MINIO+ dremio

Modern Data Infrastructure for Customers

The Modern Data Lake/Data Lakehouse

Agenda

- What are the architectural challenges facing customers around data infrastructure?
- Evaluating the options to solve these challenges.
- The power and flexibility of a modern data lake approach.
- What a modern data lake architecture provides:
  - Cost reduction
  - Foundation for AI/ML
  - Deployment flexibility
  - Scalability
- Q&A

Speaker Introduction

Brenna Buuck
Sr. Technical Evangelist
brenna@minio.io
MINIO

Alex Merced
Sr. Technical Evangelist
alex.merced@dremio.com
dremio

Challenges for Customers

Legacy Architectures

- Expensive to maintain
- Poor performance
- Not cloud-native

Build a Scalable AI/ML Foundation

- Modern data platforms are built for AI/ML
- Very difficult, even impossible to retrofit legacy architectures
- AI is here now

Flexibility and Optionality

- Mandates for multi-cloud and hybrid-cloud
- Scalability at a lower cost
- Deploy anywhere

How can we solve these challenges?

Start with Hard Requirements

AI-Ready

Performant

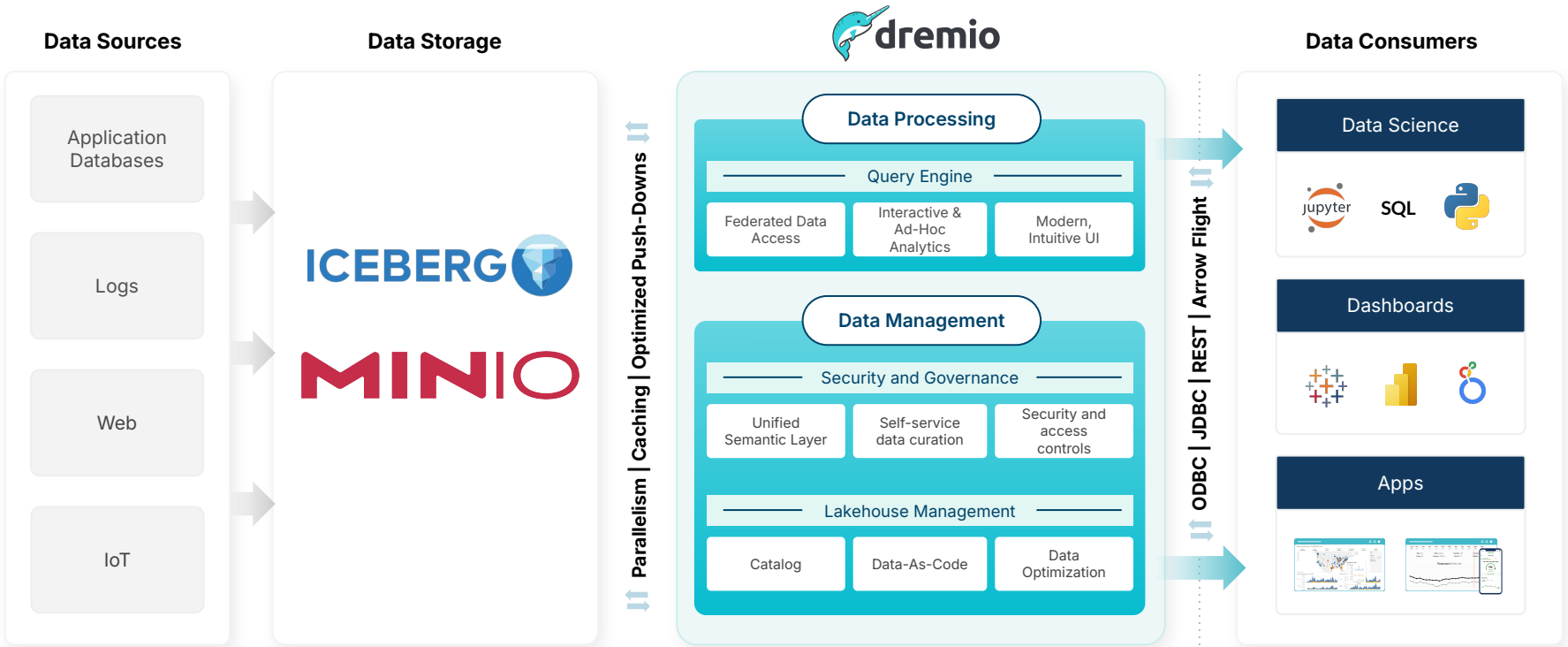
Scalable

Flexible (Data Types, Deployment Configs)

