



Unified Access for Your Data Mesh: Self-Service Data with Dremio's Semantic Layer - Q&A

Data silos and a lack of collaboration between teams have been long-standing challenges in data management. This is where data mesh comes into play as an architectural and organizational paradigm, providing an solution by enabling decentralized teams to work collaboratively and share data in a governed manner across the enterprise.

Dremio's semantic layer provides a particular useful tool for achieving both of these needs and in this episode we will discuss:

- The needs of a data mesh (Data Products, Computational Governance, Self-Service)
- The open and decentralized nature of the Dremio Open Data Lakehouse
- How data products can be created and shared with Dremio's semantic layer
- How governance can be architected centrally using fine-grained access rules
- How to unify your data products across the enterprise
- How the Dremio to Dremio connector enables sharing between domains



Alex Merced
Developer Advocate
Dremio

Show Q&A

1. Are these layers virtual?

Answer: Yes, Dremio views are logical views of the data not physical copies

2. How do you handle data that is common among all business domains? Do you place it in a separate domain?

Answer: You can add it at the Dremio admin level then add a view of the raw data into each space for each domain to evolve without making copies.

3. How does Dremio create aggregates of data across different logical business domains?

Answer: you can create joins across datasets in multiple domains and run aggregations

4. Is there any distributed computing also?

Answer: Dremio is an MPP system

5. How do you handle large query results? eg/ 10 TB? or "burst" output that spikes really high?

Answer: Depends on how you pull the data, for example if using an Arrow Flight connection you'll receive a StreamReader to read the data in batches.

6. Does Dremio work with third party IAM systems for management of user profiles

Answer: Dremio has many security integrations with Privacera, Ranger and more. I'd talk to a Dremio Solutions Ar

7. Does Dremio provide API's to write custom connectors to interface with proprietary enterprise data sources

Answer: Dremio Software has ways to write customer connectors, Dremio and Software and Cloud both have REST/JDBC/ODBC/ARROW FLIGHT interfaces for custom applications, notebooks, etc.

8. Can you enable Caching in the Dremio semantic layer?

Answer: There are several caching mechanisms throughout the Dremio platform including the Columnar Cloud Cache which accelerates Object Storage performance

9. Are you using some data-centric approaches at the Dremio for data quality enhancement?

Answer: These two presentations from SSubsurface Live 2023 should give great detail on this subject

[DataOps in Action](#)
[CI/CD on the Lakehouse](#)

10. What happens when a query fails? if a node of a query fails..

Answer: It fails and the jobs page will give you a detailed breakdown to understand why

11. Does Dremio come with a good enough native ETL capability or do you need a third party ETL tool to create complex ETL tasks?

Answer: Tools like Fivetran and Airbyte are probably tools that should handle the ingestion story then use Dremio to handle access, governance, documentation, acceleration, etc. Although, Dremio ingestion capabilities are growing with full Iceberg DML.

12. I am new to this field and how I can start by downloading Dremio. Any link plz?

Answer: Here is a guide on how to setup a local environment with Dremio and Minio so you can evaluate and learn Dremio through its community edition.

<https://github.com/developer-advocacy-dremio/quick-guides-from-dremio/blob/main/icebergminiodremio.md>

13. Is it realistic to replace the need for OLAP cubes of a EDW with all the layers of virtualized views?

Answer: Yes, the primary use case of Data Reflections is to achieve the same type of acceleration a cube would but directly on your data lake without data copies that need to be maintained and governed.