



GNARLY
Data_Waves

PRESENTED BY  **dremio**

CE Workshop

Creating an Dremio, Apache Iceberg & Nessie Data Lakehouse On Your Laptop

LIVE

Subsurface

The Data Lakehouse Conference

PRESENTED BY  **dremio**

7,000+

Global Data Experts

46+

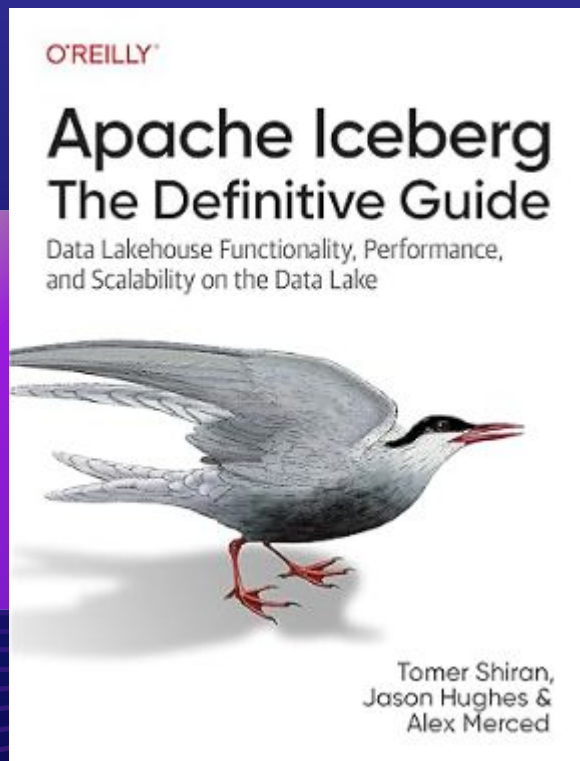
Breakout Sessions

3,000+

Companies

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

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](#)





