GNARLY Data_Waves

PRESENTED BY dremio

EPISODE 13

Making the Move: Five Factors to Consider When Migrating from Hadoop to the Data Lakehouse

Today's Agenda

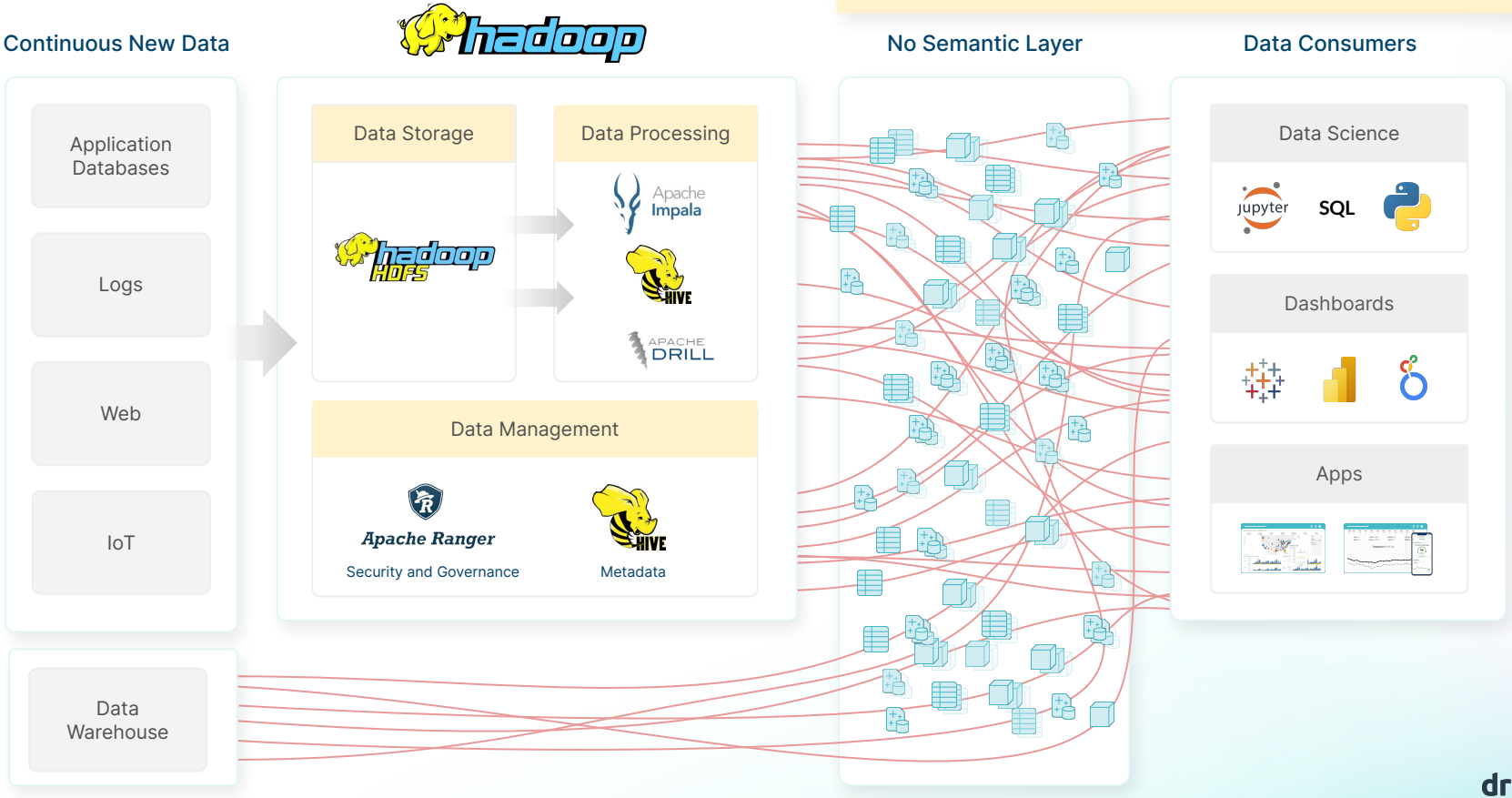
Making the Move: Five Factors to Consider When Migrating from Hadoop to the Data Lakehouse

- Why Migrate off Hadoop
- Introduction to the Dremio Lakehouse
- 5 Factors to Consider When Migrating Hadoop to the Data Lakehouse

