



## Getting Started with Dremio's Data Lakehouse - Q&A

As organizations strive to provide value faster to end users, data silos makes it difficult to provide insights on time. Learn how [Dremio's data lakehouse](#) accelerates data delivery and discovery, without copies.

In this episode, you will get:

- The fundamentals of the data lakehouse with Dremio and Apache Iceberg
- Proven use cases for unifying data access on the lakehouse
- Customer success stories



**Read Maloney**  
CMO  
Dremio



**Brock Griffey**  
Principal Master Solution Architect  
Dremio

## Show Q&A

### 1. I see apache parquet, does Dremio also support JSON file store?

**Answer:** Yes, Dremio supports JSON/CSV/PARQUET/XLXS file types. Resource link: <https://docs.dremio.com/software/data-sources/entity-promotion/#supported-file-formats>

### 2. BI at 1/10 the cost vs what?

**Answer:** This is based on TCP-DS vs. Cloud Data Warehouses, but when our customers calculate, they base it on their previous system vs. Dremio. These savings are typically more, but not necessarily replicable.

### 3. Are there benchmarks of Dremio Sonar vs AWS Athena / Trino ?

**Answer:** Yes - we benchmark against Athena as well. Results are similar...the good part of Athena is its serverless architecture. So not bad for some Ad hoc queries at times. We see some customers use us and Athena (basically replace Redshift with us)...others just use us.

**4. Is it 100x Faster than Trino uses Reflections?**

**Answer:** Yes - for BI workloads...using reflections. As I was mentioning, this is a BI performance benchmark.

**5. If we ingest data every second, how is reflection updated? On each insert ? Or by some schedule? Do you recalculate all data in reflection? Or just store some state for new data?**

**Answer:** Reflections refresh on a schedule or on demand. Depending on the type of reflection, it can be incremental or full refresh. More information can be found here: <https://docs.dremio.com/cloud/reflections/refreshing-reflections/>

**6. Does any of the data actually need to move out of the source files or does all the querying just utilize the semantic layer?**

**Answer:** The data does not need to move, so you can query it directly from the source like "postgres.tablename" but I can also curate views in my semantic layer from those sources which are logical views on the source "marketingspace.viewonpostgrestable" and data reflections can help accelerate any physical source or logical view.

**7. So, the Table "models\_and\_prices" is virtual? Just making sure I understand.**

**Answer:** No, that table was physical data in the data lake, stored in iceberg table format with arctic as the catalog.

**8. Checking my understanding - Sonar doesn't need to copy, Arctic needs to have data in Iceberg table format?**

**Answer:** Correct.

**9. How does CI/CD work with Dremio arctic ? all the tools can be used from code ? How can we manage separate environments?**

**Answer:** Using Dremio Arctic you can isolate ingestion by creating and working from a branch and merging the branch in when the data is ready. This can be done via SQL, so any tool that can be used to automate SQL being sent to Dremio or other Dremio Arctic supporting engines can automate this process.

**10. How can we define access scope for individual stakeholders ?**

**Answer:** Dremio has supports RBAC as well native column and row access policies. Users can be restricted in what they can access through these policies.

<https://docs.dremio.com/software/security/rbac/rbac-structure/>  
<https://docs.dremio.com/software/security/rbac/row-column-policies-udf/>

**11. What options to connect to Dremio from Ruby**

**Answer:** You can interact with Dremio using it's REST, JDBC/ODBC, and Arrow Flight interfaces in any language

**12. In what format does it send data via HTTP interface?**

**Answer:** The REST API returns data in JSON

**13. I was able to connect from Ruby via Arrow Flight with help from Arrow core developers, not sure if it is production ready solution**

**Answer:** Should be fine, Arrow Flight is a Standard Protocol so any client that implements it should get the job done.