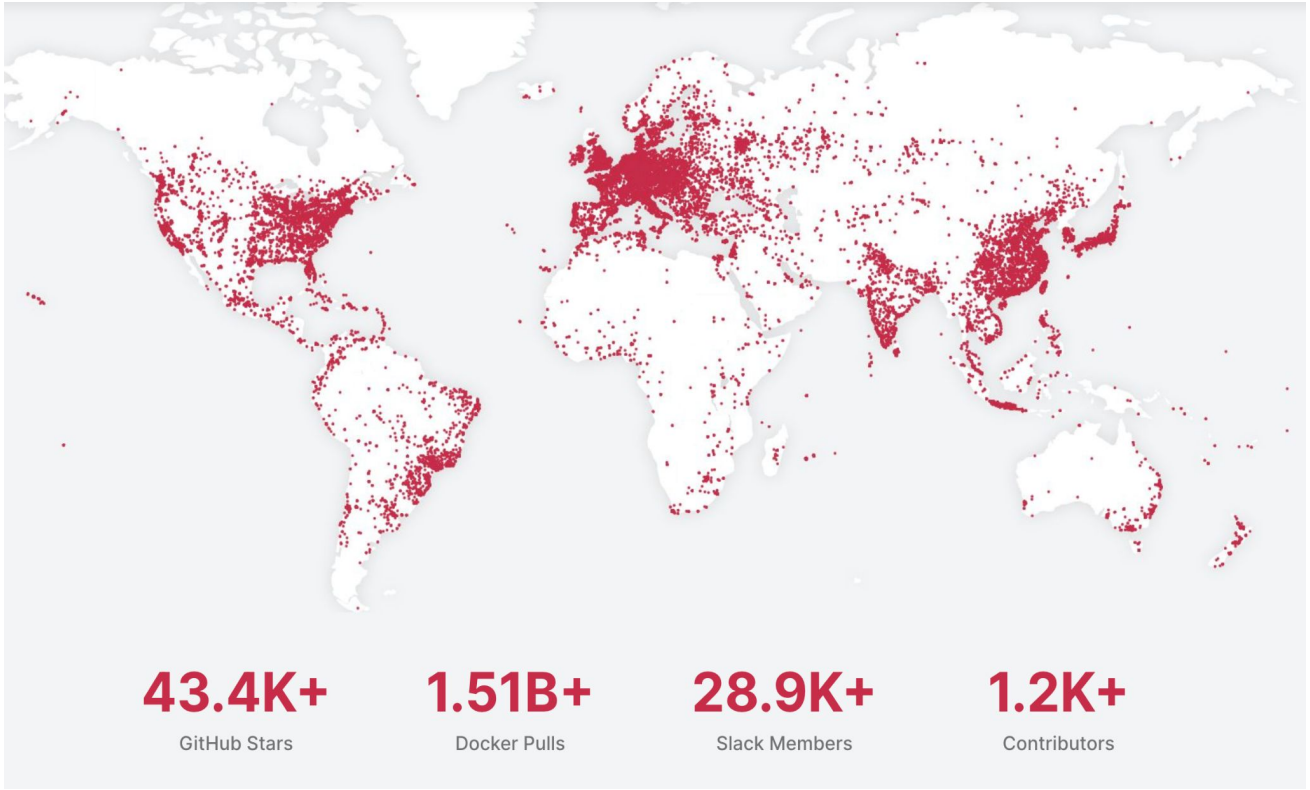
MINIO

Redefining Database Workloads

The Future with Modern Object Storage

Introduction & Credentials



MINIO

Agenda

- Challenges of Traditional Databases
- Why am I hearing about disaggregation of storage and compute?
- Two modern disaggregation strategies

Present Day: Database Challenges

Complex

Databases can become complex, sapping resources and constraining engagement.

Complexity has a material impact on scalability.

Fragmented

Lack of single point of access to data and insights in near real-time imparts risk and obscures value.

Monolithic

Tightly coupled storage and compute results in over-provisioning (generally of compute).

Vendor lock-in to inappropriate tools

Un-Scalable

Traditional systems don't scale well - from performance to economics.

Throughput is a bottleneck.

Traditional

Many traditional technologies simply cannot make the leap to the cloud.

Cloud-native is the way forward.

