GNARLY Data_Waves

PRESENTED BY dremio

Episode 37

How NetApp is Redefining the Customer Experience with Product Analytics

Aaron Sims

Technical Director, NetApp

- 26+ years of systems engineering experience
- Chief Architect for NetApp Active IQ, NetApp's product telemetry and analytics platform
- Responsible for the architecture of Active IQ's data processing platform, data lake, databases, API platforms, and front ends

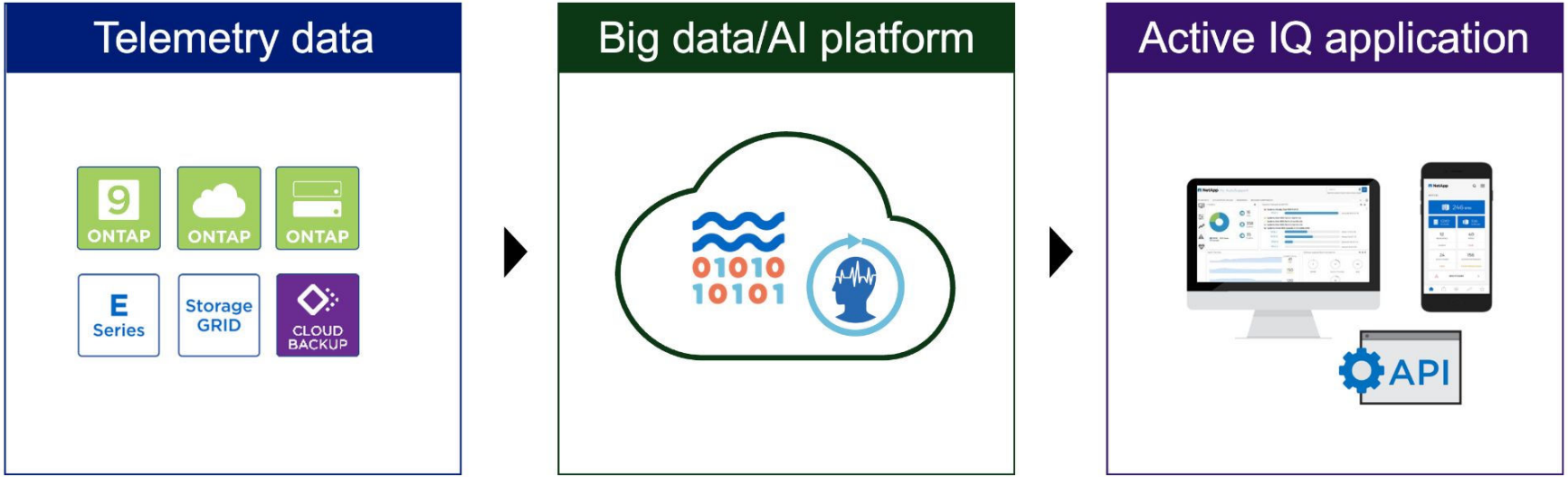
About NetApp

- The NetApp portfolio of leading data, application, and storage solutions helps organizations manage applications and data everywhere across hybrid multi-cloud environments.

- Over 30 years of innovation
  - Leader in all-flash storage
  - The only enterprise-grade storage OS available natively on the world’s biggest public clouds

