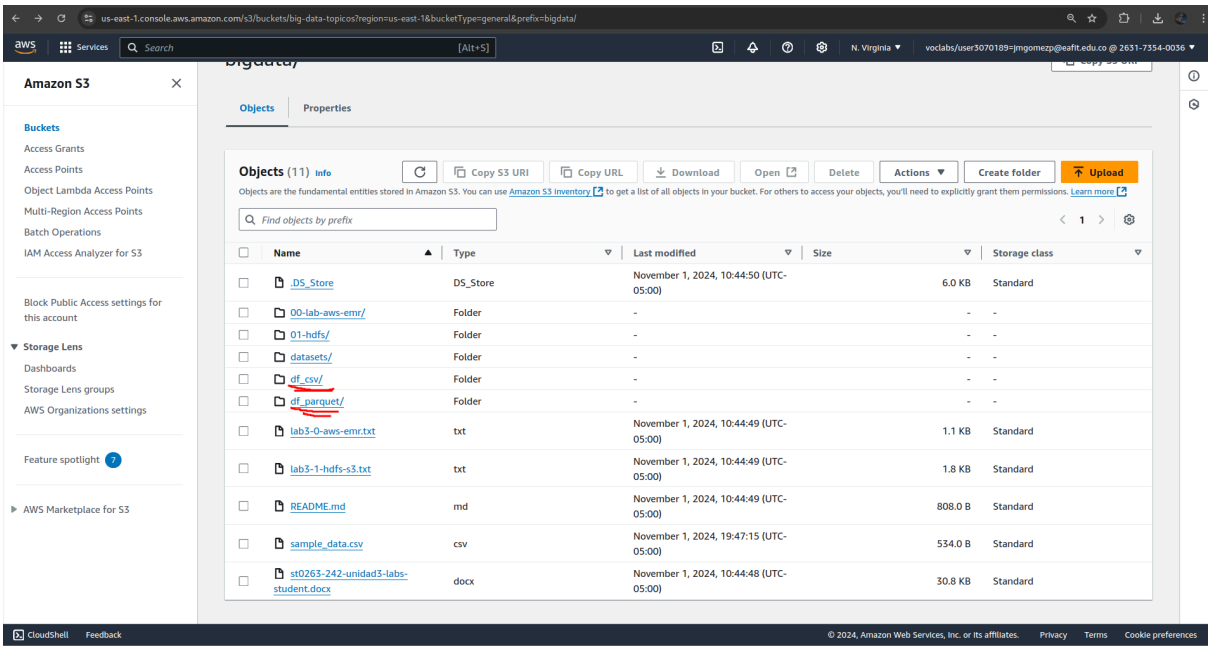
Dentro del repositorio de GitHub se encuentra el Notebook (google_colab_setup_pyspark_aws.ipynb) correspondiente a esta actividad.

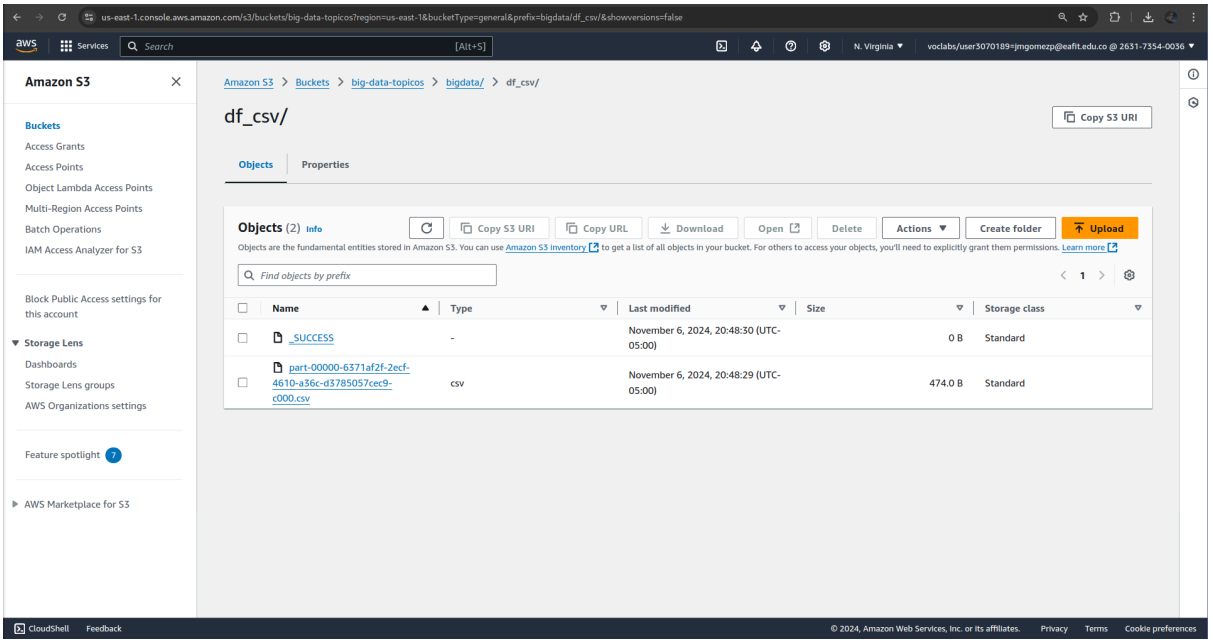
Data Processing Usando PySpark en Google Colab

