GNARLY Data_Waves

PRESENTED BY dremio

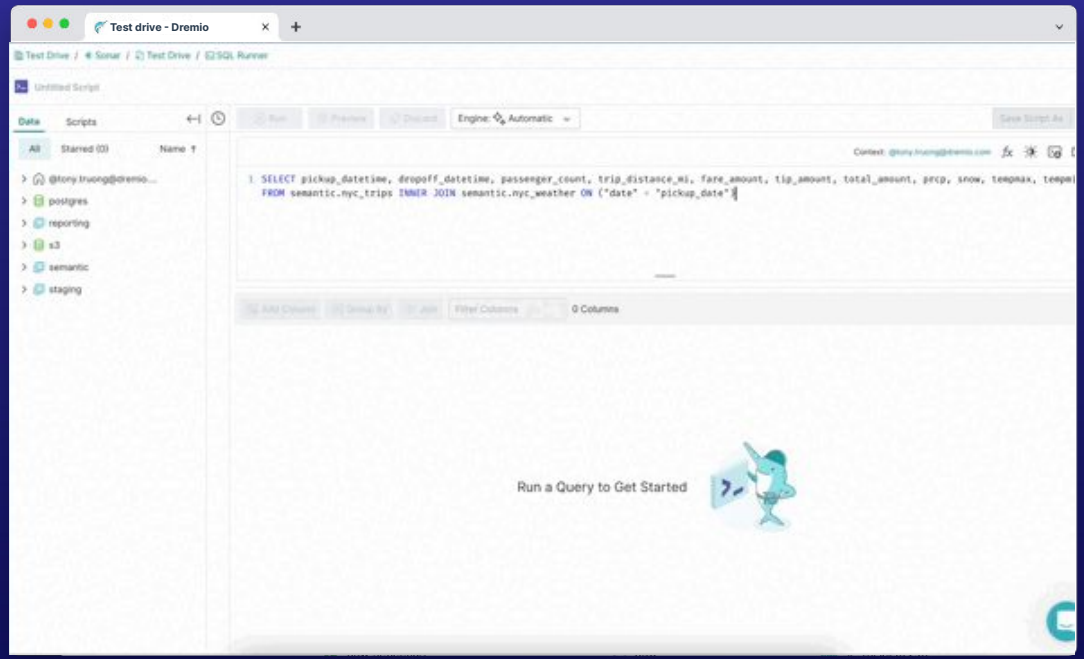
EPISODE 30

Using Dremio and DuckDB Together

Experience the data lakehouse with Dremio Test Drive

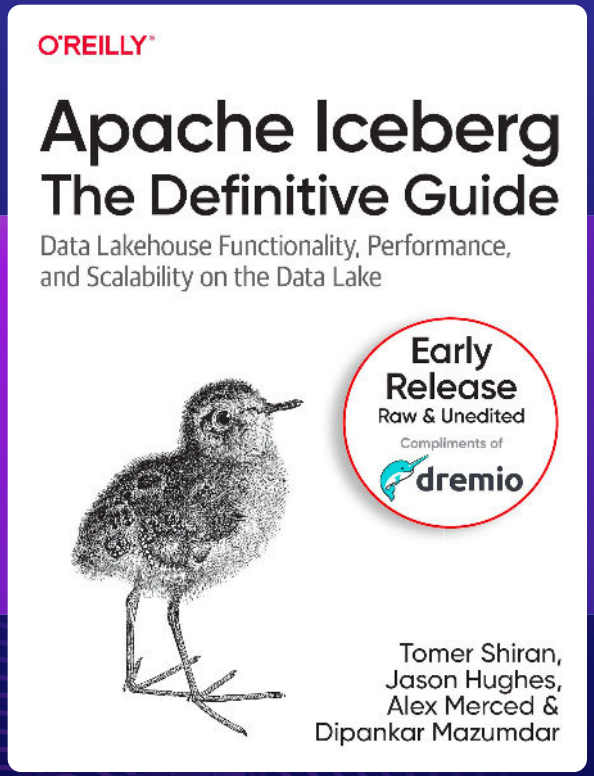
- Sub-second query on 1 billion rows of data joining Amazon S3 with a Postgres database
- Connect to Tableau or Power BI and build a dashboard with this dataset
- Everything hosted by Dremio - 100% free for you

Start Test Drive



GNARLY Data_Waves

Apache Iceberg: The Definitive Guide



Upcoming shows

Register now

EPISODE 28
Apache Iceberg Office Hours

Aug 8, 2023 at 8AM PST | 11AM EST | 4PM GMT

EPISODE 29
Simplifying Data Mesh with Dremio's Open Data Lakehouse

Aug 22, 2023 at 8AM PST | 11AM EST | 4PM GMT

EPISODE 30
Using Dremio and DuckDB together

Aug 29, 2023 at 8AM PST | 11AM EST | 4PM GMT

EPISODE 31
TBD

Sep 6, 2023 at 8AM PST | 11AM EST | 4PM GMT

EPISODE 32
It's time for an even easier data lakehouse: Learn what's new in Dremio

