

Building Smart Data Lake

Slalom Approach

May, 2018



slalom

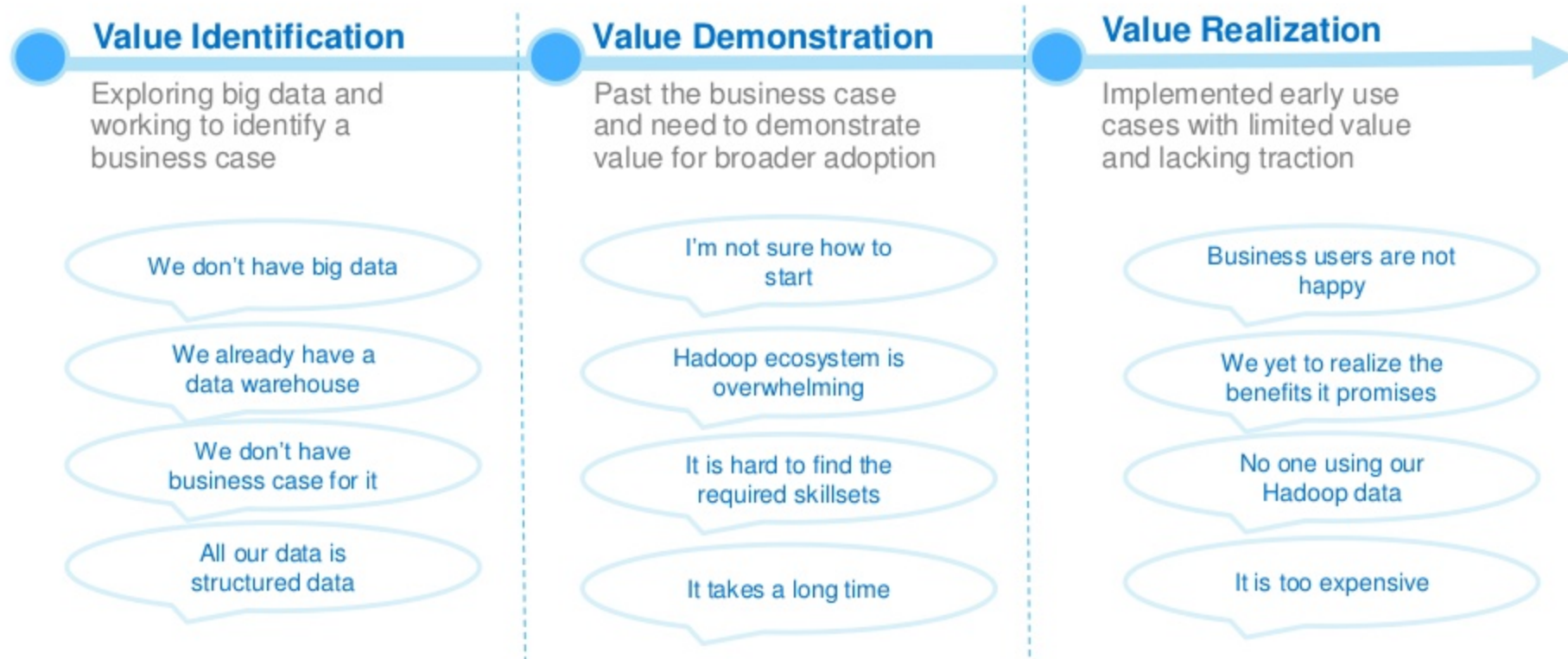
"Better to bet on cloud providers for infrastructure, Cloudera for data, analytics and security fabric, and leave the rest to the ecosystem"

Where are you with your big data journey?



Big Data Journey

Big data has come long way and the enterprises are at different phase of their journey. However, broader adoption of the computation ecosystem is still in its early stages.







Value Identification

Exploring big data and working to identify a business case



Value Identification: Traditional versus Modern Data Architecture

Traditional architectures use rigid data models, costly platforms, resource-intensive ETL and lack support for new use cases.

