

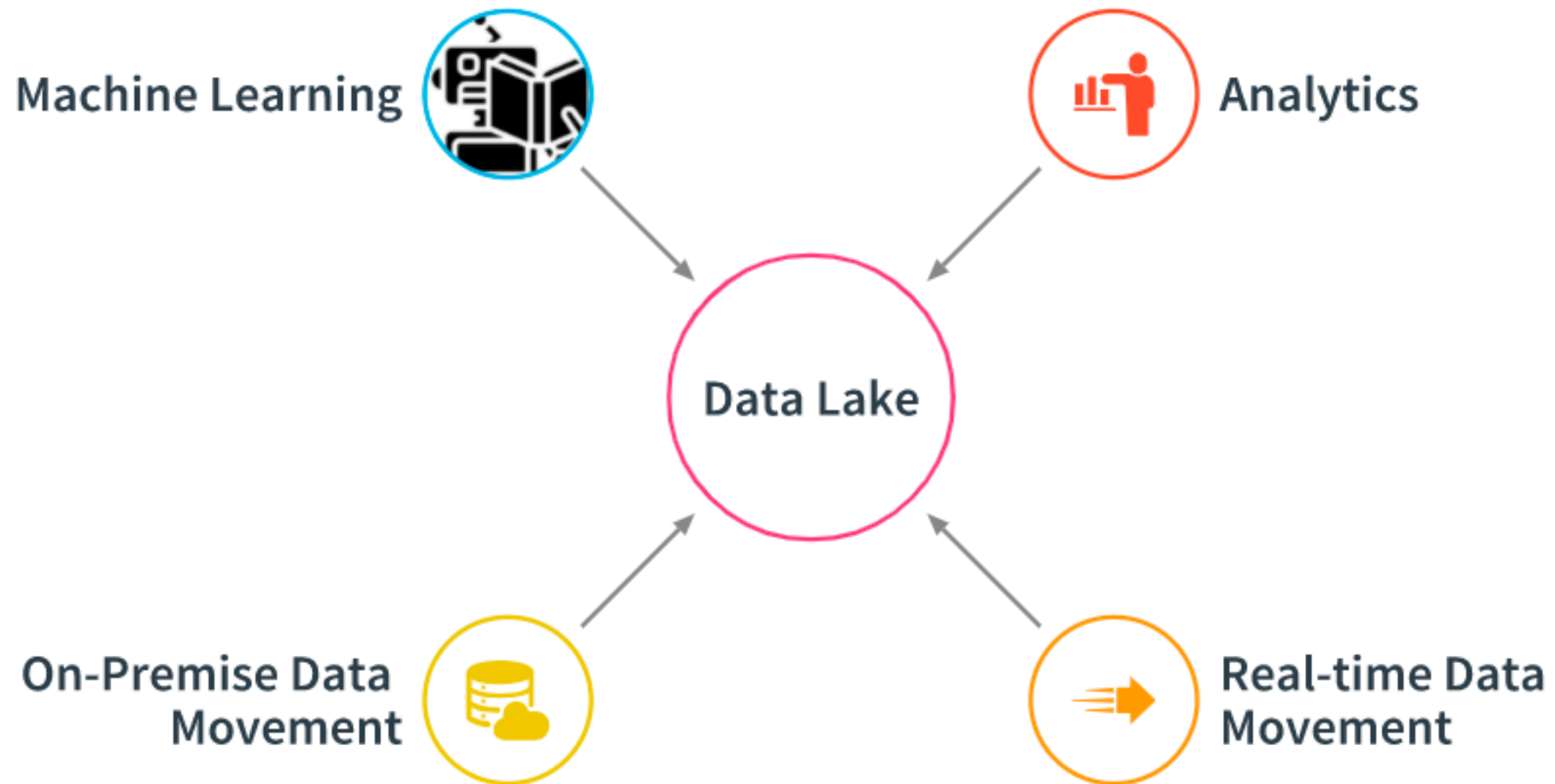
# **Data Lake Value Proposition**

**Hui Ren**

# 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

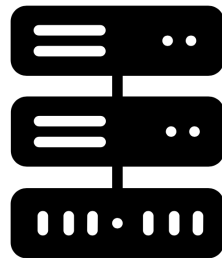
# What is a Data Lake?