- Best of class security and protection

Active IQ: Digital Advisory for Predictive Maintenance and Optimization



10 trillion data points per month

3 PB data lake

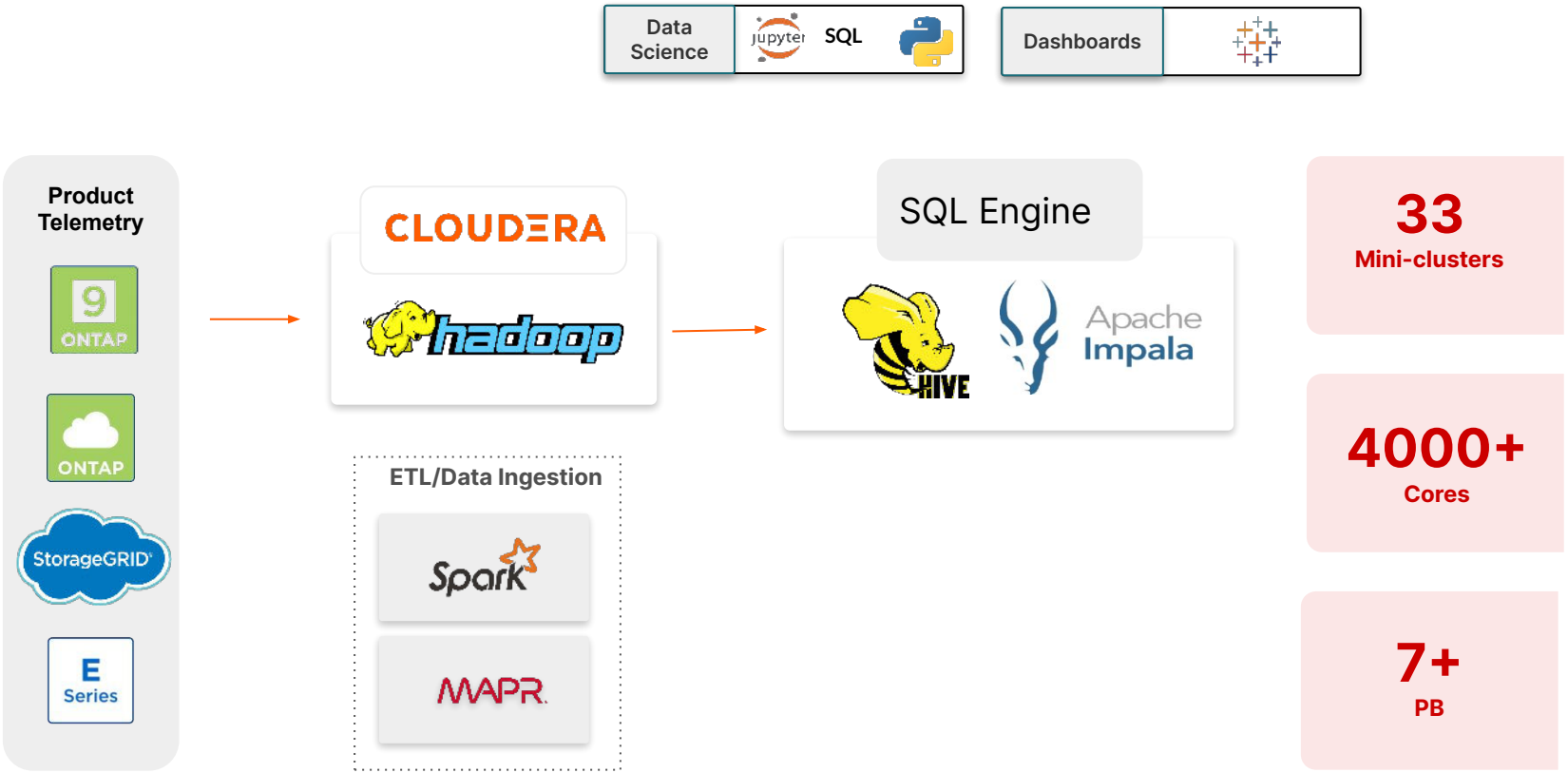
10PB shared storage

Insights and recommendations are delivered through web UI, mobile app, and APIs

Challenges

- Storage and compute tightly coupled
- Poor performance
- Operational controls
- Inefficient storage paradigm
- Data governance

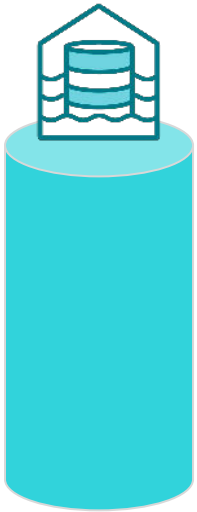
Architecture Before Dremio



Solution Considerations

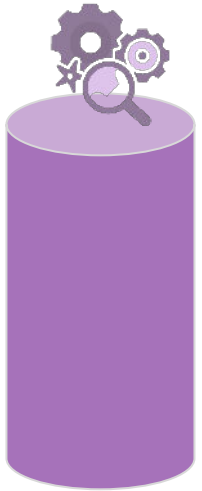
- Decouple storage and compute
- Reuse existing investment of compute and storage
- Minimize change to existing data pipelines
- Performance
- Easier Data Management
- Better Resource Management
- Disaster Recovery

Why Dremio



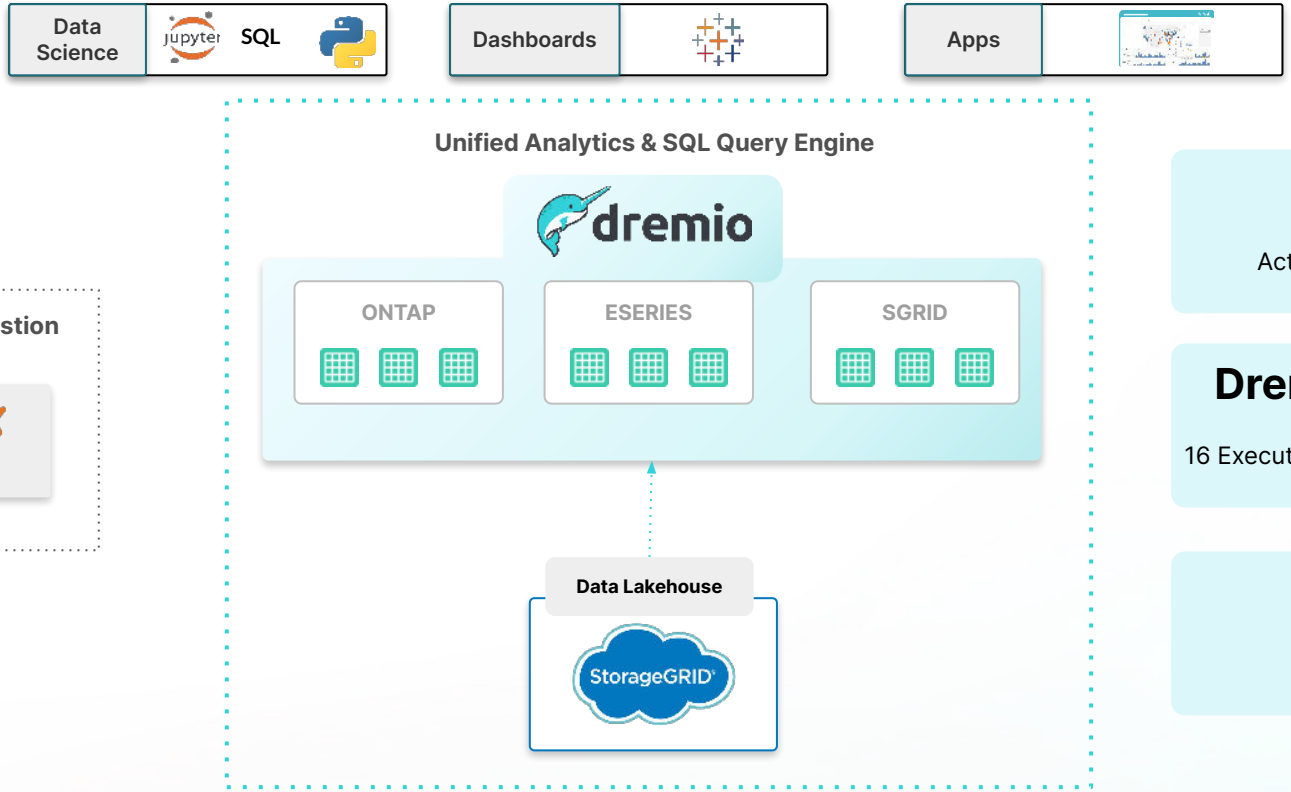
Lakehouse Query Engine

Semantic Layer



Performance

Architecture After Dremio



3 PB

Active IQ Data Lake

Dremio on K8S

16 Executor nodes on k8s cluster

8900

Tables

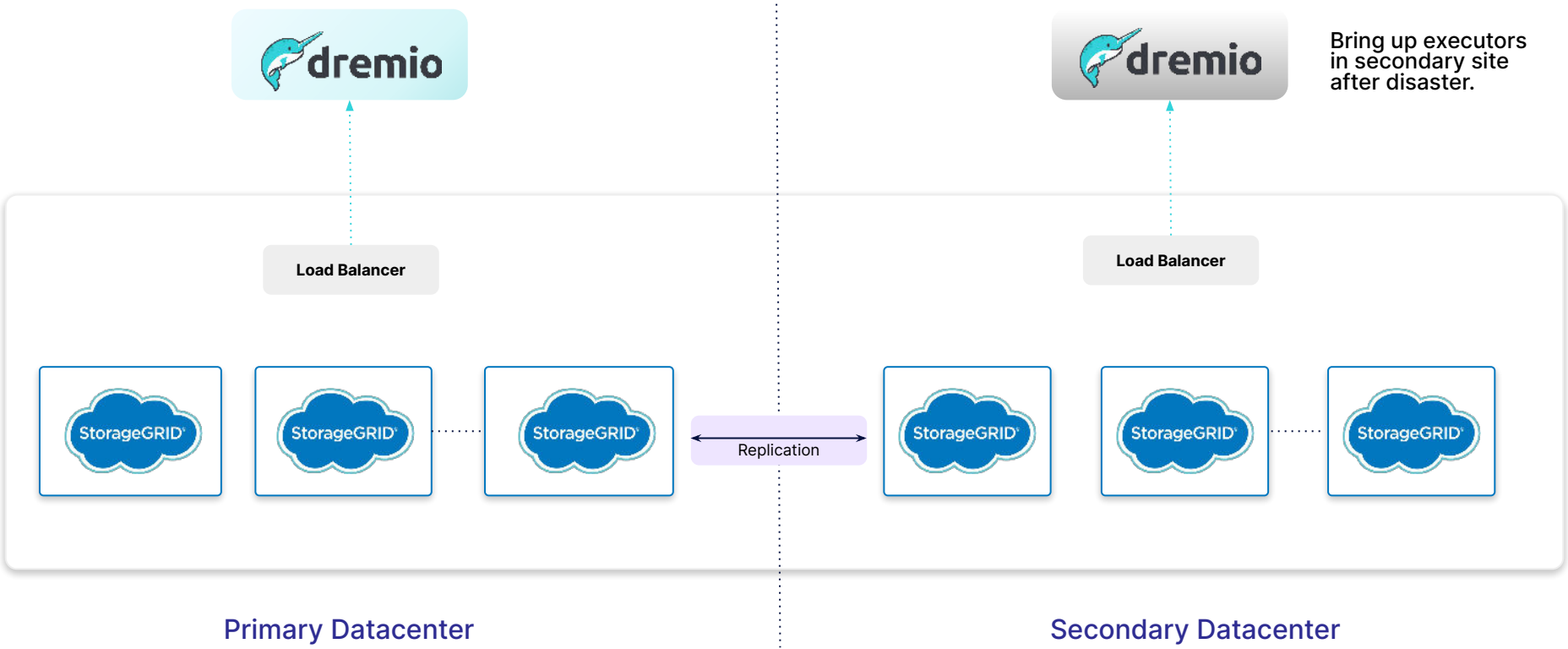
Migration Journey

- Integrate into existing data flows
- Research into current user queries
- Prepare users for migration, assist with query transformation
- Cutover users in groups each week.
- 130+ Users migrated in less than 2 months

StorageGRID ILM Makes Data Management Easy

- Erasure coding for Storage Efficiency and Performance
  - ILM policy for 4+2 Erasure Coding
- Replication for Disaster Recovery
- Retention policies allow for automatically removing old data

Disaster Recovery



Business Outcomes

Reduction in Compute

Faster Query Time

Analytics TCO Savings

Active Users

Confidential - Do Not Share or Distribute

What's Next?

Data Tiering

- Hot Data in Primary Database
- Cold Data in Dremio
- APIs query one or other based on date

Reduce / Eliminate ETLs

- Current pipelines make lots of copies of data for specific use cases
- Experiment with using Dremio as data gateway to authoritative data sources

Thank you!