Why your Database is Failing You

	Longer delivery time	Multiple reconciliations	Performance issues	Higher cost	Inability to scale
Siloed data ecosystems	✓	✓	✓	✓	✓
Multiples copies of data	✓	✓	✓	✓	✓
Difficulty identifying single source of truth		✓	✓	✓	✓
Fragmented data ownership	✓	✓			✓
Redundant application of security protocol	✓	✓	✓	✓	✓
Lack of enterprise wide data visibility		✓			✓

Why am I hearing about disaggregation?

Object Storage as Primary Storage

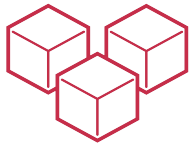
Databases


AI/ML and Stream/ Log Analytics

OS Images, Snapshots, Backup, App Artifacts

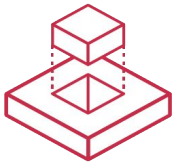
MINIO

The Future of Database Workloads



Disaggregation

Separation of storage and compute is a hard requirement.



Durability + Efficiency

Modern data lakes need to be resilient to failure, distributed and able to quickly recover.

Multi-Cloud

The data lake architecture must be extensible to any public or private cloud.

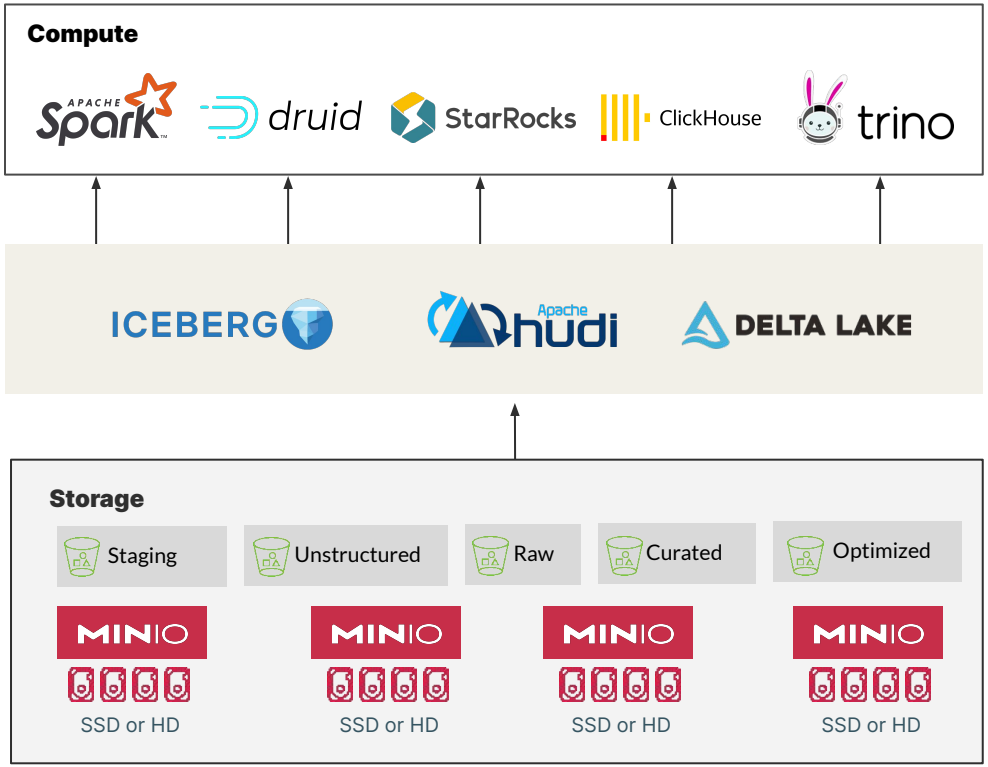
Cloud repatriation can save up to 60%.

Performance

The modern data lake needs to be faster - and it is **with the right stack**.

Two Modern Solutions

Disaggregation of Storage and Compute



Data Lake Semantics

Datalake	Raw	Curated	Optimized
■ All data (structured, semi-structured, unstructured) in one place ■ Supports fast ingestion and streamlined consumption ■ Decoupled storage and compute	■ Data stored in raw format and encrypted. ■ Ledger for transactional events ■ No generic or business transformations ■ No shared consumption, enables data stewards.	■ Data fully cataloged, authoritative source. ■ Access are policies applied. ■ Available for shared consumption ■ optimal partitions based on access patterns ■ Generic transformations.	■ Transformed and organized by consumer use cases ■ Optimized for applications, specialized analysis & performance ■ Domain-level data marts & warehouses supporting complex queries with higher speeds.