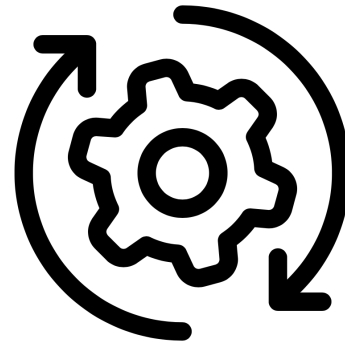
# Components of Data Lake



Ingest



Store



Process

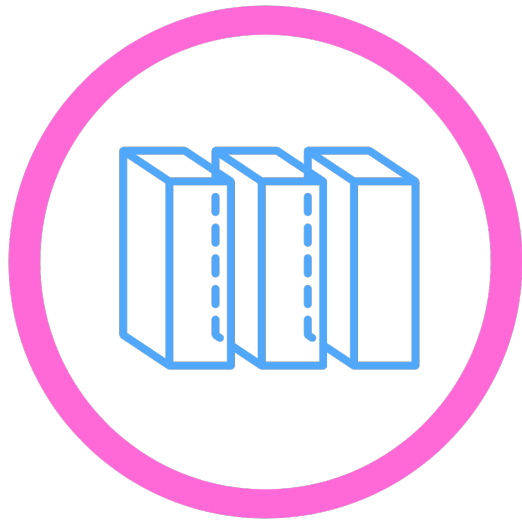


Govern & Secure

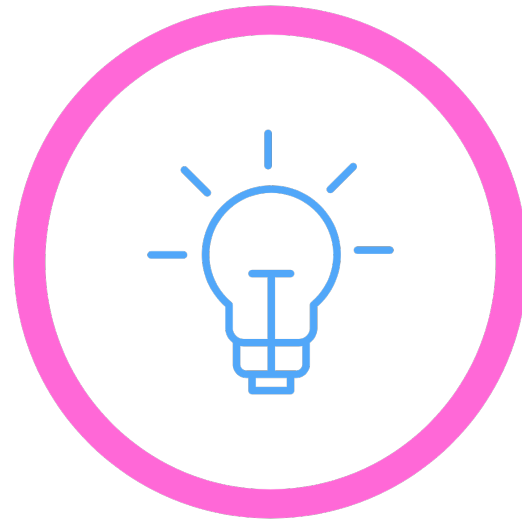
# Data Lake Vs Data Warehouse

Characteristics	Data Lake	Data Warehouse
Data	Non-relational and relational from IoT devices, web sites, mobile apps, social media, and corporate applications	Relational from transactional systems, operational databases, and line of business applications
Schema	Written at the time of analysis (schema-on-read)	Designed prior to the DW implementation (schema-on-write)
Price/ Performance	Query results getting faster using low-cost storage	Fastest query results using higher cost storage
Data Quality	Any data that may or may not be curated (ie. raw data)	Highly curated data that serves as the central version of the truth
Users	Data scientists, Data developers, and Business analysts (using curated data)	Business analysts
Analytics	Machine Learning, Predictive analytics, data discovery and profiling	Batch reporting, BI and visualizations

