



**GNARLY**  
Data\_Waves

PRESENTED BY  **dremio**

EPISODE 32

# Introduction to Dremio

## Arctic: Catalog

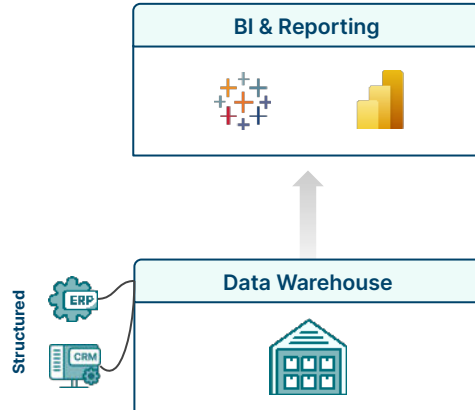
## Versioning and Iceberg

## Table Optimization

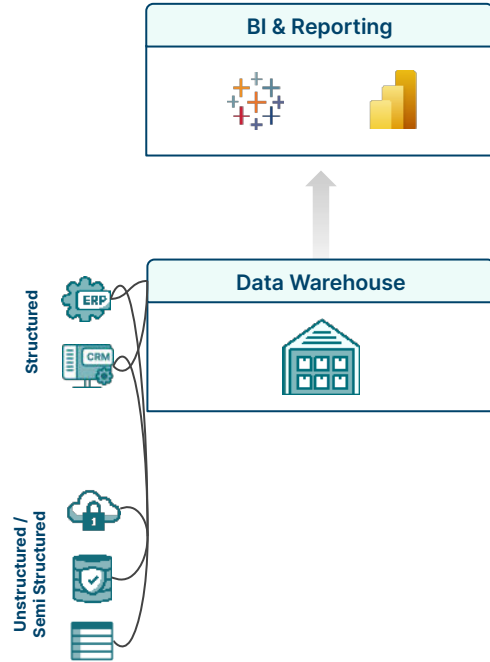
# Agenda

- A Brief Review of Data Lakes: How We Got Here
- Data Lake → Data Lakehouse: What's Needed?
- A Quick Overview of Apache Iceberg
- Dremio Arctic - Data Lakehouse Management Service
- Automatic Optimization
- Catalog Versioning with Data as Code
- Quick Demo
- Q&A

