

Simplifying Data Governance at Scale Across All Your Data



Colleen Quinn

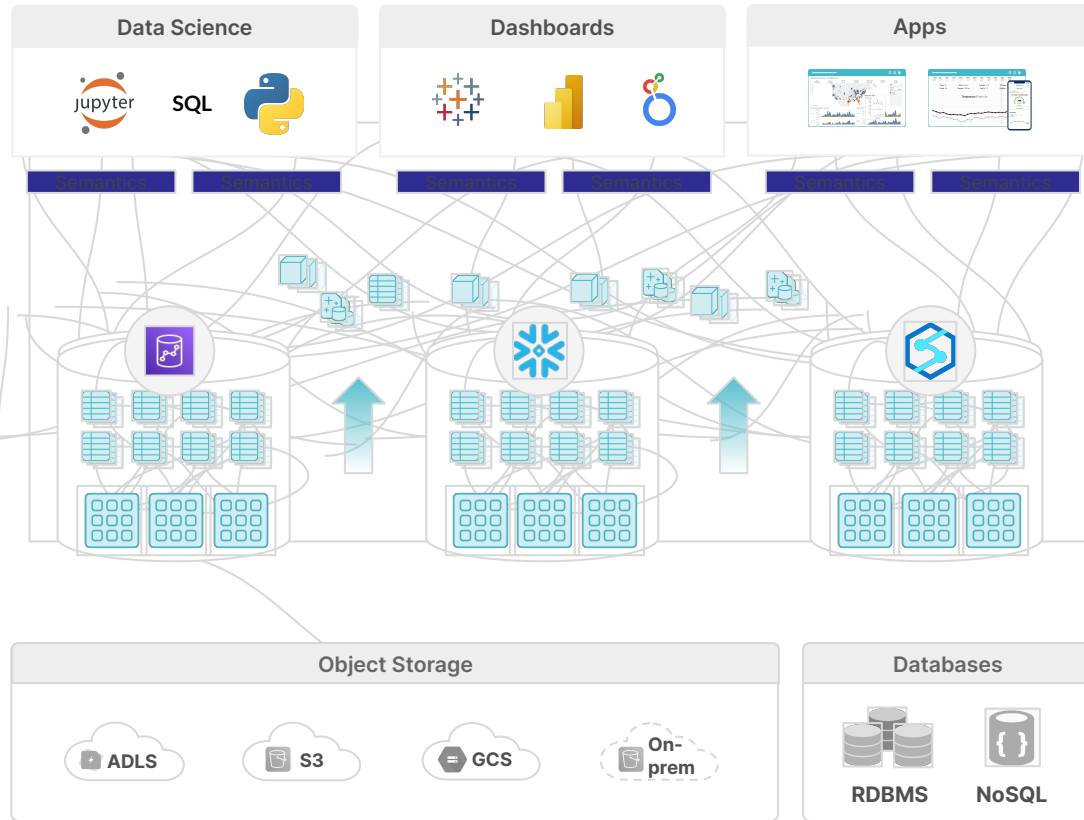
Product Marketing, Dremio





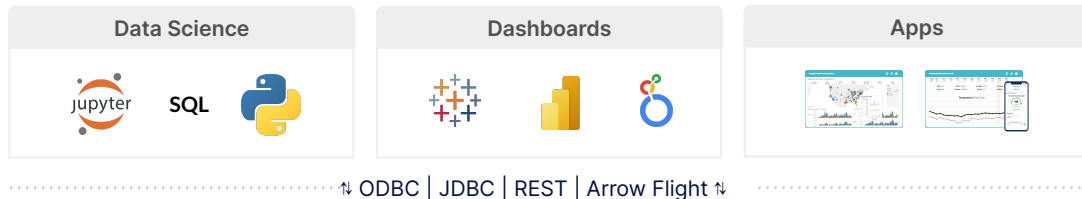
About Dremio

Data Warehouses: Complex, Proprietary, Expensive

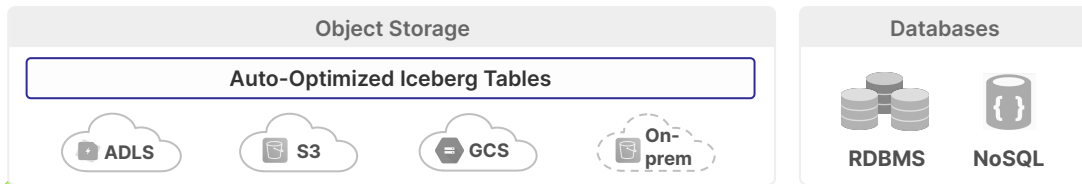


- × Complex
- × Expensive
- × Lock-in
- × Impossible to secure
- × No self-service
- × Limited data exploration
- × Inconsistent data

