Managing your data as code with Dremio Arctic - Q&A

As data lakes become the primary destination for growing volumes of customer and operational data, data teams need tools and processes that ensure data quality and consistency across data consumers and use cases. Jeremiah Morrow and Alex Merced discuss the emergence of data as code for data management, its benefits for data teams, and how Dremio customers are using it to deliver access to a consistent and accurate view of data in their data lakes.

Join us in the upcoming episode of Gnarly Data Waves and learn about:

- Why data as code is necessary for ensuring consistency and data quality for large data lakes.
- How Dremio Arctic uses Git-like concepts such as branches, tags, and commits to make data management easy.
- Some high value use cases for data as code.

Jeremiah Morrow
Director, Product Marketing
Dremio

Alex Merced
Developer Advocate
Dremio

Show Q&A

1. **Can you do this via API? Or via dbt-dremio? Ideally we would not want to manage this data manually via UI as part of a daily data load or CI/CD process?**
   Answer: Yes, any tool or service that can send SQL to Dremio via its REST/ARROW FLIGHT/JDBC/ODBC interfaces can automate these processes.

2. **When do we hope to get CI/CD solutions in Dremio?**
   Answer: Automated Optimization of tables in the near term, will follow up on job scheduling

3. **Can you connect to mainframe data, db2, IDMS etc?**
   Answer: DB2 Connector has been released or on the near term roadmap, please consult Dremio AE/SA on the existence of any particular connector.

4. **What would happen to an artic table if I used Create or Replace Table as (select query)? Does it create a new iceberg table or do we have to explicitly insert new records to the arctic table?**
   Answer: A new iceberg table is created with a new snapshot history at the table level, but the Arctic catalog would still have the historical catalog level commits pointing to the old snapshot history theoretically.

5. **Data is coming from many sources, so they constantly change and we do not need a source control for this data. Could you show one of the use cases when we need source control of data?**
   Answer: There are other reasons to embrace data as code other than source control such error resolution, multi-table transactions and experimentation.

6. **Does Dremio support all formats rdbms and NoSQL?**
   Answer: Dremio aims to support as many as possible and is constantly expanding its slate of connectors, but new platforms are releases frequently so I wouldn't say ALL, but all the most popular tools should be covered.