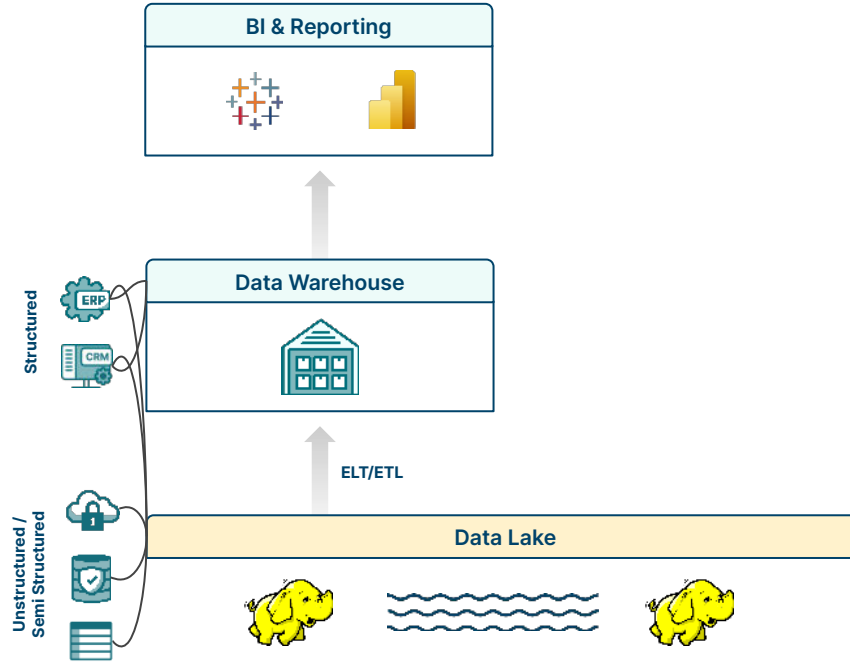
# A Brief Review of Data Lakes



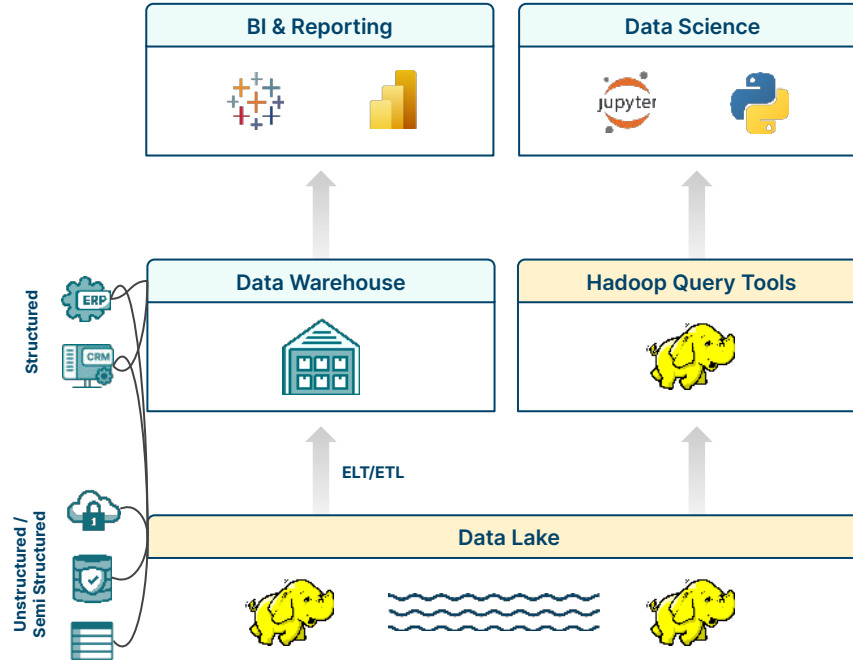
# A Brief Review of Data Lakes



# A Brief Review of Data Lakes



# Cooperative Data Architecture: Data Lakes + Warehouses



# We've Had Query Engines on Data Lakes for a While...

Data Lake Engines



Data Lake Storage



# ...But Customers Needed a Way to Manage it

Data Lake Engines



Modern Lakehouse Catalog



Table Formats



File Formats



Data Lake Storage





# Let's Talk About Iceberg

## What is it?

- An open table format for data lakes.
- Purpose-built for high-performance analytics on large datasets in the data lake
- Supported by large technology firms including Netflix, AWS, Google, Dremio, Snowflake, etc.
- It's open - accessible by multiple engines.

## What does it do?

- Iceberg makes it easy to manage your data lakehouse with optimization, schema evolution, and data versioning capabilities



