

GNARLY
Data_Waves

PRESENTED BY  **dremio**

EPIISODE 39

How To Build an Iceberg Data Lakehouse with Fivetran and Dremio



Let's Connect!



Kelly Kohlleffel

Head of Partner Sales Engineering



Rommel Garcia

Principal Partner Solutions Architect



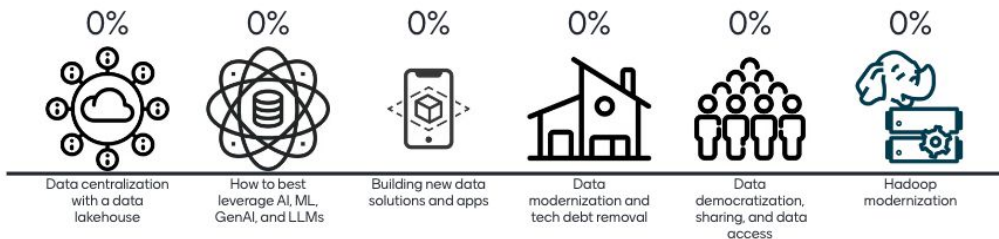
Agenda

- Fivetran & Dremio Overview
- Data Engineering and Analytics Challenges
- The Fivetran & Dremio Solution
- Integration Architecture of Fivetran & Dremio
- Data Lakehouse Demo
- Why Fivetran & Dremio
- Try Fivetran & Dremio



Perspectives

Which data workloads are you most focused on right now?



GO TO menti.com

ENTER THE CODE **7106 5172**

👤 1



AUTOMATED DATA MOVEMENT PLATFORM

Fivetran is trusted
by over **8,000**
companies to power
their data movement.



320+

Engineers building and maintaining your pipelines every day.

99.9%

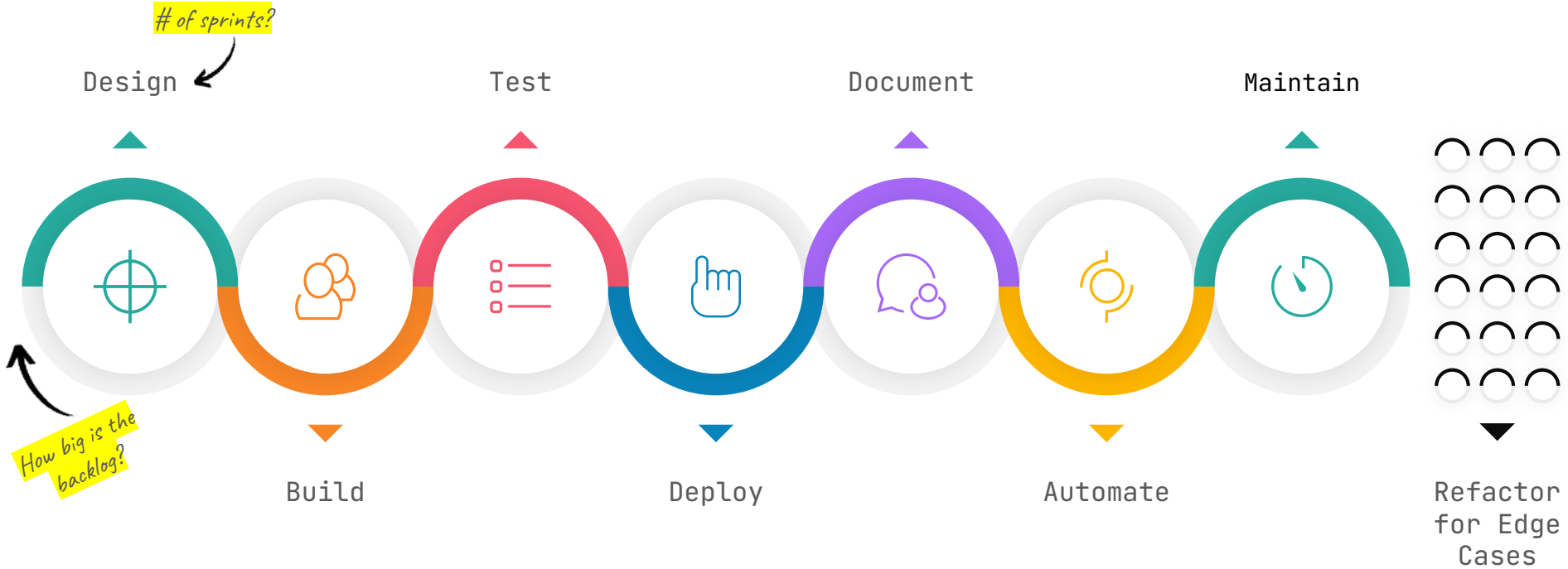
Platform uptime providing you the most reliable data pipelines.

