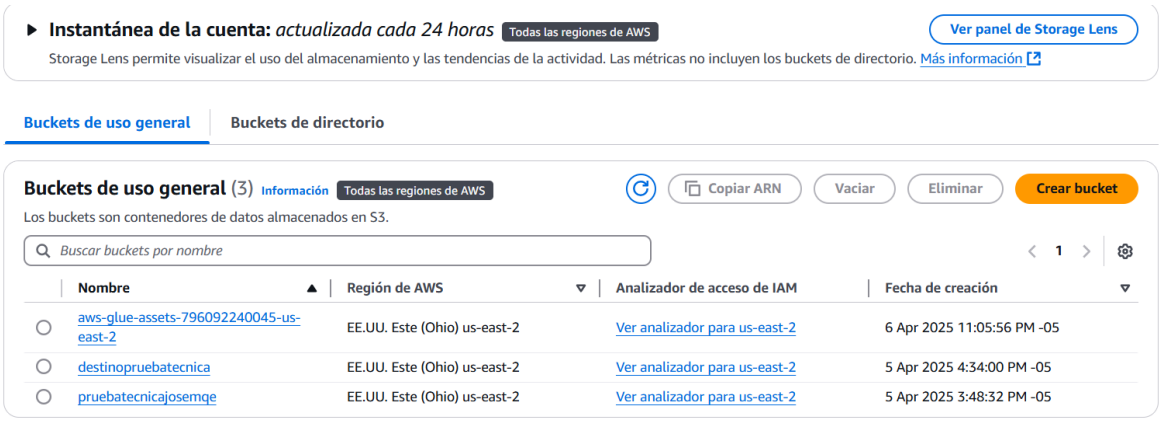
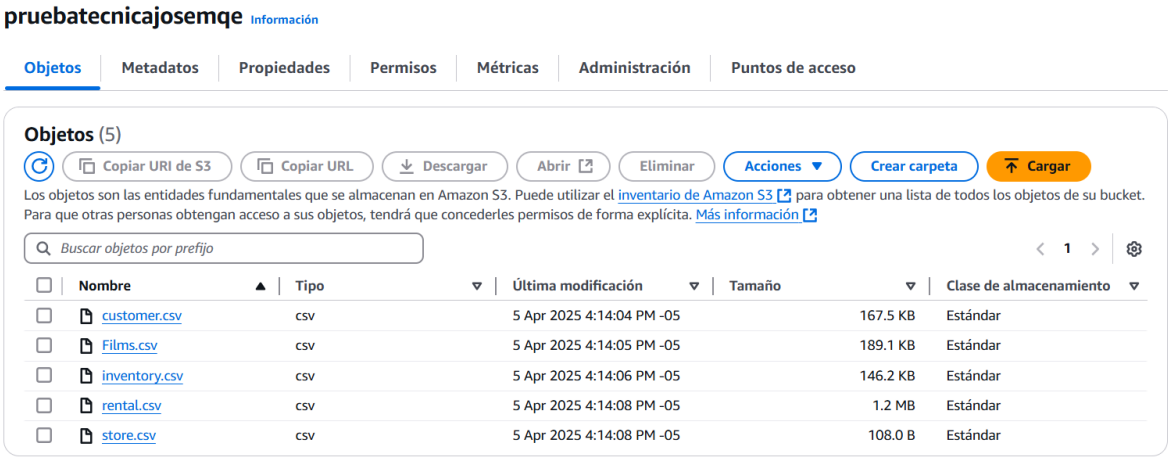
Proceso de ETL en AWS.

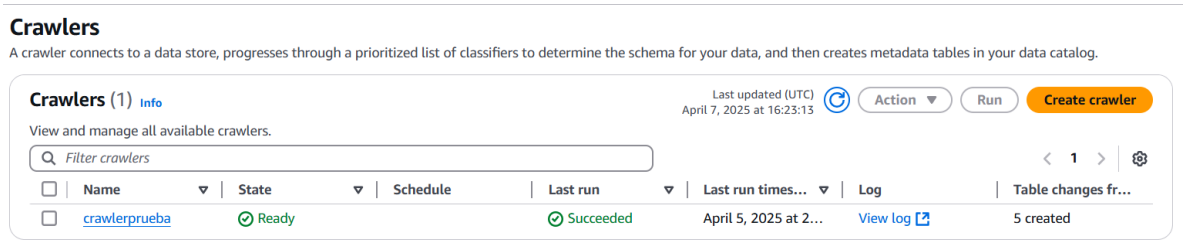
Paso 1. Creación del Bucket en S3: **pruebatecnicajosemqe**



