



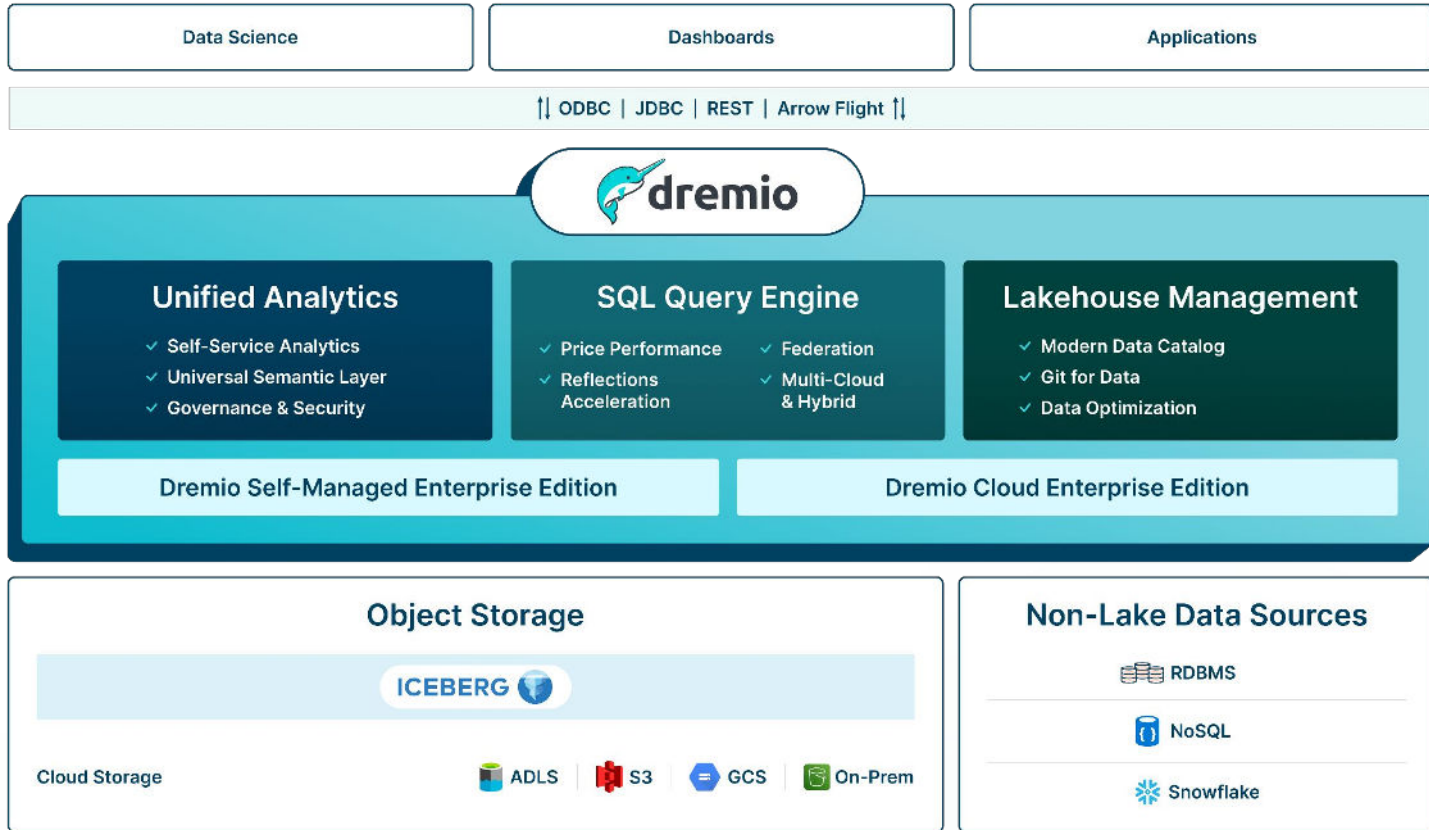
GNARLY
Data_Waves

PRESENTED BY  **dremio**

EPISODE 45

Next-Gen Data Pipelines are Virtual: Simplify Data Pipelines with dbt, Dremio, and Iceberg

Unified Lakehouse Platform for Self-Service Analytics



Shifting Left Reduces MTTI and Shortens ETL Pipelines



Shifting left requires three core pieces of innovation

Intuitive self-service experience

An intelligent query engine

Next-generation dataops capabilities

Get Started with the Dremio Lakehouse Platform

- Sub-second query on 1 billion rows of data joining Amazon S3 with a Postgres database
- Connect to Tableau or Power BI and build a dashboard with this dataset
- Everything hosted by Dremio - 100% free for you
- Either deploy a self-managed cluster with Dremio Software or enjoy a Fully Managed service with Dremio Cloud



Get started with the Dremio Data Lakehouse. Free versions available for Software or on the Dremio Cloud.