The Hadoop Ecosystem

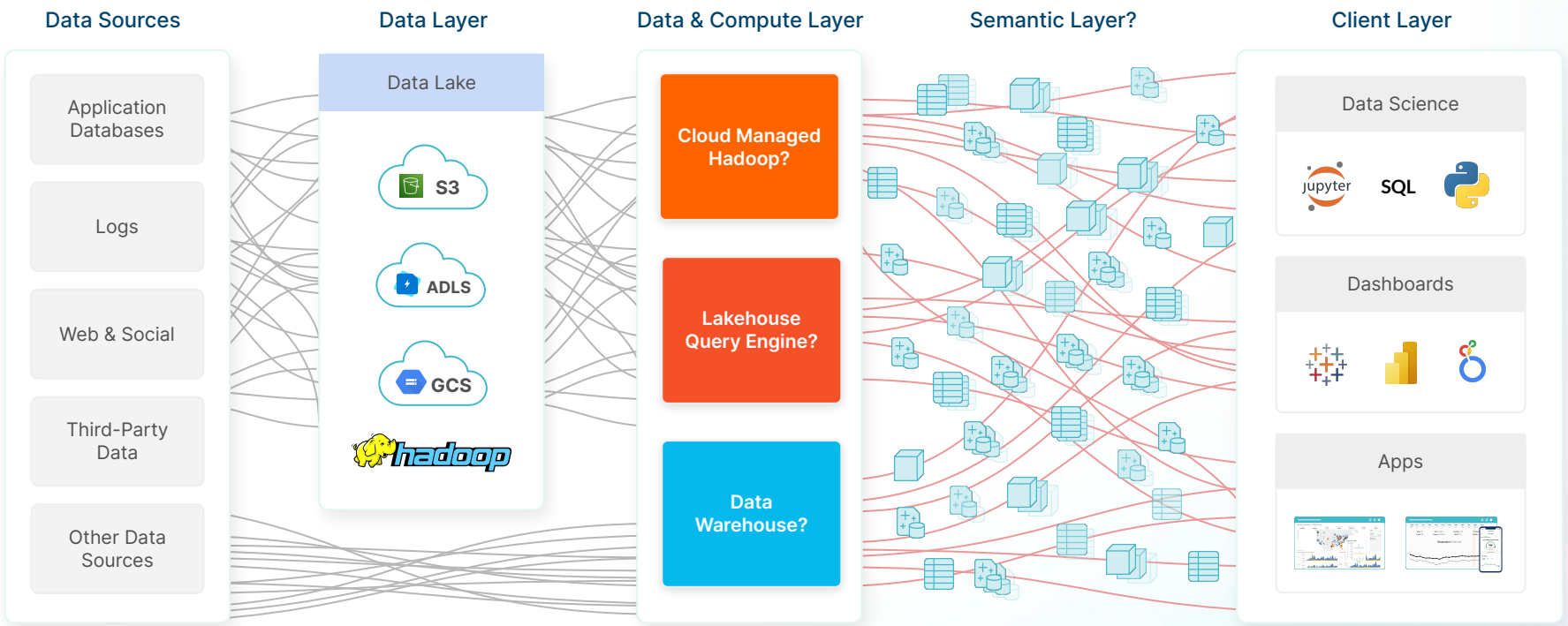
Challenges

- ✓ Requires deep expertise in the Hadoop ecosystem to maintain
- ✓ High cost of scalability as your data grows
- ✓ Query performance management
- ✓ Difficult to enable governed self-service analytics