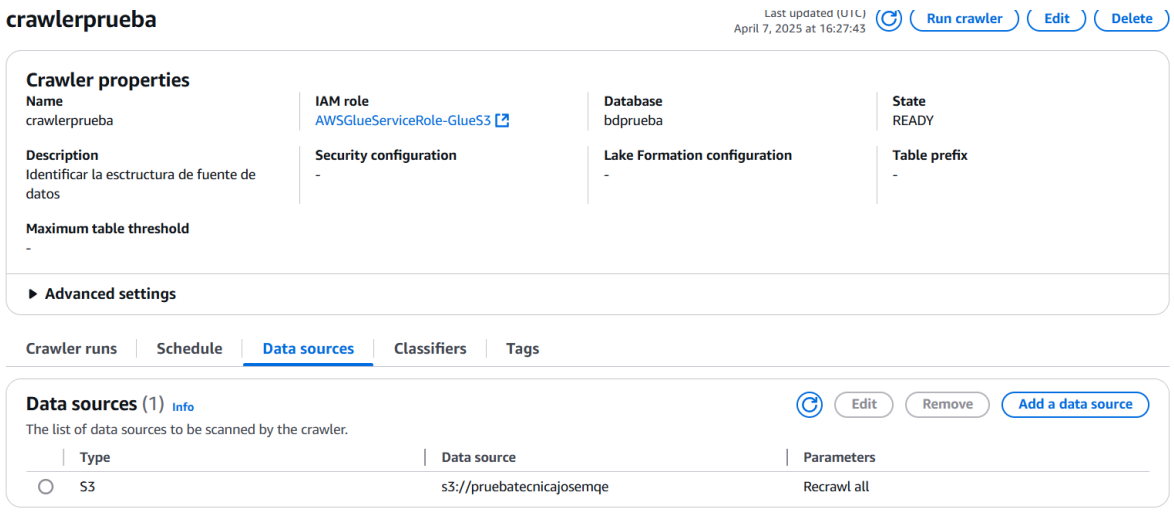
Paso 2. En el Bucket creado se cargan las tablas en formato CSV.



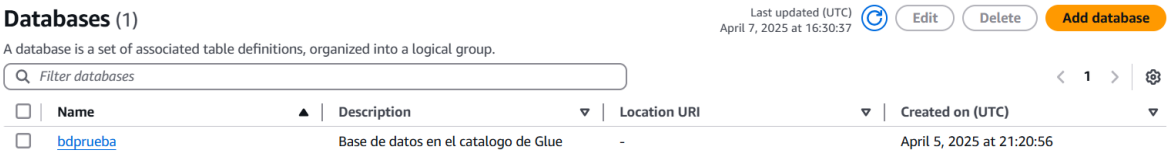
Paso 3. Se realizar la creación de Crawler: crawlerprueba este se encargará de identificar el formato de los datos para cargar en el Glue Data Catalog.



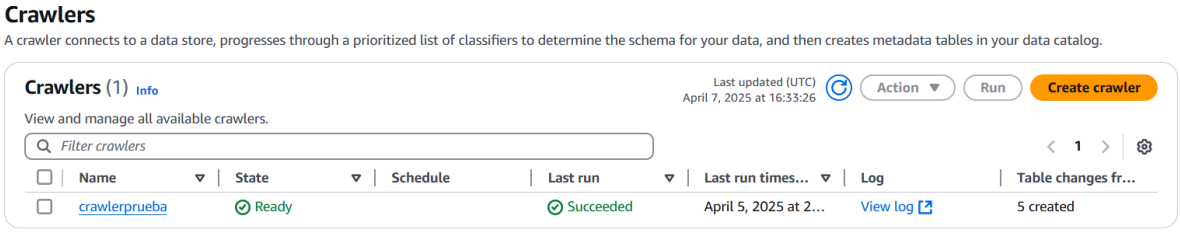
Paso 4. Al Crawler se le asigna el Bucket creado en el S3



