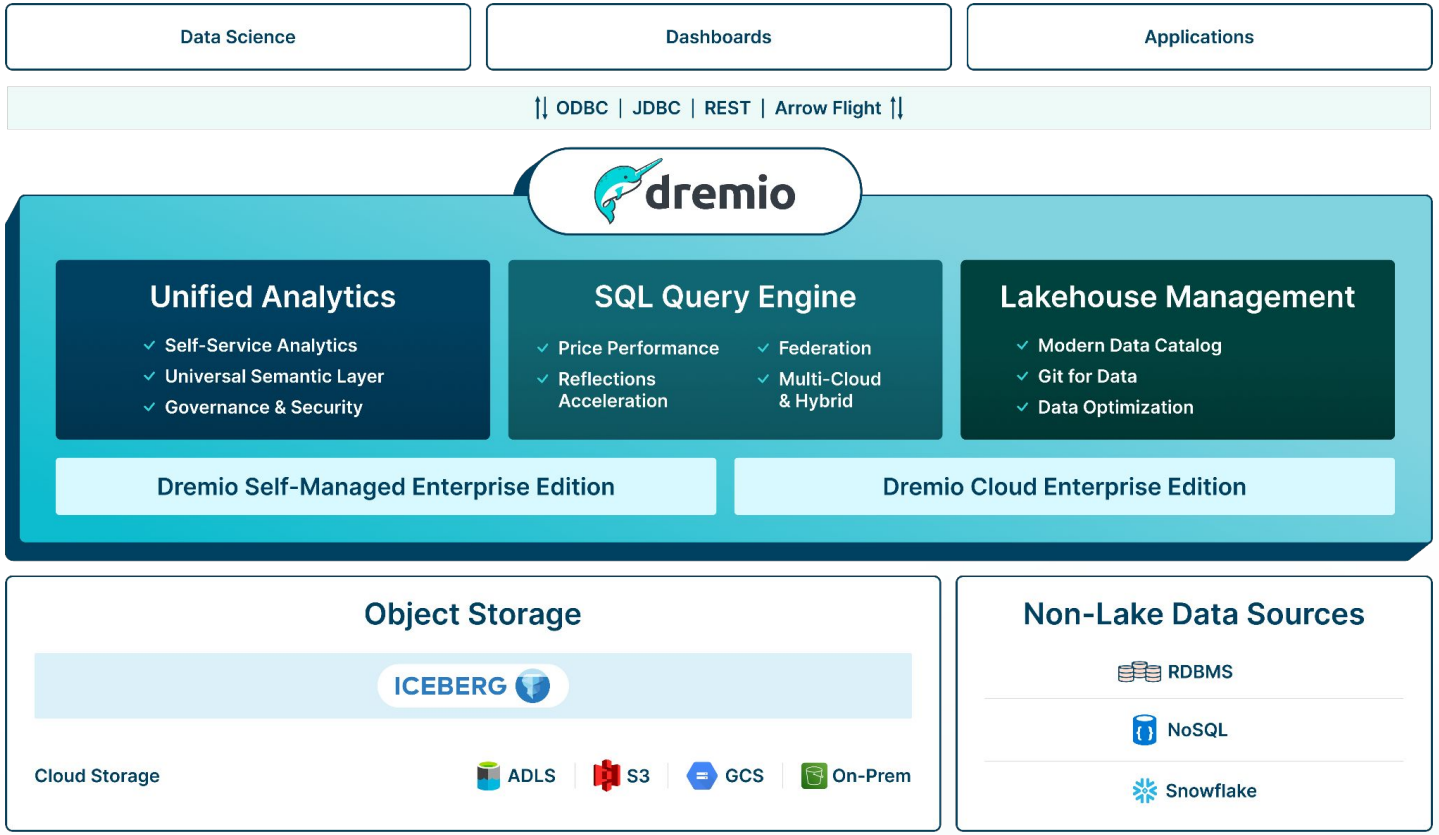
GNARLY Data_Waves

PRESENTED BY dremio

EPISODE 50

How to Optimize your Analytics with Apache Iceberg, Dremio, Snowflake and the Data Lakehouse

Unified Lakehouse Platform for Self-Service Analytics



A Iceberg/Dremio Lakehouse on your laptop exercise

Deploy Dremio Software or Dremio Cloud

Hands-On

Postgres -> Iceberg -> Dashboard

SQLServer -> Iceberg -> Dashboard

MongoDB -> Iceberg -> Dashboard

dremio.com/blog

dremio
Unlock efficiency and savings on your analytics with Dremio when using Snowflake
DREMIO + SNOWFLAKE: ICEBERG-CENTRIC ANALYTICS & AI
June 5th | 9:00am - 5:30pm
Dremio Chill Lounge @ 221 4th St, San Francisco, CA 94103

GNARLY Data_Waves

PRESENTED BY dremio

EPISODE 50

How to Optimize your Analytics Spend with Apache Iceberg, Dremio, Snowflake and the Data Lakehouse

What’s The Problem?

