

The Dremio Open Data Lakehouse on Microsoft Azure

Cloud data lakes introduce complexity, time, and risk into data architectures, and limits the value of semistructured and unstructured data. The Dremio Open Data Lakehouse simplifies data architectures and provides self-service access to data in Azure Data Lake Storage (ADLS) for a wide range of data consumers.

The Growth of Data Lakes

For decades, businesses have made large investments in Enterprise Data Warehouses - proprietary, often appliance-based systems for storing and analyzing structured data from business systems, usually housed in the data center - for enterprise Business Intelligence (BI) and reporting. As organizations experienced an explosion in terms of volume and variety of data, data warehouses struggled to scale efficiently. Data teams turned to data lakes - first Hadoop, and then cloud file and object storage - as a cheap and efficient storage solution. Yet data lakes failed to fully replace the enterprise BI & reporting capabilities of the data warehouse.

In a cooperative architecture featuring one or many data warehouses alongside one or many data lakes, accessing data in the data lake requires an Extract, Transform, and Load (ETL) pipeline to move the data into the data warehouse in its proprietary format. These ETL pipelines are often manual and ad hoc. Further, as data volumes grow, BI tools typically require custom data copies in the form of cubes and extracts to meet performance Service Level Agreements (SLAs).

As more data consumers require access to more data, ETL pipelines and data copies proliferate, and the data team becomes a bottleneck for data access. New requests can take weeks or even months to fulfill.

The Dremio Open Data Lakehouse for Microsoft ADLS

The Dremio Open Data Lakehouse simplifies data architectures by enabling direct access to data in Microsoft ADLS - a high-performance cloud file and object storage solution for Microsoft Azure - without moving or copying the data. Dremio brings all of the data management, data governance, and analytics capabilities typically associated with a data warehouse directly to the data lake.

The Dremio Open Data Lakehouse leverages Dremio Sonar, a SQL query engine built on Apache Arrow, that brings full ANSI-SQL functionality to the data lake. Dremio Sonar automatically optimizes query performance so data consumers are not limited in the amount of data they can analyze. Dremio Sonar also features a semantic layer that centralizes data governance and security while providing broad self-service access to data for technical and non-technical data consumers.

Benefits of the Dremio Open Data Lakehouse



Open: Dremio Sonar is built on Apache Arrow, and is capable of querying data in Microsoft ADLS, as well as other sources like relational and NoSQL databases, without moving or copying the data. Dremio Sonar supports multiple table formats, including Apache Iceberg, which is gaining traction for its openness and performance at massive scale. With Dremio, data remains open in ADLS and accessible by multiple engines and tools for a variety of use cases.



