



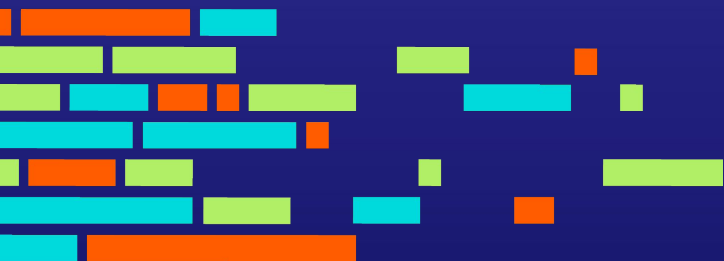
GNARLY
Data_Waves

PRESENTED BY  **dremio**

EPISODE 24

Building a Data Lakehouse on Azure Data Lake Storage

SPEAKERS



Ellie Vail

Product Manager, Azure Data Lake
Microsoft



Tony Truong

Sr. Product Marketing Manager
Dremio

Data is the most strategic asset in today's data driven digital economy



Gaining value from data is at the heart of everything we do. Data is the product.

Empower data driven digital economy with Blob and Data Lake Storage



Scalable, durable, available



Secure and compliant



Cost Effective



Comprehensive data management



Foundational for demanding workloads

Blob and Data Lake Storage key use cases



Machine Learning/AI



Analytics



HPC

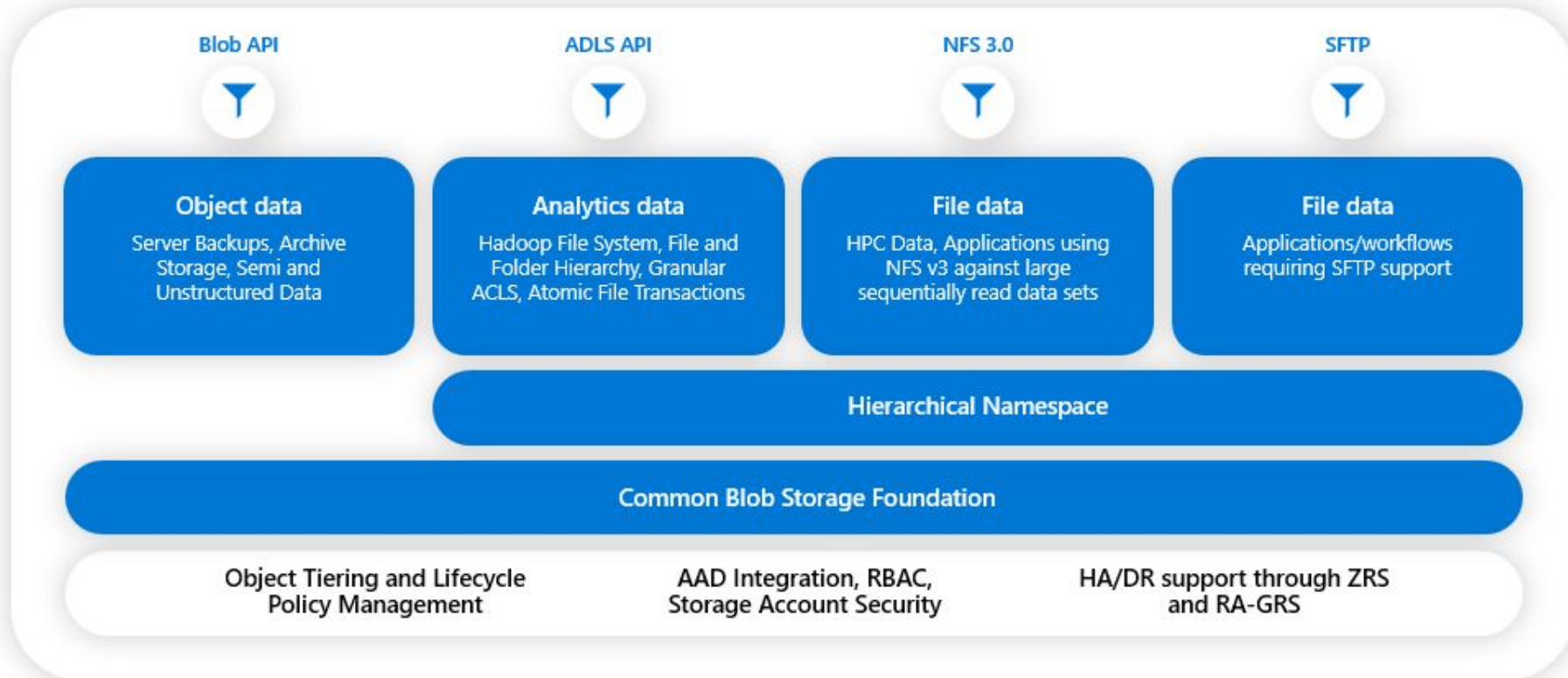


BCDR




Cloud-Native Apps

No Data Silos - Multi-protocol, single storage platform



Leading the Present and Future of Data Analytics




2010 – 2015

Big Data + Open Source

Introduced big data processing

- Fixed compute & storage capacity
- Mostly on-prem
- Harder to use & manage




The Movement

Open Data Lakehouse

Self-service analytics on the data lake

- Data in open file and table formats
- No need to copy & move data
- Multiple best-of-breed processing engines