	Traditional	Modern
	Rigid Data Architecture Early binding to the pre-defined schema makes it inflexible and costly	Flexible Architecture Data is ingested and transformed without prior knowledge of target schema
	Costly Infrastructure and Solution Data duplicated across costly platforms 50-70% spend on acquisition and integration	Simplified Infrastructure and Solution Flexible on-premise and cloud infrastructure API-based pipelines automate data ingestion
	Lacks Support for “New” Use Cases Data silo's impede real-time processing required to support modern use cases	Best Suited for “New” Use Cases Centralized hub for heterogeneous data and variety of tools enable real-time analytics
	Declining Talent Pool The new talent lacks excitement for the traditional technologies and tools	Growing Talent Pool Elevated interest in data engineering and data science work

Value Demonstration

Past the business case and need to demonstrate value for broader adoption



Value Demonstration: Initial Use Cases as MVP

Select outcome-based high impact use case(s) and deliver minimal viable product (MVP) to demonstrate immediate success.



Offload Data Warehouse

Reduce cost and increase speed-to-market by replacing costly platform and ETL processes with Hadoop and API-based integration architecture.



De-Silo Data Walls

Transition fragmented marts to Enterprise Data Hub (EDH) to enable cross functional and enterprise wide descriptive, predictive and real-time analytics.



Enable Discovery of New Use Cases

Ingest source data into the raw zone, apply basic transformations and make the data available for data analysts and data scientists to explore and model.

Enterprise data hub provides foundation for data-driven culture

The client had a vision to drive improved customer experience and engagement through personalized marketing campaigns and needed an on-premise solution that enables the initial use cases and provides foundation for enterprise-wide analytics. Slalom architected a multi-zone data lake to harness and analyze internal and external customer and product data, enabling real-time analytics and a personalized customer experience.

CLIENT

A financial services company serving over 16 million customers nationwide. They pride themselves on being able to provide a personal touch for their customers, and the size of their customer base meant they needed a solution that would be able to integrate large amounts of traditionally siloed customer and product data.

slalom.com

SOLUTIONS

Data architecture and solution design

Data governance deployment

Multi-zone data lake design and buildout

Ingestion and integration using metadata-based big data integration tool

Data discovery enabled and Tableau dashboard deployed

INDUSTRY

Financial Services

ALLIANCES



INDUSTRY

Financial Services

BIG DATA SERVICES

Big Data Startup Planning

Big Data Governance

Big Data Implementation

Enablement and Adoption

Story contact

Value Realization: Building Smart Data Lake

Implemented early use cases with limited value and lacking traction



Value Realization: Common Challenges

We think, most of the organizations lack engineering skills required to fully leverage Hadoop ecosystem and realize the potential of new technologies.



Approach

Organizations are using traditional source-to-target approach of acquiring and integrating data for known use cases



Architecture

Usually considered an IT infrastructure project, Hadoop is used as a large file system to dump data files with limited use and marginal business value



Skills

Majority of the data professionals (ETL developers, data analysts) lack engineering skills required to fully leverage Hadoop technologies



Culture

Mindset has to change from hoarding and protecting information to making it easy to access and use data as an enterprise asset

Value Realization: Smart Data Lake

Smart data lake should be....