2192.8 TB

Amount of data moving monthly.

A glimpse into data movement today (for one pipeline)

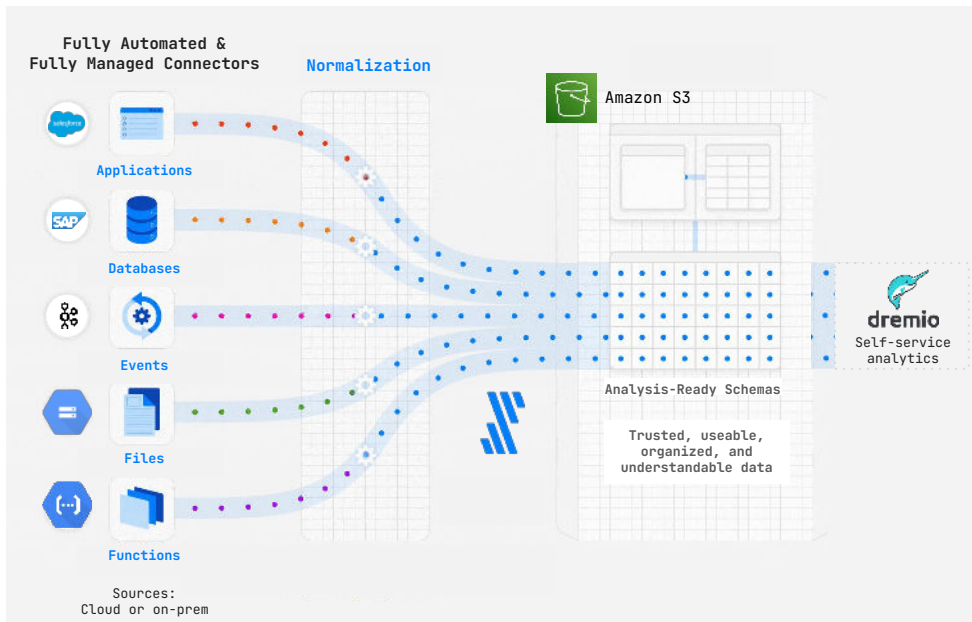
aka constant negative engineering



Fivetran is the standard for data movement.

Automation, reliability, and scale

- No-code, self-serve, connect in minutes
- Simplified & standardized architecture
- Fully automated and fully managed
- ELT approach is the way
- 450+ connectors (multiple categories)
- **Multiple deployment options from SaaS to Private Deployment**
- Log-based, Log-Free, and Agent-based CDC data replication for RDBMS
- Automated CDC for all connector types
- Centralized global security & governance
- Reduced complexity = reduced costs, duplication and errors



About Dremio

Enterprise, Open Data Lakehouse

Self-service analytics with data warehouse functionality and data lake flexibility across all of your data.

- **Enterprise Grade Lakehouse** with Self-Service Analytics
- Your Data Forever through **Open Formats**, including **No Lock In**
- Sub-Second Performance at **1/3 the Cost** of Data Warehouses or Proprietary Data Lakehouses

