GNARLY
Data_Waves
PRESENTED BY dremio

EPISODE 23

Getting Started with Dremio Data Reflections

Tiong Lee

Sr. Director, Dremio

- Senior Director of Eng, Dremio:
  - Query Planning and Reflection
- Previously:
  - Led data engineering groups at Samsung Electronic USA
  - Founder of Data Technology startup
  - Oracle
- Technologist, Entrepreneur and Engineering Leader, Coder

Today’s Agenda

Getting Started with Dremio Data Reflections

- Traditional Data Architecture
- Introduction to Open Data Lakehouse
- Dremio Reflections
- Demo!

Companies Want to Democratize Data... But How?

Data Providers

Data Lakes

S3

ADLS

GCS

hadoop

Databases

RDBMS

NoSQL

Data Consumers

Data Science

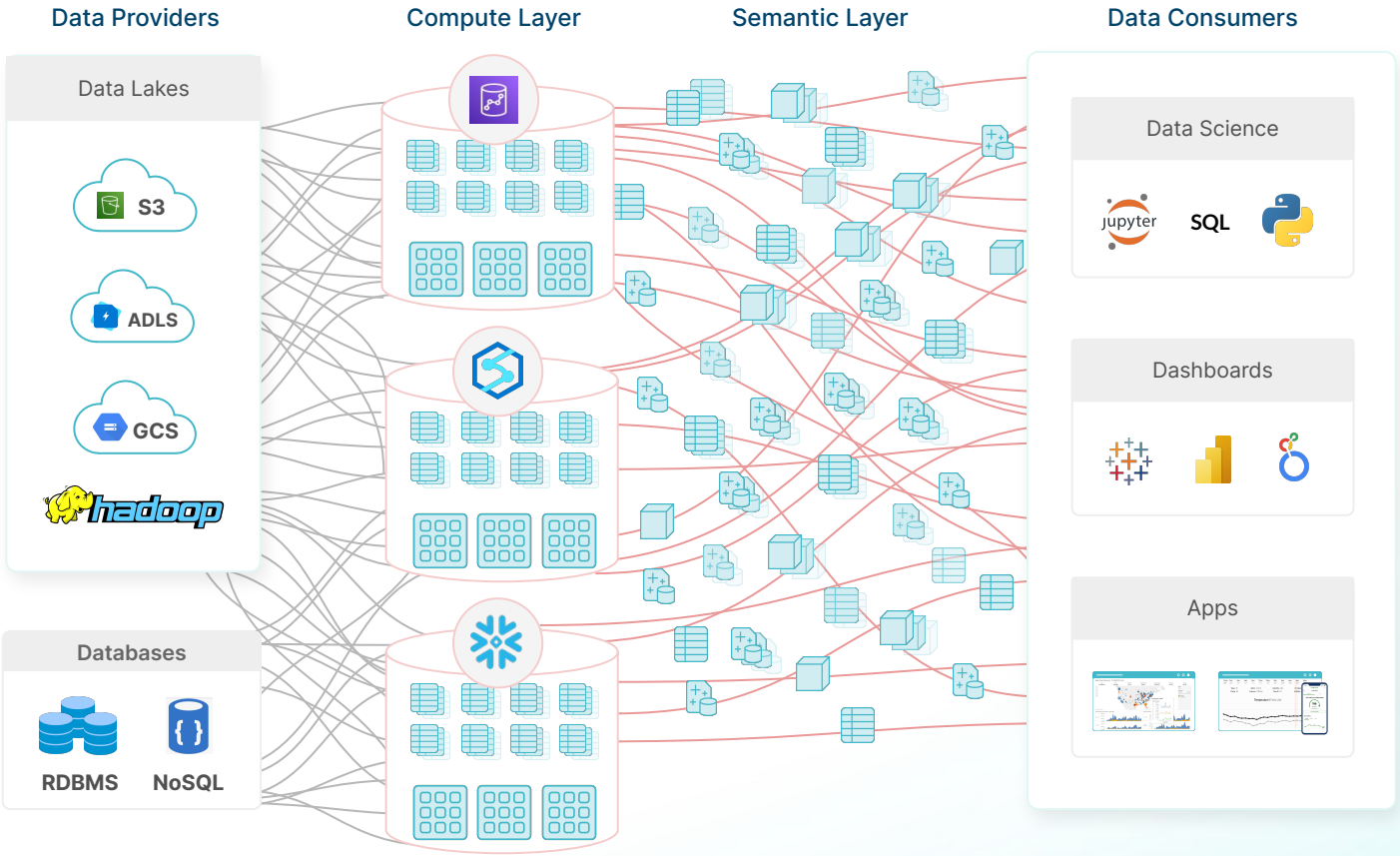
jupyter SQL

Dashboards

Apps

- Everyone wants access
- Data volumes are exploding
- Security risks
- Compliance requirements
- Limited resources

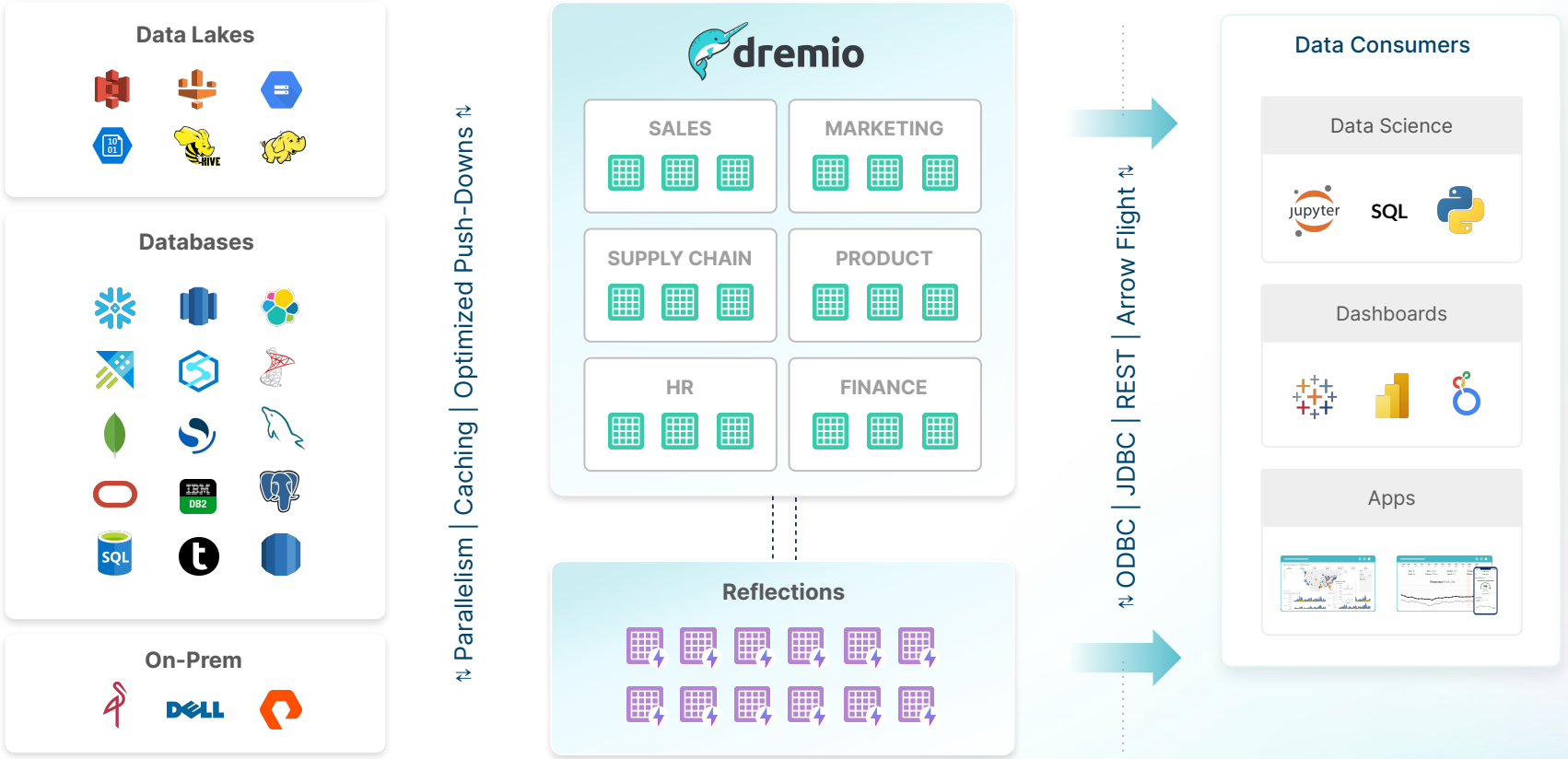
What are your options?