1980 – 2010

Enterprise Data Warehouse

Transactional and analytical workloads

- Fixed compute & storage capacity
- Mostly on-prem
- Harder to use & manage



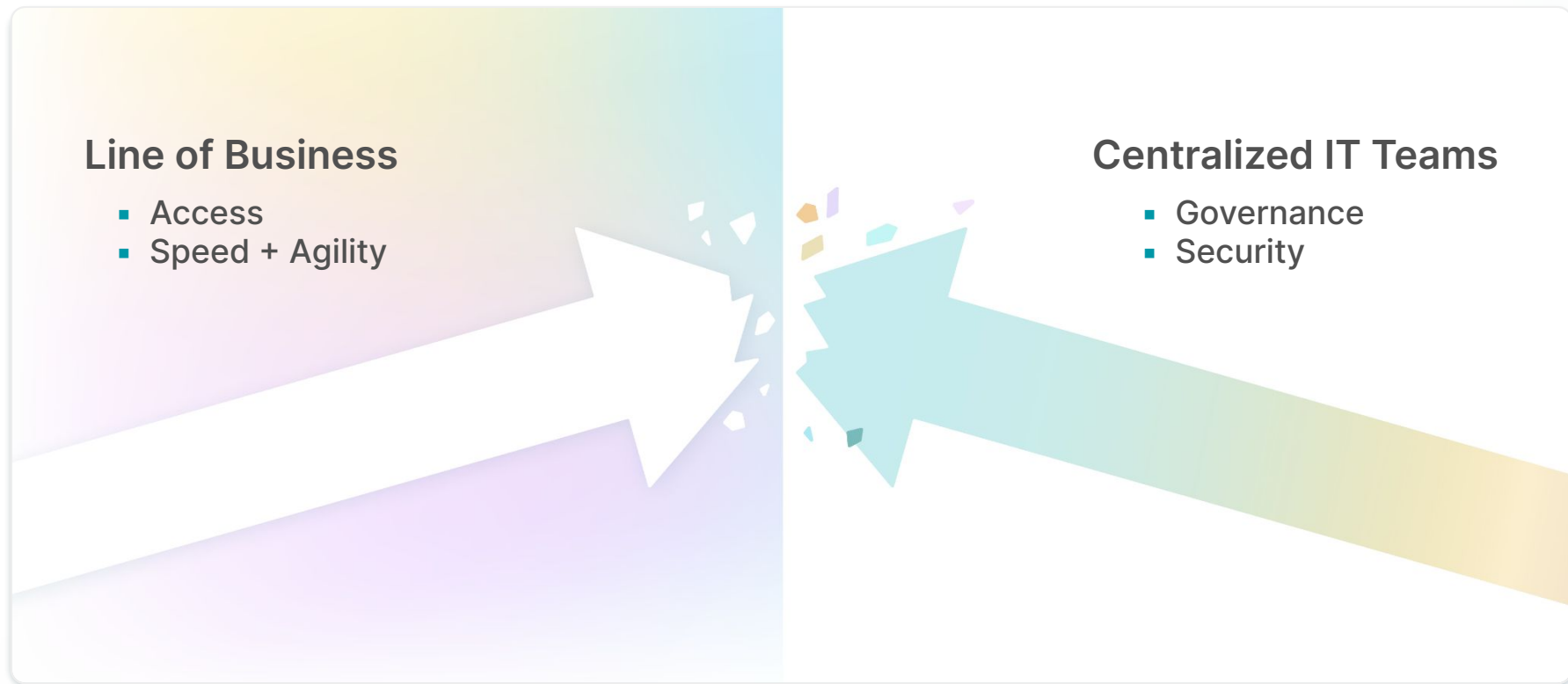
2015 – Present