Advantages of Parquet vs Table Formats

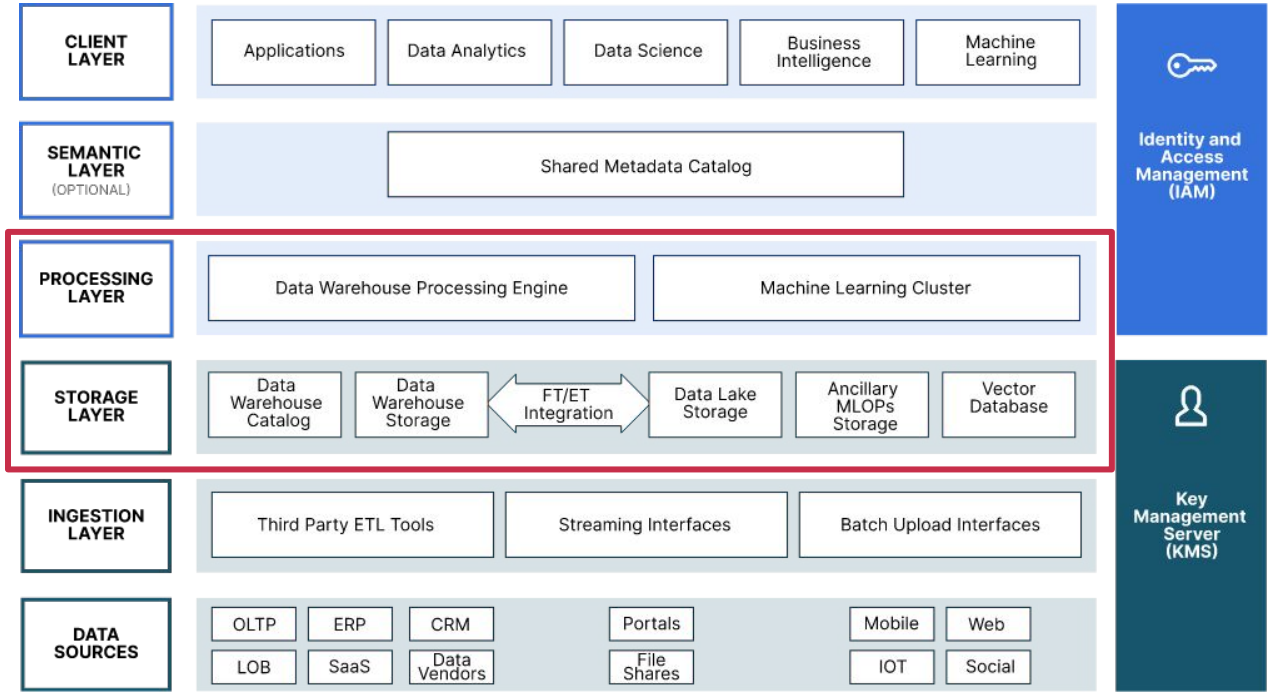
Parquet:

- **Speed:** Parquet employs column-wise compression, different encoding, compression based on data type, predicate pushdown. Better compression ratios + skipping data blocks means fewer bytes red from S3, leading to significantly better query performance
- **Cost:** Services like Macie, Athena, Okera etc. charge per amount of data scanned and Parquet compression helps reduce cost of these scans

Iceberg / Hudi / Delta Lake:

