

Apache Iceberg Crash Course

What is a Data Lakehouse and What is a Table Format?



Curriculum

July 11: What is a Data Lakehouse and What is a Table Format?

July 16: The Architecture of Apache Iceberg, Apache Hudi and Delta Lake

July 23: The Read and Write Process for Apache Iceberg Tables

Aug 13: Understanding Apache Iceberg's Partitioning Features

Aug 27: Optimizing Apache Iceberg Tables

Sep 3: Streaming with Apache Iceberg

Sep 17: The Role of Apache Iceberg Catalogs

Oct 1: Versioning with Apache Iceberg

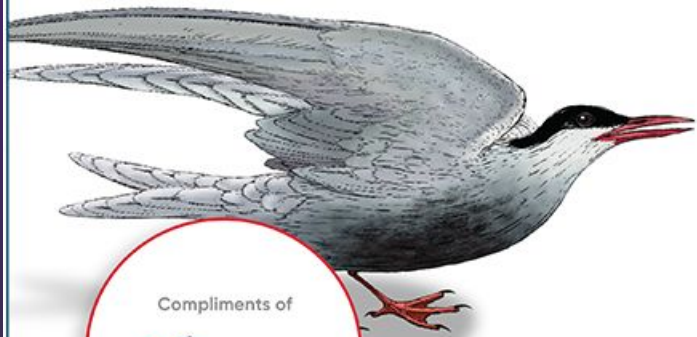
Oct 15: Ingesting Data into Apache Iceberg with Apache Spark

Oct 29: Ingesting Data into Apache Iceberg with Dremio

O'REILLY®

Apache Iceberg The Definitive Guide

Data Lakehouse Functionality, Performance,
and Scalability on the Data Lake



Compliments of



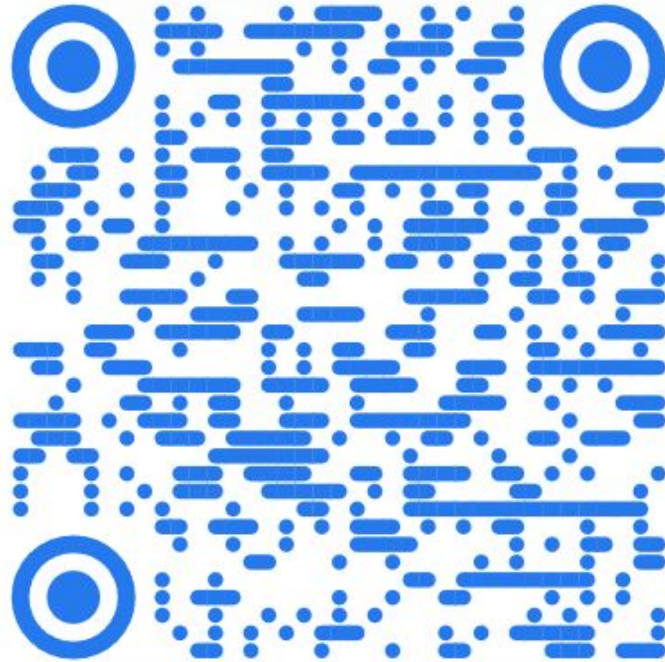
Tomer Shiran,
Jason Hughes &
Alex Merced