Designed Right

Standardized

Enterprise Scale

Secured

Governed for Adoption

Governed

Auditable

Supported

Economical to Use

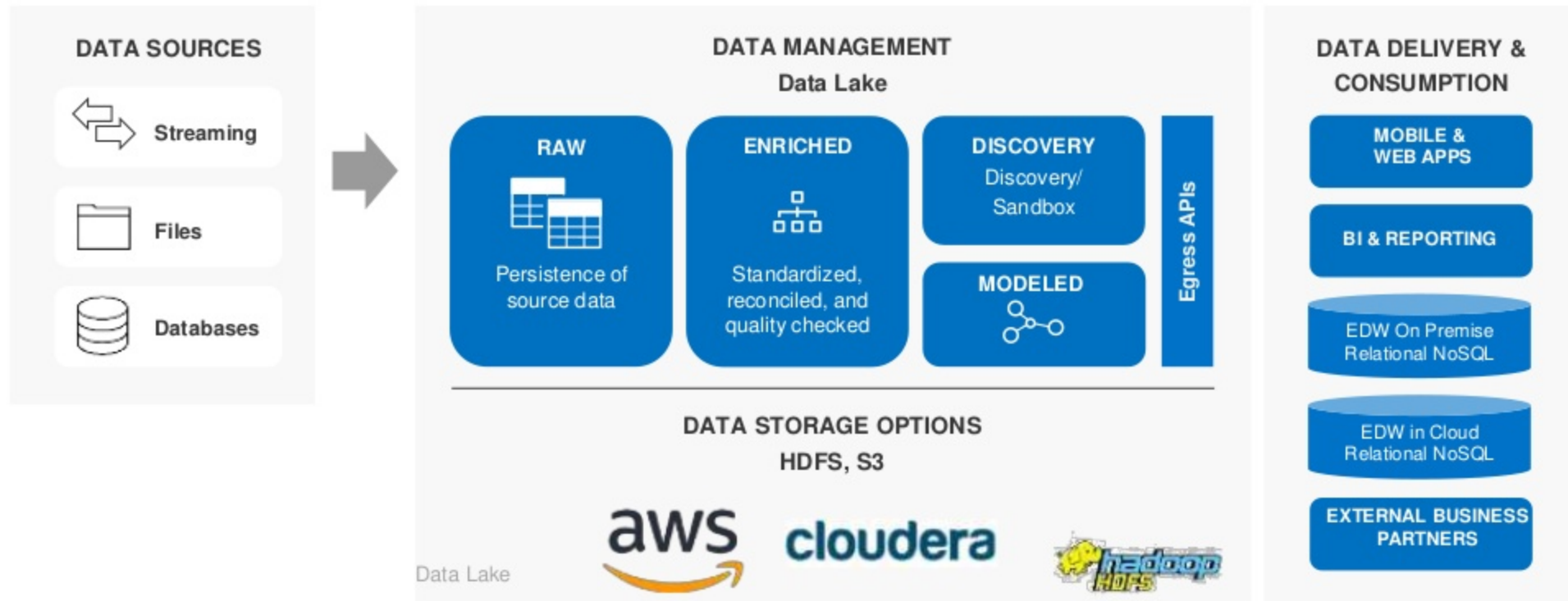
Open Source

Multi-Use Support

Extensible

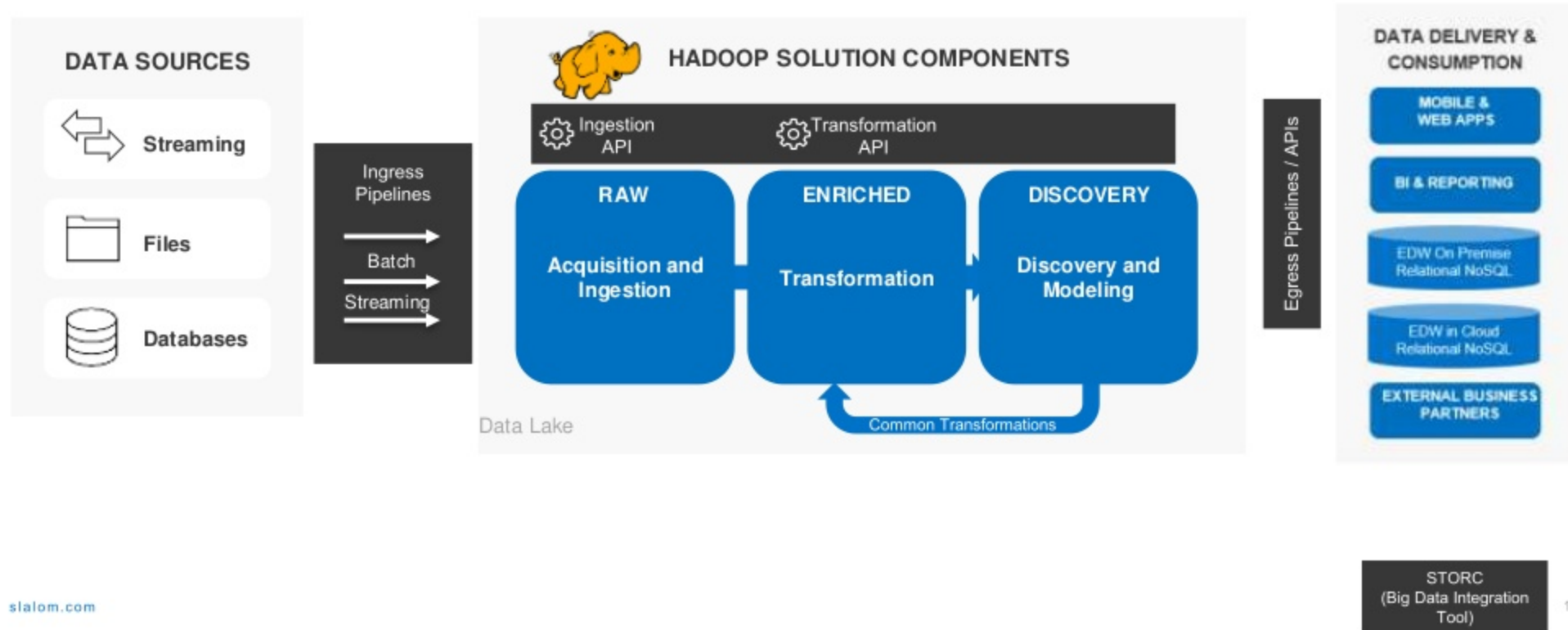
Value Realization: Designed Right