Open Source & Community

Apache Arrow
(70M+ downloads/m),
Apache Iceberg, Nessie

Creator and host of
Subsurface LIVE conference

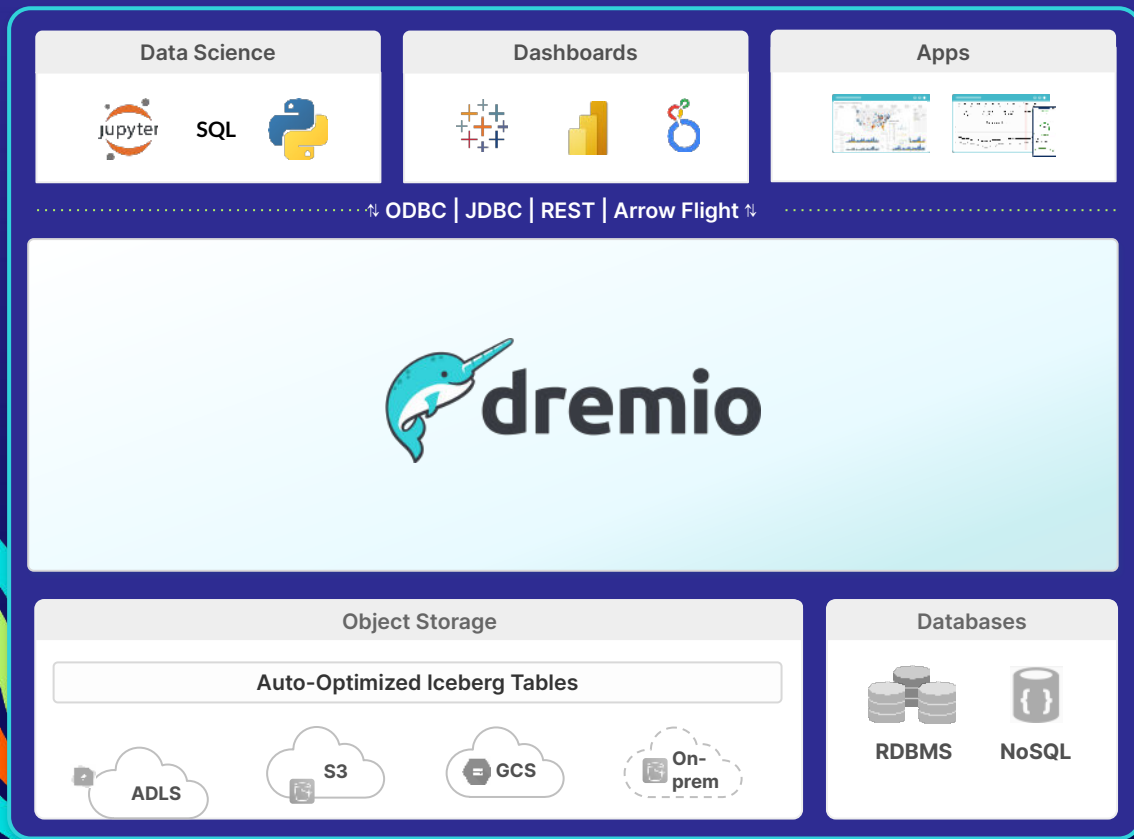
Enterprise Adoption

1000s
of companies
across all industries

5
of the Fortune 10

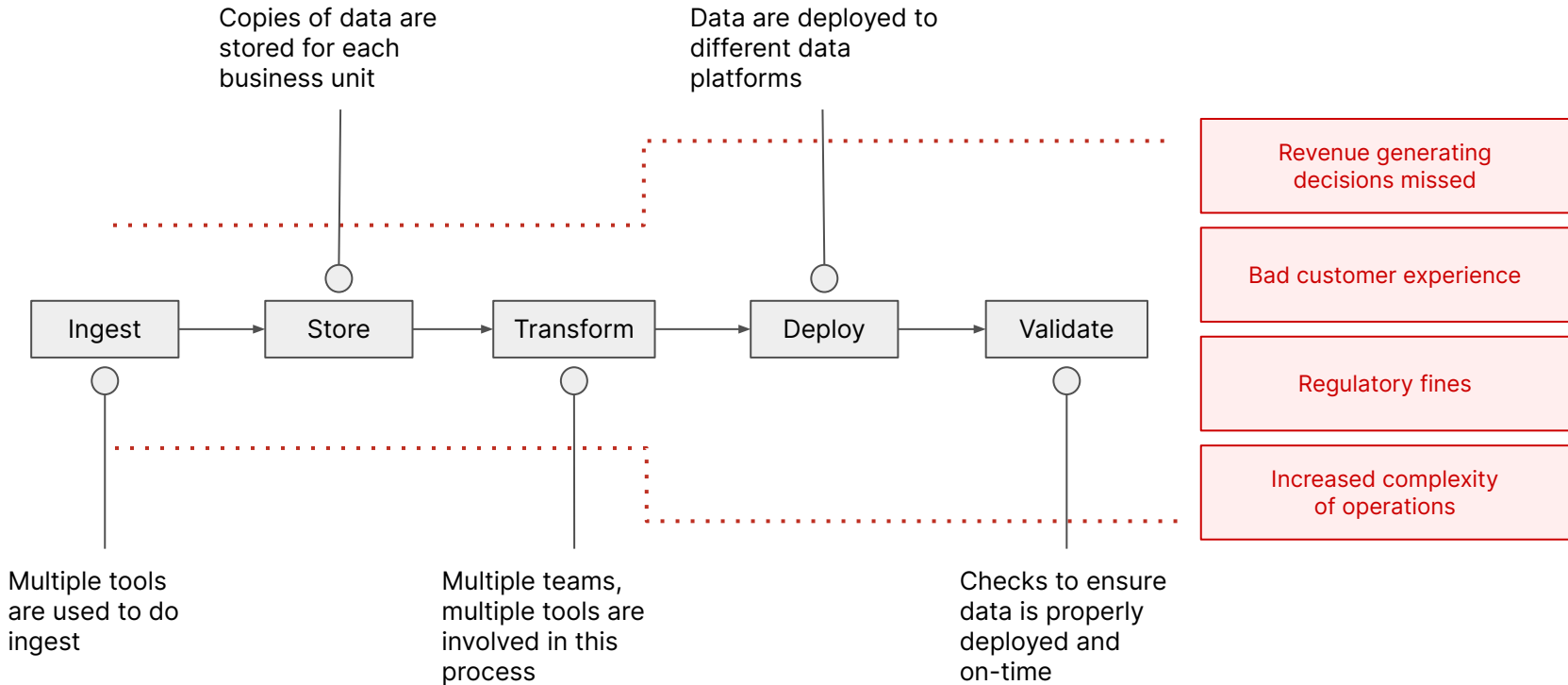


