

**CASE STUDY**

# Major Investment Management Firm Achieves Data Democratization with Dremio Lakehouse Platform

## At a Glance

### Challenge

A diversified financial services firm with \$75 billion in assets under management struggled with closed, siloed data systems primarily built around SQL Server databases for web applications. Users couldn't access raw data for analysis, creating bottlenecks when portfolio managers and research teams needed information for investment decisions. The firm lacked self-service capabilities, forcing users to wait for developers to extract data, significantly impacting time-to-insight for critical investment decisions.

### Solution

The firm implemented a comprehensive data lakehouse architecture using Dremio as the semantic layer, Azure Data Lake for storage, and Kubernetes for orchestration. They transitioned from closed SQL Server systems to an open, self-service platform that enables direct access to raw data in the lake without requiring data movement. The implementation included role-based access controls to manage data licensing restrictions and compliance requirements.

### Results

- Achieved complete data democratization across 300+ investment professionals
- Eliminated bottlenecks for portfolio managers accessing critical investment data
- Enabled self-service analytics for research teams, marketing, and decision makers
- Reduced time-to-access for data substantially
- Automated marketing reports and enabled experimentation without developer dependencies
- Maintained system stability with no unplanned outages over 4-5 years

## The Customer

A major diversified financial services firm headquartered in Chicago manages approximately \$75 billion in assets under management across multiple business divisions. The firm operates investment banking, investment management, and private wealth management divisions, serving institutional clients including endowments, pension funds, and teachers' retirement funds. With 400-500 large institutional clients at any point in time, the investment management division employs approximately 300-400 investment professionals supported by a 25-person technology team.

## The Challenge

When the Head of Data Engineering returned to the firm five years ago, the organization faced significant challenges with data accessibility and democratization. The existing data infrastructure was built primarily around Microsoft SQL Server databases that were designed for web application development rather than analytical use cases. This architecture created a closed system where data access was severely limited and controlled.

Portfolio managers, research teams, and other investment professionals had no direct access to the raw data they needed for critical investment decisions. Instead, they had to rely on developers to extract and provide data, creating substantial bottlenecks in the decision-making process. Users often didn't know who to contact or where to go to access specific datasets, leading to frustration and delays in time-sensitive investment analysis.

The closed nature of the system prevented experimentation and exploratory analysis that are crucial for investment research. Marketing teams couldn't automate their reporting processes, and research teams couldn't quickly test hypotheses or validate investment strategies. The firm's diverse user base included highly technical professionals skilled in Python, R, and C#, as well as users who preferred Excel, dashboards, and BI tools, but none of them could effectively access the underlying data.

This infrastructure significantly impacted the firm's ability to make timely investment decisions for their institutional clients. In an industry where speed and access to information can determine investment performance, the data access bottlenecks were creating competitive disadvantages and limiting the firm's analytical capabilities.



## Why Dremio?

The firm's data democratization initiative was driven from the top, with senior leadership recognizing the need to open up data systems for broader organizational use. When evaluating solutions for their planned data lake architecture, the team had the unique advantage of building from scratch rather than cleaning up an existing "data swamp."

The Head of Data Engineering brought experience with AWS technologies like Redshift Spectrum and Athena from previous roles, and was specifically looking for a data lake engine that could read data directly from the lake without requiring ingestion into other systems. This approach would minimize data movement and enable the self-service access that the organization desperately needed.

After hearing about Dremio, the team conducted a comprehensive proof of concept that lasted one to three months. They were particularly impressed with Dremio's performance when reading Parquet and Delta Lake format files directly from their Azure Data Lake. The solution demonstrated the ability to create self-service data access for their broad audience of users with varying technical skills.

The decision-making process involved formal ROI analysis and evaluation of alternatives, though few solutions offered the lakehouse architecture approach they were committed to pursuing. The user interface that enabled self-service data access was a critical factor, as was Dremio's role-based access controls that could handle their complex data licensing requirements where some datasets could only be accessed by specific numbers of users.

The firm conducted internal SWAT team assessments to identify pain points, and leadership was convinced that Dremio could solve the identified challenges while supporting their vision for data democratization.

## The Solution

The firm implemented a comprehensive lakehouse architecture centered around Dremio as the semantic layer and query engine. Their new data platform architecture leverages several key components working together to provide seamless data access and analytics capabilities.