Cloud-Native

Secure

Dremio & MinIO: Next-Gen Data Lakehouse



Advantages of Data Lakehouses

	Data Warehouse	Data Lake	Data Lakehouse
Structured Data	✓	✗	✓
Unstructured Data	✗	✓	✓
ETL	Required	Not required	Optional
Schema support	Schema on write	Schema on read	Schema on write & Schema on read
Schema evolution	✓	Not applicable	✓
Scalability	Vertical	Horizontal	Horizontal
Storage	Relational Tables	Object Storage or File System	Object Storage
Time Travel	✓	✗	✓
Data Versioning	Not Supported	✗	✓

Modern Data Lake Components

The Components of a Modern Data Lake

Storage

Performant, secure object storage suitable for all types of application workloads that can handle structured and unstructured data.

Open Table Formats

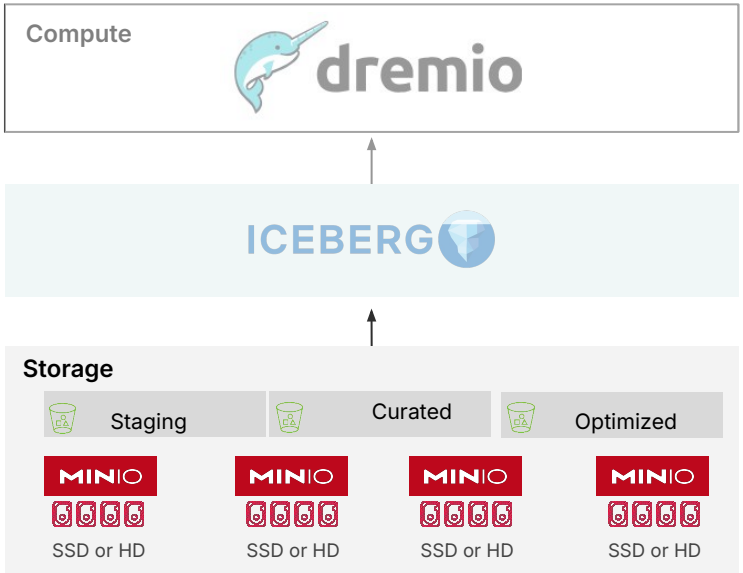
Provides a layer of abstraction on top of your data lake, allowing for a more traditional database like interaction, while also handling things like data version control for different applications.

Compute

A distributed analytics engine that offers an intuitive user interface for self-service data exploration, transformation, and collaboration.

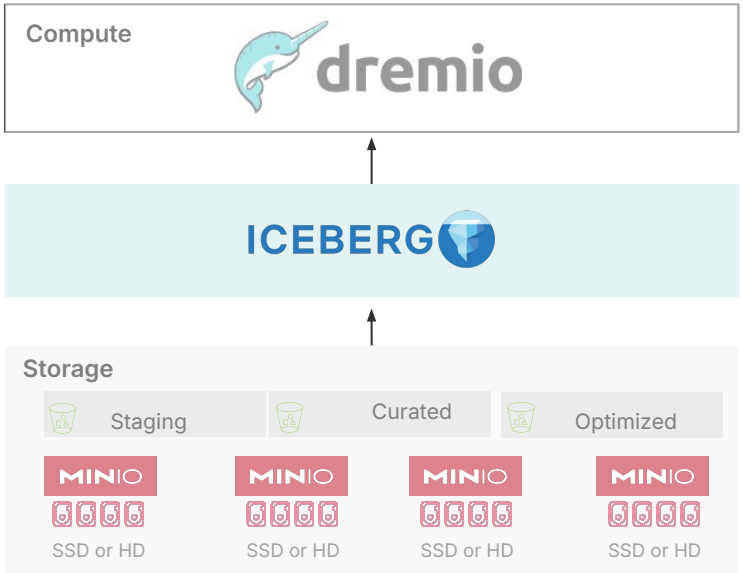
Primary Storage - MinIO Object Store

- High-performance, S3-compatible object store built for AI/ML, advanced analytics, databases, data lakes and HDFS replacement.
- Offers a rich suite of enterprise features targeting security, resiliency, data protection, observability and scalability.
- Exclusively software defined, truly cloud-native and remarkably simple to install, manage and scale.
- Deployable anywhere from public and private clouds, to colos and the edge.