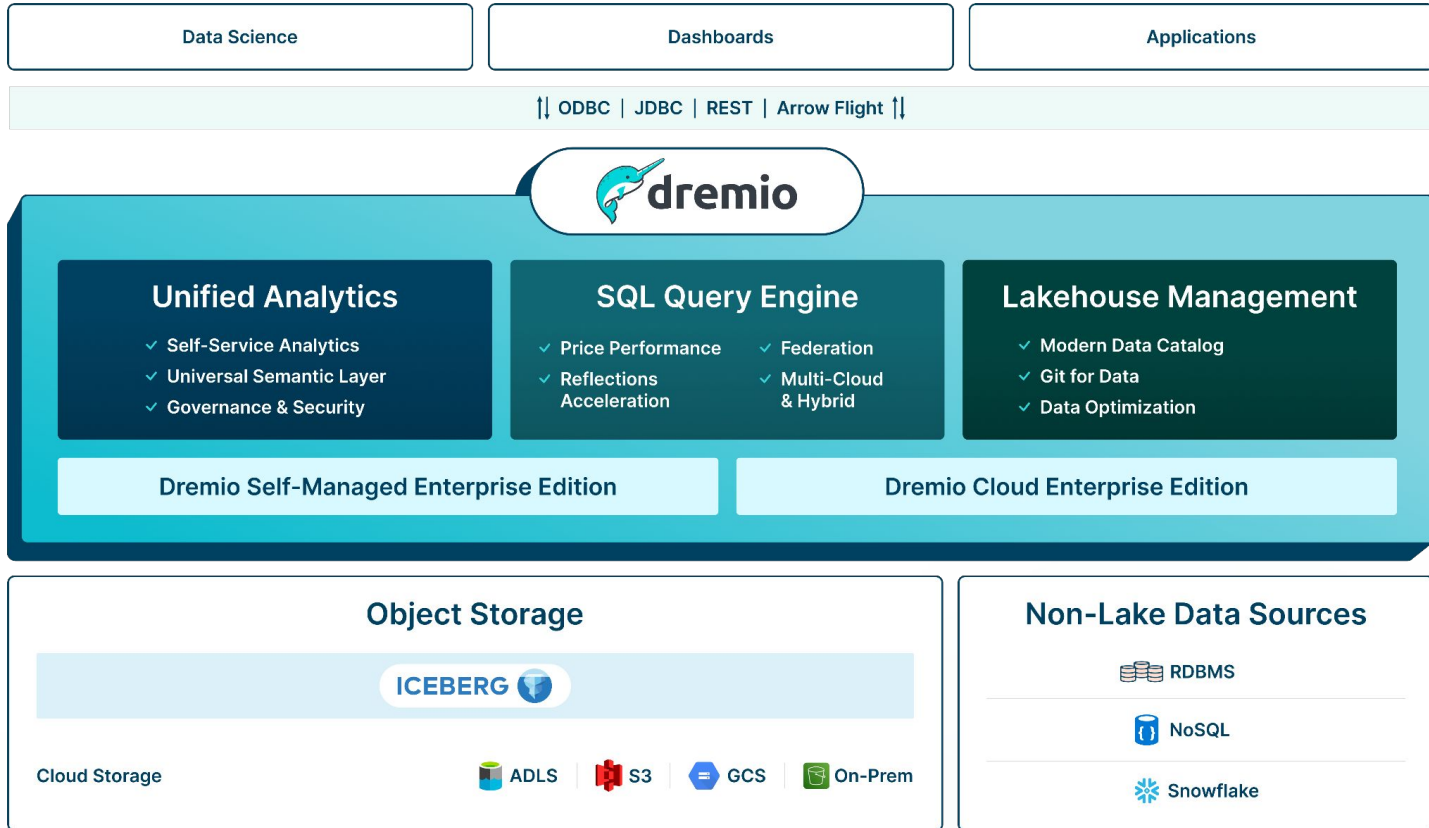
GNARLY
Data_Waves

PRESENTED BY  **dremio**

CE Workshop

Creating a Data Lakehouse On Your Laptop

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



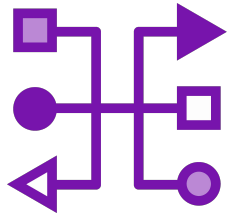
GNARLY
Data_Waves

PRESENTED BY  **dremio**

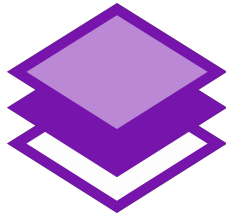
The Value of Dremio

Unified Analytics

Self-Service



Any Data and Any Client



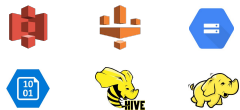
Governed Semantic Layer



