

# Apache Iceberg Crash Course

Versioning with Apache Iceberg



# 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**

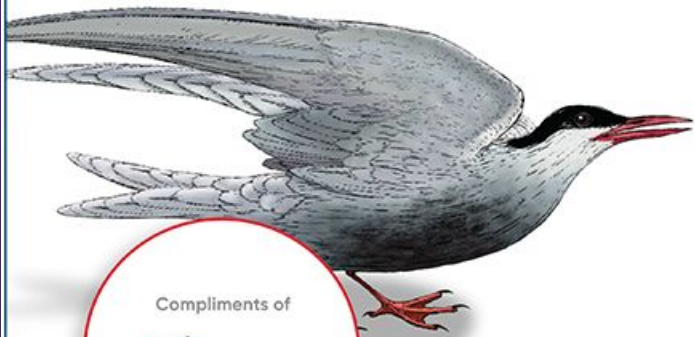
Oct 15: Ingesting Data into Apache Iceberg with Apache Spark

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



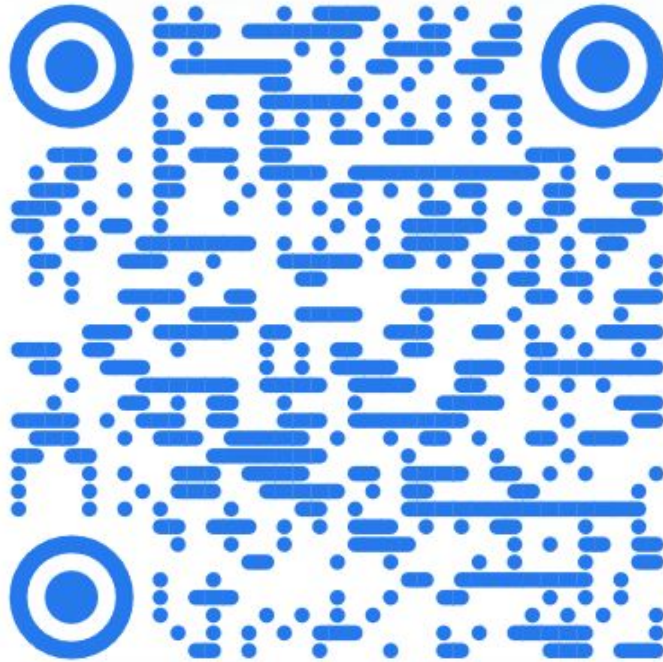
Tomer Shiran,  
Jason Hughes &  
Alex Merced

Forewords by Gerrit Kazmaier,  
Raghu Ramakrishnan & Rick Sears





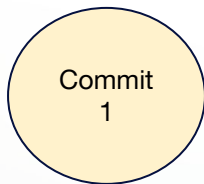
[dremio.com/gnarly-data-waves](https://dremio.com/gnarly-data-waves)  
Youtube | Spotify | iTunes



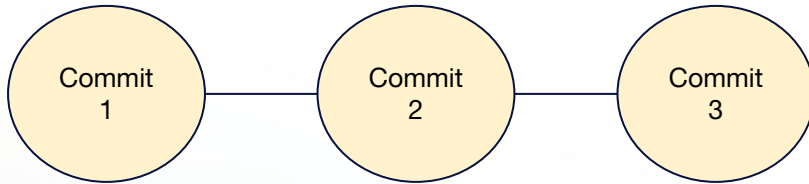
**community.dremio.com**  
**Apache Iceberg Category**