Dremio Data Lakehouse: Easy, Open, 1/10th the Cost

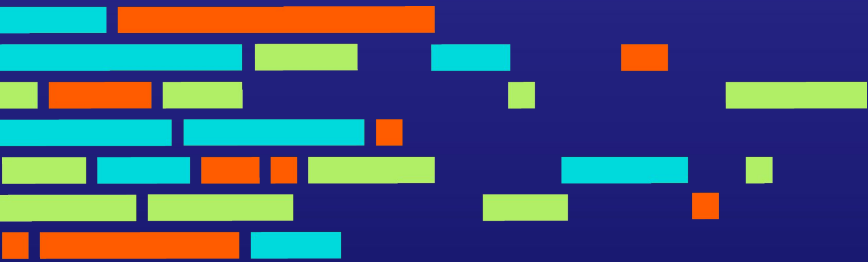


- ✓ No data copies
- ✓ No complex ETL pipelines
- ✓ Inexpensive
- ✓ Self-service analytics
- ✓ Data engineering productivity
- ✓ Security
- ✓ Governance
- ✓ Consistent data



Agenda

- Why is Data Governance important?
- Zoom-in on key technology considerations
 - Data Architecture
 - Data Security
 - Compliance & Data Quality



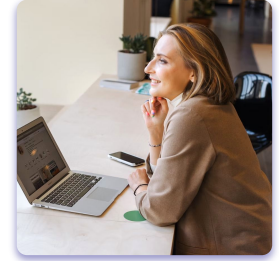
A data story

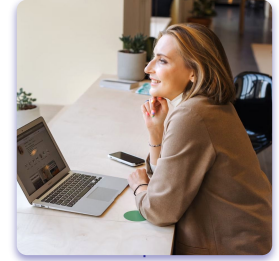


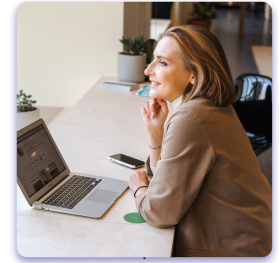


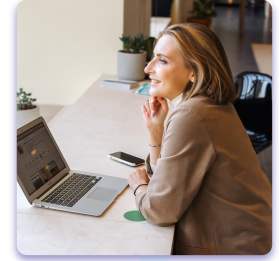










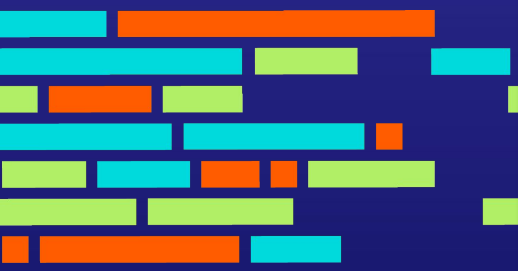




5

Copies





What have we done?

- More data storage
- More data to govern
- No single source of truth
- Errors introduced to the data
- Increased compliance & regulatory risk



Why does Data Governance matter?

- **Improved Data Quality:** Data accuracy, completeness, and consistency, for better decision-making
- **Compliance & Regulatory Requirements:** Comply with relevant laws and regulations
- **Data Security and Privacy:** Safeguard sensitive information and protect against unauthorized access