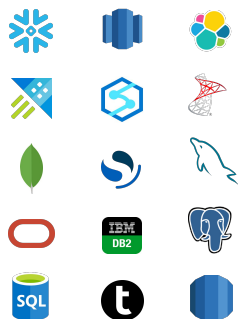
Intuitive User Interface

Any Data and Any Client

Data Lakes



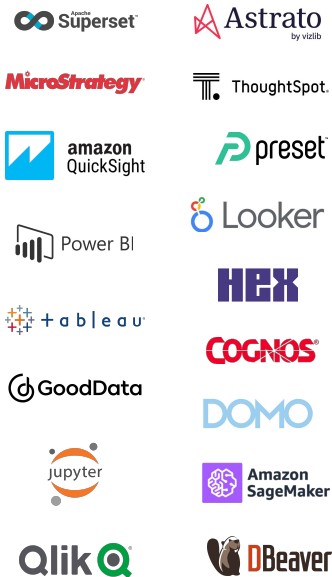
Databases



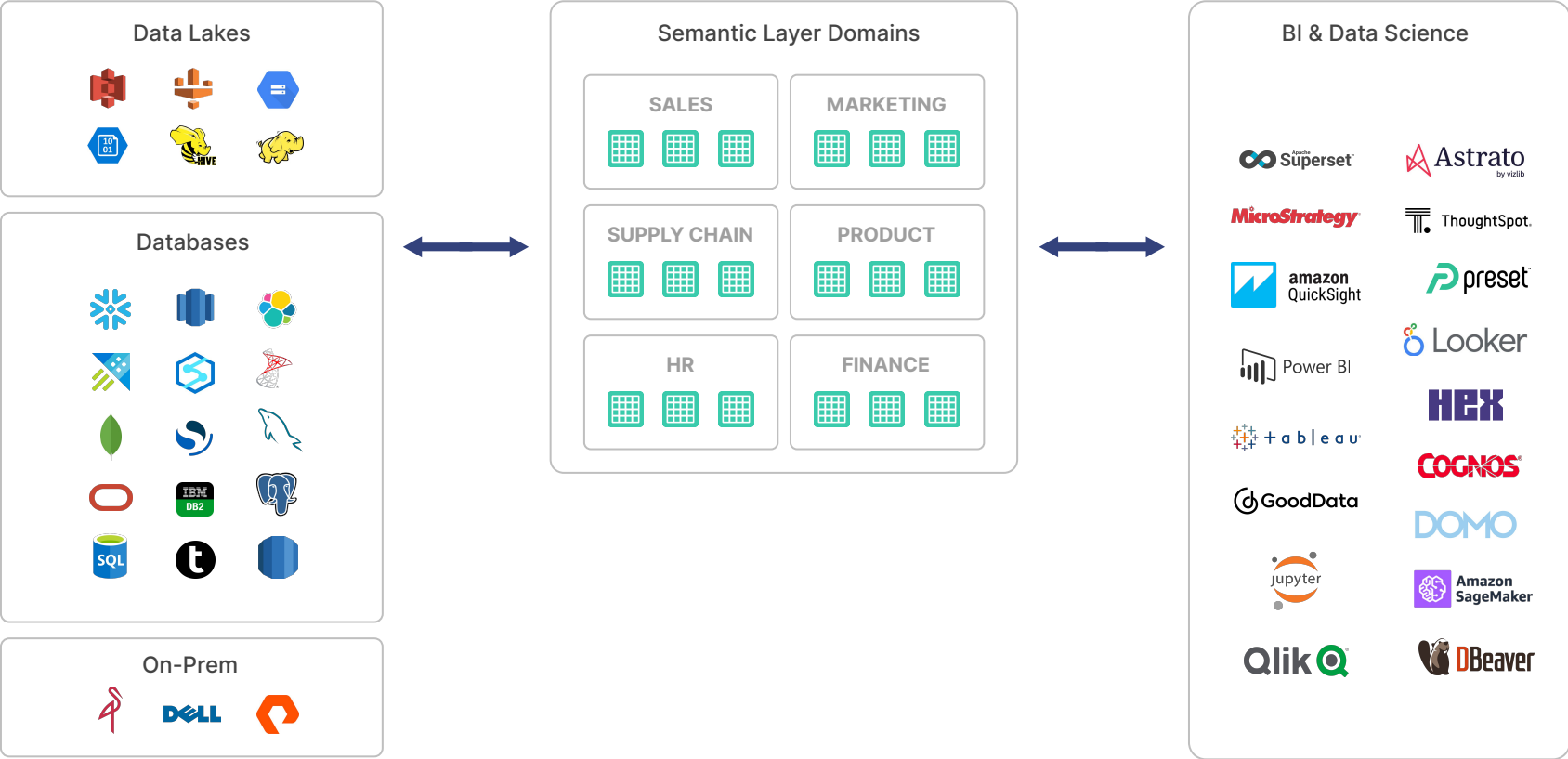
On-Prem



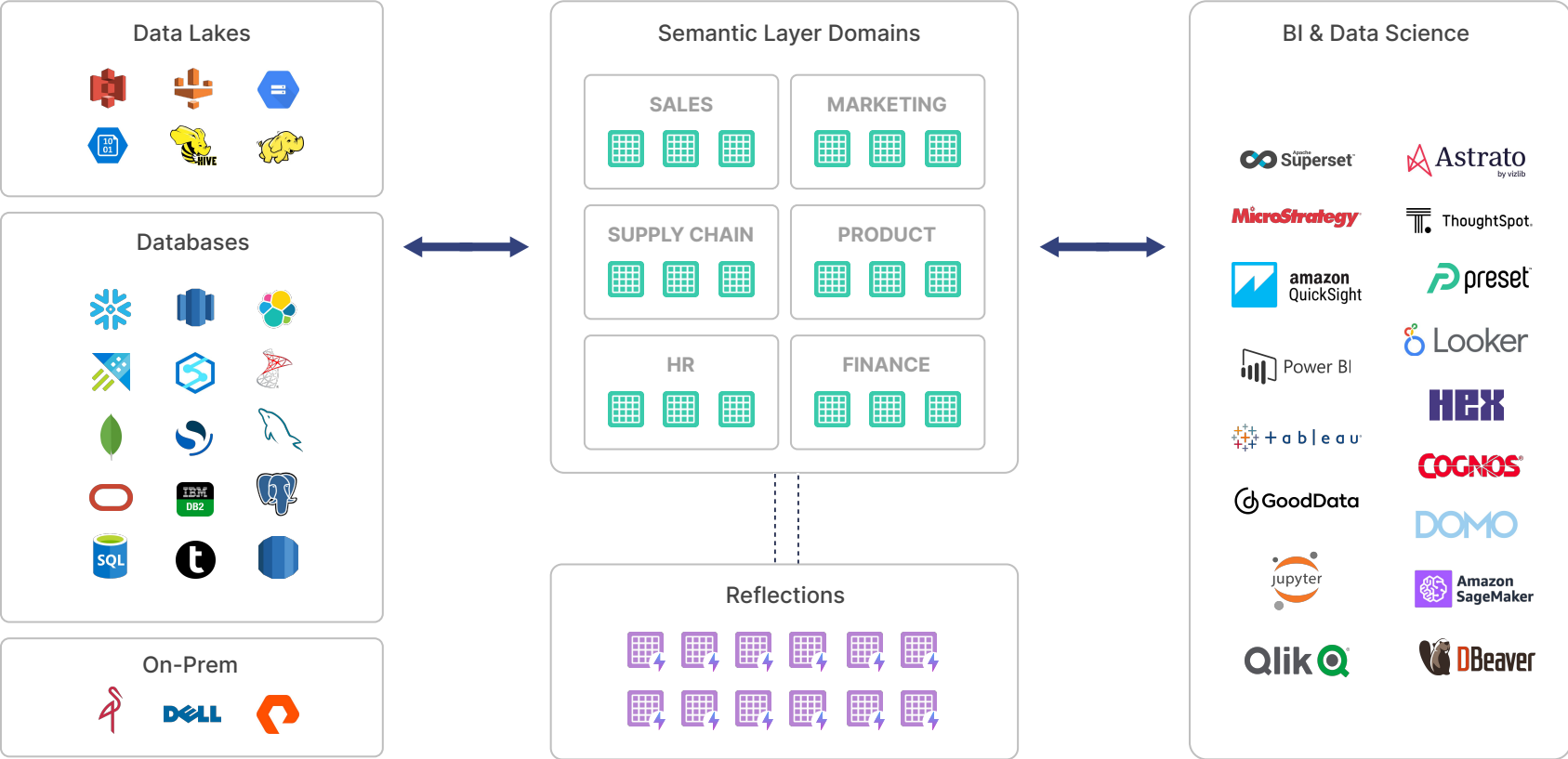
BI & Data Science



Governed, Domain-based Semantic Layer



Governed, Domain-based Semantic Layer



Security and Governance



Fine-Grained
Access Control



End-to-end
Authentication



Compliance

Fine-Grained Access Controls

First-class SQL UDFs

f_x

```
mask_ssn(  
  ssn VARCHAR,  
  co VARCHAR)  
RETURNS VARCHAR
```

```
ALTER TABLE customers  
ALTER COLUMN ssn  
SET MASKING POLICY  
mask_ssn(ssn, country)
```

f_x

```
restrict_location(  
  st VARCHAR,  
  co VARCHAR)  
RETURNS BOOLEAN
```

```
ALTER TABLE customers  
ADD ROW ACCESS POLICY  
restrict_location(  
  state, country)
```

customer_id	ssn	state	country
	***-**-6543	CA	US
	***-**-1212	CA	US
	***-**-9831	CA	US
	***-**-3289	CA	US

