

BUCKETS CON AWS CLI

1) APARTADO A

- Crear tres buckets en la región por defecto (con nombres similares a: primero, copia, sincro).

```
C:\Users\Mañana>aws s3api create-bucket --bucket primero-pablomr-17
{
    "Location": "/primero-pablomr-17",
    "BucketArn": "arn:aws:s3:::primero-pablomr-17"
}

C:\Users\Mañana>aws s3api create-bucket --bucket copia-pablomr-17
{
    "Location": "/copia-pablomr-17",
    "BucketArn": "arn:aws:s3:::copia-pablomr-17"
}

C:\Users\Mañana>aws s3api create-bucket --bucket sincro-pablomr-17
{
    "Location": "/sincro-pablomr-17",
    "BucketArn": "arn:aws:s3:::sincro-pablomr-17"
}
```

- Listar todos los cubos.

```
C:\Users\Mañana>aws s3 ls
2025-12-10 11:26:31 amazon-pablomr
2025-12-17 11:44:07 copia-pablomr-17
2025-12-17 11:44:06 primero-pablomr-17
2025-12-17 11:44:09 sincro-pablomr-17
```

- Ponte en una carpeta donde tengas varios archivos, mejor CSV's o JSON, sube al menos dos archivos de esa carpeta al bucket primero.

```
C:\Users\Mañana>cd "C:\Users\Mañana\Desktop\Datasets"
C:\Users\Mañana\Desktop\Datasets>aws s3 cp "C:\Users\Mañana\Desktop\Datasets\onu.csv" s3://primero-pablomr-17/
upload: ./onu.csv to s3://primero-pablomr-17/onu.csv
C:\Users\Mañana\Desktop\Datasets>aws s3 cp "C:\Users\Mañana\Desktop\Datasets\r3000.csv" s3://primero-pablomr-17/
upload: ./r3000.csv to s3://primero-pablomr-17/r3000.csv
```

- Copia los archivos del bucket primero a copia de uno en uno.

```
C:\Users\Mañana\Desktop\Datasets>aws s3 cp s3://primero-pablomr-17/onu.csv s3://copia-pablomr-17/onu.csv
copy: s3://primero-pablomr-17/onu.csv to s3://copia-pablomr-17/onu.csv
C:\Users\Mañana\Desktop\Datasets>aws s3 cp s3://primero-pablomr-17/r3000.csv s3://copia-pablomr-17/r3000.csv
copy: s3://primero-pablomr-17/r3000.csv to s3://copia-pablomr-17/r3000.csv
C:\Users\Mañana\Desktop\Datasets>aws s3 ls s3://copia-pablomr-17/
2025-12-17 11:51:40      646160 onu.csv
2025-12-17 11:51:48      463790 r3000.csv
```

- 5) Mover todos los objetos de primero a copia dentro de la carpeta datos de forma recursiva.

```
C:\Users\Mañana\Desktop\Datasets>aws s3 mv s3://primero-pablomr-17/ s3://copia-pablomr-17/datos/ --recursive
move: s3://primero-pablomr-17/r3000.csv to s3://copia-pablomr-17/datos/r3000.csv
move: s3://primero-pablomr-17/onu.csv to s3://copia-pablomr-17/datos/onu.csv
```

- 6) Listar los objetos de una copia/datos.

```
C:\Users\Mañana\Desktop\Datasets>aws s3 ls s3://copia-pablomr-17/datos/
2025-12-17 11:58:27      646160 onu.csv
2025-12-17 11:58:27      463790 r3000.csv
```

- 7) Listar recursivamente todos los objetos el cubo copia.

```
C:\Users\Mañana\Desktop\Datasets>aws s3 ls s3://copia-pablomr-17/ --recursive
2025-12-17 11:58:27      646160 datos/onu.csv
2025-12-17 11:58:27      463790 datos/r3000.csv
2025-12-17 11:51:40      646160 onu.csv
2025-12-17 11:51:48      463790 r3000.csv
```

- 8) Lista recursivamente los datos del cubo en formato legible por humanos.

¿Qué significa eso?

