



How to Modernize Hive to the Data Lakehouse with Dremio and Apache Iceberg - Q&A

So you want to base your data lakehouse around Apache Iceberg to take advantage of its features, performance and vast ecosystem of platform/tool compatibility. You'll need to take your current Hive tables and convert them to Iceberg. Iceberg offers several tools to do so depending on your needs and in this webinar we'll explore that migration process.

How to do an in-place migration and avoid rewriting your data

- How to do a shadow migration where you can update your data's schema and partitioning
- How to move Apache Iceberg tables from one catalog to another



Alex Merced
Developer Advocate
Dremio

Show Q&A

1. **Once you copy metadata into Nessie, how do you keep it synchronized with the changes made in the source catalog?**

Answer: You should only have your tables in one catalog at a time, the register table procedure is just for migrating between catalog not to sync between two catalogs.

2. **How do you coexist Hive tables and iceberg tables for a while since you have applications running on both for a (long) time until applications are migrated?**

Answer:

In a Hive catalog, Hive and Iceberg tables can coexist.

3. You said a new catalog becomes your new catalog. That does not work with the coexistence situation I mentioned above.

Answer: multiple tables can exist in one catalog, although a table should not be in multiple catalogs.

4. With Iceberg, how does metadata update work with writes happening from different processes / engines?

Answer: Iceberg uses optimistic writes, so all writes will be attempted and if write A completes before write B, write B will be reattempted

5. How is compaction triggered in scenarios like streaming ingest where we're accumulating lots of small parquet files?

Answer: You can use Dremio Arctic to schedule regular table automation when using the Dremio Arctic catalog.

6. Is it possible to write from Kafka to on-prem Nessie + Iceberg? Did not see that in your Nessie compatibility slide.

Answer: Use flink to connect to Kafka, as Flink has an Iceberg sync.

7. Does the community edition have Nessie and arctic?

Answer: Dremio Software will soon have a Nessie connector.

8. So there's no way to use Nessie from the community with the current edition?

Answer: Dremio Software will soon have a Nessie connector.

9. Is Nessie itself in the community version?

Answer: Nessie is a separate piece of software you can run yourself or you can get the cloud SaaS version via Dremio Arctic.

10. OK, so Nessie is a separate open source that users can stand in their own environment? correct?

Answer: Yes

11. If I migrate to Iceberg how would I keep my tables optimized, must I do this manually?

Answer: This can be done via SQL queries to Spark or Dremio which can be automated using python scripts calling pySpark or sending HTTP/Arrow/JDBC requests to Dremio. Services like Dremio Arctic have table optimization available.

12. How can I get started with Dremio, how much would it cost?

Answer: Dremio Cloud and Dremio Cloud edition are free to start and you can later upgrade to an enterprise tier.