Dremio and Data Lakehouse Table Formats (Apache Iceberg, Delta Lake and Apache Hudi & Dremio) - Q&A

In the search to implement a data lakehouse many have been adopting one of the three major data lakehouse table formats. In this presentation you'll learn about how the different formats can be used with Dremio's lakehouse platform.

- What is a Table Format
- What is Iceberg, Delta Lake and Hudi
- Reading with Dremio
- Using Multiple Formats with Dremio
- Accelerating Queries with Dremio

Alex Merced
Developer Advocate
Dremio

Show's Q&A

1. **Does Delta Lake need a catalog?**
   Answer: It doesn't need one, since it's contained in a single directory, although catalogs always help in bundling all your tables and bringing them from engine to engine instead of having to load them one by one into each tool.

2. **Can you explain with an example what happens when you store a transactional table in an Iceberg format? Does Iceberg read the table and create metadata tables?**
   Answer:

   Engines will use their own parquet writers then
   - pass list of parquet files to manifest writer
   - pass list of manifests to to manifest list writer

- pass manifest to metadata writer
- update catalog with reference to new metadata file

3. **Often the best technology fails to win in the marketplace. How does Iceberg plan to counter all the industry momentum towards Delta Lake by offerings like Databricks, Microsoft Fabric, Parquet users, etc.?**
   Answer: I think there are a lot of big events spurring adoptions of Delta and Iceberg, I think they'll be able to coexist and the ability and friction between working between them will continue to reduce.

4. **How does Dremio operate in the context of encryption and data governance, especially in a multitenant environment with a variety of customer keys and access grant levels?**
   Answer: Recommend sitting with a solutions architect to explore for a particular use case.

5. **Does Dremio integrate with the open source metadata management Egeria project?**
   Answer: Recommend sitting with a solutions architect to explore for a particular use case.

6. **How does one keep themselves up-to-date on the capabilities of each of these table formats, esp. iceberg being more open-source AND community driven, while Delta Lake being a little more tied to its parent.?**
   Answer: Follow people on linked in, follow repos, and twitter

7. **Where can one test drive all 3 in 1 env ?**
   Answer: Refer to my blog here for getting started in spark then refer to Hudi docs and use the same image to try Hudi.
   https://www.dremio.com/blog/3-ways-to-convert-a-delta-lake-table-into-an-apache-iceberg-table/

8. **Is there anything of interest for domains with structured, but not exactly general purpose data formats, e.g. BAM and VCF (Gene/Protein -omics data sets)?**
   Answer: Recommend sitting with a solutions architect to explore for a particular use case.

9. **Can Dremio work with Databricks Unity Catalog?**
   Answer: Dremio supports the Hive API, so theoretically yes, haven't tried it yet

10. **If I was answering the question about Market momentum I would refer to the Dremio partnership page. https://www.dremio.com/partners/**
    Answer: Agreed

11. **Does Dremio have a Data Lineage service?**
    Answer: Yes, Dremio has lineage for views created in Dremio

12. **Does Dremio have a preference between ETL and ELT (like the "Medallion Architecture)?**
   Answer: Dremio is built to progressively enhance all data workflows