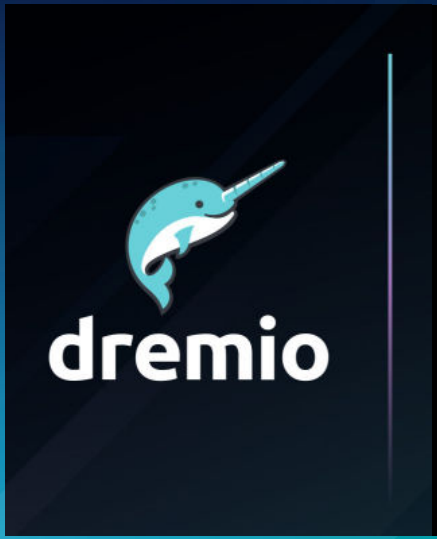



# What's New in Dremio?

Mark Shainman, Dremio Product Marketing

# Agenda

- Dremio Release Focus
- What’s New in Dremio
  - Dremio Cloud
  - Data Ingestion & Migration to Iceberg
  - Reliability & Performance Improvements
  - Reflections
  - Integrated Observability
  - SQL Functions
  - Security
- Get Started

# Fastest path to an Iceberg Lakehouse

# We’re doubling down on offering the fastest path to an Iceberg Lakehouse

- Data Ingestion
- Data Processing
- Data Optimization
- Data Observability


Accelerated Path to an Iceberg Data Lakehouse

# Apache Iceberg

## An Open Table Format for Enterprise Data Lakes

**High-performance queries -** Purpose-built for high performance queries on massive datasets.

**Data warehouse functionality on the data lake -** ACID transactions, time travel, and schema evolution enable more data warehouse workloads directly on data lake storage.

**Easy data operations -** Reduce overhead costs with table optimization, garbage cleanup, and more.

### The Largest Open Source Community

**More** individual companies with **contributions** than any other open table format

**More OSS integrations** than any other open table format.

### Enterprise Companies Using Iceberg

NETFLIX Expedia stripe

airbnb twillo

LinkedIn Adobe Tencent

### Commercial Support for Iceberg

aws dremio snowflake

Google Cloud CLOUDERA

Starburst

# Dremio Cloud

# Now Available: Dremio Cloud on Microsoft Azure

Data Science | Dashboards | Applications

ODBC | JDBC | REST | Arrow Flight

dremio

**Unified Analytics**
- Self-Service Analytics
- Universal Semantic Layer
- Governance & Security

**SQL Query Engine**
- Price Performance
- Reflections Acceleration
- Federation
- Multi-Cloud & Hybrid

**Lakehouse Management**
- Modern Data Catalog
- Git for Data
- Data Optimization

**Dremio Cloud Enterprise Edition on Microsoft Azure**

**Object Storage**

ICEBERG

Cloud Storage: ADLS | S3 | GCS | On-Prem

**Non-Lake Data Sources**

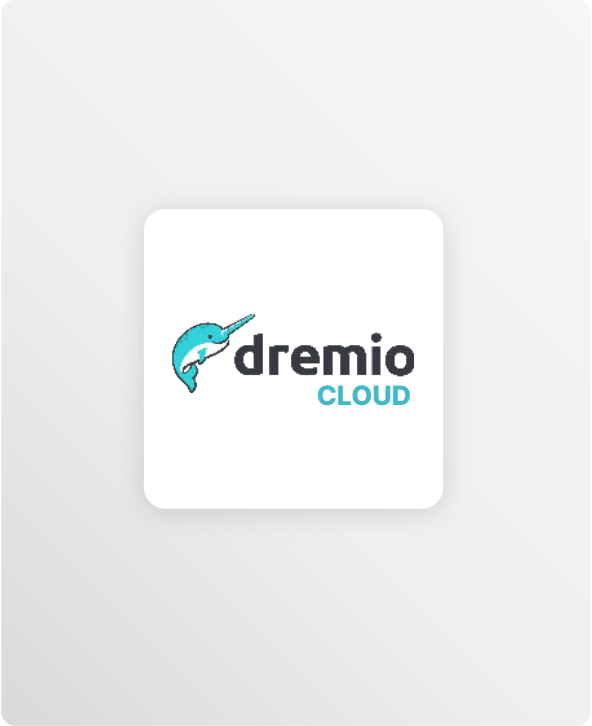
RDBMS

NoSQL

Snowflake

- ✓ Fully Managed Solution
- ✓ Next Generation Architecture
- ✓ Zero Downtime Upgrades
- ✓ Zero Setup & Management Overhead
- ✓ Consumption Pricing
- ✓ Latest Functionality
- ✓ Quickest Time to Value