Cloud Data Warehouse

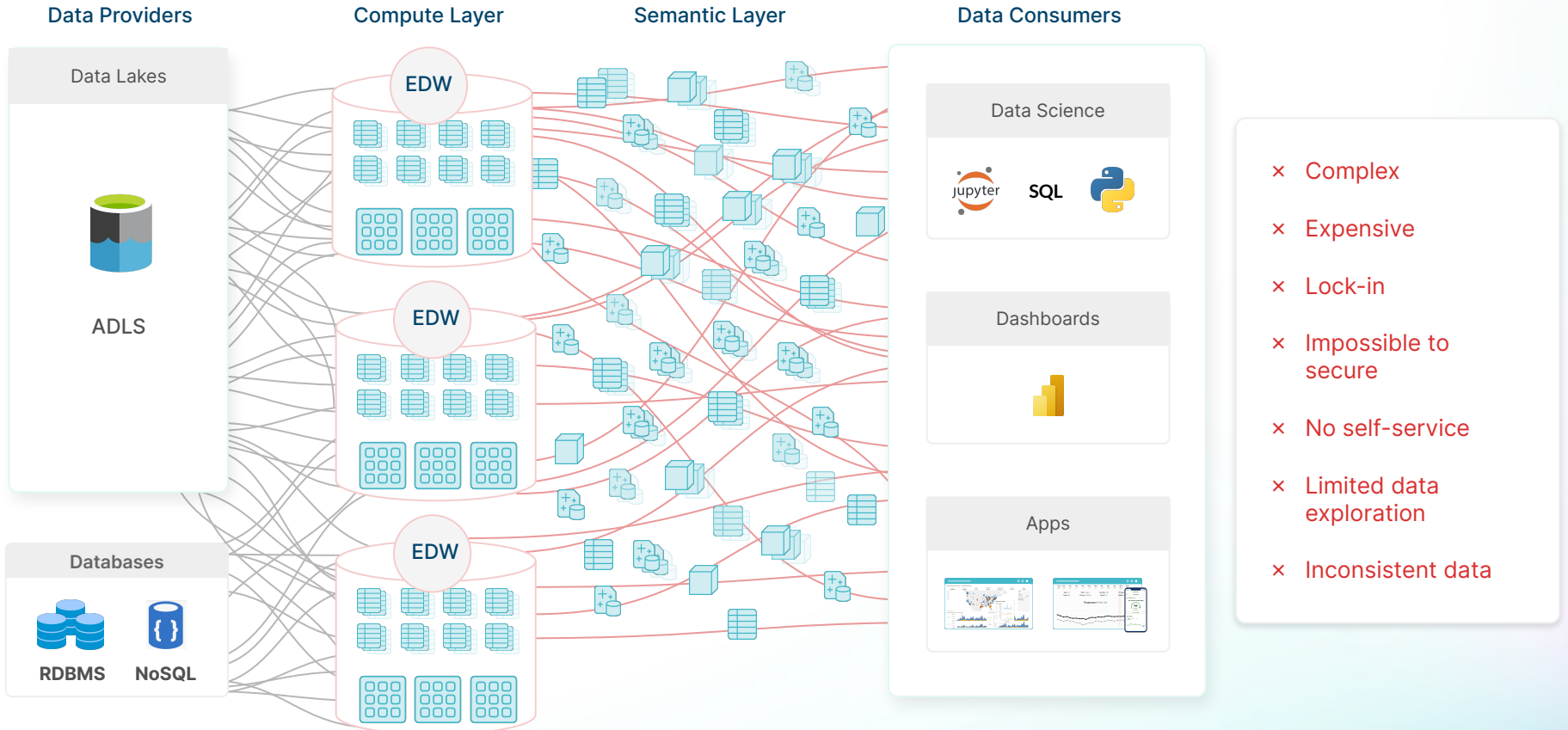
Analytics on cloud data warehouse

- Scale storage and compute independently
- Must load data into proprietary system
- Limited to one processing engine
- Cost prohibitive