Broken Pipelines that require tedious backfilling

Angry Consumers from Late and Inconsistent Data

Cost of these Movements

1. Storage Costs
2. Compute/Processing Cost
3. Network & Egress Costs
4. Lost productivity in time it takes for all pipelines to proliferate data
5. Regulatory fees from governance and security risks in too many copies
6. Data Model Drift from as data models may become inconsistent over several movement
7. Cost of Bad Insights from Inconsistent Data from Data Copy Sprawl

Snowflake + Dremio Solution

#1 - Dashboard Optimization

Problem

Want to serve a dashboard that uses a lot of aggregate functions and GROUP BYs under the hood that can be expensive to run real-time as the board is accessed.

Possible Solution

We can create a new table of the pre-computed results using Snowflakes GROUP BY CUBE or GROUP BY ROLLUP to materialize a new table.

The creates a new namespace and you'll have to rebuild this materialization when the data changes.

Possible Solution

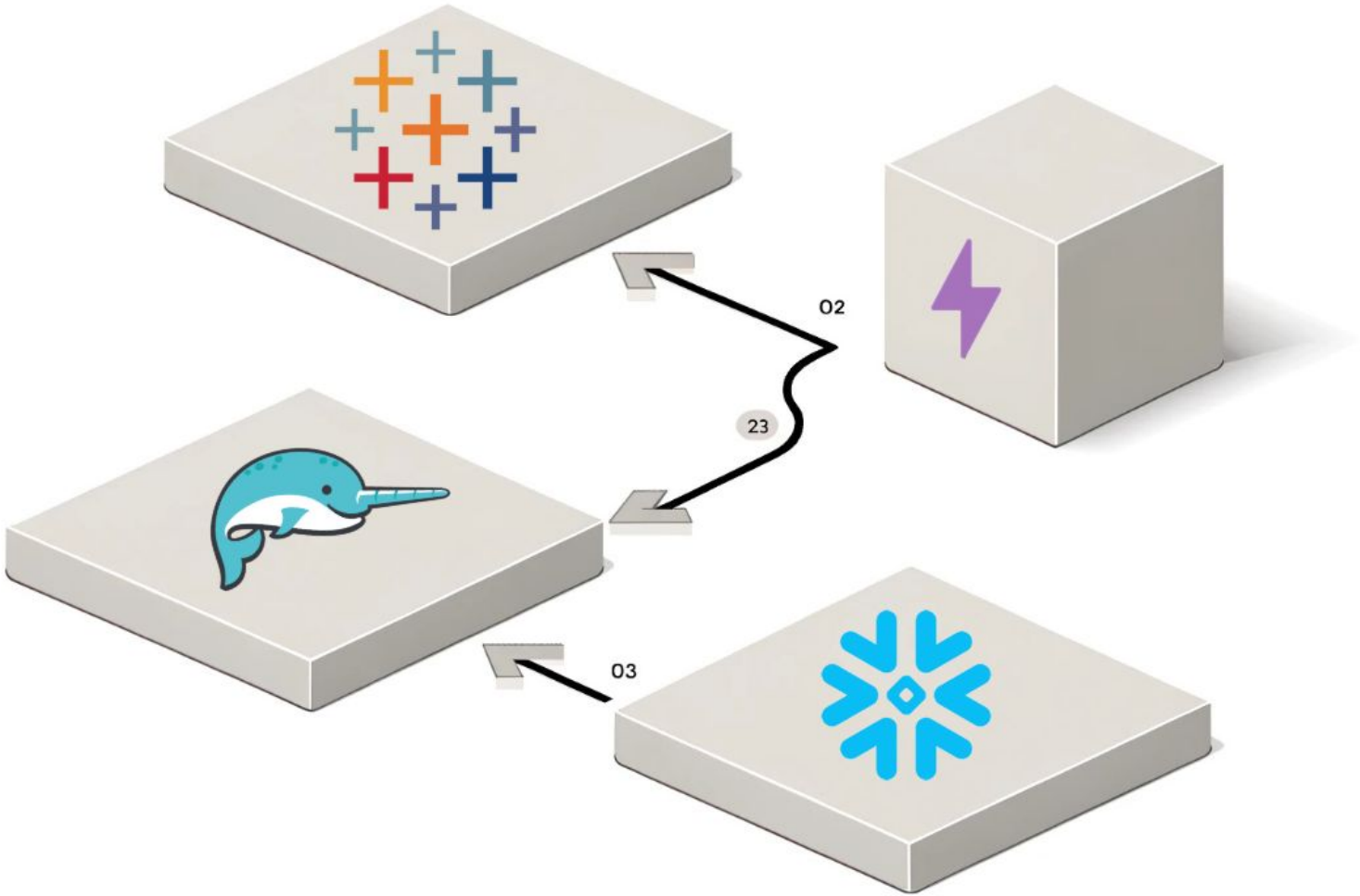
We can possibly use Snowflakes Dynamic Tables which will automatically update the result set periodically.