Dremio Data Lakehouse: Easy, Open, 1/10th the Cost



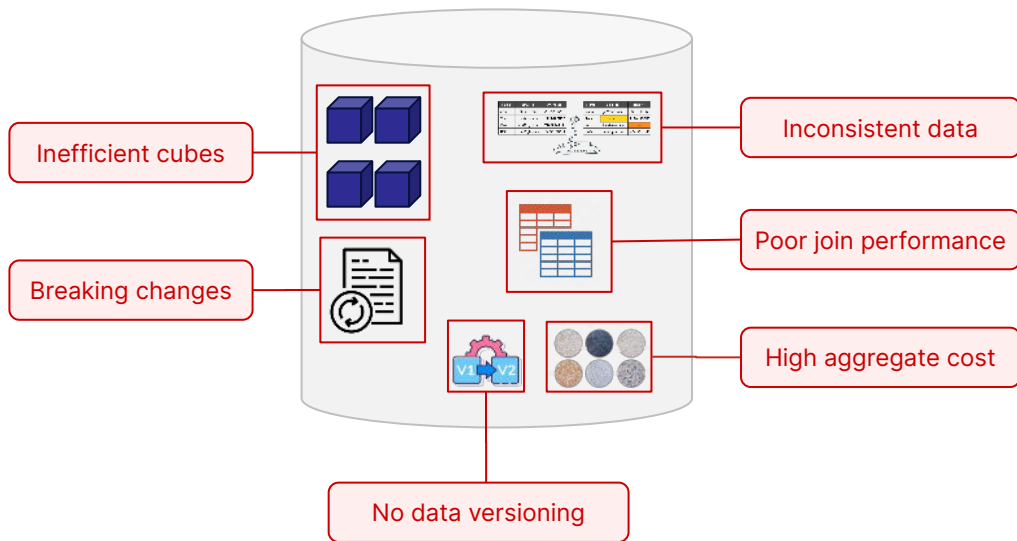
- ✓ No data copies
- ✓ No complex ETL pipelines
- ✓ Inexpensive
- ✓ Self-service analytics
- ✓ Data engineering productivity
- ✓ Security
- ✓ Governance
- ✓ Consistent data