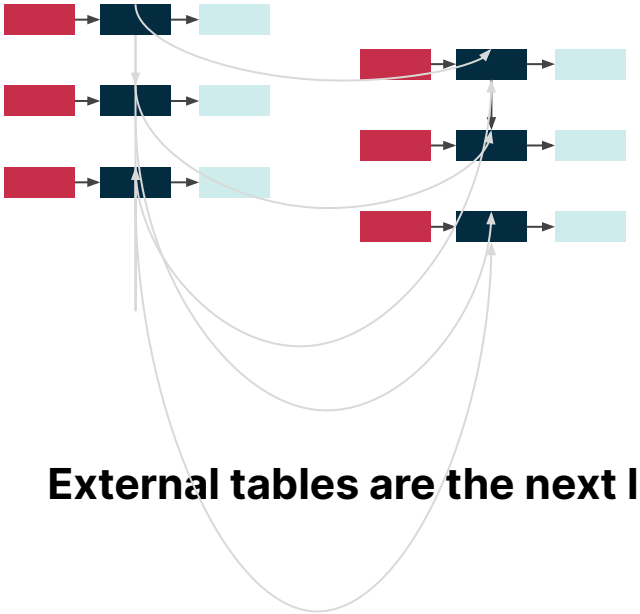
- Provides a table view on S3, similar to a database with indexes.
- Provides upsert / delete capabilities with indexing on the trusted S3 bucket.
- Transactional (ACID) semantics on object storage
- Merge-on-read (optimized write) or copy-on-write (optimized read)
- De-duplication of data with simple and compound keys
- Data versioning (aka Time Travel)
- Schema Evolution and Partition Evolution

The Modern Data Stack

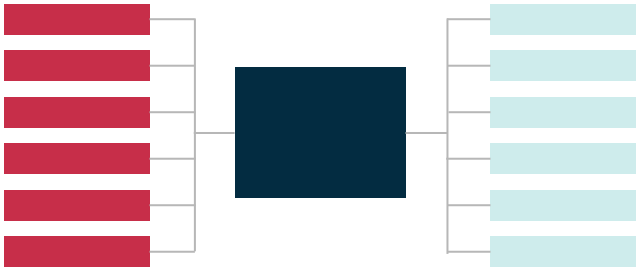


MINIO

Next Generation of Disaggregation

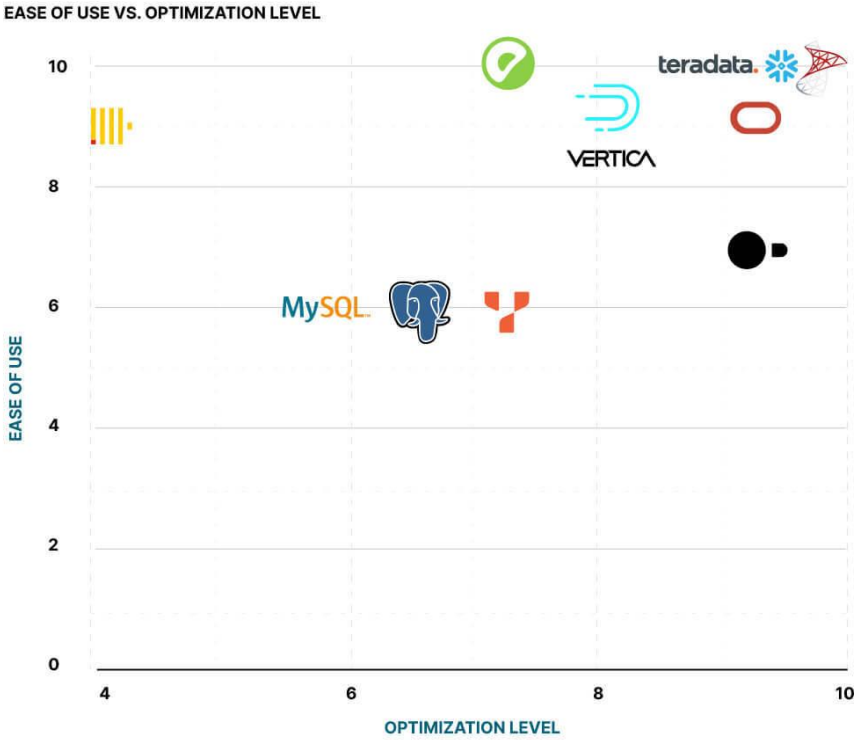


To a cloud-native future state:



External tables are the next logical step in the disaggregation of storage and compute

Databases Optimized for Object Storage



MINIO

Summary

- Present Challenges of Traditional Databases
- The Future of Database Workloads
- Object Storage as Primary Storage for databases
- Sample Architectures & Design
- External Tables as the natural extension of disaggregation

MINIO

Thank you!

Brenna Buuck
SME Databases and Datalakes
brenna@min.io

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