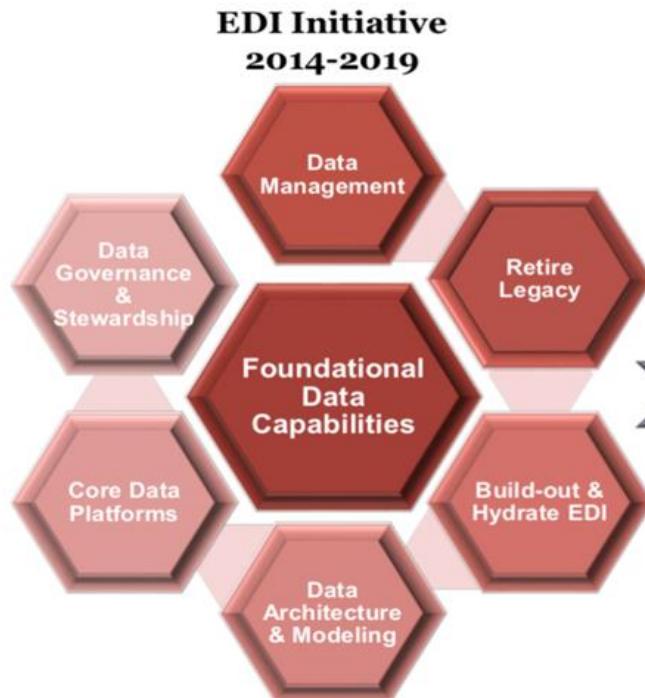


iSolvers Inc

Enterprise Data Lake

The Data Strategy is evolving ...



Enable



Enterprise Data Lake Definition

What is a data lake?

A repository for large quantities and varieties of data, both structured and unstructured.

Data generalists/
programmers can tap
the stream data for
real-time analytics.

The lake can serve as a staging area for the data warehouse,
the location of more carefully
“treated” data for reporting
and analysis in batch mode.

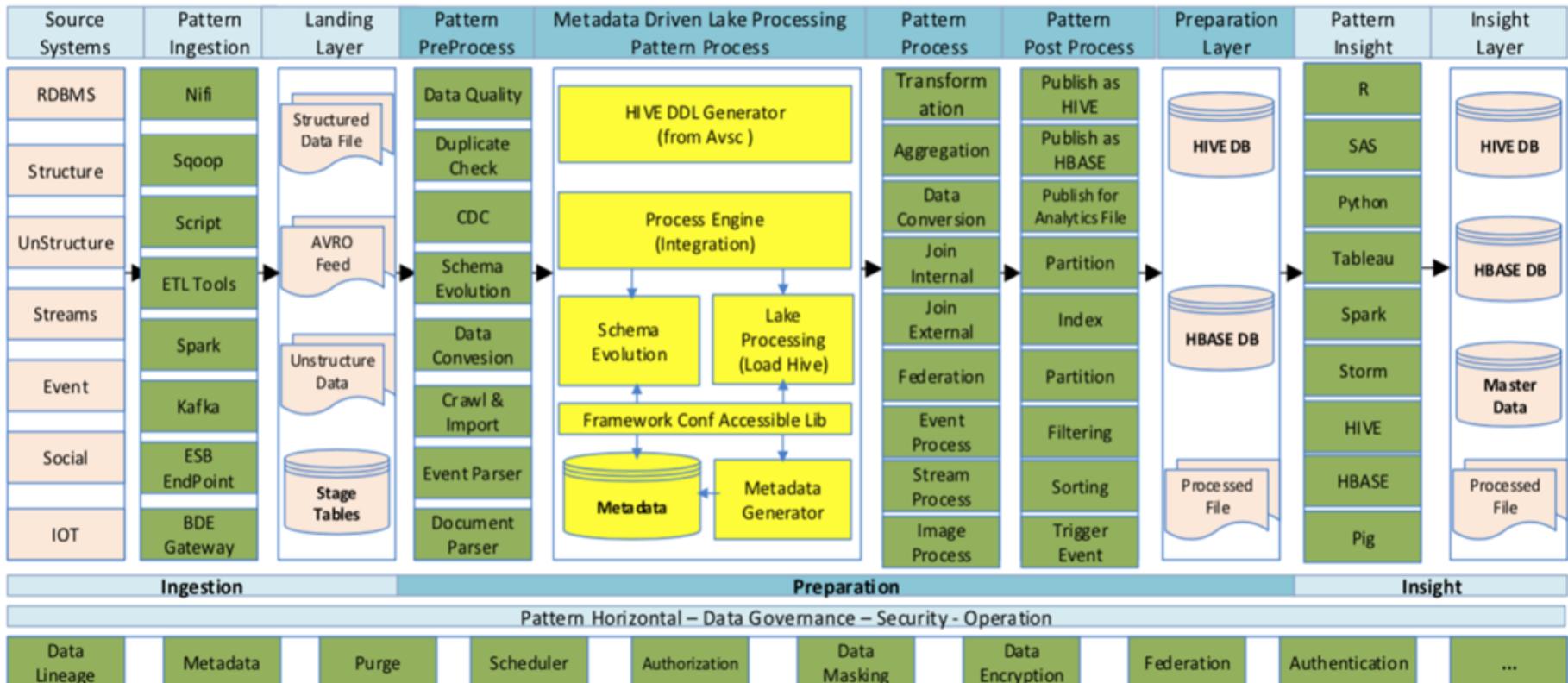
Data scientists
use the lake for
discovery and
ideation.