Data Engineering Challenges



Data Analytics Challenges

Data Warehouse Issues



Data Consumption Issues

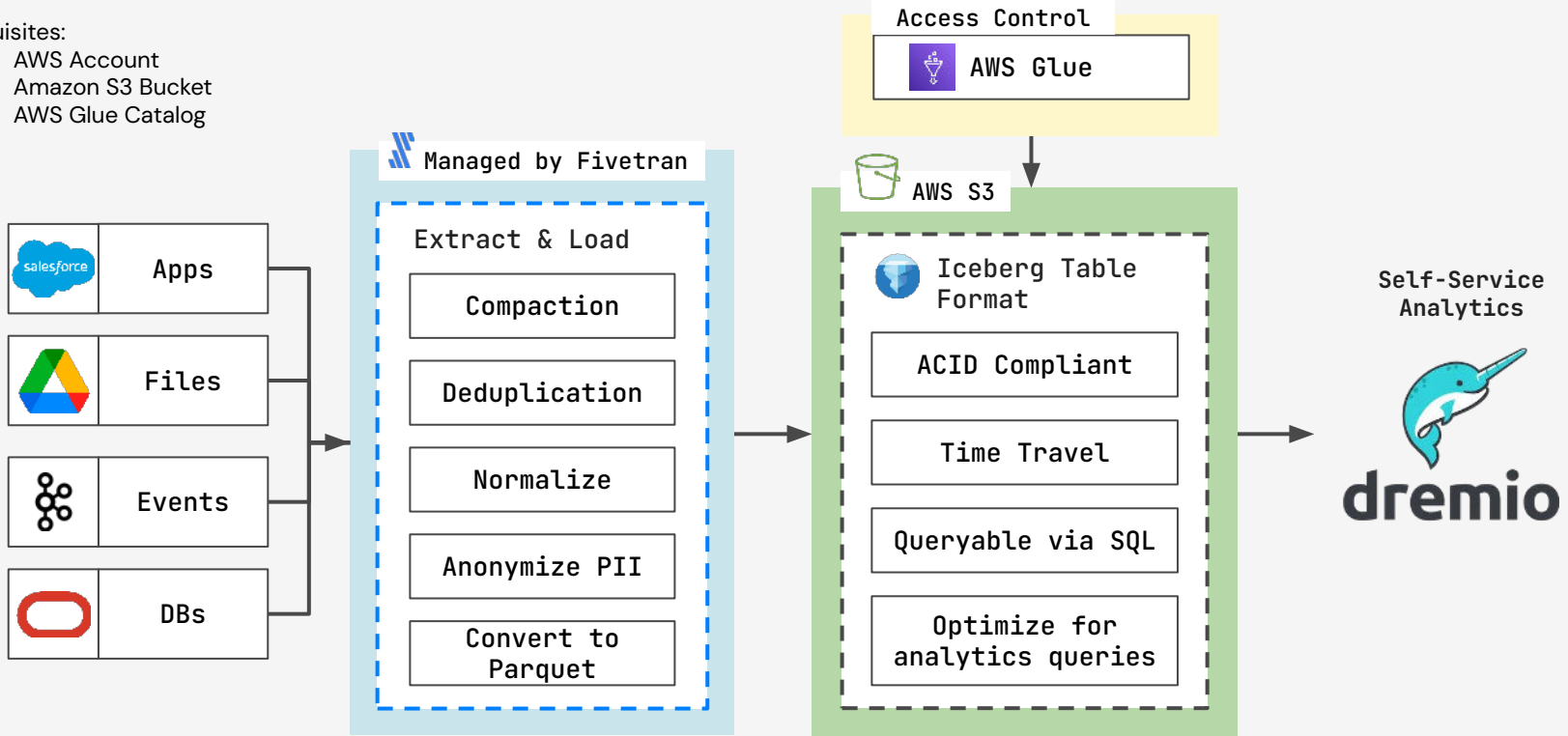


- Slow Queries
- Scaling issues
- Low Query Portability
- Concurrency Issues

Fivetran & Dremio Architecture

Prerequisites:

- AWS Account
- Amazon S3 Bucket
- AWS Glue Catalog



Demo

- Fivetran & Dremio Integration Setup
- Extract & Load data using Fivetran from Workday, Oracle & Postgres to AWS S3
- Parquet to Iceberg Table Format
- Querying Iceberg Table
- Applying Reflections

Why You Should Adopt Apache Iceberg?