**Easy Data Management**



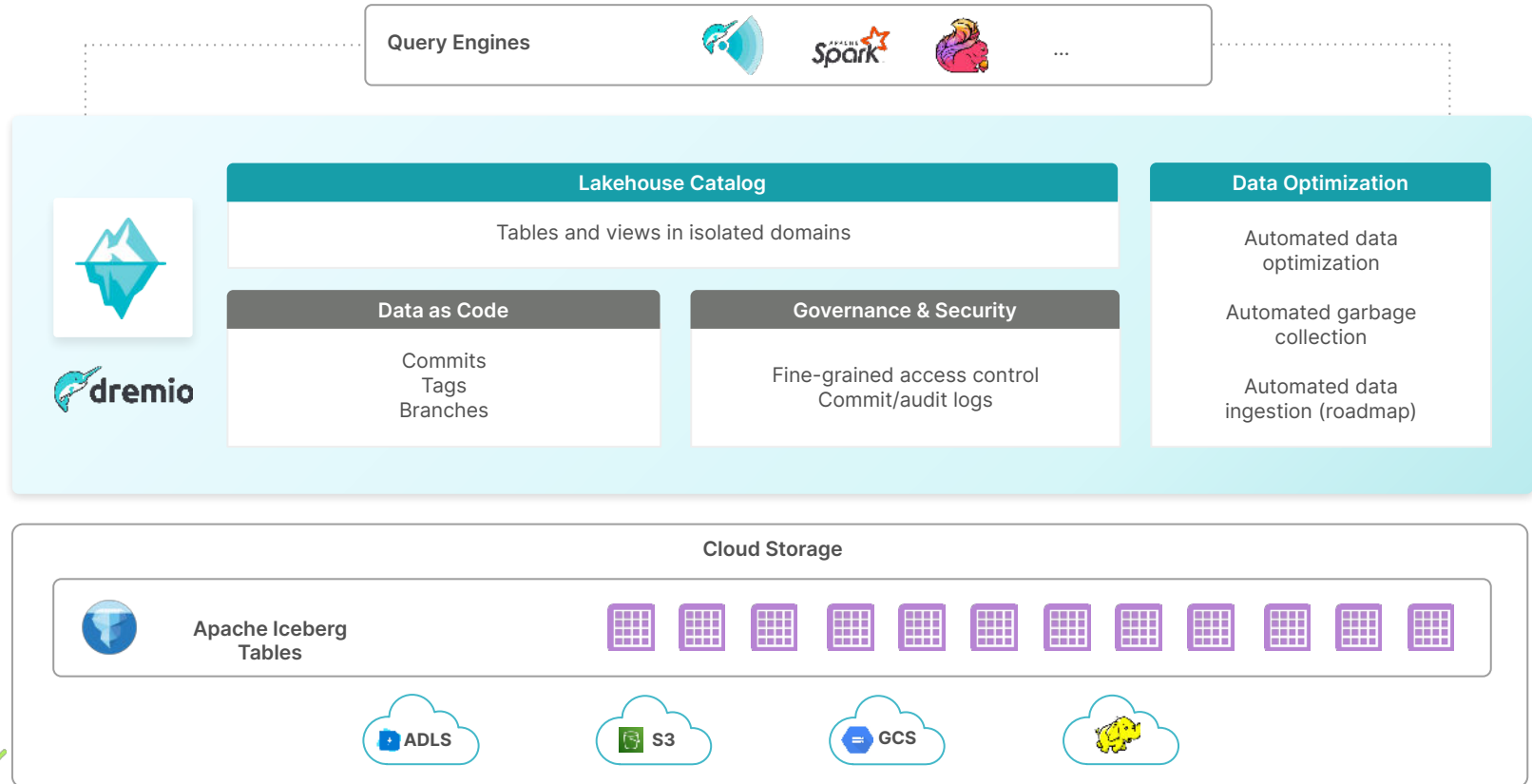
**Transactional Consistency**



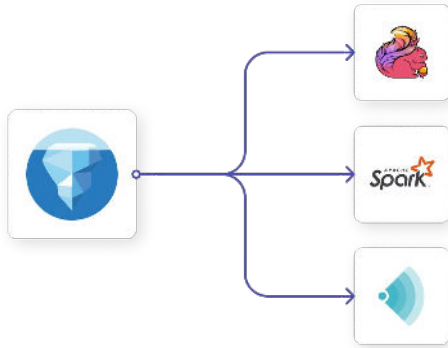
**High Performance**

**What if you could wrap up the benefits of Iceberg in an Iceberg-native lakehouse catalog that automates its optimization features, provides enterprise-level security and governance, and extends the data management and versioning capabilities of Iceberg with data as code?**

# Dremio Arctic is a Data Lakehouse Management Service

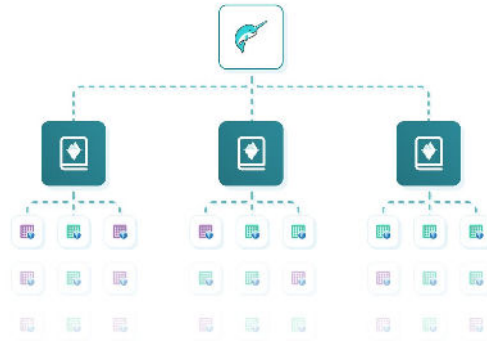


# Dremio Arctic is a Modern Lakehouse Catalog



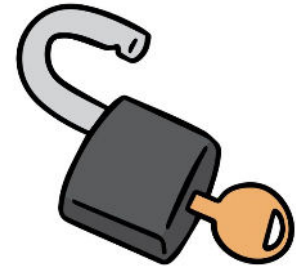
## ICEBERG-NATIVE

- Nessie (the Arctic catalog) is built into the open source Apache Iceberg project
- Use a variety of Iceberg-compatible engines including Dremio Sonar, Spark and Flink



