Build your open data lakehouse with Fivetran & Dremio

Brett Roberts | Principal Partner Solutions Architect | Dremio
Coral Trivedi | Destination Product Manager | Fivetran
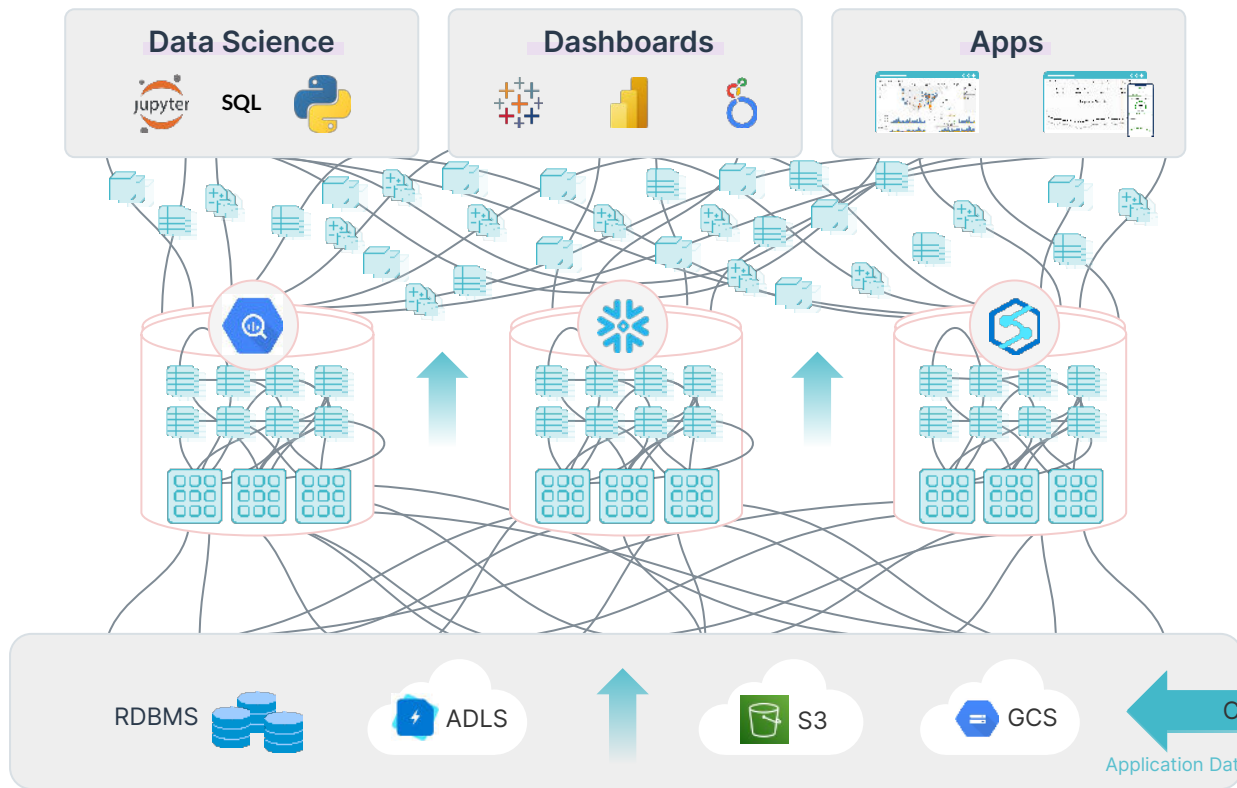
What is a Data Lakehouse?

Data Warehouse

Data Lake

Fivetran dremio

Two-Tiered Architectures: Expensive, Proprietary, Complex

Data Science
jupyter SQL python

Dashboards

Apps

× Skyrocketing costs

× Vendor lock-in

× Exploding requests backlog

× Can't explore data

× Not self-service

RDBMS ADLS S3 GCS Continuous New Data

Application Databases | IoT | Web | Logs

Fivetran dremio

Data Warehouse

+

Data Lake

Transactional Consistency
Files can be added, removed, or modified atomically, with full read isolation

Flexible
Object or file-based storage architectures based give users greater flexibility when storing, using, and managing data

Data Quality
Track table changes over time & roll back to prior versions of your data to quickly correct issues

Scalable
Highly parallelized & decoupled storage, allows data growth as volume, variety, and workload needs evolve over time

Reliability
Provides most reliable way for many concurrent business users to access data for BI & reporting

Cost Effective
Built on scalable, low-cost commodity servers or leverage cloud-based object storage