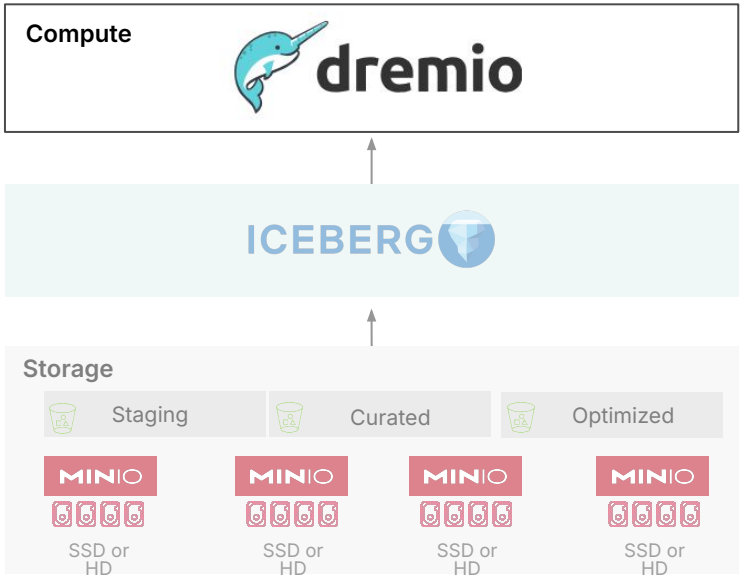
Open Table Format - Apache Iceberg

- Open standard built for the modern data lake.
- Supports Java, Python, Rust, and Go for broad ecosystem integration.
- Enables ACID transactions, time travel, schema/partition evolution and more
- Optimized for speed with hidden partitioning, efficient metadata, and advanced indexing for quick query performance, even on large datasets



Compute - Dremio

- Bring users closer to the data with Unified Self-Service Analytics
- Optimized SQL Query Engine price performance with acceleration for sub-second BI
- Centralized Data Governance enables faster access to data
- Lakehouse Management through an integrated enterprise iceberg catalog and ability to optimize and maintain Iceberg tables, so they just work.



