



GNARLY
Data_Waves

PRESENTED BY  **dremio**

EPISODE 54

Mastering Semantic Layers: The Key to Data-Driven Innovation



Andrew Madson
Technical Evangelist, Dremio



Gnarly Data Waves Host — Andrew Madson

Developer Advocate - Tech Evangelist
Dremio

Andrew is an experienced data leader with 17 years experience leading technical teams. Andrew has held senior leadership positions at institutions such as JP Morgan, LPL Financial, MassMutual, and Arizona State University.

In addition to leading data teams, Andrew is a professor of data science and analytics at several universities where he teaches graduate courses in machine learning, statistics, SQL, R, python, Tableau, and Power BI.

Let's Connect!



Andrew Madson

Developer Advocate - Evangelist



Today's Agenda

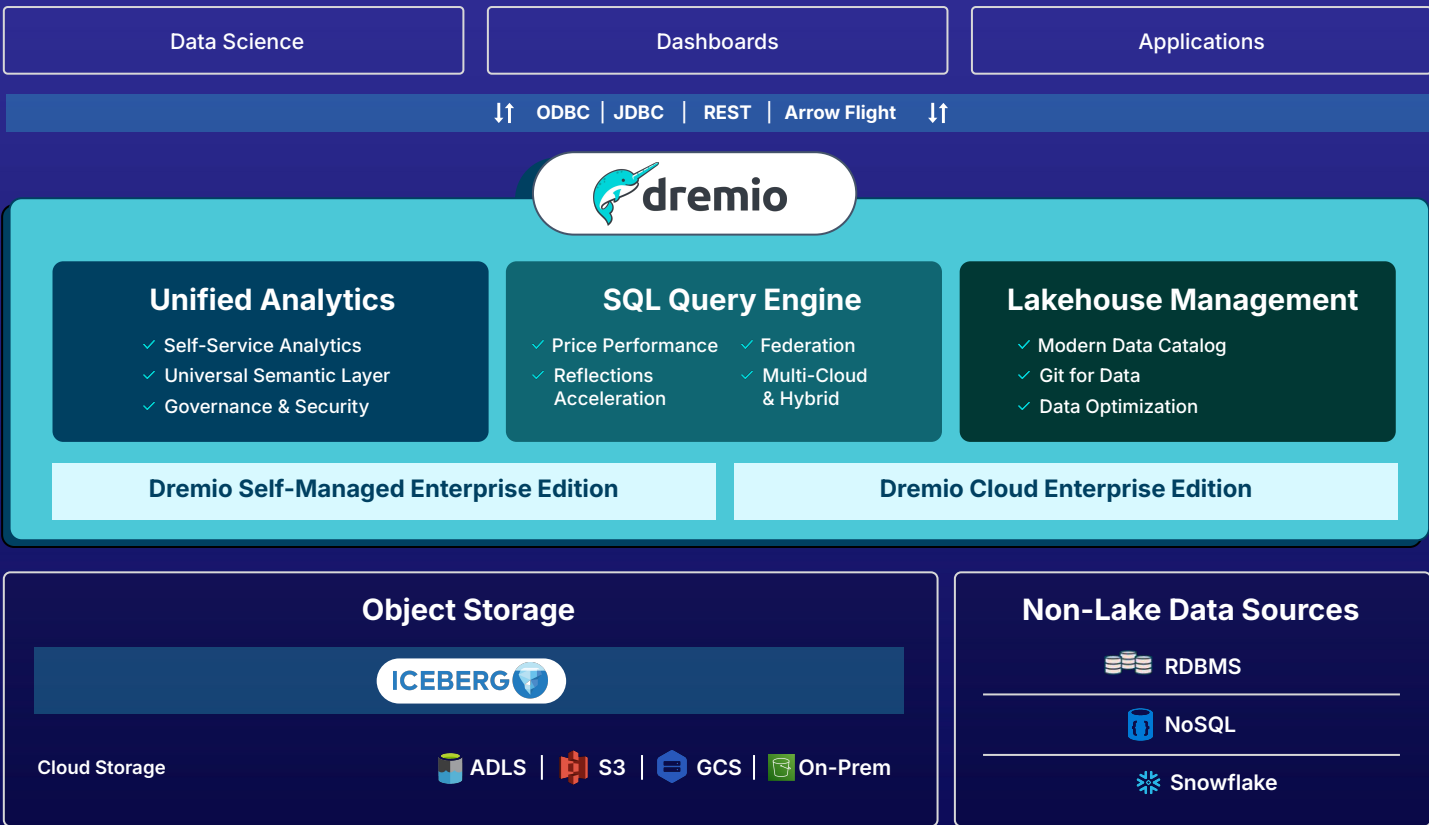
Semantic Layers

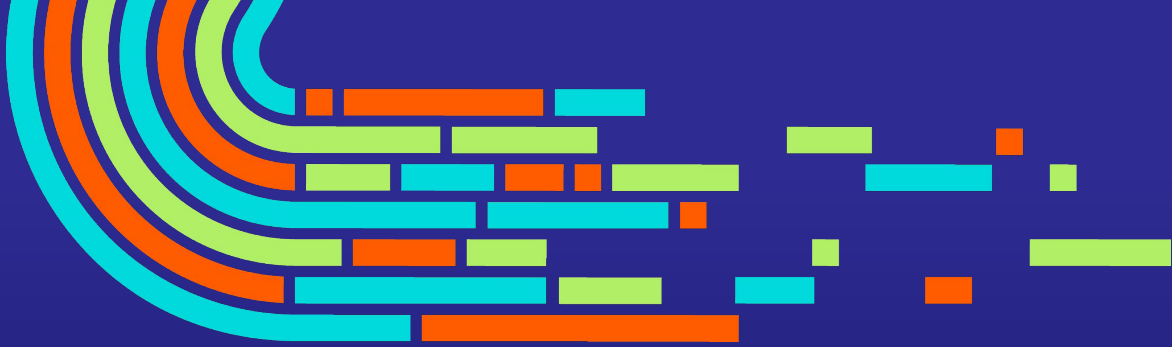
Today we'll cover

1. Current state of business intelligence
2. What is a semantic layer?
3. How do semantic layers enable decision making?
4. Industry trends
5. Dremio's approach to semantic layers



The Unified Lakehouse Platform for Self-Service Analytics & AI

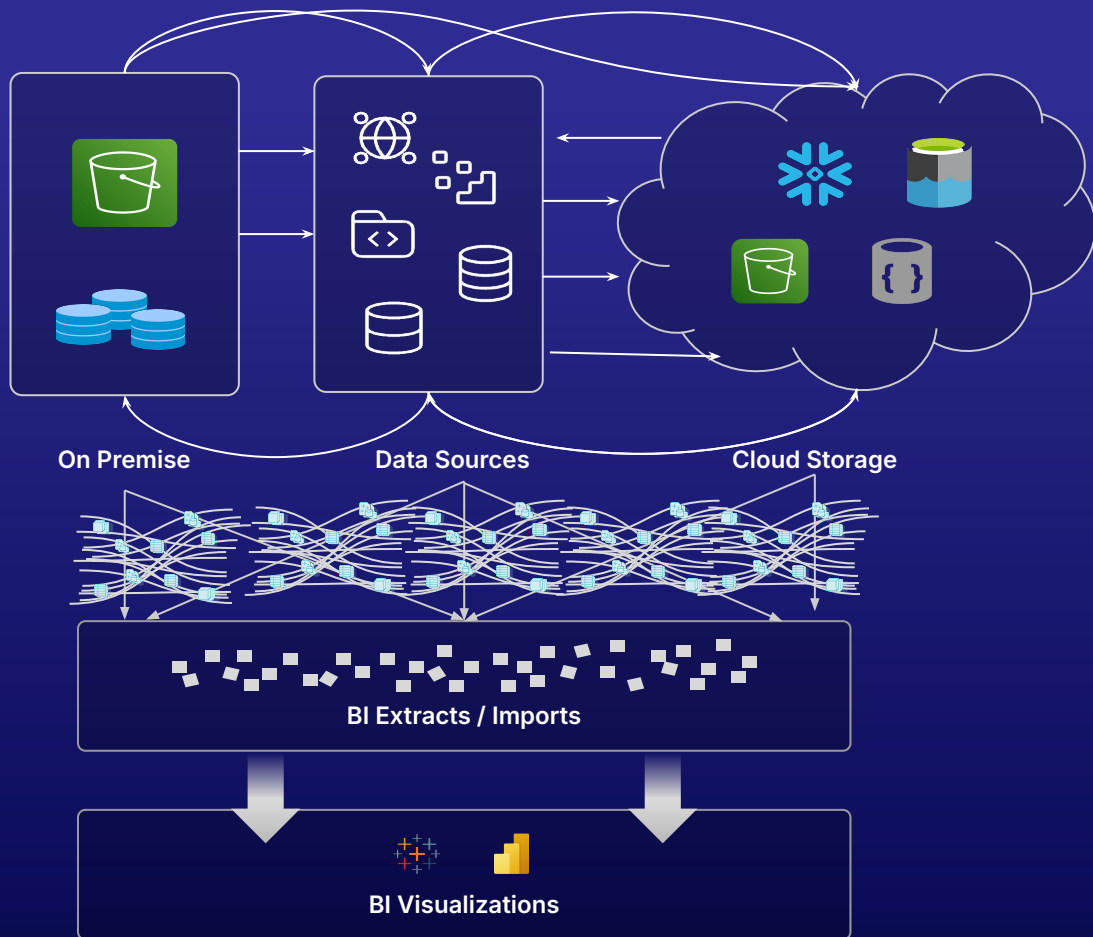
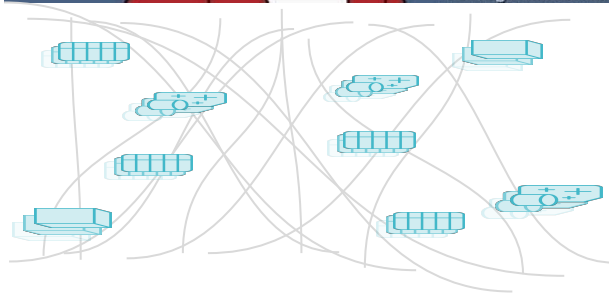




What's the Problem?

