

Make Dell EMC PowerScale and ECS with Dremio the Centerpiece of your Data Analytics Infrastructure

Your customer's most trusted data resides in Dell EMC PowerScale and ECS. Unlock the full potential and maximum efficiency of Dell EMC PowerScale and ECS for business intelligence, analytics, and data science workloads with the Dremio Data Lake Query Engine. Together, Dell Technologies and Dremio offer a solution that enables your customers to query the data residing on Dell EMC PowerScale and ECS directly, eliminating the need for ETL, cubes, extracts or any data movement away from Dell EMC PowerScale and ECS. The solution is easy to deploy, highly-scalable and provides significant performance and cost benefits to customers. A self-service semantic layer makes it easy for non-technical users to access and analyze data. With Dremio, you can drive your customer's data migration from legacy HDFS data warehouses and legacy HDFS database systems to Dell EMC PowerScale and ECS.

Dremio's Data Consolidation Engine with Dell EMC PowerScale and ECS



Solution Highlights

- Lightning-fast queries directly on top of Dell EMC PowerScale and ECS
- Up to 4X faster ad-hoc queries vs. traditional data consolidation engines
- 10-100X BI query cost reduction
- No data movement or transformation
- Built-in Dell EMC connector
- Easy-to-manage Dell EMC infrastructure

Faster time to insight

- ✓ Lightning-fast queries directly on Dell EMC PowerScale and ECS storage.
- ✓ No need to move data into AWS S3, ADLS, Hadoop, MPP databases or data warehouses.
- ✓ Data analytics workloads are compute and storage-intensive, requiring Dell EMC PowerScale and ECS performance, manageability, and scalability.

Best-of-breed query acceleration technologies

Lightning-fast queries, directly on data consolidation storage

Dremio technologies like Data Reflections, Columnar Cloud Cache (C3) and Predictive Pipelining work alongside Apache Arrow to execute queries directly on data consolidation storage at interactive speed. These query acceleration technologies combine to deliver 4-100x faster performance compared to traditional data consolidation engines¹.

Self-service semantic layer

This abstraction layer enables data engineers to apply security and business meaning while enabling analysts and data scientists to easily explore data and derive new virtual datasets.

Flexible and open

Dremio enables your customers to avoid vendor lock-in—they can query data directly in the cloud or on-prem and keep their data in storage that they own and control. Dremio maximizes customer flexibility and freedom to use their data as they see fit.

¹Based on Dremio internal performance benchmarking, May 2020.

Case Study: Automotive Manufacturer

Despite the massive increase in cloud data lakes and technology, enterprises continue to rely on very large on-premises data ecosystems that are either already at capacity or tightly-coupled. Investing further in these rigid legacy solutions is costly and will likely add to the complexity of a data pipeline that's already unwieldy. Enterprises need to take advantage of the economics that exist with a decoupled architecture which separates compute from storage. Dremio customers leverage Dell EMC PowerScale and ECS with Dremio to make this a reality.

The Customer's Challenges:

Complex ETL. The customer's data pipelines consisted of multiple hops across different systems. In some cases, data originated in a data warehouse system, was transferred to HDFS for historical purposes and, finally, then transferred back to a data warehouse to support workloads that could not run on the saturated HDFS analytics environment. This was not a sustainable solution.

Data Latency. SQL engines (e.g., Hive and Presto) were deployed with limited success. The performance was simply inadequate (sometimes by orders of magnitude), the data was too messy, and the infrastructure prohibitively expensive. Additionally, users were unable to directly connect their BI tools to the data lake.

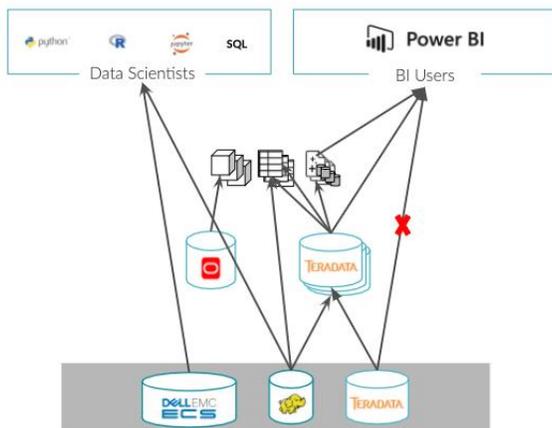
Unscalable. The legacy data consolidation infrastructure and toolset were far too expensive and complex to scale to support the growing number of data consumers who were emerging throughout the organization.

The Solution: Dell EMC ECS + DREMIO

Below is an architectural diagram of the "before" architecture where many ETL workloads were needed to deliver data to various systems and users. As you can see in the "after" architecture, Dremio is able to provide performant access to all the data directly within Dell EMC ECS storage. As a result, our joint customer was able to:

1. Greatly simplify their IT infrastructure;
2. Eliminate the data latency introduced when delivering consumable data to end users, and;
3. Deliver performant access to the analysts whenever they need it.

Existing Unscalable Architecture



Simplified, Performant Access to the Data Lake



To learn more about our offerings visit our website
DellTechnologies.com/StorageforDataAnalytics