# What is Versioning

# What is a Data-as-Code/DataOps?

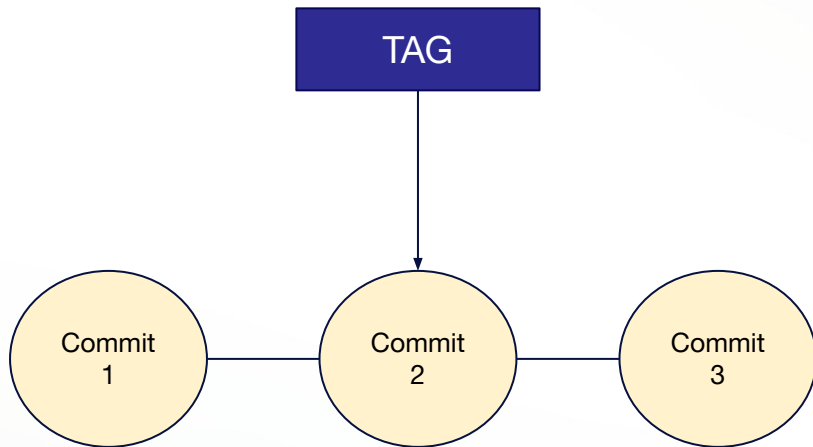


# What is a Data-as-Code/DataOps?