- × Complex
- × Expensive
- × Lock-in
- × Impossible to secure
- × No self-service
- × Limited data exploration
- × Inconsistent data

dremio

Dremio's Data Lakehouse is the Unified Access Layer

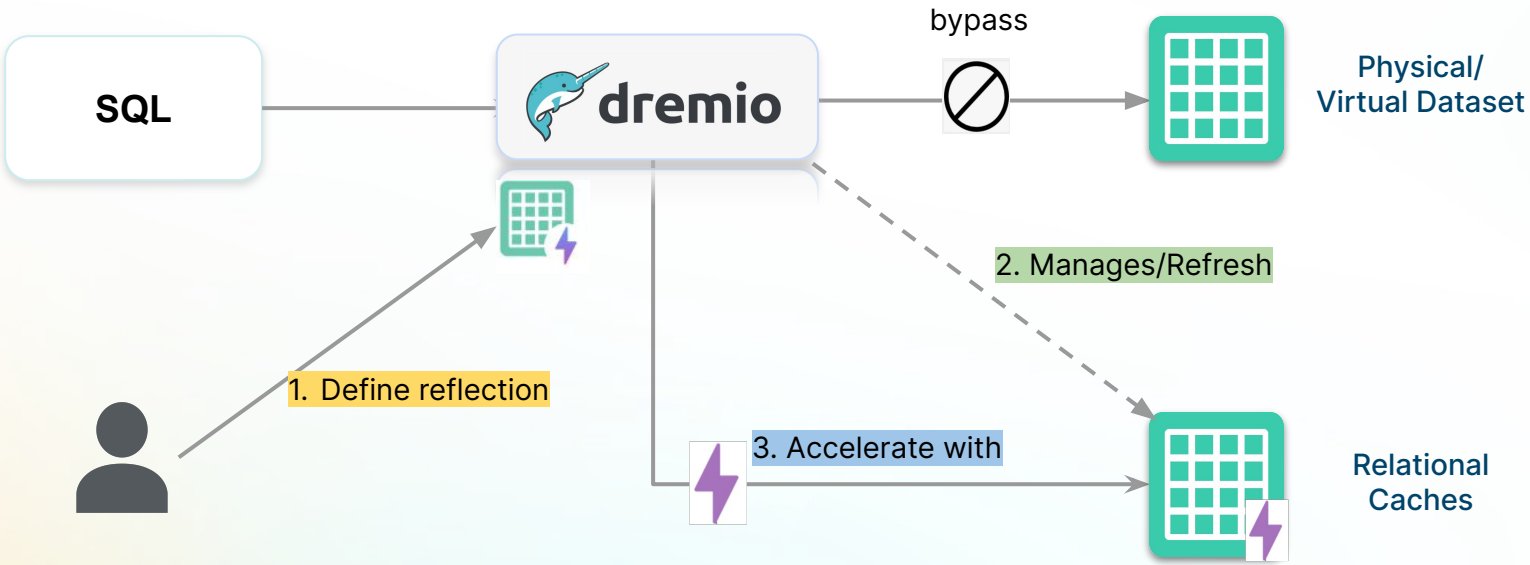


What is Data Reflection in Dremio

- “Enhance query performance using user-defined relational caches”
- Relational caches = materialized-view = query plan + materializations