## MULTIPLE DOMAINS

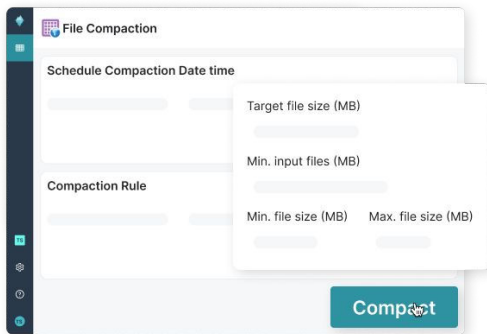
- Multiple isolated domains/catalogs in an organization, each containing a folder hierarchy of tables and views
- Designed to enable data mesh (including federated ownership and data sharing)



## ACCESS CONTROL

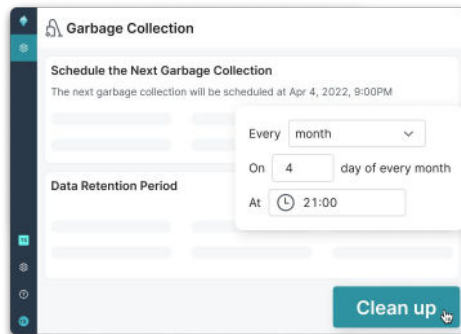
- Table, column- and row-based access control
- Custom roles and integration with existing user/group directories (AAD, Okta, etc.)

# Automatic Optimization Makes Data Lake Management Easy



## TABLE OPTIMIZATION

- Dremio Arctic automatically rewrites smaller files into larger files and groups similar rows in a table together.
- Table optimization significantly accelerates query performance.