Sep 12, 2023 at 8AM PST | 11AM EST | 4PM GMT

EPISODE 33
ELT, ETL and the Dremio Data Lakehouse

Sep 19, 2023 at 8AM PST | 11AM EST | 4PM GMT

BIGDATA & AI by Xcorp PARIS

TIME TO ACCELERATE

September 25 & 26, 2023

Paris Convention Center

dremio

Coalesce by dbt

Oct 16-20, 2023
Hilton Bayfront San Diego

GNARLY Data_Waves

PRESENTED BY dremio

EPISODE 30

Using Dremio and DuckDB together

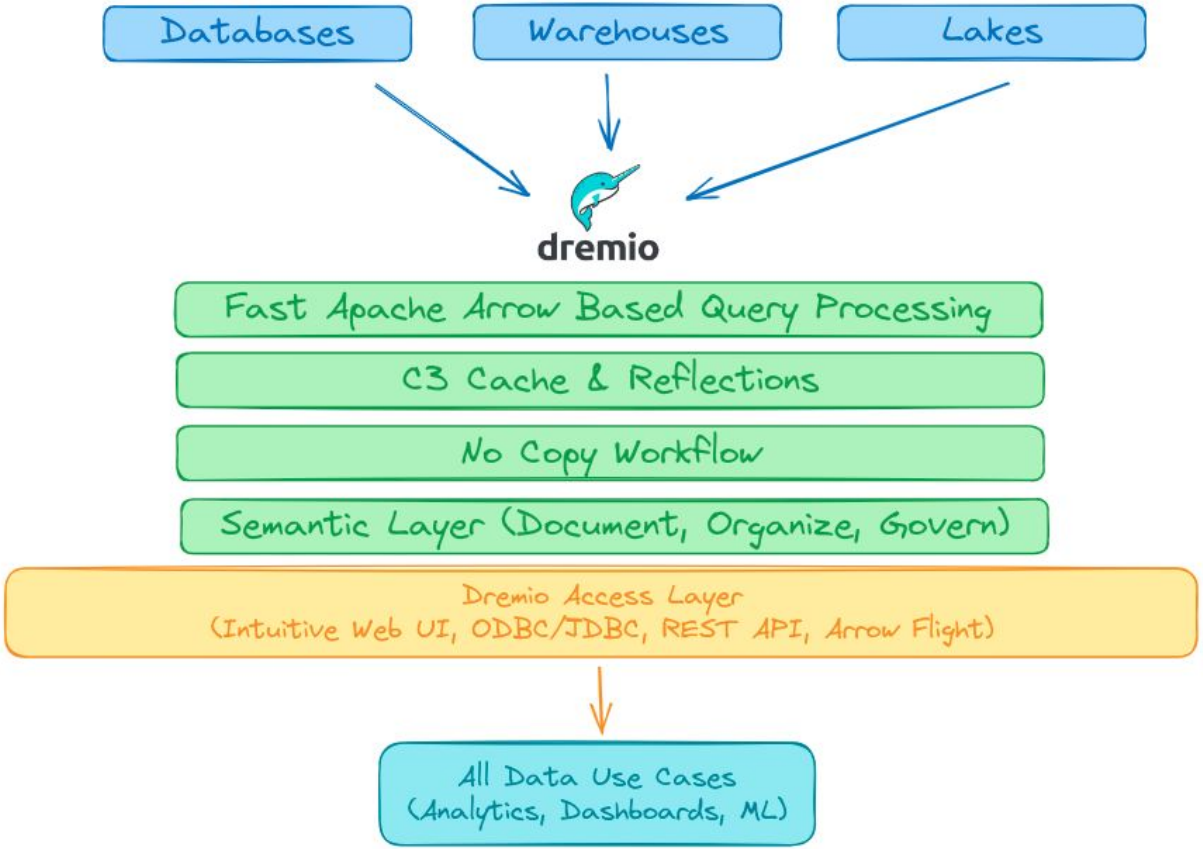
August 29 at 8AM PST | 11AM EST | 4PM GMT

Alex Merced
Developer Advocate, Dremio

WHat Good Data Architecture Should Provide...

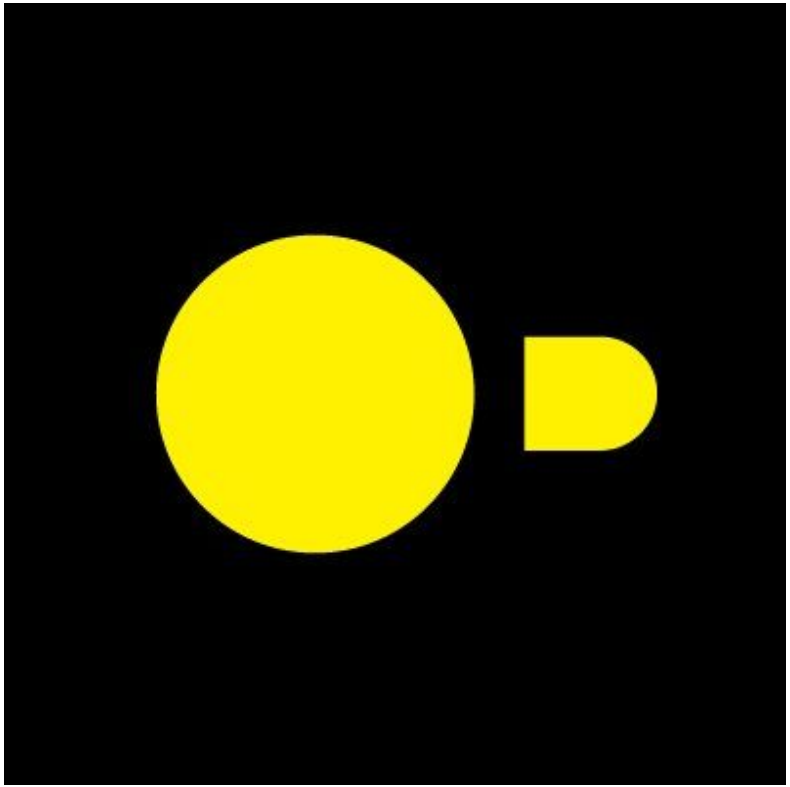
Better Performance
Lower Costs
Easier/Self-Service Usage

What is Dremio?
The East, Fast and Open Lakehouse Platform



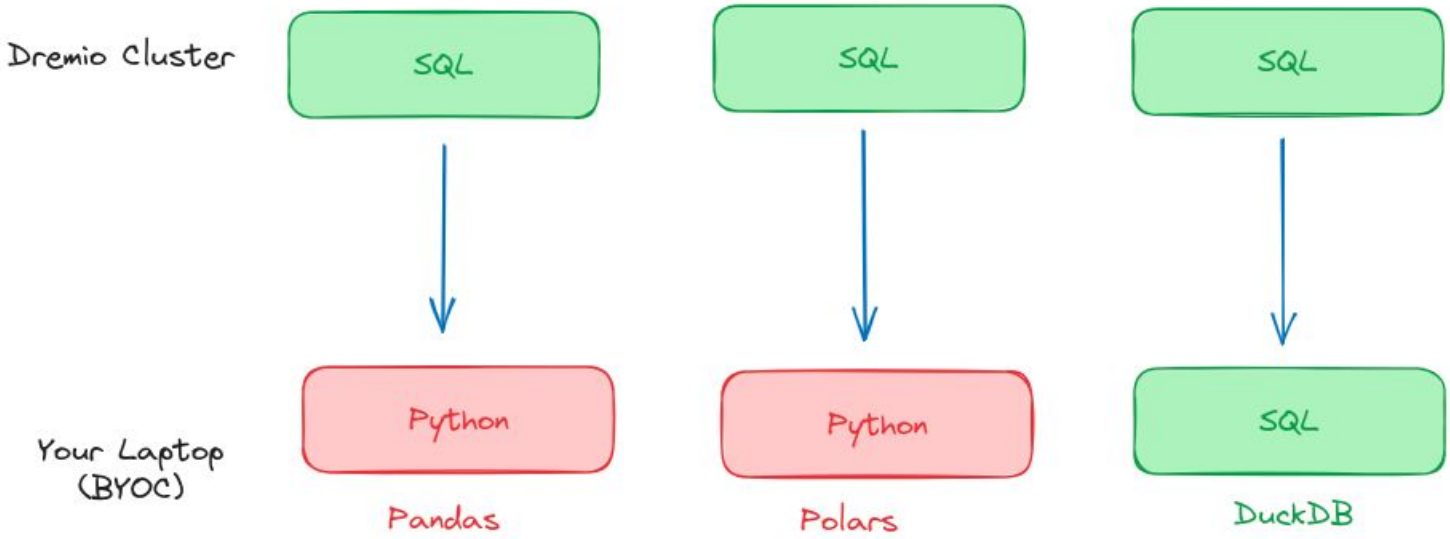
What is DuckDB?

- In-process database
- Columnar Storage
- Vectorized Query Execution
- Standalone
- ACID Transactions



GNARLY Data Waves

SQL All The Way

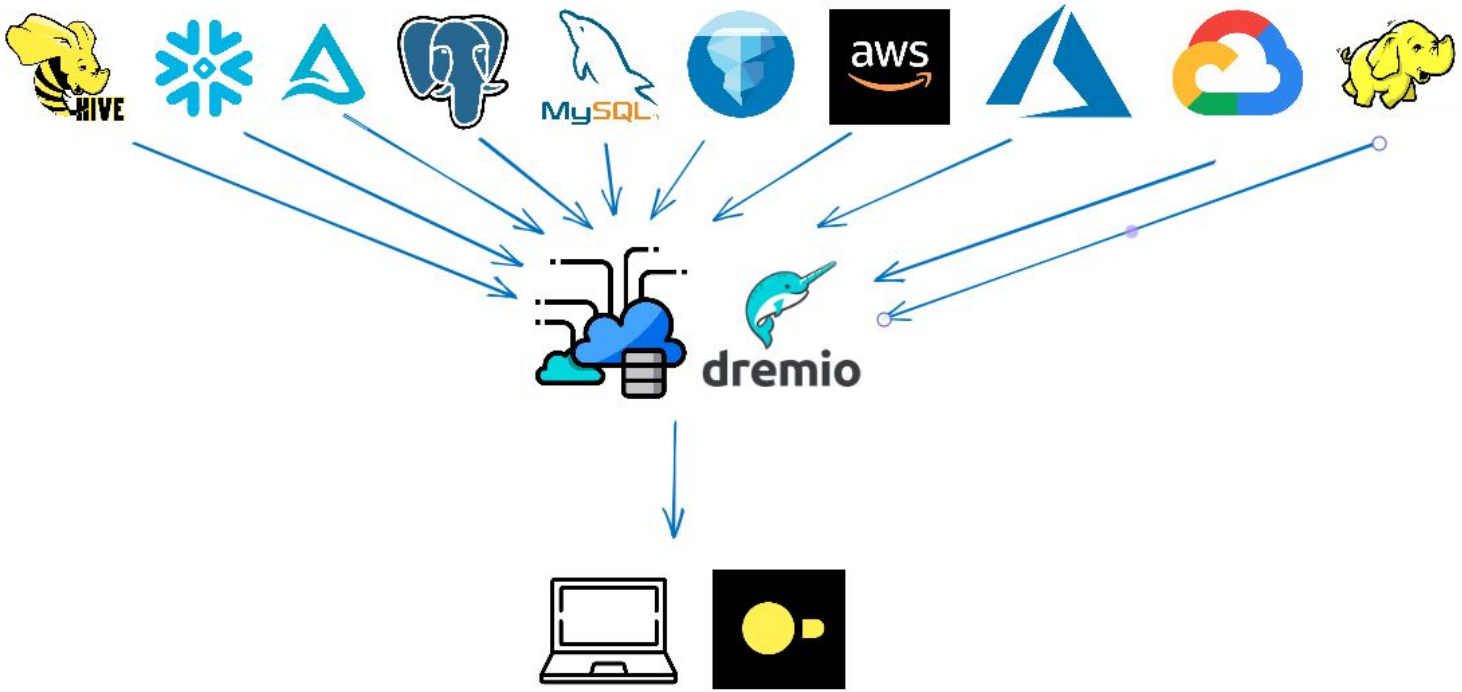


```
pip install dremio-simple-query
```

Dremio -> Arrow/Pandas

```python
from dremio_simple_query.connect import DremioConnection
from os import getenv
from dotenv import load_dotenv

load_dotenv()

## Dremio Person Token
token = getenv("TOKEN")

## Arrow Endpoint (See Dremio Documentation)
uri = getenv("ARROW_ENDPOINT")

## Create Dremio Arrow Connection
dremio = DremioConnection(token, uri)

## Get Back Arrow Stream from Dremio Query
stream = dremio.toArrow("SELECT * FROM arctic.table1;")

## Convert to Arrow Table
arrow_table = stream.read_all()

## Convert to Pandas Dataframe
df = stream.read_pandas()
```

Dremio -> DuckDB

```python
from dremio_simple_query.connect import DremioConnection
from os import getenv
from dotenv import load_dotenv

load_dotenv()

## Dremio Person Token
token = getenv("TOKEN")

## Arrow Endpoint (See Dremio Documentation)
uri = getenv("ARROW_ENDPOINT")

## Create Dremio Arrow Connection
dremio = DremioConnection(token, uri)

## Get Back DuckDB Relation
duck_rel = dremio.toDuckDB("SELECT * FROM arctic.table1")

## Query DuckDB Relation
result = duck_rel.query("table1",
"""
SELECT * from table1 WHERE field1 > 1 and field2 = 'label'
"""
).fetchall()
```

What Becomes Possible