Forewords by Gerrit Kazmaier,
Raghu Ramakrishnan & Rick Sears





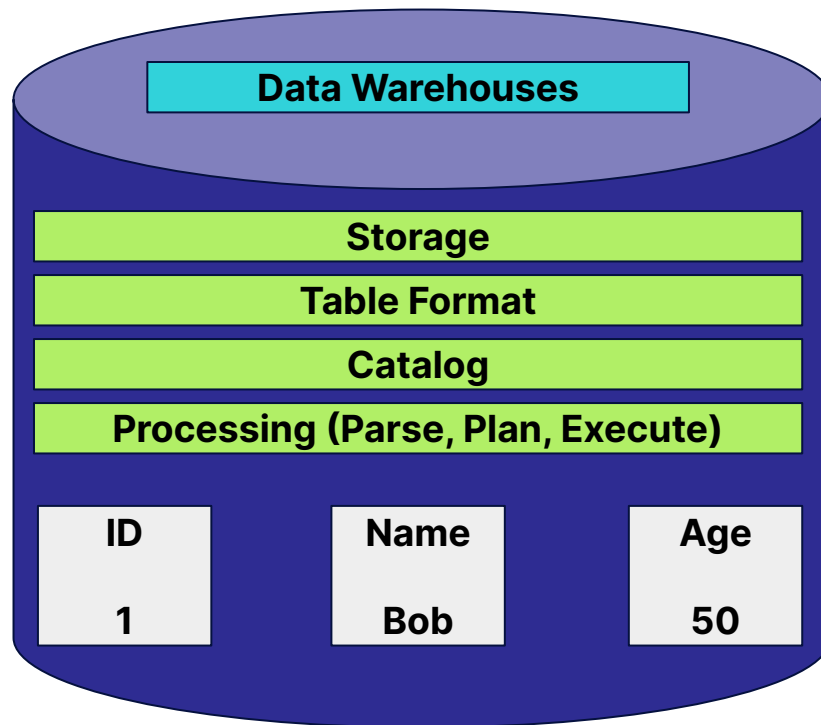
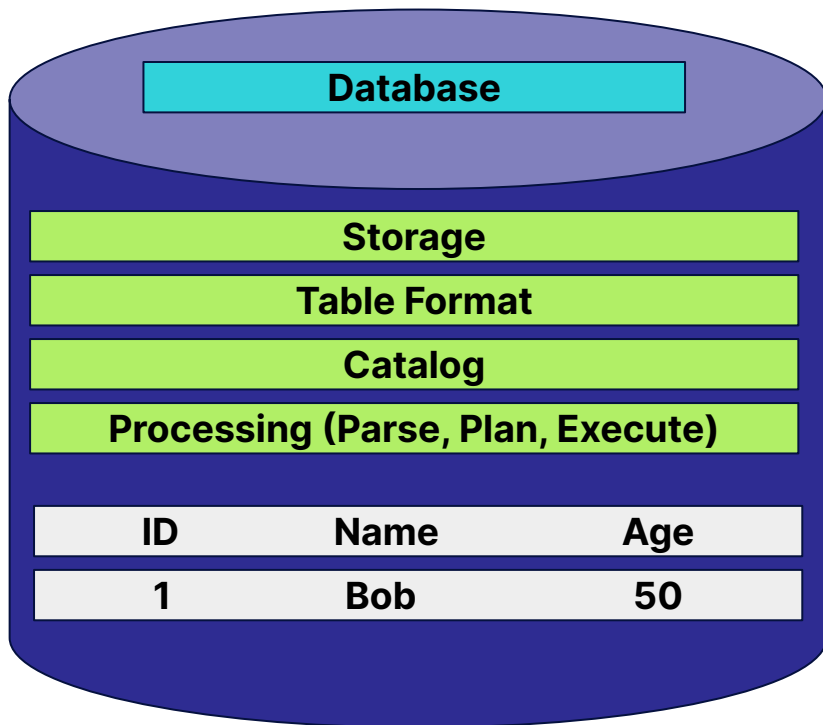
dremio.com/gnarly-data-waves
Youtube | Spotify | iTunes