The data lake accepts input from various sources and can preserve both the original data fidelity and the lineage of data transformations. Data models emerge with usage over time rather than being imposed up front.

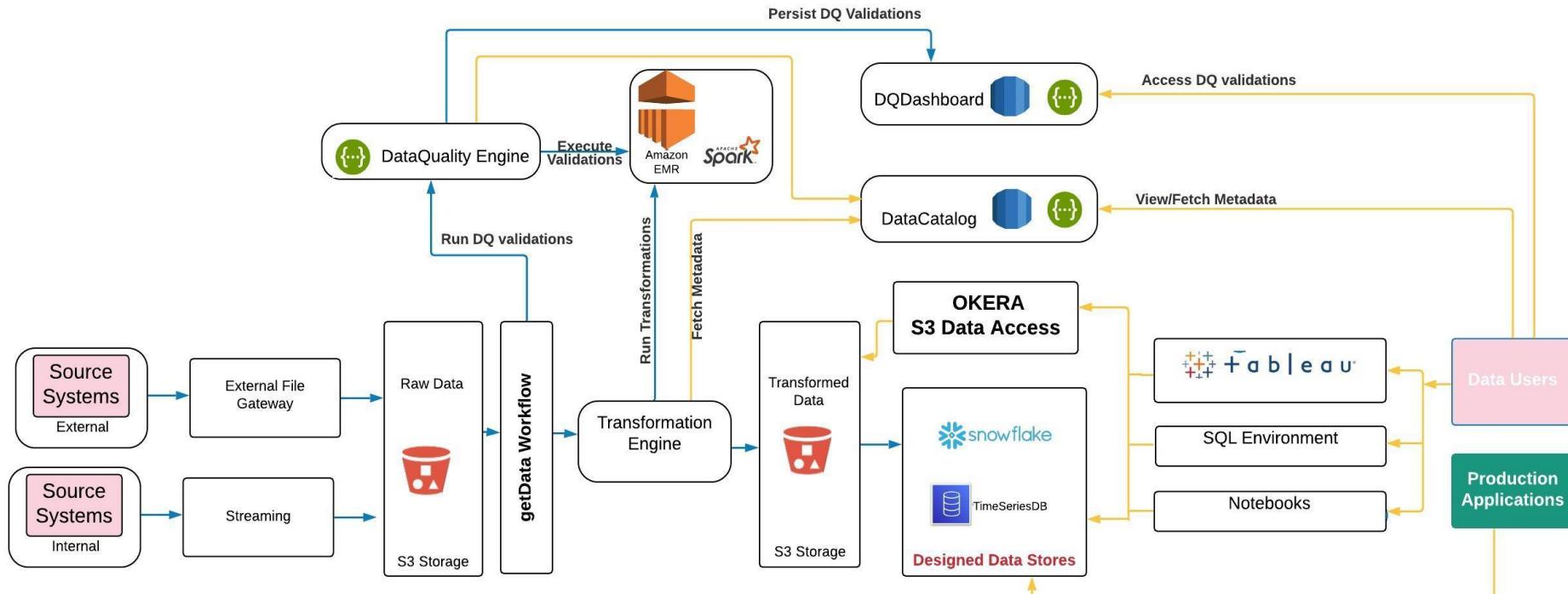


Data lakes take advantage of commodity cluster computing techniques for massively scalable, low-cost storage of data files in any format.

Metadata Driven Lake Processing Framework



Cloud Native Data Lake Ecosystem



Producer Pattern

Consumer Pattern

Data Lake Components



- Drive faster insights and customer centricity by modernizing architecture to provide capabilities like 'schema-on-read/demand', analytics and stream processing
- Lower cost of data management by leveraging object storage in a distributed environment to achieve industry grade performance at rock bottom price
- Lower cycle time for standardizing across source-store-vend layers and access/consumption