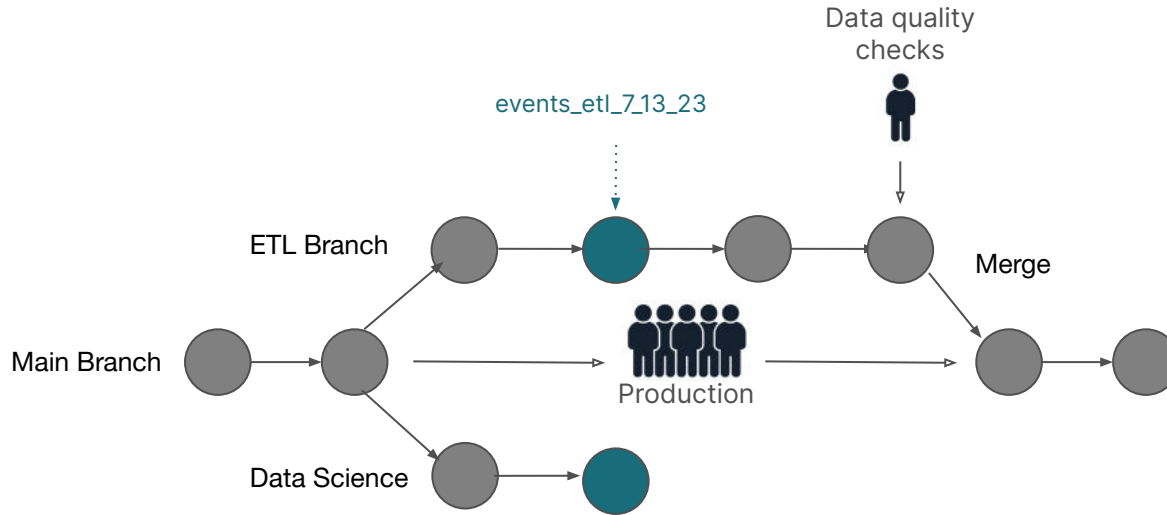


## TABLE CLEANUP

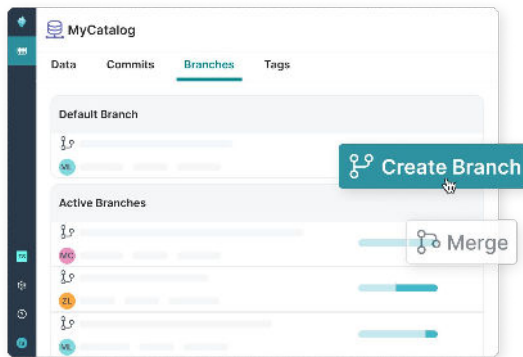
- Dremio Arctic automatically removes unused manifest files, manifest lists, and data files.
- Cleanup runs in the background and ensures efficient use of data lake storage.

# What is Data as Code?

Software Development Principles Like CI/CD & Versioning for Data

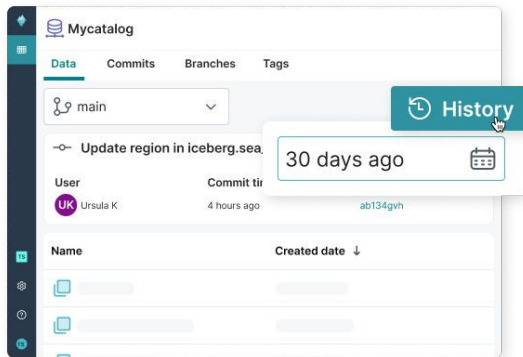


# Catalog Versioning with Data as Code



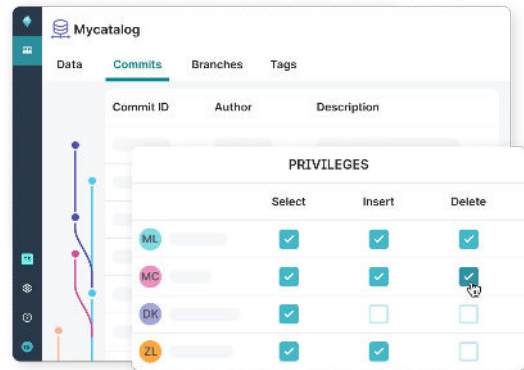
## ISOLATION

- Experiment with data without impacting other users
- Ingest, transform and test data before exposing it to other users in an atomic merge