Better but still creates a new namespace and has limitation regarding joins and window functions that may not work for some use cases.

Possible Solution

Connect Dremio to your Snowflake account and create an aggregation reflection on the dataset. Serve the BI dashboard from live queries on Dremio

Reduce costs, no new namespace, automatically updates and works for all joins.


02
23
03

Snowflake + Dremio Solution

#2 - Data Unification

Problem

You have your own data in Snowflake you want to join with data you have elsewhere in a database, data lake or other data warehouse.

Possible Solution

You can copy the data into a table on Snowflake.

You will have to create a data pipeline to regularly ingest the data that can break, be expensive and possibly result in consistency issues.

Possible Solution

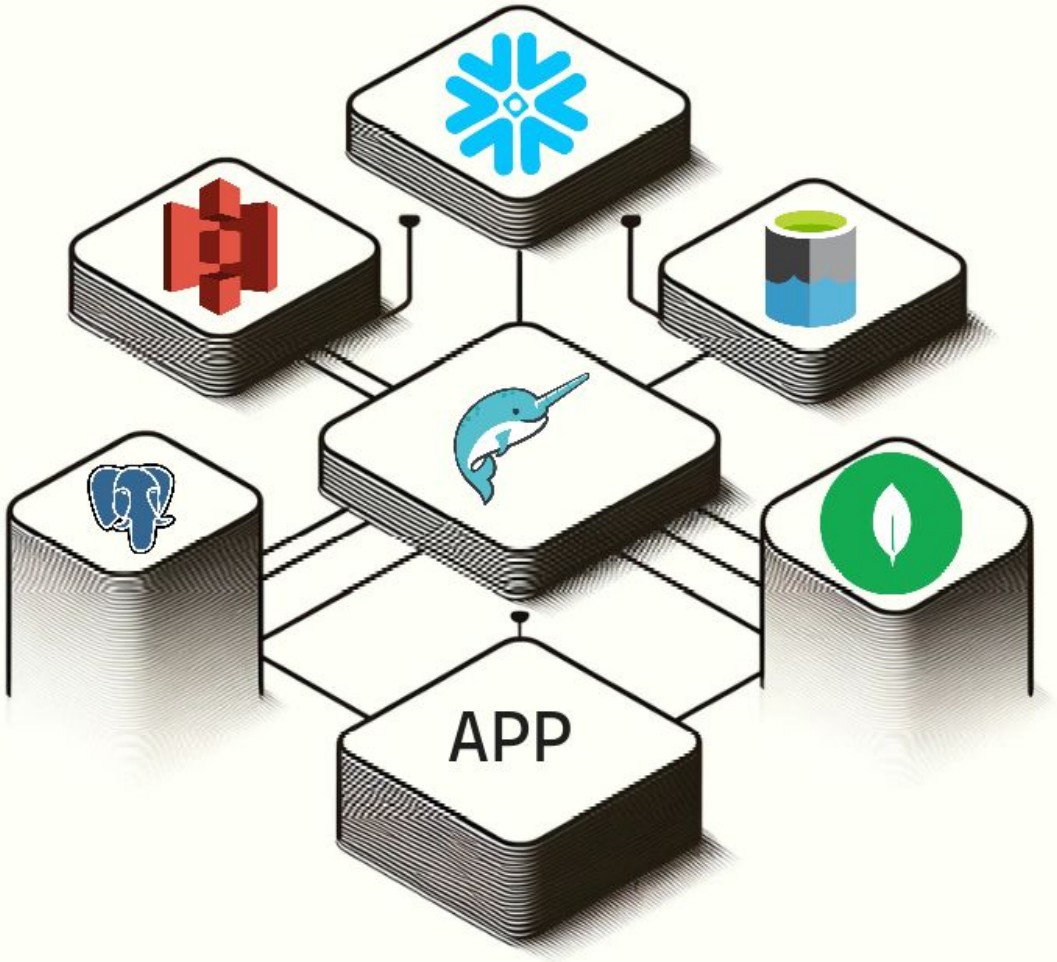
You can connect your Snowflake account and the other source to Dremio and join them from Dremio without the need to move data.