- Enable availability of data of all grades (raw, semi integrated and integrated) in production at a faster timeline than the traditional data processing

Enterprise Data Lake-Conceptual View

A self-serve centralized storage and federated compute platform

Core Capabilities:

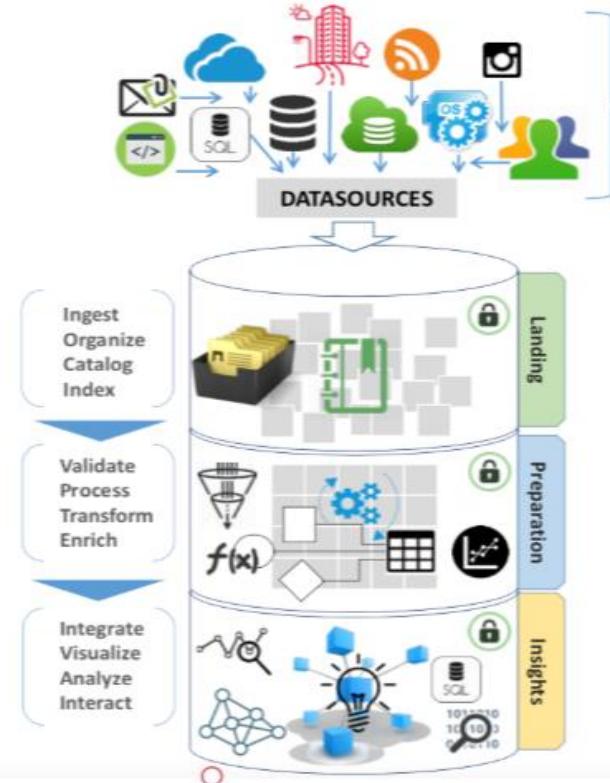
- Data Monetization through Predictive Analytics
- Decision Support & BI Reporting
- Common Landing for Sourced Aligned Data
- Cost effective Data Management
- Fully Governed Active Data Reservoir
- Workload off-load & Ad-hoc Reporting

Foundational Elements:

- Architecture Patterns
- Center of Excellence – Coaching & Guidance
- Data Governance
- Infrastructure Components
- Automated Provisioning
- Data Cataloging & Discovery
- CI-CD Process
- Data Security & Access Management

Note:

Enterprise Data Lake (EDL) is not intended to serve OLTP or be in the middle of business transaction.



