Data Lake Solution Architecture

Proposed Solution for Medical Data Processing Company

aws + MongoDB + APACHE

Ingestion Layer

Medical Facility Data Sources

Databases

Customer APIs

Documents (XML, TXT, CSV)

Connectors

Connectors

Secure FTP

kafka
Continous Data Streaming

Connector

Storage Layer

Metadata

MongoDB
Atlas Clusters

Archiving

Amazon S3 **Buckets**

MongoDB
Atlas Data Lake

Processing Layer

MongoDB
Data Federation

MongoDB
Atlas SQL Interface

Connector

APACHE **Spark**

Serving Layer

Ad-hoc and/or scheduled interactive querying capability using SQL

Tableau

Microsoft **Power BI**

PyTorch

scikit learn

AI / Machine Learning

Users

Scheduling

Apache Airflow