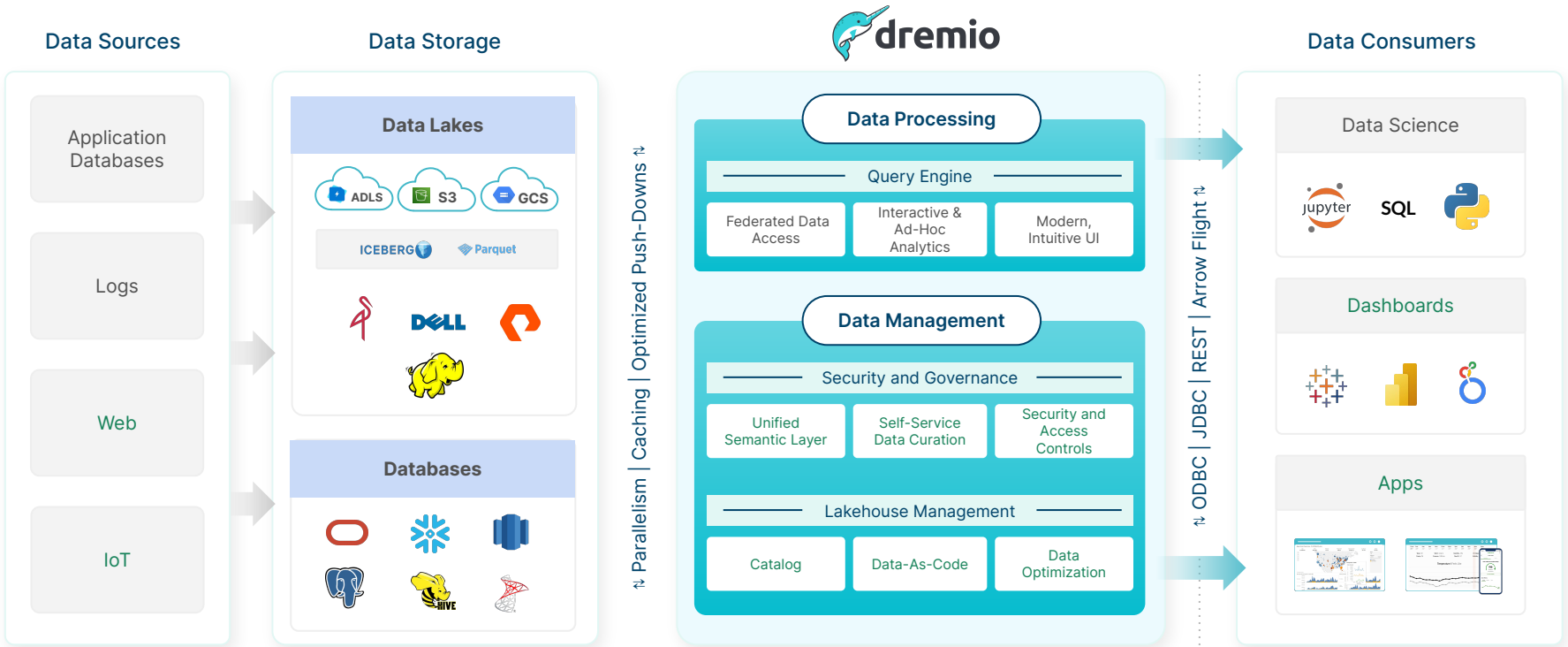
dremio

What are your options?



dremio

Dremio Data Lakehouse



#1 - Digital and AI Transformation

Supply Chain Optimization

- Identify and resolve supply chain issues
- Increase market share and profitability

Customer 360

- Drive insights before insurance policies lapses
- Understand patient history and identify interactions with their network faster

Simplify IT Efficiency

- Consolidate data and analytics platform
- Provide customers easy access and integrations to data and tooling

Who We've Helped

Memorial Sloan Kettering Cancer Center

Creating a purposeful scientific data management system that accelerates cancer research for scientists