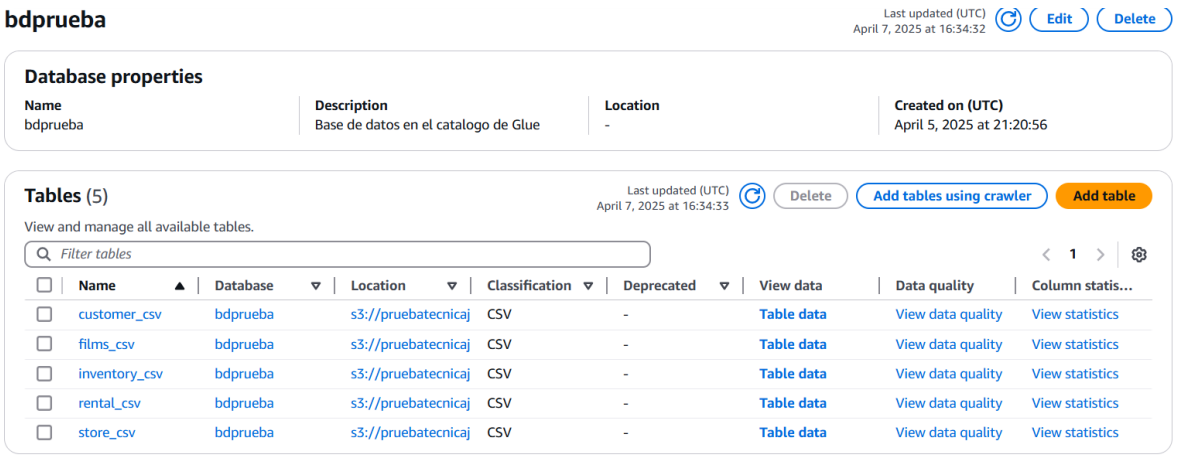
Paso 5. Se crea la base de datos: **bdprueba** esta base de datos también se le asigna al Crawler creado.