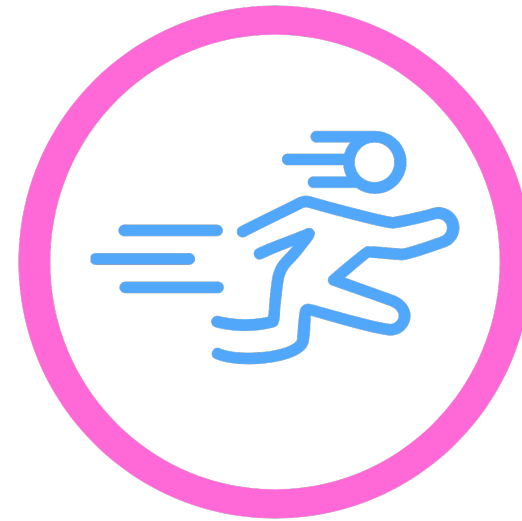
# Business Value of Data Lake



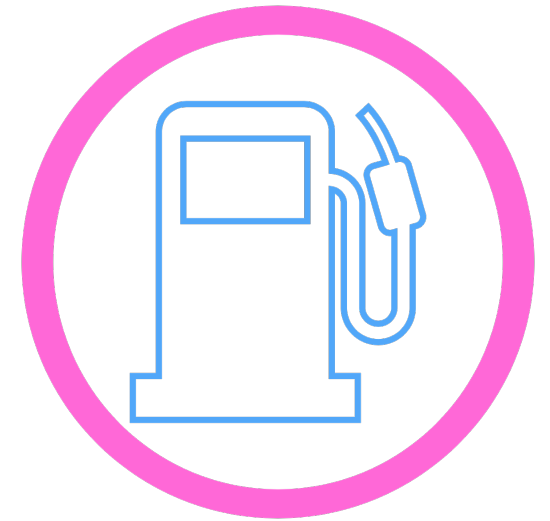
Break Silos



Innovation

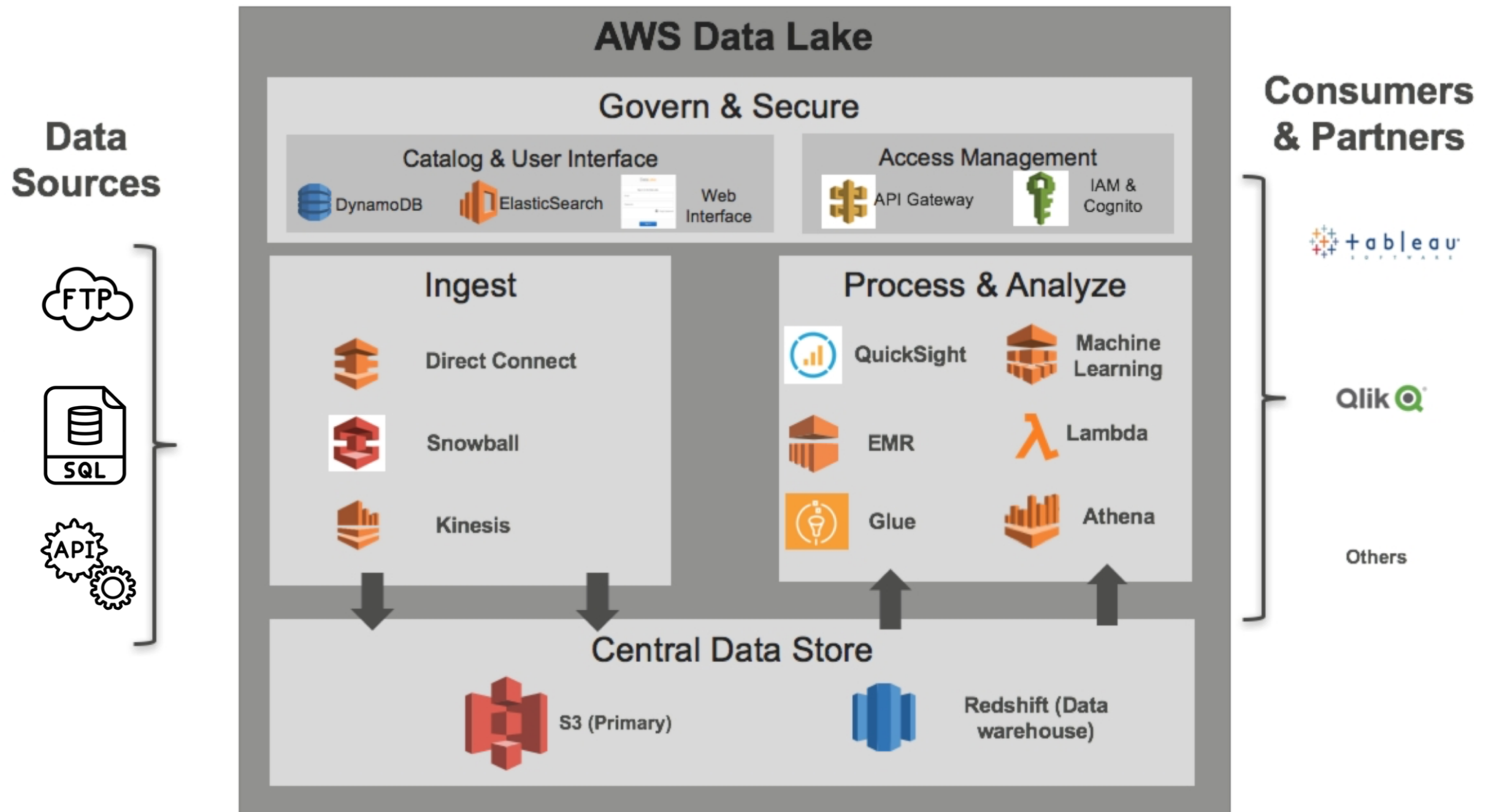


Speed



Self Service

# Data Lake Architecture



**Thank You!**