## VERSION CONTROL

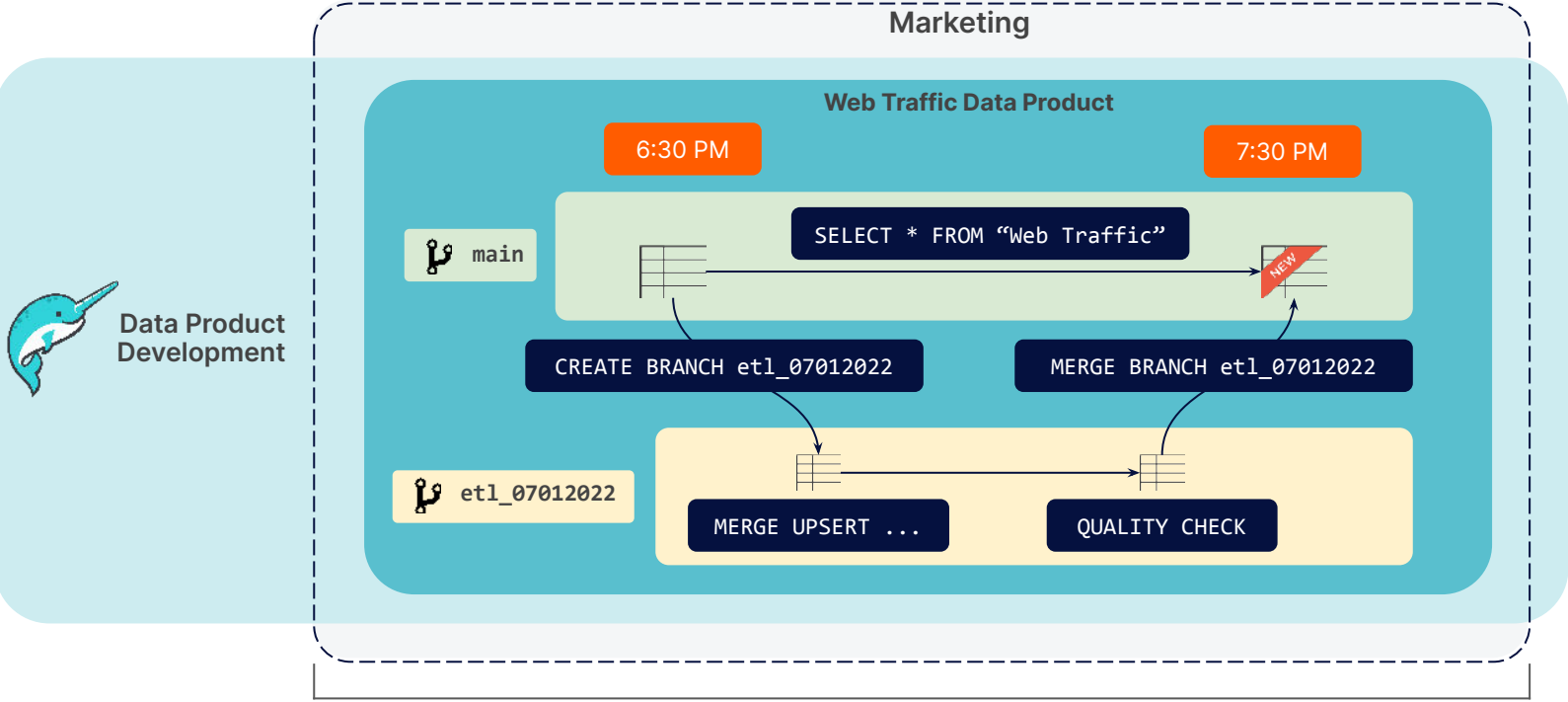
- Reproduce models and dashboards from historical data based on time or tags
- Recover from any mistake by instantly undoing accidental data or metadata changes



## GOVERNANCE

- All changes to the data and metadata are tracked: who accessed what data and when
- Fine-grained privileges to control access to the data at the table, column and row level

# Use Case: Data Product Development







Demo

# Ready to Give it a Try?

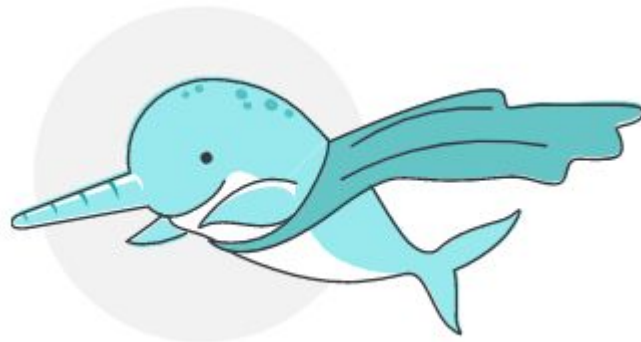
Dremio Arctic is now the default catalog in Dremio Cloud, so every new project will get a Dremio Arctic catalog!

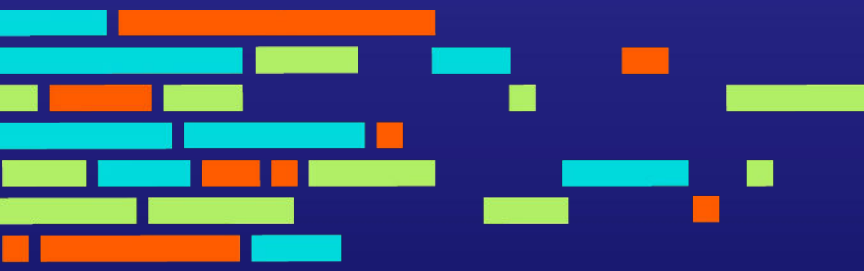
Expand what your semantic layer can do for you with:

- Automatic Optimization of your Iceberg tables
- Catalog versioning with data as code

Sign up for Dremio Cloud for free at

[www.dremio.com/get-started](http://www.dremio.com/get-started)





# Q&A