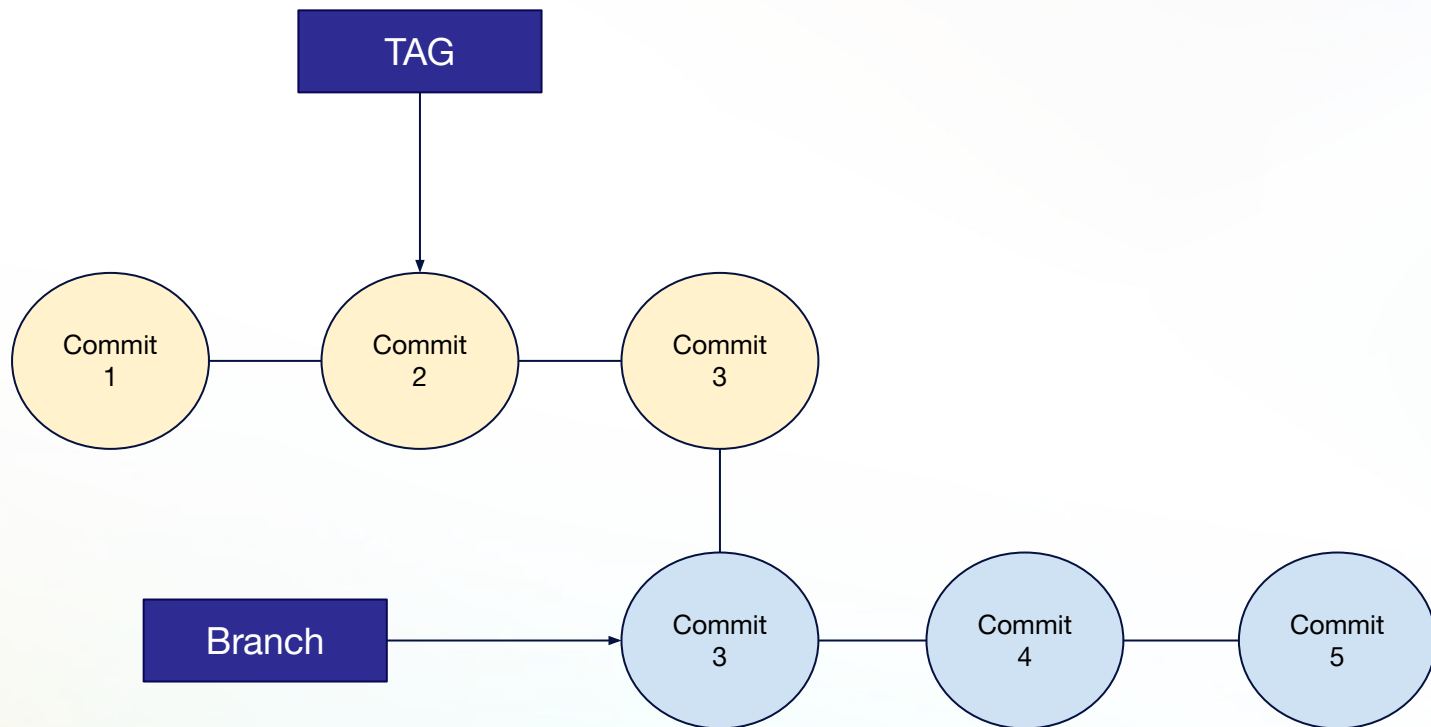




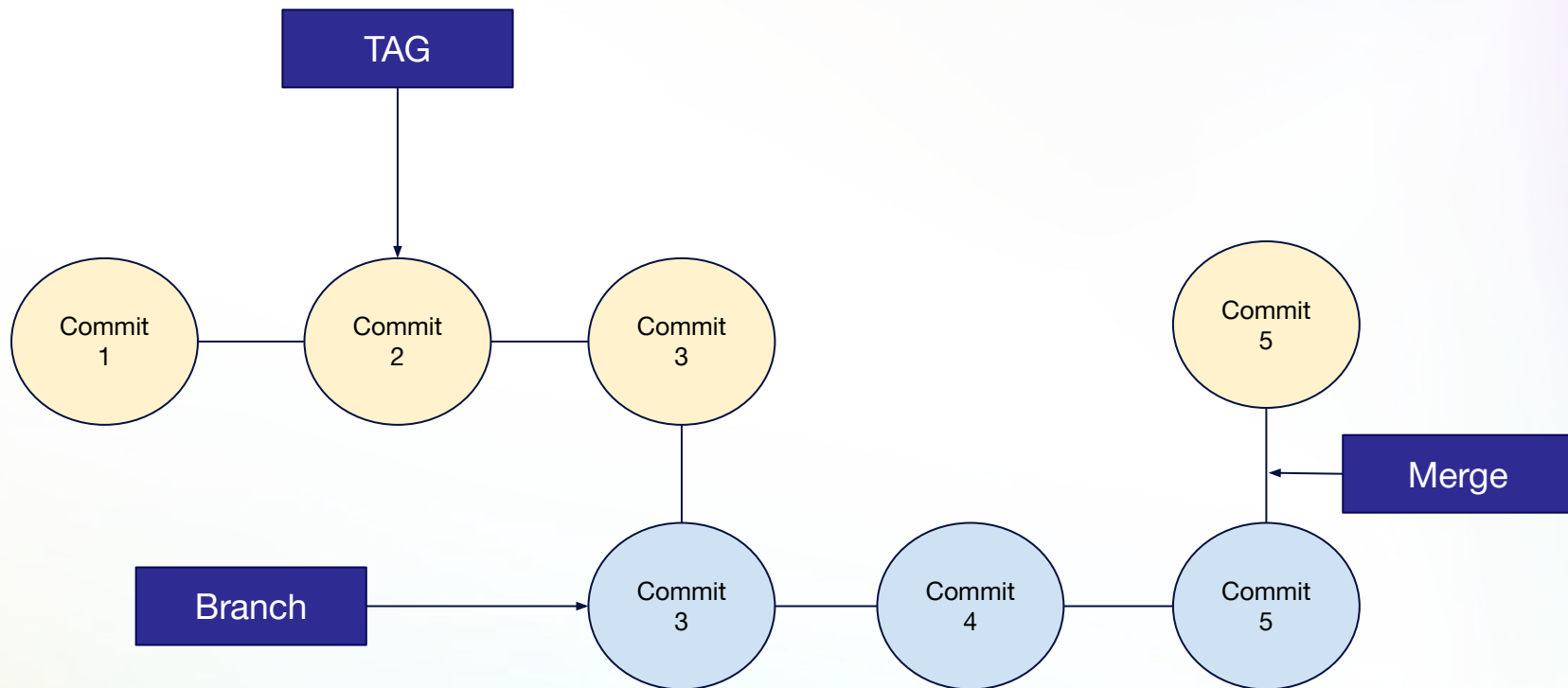
# What is a Data-as-Code/DataOps?