Raw Data Ingest

Ability to ingest structured and unstructured data in batch, micro-batch and real-time modes from heterogeneous sources. Supports ingestion of high volumes of data at a high rate (velocity).

Data Management

- The single, secure and cost effective data repository for the enterprise
- Ability to interact with heterogeneous data sources both on cloud and intranet

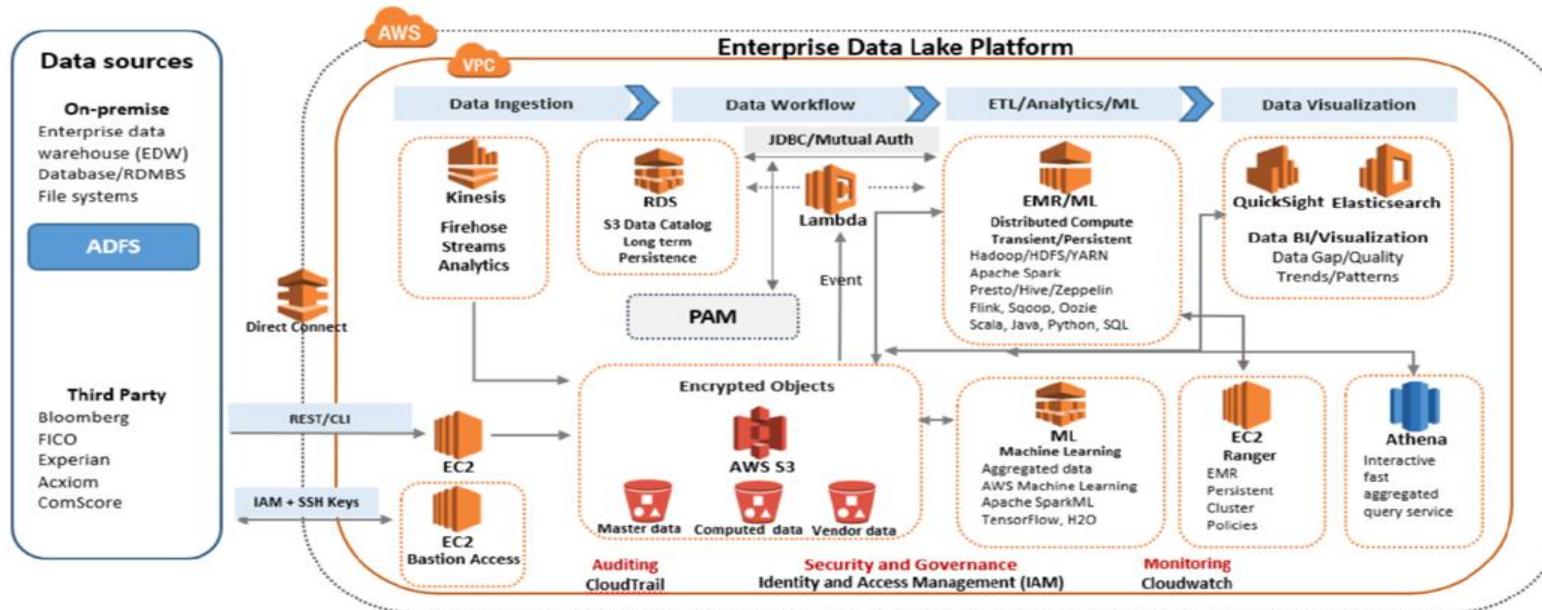
Data Preparation

- Advanced distributed compute capability to help process volume at high-velocity
- Schema evolution support for time-series analysis & advanced BI

Analytics and Modeling

- Visualization, Reporting Dashboards, and Query Interfaces
- Predictive Analytics & Exploratory Science
- Push Down Optimization (PDO) support

Enterprise Data Lake-Reference Architecture



Questions?