Transaction Consistency

- Consistent across multiple applications
- Full read isolation
- Multiple concurrent writes

Schema Evolution

- Unique IDs to track each column in a table
- Does not require rewrites or migration to a new table
- Metadata changes

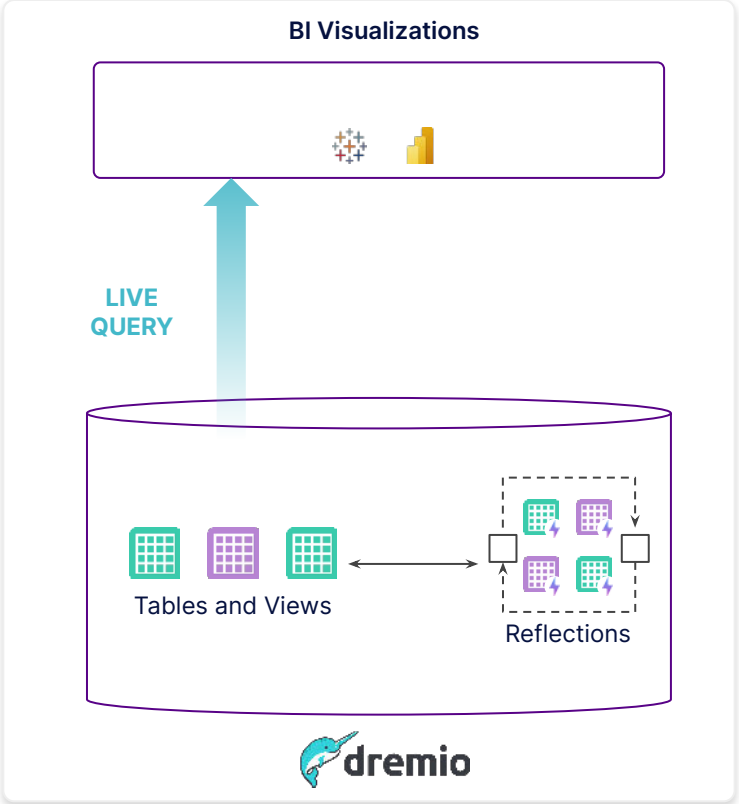
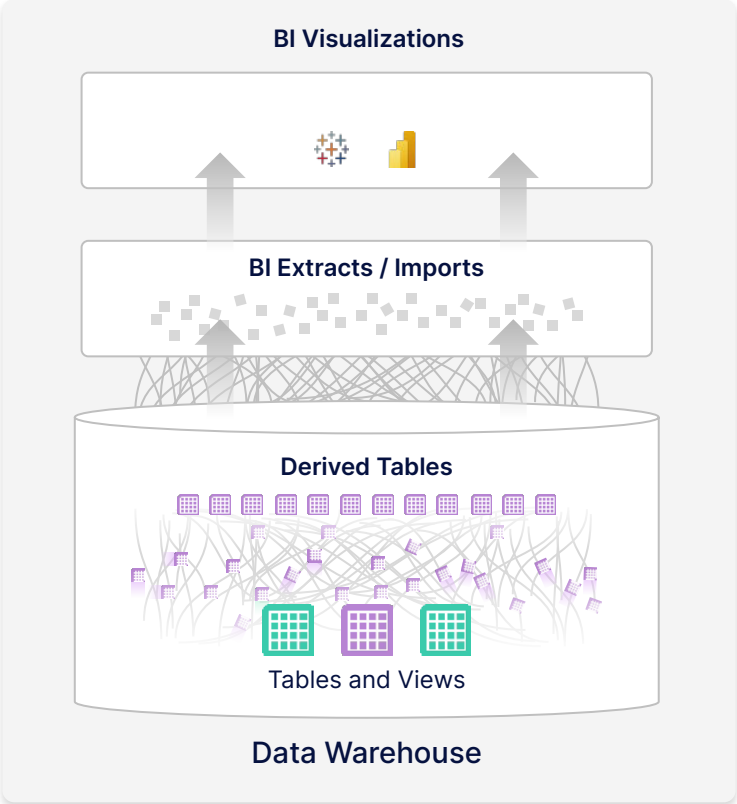
Time Travel

