



Versioning Data in the Data Lakehouse (File, Table and Catalog Versioning) - Q&A

Versioning is a technique that has helped software developers to develop many practices that allow them integrate and deploy new code continuously allowing for more rapid development of software. In a world where data is being generated faster than ever, the data community needs technology that allows for rapid integration and deployment of new data.

In this presentations we'll discuss:

- 3 Levels of Versioning on the Data Lakehouse (File, Table and Catalog)
- Pros and Cons to each versioning paradigm
- When should you use each?



Alex Merced
Developer Advocate
Dremio

Show's Q&A

1. **Are the three levels (Nessie, Iceberg, LakeFS) themselves transactionally linked in the sense that if there is an error on one level within an operation will rollback the change across the layers??**

Answer: No, they aren't connected. How they effect each other varies, refer to Q&A Video

2. **File Level Version question - can they be stored on AWS File storage ? The slide stated 'Required S3 Compat object storage.'?**

Answer: My understanding is LakeFS requires S3 compatible storage, so AWS S3 should work fine, but other types of storage AWS may offer like spinning up a Hadoop cluster with EMR wouldn't support LakeFS on those files.