Self-service analytics with data warehouse functionality at the lowest cost



Streamlined self-service analytics for data-driven innovation

Seamless unified data access and analytics across all your data with no data movement.



Revolutionary SQL performance

Optimized price performance for SQL query speed and acceleration, with sub-second analytics across your data.



Centralized data governance to balance data access and control

Simplify and scale data governance across organizational boundaries with role-based and fine-grained access control at every level.



Frictionless connector integrations for all your data

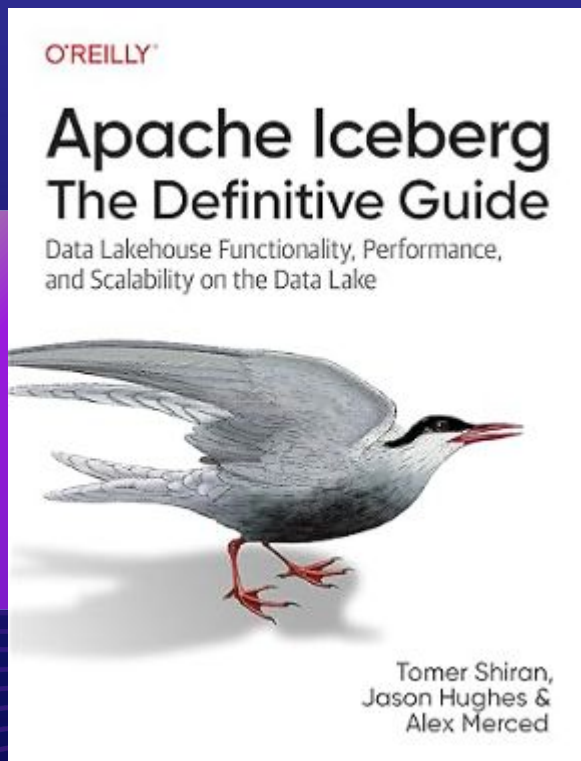
Dozens of integrations across an array of sources, including object storage, metastores, and databases in the cloud and on-premises.



Seamless lakehouse data management

Easy and efficient Iceberg data lakehouse management with Git-inspired data versioning. Available in Dremio Cloud.

Apache Iceberg: The Definitive Guide





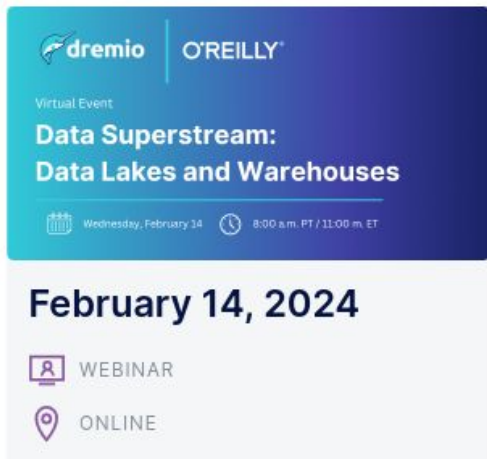
Survey Report



2024 State of the Data Lakehouse

Gain unprecedented insights into the evolving landscape of data lakehouses and benchmark your organization with Dremio's State of the Data Lakehouse Survey Report. This survey of 500 full-time IT and data professionals from large enterprises offers fresh insights on data lakehouse trends, adoption and associated benefits.

[Download Free](#)








 | 


Virtual Event

**Data Superstream:
Data Lakes and Warehouses**

 Wednesday, February 14  8:00 a.m. PT / 11:00 a.m. ET

February 14, 2024

 WEBINAR

 ONLINE

Data Superstream: Data Lakes and Warehouses

Join Dremio as they partner with O'Reilly to discuss the storing, processing, and moving of data efficiently and cost-effectively in the cloud. These expert-led sessions deliver insight into how to increase the scalability, speed, and availability of your data and equip you with a set of best practices for utilizing your data warehouse, data lake, or data lakehouse.

