



Data as Code with Dremio Arctic: ML Experimentation & Reproducibility on the Lakehouse - Q&A

As more data consumers require access to critical customer and operational data in the data lake, data teams need solutions that enable multiple users to leverage the same view of the data for a wide range of use cases without impacting each other. In this episode of Gnarly Data Waves, we will discuss how the data as code capabilities in Dremio Arctic enable data scientists to:

- Create a data science branch of the production branch for experimentation without creating expensive data copies or impacting production workloads
- Easily work and collaborate cross-functionally with other data consumers and line of business experts
- Quickly reproduce models and results by returning to previous branch states with tags and commit history



Alex Merced
Developer Advocate
Dremio



Dipankar Mazumdar
Developer Advocate
Dremio

Show's Q&A

1. While merging, how are potential data differences in the two branches handled?

Answer: Data differences are essentially handled by applying the changes based on the manifests (manifest list) on whether a file was appended or removed.

2. Is there a way we can see all the commits in Dremio Arctic UI and manage them?

Answer: Dremio Arctic UI presents a unified view of all the commits made by a particular user. The 'commits' section allows us to view them.

3. Do we have to only use Dremio Sonar as the compute engine with Arctic?

Answer: No. Arctic is open to other engines such as Apache Spark.