ICEBERG

Data Lakehouse

Fivetran dremio

Key Components of a Lakehouse

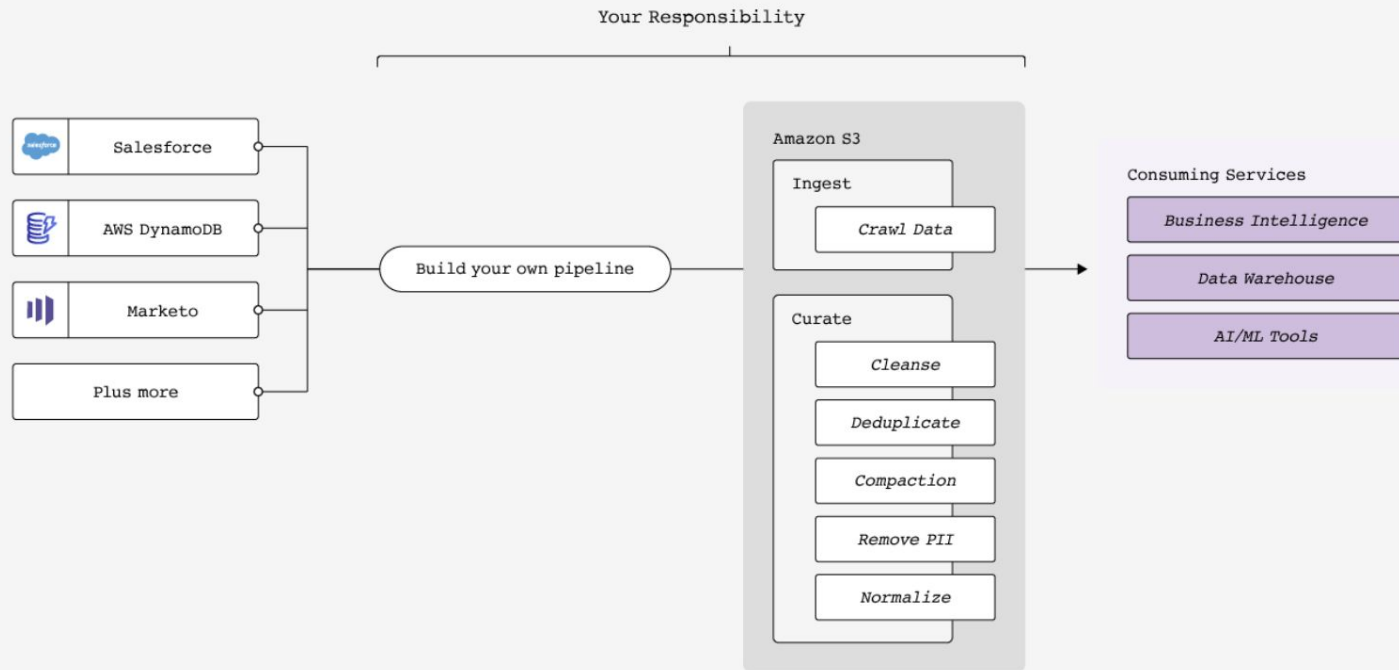|		
Data Pipelines	Reliable Automated Connects to a variety of sources	
Data Lake Storage	Highly scalable Stores data in open file and table formats Always Available	
Transactional Tables	ACID transactions on the data lake Time travel/versioning Easy schema/partition evolution	
Query Engine	Query data in place Warehouse performance on the data lake Self-Service Semantic Layer	