The Origin Story

Data Chaos



Technical & Business Bottlenecks



Centralized Data Responsibility

Centralized data teams become bottlenecks when managing multiple data sources and complex data consumption use cases.

Data Silos and Friction

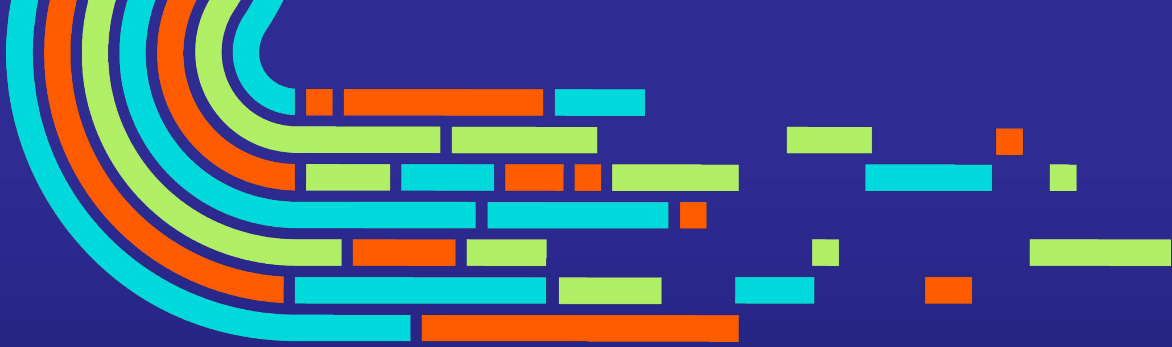
Traditional data architectures lead to silos, misunderstandings, and friction between data producers and consumers.

Governance and Quality Assurance

Ensuring data quality, metadata documentation, and compliance across the organization is challenging with data sprawls and siloed processes.

Lack of Data Utilization

Data often remains underutilized, leading to missed opportunities for insights and delivering on shared OKRs.

