DATASHEET

Dremio for AWS

Enable Interactive BI on Amazon S3

Dremio is a high-performance SQL lakehouse platform built on an open data architecture that enables you to drive high-performance, mission-critical BI and interactive analytics directly on Amazon S3 from your AWS account — in just minutes. Dremio simplifies data engineering and eliminates the need to copy and move data into proprietary data warehouses, or create cubes, aggregation tables or BI extracts, providing flexibility and control for data architects and data engineers and self-service for data consumers. Your data stays in your S3 buckets, and your EC2 instances run in your account. Data teams gain flexibility and maintain full control, while providing analysts with easy access to data and reducing time to insight. Dremio enables lightning-fast BI dashboards and interactive analytics directly on Amazon S3 through seamless integrations with Tableau, Power BI and other BI tools, using groundbreaking Dremio technologies and innovations.

Dremio Is a High-Performance SQL Lakehouse Platform

Why Dremio:

Enable a Delightful BI Experience

- Enables mission-critical BI dashboards and interactive analytics directly on cloud data lake storage such as Amazon S3

- Doesn't require copying data into proprietary warehouses, marts, extracts or cubes

- Empowers data analysts and data scientists to discover, curate, analyze and share datasets in a self-service manner

- Allows users to build interactive dashboards that bring their data lake to life through native Dremio connectors in tools such as Tableau and Power BI

Simplify Data Engineering

- Eliminates the need for costly, complex and rigid data pipelines to move and copy data into proprietary data warehouses

- Enables data engineers to focus their time on strategic projects instead of responding to data access and ETL/ELT requests

- Helps data engineers eliminate data sprawl and inconsistent reports

- Centralizes security and governance through a self-service semantic layer

Future-Proof Your Data Architecture

- Enables you to build an open data architecture that allows you to use best-of-breed tools directly on cloud data lake storage

- Eliminates vendor lock-in and resulting long-term financial and technological dependence

- Allows you to easily adopt new technologies that may emerge in the future, as they become available in the market

Differentiators

Open

Separate data, not just storage, from your compute to future-proof your analytics architecture and leverage best-of-breed applications and engines today — and tomorrow.

Lightning-fast

Accelerate ad hoc and BI queries 100x versus SQL engines, eliminating the need for cubes, extracts or aggregation tables, or even to ETL your data into a data warehouse.

Productive

Provision new datasets with consistent KPIs and business logic in minutes, not days or weeks. Empower analysts to create their own derivative datasets, without copies.

Efficient

Reduce compute infrastructure and associated costs by up to 60% when running on AWS, by eliminating the need to over-provision infrastructure, with Dremio's right-sized, auto-start/stop compute engines that dynamically scale based on current workload demands.

> "Dremio is helping us democratize data and deliver analytical solutions far quicker than normal.
>
> **- G2 Customer**

100x
faster ad hoc queries

100x
Faster BI workloads

0
cubes, extracts or aggregaton tables

Lightning-Fast Queries

Dremio's high-performance execution engine, which is powered by Apache Arrow and accelerated by several proprietary technologies, speeds queries by 100x and eliminates up to 60% of compute costs.

Consistent Semantics Across All Users, Apps and Tools

Dremio's vertically integrated semantic layer, with shared business logic, data access, KPIs and security, enables exploration and sharing of data in real time and provides a single, consistent view of the data for all teams and tools and applies security and data governance across the company.

> "We found the following capabilities of Dremio to be of particular value to The Hartford: Dremio enables us to easily access and query data that previously was not possible or very difficult. Dremio natively supports the following cloud-based data sources: Amazon Elasticsearch, Amazon Redshift, Amazon S3 and other data sources.
>
> **- The Hartford Insurance Group**

Dremio allows NewDay, a financial services organization, to improve productivity for business users and data scientists through self-service access to data on S3. The solution is considerably faster with most queries returning in seconds versus minutes or hours on the legacy systems.

Read more about the NewDay case study.

Self-Service Data Access

Dremio's no-copy architecture eliminates long backlogs of requests and empowers data scientists, data analysts and data consumers through self-service access to business-ready data in real time.