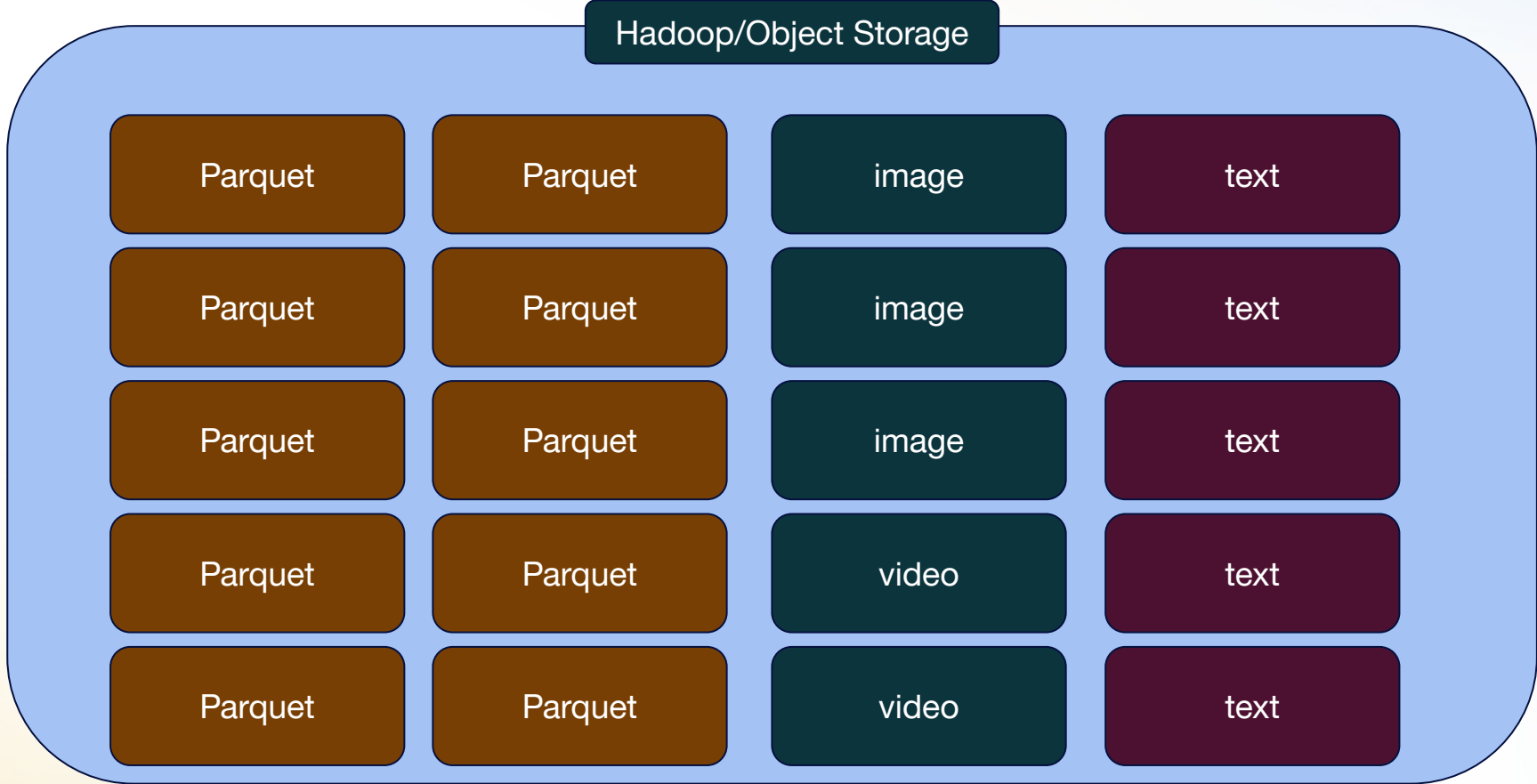
community.dremio.com
Apache Iceberg Category

What is a Data Lake?

Traditional Data Systems



What is a Data Lake?



What is a Data Lakehouse?

What is a Data Lakehouse?