# **Dremio Cloud in Azure:** Eliminates the pain of managing infrastructure




## Dremio Cloud

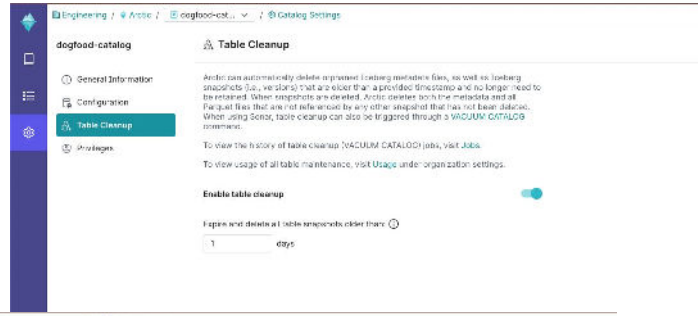
**Split plane architecture with control plane hosted by Dremio and execution hosted in customer tenant.**

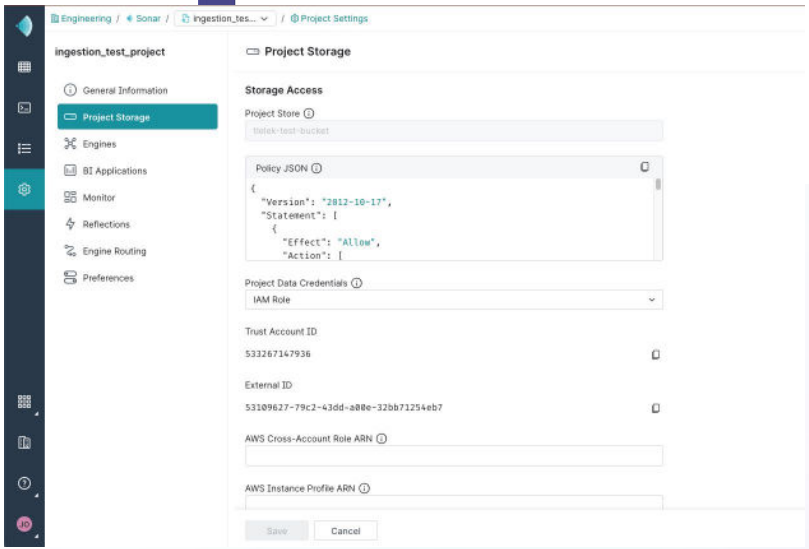
### Cloud Benefits

- Always the latest functionality
- No management overhead
- Zero downtime automatic upgrades
- Automatic scalability
- Passwordless experience integrating with Enterprise SSO IdPs
- Managed Dremio environments with end-to-end encryption
- Capacity-based pricing

# Dremio Cloud Improvements

- **No tedious data management**: Automated VACUUM & clean-up for Iceberg on AWS & Azure




- **More flexible security and access control:** Edit IAM role or access keys for Dremio projects

Automating & Simplifying Maintenance Processes





dremio

# Data Ingestion and Migration into Iceberg

# New: Near-Real-Time Streaming Data Support

- High-speed streaming from Kafa into Iceberg
- Reads event data from Kafka topics and writes it to an Iceberg table
- Deploy and manage connector on your Kafka Connect cluster


Enabling Real-time Analytics on Iceberg Lakehouses

# Faster, Easier Migration to Iceberg

- Seamless **conversion to Apache Iceberg**
- Improved support for **Parquet, JSON & CSV migration**
- Continues to expand and simplify existing **COPY INTO** functionality



Dremio Makes it Quick and Easy to Move to a Iceberg Lakehouse

# Two-step conversion (and some housekeeping!)

Step 1: **Create Your Table** using CREATE TABLE

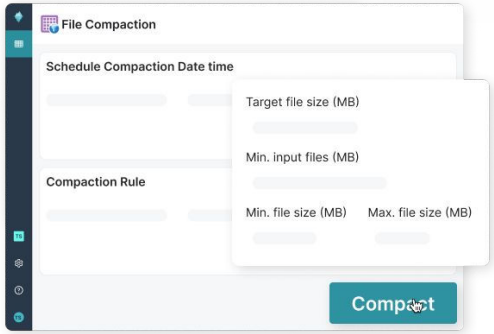
Step 2: **Copy data into Iceberg** using one of the COPY INTO commands

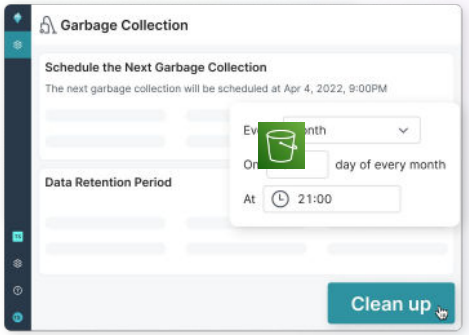
Step 3: **Optimize your tables** using OPTIMIZE TABLE