Can result in some egress expenses, turn on reflections on these sources and the egress costs go away too!

aws

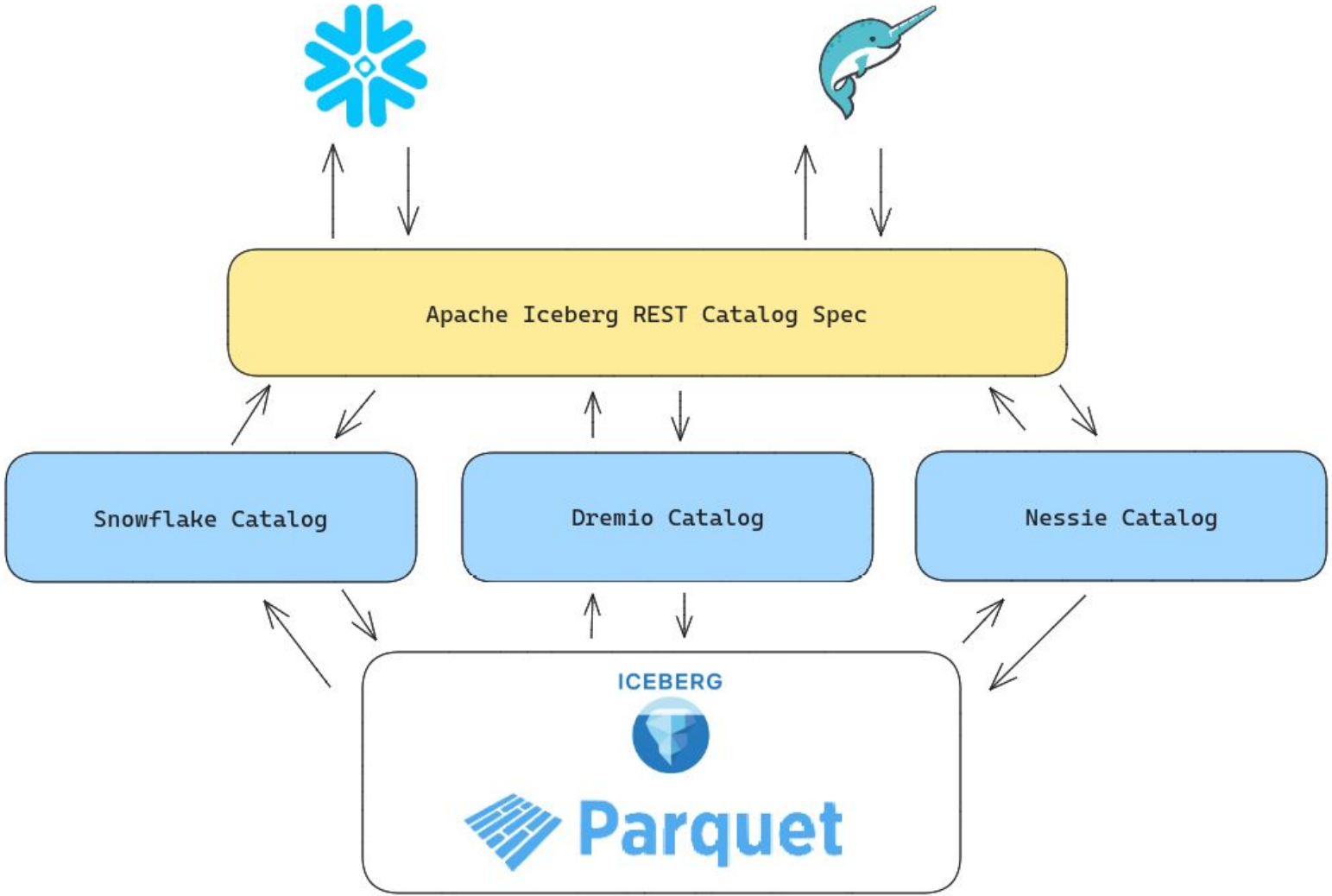
Snowflake + Dremio Solution

#3 - Simplifying Ingestion


APP

Snowflake + Dremio Solution

#4 - Single Source of Truth (future looking)


Apache Iceberg REST Catalog Spec
Snowflake Catalog
Dremio Catalog
Nessie Catalog
ICEBERG
Parquet

dremio
Unlock efficiency and savings on your analytics with Dremio when using Snowflake
DREMIO + SNOWFLAKE: ICEBERG-CENTRIC ANALYTICS & AI
June 5th | 9:00am - 5:30pm
Dremio Chill Lounge @ 221 4th St, San Francisco, CA 94103

Unifying Snowflake, Azure, AWS and Google Based

Data Marketplaces and Data Sharing with Dremio

Dremio Blog

A Iceberg/Dremio Lakehouse on your laptop exercise

Deploy Dremio Software or Dremio Cloud

Hands-On

Postgres -> Iceberg -> Dashboard

SQLServer -> Iceberg -> Dashboard

MongoDB -> Iceberg -> Dashboard

dremio.com/blog