Multi-zone, self-governed data lake to provide secure and flexible data architecture to harness enterprise data for accelerated speed to insight.



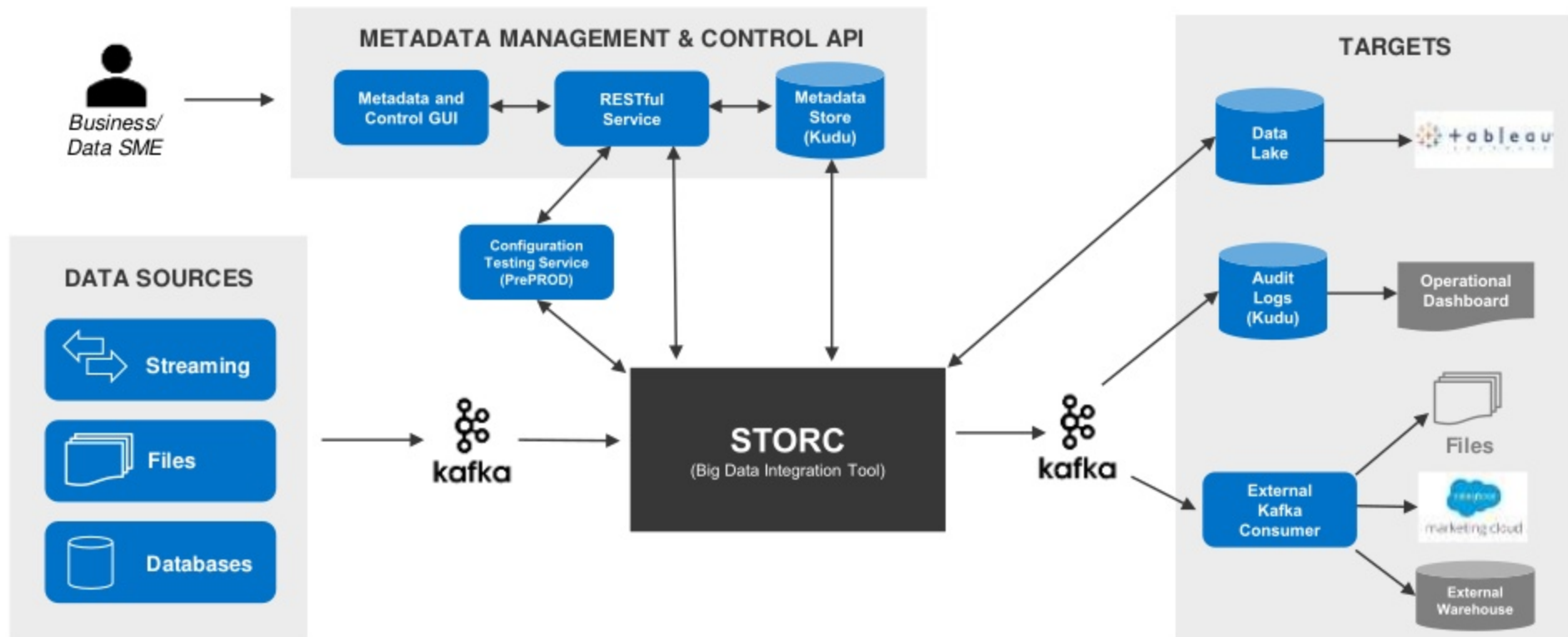
Value Realization: Accelerated Implementation

The architecture implements data pipelines using our purpose-built open source integration APIs accelerating implementation by 9-12 weeks.



