

Data Sources



Databases



Flat Files



APIs



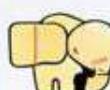
SFTP



Ingestion Layer



Apache **nifi**



Apache **Sqoop**



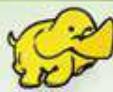
Apache **kafka**

Batch & Real-Time Ingestion



Storage Layer

Raw Zone



Apache HDFS

Curated Zone



Apache Hive Metastore

Trusted Zone



Apache Ranger

Processing Layer



Batch, Streaming & Ad-Hoc Processing



Serving Layer



BI & Reporting



Machine Learning



Data APIs

Analytics, ML Models, & APIs

Data Lake Architecture for Medical Data Processing Company