Step 4: **Clean up your tables** using VACUUM CATALOG

```
-- CREATE AN ICEBERG TABLE FROM OUR SALES TABLE FROM A
LEGACY VERSION OF THE TABLE IN POSTGRES
-- PARTITIONED BY SALES MONTH, SORTED BY SALES TIMESTAMP
CREATE TABLE arctic.db.sales
AS SELECT * FROM postgres.legacy_sales_table
PARTITIONED BY (month(sales_ts))
LOCALSORT BY (sales_ts);

-- COPY INTO SALES DATA FOR DECEMBER 2023 INTO TABLE
-- SALES DATA STORED AS CSV FILES ON OBJECT STORAGE
COPY INTO arctic.db.sales
FROM '@SOURCE/sales/2023/december'
FILE_FORMAT 'csv';

-- OPTIMIZE ALL NEW DATA INGESTED FOR THE LAST MONTH OF
SALES
OPTIMIZE TABLE arctic.db.sales
   REWRITE DATA USING BIN_PACK
   FOR PARTITIONS (sales_ts BETWEEN TIMESTAMP '2023-12-01
00:00:00' AND TIMESTAMP '2023-12-31 00:00:00');

-- VACUUM ALL DATA FROM BEFORE THE 90 DAY DATA RETENTION
POLICY
VACUUM TABLE arctic.db.sales
    EXPIRE SNAPSHOTS older_than '2023-10-03
00:00:00.000';  -- 90 days from January 1st, 2024
```

# Ingest and optimize data into a Iceberg Lakehouse


## INGESTION

- **Event-driven pipelines:** Automatically ingest from Amazon S3, ADLS, GCS (Q3)
- **Continuous ingestion:** Automatically write from Kafka topics into Iceberg

## TABLE OPTIMIZATION

- Automatically **compact small files** and **group similar rows** together
- Table optimization significantly accelerates query performance

## GARBAGE COLLECTION

- Automatically **remove unused data files, manifest files, and manifest lists** (Q4)
- Background cleanup ensures efficient use of data lake storage

## Implementing an Iceberg Lakehouse has Never Been Easier!

# Business-Critical Reliability & Performance

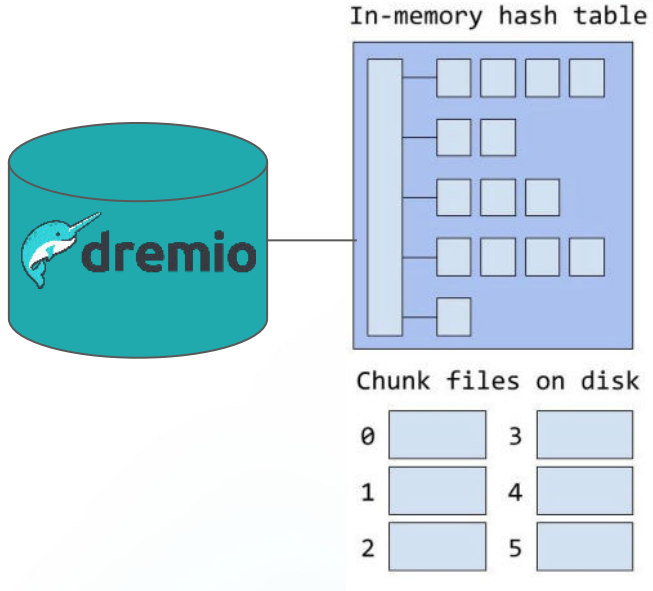
# Mission Critical Reliability with Dynamic Memory Management

- Supports queries at massive scale
- Eliminates out-of-memory failures with dynamic data spilling to disk
- Delivers the most durable and deterministic query experience
- Fewer to no queries will fail at massive scale

Ensuring Business-Critical Analytical Workload Reliability, Stability, and Consistency

# Spillable Hash Joins

- Support for spilling part of hash table to disk when memory is exceeded
- Allows system to support extremely large hash joins
- Eliminates hash join failures when memory is limited