Data Lake

Storage

**Compute/Processing
(Dremio, AWS Athena)**

Data Lakehouse

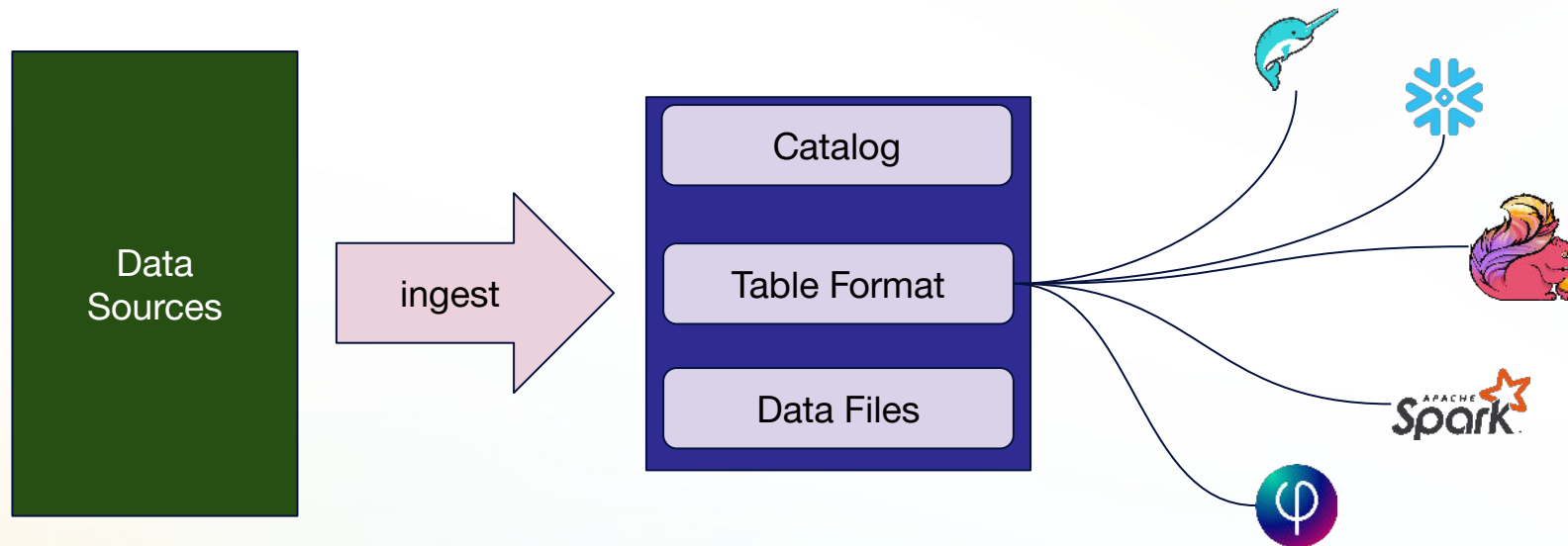
Storage

Table Format

Catalog

**Compute/Processing
(Dremio, AWS Athena)**

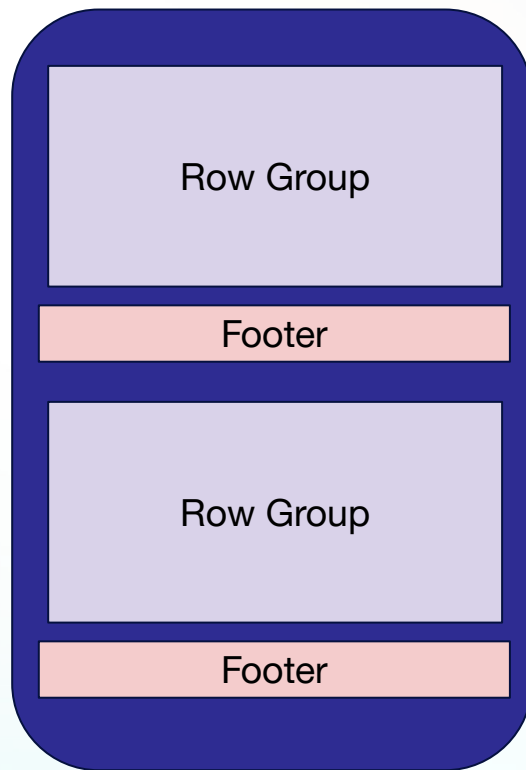
What is a Data Lakehouse?



