dremio

Data-as-a-Service Platform

The demand for data is exploding

Software Data Analytics Market Size ($75B) - 2022
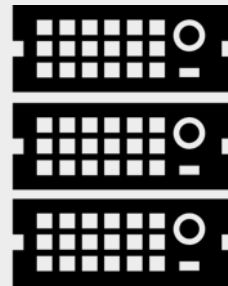
36%
27%
17%
15%
5%
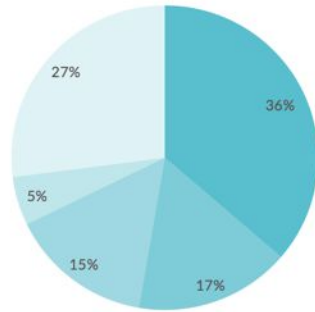
BI · Data Integration · Data Lake · EDW Acceleration · EDW

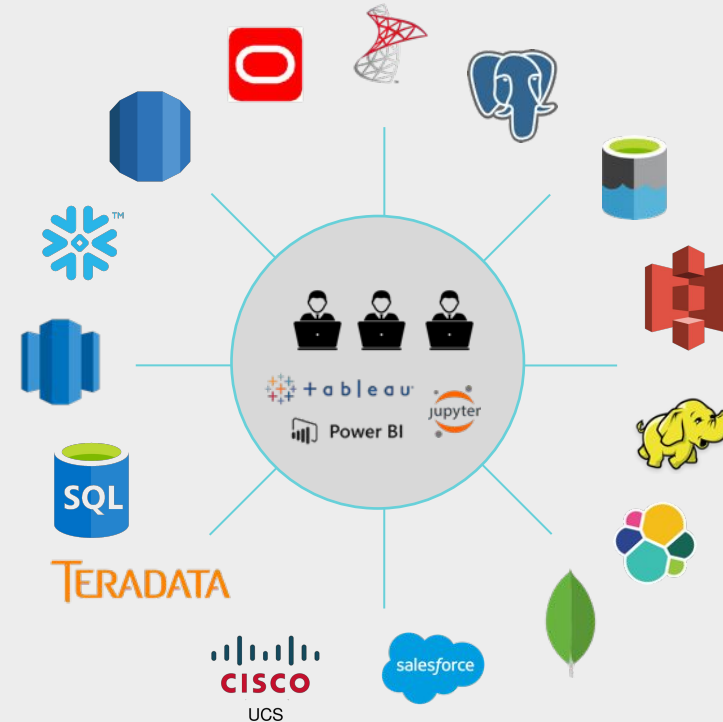$75B annual spend on traditional tools

150M data consumers
60% under-utilized = $9T

THEN

Orders of magnitude more data, more sources, more tools/apps, more data consumers, ...

tableau
Power BI
jupyter

SQL
TERADATA
CISCO
UCS
salesforce

NOW

dremio

Everything-as-a-Service

Infrastructure

Dev Tools

Applications

What about my data?

✔ It's an engineering project every time

✔ It's only getting harder...

✔ What if business analysts could be more self-sufficient?