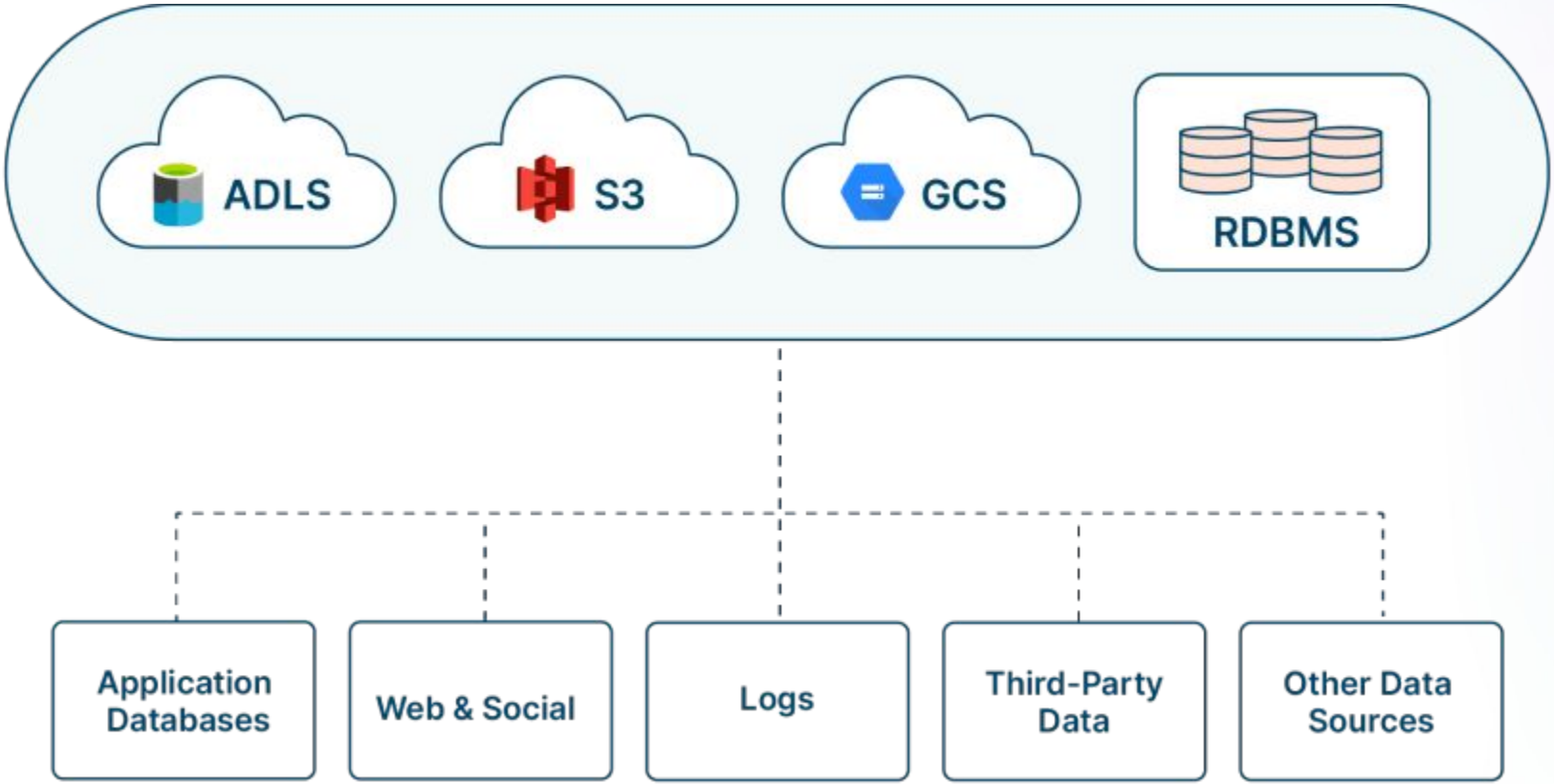
Mapping customer experience to identify bottlenecks in order processing to consistently meets delivery commitments

Building products for financial inclusion and promote economic opportunity

"Dremio is literally going to help MSK save the lives of cancer patients."

Arfath Pasha, Lead Software Engineer at Memorial Sloan Kettering Cancer Center

#2 - Data Estate Vision



#3 - Hadoop Migration Plan

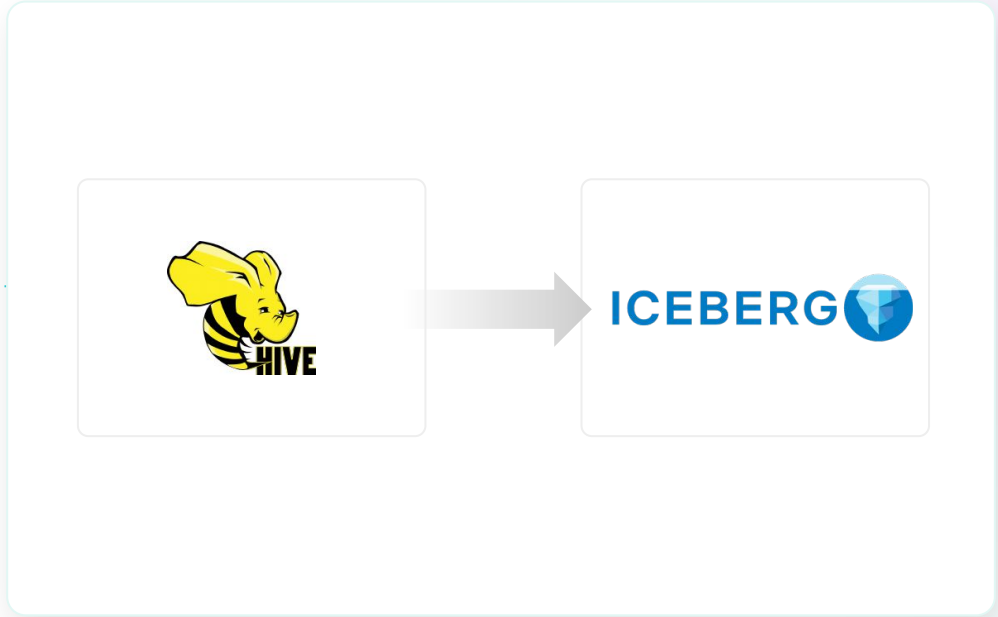
Phase 01	Phase 02	Phase 03
Modernize Hadoop Query Engine & Provide Self-Service Analytics	Migrate HDFS to object storage	Create Your Open Data Lakehouse

