Data as Code: Project Nessie brings a Git-like experience for Apache Iceberg Tables

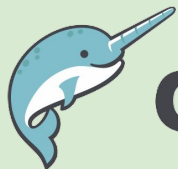Presented by Alex Merced

Alex Merced

Developer Advocate, Dremio

Alex Merced is a developer advocate for Dremio, a developer, and a seasoned instructor with a rich professional background. Having worked with companies like GenEd Systems, Crossfield Digital, CampusGuard, and General Assembly.

Alex is a co-author of the O'Reilly Book "Apache Iceberg: The Definitive Guide." With a deep understanding of the subject matter, Alex has shared his insights as a speaker at events including Data Day Texas, OSA Con, P99Conf and Data Council.
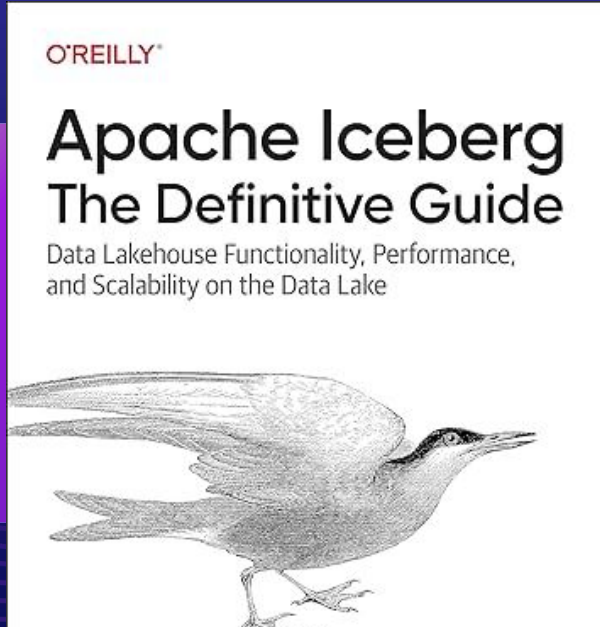
Driven by a profound passion for technology, Alex has been instrumental in disseminating his knowledge through various platforms. His tech content can be found in blogs, videos, and his podcasts, Datanation and Web Dev 101.
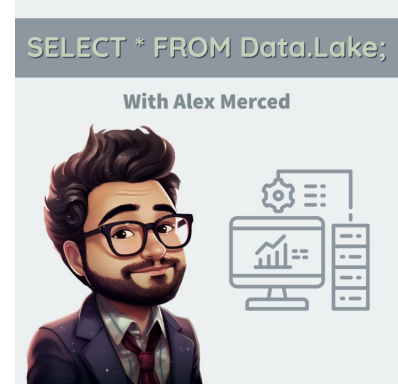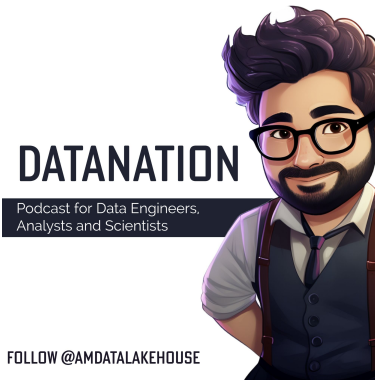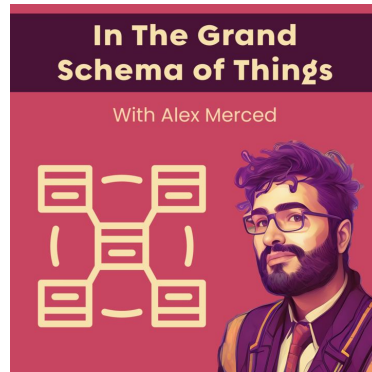
Moreover, Alex Merced has made contributions to the JavaScript and Python communities by developing a range of libraries. Notable examples include SencilloDB, CoquitoJS, and dremio-simple-query, among others.

dremio

Apache Iceberg: The Definitive Guide

O'REILLY®

Apache Iceberg
The Definitive Guide

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

GNARLY Data_waves

In The Grand Schema of Things
With Alex Merced

DATANATION
Podcast for Data Engineers,
Analysts and Scientists

FOLLOW @AMDATALAKEHOUSE

SELECT * FROM Data.Lake;
With Alex Merced

Subscribe on Spotify/iTunes

dremio

Data Lake

```
┌─────────────────────────────┐
│   Compute (Read Only)       │
└─────────────────────────────┘
        │              │
        ▼              ▼
┌─────────────────────────────┐
│     Storage of Files        │
└─────────────────────────────┘
```