Technology Considerations for Data Governance

Challenges with traditional analytics & data architecture



Lose agility with complex data architecture



No single source of truth



Lack of comprehensive data inventory

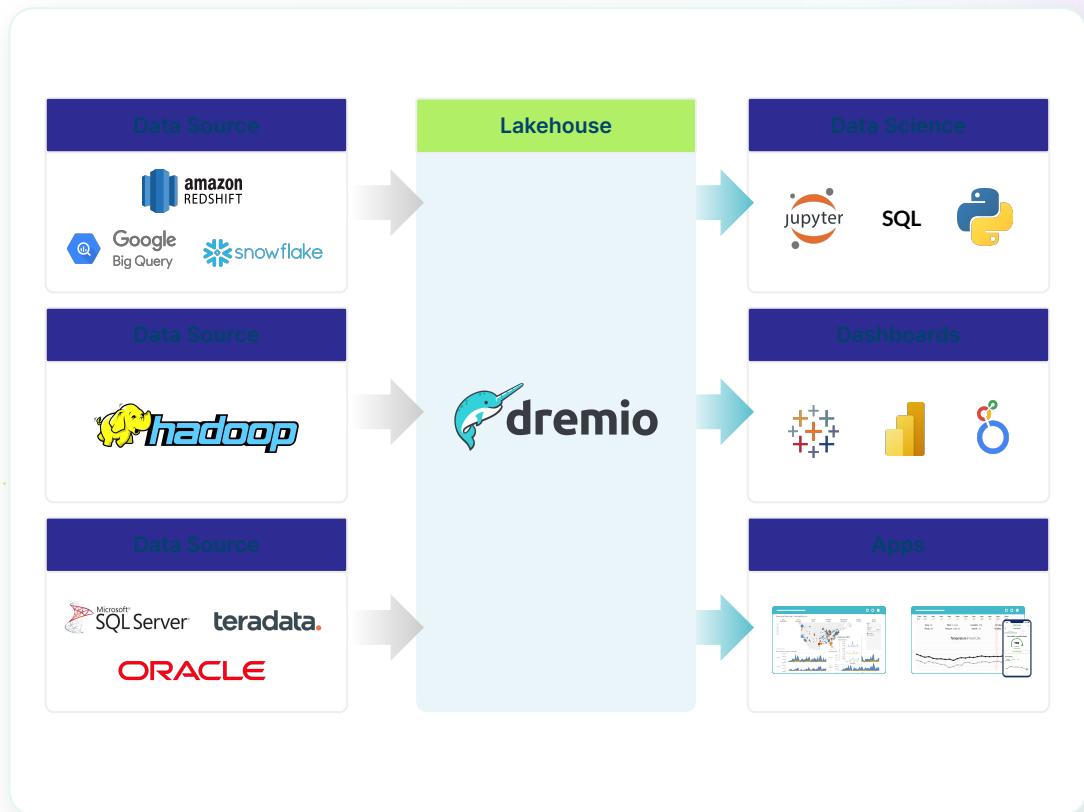
Data Architecture for Data Governance

What does it look like?

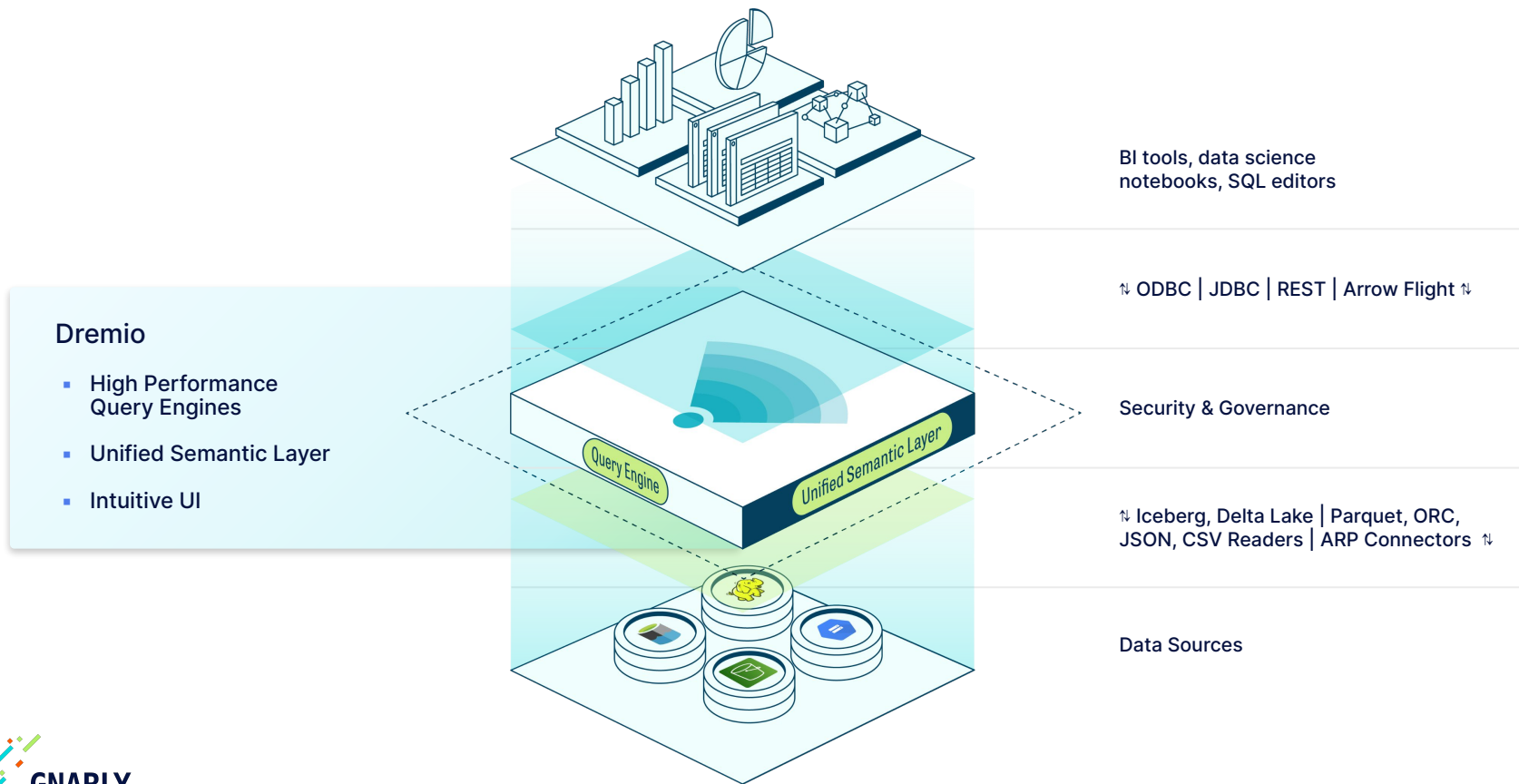
- ✓ Make your data lake transactional
- ✓ Unify all your data for federated governance
- ✓ Reduce ETL and data movement
- ✓ Enable easier data discovery