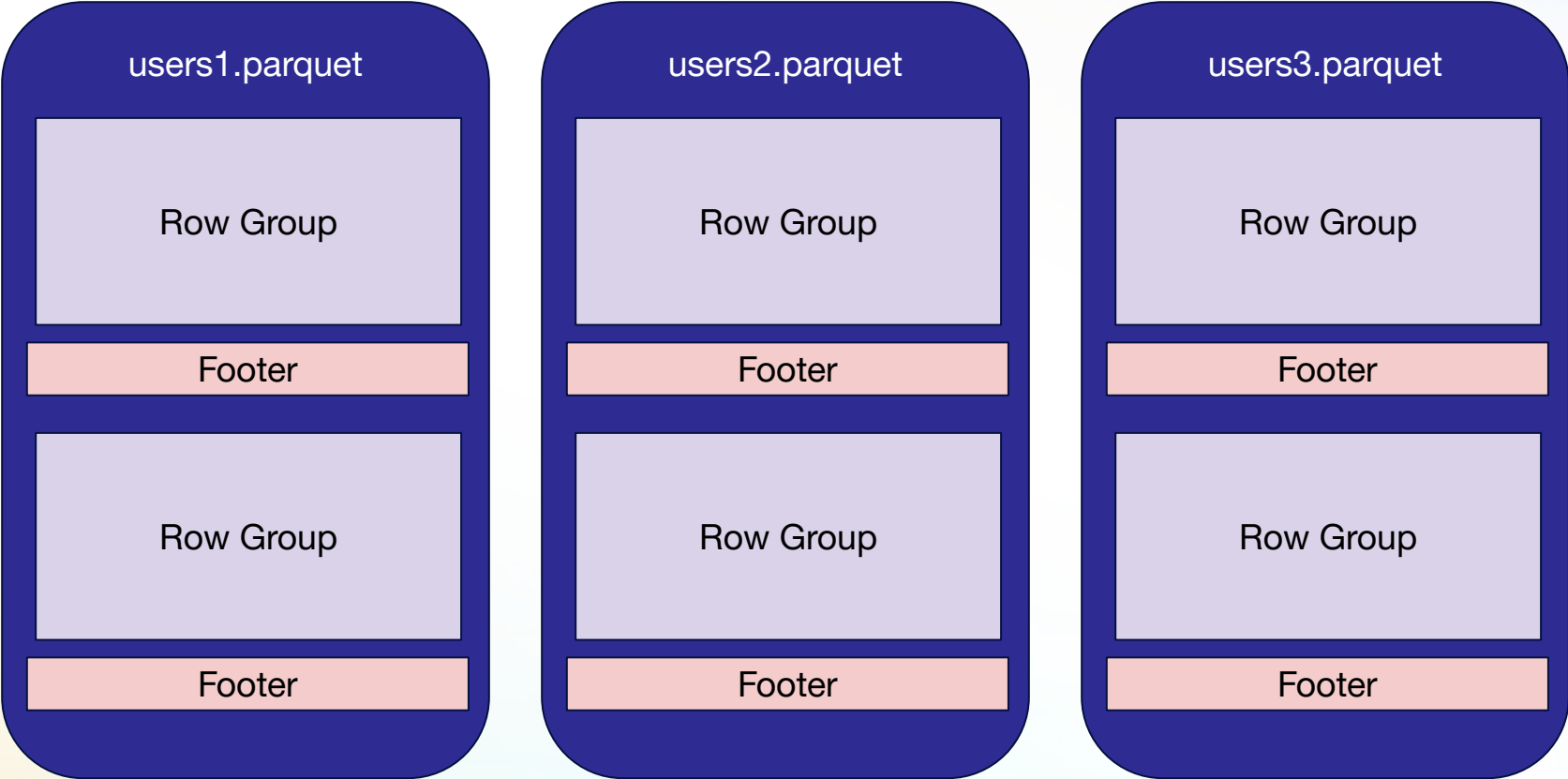
What is a Table Format?