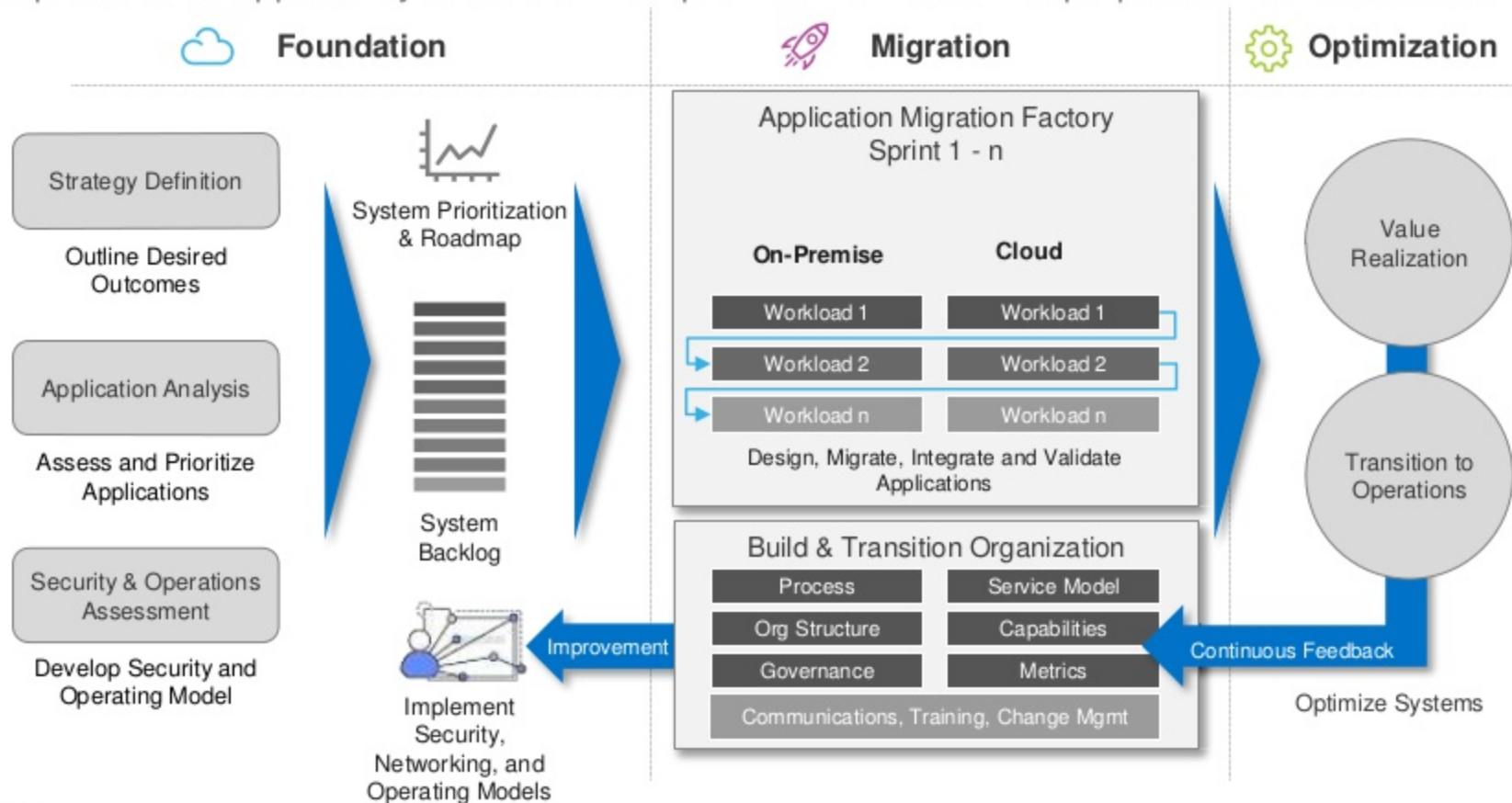
Value Realization: Accelerator Architecture

The accelerator enables self-service by allowing data analysts and data SMEs to ingest new data sources and promote data through the lake with limited to no IT dependencies.