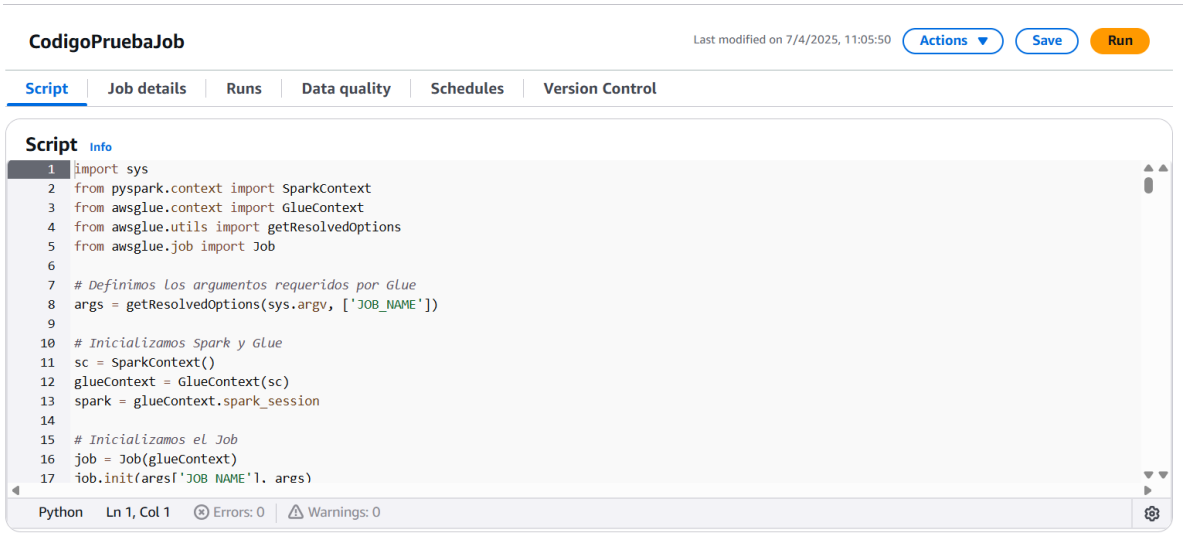
Paso 6. Se ejecuta el Crawler



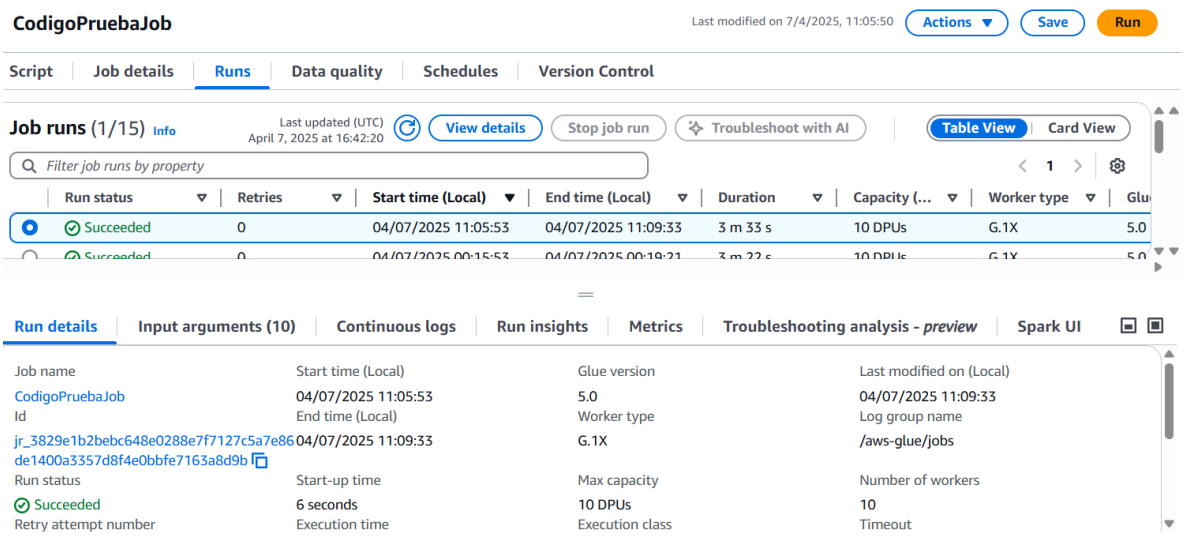
Paso 7. El Crawler crea en la base de datos con los data frames correspondiente a cada archivo CSV.



Paso 8. Se realizar la creación del Script en el ETL Job, allí copiamos el código realizado en Google Colab utilizando Python como lenguaje de programación y framework de procesamiento de datos PySpark, El código de adecuarse para cargar las tablas que fueron creadas en la base de datos de S3.



Paso 9. Se compila el código, comprobando que no existan errores de compilación.



Paso 10. Se obtienen y analizan los resultados.

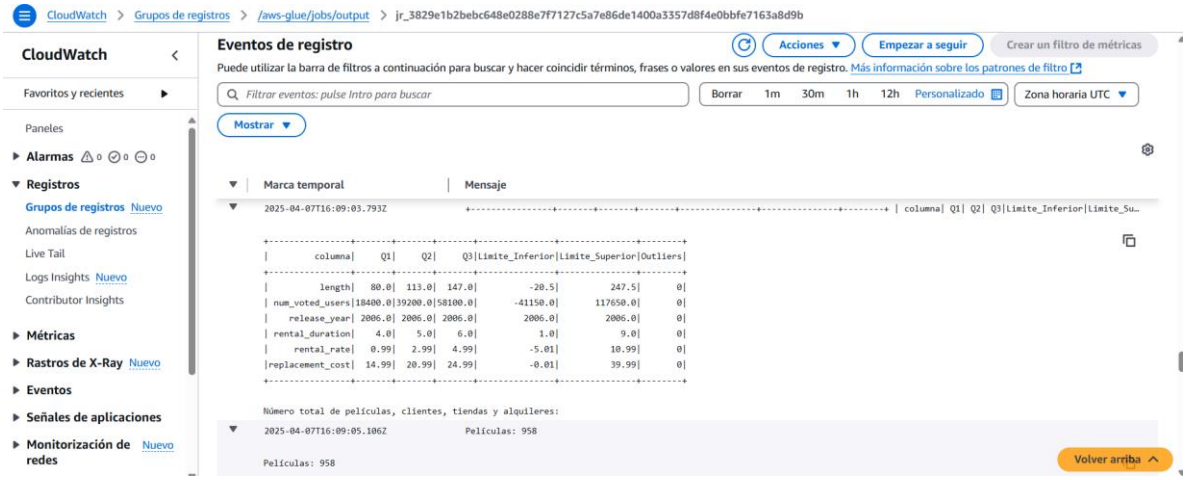


Diagrama de flujo proceso en AWS.