MINIO + dremio

The Benefits of a Modern Data Lake

Massive Cost Savings over Legacy Hadoop

TransUnion

10x

performance gain with Dremio compute

NCR

30x

faster dashboard speeds

Leading Financial Group

60%+

cost to performance gain with MinIO over HDFS

Leading Financial Group

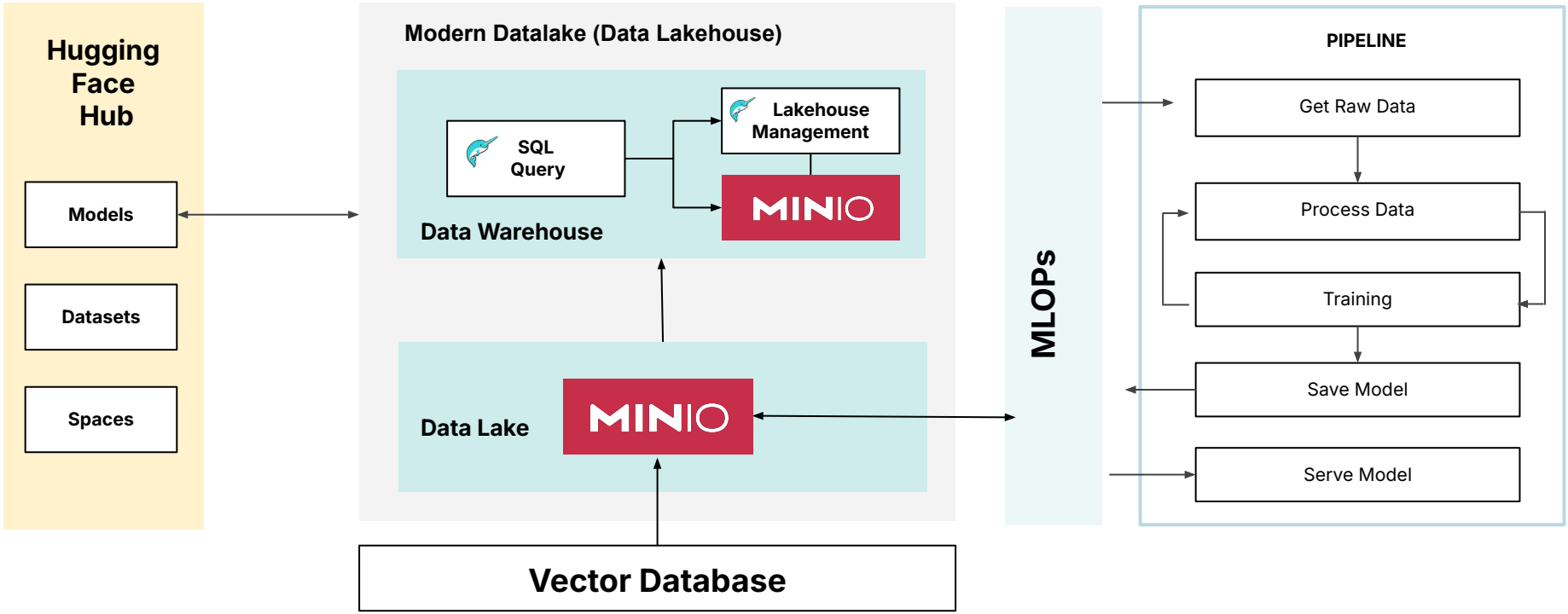
20%

immediate performance gain for MinIO vs HDFS w/ MinIO choking existing HW

> "Using MinIO and Dremio has allowed us to go from a pretty much defunct technology (Hadoop) to a forward-looking one that, importantly, actually supports a true hybrid - on-prem and cloud - architecture."
>
> **A Global Financial Company**