[Register Now](#)

LIVE

Subsurface

The Data Lakehouse Conference

PRESENTED BY  **dremio**

7,000+

Global Data Experts

46+

Breakout Sessions

3,000+

Companies

May 2/3rd - Learn more at [Dremio.com/subsurface](https://dremio.com/subsurface)



GNARLY
Data_Waves

PRESENTED BY  **dremio**

EPISODE 45

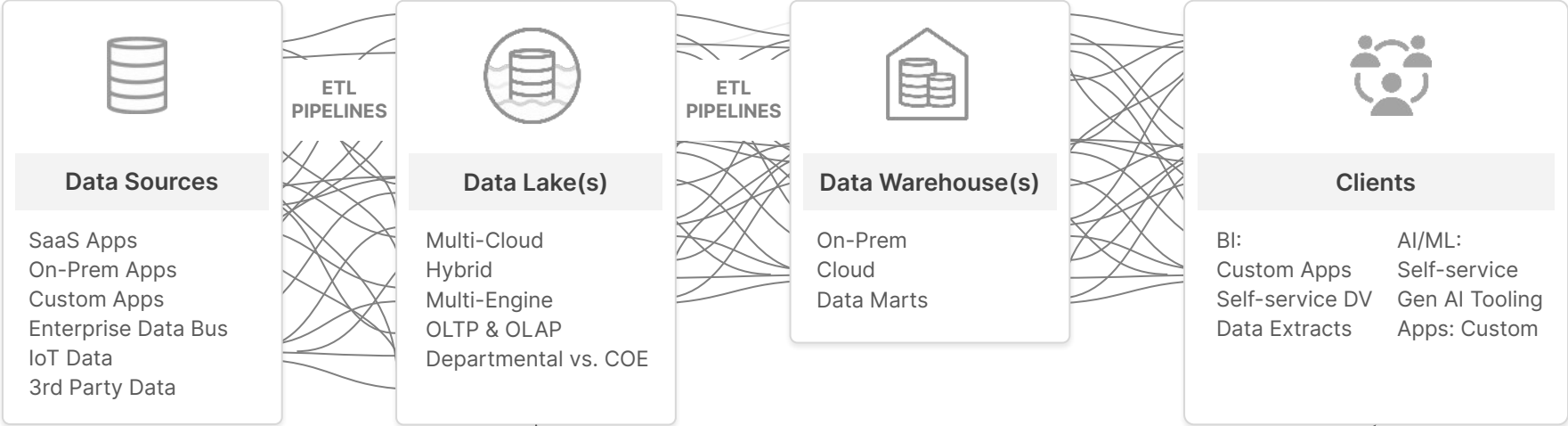
Next-Gen Data Pipelines are Virtual: Simplify Data Pipelines with dbt, Dremio, and Iceberg



Introduction to Dremio Arctic: Catalog Versioning and Iceberg Table Optimization

Presented by Alex Merced

Traditional Chain of Data Movement Pipelines



What's The Problem?

Broken Pipelines that require tedious backfilling



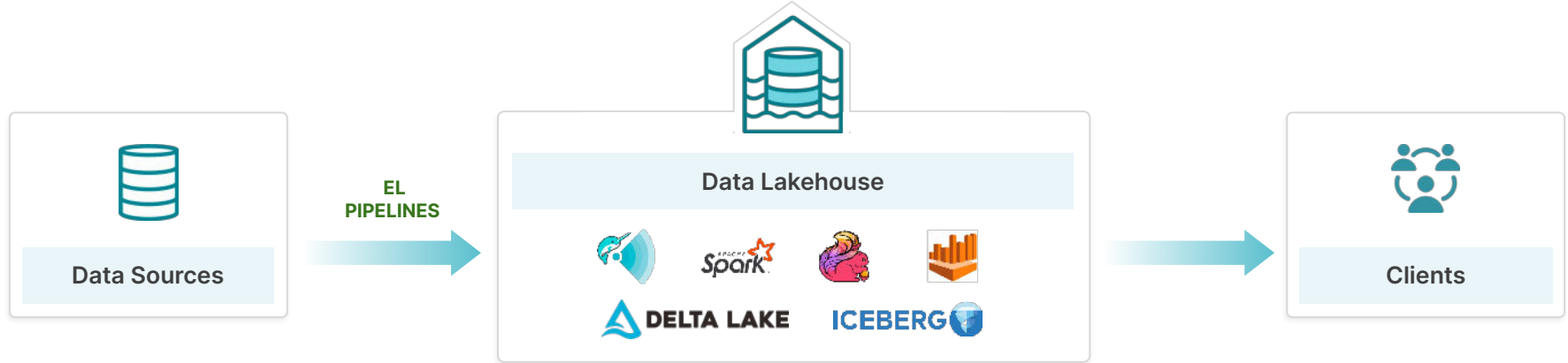


Angry Consumers from Late and Inconsistent Data

Cost of these Movements

1. Storage Costs
2. Compute/Processing Cost
3. Network & Egress Costs
4. Lost productivity in time it takes for all pipelines to proliferate data
5. Regulatory fees from governance and security risks in too many copies
6. Data Model Drift from as data models may become inconsistent over several movement
7. Cost of Bad Insights from Inconsistent Data from Data Copy Sprawl

The Dremio Approach (ZeroETL & Virtual Data Marts)



Federated Data Sources connected directly to Dremio (ZeroETL, Low-ETL)

Use Dremio's Semantic Layers to Model data for each business line into virtual data marts.