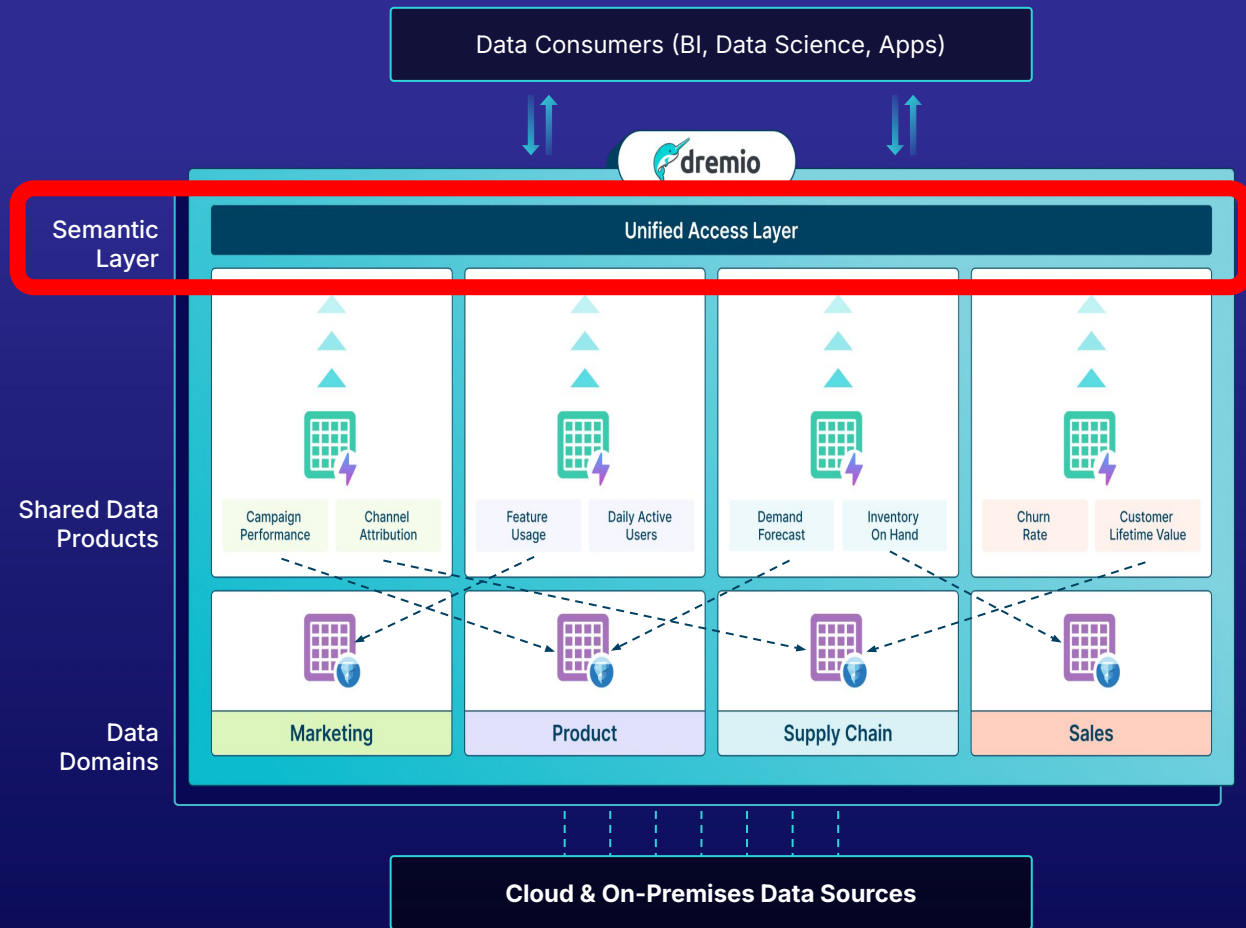


Demystifying Semantic Layers

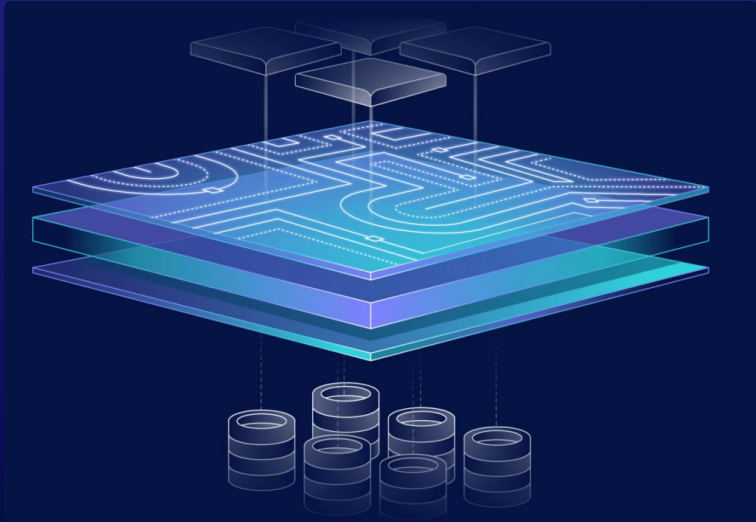
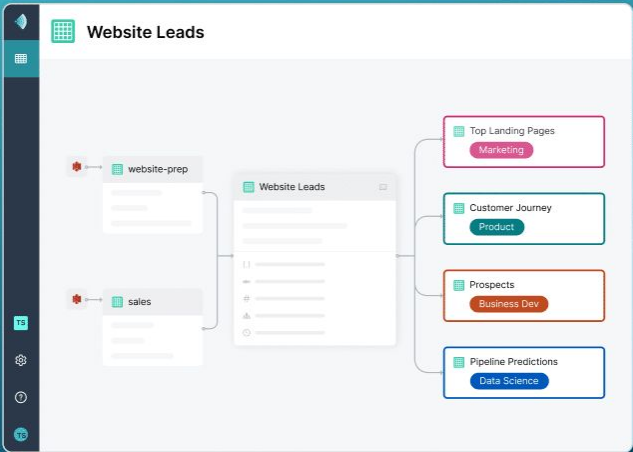
The Origin Story

What is a Semantic Layer?

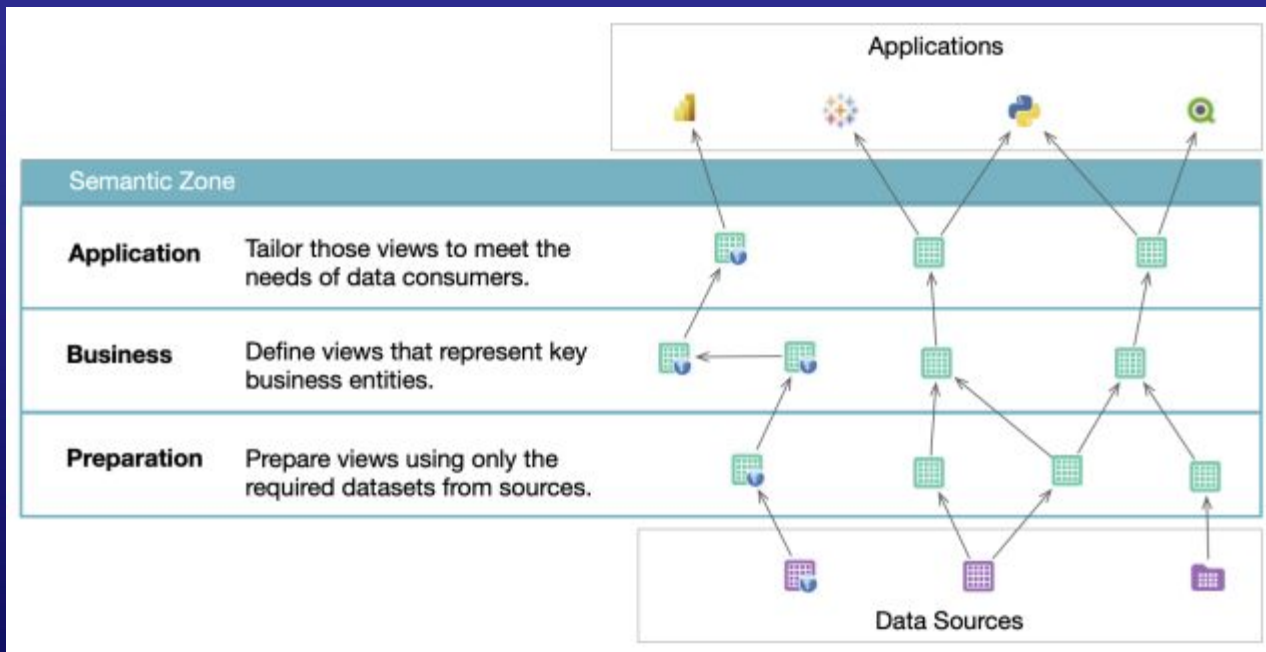
The Semantic Layer serves as an abstraction layer that translates complex, technical data into user-friendly, business-oriented terms.