# What is a Data-as-Code/DataOps?



# What is a Data-as-Code/DataOps?



# Benefits of Versioning

Isolation

Zero-Copy

Multi-Table

Rollbacks

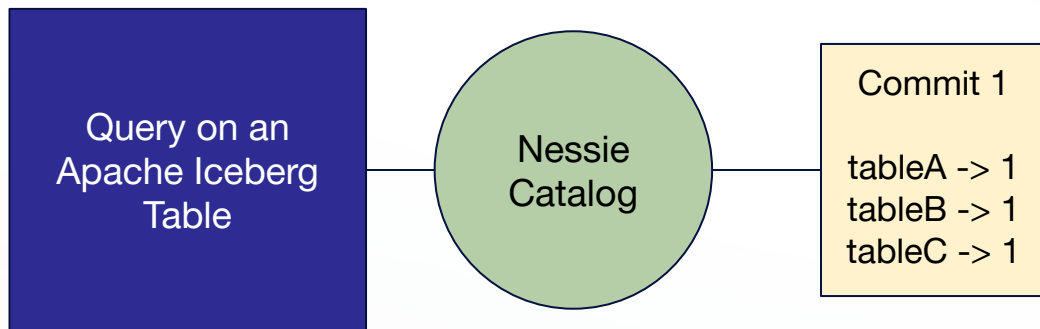