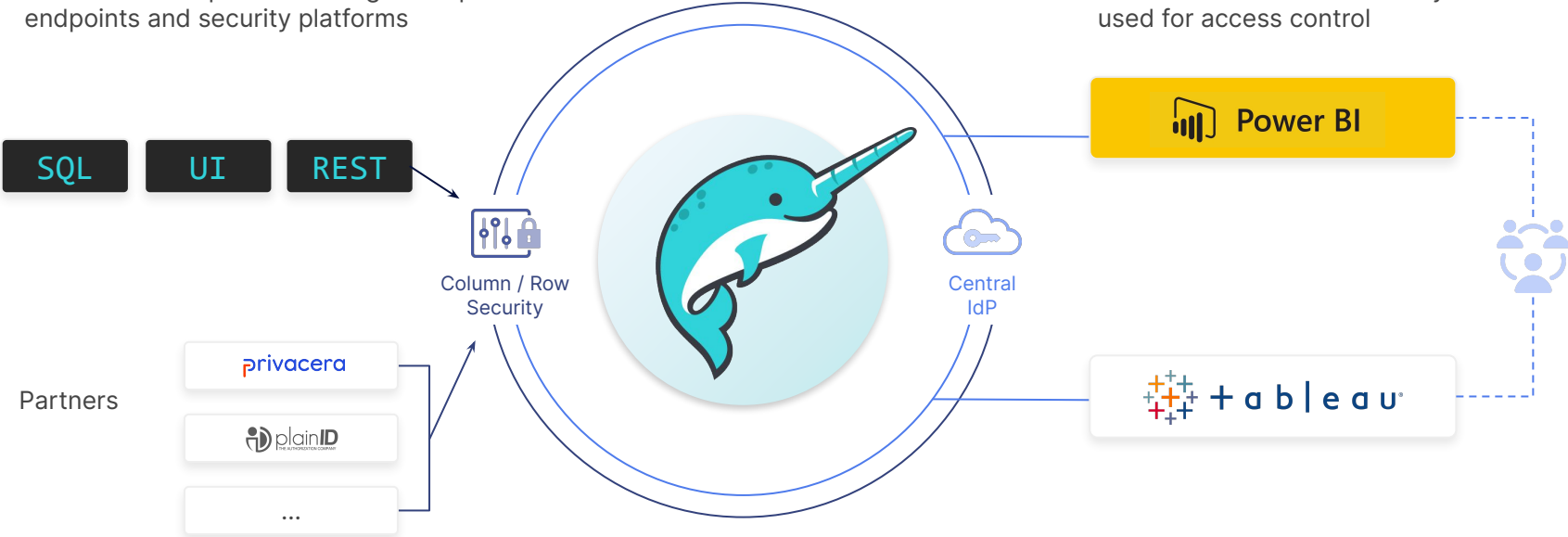
End-to-End Authentication

Policy Control

Flexibility to control native table-, column- and row-based policies through multiple endpoints and security platforms

Single Sign On

Token-based SSO from BI tools ensure the end-user's identity is used for access control