Azure Data Lake serves as the primary storage layer, hosting all of their structured and semi-structured data in formats optimized for analytical workloads. Databricks handles the data ingestion and ETL processes, bringing data from various sources into the lake in a clean, organized manner. Dremio sits as the unified semantic layer, providing direct access to data in the lake without requiring movement to separate analytical systems.

The entire platform runs on Kubernetes using Azure Kubernetes Service (AKS), providing scalability and reliability for their analytical workloads. The team maintains a three-node cluster that generally stays consistent in size, though they've experimented with auto-scaling features to optimize costs during off-hours and development periods.

The firm ingests data from numerous common financial services vendors including benchmark providers like S&P and Russell, stock exchange data providers, and financial statement data vendors. They also capture internally generated research and analysis from their investment teams, creating a comprehensive mosaic of structured and unstructured data sources that can be combined for unique insights.

This architecture enables reading data directly from the data lake without the bottlenecks of traditional data warehouse approaches. Users can access raw data through SQL interfaces, connect their preferred BI tools, or integrate with custom applications through APIs. The role-based access controls ensure compliance with data licensing agreements that may restrict certain datasets to specific numbers of users.

## The Impact

The implementation of Dremio and the lakehouse architecture has transformed how the investment management firm operates, delivering substantial benefits across multiple dimensions of their business.

Data democratization has been achieved across the organization, with 300-400 investment professionals now having self-service access to the data they need for decision-making. Portfolio managers no longer experience delays when they need to access critical investment data for time-sensitive decisions. Research teams can now experiment freely with different datasets and conduct exploratory analysis without waiting for developer intervention.

Time-to-access for data has been reduced substantially, enabling more agile investment decision-making processes. Marketing teams have automated their reporting processes, whether using Excel, Python, Tableau, or Power BI, freeing up valuable time for strategic activities. The ability to combine internal research with external market data has created new analytical possibilities that weren't feasible under the previous architecture.

The system has demonstrated remarkable stability over its 4-5 year operational period, with no unplanned outages that weren't caused by user error or intentional system changes. This reliability has been crucial for a financial services organization where system availability directly impacts business operations and client service.



Cost optimization has been realized through the elimination of complex data movement processes and the reduction in developer time required to support data access requests. The firm has been able to support a much larger user base with the same core technology team, improving overall operational efficiency.

The platform has also enabled the firm to better manage their complex data licensing requirements, with role-based access controls ensuring that datasets with restricted user limits are properly governed while still being accessible to authorized personnel.

## Looking Ahead

The firm continues to evolve their data platform with several strategic initiatives planned for the future. Migration to Apache Iceberg is a high priority that the team has been discussing for two years, though resource constraints have prevented immediate implementation. They recognize that adopting a unified data format will become increasingly necessary as their data volumes and complexity continue to grow.

The team is actively exploring Unity Catalog implementation to improve their data governance and metadata management capabilities. This enhancement would provide better data discovery and lineage tracking, further supporting their self-service analytics objectives.

Auto-scaling improvements for their Kubernetes deployment remain a focus area, with the goal of optimizing infrastructure costs by scaling down during off-hours and development periods when usage is lower. The team continues to work on scaling up capabilities to handle peak demand periods more effectively.

Version upgrades are planned to take advantage of new Dremio features, though the team approaches these carefully with thorough testing and backup procedures. They're currently running version 24 and plan to upgrade to versions 25 or 26 to access enhanced capabilities.

The firm is in the early stages of exploring AI applications, particularly around building retrieval augmented generation (RAG) systems for document search and summarization. They're developing chatbot and knowledge assistant capabilities to help users search large corpuses of unstructured data like PDFs, though agentic AI applications remain under evaluation.

These initiatives demonstrate the firm's commitment to continuously improving their data platform while maintaining the stability and reliability that their investment operations depend on. The success of their Dremio implementation has created momentum for further modernization efforts across their technology stack.

### ABOUT DREMIO

Dremio is the intelligent lakehouse platform for the business, serving hundreds of global enterprises, including Maersk, Amazon, Regeneron, NetApp, and S&P Global. Based on open-source technologies like Apache Iceberg and Apache Arrow, Dremio provides an open lakehouse architecture enabling the fastest time to insight and platform flexibility at a fraction of the cost.

Learn more at [www.dremio.com](http://www.dremio.com).

Dremio and the Narwhal logo are registered trademarks or trademarks of Dremio, Inc. in the United States and other countries. Other brand names mentioned herein are for identification purposes only and may be trademarks of their respective holder(s). © 2025 Dremio, Inc. All rights reserved.