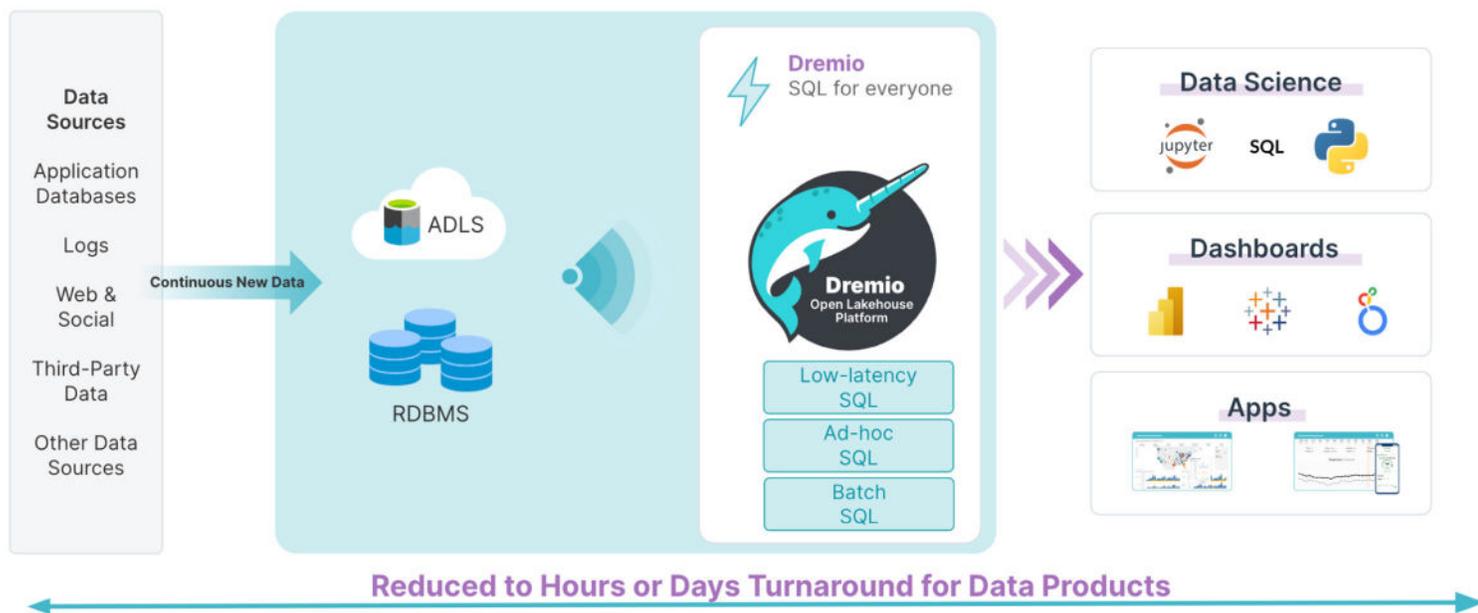
Built for SQL: SQL remains the most common and widely used data skill set in the enterprise. Dremio is built for SQL users so technical and non-technical data consumers can access data for analytics.



Self-Service Analytics: Dremio's semantic layer gives data consumers the ability to join and query data without requiring complex ETL pipelines or data copies. Data requests take just hours or days to fulfill, and data teams are free to focus on innovating with data instead of managing a backlog of requests.



Performance at Scale: Dremio automatically optimizes queries, and delivers interactive analytics on billions of rows of data, so users are not limited in how much data they analyze.



Use Cases

011001 Data Democratization: Technical and non-technical data consumers need access to a consistent and near-real-time view of data to satisfy a wide range of analytic use cases. Dremio delivers direct access to the data lake and other data stores, gives users the ability to join and query data in place, and provides a consistent and accurate view of the data with its no-copy architecture.

Data Lake Modernization: Data teams invested in the data lake as a cheap and efficient means of storing large volumes of data, but data lakes lacked the enterprise BI & reporting capabilities of the data warehouse. The Dremio Open Data Lakehouse combines the scalability and flexibility of the data lake with the data management, data governance, and analytic capabilities of the data warehouse. Dremio's semantic layer delivers access to disparate data volumes so businesses don't have to wait to have a unified view of their data.

Data Exploration: Dremio centralizes data management and governance while providing broad access to data for a wide range of analytic use cases, including interactive dashboards and ad hoc queries on large volumes of data. Users can explore data in tools like Power BI, using native connectors that deliver a seamless experience across environments.

Getting started with Dremio is easy.

Visit <https://azuremarketplace.microsoft.com/en-us/marketplace/apps/dremiocorporation> and get started with your Forever Free edition of Dremio today. All you pay for is Azure infrastructure!

About Dremio

Dremio is the lakehouse company. Hundreds of enterprises around the world, including brands like Allianz Global Investors, FactSet, Knauf Insulation, Nutanix and OTP Bank, use Dremio's SQL engine to deliver mission-critical BI on the lake. As the original creator of Apache Arrow, Dremio is on a mission to reinvent SQL for data lakes and meet customers where they are in their cloud journey. Dremio was founded in 2015 and is headquartered in Santa Clara. To learn more, follow the company on [GitHub](#), [LinkedIn](#), [Twitter](#), and [Facebook](#), or visit www.dremio.com

Dremio, Sonar, Arctic, and the Narwhal logo are registered trademarks or trademarks of Dremio Corporation 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). © 2022 Dremio Corporation. All rights reserved.