

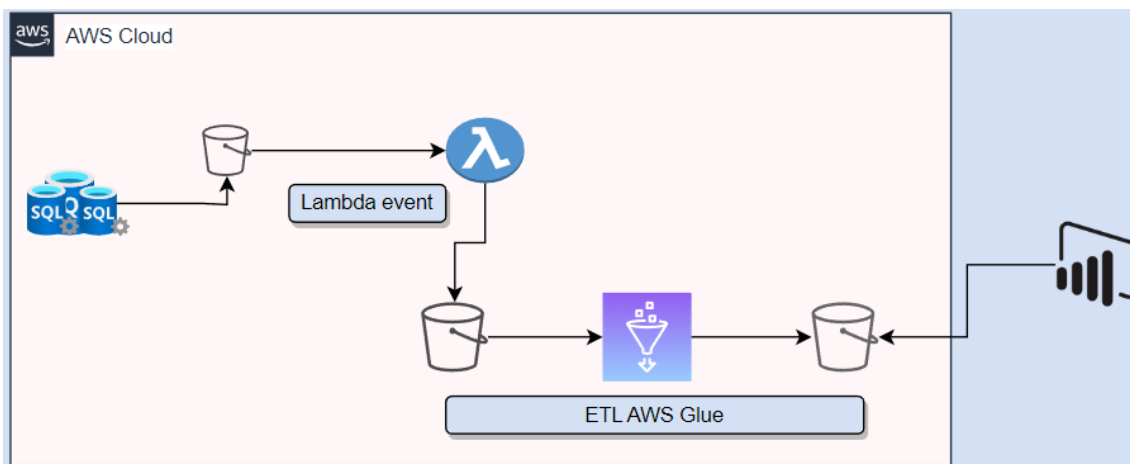
Explicación arquitectura:

3-Sql server exporta los files xlsx a s3.

2- Se crea una función lambda para transformar los xlsx en csv y se guardan en otro directorio el cual será nuestro stg.

3-Una vez cargado stg se configuran etls de aws Glue, se realizan las transformaciones necesarias y se cargan en un s3 de publicación en formato parquet.

4- Power BI se conecta a s3, se restringe el acceso a los usuarios según rol con RLS.



Porque utilizamos

S3:

Las razones principales son que s3 escala según la necesidad de la data es decir no se va a interrumpir la carga porque se quede sin almacenamiento, tiene baja latencia y se puede integrar con múltiples sistemas fuera de aws en nuestro caso por ejemplo Power BI además de ser muy amigable para administrar los objetos.

Lambda: Es sin servidor, se activa por eventos en este caso por los files que llegan a s3, es flexible ya que te permite desarrollar en varios lenguajes.

Aws Glue: Es una herramienta pensada para desarrollo de ETL, nos da la facilidad de orquestar por eventos al igual que lambda, podemos utilizar tecnologías de big data como por ejemplo Spark para realizar las transformaciones de negocio.