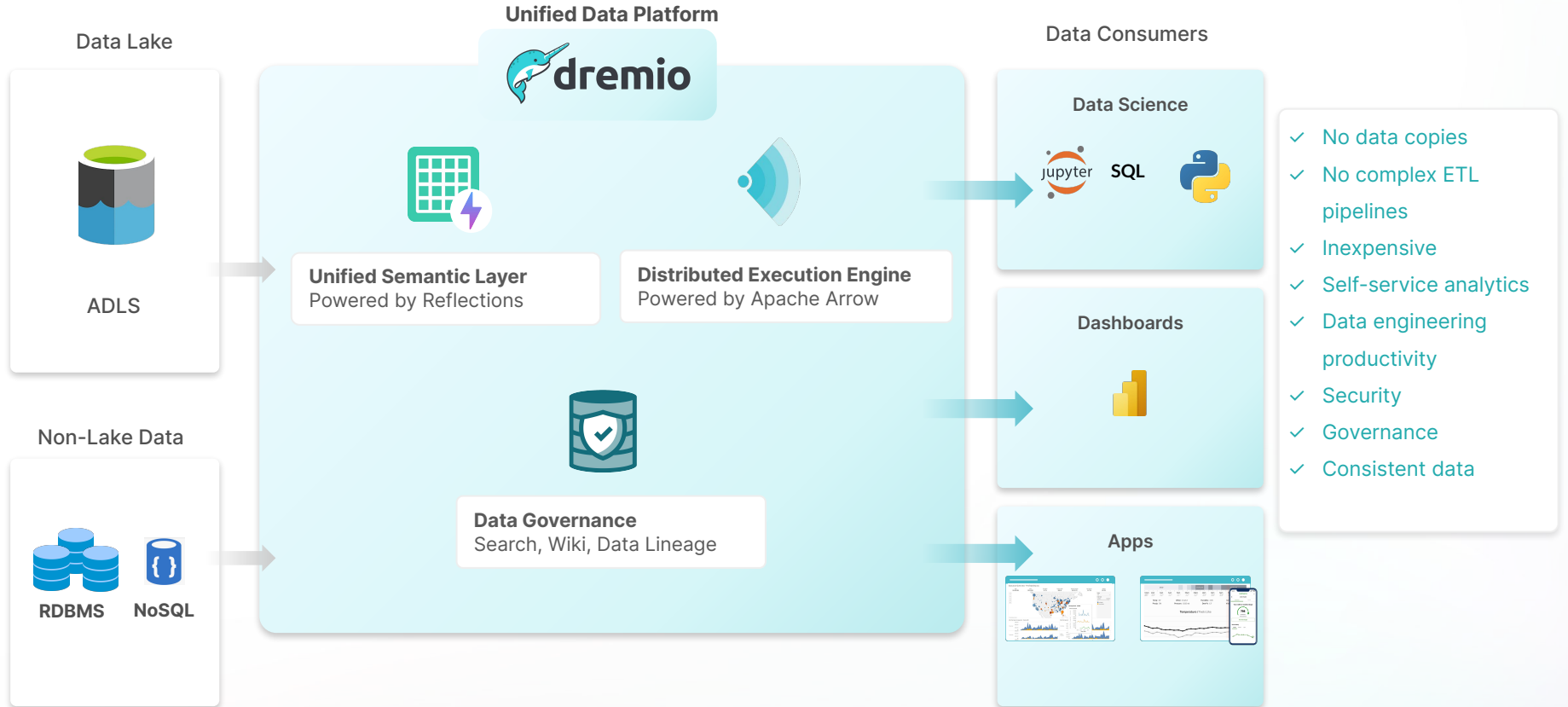
Intensity of Competing Data Priorities is Increasing



Traditional Analytics Architecture



Data Lakehouse on ADLS with Dremio



Growth in Digital & AI Transformation

Data Analytics Modernization

Make your current system work better

Data Mesh

Migrate Hadoop

Modern Data
Virtualization

Migrate Data
Warehouses to
Lakehouses

Cloud Data Lake to
Cloud Data Lakehouse

New Projects

Build new analytics capabilities

Departmental
Lakehouses

- Customer 360 / CDP
- Supply Chain
- Trading Data for Quants
- Product Analytics