Compliance





GNARLY
Data_Waves

PRESENTED BY  **dremio**

The Value of Dremio

SQL Query Engine

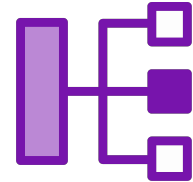
Performance



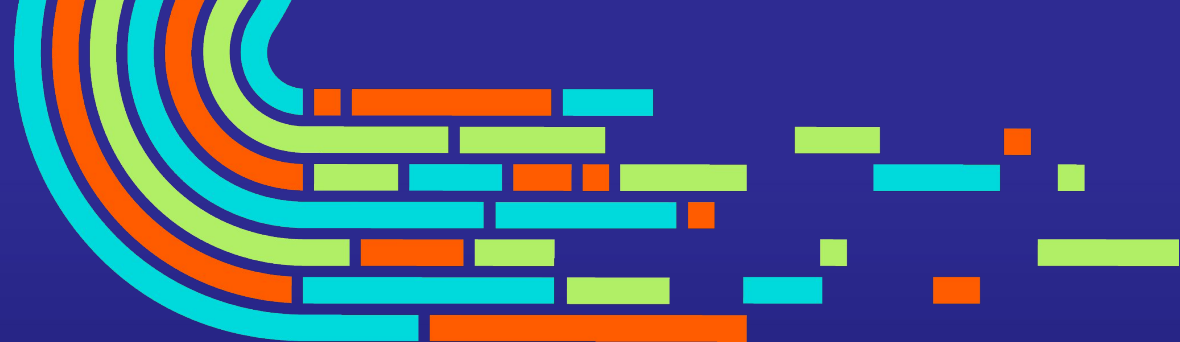
Apache Arrow



Query Acceleration



Unlimited Concurrency



GNARLY
Data_Waves

PRESENTED BY  **dremio**

The Value of Dremio

Lakehouse Management

Automatic Data Optimization

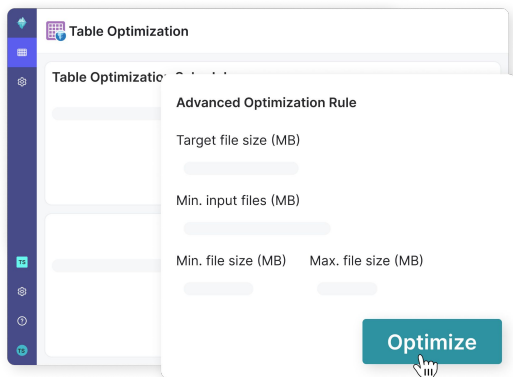


TABLE OPTIMIZATION

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

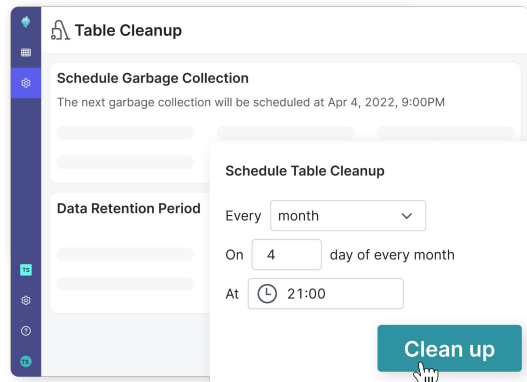
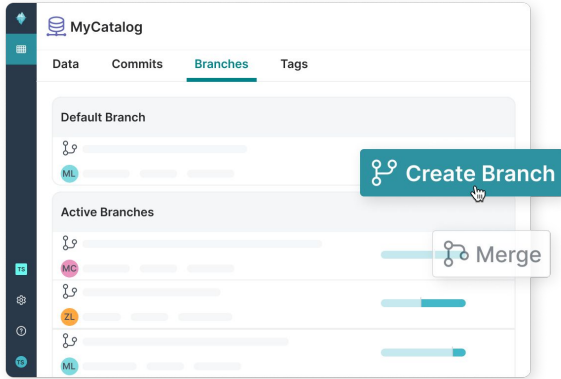


TABLE CLEANUP

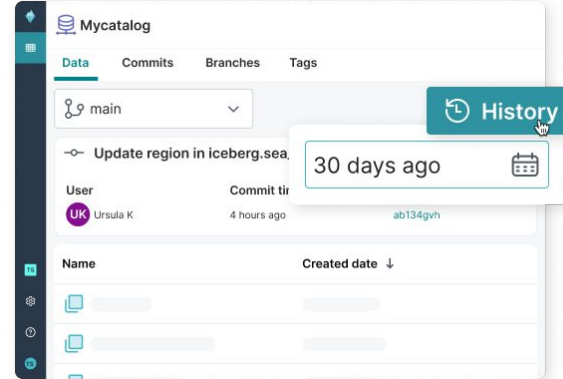
- Dremio Arctic automatically removes unused manifest files, manifest lists, and data files
- Cleanup runs in the background and ensures efficient use of data lake storage