> " I want my teams to be much faster to quantify hypotheses in terms of costs, quality and time. Dremio helps us to access all data assets with the tools of choice to provide true self-service at BI.
>
> **- Paul Petraro, Global Head of RWE, Boehringer Ingelheim**

Elastic Engines

Dremio enables independently sized query engines with on-demand, elastic scale and automated start and stop. Dremio's elastic, multi-engine architecture scales infinitely and helps manage workloads with the highest concurrency and lowest latency. This helps companies tackle any level of concurrency while maintaining consistent performance.

Your data stays in your control

All data is stored and processed within the customer account, ensuring that customers have full control of their data. There are also no inbound connections into the data plane, so customers don't have to poke holes in their firewalls/security groups, ensuring more security and governance.

> " How to handle maintenance and memory heap - Drill gave me severe headaches multiple times. The elastic engines from Dremio and its native AWS integration convinced us already in the early stage.
>
> **- Damon Kheir-Eldin, Principal Analyst, Boehringer Ingelheim**

Advanced Security

Dremio provides LDAP authentication, role-based authorization, Apache Ranger integration, on-the-wire TLS encryption and secure impersonation. Dremio is designed from the ground up to deliver industry-leading, enterprise-grade infrastructure security and is purpose-built to provide the industry's highest levels of security with multiple layers of protection. Data is encrypted both at rest and on the wire, with state-of-the-art key management that ensures secure communication between Dremio and client applications. Advanced role-based access control (RBAC) enables companies to define privileges on every dataset and object in the system.

Parallel Projects

Create multi-tenant Dremio instances with end-to-end automation across deployment, configuration, and upgrades.

> " Made us rethink our whole architecture!
>
> **- G2 Customer**

How It Works

The Dremio provisioning process is based on an AWS CloudFormation template (CFT) that is launched from the AWS Marketplace. Users leverage the existing VPC and subnet within their own tenancies to seamlessly provision all required Dremio resources defined by the CFT.

The Dremio coordinator is responsible for creating and launching a Dremio project which contains all customer-specific definitions, metadata, logs and admin settings stored in the dedicated S3 bucket. It also manages and controls Dremio elastic engines, which are responsible for execution and processing of queries and can be dynamically provisioned, started or stopped on demand as well as scaled up or down.

The Dremio coordinator and Dremio engines leverage direct-attached elastic block storage (EBS) to store configuration, logs and operational data. Dremio engines also utilize EBS to create and store data in the columnar cloud cache (C3) which is used to accelerate query performance.

Both the Dremio coordinator and engines discover and access datasets directly from S3 cloud storage or through the AWS Glue Catalog.