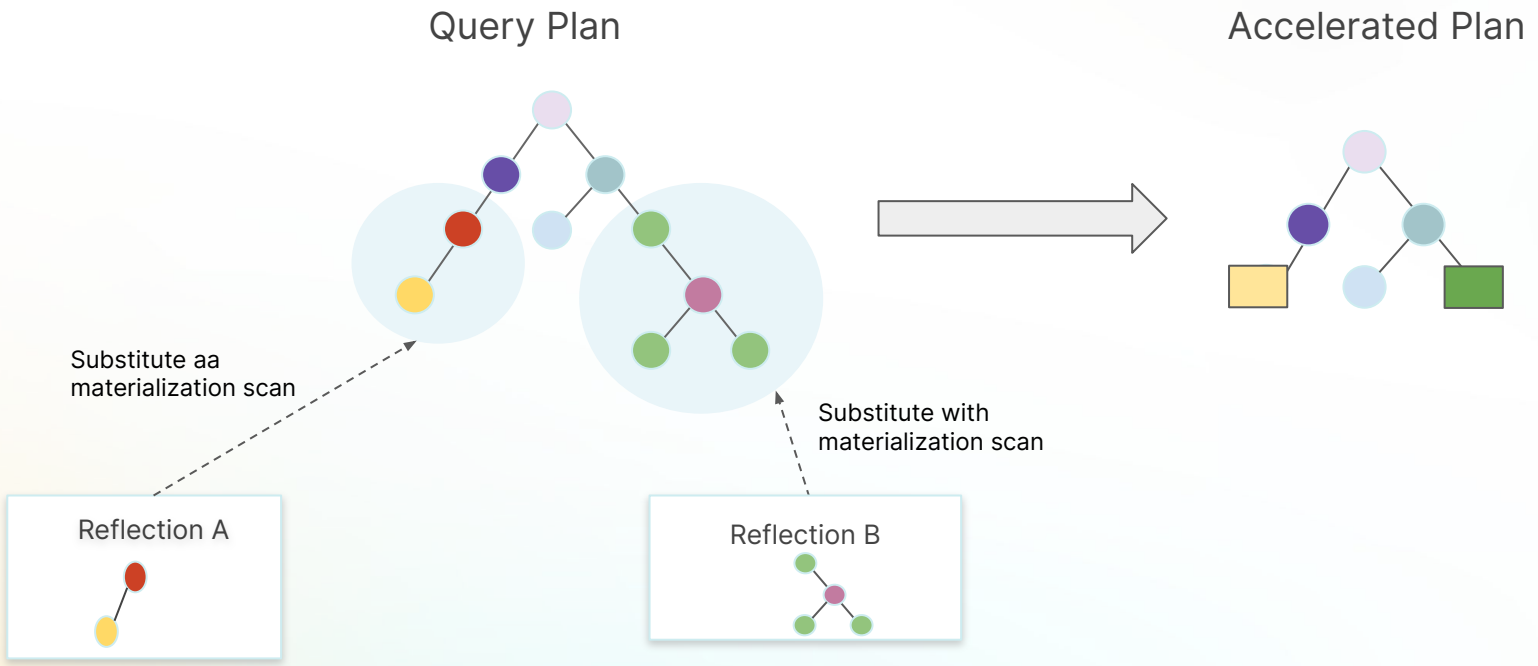
What is Data Reflection in Dremio

- “Enhance query performance using user-defined relational caches”
- Relational caches = materialized-view = query plan + materialization



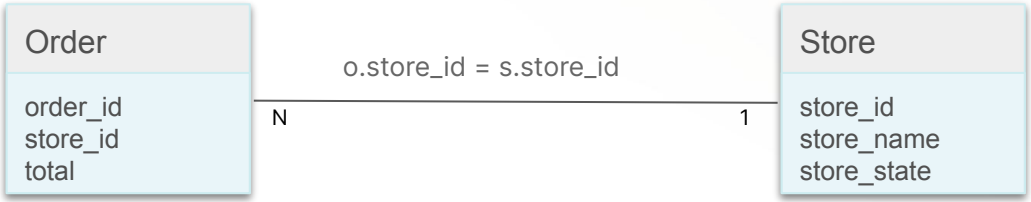
Query Plan Acceleration (Substitution) Process

Algebraic Matching



dremio

What can be accelerated



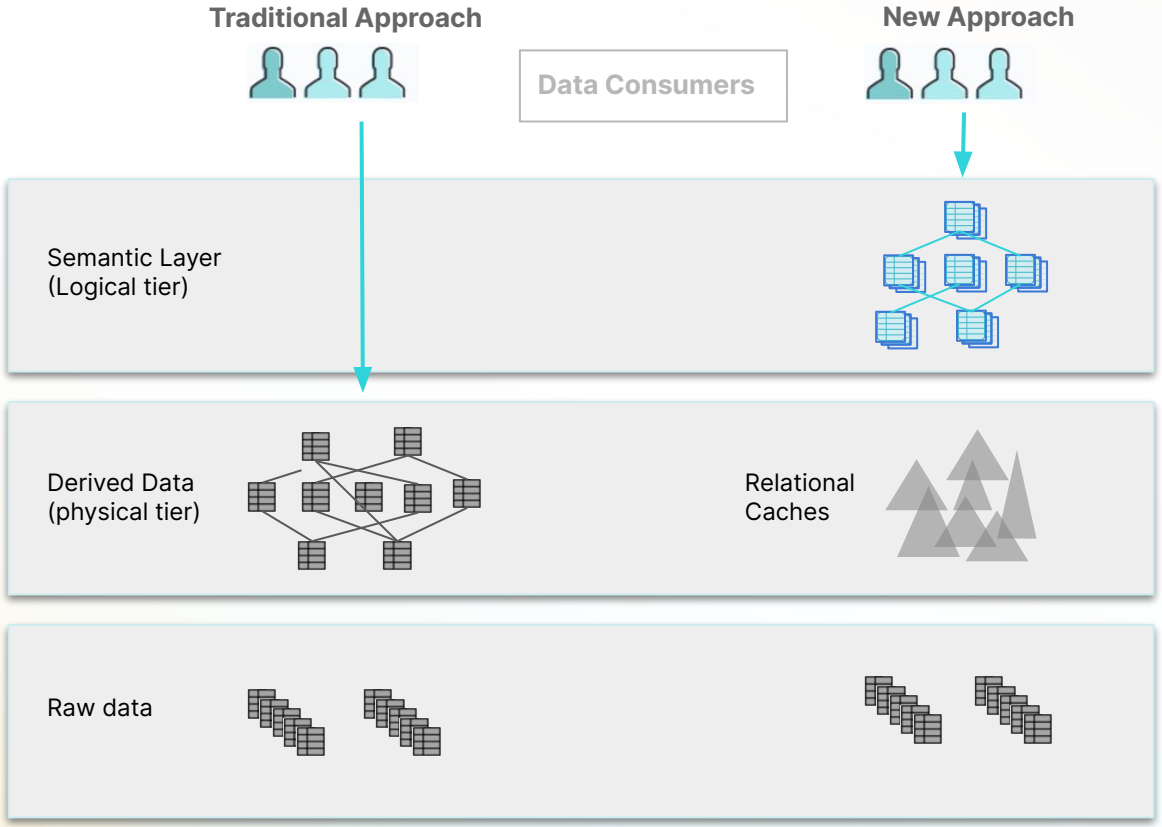
Reflection definition (total_by_store_id)