Formato legible por humanos significa que se imprimirá el tamaño del archivo en Bytes, KiB, MiB... para hacerlo más comprensible por el usuario.

```
C:\Users\Mañana\Desktop\Datasets>aws s3 ls s3://copia-pablomr-17/ --recursive --human-readable
2025-12-17 11:58:27  631.0 KiB datos/onu.csv
2025-12-17 11:58:27  452.9 KiB datos/r3000.csv
2025-12-17 11:51:40  631.0 KiB onu.csv
2025-12-17 11:51:48  452.9 KiB r3000.csv
```

- 9) Mostrar la información resumida, incluido el número de objetos y el tamaño total del cubo copia.

```
C:\Users\Mañana\Desktop\Datasets>aws s3 ls s3://copia-pablomr-17/ --recursive --human-readable --summarize
2025-12-17 11:58:27  631.0 KiB datos/onu.csv
2025-12-17 11:58:27  452.9 KiB datos/r3000.csv
2025-12-17 11:51:40  631.0 KiB onu.csv
2025-12-17 11:51:48  452.9 KiB r3000.csv

Total Objects: 4
Total Size: 2.1 MiB
```

- 10) Muévete a otra carpeta en local y descarga todos los objetos de copia a ella. Borra en local los archivos descargados.

```
C:\Users\Mañana\Desktop\Datasets>cd "C:\Users\Mañana\Desktop\cubiletes"
C:\Users\Mañana\Desktop\cubiletes>aws s3 cp s3://copia-pablomr-17/ . --recursive
download: s3://copia-pablomr-17/datos/r3000.csv to datos\r3000.csv
download: s3://copia-pablomr-17/r3000.csv to .\r3000.csv
download: s3://copia-pablomr-17/datos/onu.csv to datos\onu.csv
download: s3://copia-pablomr-17/onu.csv to .\onu.csv
```

```
C:\Users\Mañana\Desktop\cubiletes>dir
El volumen de la unidad C no tiene etiqueta.
El número de serie del volumen es: 1494-2731

Directorio de C:\Users\Mañana\Desktop\cubiletes

17/12/2025 12:07    <DIR>      .
17/12/2025 12:05    <DIR>      ..
17/12/2025 12:07    <DIR>      datos
17/12/2025 11:51          646.160 onu.csv
17/12/2025 11:51          463.790 r3000.csv
                2 archivos     1.109.950 bytes
                3 dirs   238.178.037.760 bytes libres

C:\Users\Mañana\Desktop\cubiletes>cd ..

C:\Users\Mañana\Desktop>rmdir "cubiletes" /s /q
```

11) Sincronizar el contenido del cubo copia con el cubo sincro.

```
C:\Users\Mañana\Desktop>aws s3 sync s3://copia-pablomr-17/ s3://sincro-pablomr-17/
copy: s3://copia-pablomr-17/datos/onu.csv to s3://sincro-pablomr-17/datos/onu.csv
copy: s3://copia-pablomr-17/r3000.csv to s3://sincro-pablomr-17/r3000.csv
copy: s3://copia-pablomr-17/onu.csv to s3://sincro-pablomr-17/onu.csv
copy: s3://copia-pablomr-17/datos/r3000.csv to s3://sincro-pablomr-17/datos/r3000.csv

C:\Users\Mañana\Desktop>aws s3 ls s3://sincro-pablomr-17/ --recursive
2025-12-17 12:11:16      646160 datos/onu.csv
2025-12-17 12:11:16      463790 datos/r3000.csv
2025-12-17 12:11:16      646160 onu.csv
2025-12-17 12:11:16      463790 r3000.csv
```

12) Sincronizar el contenido del cubo copia con la carpeta local.

```
C:\Users\Mañana\Desktop>aws s3 sync s3://copia-pablomr-17/ "C:\Users\Mañana\Desktop\cubiletes"
download: s3://copia-pablomr-17/datos/r3000.csv to cubiletes\datos\r3000.csv
download: s3://copia-pablomr-17/r3000.csv to cubiletes\r3000.csv
download: s3://copia-pablomr-17/datos/onu.csv to cubiletes\datos\onu.csv
download: s3://copia-pablomr-17/onu.csv to cubiletes\onu.csv

C:\Users\Mañana\Desktop>dir "C:\Users\Mañana\Desktop\cubiletes" /s
El volumen de la unidad C no tiene etiqueta.
El número de serie del volumen es: 1494-2731

Directorio de C:\Users\Mañana\Desktop\cubiletes

17/12/2025 12:13    <DIR>      .
17/12/2025 12:11    <DIR>      ..
17/12/2025 12:13    <DIR>      datos
17/12/2025 11:51          646.160 onu.csv
17/12/2025 11:51          463.790 r3000.csv
                2 archivos     1.109.950 bytes

Directorio de C:\Users\Mañana\Desktop\cubiletes\datos

17/12/2025 12:13    <DIR>      .
17/12/2025 12:13    <DIR>      ..
17/12/2025 11:58          646.160 onu.csv
17/12/2025 11:58          463.790 r3000.csv
                2 archivos     1.109.950 bytes

Total de archivos en la lista:
        4 archivos     2.219.900 bytes
        5 dirs   238.405.459.968 bytes libres
```

13) Haz al menos tres consultas SQL cualesquiera sobre algunos de los archivos CSV o JSON que hayas subido.

BUCKETS CON AWS CLI

1 /* Para crear un punto de referencia para escribir consultas SQL, puede mostrar los primeros 5 registros de datos de entrada ejecutando la siguiente consulta SQL: SELECT * FROM s3object s LIMIT 5 */
2 SELECT * FROM s3object s LIMIT 5

SQL Ln 2, Col 35 | Errores: 0 | Advertencias: 0 | Descargar resultados

Resultados de la consulta

Los resultados de las consultas no están disponibles después de elegir Close (Cerrar) o salir de la página. Elija Download results (Descargar resultados) para descargar una copia de los siguientes resultados de la consulta.

Estado

Se han devuelto correctamente 5 registros en 390 ms

Bytes devueltos: 1209 B

Sin procesar Formateado

dt_year	dt_date	country	región	level_of_development	greenhousegas_emissione_mtco2equivalent	total_government_revenue_proportion_of_gdp_perc	fdi_inflows_millik
2002	2002-01-01	Afghanistan	Central and Southern Asia	Least Developed			0.68
2003	2003-01-01	Afghanistan	Central and Southern Asia	Least Developed		0.12095510000000001	50.0
2004	2004-01-01	Afghanistan	Central and Southern Asia	Least Developed		0.1762145	57.8

14) Elimina los objetos de primero de uno en uno.

```
C:\Users\Mañana>aws s3 ls s3://primero-pablomr-17/
C:\Users\Mañana>aws s3 rm s3://primero-pablomr-17/onu.csv
delete: s3://primero-pablomr-17/onu.csv

C:\Users\Mañana>aws s3 rm s3://primero-pablomr-17/r3000.csv
delete: s3://primero-pablomr-17/r3000.csv
```

15) Elimina todos los objetos de forma recursiva.

```
C:\Users\Mañana>aws s3 rm s3://copia-pablomr-17/ --recursive
delete: s3://copia-pablomr-17/datos/onu.csv
delete: s3://copia-pablomr-17/onu.csv
delete: s3://copia-pablomr-17/datos/r3000.csv
delete: s3://copia-pablomr-17/r3000.csv
```

16) Elimina los cubos primero y copia.

```
C:\Users\Mañana>aws s3 rb s3://primero-pablomr-17
remove_bucket: primero-pablomr-17

C:\Users\Mañana>aws s3 rb s3://copia-pablomr-17
remove_bucket: copia-pablomr-17

C:\Users\Mañana>aws s3 ls
2025-12-10 11:26:31 amazon-pablomr
2025-12-17 11:44:09 sincro-pablomr-17
```

17) Fuerza la eliminación del cubo sincro sin vaciarlo previamente.

```
C:\Users\Mañana>aws s3 rb s3://sincro-pablomr-17 --force
delete: s3://sincro-pablomr-17/onu.csv
delete: s3://sincro-pablomr-17/datos/r3000.csv
delete: s3://sincro-pablomr-17/datos/onu.csv
delete: s3://sincro-pablomr-17/r3000.csv
remove_bucket: sincro-pablomr-17

C:\Users\Mañana>aws s3 ls
2025-12-10 11:26:31 amazon-pablomr
```