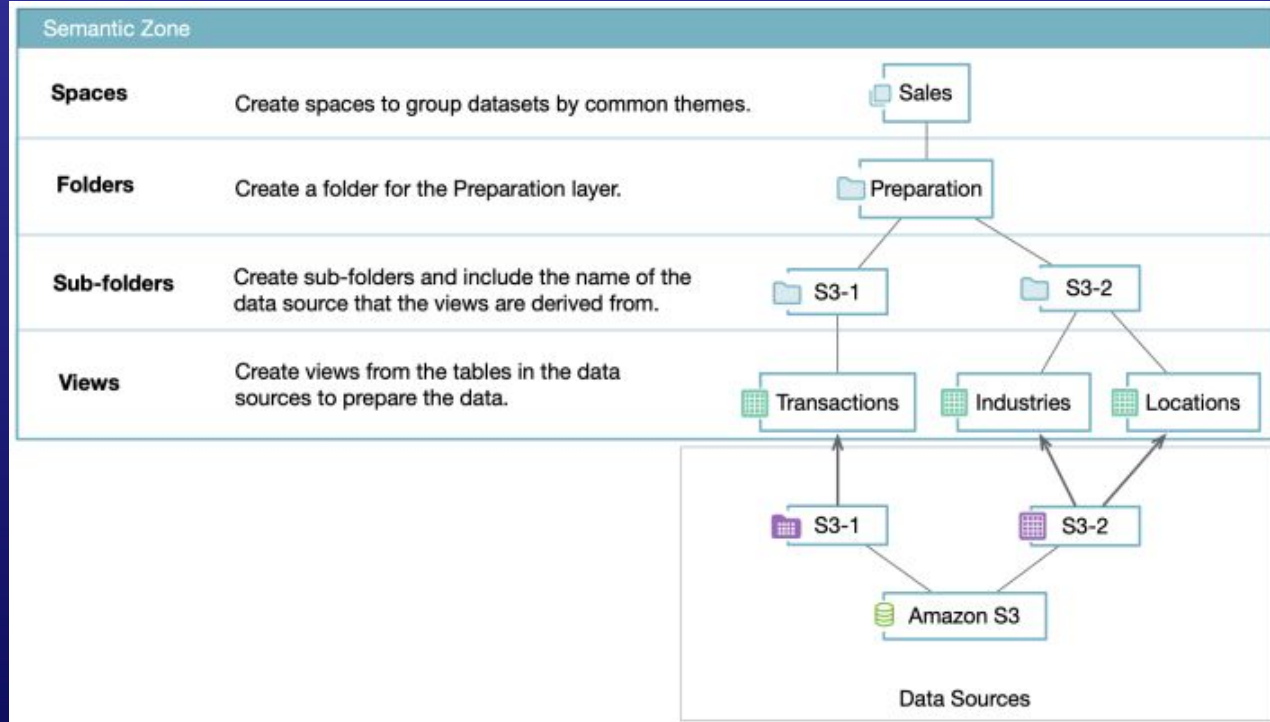
Benefits of a Semantic Layer



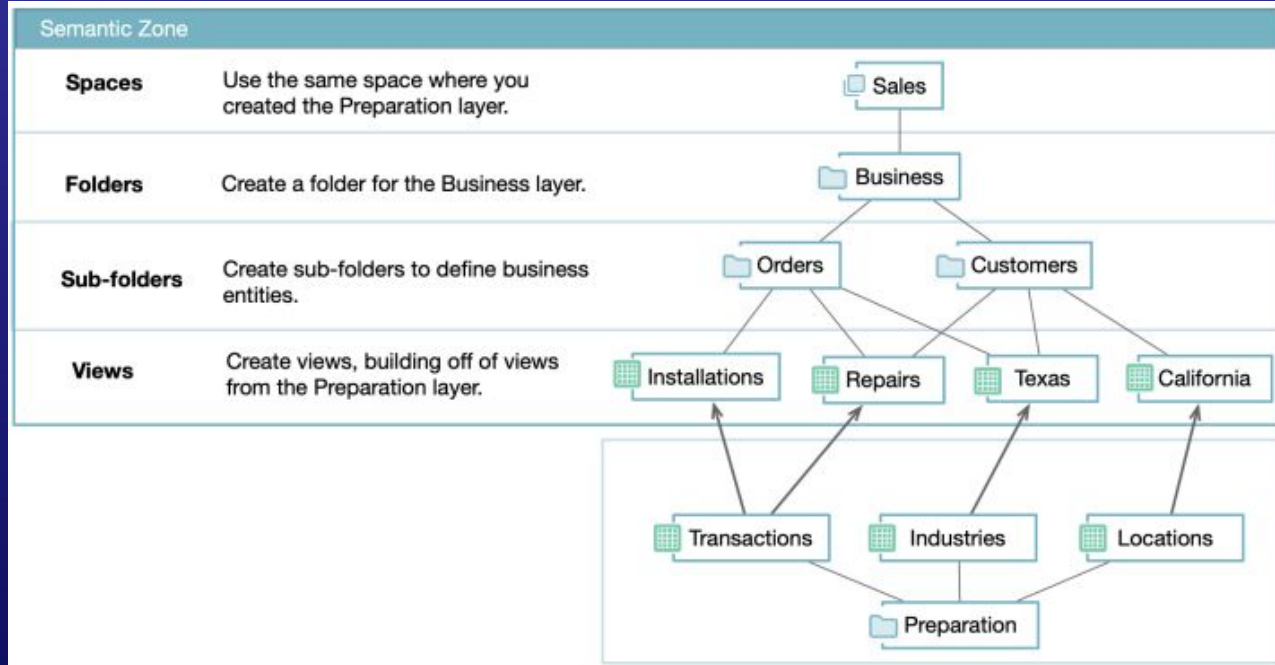
What Are the Components of a Semantic Layer?



