



What's New in the Apache Iceberg Project: Version 1.2.0 Updates, Pylceberg, Compute Engines - Q&A

The Apache Iceberg project has made tremendous strides, evolving on various fronts such as usage, ecosystem adoption, community growth, and capabilities. In the past few months, the project has introduced many exciting new features and performance improvements around the core library, compute engines and standalone libraries (such as Pylceberg) that makes this lakehouse technology robust & valuable for organizations. In this episode of Gnarly Data Waves, we will go over some of the notable new capabilities of Apache Iceberg.

Specifically, we will discuss about:

- Version 1.2.0 release
- Features such as : Branching/Tagging, New write-distribution-mode, Change Data Capture, Catalog Migrator Tool, Delta to Iceberg migration
- Pylceberg (What's happening in the Python library)
- Compute Engine-specific features: Dremio, Apache Spark, Flink



Alex Merced
Developer Advocate
Dremio



Dipankar Mazumdar
Developer Advocate
Dremio

Show Q&A

1. Delta to iceberg, does it retain the z-order as well?

Answer: It creates a table using the existing files, so the ordering of the files should be retained. As new data is added to the table you'll want to run future compaction jobs using the z-order strategy.

2. How do I implement data quality checks like write-audit-publish in Iceberg using branches?

Answer: Apache Iceberg's new branching capability allows you to create isolated branches from a particular snapshot and allows writing data to that branch. So, we can basically ingest new data to that isolated branch (separate from production), run our data quality checks and then publish it to the main branch. More details in this blog: <https://www.dremio.com/blog/streamlining-data-quality-in-apache-iceberg-with-write-audit-publish-branching/>