GNARLY Data_Waves

PRESENTED BY dremio

EPISODE 44

How S&P Global is Building an Azure Data Lakehouse with Dremio

Tian de Klerk
Director of Business Intelligence, S&P Global

Tony Truong
Sr. Product Marketing Manager, Dremio

January 30 at 8AM PST | 11AM EST | 4PM GMT

dremio | S&P Global

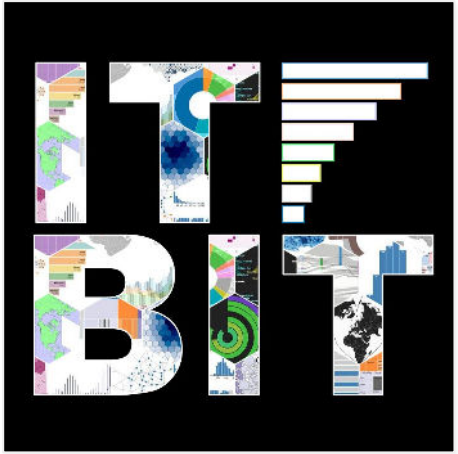
About Me

- From South Africa - Living in the Netherlands
- Background in Cloud Reporting assisting customers in cloud financial allocations and savings
- Started working with data about 7 years of Power BI experience
- Have been building and working with ETL and data pipelines for nearly as long
- Brought over to IHS Markit now S&P Global in 2020 to lead the IT business intelligence team

GNARLY Data_Waves

IT Business Intelligence Team

- Team is responsible for IT reporting to Corporate as well as to divisions
  - Service Management
  - Cloud Financial
  - Asset Inventory
  - And more
- Our developers connect to and import data from source systems for example
  - AWS (Inventory and Billing)
  - CMDB (Core for business mapping)
- Report primarily through Power BI, but we also deliver data to teams



GNARLY Data_Waves

The Business

Optimize Data Lake

- Needed to improve and optimize existing data lake architecture
- Need to move modeling out of PowerBI so it is reusable

Financial Operations

- Improve financial operations analysis on cloud services and ServiceNow data, including trend analysis and usage optimization

Improve Analytics TCO

- Division initiative to decrease costs and improve data access efficiency in analytics infrastructure

Data Lake Challenges

Rising Cost & Complexity

CosmosDB used traditional data engineering processes, increasing number of complex ETL pipelines for data extractions

Lack of Self-Service Analytics

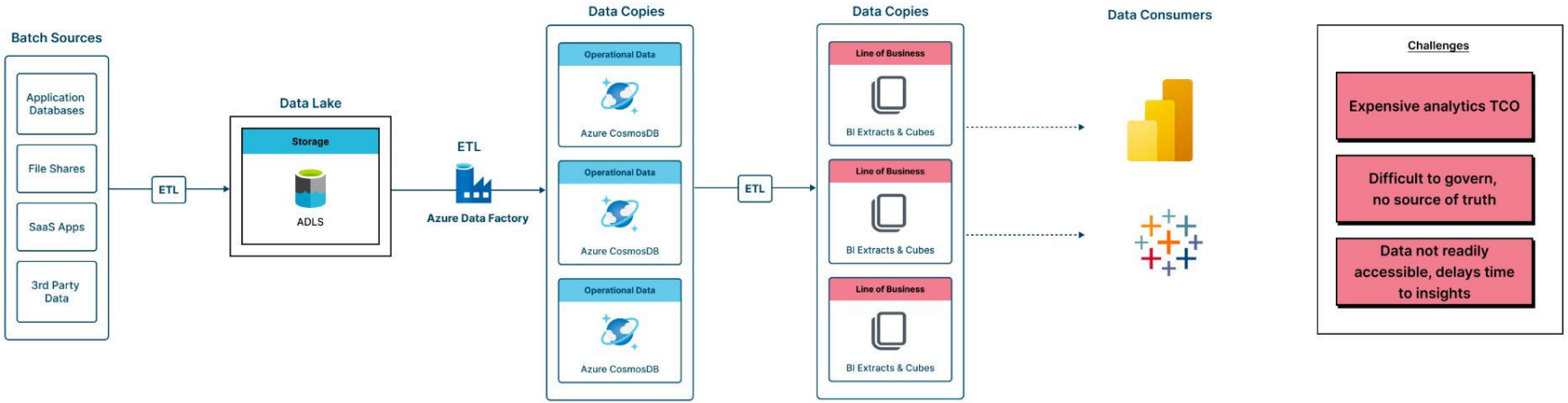
Reliance on BI extracts rather than direct query access for data exploration and analysis

Siloed Data Architecture

Placed heavy burden on data engineering team, required lots of manual work to maintain data consistency and governance

Challenges with existing architecture

FinOps Cloud Data Analytics



Why Dremio?

Semantic Layer - Ease of Use

Intuitive UI, non-technical users can access the data and self-service without engineering overhead

Fast Performance & Low Compute Costs

High performance query engine over ADLS and more cost effective than current approach

Data Security

Governed data access with role-based security. Data is more accessible via ODBC, Power BI, and from the portal

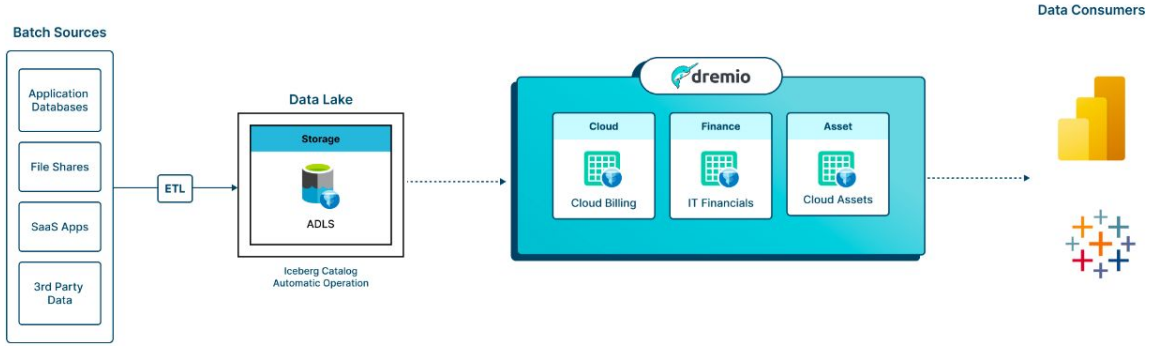
No More Data Copies

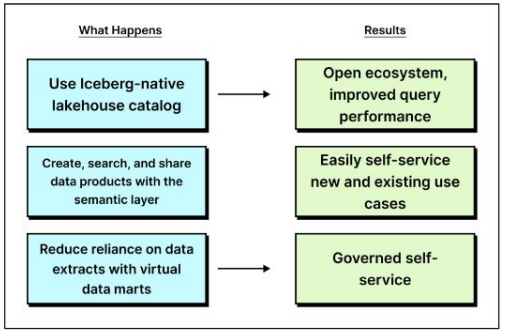
Virtual data marts with the semantic layer and replaced extracts with Reflections (Dremio's query acceleration)

GNARLY Data_Waves

Architecture After Dremio

Open Data Lakehouse for FinOps Analytics





Business Outcomes

Analytics TCO Savings in Year 1

To live data in ADLS from Power BI

Faster Query Time

What's Next?

Centralize IT - Dremio enables us to provide a single platform to query the data

Robust access controls and row level security enable us to share data to the right people

Ability to route queries to engines, so we can allocate costs

ICEBERG

Git for Data - Use Dremio for lakehouse management and version our data like a product

GNARLY Data_Waves

Thank you!