

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

## Preguntas de negocio sobre del dataset

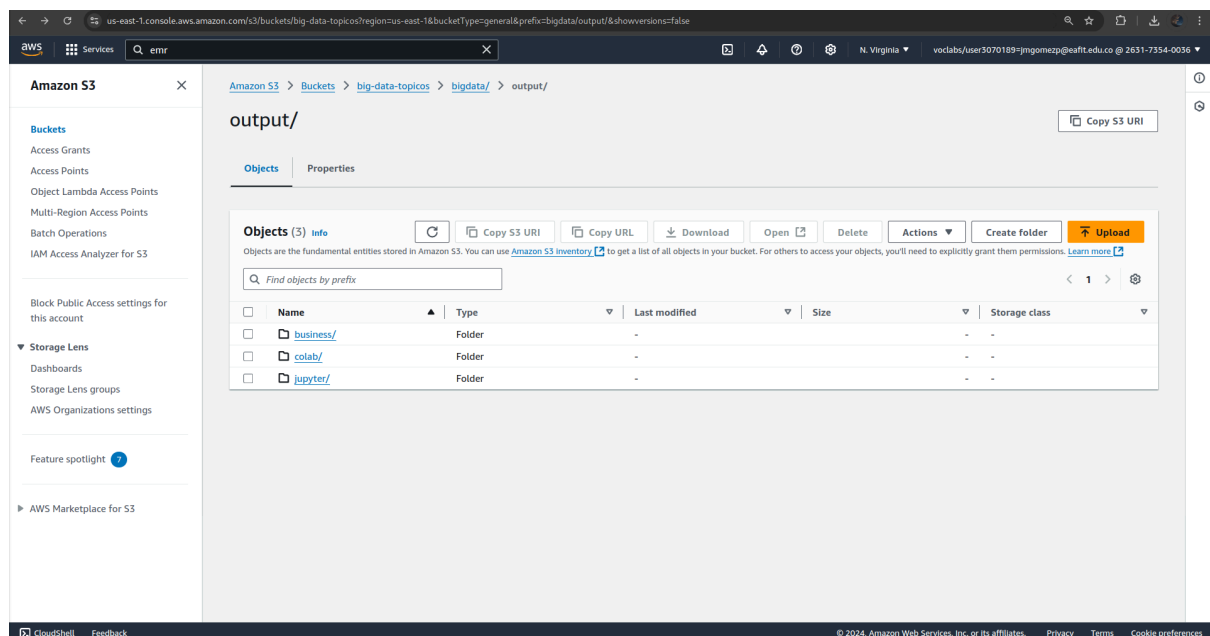
### Casos\_positivos\_de\_COVID-19\_en\_Colombia-100K.csv usando Spark en Jupyter Hub

En esta ocasión se respondieron unas preguntas de negocio sobre los casos positivos de COVID-19 dentro del país. Se utilizaron dos formas de consultar los datos, la primera fue como DataFrame y la segunda usando SparkSQL.

En este caso se dio respuesta a los siguientes puntos:

- Los 10 departamentos con más casos de covid en Colombia ordenados de mayor a menor.
- Las 10 ciudades con más casos de covid en Colombia ordenados de mayor a menor.
- Los 10 días con más casos de covid en Colombia ordenados de mayor a menor.
- Distribución de casos por edades de covid en Colombia.
- Realice la pregunta de negocio que quiera sobre los datos y respondala con la correspondiente programación en Spark. en este caso la pregunta fue: ¿cuántos casos hay por cada género en Colombia?

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 (business\_questions.ipynb) con los resultados de este laboratorio.



us-east-1.console.aws.amazon.com/s3/buckets/big-data-topics/region=us-east-1&bucketType=general&prefix=bigdata/output/business/&showversions=false

Services Search [Alt+S]

N. Virginia voclabs/user3070189+jngomezp@eafft.edu.co @ 2631-7354-0036

Amazon S3 Buckets big-data-topics bigdata/ output/ business/

business/ Copy S3 URI

Objects Properties

Objects (2) info Copy Copy S3 URI Copy URL Download Open Delete Actions Create folder Upload

Find objects by prefix

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	csv/	Folder	-	-	-
<input type="checkbox"/>	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-topics/region=us-east-1&bucketType=general&prefix=bigdata/output/business/csv/&showversions=false

Services Search [Alt+S]

N. Virginia voclabs/user3070189+jngomezp@eafft.edu.co @ 2631-7354-0036

Amazon S3 Buckets big-data-topics bigdata/ output/ business/ csv/

csv/ Copy S3 URI

Objects Properties

Objects (2) info Copy Copy S3 URI Copy URL Download Open Delete Actions Create folder Upload

Find objects by prefix

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	_SUCCESS	-	November 7, 2024, 20:41:31 (UTC-05:00)	0 B	Standard
<input type="checkbox"/>	part-00000-6c202930-ab96-477c-8693-a3255301f165-c000.csv	csv	November 7, 2024, 20:41:28 (UTC-05:00)	16.3 MB	Standard

us-east-1.console.aws.amazon.com/s3/buckets/big-data-topics?region=us-east-1&bucket=general&prefix=bigdata/output/business/parquet/&showversions=false

ServicesSearch[Alt+S]

N. Virginiavoclabs/user3070189-jngomezp@eafit.edu.co @ 2631-7354-0036

Amazon S3> Buckets> big-data-topics> bigdata/> output/> business/> parquet/

parquet/

Copy S3 URI

ObjectsProperties

Objects (5) info

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

Copy Copy S3 URI Copy URL Download Open Delete Actions Create folder Upload

	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	_SUCCESS	-	November 7, 2024, 20:41:41 (UTC-05:00)	0 B	Standard
<input type="checkbox"/>	part-00000-d89bce61-dc94-48f6-8d80-c0d450f40e25-c000.snappy.parquet	parquet	November 7, 2024, 20:41:39 (UTC-05:00)	693.6 KB	Standard
<input type="checkbox"/>	part-00001-d89bce61-dc94-48f6-8d80-c0d450f40e25-c000.snappy.parquet	parquet	November 7, 2024, 20:41:39 (UTC-05:00)	417.8 KB	Standard

CloudShellFeedback

© 2024, Amazon Web Services, Inc. or its affiliates. PrivacyTermsCookie preferences