dremio

Getting Started With Dremio Data Reflections - Q&A

For analytical workloads, data teams today have various options to choose from in terms of data warehouses and lakehouse query engines. To enable self-service, they provide a semantic layer for end users, usually with materialized views, BI extracts, or OLAP cubes. The problem is, this process creates data copies and requires end users to understand the underlying physical data model.

Join the Dremio engineering team in this episode of Gnarly Data Waves to learn about accelerating your queries with data reflections. Get answers to business questions faster without the challenges that come with today's approach, such as governing data copies or managing complex aggregate tables and materialized views.

In this episode, you will learn:

- The importance of data reflections and how it removes the need for data copies
- When to use raw reflections and aggregate reflections
- Best practices on data reflection refreshes

Tiong Lee
Sr. Director, Engineering
Dremio

Alex Merced
Developer Advocate
Dremio

Tong Truong
Sr. Product Marketing Manager
Dremio

Show Q&A

1. **Does Dremio have materialized view support?**
   Answer: A view with a RAW reflection that has all fields is essentially a materialized view. That means whenever a query refers to the view, we will use the materialization automatically by default. Better than that, we will still find potential reflections that will accelerate your query further than your RAW reflection. That means reflection is strictly better than the materialized view in terms of performance. On top of that, your “RAW

reflection" for the view can be used to accelerate other queries even if the quote and quote "materialized view" is never referenced in any query.

2. **Does reflection materialization require computing and storage costs?**
   Answer: The reflection refresh requires proper resourcing of compute and long-term storage - and it's probably no different than your typical pipelines that create derived datasets. One of the coolest things though is a reflection refresh can be accelerated by other reflections during refresh. For example, if you have a day sales total reflection, it will be used for weekly sales total reflection or year sales reflection internally during the refresh. So your data operation will always be most optimal in terms of efficiency and cost.

3. **Can reflection be used for ETL, as in the talk you are moving data from external sources? Where is reflection materialized stored, what is the format, and can it be accessed by other systems?**
   Answer: We have some customers who use reflections that way so sure. The reflection materialization is stored in cloud storage as Nessie/Iceberg data.

4. **We just migrated to Dremio now but our system is quite complex and hasn't used reflection yet. Where do we start?**
   Answer: It is true that when you use Dremio, you can still use your data system in the traditional way - the reflection usage is completely optional. The semantic layer and reflection is a paradigm shift that could take some time to get used to if you already have an existing system in place. You should start by taking a smaller isolated area - migrate it to start using the new way of working with data with Reflection. Once you get comfortable with Reflection usage, envision how it can work on a larger scale and plan for the larger refactoring of your data workflow. In fact, you will find that most of the time it is about simplification of all your data processes, integration, and usage for the long haul.

5. **I am curious about the potential cost savings when using Dremio with Snowflake for example. Do you have any point of view on an order of magnitude? I believe both are priced based on usage right?**
   Answer: You can compare the pricing with https://www.dremio.com/pricing/. Both Dremio/Snowflake are priced based on usage.
   But what makes Dremio stand out is you practically just point to the data and start querying the data - unlike Snowflake where there needs to be a lot of data movement before you can do anything with your data. And with Dremio Reflection, you will focus less on creating a lot of Derived Dataset for your reports and focusing more on accelerating the data with the common reflections which make many of the reports Dashboard faster. The time-to-market for the data is much quicker, and the ROI is much higher, not to mention the simplicity of it which is priceless.