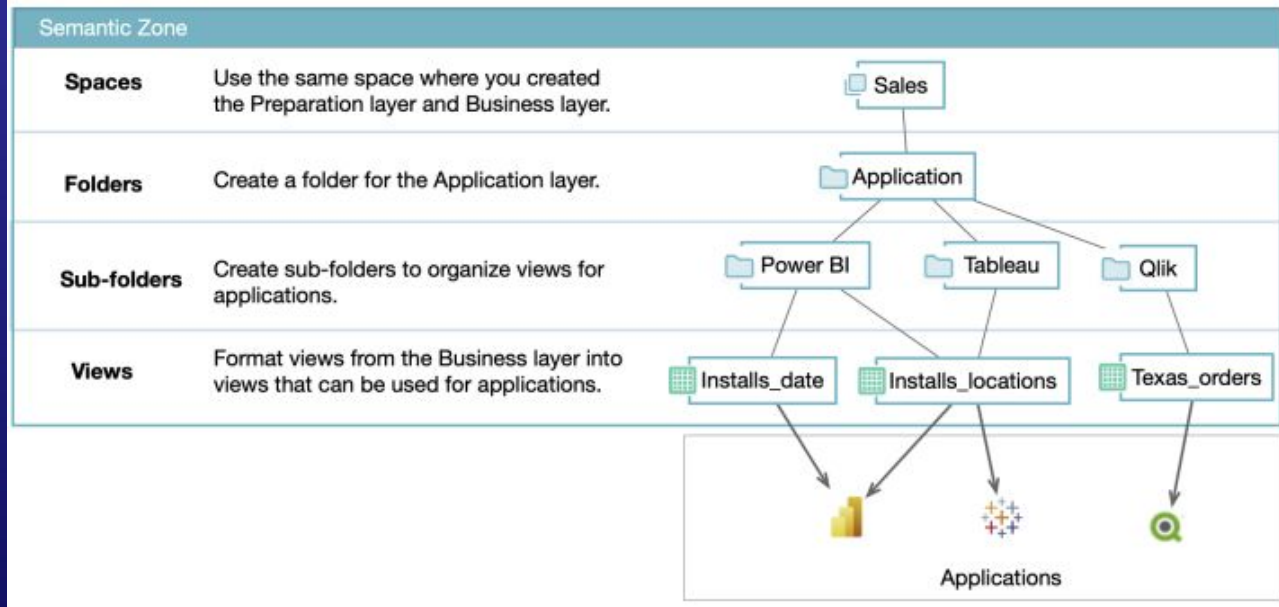
Best Practices - Preparation Layer



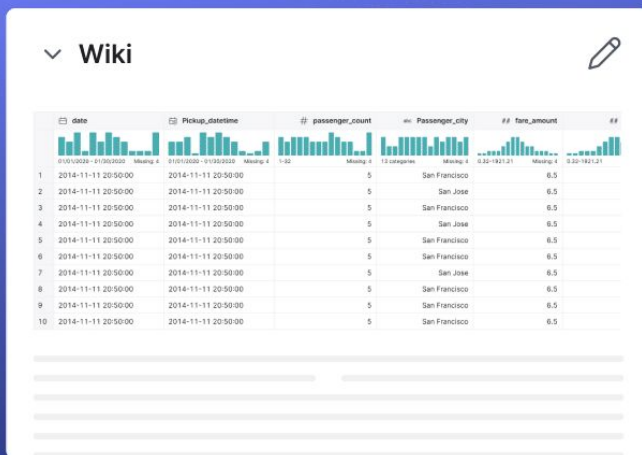
Best Practices - Business Layer



Best Practices - Application Layer



Best Practices - Wiki and Tags



NYC Taxi Trips

Edit wiki

Generated wiki (Preview)

Volume

1 Billion records stored in the data lake

Data Quality

This dataset has been validated to meet the gold data quality standards by our data quality team

Sample Query - get the number of the countries

```
SELECT
  country.country_name_eng,
  SUM(CASE WHEN call_id IS NOT NULL THEN 1 ELSE 0 END) AS calls,
  AVG(CAST(SPLIT(SPLIT(SOURCE, call_start_time, call_end_time), #)) AS avg_difference)
FROM country
ORDER BY calls DESC, country_id ASC;
```

Description

If you have questions about this dataset, please contact Tomer (tomerd@dremio.com).