Dremio AWS Edition deployment architecture
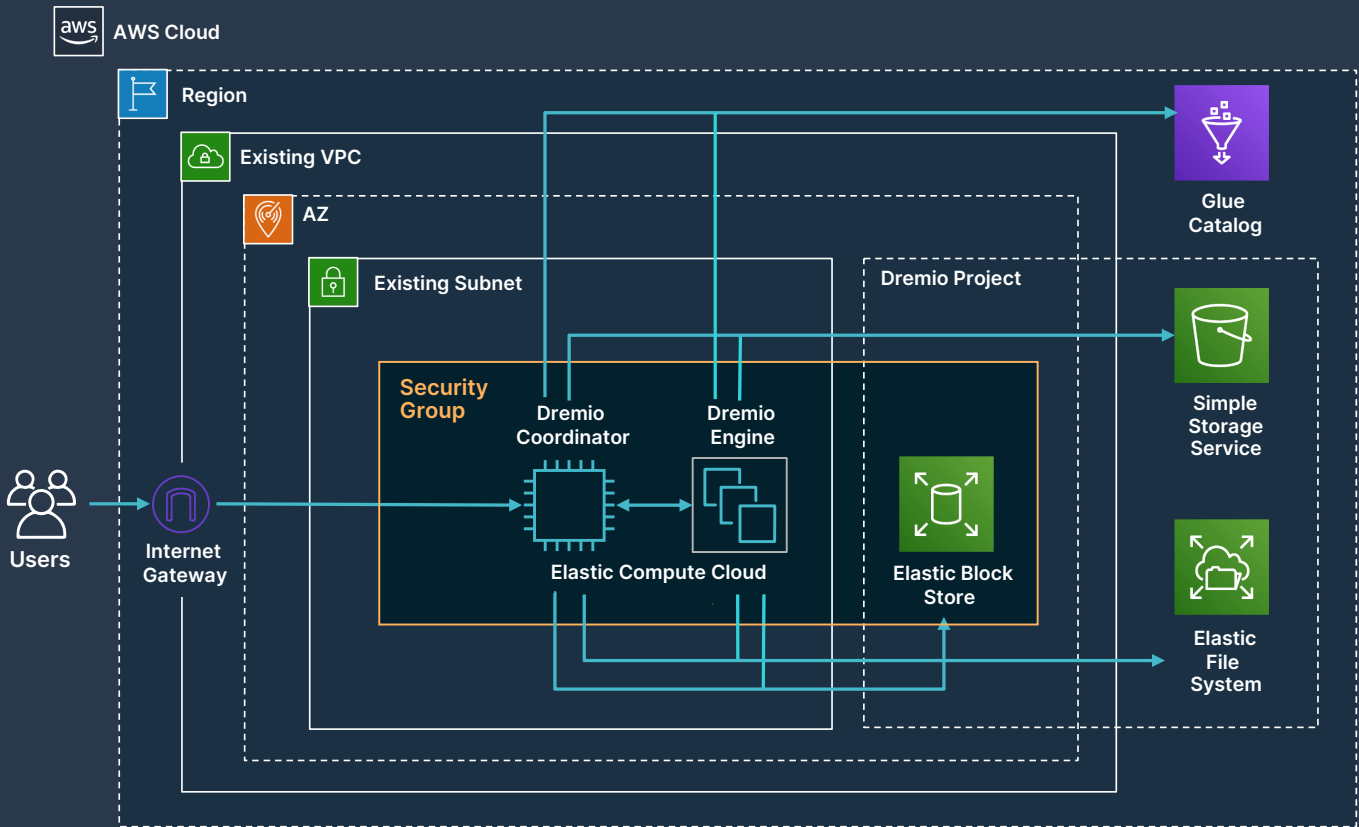
60%
Lower compute costs

Eliminate the need to over-provision infrastructure with right-sized engines that automatically start, stop and scale based on current workload demands.

0
No noisy neighbors

Engines are physically isolated, so workloads can run independently without bottlenexks and resource contention.

100%
Control of resources

Manage resource allocation by using workload management rules to route queries to engines.

Differentiators

- **Open** – Separate data (not just storage) from your compute to future-proof your analytics architecture and leverage best-of-breed applications and engines today — and tomorrow.

- **Lightning-fast** – Accelerate ad hoc and BI queries 100x versus SQL engines, eliminating the need for cubes, extracts or aggregation tables, or even to ETL your data into a data warehouse.

- **Productive** – Provision new datasets with consistent KPIs and business logic in minutes, not days or weeks. Empower analysts to create their own derivative datasets, without copies.

- **Efficient** – Easily size the minimum compute you need for each workload, and only consume compute when running queries. Reduce compute infrastructure and associated costs by up to 60% when running on AWS, by eliminating the need to over-provision infrastructure, with Dremio's right-sized compute engines that automatically start, stop and dynamically scale based on current workload demands.

Dremio solution is available in [AWS Marketplace](#).

ABOUT DREMIO

Dremio is a SQL Lakehouse Platform company enabling organizations to leverage open data architectures. Dremio's SQL Lakehouse Platform simplifies data engineering and eliminates the need to copy and move data to proprietary data warehouses or create cubes, aggregation tables and BI extracts, providing flexibility and control for data architects and data engineers, and self-service for data consumers. Dremio Cloud, a frictionless, infinitely scalable service, enables high performance SQL workloads directly on cloud storage, eliminating the cost and complexity of copying and moving data. Dremio Cloud reimagines the traditional data lake by combining the best of traditional data warehouses and data lakes into a SQL lakehouse, while removing the limitations of traditional data warehouses resulting from closed data architectures.