![Dremio logo]

Apache Iceberg Office Hours - Q&A

Join the Dremio developer advocacy and engineering teams for the first installment of Apache Iceberg office hours. During this time we'll have a brief Iceberg presentation on copy-on-write vs merge-on-read in Iceberg and then lots of dedicated time for Q&A on the presented topic or any other questions or guidance you're looking for help on in learning about Apache Iceberg or architecting your data lakehouse around Apache Iceberg.

Examples of questions you can come to ask:
- How can I optimize my Iceberg tables for my different use cases?
- What tools will best handle my ETL job to write to Iceberg?
- How can I control access to my Iceberg tables?
- How can I convert data from X into an Iceberg table?
- How can I get started with Iceberg in Databricks?

Join us for the first installment of our regular Apache Iceberg office hours to help build your low-cost high performance data lakehouse!
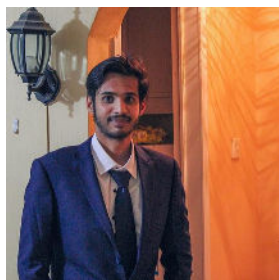
Jason Hughes
Director of Product Management
Dremio

Alex Merced
Developer Advocate
Dremio

Dipankar Mazumdar
Developer Advocate
Dremio

Ajantha Bhat
OSS Engineer
Dremio

Dmitri Bourlatchkov
Sr. OSS Developer
Dremio

Scott Cowell
Software Engineer
Dremio

Webinar Q&A

1. **How do you handle replication failures on AWS S3 with Iceberg? Say you commit a transaction, the data and metadata files are successfully written to one region, commit succeeds, metadata file location is updated in the metastore, and then AWS S3 fails to replicate the new files to other region(s). What happens to your Iceberg table?**
   Answer: The region of the bucket isn't in the address stored by Iceberg so in the worst case scenario it will search different regions to find where the files do exist if it doesn't exist in region X.

2. **What is the best way to get data into Dremio after an AWS ETL Glue job? There are 3 schemas that I need to have updated simultaneously into 3 VDS's.**
   Answer: Should be based on intervals of metadata refresh, can trigger manual metadata refresh

3. **I don't want some to be updated before others....it's important to have consistency.**
   Answer: refer to previous answer but also using catalog level branches with Dremio Arctic/Project Nessie you can easily achieve multi-table consistency.

4. **How can I tune metadata refreshes? I thought I didn't have much control over that.**
   Answer: Refer to the [ALTER TABLE x REFRESH METADATA](#) command

5. **Yes, I did hear about Arctic that can solve this problem precisely. I just don't use it currently, and wasn't sure if AWS Glue could do it as well.**
   Answer: AWS GLUE doesn't have versioning features for this problem at the moment, there is discussion in the Iceberg community of updating the Iceberg catalog spec to enable multi-table consistency with any catalog.

6. **If you want to go into more detail on external reflections, and when they should be used? (or point to documentation, I didn't find too much online)**
   Answer:
   Here is where you can find more info about [External Reflections](#), which are reflections a View can be configured to use for acceleration that are managed outside of Dremio. An external reflection can only accelerate a particular view and must match the current schema of the data.