Volume

1 Billion

Details

This dataset powers a number of dashboards, for example the credit v. cash analysis report.

Reports

ETL Status

Dataset updated at 18:30 by ETL Process XYZ

Cancel Save

Use Case Analytics



Unify All Data

Dremio allows seamless unified data access with virtualization, federation, and interactive analytics across all your data, both inside Snowflake and out.

Optimize Analytics

Maximize cost efficiency and enhance performance for BI and analytics across Snowflake data and beyond.

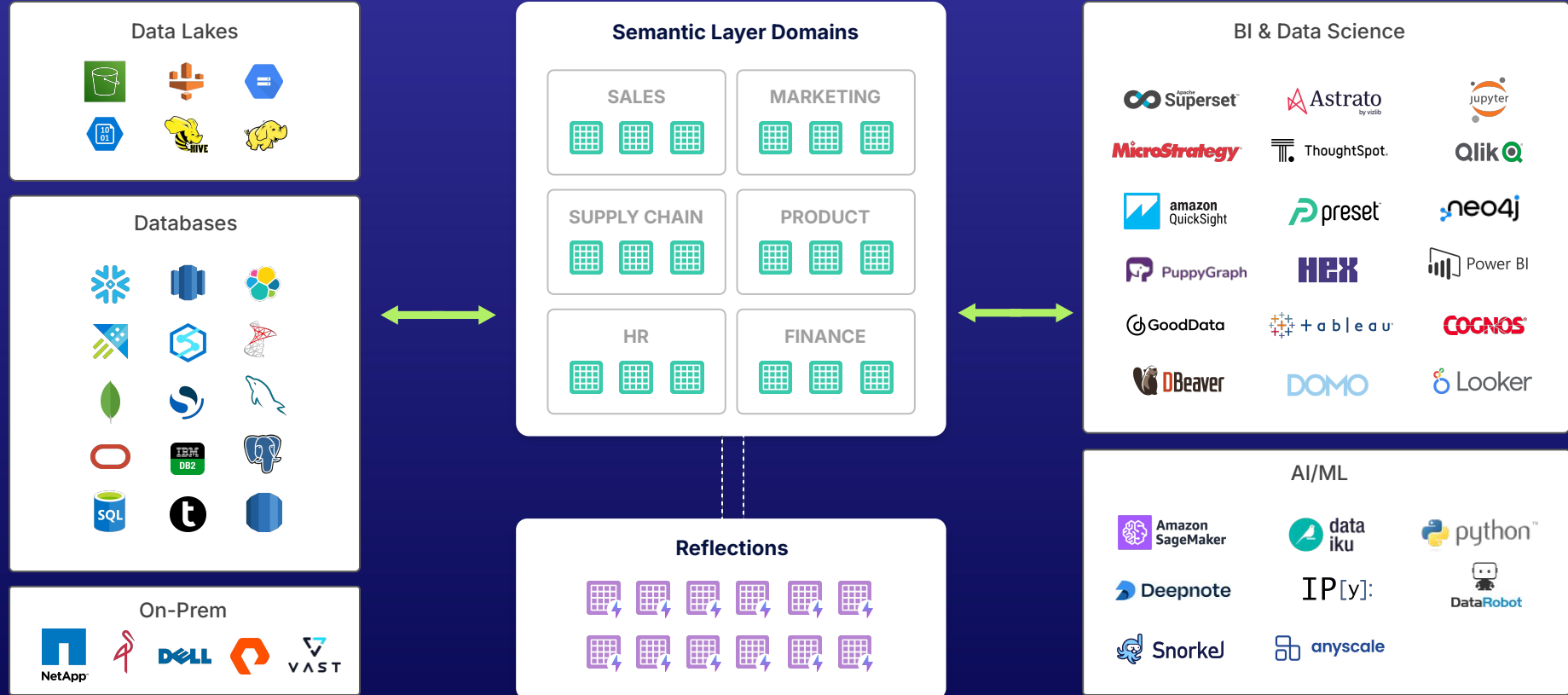
Self-Service Analytics

Dremio's universal Semantic Layer makes data discovery and analysis easy and intuitive. Lightning-fast, intuitive, self-service data access across all your data, source for faster time-to-insight.

Apache Iceberg Native Compatibility

Dremio is Apache Iceberg native, so data is always yours, always open, and always compatible with other analytics tools, engines and catalogs in the Iceberg ecosystem.