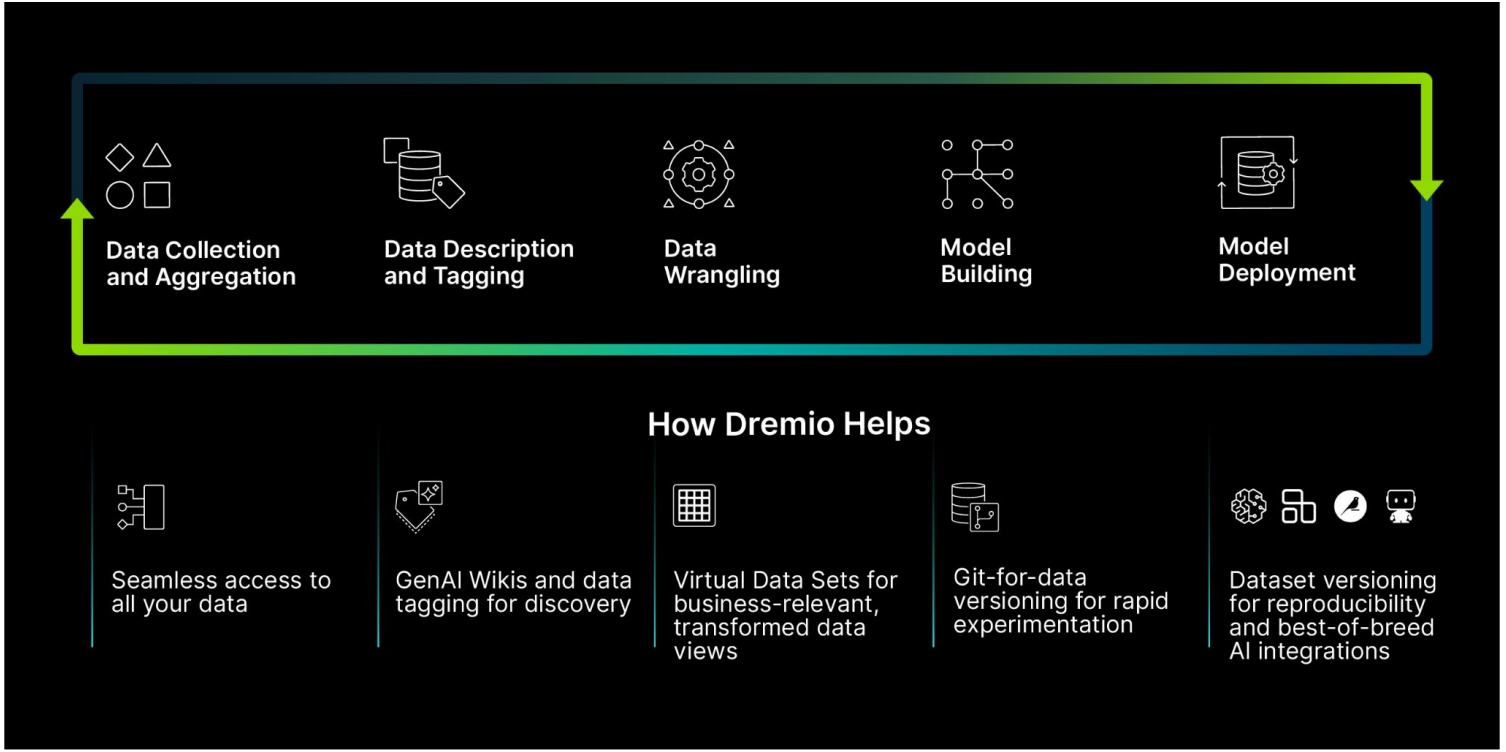
MINIO + dremio

The Foundation for AI/MLq



MINIO + dremio

Flexibility and Speed for AI/ML



Graphic from: www.dremio.com

Deployment Flexibility and Control

MINIO + dremio

Scale Intelligently

Multi-/Hybrid-Cloud

- MinIO can be deployed anywhere.
- Avoid cloud vendor lock-in by treating the cloud as an operating model.
- Dremio can connect to storage and data sources in the cloud and on-prem

Scaling where you need it

- Scale your compute and storage layers independently.
- Add additional storage capacity to MinIO simply by adding new server pools.
- Dremio auto-scaling features help keep your compute needs managed.

Minimize Costs

- Ease of administration and management allows for less administration overhead.
- Extend and accelerate compute performance on-demand with Dremio Reflections.
- Advanced erasure coding helps protect data integrity with less infrastructure.

Putting This Into Play

Stage 1	Stage 2	Stage 3
Modernize Data Lake Query Engine with Dremio & Provide Self-Service Analytics	Migrate HDFS and other legacy storage to MinIO object storage	You Have Your Modern Data Lake!

Scan for Lakehouse Migration Resources

Summary

Modern Data Lake

A modern data lake consists of a data computation layer, an open table format, and a performant object store.

With Dremio, Iceberg and MinIO, you can create a modern data lake architecture that will future-proof your data storage and processing needs.

Updating Legacy Architecture

Older architectures are not only expensive to expand to fit your current/future needs, they're expensive to maintain and manage.

A modern data lake with Dremio, Iceberg and MinIO will jump you into the future with a high-performance, cloud-native data architecture that are easy to maintain and even easier to scale.

AI/ML Foundation

AI/ML applications aren't just coming, they're here.

Legacy architectures weren't built to handle these types of workloads. A moderna data lake with Dremio, Iceberg, and MinIO has the performance, versioning and everything else you need to manage your data through the entire AI/ML pipeline, from training to testing to refining.

Flexibility & Scalability

With cloud-native solutions like Dremio, Iceberg and MinIO, you have the flexibility to deploy anywhere, from public and private clouds, to colos, and even the edge. Hybrid and multi-cloud requirements are also easily satisfied.

Auto scaling capabilities of Dremio and MinIO can also happen independently, allowing you to expand capacity only where you need it.

Questions?

- @minio
- https://github.com/minio/minio
- https://slack.min.io
- https://min.io

- @dremio
- https://github.com/dremio/dremio-oss
- https://www.dremio.com