Reproducibility

Validation

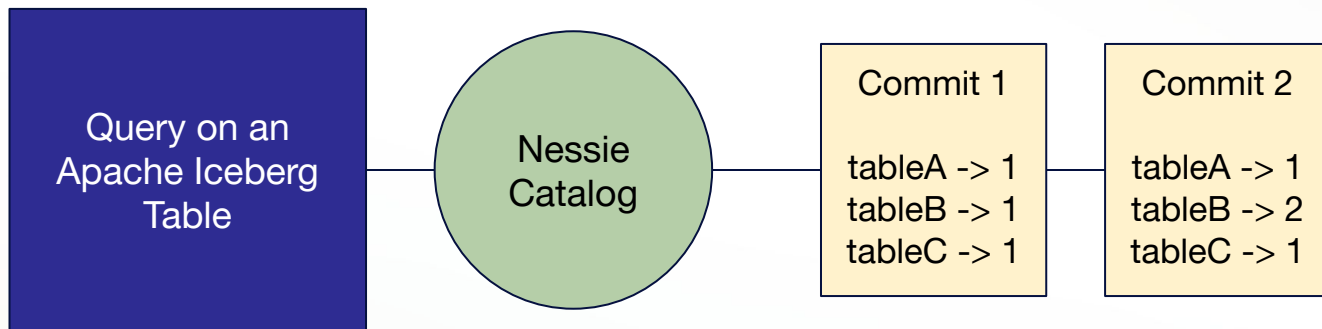
# Table Level Versioning

# Catalog Versioning (Nessie)

# What is Project Nessie?

