Data Lake Value Proposition

Medical Data Processing Company

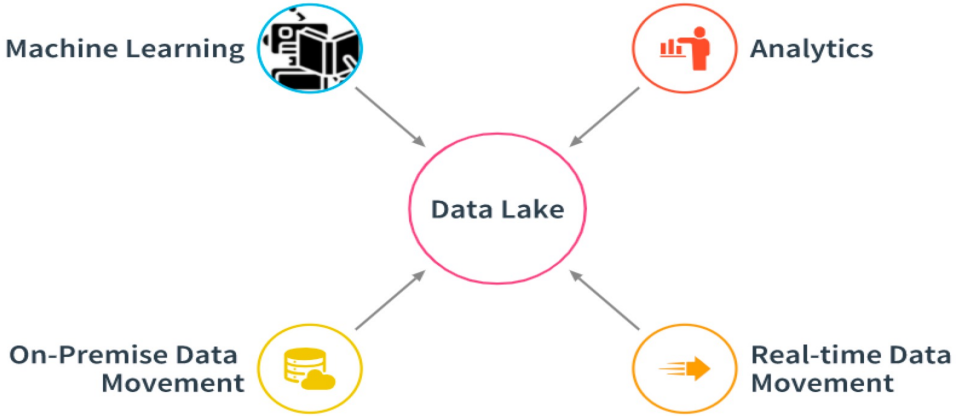
Mohit Bansal

Agenda

- What is a Data Lake
- Components of a Data Lake
- Data Lake vs Data Warehouse
- Business Value of Data Lake Solution
- Proposed Data Lake Architecture for Medical Data Processing system

What is a Data Lake

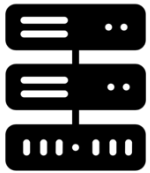
Big Data represent large amounts of data. A Data Lake represents where and that data is stored



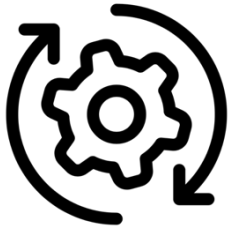
Components of Data Lake

- Ingestion - tools to ingest data in Data Lake
- Processing - tools to process the data in Data Lake
- Storage - Store the data at scale in a data lake. Mostly leveraging Object Storage solutions
- Governance - Broad term, including management of the data lake of the whole. Includes availability, monitoring, access control policies

Ingest



Store



Process

Govern & Secure

Data Warehouse

- Schema designed at the beginning
- Fastest query results using higher cost storage
- Relational from transactional systems, operational databases, and line of business applications
- Used for historical analytics, visualizations, BI
- Collects similar data from multiple resources

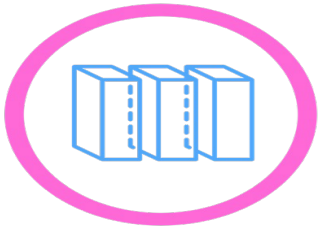
Data Lake

- schema designed at the end
- Query results getting faster using low-cost storage
- Non-relational and relational from IoT devices, web sites, mobile apps, social media, and corporate applications
- Used for predictive analytics, machine learning
- Connecting various types of data from a wide variety of sources

Business Value of Data Lake

Data Lake promotes organizations where

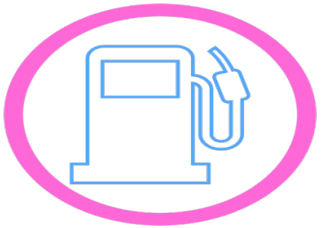
- Innovation is enabled
- Real-time, integrated-data Analytics is possible
- Data silos are broken, which allows new collaboration and interactions
- Self-service access instead of weeks months of work to get what they need
- Data, even new data, can be dynamic, easy to incorporate, and quick to access



Break Silos

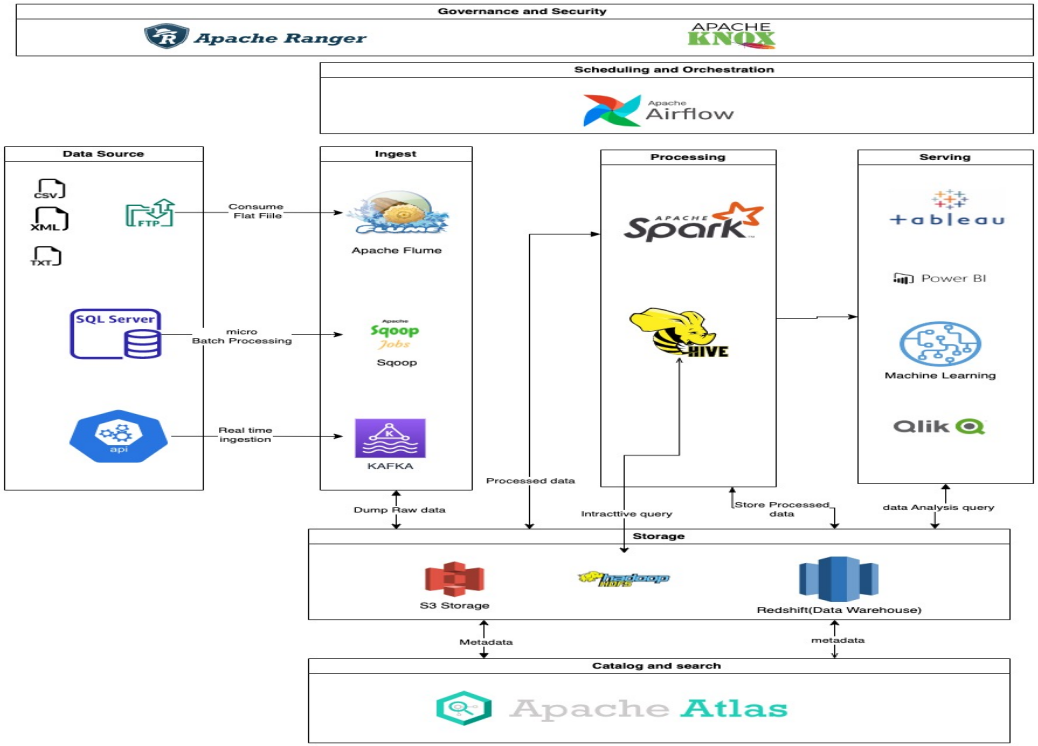
Innovation

Speed



Self Service

Data Lake Architecture



THANK YOU