Data Lakehouse

```
┌─────────────────────────────┐
│   Compute (Read/Write)      │
└─────────────────────────────┘
        │              ▲
        ▼              │
┌─────────────────────────────┐
│   Table Format + Catalog    │
└─────────────────────────────┘
        │              ▲
        ▼              │
┌─────────────────────────────┐
│     Storage of Files        │
└─────────────────────────────┘
```

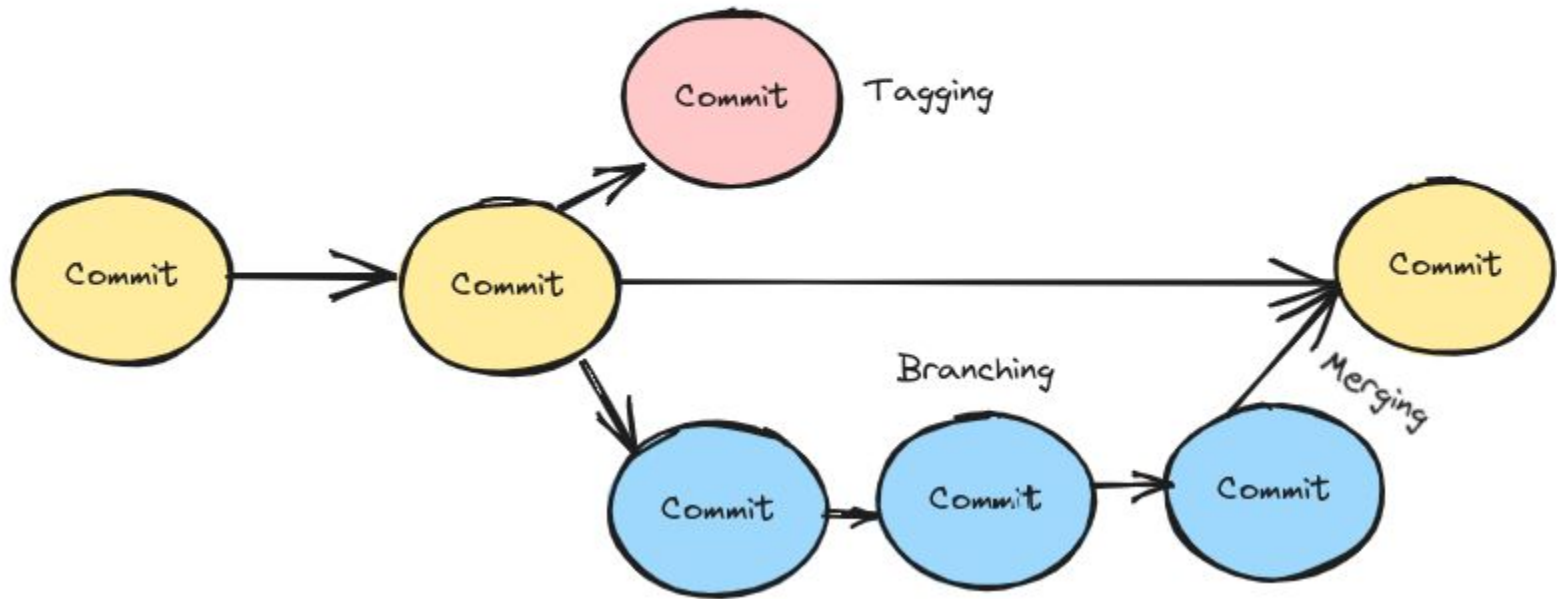dremio
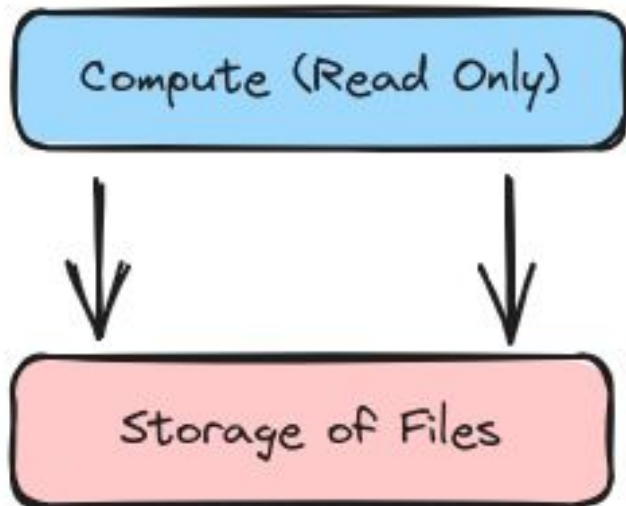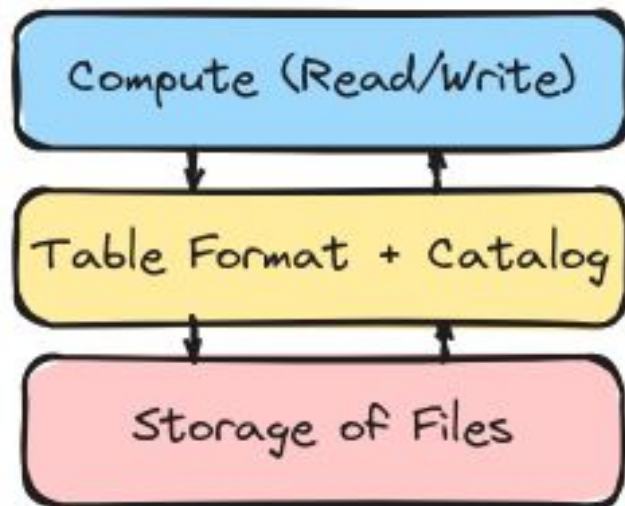
What is Data as Code?

