COVID-19 Data Analytics Pipeline Architecture

Ingestion & Storage

AWS S3
Data Lake for
Raw Data Storage

→ Data Upload →

AWS Glue Crawler
Automated Schema Inference and Data Cataloging

→ Schema Inference →

Processing Pipeline

Amazon Athena
Serverless SQL Querying
for Data Exploration

→ Querying →

Python ETL Scripts
Data Transformation
and Preprocessing

→ Transformation →

AWS S3
Storage for Transformed Data

→ Data Transfer →

Warehousing & Visualization

Amazon Redshift
Scalable Data Warehousing

→ Data Load →

BI Tools (e.g.,
PowerBI, Tableau)
Data Visualization and Reporting

Orchestration

AWS Glue Jobs
Orchestrating
ETL Processes

ETL Control

Trigger ETL Jobs