Parquet Data Files

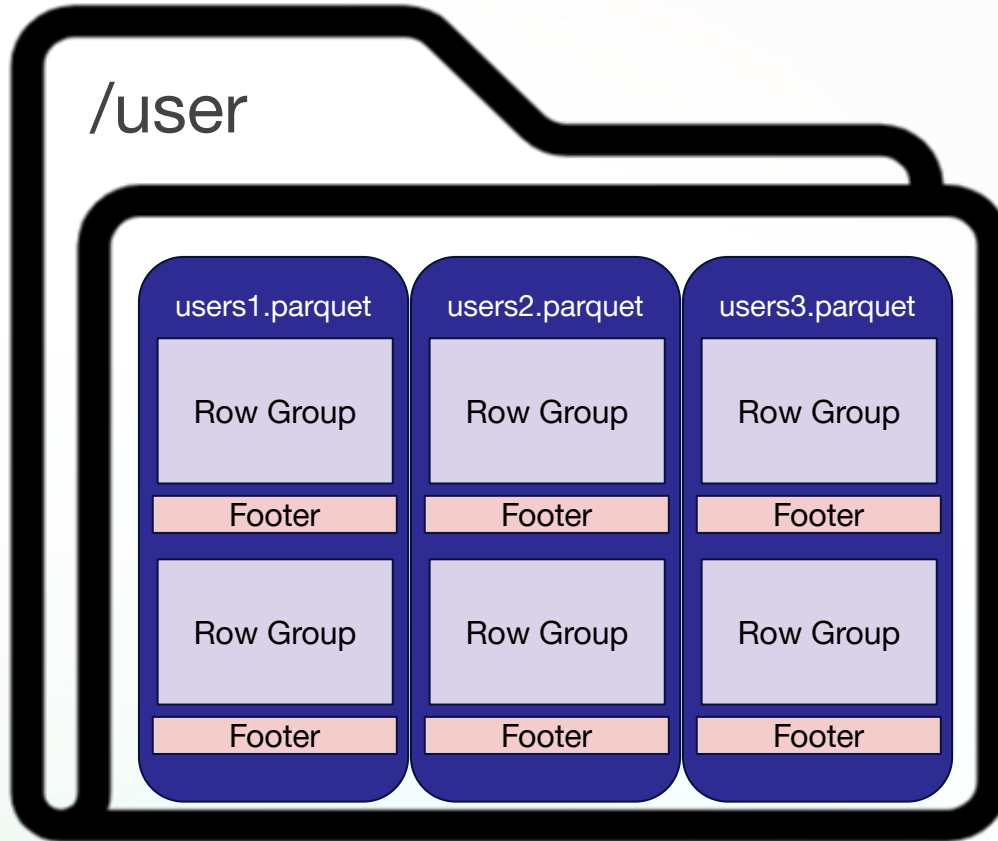
Parquet



Defining Multi-File Datasets



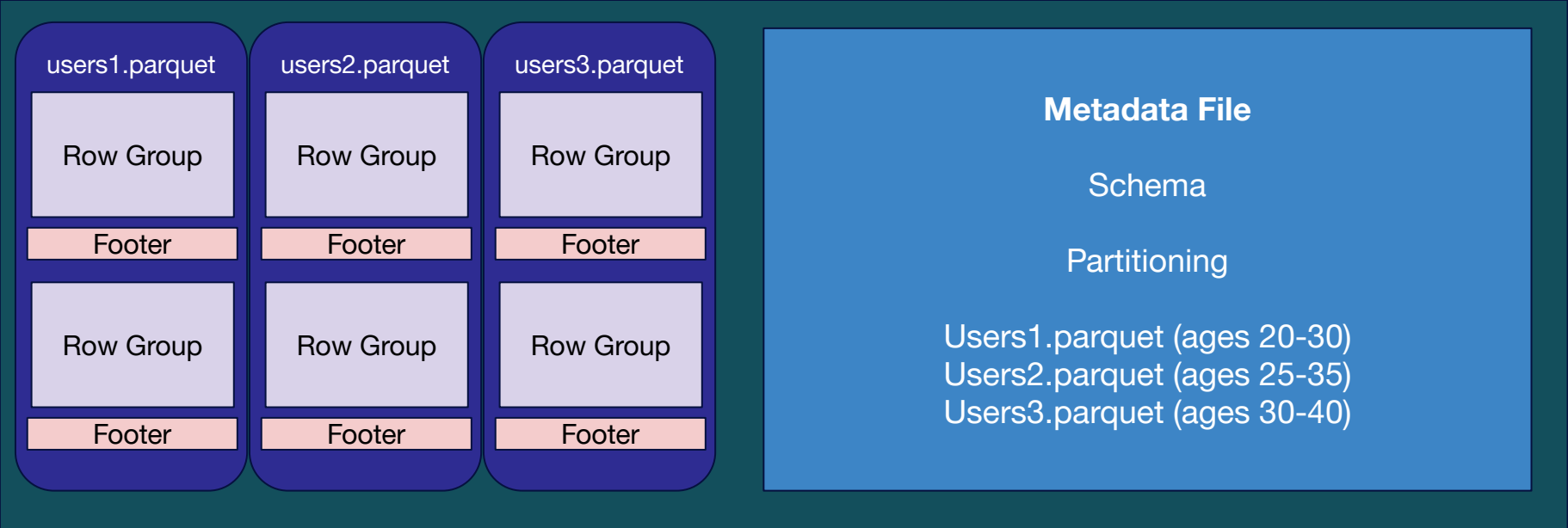
The Hive Approach to Table



Problems

- File Listing
- Granular Updates
- limited/No table evolution

Modern Approach to Table



Benefits

- Faster Scans
- Time Travel
- ACID Transaction
- Efficient Updates