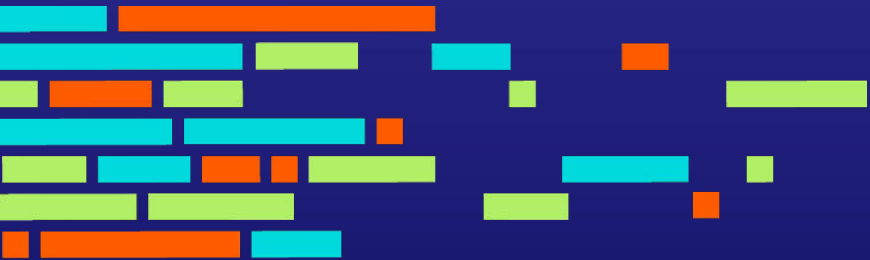
- View data as it existed at any point in time
- Leverages snapshots created via update processes
- Query historical data ; verify changes between updates

Partition Evolution

- Updates partition scheme for new data
- Avoids rewriting all previous data
- Split planning

Reflections Eliminate the Need for BI Extracts/Imports





Why Use Fivetran and Dremio Together?

Go Self-Service!



Data Democratization

- Enable non-technical business users to explore data independently
- Breaks down existing data access barriers and allows business users to obtain insights in their moment of need



Data Driven Decisions

- Drill into the source data to determine root causes and actionable insights, all without experiencing data access limitations or using IT resources.
- Obtain customized data insights to drive critical decisions



Data Engineer Efficiency

- Data Engineers focus on more strategic initiatives
- Supports scale of data growth and requests

Try out Fivetran and Dremio Today!

fivetran.com/signup

dremio.com/get-started/

community.dremio.com

community.fivetran.com



LIVE

Subsurface

The Data Lakehouse Conference

PRESENTED BY  **dremio**

7,000+ Global Data Experts	46+ Breakout Sessions	3,000+ Companies
--------------------------------------	---------------------------------	----------------------------

MAY 2-3, 2024 LIVE IN NEW YORK CITY + VIRTUAL [Get Notified](#)



DEVintersection
Conference
POWERED BY
Microsoft & NextGen

Microsoft Azure
+ AI Conference
POWERED BY
Microsoft & TechIntersection

Dec.
5-7

Walt Disney World Swan and Dolphin Resort, Orlando, FL
Workshops: Dec. 3rd, 4th & 8th

[REGISTER](#) [BOOK HOTEL](#)



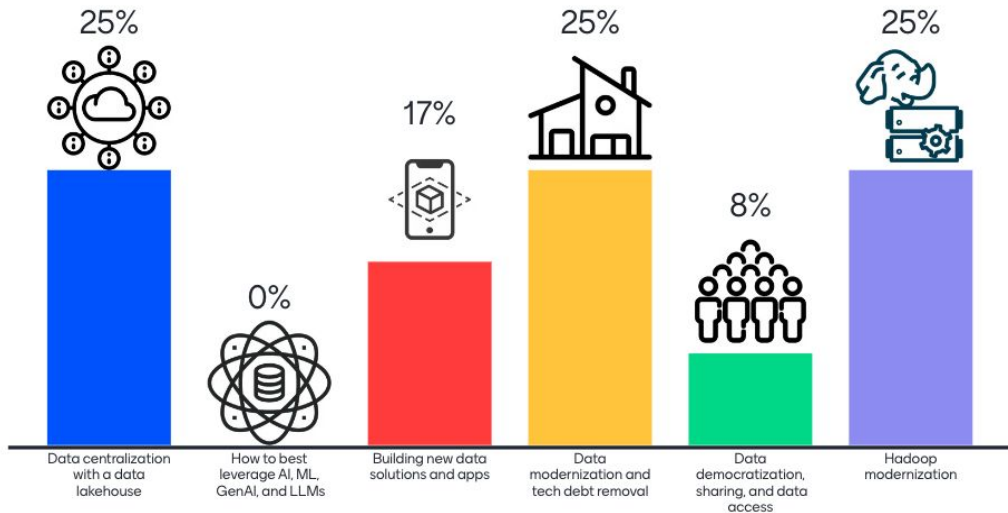
AWS re:Invent

NOV. 27 - DEC. 1, 2023 | LAS VEGAS, NV



Poll Results

Which data workloads are you most focused on right now?



GO TO [menti.com](https://www.menti.com)

ENTER THE CODE **7106 5172**