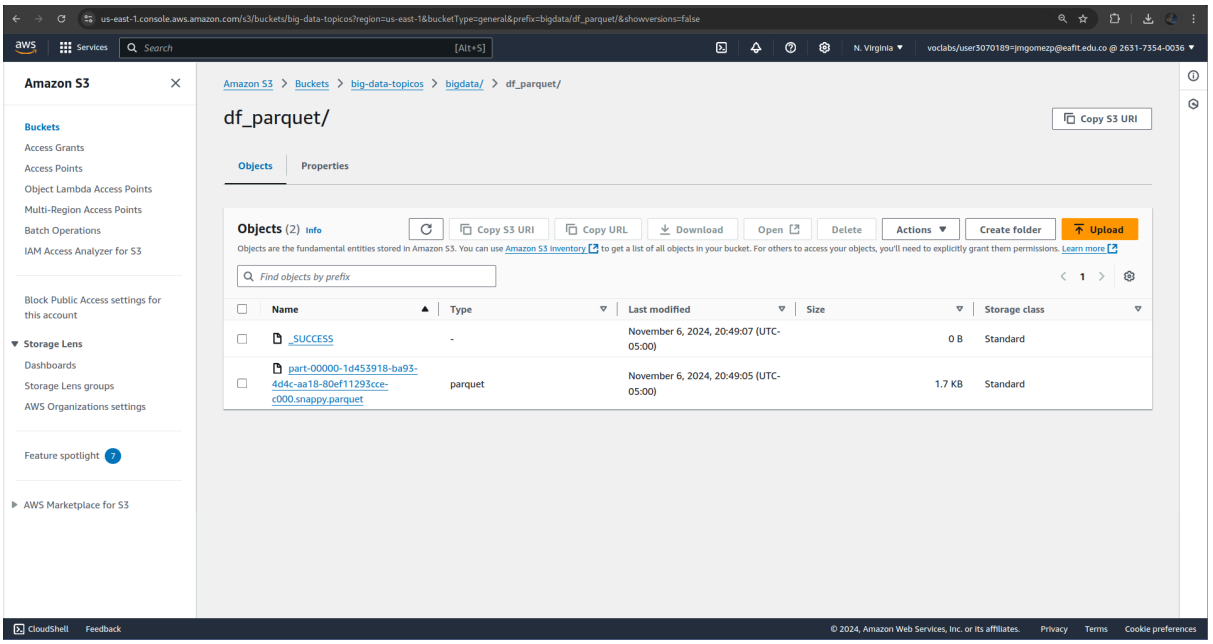
En este caso se hizo una exploración de los datos usando PySpark en Google Colab, leyendo los datos de un archivo *csv* almacenado en Google Drive. Además, se estableció una conexión a AWS para poder almacenar los datos una vez se finalice con la exploración de los datos.

Dentro del repositorio de GitHub se encuentra el Notebook (Data_processing_using_PySpark_google_colab.ipynb) correspondiente a esta actividad.

A continuación se muestran los datos guardados en S3.




us-east-1.console.aws.amazon.com/s3/buckets/big-data-topicos?region=us-east-1&bucketType=general&prefix=bigdata/df_csv/&showversions=false
Services
Search
[Alt+S]
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/ > df_csv/
df_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
1
Name
Type
Last modified
Size
Storage class
_SUCCESS
-
November 6, 2024, 20:48:30 (UTC-05:00)
0 B
Standard
part-00000-6371af2f-2ecf-4610-a36c-d3785057cec9-c000.csv
csv
November 6, 2024, 20:48:29 (UTC-05:00)
474.0 B
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/df_parquet/&showversions=false
Services
Search
[Alt+S]
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/ > df_parquet/
df_parquet/
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
1
Name
Type
Last modified
Size
Storage class
_SUCCESS
-
November 6, 2024, 20:49:07 (UTC-05:00)
0 B
Standard
part-00000-1d453918-ba93-4d4c-aa18-80ef11293cce-c000.snappy.parquet
parquet
November 6, 2024, 20:49:05 (UTC-05:00)
1.7 KB
Standard
CloudShell
Feedback
© 2024, Amazon Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences