

GNARLY
Data_Waves

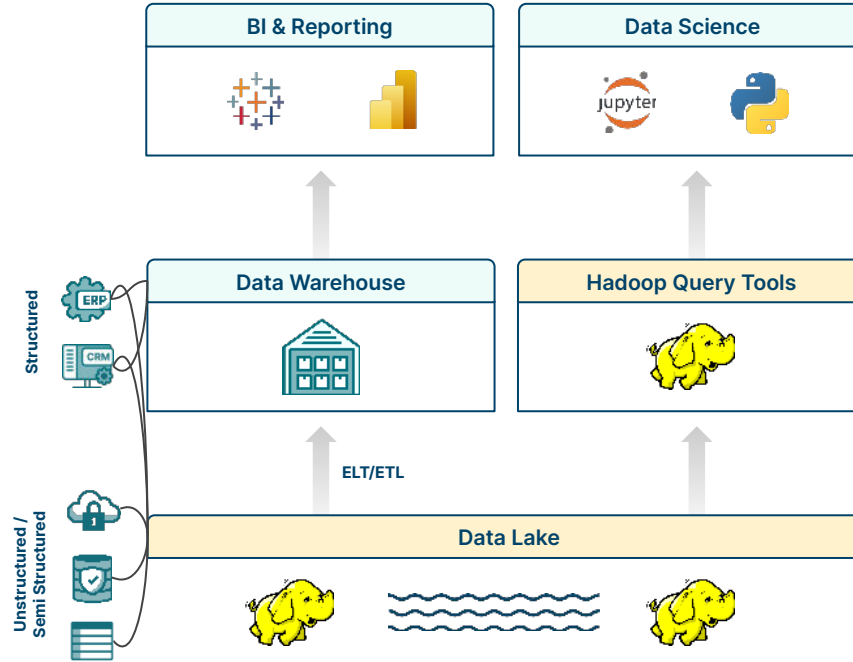
PRESENTED BY  **dremio**

Easy Data Lakehouse Operations with Zero-Copy Clones and Multi-Table Transactions

Agenda

- Data Lakes → Data Lakehouses
- What is Dremio Arctic?
- Zero-Copy Clones with Git for Data
- A Git for Data Use Case: Multi-Table Transactions
- Demo
- Q&A

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



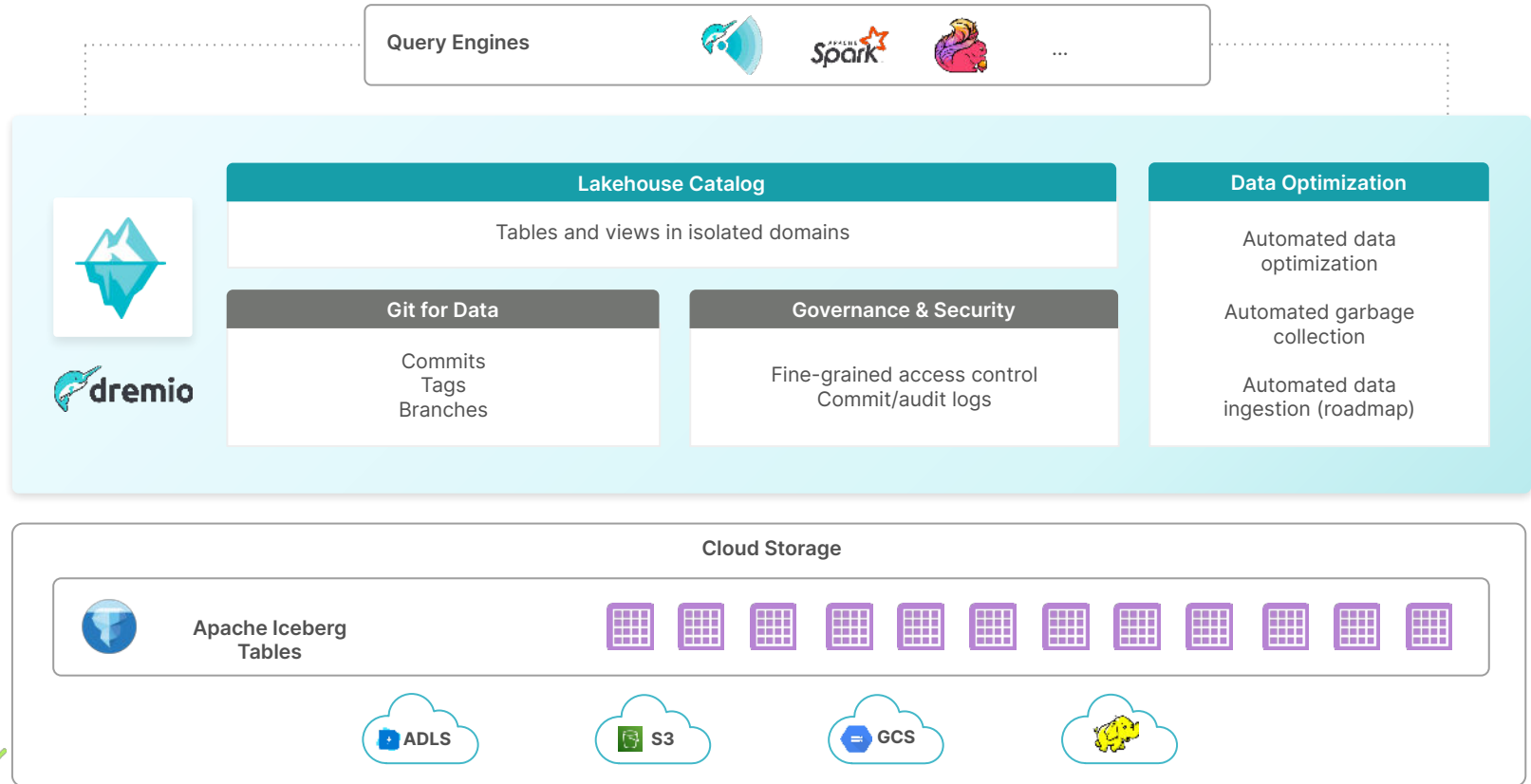
File Formats



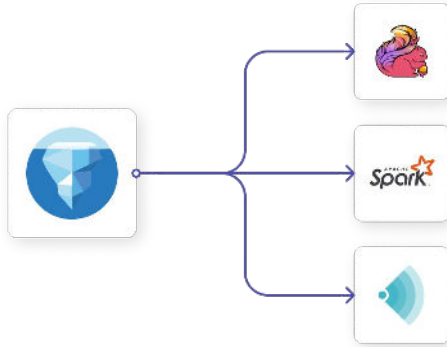
Data Lake Storage



Dremio Arctic is a Data Lakehouse Management Service

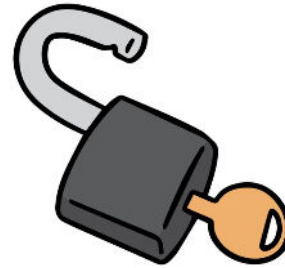


Dremio Arctic is a Modern Lakehouse Catalog



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



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

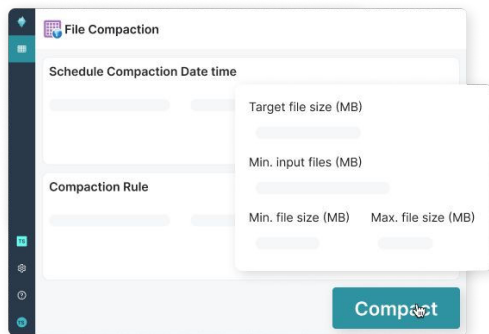


TABLE OPTIMIZATION

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

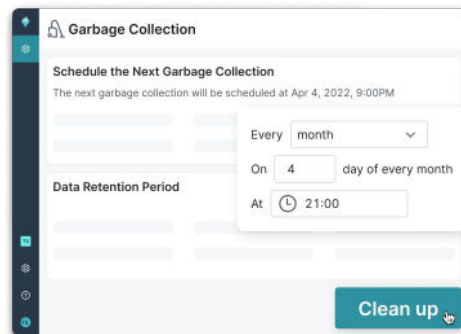
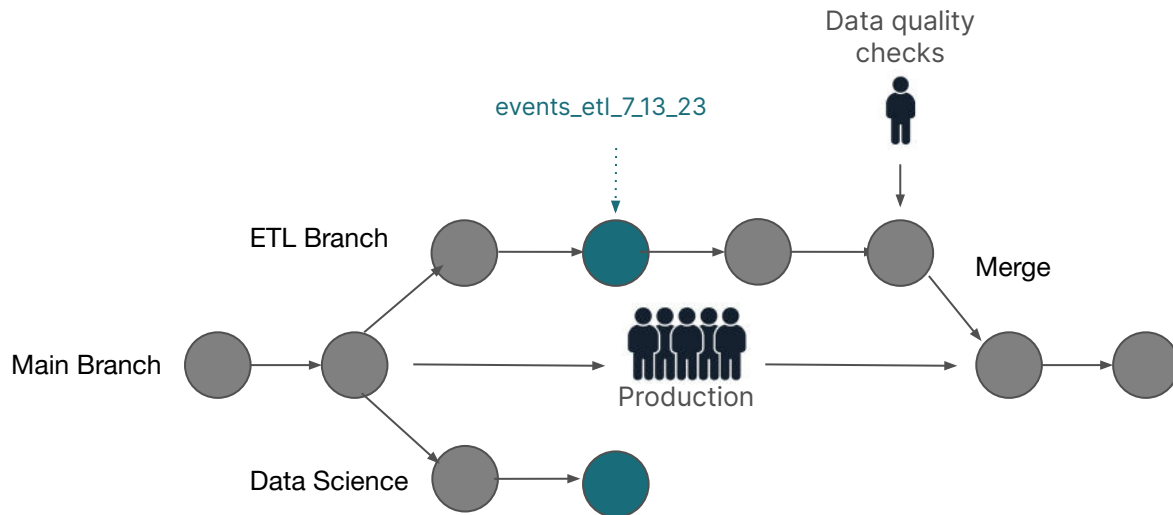


TABLE CLEANUP

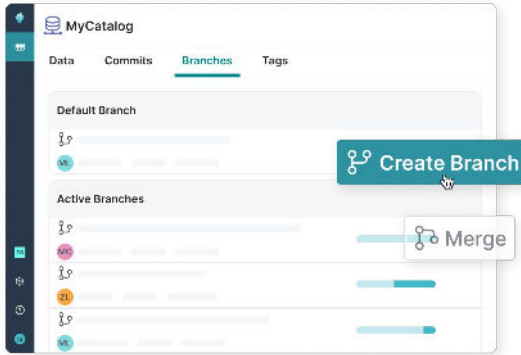
- Dremio Arctic automatically removes unused manifest files, manifest lists, and data files.
- Cleanup runs in the background and ensures efficient use of data lake storage.