SELECT SUM(total) **FROM** orders o **GROUP BY** o.store_id - - on single table

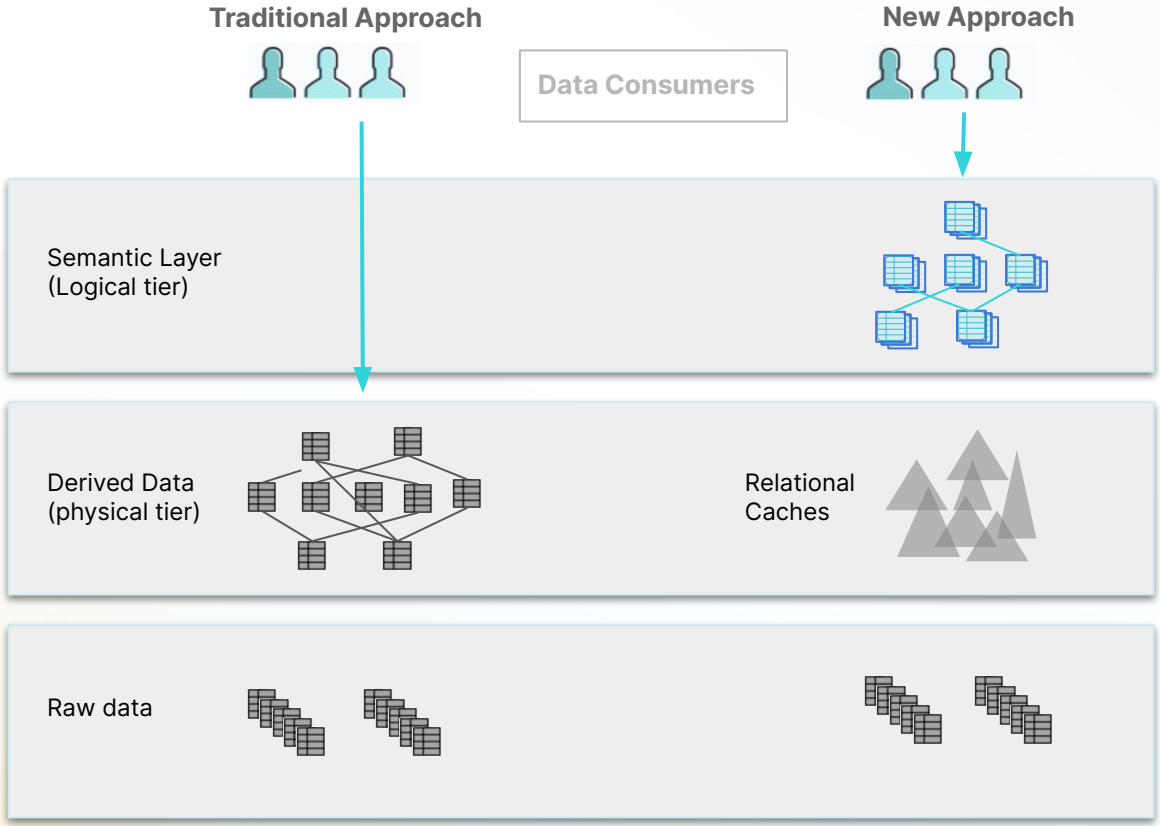
Example of queries that can be accelerated

1. **SELECT SUM**(total) **FROM** orders **GROUP BY** store_id - - identity query
2. **SELECT SUM**(total) **FROM** View1 **GROUP BY** store_name - - joinable
3. **SELECT SUM**(total) **FROM** View1 **GROUP BY** store_state - - joinable with further aggregation
4. **SELECT SUM**(total) **FROM** View1 - - singularity

Advantages of Data Reflection



Advantages of Data Reflection



1. **Simple**

   Usage and integration

2. **Flexible**

   Optimize anytime

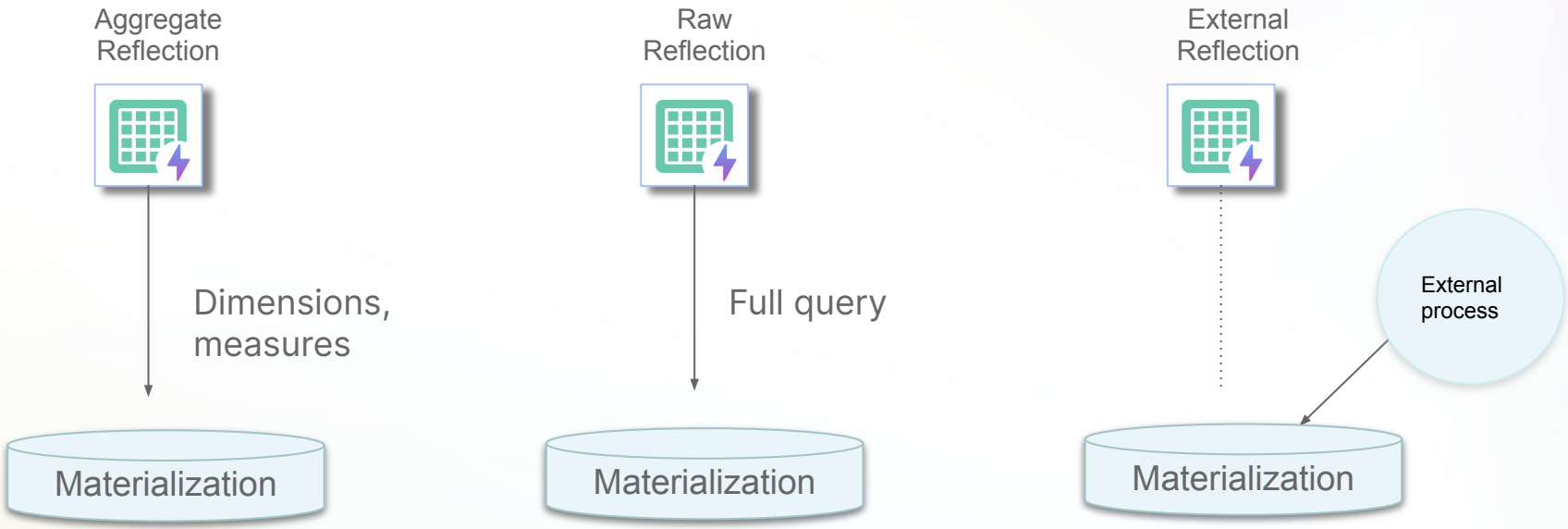
3. **Cost**

   Everyone can share relation caches

4. **Easy**

   Less dataset, less pipelines

Type of Reflections

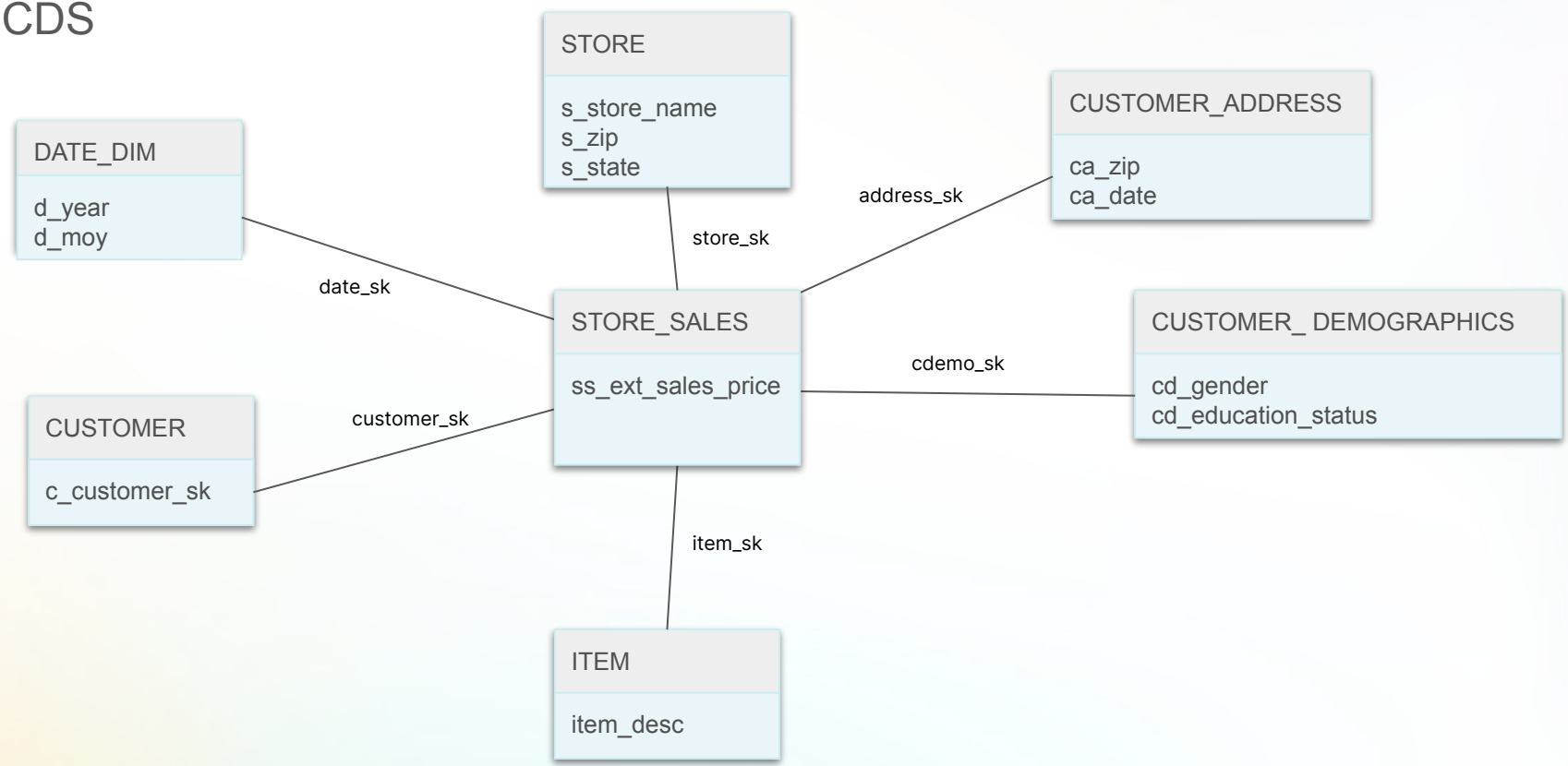


dremio

Demo

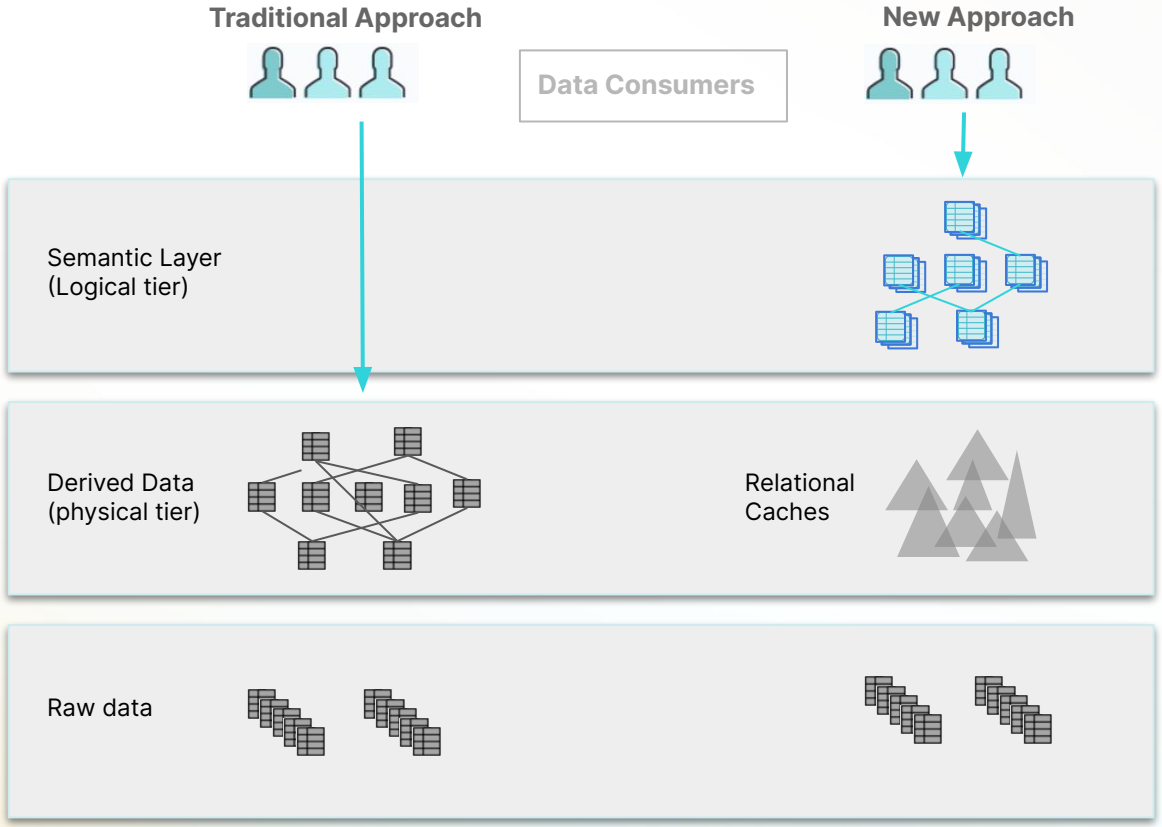
Meet the data

TPCDS



dremio

Advantages of Data Reflection



1. **Simple**

   Usage and integration

2. **Flexible**

   Optimize anytime

3. **Cost**

   Everyone can share relation caches

4. **Easy**

   Less dataset, less pipelines

dremio

Thank you!