# Arquitectura AWS Prueba Fondeadora



26- octubre-2020



#### **Servicio**







- Para este reto se seleccionaron estos tres servicios para generar un proceso ETL desde Bases de Datos Athena hacia Redshift.
- Athena es de Tipo MySQL usando InnoDB com motor
- La instancia EC2 de del tipo Ubuntu con Python 3.8 para ejecutar el script principal
- Redshift es una instancia estándar.
- El proceso involucra crear una base de datos de nombre fondeadoradb con 3 tablas (envios, logística y rutas)
- La instancia EC2 es el intermedio entre los servicios y actua como intermediario para la transformación y almacenaje temporal
- Redshift es el punto de entrega con las tablas: rutas y cargo



#### **Servicio**







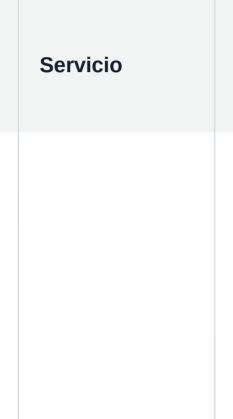
- Para ejecutar:
  - De la carpeta auroradb, generar las tablas y colocar los datos semilla en MySQL
  - Modificar el acceso de MySQL de manera local en el script: auroramysql-to-redshift.py
  - Ejecutar en Python 3.8 el script.
  - En caso de que Redshift no sea disponible, el script se puede ajustar para que los resultados se entreguen a PosgreSQL



## **Estructura Base**









, .........

#### Datos "Raw"

Los datos se consideran en crudo, sin tratamiento ni limpieza



Amazon EC2

#### Datos "Staging"

Se realiza la carga de las fuentes de datos y se realizan operaciones entre ellos para enriquecerlos



Amazon Redshift

#### Presentación

Los datos se tienen en forma tabular para su aprovechamiento en ciencia de datos o inteligencia de negocio.



Servicio	Amazon Athena	Amazon EC2	Amazon Redshift
	Datos "Raw"  Base de datos: fondeadoradb  Tablas: logistica envios rutas	Datos "Staging"  Aloja el script: auroramysql_to_r edshift.py  y las tablas temporales CSV	Presentación Receptor de los datos tratados en las tablas: - rutas - cargo