Dremio's Data Reflections eliminates the need for Cubes and Extracts

The Benefits of the Dremio Approach?

Less Pipelines,
Less Backfilling,
Fresher Data Sooner





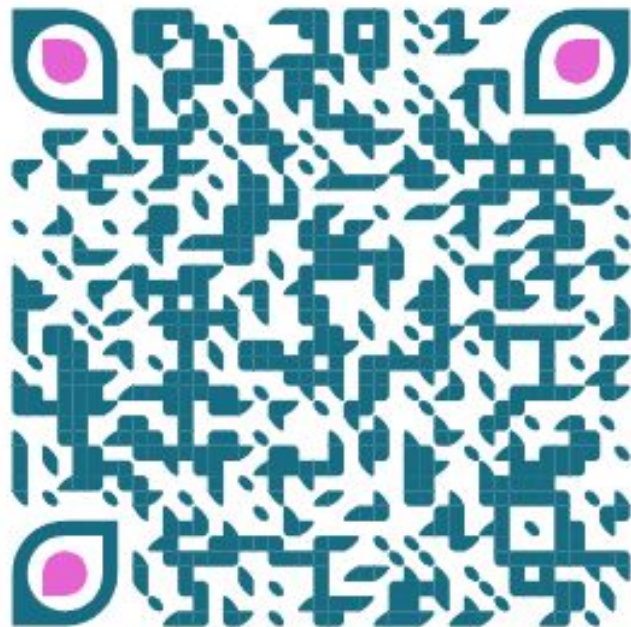
Data Consumers
getting fresh,
correct and
fast data

Cost Benefits of this Approach

1. Less Copies, Less Storage Costs
2. Less Pipelines, Less Compute
3. Less Data Movement, Less Network Costs
4. Less Pipelines, All Datasets Remain Fresh
5. Less Copies in One Place, easier to comply with regulations
6. Less Copies, Less Chances of Data Models Changing
7. Fresher Data Means Better Insights



**A Iceberg/Dremio Lakehouse on
your laptop exercise**



**Deploy Dremio Software or
Dremio Cloud**

Branching and Merging with Dremio's Lakehouse Catalog

Video

Learn how to curate data using virtual data marts accessed through a universal Semantic layer.

Video

Better Orchestration with dbt: Discover the benefits of orchestrating DML and view logic, optimizing data workflows.

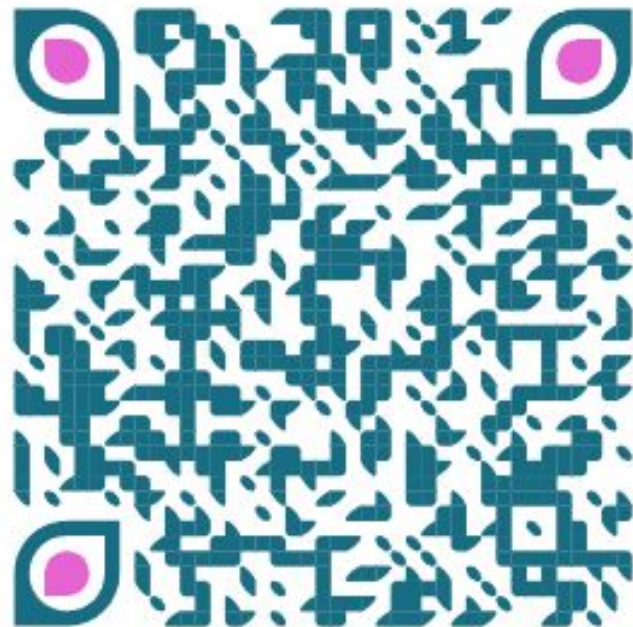
Video

Elevating Data Quality: Learn techniques to automate lakehouse maintenance and improve data integrity.

Video



**A Iceberg/Dremio Lakehouse on
your laptop exercise**



**Deploy Dremio Software or
Dremio Cloud**