12/12/2023

Bases de datos en AWS

Se hacen a través del servicio RDS

09/01/2024

Redshift es un cluster de RDS. Necesita al menos un par de nodos y consue bastante. Características:

* Almacenamiento en columnas:
* Escalabilidad

Athena

Analizar datos almacenados en S3 utilizando el estándar de SQL

Para la practica crear una carpeta de datos y otra de resultadso. En datos esta el .csv original y en resultados lo que genera

16/01/2024

AWS Glue

Tiene ETL visuales, transformaciones, notebooks y Jobs (tanto visual como jupyter notebook)

Los elementos que tiene el Glue:

* Crawler: sobre un esquemapuede inferir un esquema de datos dado un csv en un S3
* Data catalog: una vez el crawler infiere la estructura la almacena en un data catalog, desde ahí se puede utilizar Athena para hacer consultas o quickview
* Job
  + Notebook
  + Visual:

Ejemplo practico

Crear un bucket con una carpeta de resultados (para athena)

23/01/2024

AWS Glue

* Crawler
* Jobs
  + Visual ETL
* Data catalog
  + Lambda
* Workflow