Fivetran dremio

The Current State of Data Lake Management

Why 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
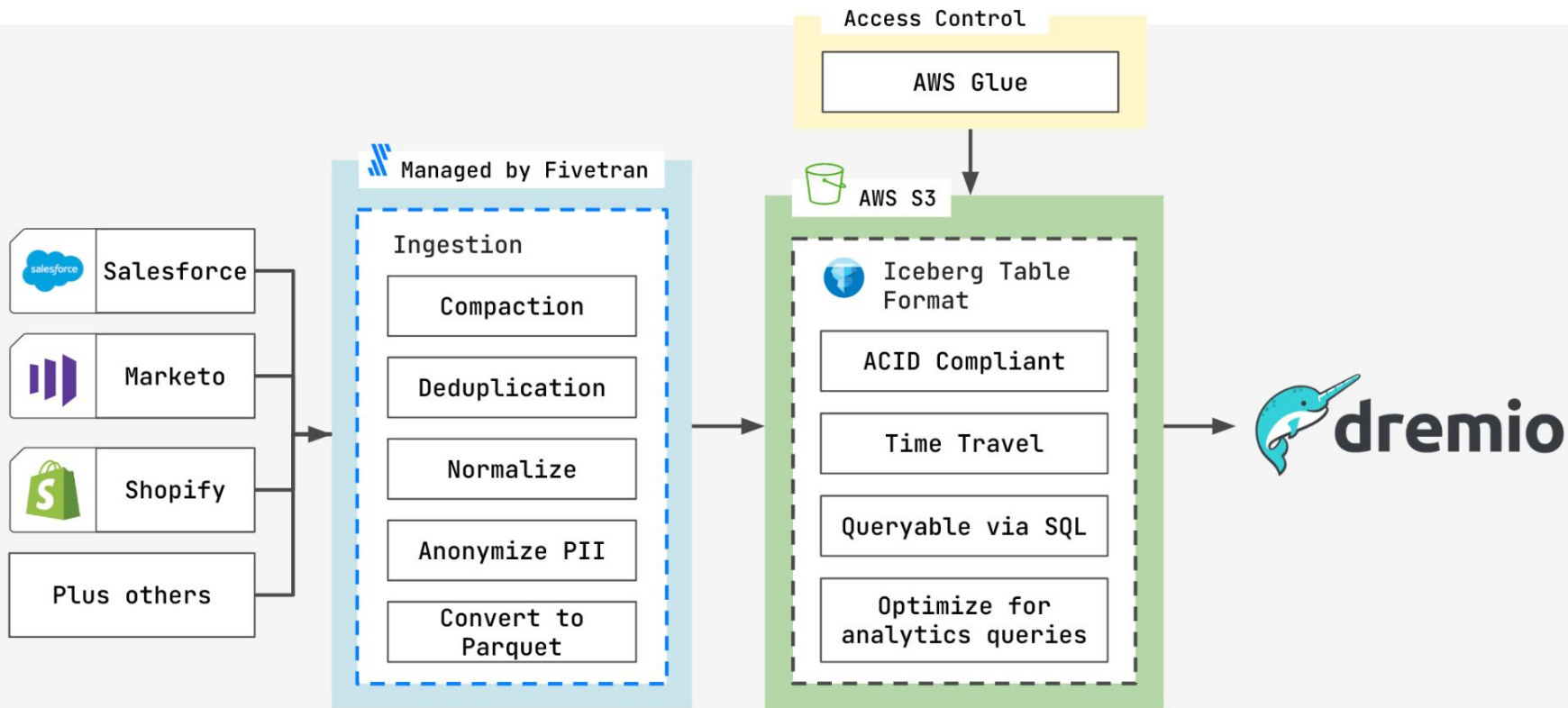
Avoids rewriting all previous data

Split planning

Fivetran dremio

Fivetran's S3 Destination

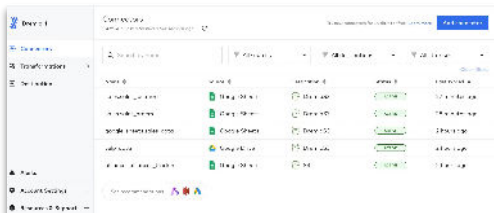|	Compliance	Security	Ease of Use	Reliability
	Hash or block your data before it enters the pipeline.	Fully encrypt the data as it flows through the pipeline to ensure security from start to finish.	With 200+ fully-managed data pipelines that take minutes to set up, your data lake just got a whole lot easier to use.	Automated compaction for better performance.
	With AWS Glue Data Catalog providing metadata to Iceberg, your data lake can now handle compliance with ease.	AWS Lake Formation provides granular access control with AWS Glue.	Iceberg's time travel feature ensures you don't have to worry about managing versions.	Iceberg's ACID compliance feature ensures data correctness and consistency

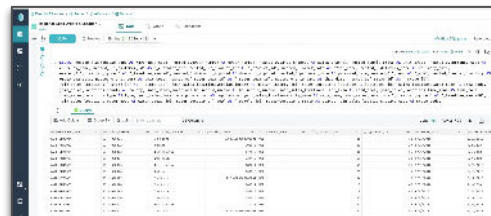# Fivetran & Dremio Architecture

Demo

Fivetran dremio

How can you get started?

Try Fivetran for free!

Create connections to your AWS data lakehouse in minutes!

Use Dremio for free!

Connect Dremio to AWS Glue to build your data lakehouse

Dremio Webinars

Thanks!

Brett Roberts | brett.roberts@dremio.com
Coral Trivedi | coral.trivedi@fivetran.com