Results:

- ✓ All your data is available for business decision-making no matter where it lives, increasing business agility



Unifying and managing data with Dremio



Dremio Partner Ecosystem

DATA PRIVACY & GOVERNANCE

PROTEGRITY

OKERA



PRIVAGERA



Collibra



plainID



ETL/TRANSFORMATION

upsolver

dbt Labs

StreamSets

Airbyte

DREMIO Lakehouse

An open data lakehouse engine powered by Apache Arrow

Semantic Layer

Query Acceleration

ANSI SQL Engine

DREMIO Data Management

A data lakehouse management service

Data-as-Code

Automatic Data Optimization

Data Catalog

BI/VISUALIZATION

+ a b l e a u

ThoughtSpot

IBM Cognos Analytics

HEX

Power BI

MicroStrategy

GoodData

Looker

preset

CLOUD AND INFRASTRUCTURE



Azure



VAST

DELL Technologies

PURE STORAGE

Hewlett Packard Enterprise



NetApp

Challenges with Data Security

- Managing permissions at scale is complex
- Access control is coarse-grained
- It's hard to understanding where data comes from and how it's being used



Data Security for Data Governance

- Role-based access control to simplify and scale permissions
- Fine-grained access control: table, column, row
- Seamless support for data sharing architectures, like data mesh

Fine-grained access control in Dremio

Source Settings

General

Advanced Options

Reflection Refresh


Metadata

Privileges

Add User/Role

Business Users

Add to Privileges

Users	Select	Alter	Create Table	Drop
 All Users	✓	✓		

dremiocloud-demo / Sonar / Field D

Field Demos

colleen.quinn@dremio.com

Spaces (12)	
Application	14
Business	2
Citibike_space	4
demo	8
demo1	6
NYC Taxi	2
Preparation	7
Security	.1
Shared	0

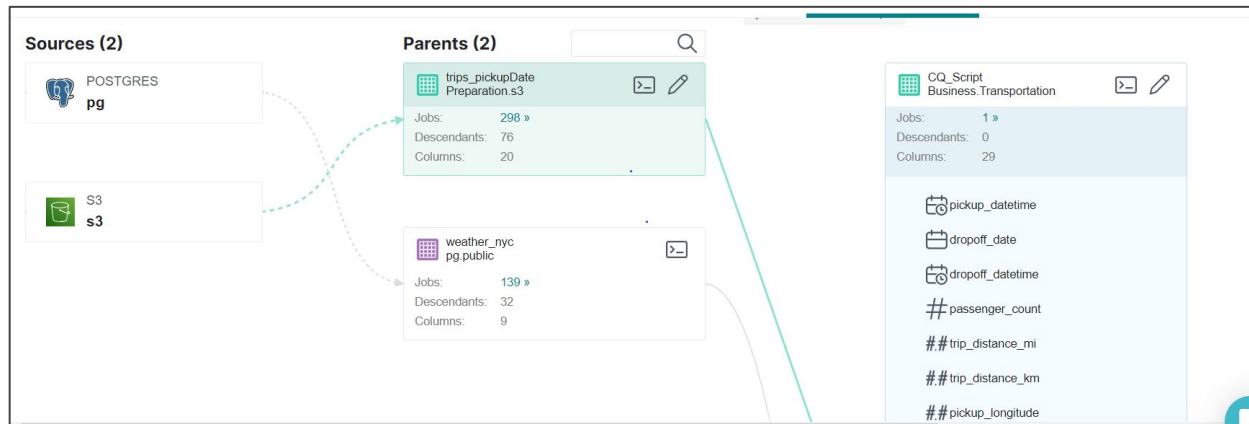
Challenges with Compliance & Data Quality

- Complex and shifting regulatory policies
- Incomplete auditing picture
- Risk of sensitive data being shared
- Data lakes becoming data swamps
 - Lack of lineage
 - Data errors

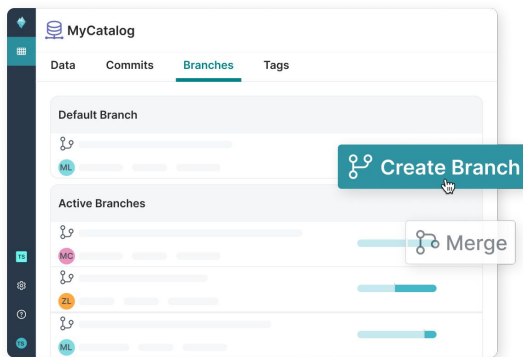


Compliance & Data Quality for Data Governance

- Comprehensive data lineage
- Complete data access auditing
- Fine-grained access control for sensitive data
- Data product owner management

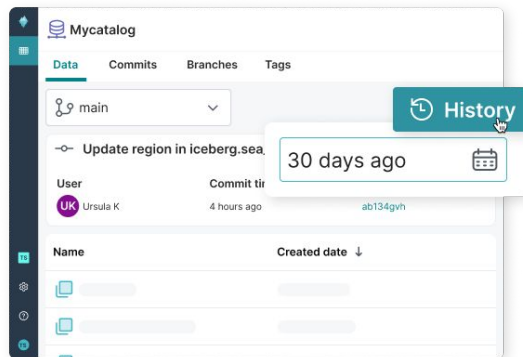


Simplify Data Engineering with Data-as-Code



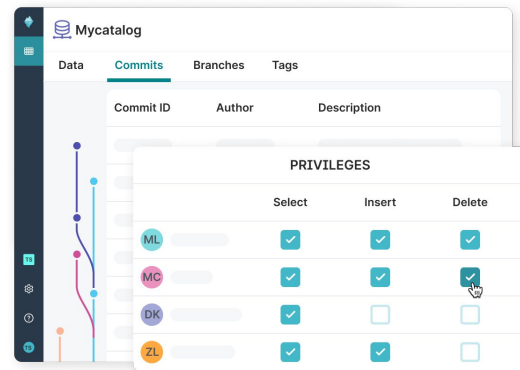
BRANCHING

- **Ingest, transform, and validate data** on isolated branches and **instantly promote changes** to production
- **Experiment with data immediately** using branches instead of creating data copies



VERSION CONTROL

- **Reproduce models and dashboards** from historical data based on time or tags
- **Roll back changes instantly** and undo mistakes without downtime



GOVERNANCE

- **Track all changes** to data and metadata: who accessed what data and when
- **Control access** using role-based access control (Q3) and fine-grained privileges (Q4)

Data Governance to be Data Driven

85%

of businesses
want
to be data
driven

37%

have been
successful

“Data governance is **no longer optional** for enterprise organizations. They are finally realizing the value of data as an asset that needs to be protected, managed, and maintained to increase asset value.”

IDC

“Organizations lack data knowledge for efficient and effective data governance activities; **30% of time spent on data governance is wasted.**”

IDC



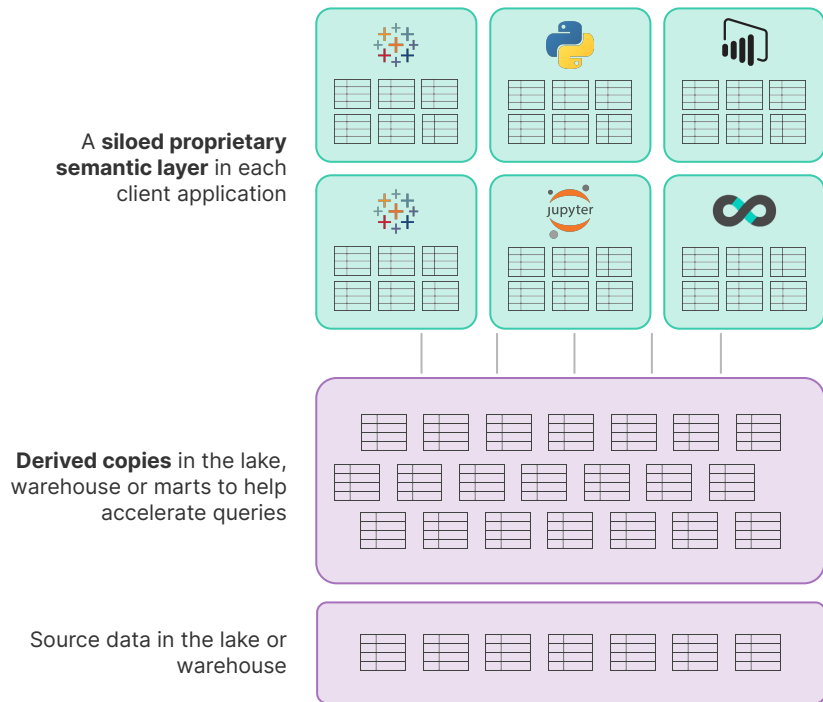
Thank you!



Questions?

Shared Semantic Layer for All Users & Tools

Without Dremio

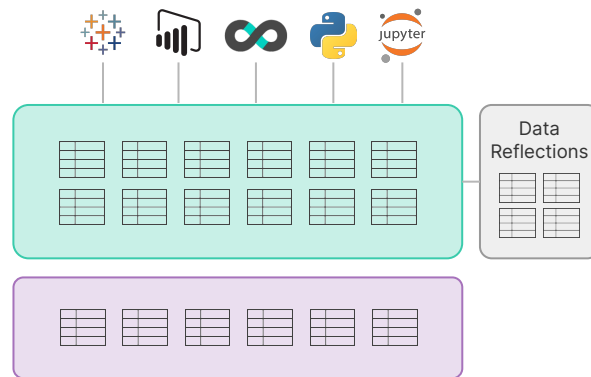


With Dremio

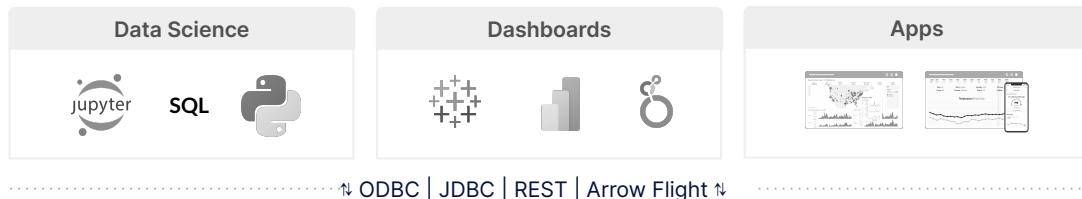
- ✓ Consistent semantics across all client applications
- ✓ Centralized security & governance
- ✓ Fast time to value

BI tools are used as **thin clients** (no semantic layer and no extracts)

A **shared semantic layer** for all users and tools, with decoupled acceleration



Dremio Data Lakehouse: Easy, Open, 1/10th the Cost



Third Party Engines

Spark
Flink
Presto
DuckDB
....



Dremio Sonar

Unified Semantic Layer
Powered by Reflections

Distributed Execution Engine
Powered by Arrow



Dremio Arctic

Lakehouse Management Service with:

- Data as Code
- Data Governance
- Table Optimizer

- ✓ No data copies
- ✓ No complex ETL pipelines
- ✓ Inexpensive
- ✓ Self-service analytics
- ✓ Data engineering productivity
- ✓ Security
- ✓ Governance
- ✓ Consistent data

Object Storage

Auto-Optimized Iceberg Tables



Databases



RDBMS



NoSQL

Data Architecture

Challenges

- Legacy architectures
- Data silos
- Lack of comprehensive data inventory
- Disorganized data

Needs

- Data lake and lakehouse first
- 360-degree data view and access
- Easy to inventory and discover data
- Ability to logically organize and describe data

What is Data Governance?

Data Governance is the collection of policies, processes, and systems that organizations use to ensure the quality and appropriate handling of their data throughout its lifecycle for the purpose of generating business value.