Data-as-a-Service is a $100B opportunity

1999	2006	2009	Now

Software as a Service

Infrastructure as a Service

Mobility as a Service

Data as a Service

dremio

About Dremio

dremio

There's a better way,

dremio

✓ Founded 2015, Dremio 1.0 July 2017

✓ Co-creators of Apache Arrow, 2.5m+ downloads/month

✓ New category: Data-as-a-Service Platform

✓ Headquartered in Santa Clara, CA

✓ Open core

Fortune 500/Global 2000 Customers in all industries

dremio

Breakthrough in interactive performance with Cisco UCS
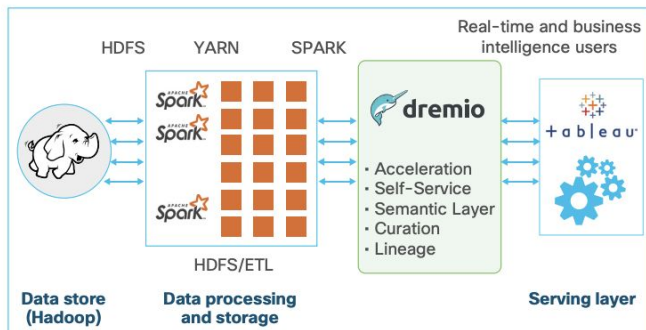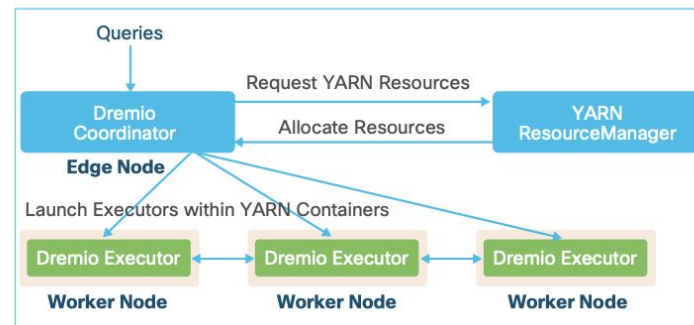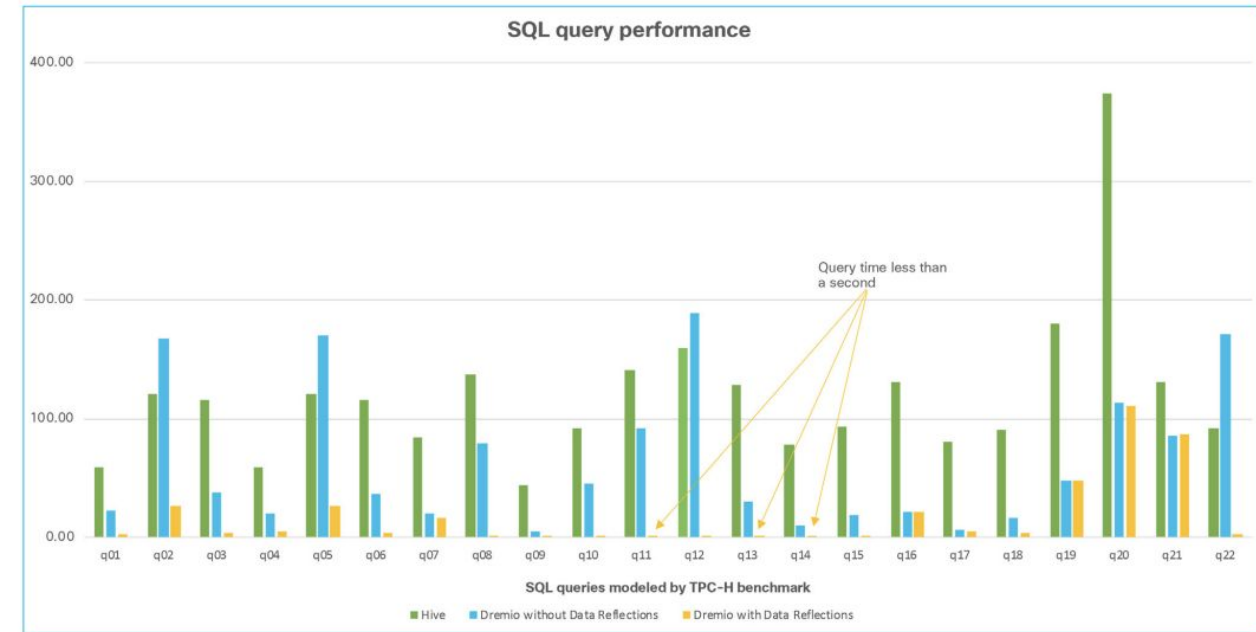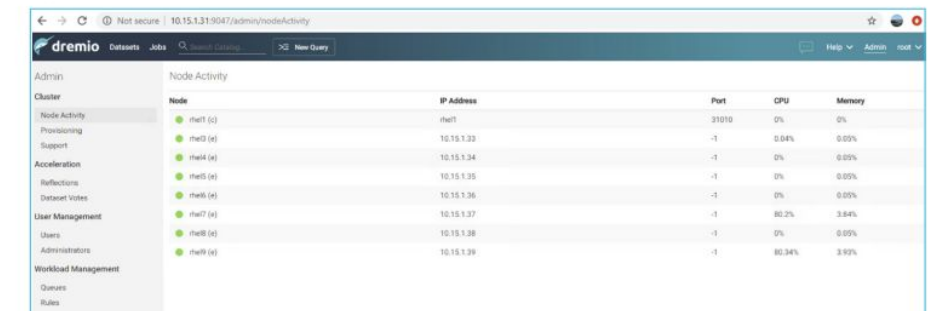
Customer Examples

New self-service app for tier 1 customers

TransUnion's PRAMA platform allows customers to explore a deep history of large volumes of data themselves rather than requiring IT intervention. The goal is to make PBs of data easy to access, use with a fast experience.