Data as Code



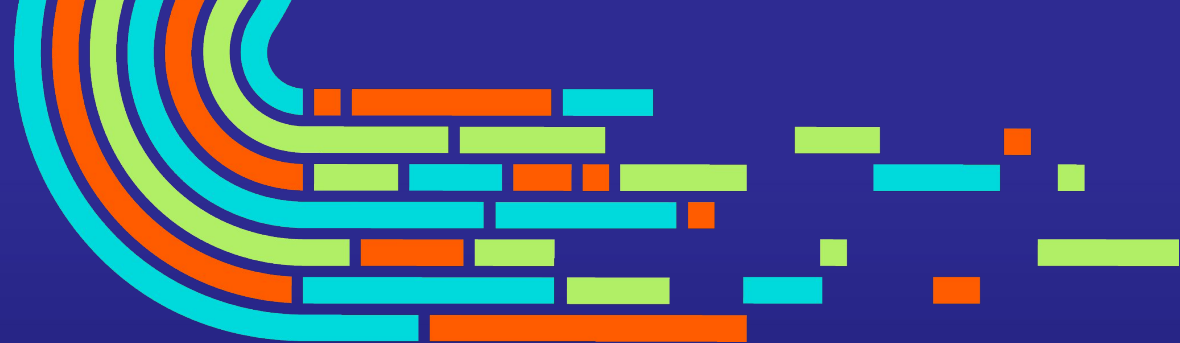
ISOLATION

- 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



GNARLY
Data_Waves

PRESENTED BY  **dremio**

The Value of Dremio

The Exercise



A Iceberg/Dremio Lakehouse on your laptop exercise

Step 1 - Spin Up Our Environment with Docker Compose

Dremio - Lakehouse Platform

Nessie - Open Source Lakehouse Catalog

Minio - Data Lake/Object Storage

For Production Deployments

Cloud Based - Use Dremio Cloud

Self-Managed - Use Our K8's Help Chart

Step 2 - Create a Warehouse Bucket in Data Lake

This will be the bucket where newly created tables will be written to

Step 3 - Login to Dremio

Login to our Dremio Instance

Step 4 - Connect Sources

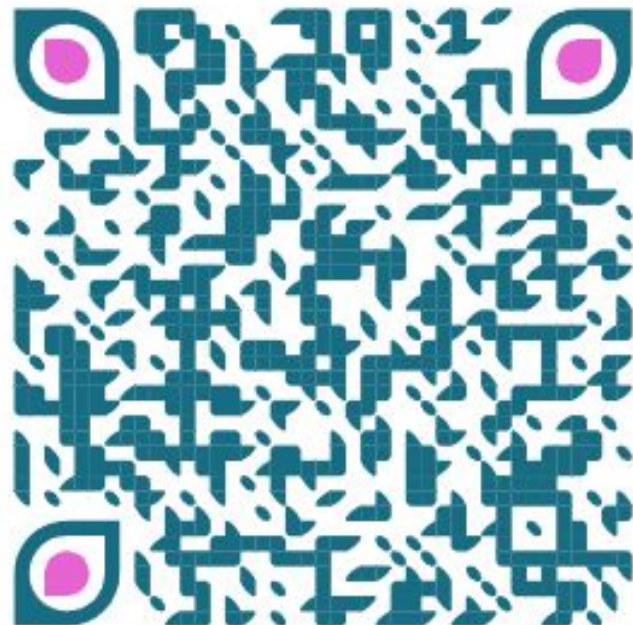
Connect Our Data Sources

Step 5 - Enjoy Your New Lakehouse!

Create Tables, Views, and run analytics!



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



**Deploy Dremio Software or
Dremio Cloud**