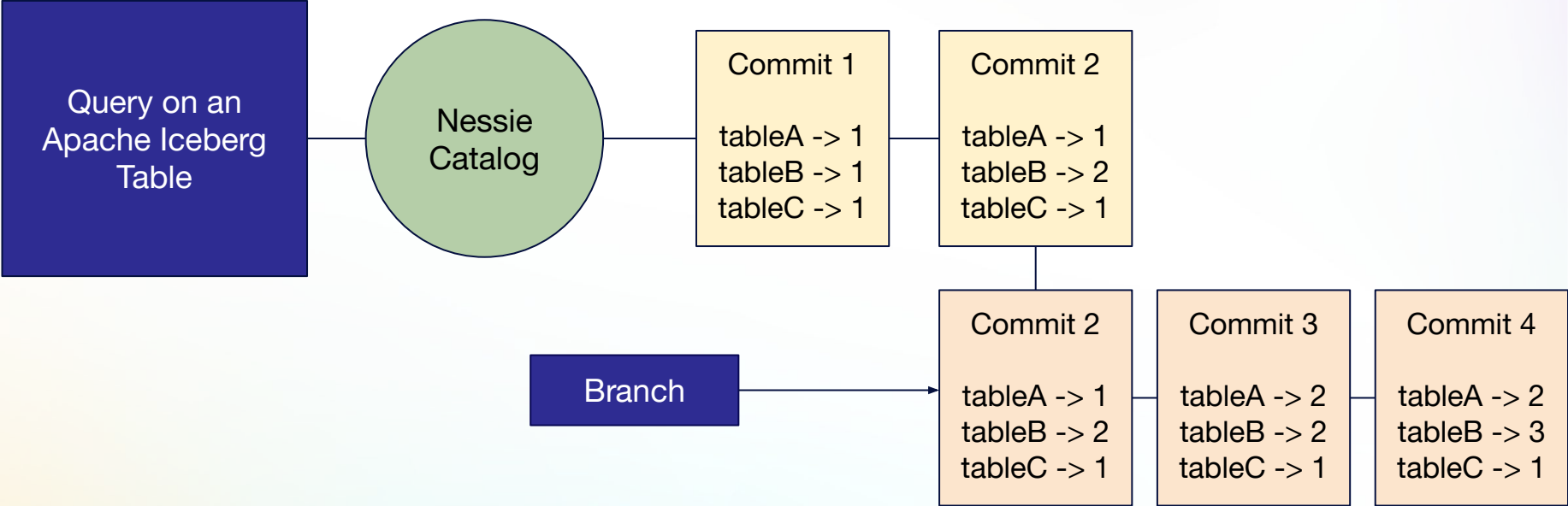


# What is Project Nessie?

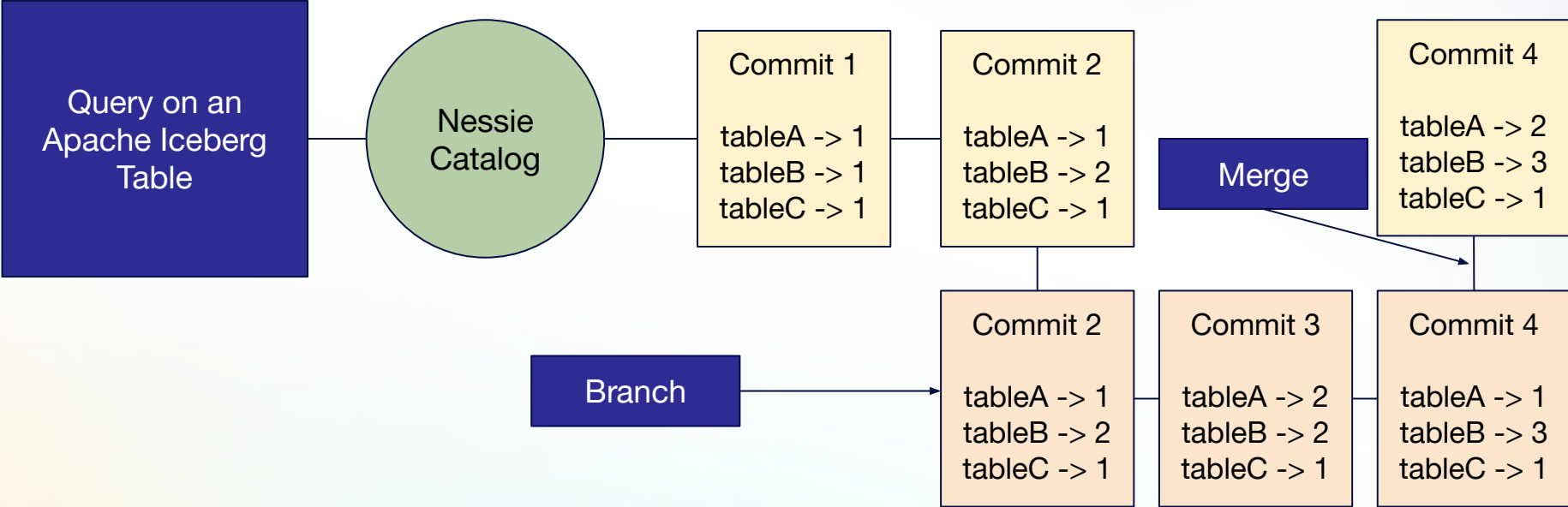




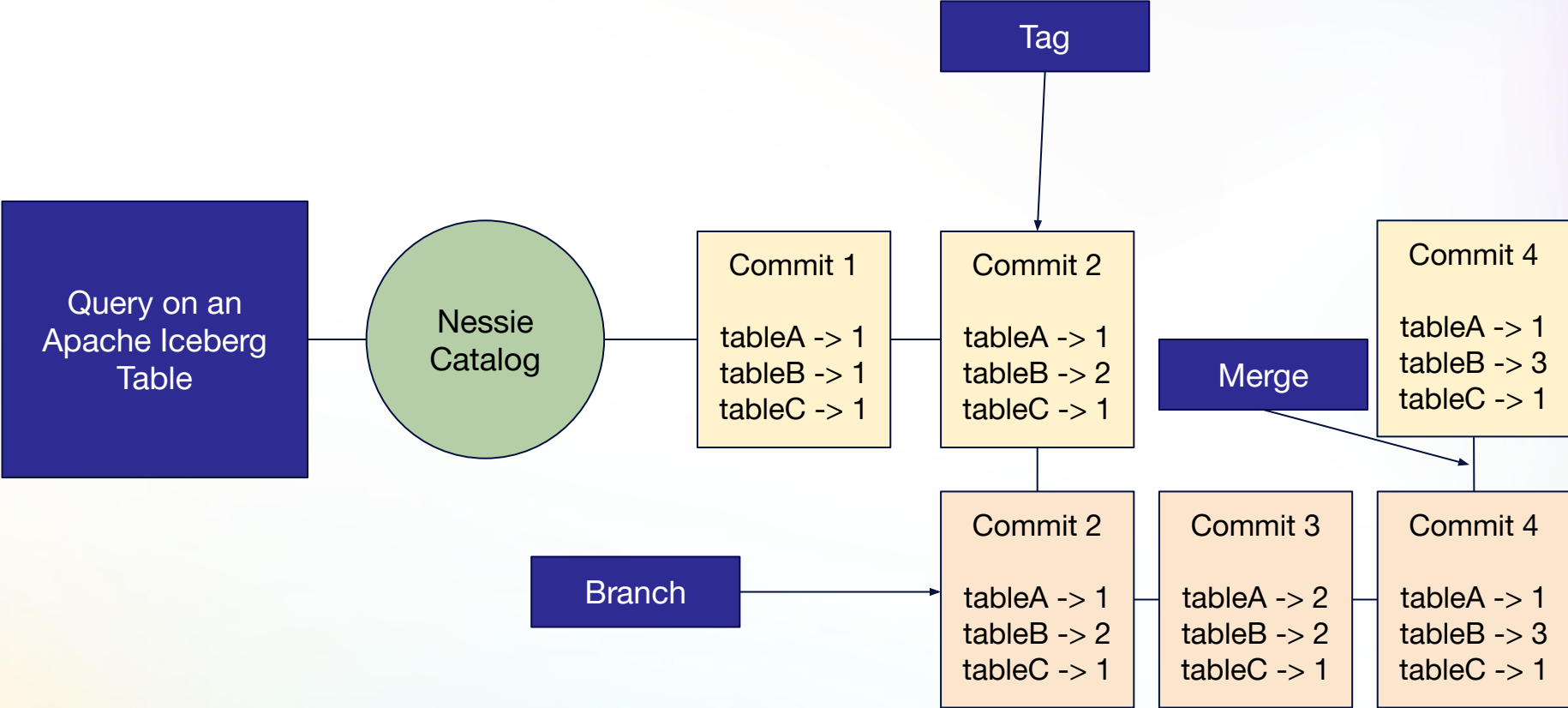
# What is Project Nessie?



