Enterprise Data Lake for Medical Data Processing Company Architecture

Ingestion Layer   Storage Layer   Processing Layer   Serving Layer

Streaming Data

Amazon Kinesis Data Streams

Amazon Kinesis Data Firehose

AWS Transfer Family for SFTP

AWS Database Migration Service

AWS Glue ETL

AWS Glue crawler

Glue Data Catalog tables

AWS Lake Formation

AWS Glue crawler

Apache Hudi

Apache Airflow

AWS KMS

Amazon EMR Hadoop Spark HIVE Presto

AWS S3 Raw

AWS S3 Curated

Amazon Glacier

Amazon Kinesis Data Analytics

Real-time queries

Amazon Athena

Amazon SageMaker

Users

Tableau Dashboards

Amazon Aurora cluster PostgreSQL

Amazon Aurora DB Cluster

Availability Zone a — Primary Instance (M)

Availability Zone b — Aurora Replica (R)

Availability Zone c — Aurora Replicas (R) (R)

Reads Writes Writes Reads Reads Reads

Data Copies   Data Copies   Data Copies

Cluster Volume

Application Layer

Web Server Layer

Load Balancer

Users Users Users Users Users Users