Use Case - Machine Learning

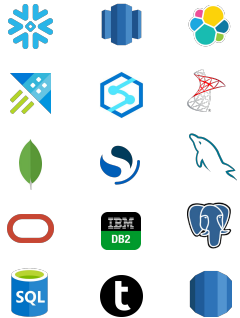


Use Case Machine

Data Lakes



Databases



On-Prem



Semantic Layer Domains

SALES



MARKETING



SUPPLY CHAIN



PRODUCT



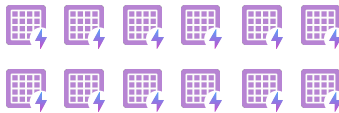
HR



FINANCE



Reflections



BI & Data Science

Superset

Astrato



MicroStrategy

ThoughtSpot

Qlik

amazon QuickSight

preset

neo4j

PuppyGraph

HEX

Power BI

GoodData

tableau

COGNOS

DBBeaver

DOMO

Looker

AI/ML

Amazon SageMaker

dataiku

python

Deepnote

IP[y]:

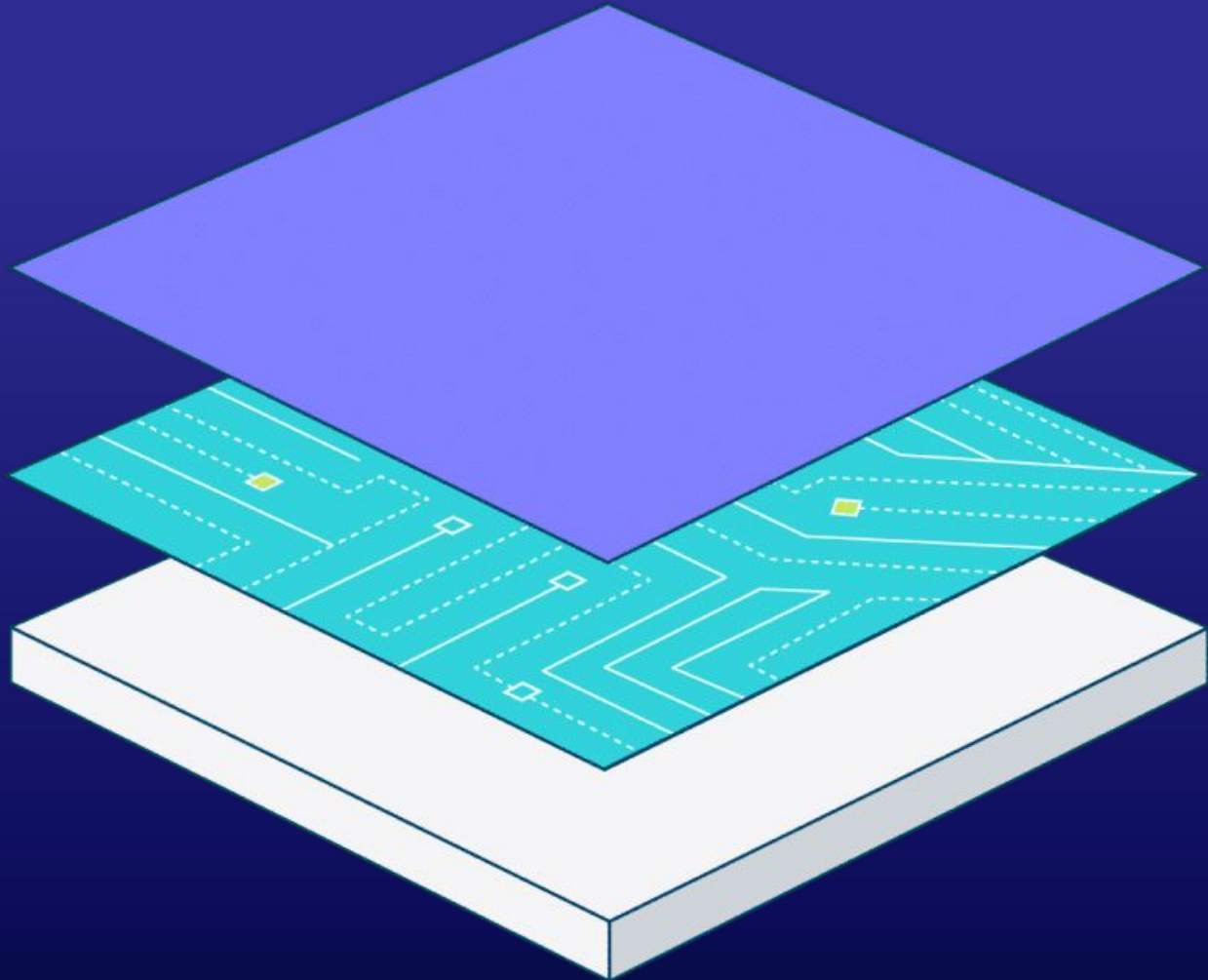
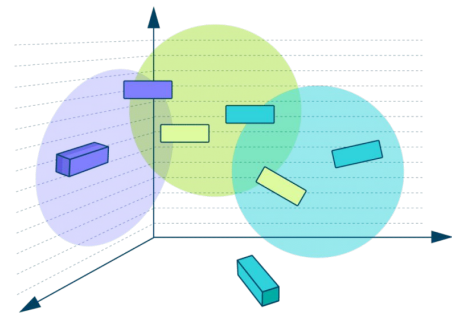
DataRobot

Snorkel

anyscale

Use Case

AI



Questions?



GNARLY Data_Waves

PRESENTED BY  dremio



Thank you!