# What is Project Nessie?



# What is Project Nessie?



# Benefits of Versioning

Isolation

Zero-Copy

Multi-Table

Rollbacks

Reproducibility

Validation

# Benefits of Nessie

Catalog Level

Open Source

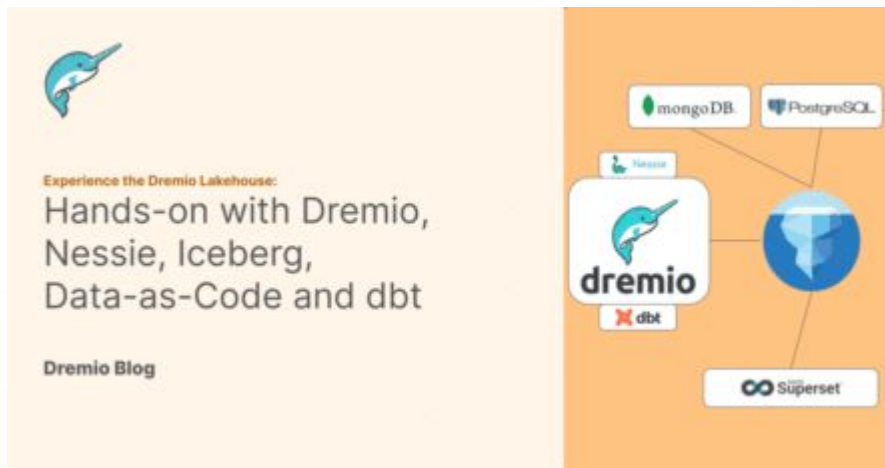
Managed Service

Cloud Agnostic

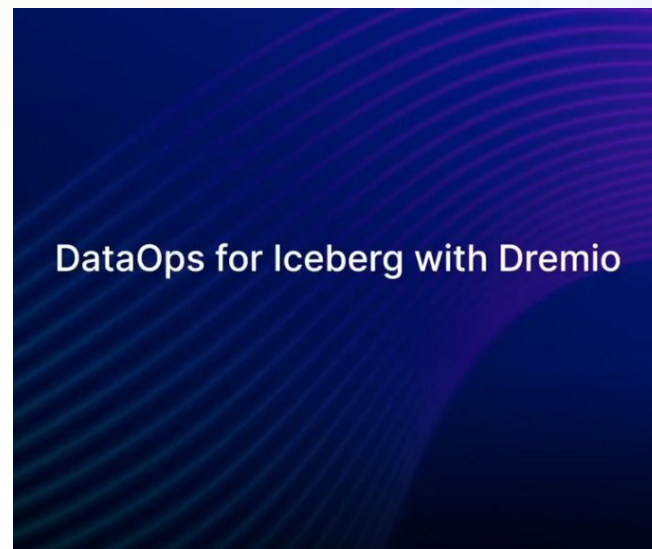
Storage Agnostic

Optimization

# See It All Put Together



The image is a blog post thumbnail. On the left, it features the Dremio logo (a blue whale) and the text: "Experience the Dremio Lakehouse: Hands-on with Dremio, Nessie, Iceberg, Data-as-Code and dbt". Below this, it says "Dremio Blog". On the right, there is a diagram with a central Dremio logo and a globe icon. Lines connect this central hub to five surrounding boxes: "mongoDB", "PostgreSQL", "Nessie", "dbt", and "Superset".

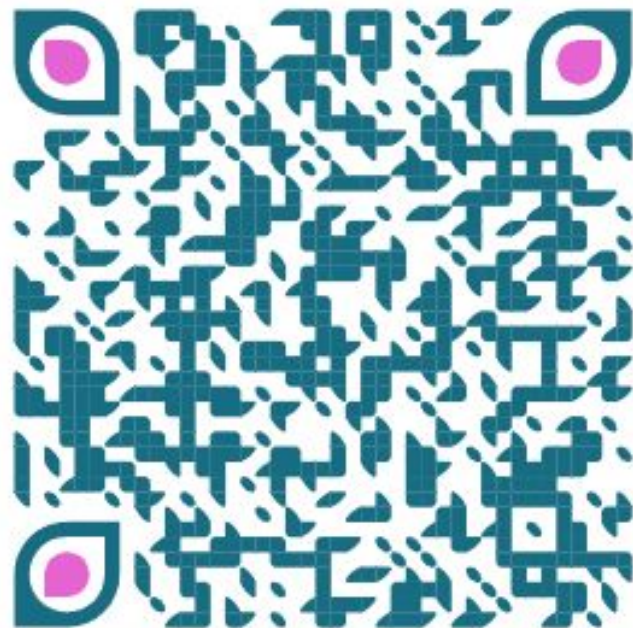


The slide has a dark blue background with a pattern of concentric, wavy lines in a lighter shade of blue. The text "DataOps for Iceberg with Dremio" is centered in white.





A Iceberg/Dremio Lakehouse on  
your laptop exercise



Deploy Dremio Software or  
Dremio Cloud



Postgres -> Iceberg -> Dashboard



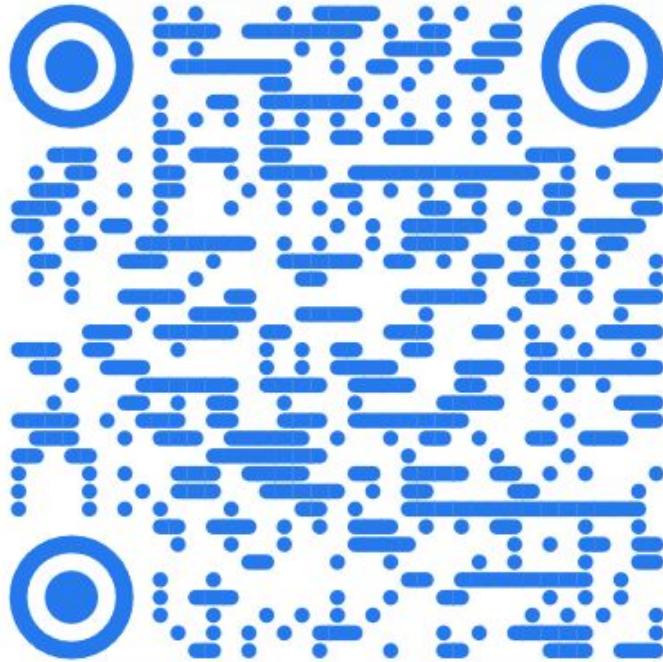
SQLServer -> Iceberg -> Dashboard



MongoDB -> Iceberg -> Dashboard

**[dremio.com/blog](https://dremio.com/blog)**





**community.dremio.com**  
**Apache Iceberg Category**