Data Science

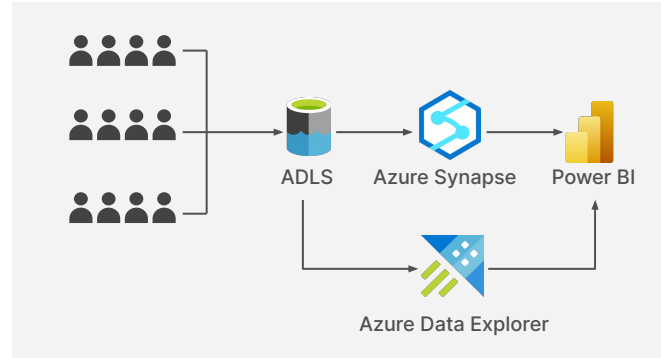
Internal Analytics
Apps / Custom
Dashboards

Customer-Facing
Analytics Apps

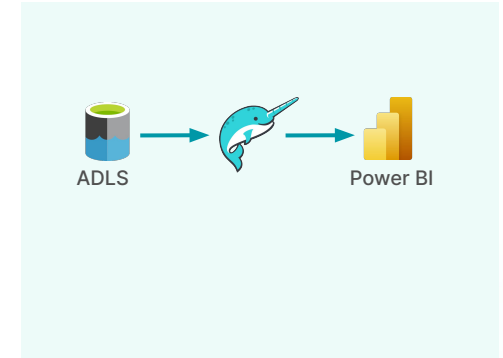
Delivering Product Analytics for Customer 360 on ADLS

Top Fortune 15 Company

Before Dremio



After Dremio



Business Problem

- Needed to analyze product usage across portfolio cloud offerings
- **50 TB of data today, growing 5 TB per day**
- Lots of ETL pipelines and Spark jobs, Presto not an effective solution
- Internal tools required data movement
- Adding more data sources means adding data engineers
- **Data pipelines were hard to create, and even harder to maintain**

Why Dremio?

- Rich semantic layer
- Run fast BI queries directly on ADLS without copying into data warehouse and BI cubes
- Support Parquet and Delta Lake tables

Results

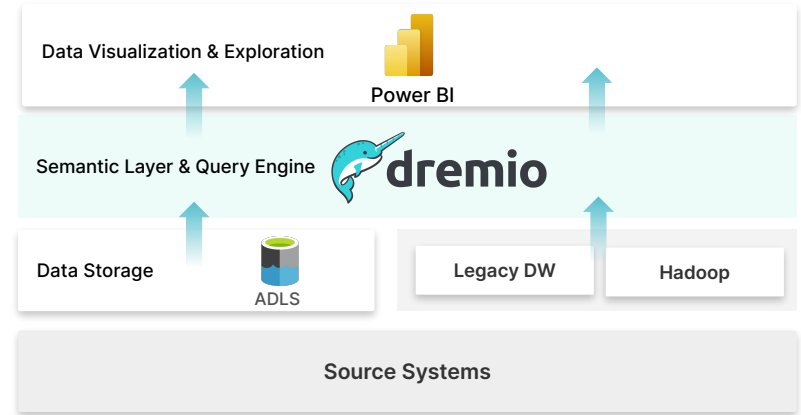
Self-Service Access

- Analysts now have self-service access to data at interactive speed using Power BI directly on ADLS
- Dynamic security controls govern access and maintain data lineage

Reduced Data Engineering Workload

- Reduced complexity and costs associated with building and maintaining ETL pipelines
- Only need single copy of data

Dashboards Running Up To 30x Faster



Business Problem	Why Dremio?	Results
<ul style="list-style-type: none">▪ Support the business's ability to cross-sell, up-sell, and service their customer base▪ Moving data pipelines took 2-3 months for critical and large datasets▪ Slow analytics development due to functional silos created among experts in different data repositories▪ Long turnaround time for data requests	<ul style="list-style-type: none">▪ Self-service data analytics▪ Modernize data infrastructure on data lake▪ Cost-effective solution that replaces expensive on-prem DW▪ Immediate performance gains on Hadoop	<p>Cost reduction</p> <ul style="list-style-type: none">▪ Reduced cost & dependency on external data engineering consultants▪ Retire EDW in 2 years and decommissioning Hadoop infrastructure <p>Faster time-to-insight</p> <ul style="list-style-type: none">▪ Minimize "revenue leakage" by not having to wait to run analyses



Demo

GNARLY Data_Waves

PRESENTED BY  **dremio**



Thank
you!