Create Zero-Copy Clones of Production Data with Git for Data

Software Development Principles Like CI/CD & Versioning for Data

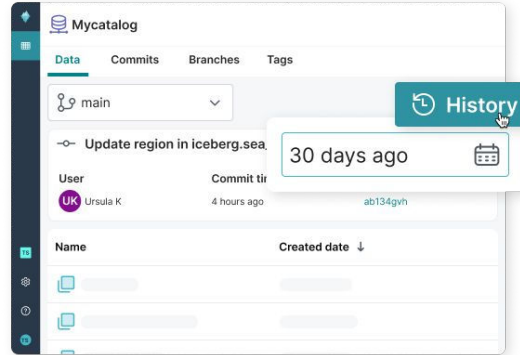


Catalog Versioning with Git for Data



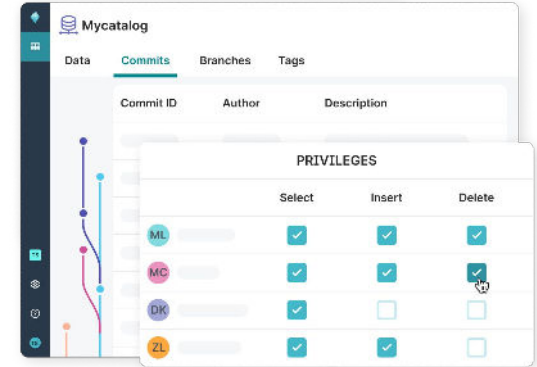
ISOLATION

- 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

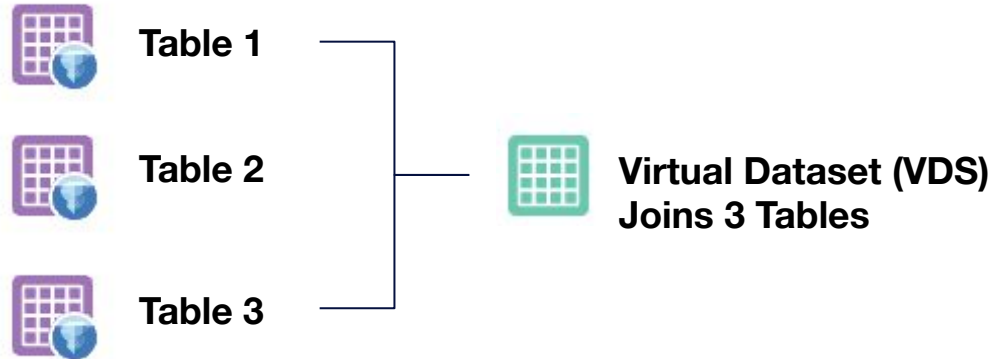


GOVERNANCE

- 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

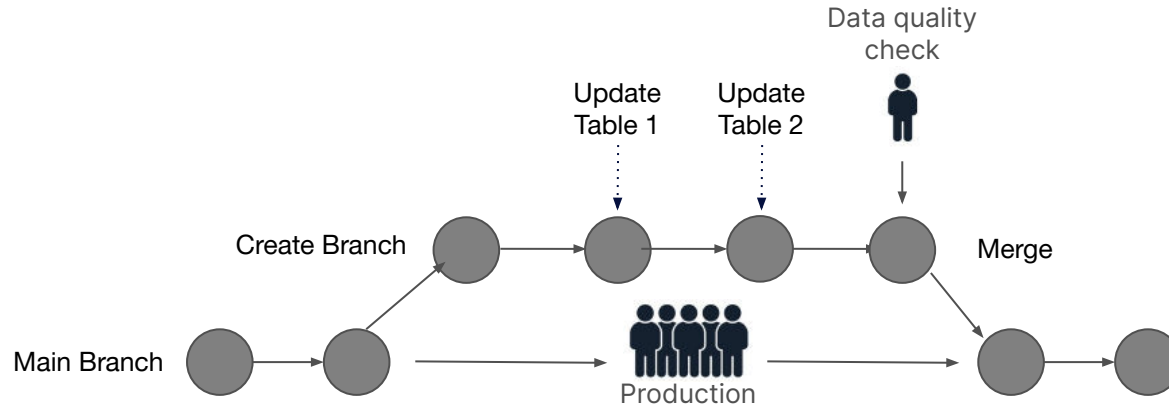
Use Case - Multi-Table Transactions

Problem: I have multiple tables that I need to update, but I don't want to expose my users to the changed data until all of the tables have been updated.



Use Case - Multi-Table Transactions

Solution: Create a branch, make the necessary changes to your tables, and then perform a merge.



Changes to the main branch are made atomically, so no one sees an impartial view of the data.



Demo

Ready to Give it a Try?

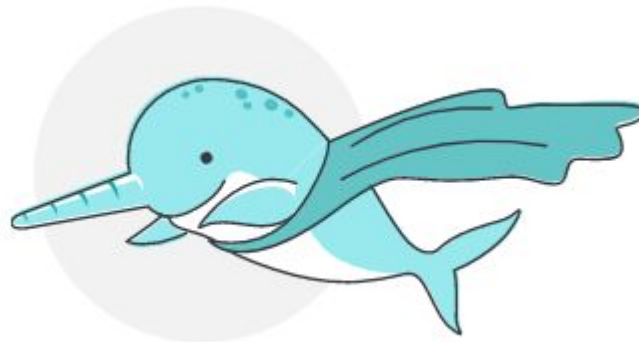
Dremio Arctic is now the default catalog in Dremio Cloud, so every new project will get a Dremio Arctic catalog!

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





Q&A