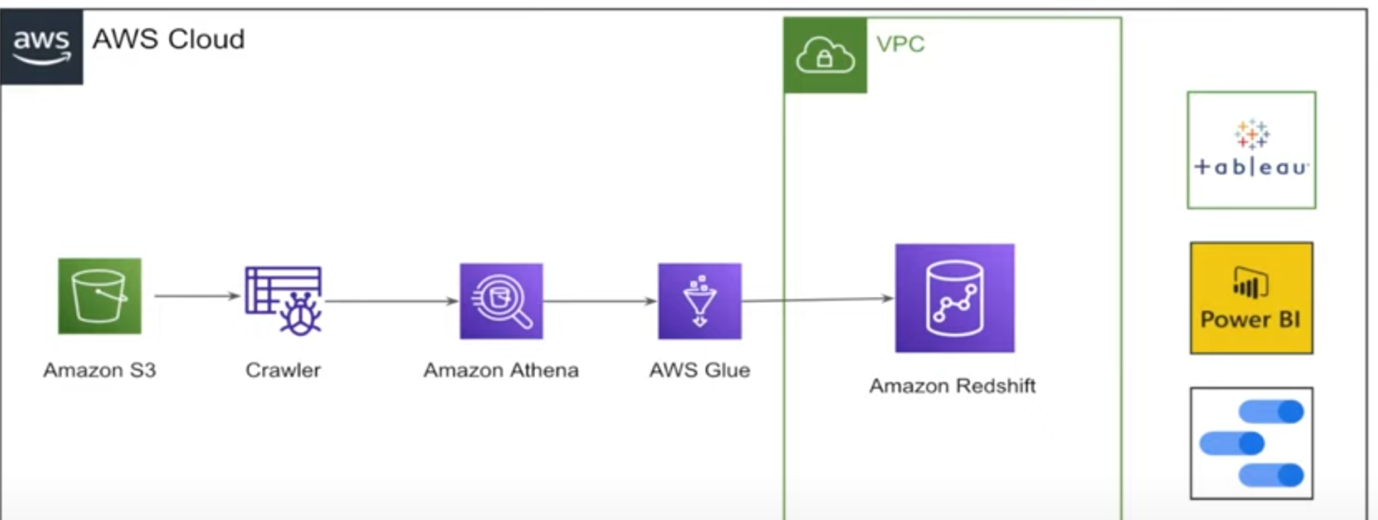
**COVID19 DATA ANALSIS PROJECT AWS ARCHITECTURE**



* **VPC & Security Setup**
  + Created a **Security Group** within the default VPC to allow only **SSH traffic from the internet**.
  + Configured **IAM user** and **S3 bucket** for access control and data storage.
* **Data Ingestion**
  + Used **AWS Open Data** to source public datasets.
  + Stored raw data files (e.g., CSV, JSON) in **Amazon S3**.
* **Schema Discovery**
  + Ran an **AWS Glue Crawler** to infer schema and populate the **Glue Data Catalog**.
* **Ad-hoc Analysis**
  + Used **Amazon Athena** to run SQL queries directly on S3 data for exploration.
* **Data Transformation**
  + Applied ETL logic using **AWS Glue** to clean and structure the data.
* **Data Warehousing**
  + Deployed an **Amazon Redshift** cluster within the VPC.
  + Loaded transformed data into Redshift for analytics.
* **Optional Visualization**
  + Enabled support for downstream **BI tools** (e.g., QuickSight) for data visualization.