Value Realization: On-Premise to Cloud

Cloud presents an opportunity to transform on premise workloads into purpose driven scalable solutions



Next Generation Data Platform for Healthcare Analytics

BACKGROUND

Slalom partnered with a Fortune 500 healthcare company to deliver a next generation data platform. The client's existing platform could not support increasing data volumes and a growing need for advanced analytics workloads. The new platform not only addressed these scalability concerns but also allowed the client to host both structured and unstructured data in near-real time. Most importantly, this data platform opened doors for new monetization opportunities

PROJECT

Slalom built a next generation Hadoop data platform to meet the client's needs. Leveraging the cloud enabled a quick turnaround time as well as security features ideal for storing PII and PHI data. Slalom team migrated and optimized existing data to leverage Hadoop high-performance features. Slalom also built a near-real time platform that can ingest HL7 messages from several hospitals and provide event-driven alerting.

slalom.com

RESULTS

PEM delivery methodology was used to deliver a cost effective and scalable solution

Client is exploring opportunities to monetize the solution as an analytics workbench

The data science team can leverage both SAS and R integration with the platform for advanced analytics

Sunset existing platforms, reducing licensing and support maintenance costs

TECHNOLOGY



INDUSTRY

Healthcare

BIG DATA SERVICES

Agile Delivery Approach

Big Data Implementation

Story contact:
Maz Chaudhri

Real-time Transaction Streaming using Cloudera in AWS for a QSR

BACKGROUND

Our client in the fast-casual food industry was having widespread challenges accurately capturing and measuring key business metrics. Due to inconsistent data integrity in the nightly batch process, executives and leaders were growing skeptical of the reliability of reporting and analytics built from the data. Leaders were clamoring for timely visibility to better, cleaner data.

PROJECT

The Slalom team served as Scrum Master, Product Owner and Analyst during the architecture and delivery of the AWS-based Cloudera platform. Using a Kafka-based publish-subscribe architecture, each restaurant location in addition to the online ordering platform was set up to stream data feeds to the unified Cloud platform.

RESULTS

Agile Delivery Methodology

Real-time data platform

Self-service enablement

Up-to-the-minute view into the operations of over 6,000 restaurant locations nation-wide.

Ability to monitor KPIs and react with targeted efforts to boost sales exactly where it is needed.

TECHNOLOGY



INDUSTRY

Food Service

BIG DATA SERVICES

Agile Delivery Approach

Big Data Startup Planning

Platform Evaluation & Selection

Big Data Implementation

Next Generation Data Platform & Supply Chain visibility

BACKGROUND

A top 10 Pharmaceutical company, and top 150 Fortune 500, sought to implement a next generation modern data platform. The platform needed to not only provide end to end supply chain visibility, but also be flexible and scalable to handle a heavy volume of serialized data. The client also wanted to establish a data lake so as to be able to predict and prescribe their inventory and shipments to better serve their customers.

PROJECT

Slalom utilized AWS and Cloudera Hadoop to build this next generation data platform. The data platform gave visibility to inventory levels to help drive the development of inventory optimization strategies and integrated multiple disparate sources to give end to end shipment visibility of the client's supply chain.

RESULTS

A scalable and flexible Big Data Platform

A universal XML ingestion framework

HDFS Data lake that ingests and persists all data from source system

Allowed the client to sunset a reporting product that saved over \$1MM annually in support maintenance cost

Qlik BI & Operational reports utilizing Hadoop as the backend

ALLIANCES



INDUSTRY

Pharmaceuticals

BIG DATA SERVICES

Agile Delivery Approach

Big Data Implementation

Why Slalom?



Time to Value



Pre-Built Accelerators



Proven Approach and Experience



AGILE ENGINEERING APPROACH

Start small, deliver value and evolve your Big Data program



BIG DATA INTEGRATION TOOL

Open-source meta-data driven integration API



DATA GOVERNANCE in a BOX

Multi-faceted data governance deployment and tools



BIG DATA STARTUP PLANNING

Pre-defined epics and stories for big data startup



PLATFORM SELECTION

Best practices-based evaluation toolset



READINESS AND ADOPTION

Org readiness and change strategy and enablement