

Data Provisioning for (COVID-19) AI/ML Models

Leading Biotechnology Organization

Business Problems

- Build **AI/ML-based models** to carry out multiple projects on COVID-19 diagnosis
- Capture data** around various subject areas and from multiple data sources
- Enable **data pre-processing and mapping** to develop predictive models

Solution(s) & Value Delivered

- ▶ **Acquired data** from primary data sources to client landing zones
- ▶ **Transferred data** and loaded into **AWS Redshift**
- ▶ Performed **virus scan, data quality checks** and **PHI/PII classification**
- ▶ **Updated metadata repository** and **executive dashboard** to show newly loaded data
- ▶ **Ensured availability of fresh data** to build models by mapping the automated pipeline to **ingest regular COVID data**



AI/ML
models for predictive analytics

Interactive dashboards

Regular checks
for data quality

Metadata repository
for fresh data