PROBLEM

- Total volume = PBs of Data
- Individual data sets = 100's of billions of rows
- Feeling that IT was in the way
- Forced to have entire agile teams devoted to tuning tools and machines
- Poor analyst experience

SOLUTION

- Dremio implemented between BI and Data Science tools and TU Data Lake
- Data Reflections & Virtual Datasets provide acceleration and a "handoff between IT and the end users"
- True technology partnership with Dremio

BUSINESS OUTCOME

- Empowered Analysts to be self-service for the first time
- Individualized interactive dashboards at speed of thought
- 5-10x immediate performance gain – before implementing reflections.
- Able to refocus ~14 Data Engineers to building new products that impact the business

dremio

Tableau and SQL on HDFS

Standard Chartered's Data Fabric team was creating a virtual data layer to simplify data access for analysts, but performance of SCB's SQL, BI and Ad Hoc analytics tool made the approach impractical for their typical data consumer.

PROBLEM	SOLUTION	BUSINESS OUTCOME
• Multiple EDW & Data Lakes. Cannot efficiently analyze information across these systems • SQL on Big Data is too slow and cannot handle enterprise needs • Tableau/Microstrategy on HDP are slow; can't connect to Search, NoSQL • Business cannot easily discover datasets across enterprise • Limited visibility into usage, lineage to improve governance	• SCB found Dremio was both easier to use and outperformed Presto for cross-platform analytics • Users of Tableau, Microstrategy and SQL could search the enterprise catalog creating virtual data sets • For Tableau and SQL, Dremio delivered queries 20 or more times faster than Hive/LLAP	• Tableau and SQL can be used at interactive speeds on big data and previously incompatible data sets • Analysts can perform Ad hoc analysis faster and more easily new data sets • SCB can cost effectively scale the analytic environment to keep pace with growing data volume • Can eliminate many intermediate tables and cubes

Teradata Retirement

NCR, once the parent of Teradata, was seeking to shift workload away from their higher-cost Teradata EDW to a "modern" Hadoop-based analytics infrastructure. They identified Dremio as a way to accelerate the adoption of the new platform.

PROBLEM

- Moving data pipelines to modern data analytics infrastructure (Hadoop) was taking 2-3 months
- Migration involved many outside consultants mapping legacy schemas to produce the required dashboards
- Multiple BUs were not satisfied with rigidity, turnaround time for data requests, and poor performance of the new data lake-based analytics
- Due to Hadoop performance problems, business users would not be able to operate on their business semantic model

SOLUTION

- NCR deployed Dremio on HDP cluster sitting between data stored in HDFS and client analysis tools
- Complex table schemas reproduced via Virtual Data Sets in days by data engineers brand new to Dremio
- Interactive performance on the data lake, with Tableau dashboards performing better than in Teradata

BUSINESS OUTCOME

- Data pipelines and associated queries / reporting / dashboards created in days vs. months
- Performance significantly increased 3 minutes in Hive and 10 seconds in Teradata to 1-2 seconds on complex queries
- High priced consultants no longer needed to develop and deploy use cases on HDP

OneNote: replatform analytics to Azure data lake

Microsoft

Focused on analyzing how customers are using their products to identify potential enhancements. IT's goal was to enable product managers and analysts to directly query and visualize large amounts of data in ADLS using Power BI without requiring IT.

PROBLEM	SOLUTION	BUSINESS OUTCOME
• Poor performance forced building of cubes using SSAS • SSAS not built for big data and building cubes is cumbersome & time consuming • Using cloud data warehouses to pre-process data was costly • Expensive to maintain multiple copies of data • No semantic layer – consumers don't know which cube or dataset to use	• Microsoft team found Dremio after investigating multiple tools that did not satisfy their needs • Installed and tested in weeks with limited assistance • Implemented Dremio between Power BI and ADLS	• Single copy of data • Reduced complexity and costs associated with building and maintaining cubes • Interactive speed for Analysts using Power BI directly on ADLS • Dynamic Security Controls to govern access and maintain data lineage

dremio

Data exploration & analysis ("Customer 360")

Building a digital platform by migrating to the cloud (Azure) & making data accessible to business analysts and data scientists. IT's goal was to empower data consumers while minimizing the need to build data warehouses and data marts.

PROBLEM	SOLUTION	BUSINESS OUTCOME
• RCCL wanted to Migrating data to the cloud (Azure) with ADLS as the landing zone/lake • Focused on making data accessible to the business analysts • Didn't want to spend months & $$$ modeling and loading all the data into data warehouses	• RCCL team deployed Dremio on Azure with ADLS as the primary data source • Up and running immediately without the need to model data • Implemented Dremio between BI tools and ADLS, SQL Server and MongoDB	• Azure-based data lake with Dremio & Spark as the primary processing technologies • Business analysts use Power BI and Dremio to query data and collaborate (virtual datasets) • Interactive speed for Analysts using Power BI directly on ADLS • Raw reflections offload analytical queries from SQL Server & MongoDB

dremio