2) APARTADO B

- 1) Crea un nuevo bucket en S3 y utilizando el comando anterior, copia el archivo ventas.csv que tenemos en HDFS y obtuviste en la práctica 7.1 a él.

```
C:\Users\Mañana>aws s3api create-bucket --bucket ventas-pablo-18
{
    "Location": "/ventas-pablo-18",
    "BucketArn": "arn:aws:s3:::ventas-pablo-18"
}

C:\Users\Mañana>aws s3 cp "C:\Users\Mañana\Desktop\Datasets\ventas.csv" s3://ventas-pablo-18/
upload: Desktop\Datasets\ventas.csv to s3://ventas-pablo-18/ventas.csv
```

- 2) ¿Cómo podrás desde HIVE lanzar las mismas consultas que hicimos en los apartados D y E de la práctica 7.1, pero en esta ocasión sobre ficheros en S3?

Importante antes de nada definir los ajustes del dataset

Configuración de entrada

Ruta
s3://ventas-pablo-18/ventas.csv

Tamaño
104.9 KB (107396.0 B)

Formato
 CSV
 JSON
 Apache Parquet

CSV delimitador
 Coma
 Tabulador
 Personalizado

Delimitador CSV personalizado

Excluir la primera línea de CSV datos
Habilite esta configuración si CSV contiene una fila de encabezado.

Compresión
 Ninguno
 GZIP
 BZIP2

BUCKETS CON AWS CLI

- i. Mostrar registros con cantidades mayores o iguales a cero. A partir de la consulta anterior, mostrar registros precio mayor a cero. A partir de la consulta anterior, mostrar solamente los registros con algún valor en el campo CustomerID.

```
SELECT * FROM s3object s WHERE CAST(s.Quantity AS DECIMAL) >= 0 AND  
CAST(s.UnitPrice AS DECIMAL) > 0 AND s.CustomerID IS NOT NULL
```

536370	POST	POSTAGE	3	01/12/2010 8:45	18	12583	France
536392	22827	RUSTIC SEVENTEEN DRAWER SIDEBOARD	1	01/12/2010 10:29	165	13705	United Kingdom
536403	POST	POSTAGE	1	01/12/2010 11:27	15	12791	Netherlands
536527	POST	POSTAGE	1	01/12/2010 13:04	18	12662	Germany
536540	C2	CARRIAGE	1	01/12/2010 14:05	50	14911	EIRE
536779	BANK CHARGES	Bank Charges	1	02/12/2010 15:08	15	15823	United Kingdom
536835	22655	VINTAGE RED KITCHEN CABINET	1	02/12/2010 18:06	295	13145	United Kingdom
536840	POST	POSTAGE	1	02/12/2010 18:27	18	12738	Germany
536852	POST	POSTAGE	1	03/12/2010 9:51	18	12686	France
536858	POST	POSTAGE	2	03/12/2010 10:36	40	13520	Switzerland
536861	POST	POSTAGE	3	03/12/2010 10:44	18	12427	Germany
536967	POST	POSTAGE	1	03/12/2010 12:57	18	12600	Germany
536974	POST	POSTAGE	2	03/12/2010 13:59	18	12682	France
536983	POST	POSTAGE	1	05/12/2010 14:30	18	12712	Germany
536990	POST	POSTAGE	1	05/12/2010 15:14	28	12793	Portugal
537022	POST	POSTAGE	2	05/12/2010 15:45	28	12725	Italy
537026	POST	POSTAGE	2	05/12/2010 16:35	18	12395	Belgium

- ii. ¿Cuántos registros tiene la última consulta?

```
SELECT COUNT(*) FROM s3object s WHERE CAST(s.Quantity AS DECIMAL) >= 0 AND  
CAST(s.UnitPrice AS DECIMAL) > 0 AND s.CustomerID IS NOT NULL
```

1 1712
2 |