Bringing Code Like Practices like CI/CD & Versioning to Data

dremio

Benefits of Versioning

Isolation

No-copy Experimentation
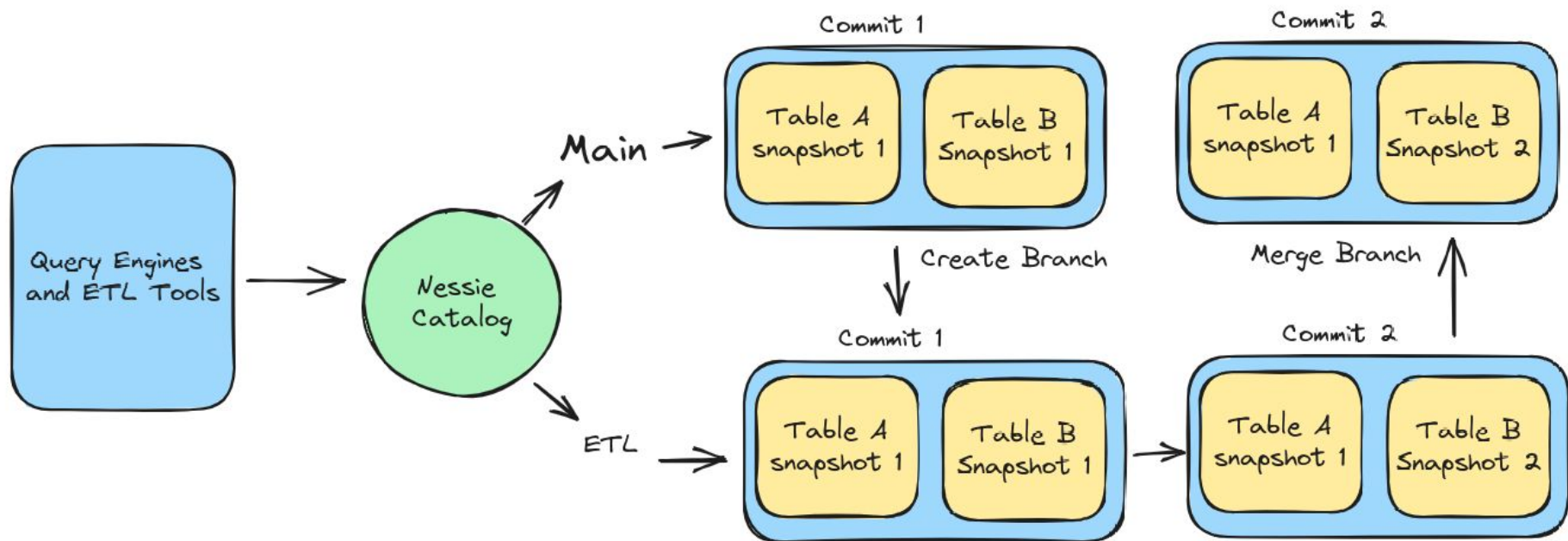
Multi-Table Transactions

Rollbacks

Reproducability

Consistency, Quality and Data Validation

dremio

What is Project Nessie?

dremio

Benefits of Nessie

Catalog Level Operations (Branch, Merge, Tagging, Rollback)

Open Source

Cloud Managed Service (Dremio Arctic)

Cloud Agnostic (Azure, AWS, Google, etc.)

Storage Agnostic (Object Storage, HDFS, etc.)

Access Control and Clean-up Features

dremio

Alex Merced
follow @amdatalakehouse

dremio