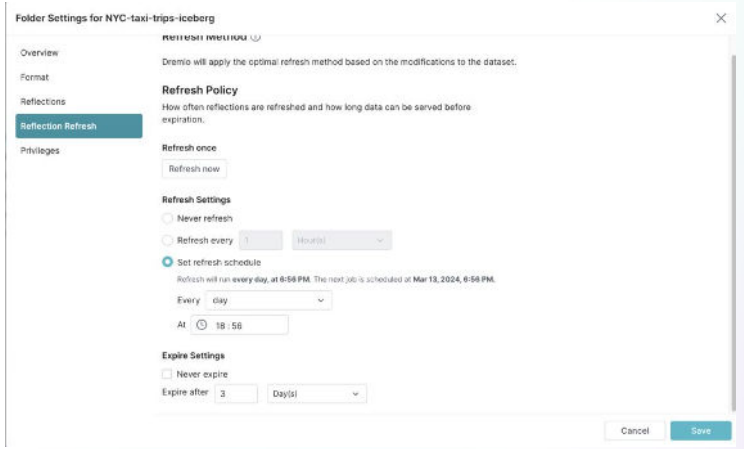

Improving Join Reliability and Scalability

# Query Acceleration with Dremio Reflections

# Smarter, more cost-effective Reflections refresh

- **Smarter incremental Reflections refresh** across complex queries
- **Easily schedule Reflections refresh**
  - Set all days at a set time
  - Set different days of the week at a set time
- **Smarter and streamlined** to bypass unchanged tables

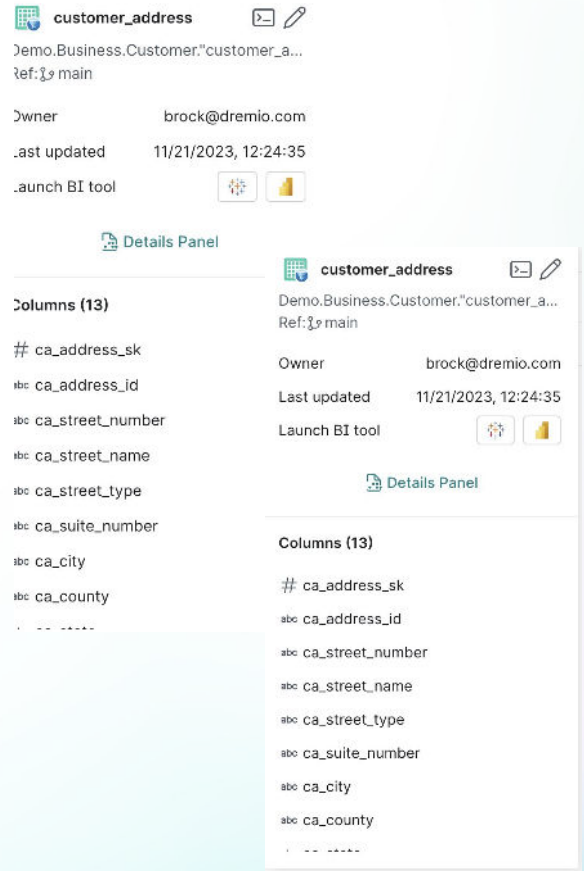
Refresh Scheduler




Market-Leading Query Performance that is Easy to Manage

# Query Acceleration that Supports Governance

- Granular access and governance even with Reflections
- Fine-grained governance supported by Reflections with row & column access control
- Accelerate queries with Reflections on governed tables with defined Row Column Access Control (RCAC) policies








dremio

# Integrated Observability

# Integrated Observability to Improve Administration and Monitoring


## Out-of-the-box metrics for Jobs & Catalogs

- View most active users, queried datasets, queried spaces / folders
- Monitor fluctuation in query volumes, identify resource intensive queries, and longest running queries

## Easier integration with monitoring tools

- Tested, proven support for integrations with Splunk and Datadog

# Expanded SQL Functions

# Comprehensive SQL Coverage… and Growing

**New SQL functions introduced:**

- **Arrays_Overlap()**: Compare if two arrays have at least one element in common
- **Cumulative window framing**: Compute rolling values from beginning of window to current row or from current row to the end of window
- **Sliding window framing**: Compute rolling values between any two rows in window, relative to current row

# Security

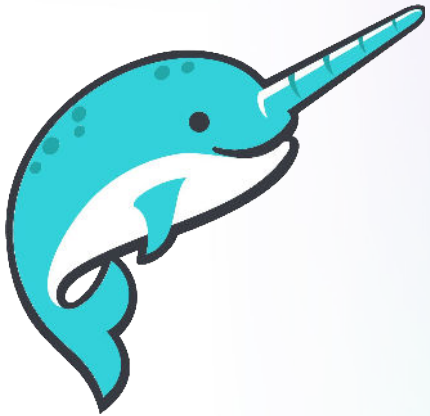
# Secrets Management for Stronger Security

- Hashicorp Vault secrets management integration
- Ensure secure, auditable and restricted access to secrets (aka data source passwords)
- Dremio environment is even more secure and security is easier to manage.

# Get Started

# Ready to get started?

- Current Dremio Software customer? Visit Dremio Support Portal to download.
- Current Dremio Cloud customer? It's live!
- New to Dremio? Try it for free using Dremio Cloud or the Community Edition for Dremio Software



# Thank You