Accelerate value and transform your business in weeks, not years

#4 - Open Data

What happens?

- Data is in cloud object storage
- Migrate Hive tables to open table format like Apache Iceberg



Results:

- Future proof your data architecture
- Achieve higher performance, DDL, DML, schema evolution, time-travel, and other data warehouse functionality
- Avoid vendor lock-in from proprietary table formats
- Make data accessible to your query engine(s)

Open Data Lakehouse with Dremio

Community-Built

Apache Iceberg

- ACID Transactions
- Partition Evolution
- Schema Evolution
- Time Travel

Unified Data Access

Dremio Sonar

- High Performance Data Lakehouse Query Engine
- Data federation
- Unified Semantic Layer

Easy Lakehouse Management

Dremio Arctic

- Lakehouse Catalog
- Manage your data as code
- Multiple Domains

#5- Self-Service for Data Consumers

Data has become a critical asset to every company

Data is pervasive so it's no longer practical for one IT team to centrally manage a company's data

Discoverable
Data must be exposed through a catalog with rich metadata

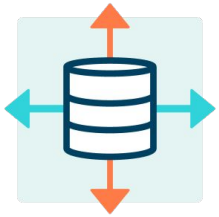
Addressable
Data must be easily accessible to a variety of users and tools

Trustworthy and truthful
Data quality must be ensured before it is consumed

Self-describing
Schemas with clear semantics and syntax to enable self-service

Secure
Data must be secure and governed by global access control

Dremio Makes Self-Service Easy for Data Consumers

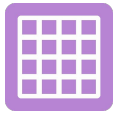


Unified Data Access

A single, unified access point for all enterprise data for BI and ad-hoc self-service.

Based on Apache Arrow

Deliver data faster from object storage to your favorite BI tools such as Power BI and Tableau.



Data Reflections

Get answers to your business questions with sub-second latency, without needing BI cubes, data copies, or complex materialized views.

Business Friendly Semantic Layer

Provide an easy, no/low code experience to interact with, analyze, and serve data.

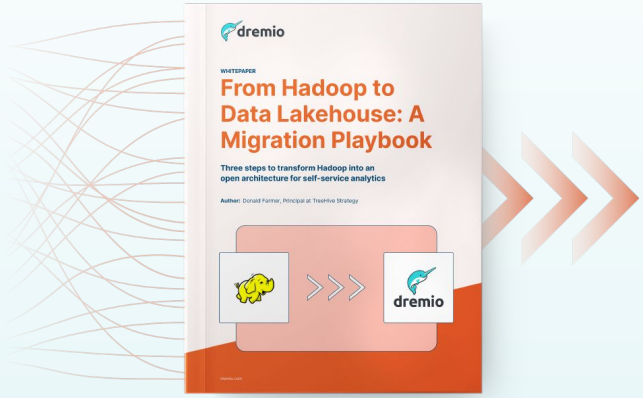
Enable Data Mesh and Data Products

Quickly publish data assets and products into a catalog experience from data stored across the enterprise.

Open Data Lakehouse

With open source at our core, Dremio provides easy data lakehouse management for multiple domains and table formats like Apache Iceberg.

Data Lake Modernization with Dremio



Phase 01

Modernize Hadoop Query Engine & Provide Self-Service Analytics

Phase 02

Migrate HDFS to object storage

Phase 03

Create Your Open Data Lakehouse

Thank you!