- Schema Evolution



**A Iceberg/Dremio Lakehouse on
your laptop exercise**



**Deploy Dremio Software or
Dremio Cloud**



Postgres -> Iceberg -> Dashboard

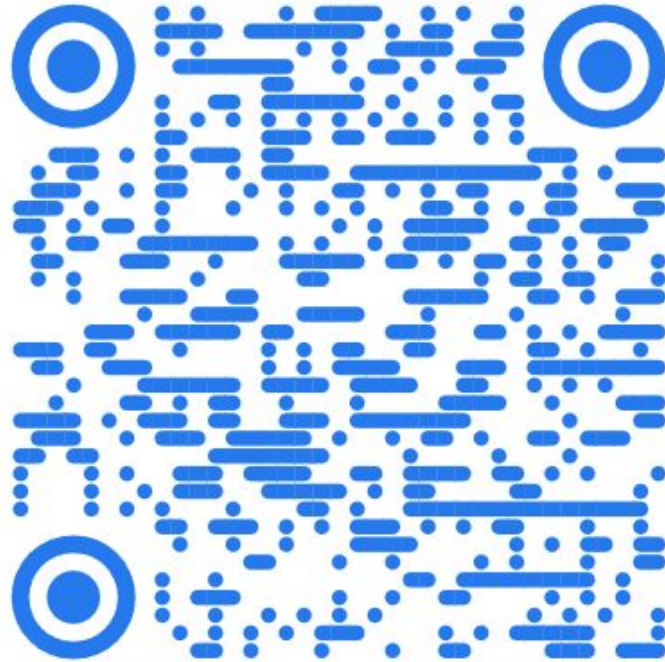


SQLServer -> Iceberg -> Dashboard



MongoDB -> Iceberg -> Dashboard

dremio.com/blog



community.dremio.com
Apache Iceberg Category