

Databricks vs. Snowflake: Clash of the Datalakes

October 21, 2026

Priya Stanley, Chief Technology Officer

Rob Robertson, President

Pat Tighe, Head of Product

WHAT'S INCLUDED

- Episode Overview
- Warehouse vs. Lake vs. Lakehouse: The Core Concepts
- Where Each Platform Started, and Where It's Headed
- Which Platform Fits Which Firm
- Why Some Firms Use Both
- Accessing Dakota Marketplace Data in Snowflake and Databricks
- Key Takeaways
- Frequently Asked Questions

EPISODE OVERVIEW

*In this webinar, Priya Stanley, Rob Robertson, and Pat Tighe broke down the two platforms investment firms keep asking about: **Databricks** and **Snowflake**. Priya covered the technical mechanics, Rob brought in what Dakota is hearing from clients, and Pat moderated and translated it into practical guidance for the data and technology leaders running infrastructure at investment firms.*

The framing that anchored the session: Snowflake is a data warehouse world evolving toward AI, and Databricks is a data lake world that built the lakehouse and is now pushing further into AI on top of it. Same destination, different starting point. And wherever a firm lands in the clash, Dakota's data can meet them there.

WAREHOUSE VS. LAKE VS. LAKEHOUSE: THE CORE CONCEPTS

Data Warehouse

- Stores structured data in rows and columns, like transaction records or sales figures.
- Data is cleaned and organized before it goes in, known as schema-on-write.
- Fast, reliable SQL queries that are easy to report against.

***Examples:** Snowflake is the cleanest example, along with Amazon Redshift and Google BigQuery.*

Data Lake

- Stores any type of data: structured, semi-structured, or unstructured (text, images, video) at massive scale.
- Structure is applied when someone queries it, known as schema-on-read.
- The flexibility is powerful, but without discipline, a lake can turn into a “data swamp.”

***Examples:** Amazon S3 and Azure Data Lake Storage.*

Data Lakehouse

- Combines the flexible structure of a lake with the reliability and fast query performance of a warehouse, all layered on the same raw storage.

***Examples:** Databricks is the cleanest example, and Microsoft Fabric spans both categories.*

The Practical Shorthand

- As Pat summarized it: when data comes in from a provider like Dakota, unstructured data that needs more cleaning goes to a data lake, while data that is already structured and ready to use goes straight to a warehouse.

WHERE EACH PLATFORM STARTED, AND WHERE IT'S HEADED

Origins

- Snowflake started as a data warehouse built for structured data, fast SQL, and BI dashboards.
- Databricks started as a lakehouse built on Apache Spark, designed for heavy, messy, large-scale unstructured data with machine learning running directly on it.
- Simplest version: Snowflake is for the analytics team. Databricks is for the data science and engineering team.

How Both Platforms Have Evolved

- **Snowflake adopted Apache Iceberg**, an open format that gives firms more flexibility to use their data across other tools.
- **Databricks added Databricks SQL** and more BI-friendly tooling, making it easy for analysts, not just engineers, to work in the platform.
- Both companies have focused heavily on AI over the last two years, and today they compete head to head. That's what makes this a real clash instead of two products in separate lanes.

WHICH PLATFORM FITS WHICH FIRM

Where Databricks Tends to Win

- Risk and model governance. Priya pointed to Morgan Stanley running regulatory risk calculations on Databricks, and to firms using Unity Catalog to trace exactly what data fed into an investment decision.
- Real-time work. Use cases like fraud detection, with Capital One cited as an example.

Where Snowflake Tends to Win

- Regulatory reporting and compliance. Priya cited BlackRock's Aladdin platform, where the risk analytics and reporting run on Snowflake. Snowflake governs access so portfolio managers, risk, and compliance teams can all work off the same data without stepping on each other's toes.

The Toss-Up: Quant Research

- Simpler, SQL-based research teams lean toward Snowflake. Heavier ML operations teams lean toward Databricks.

What Dakota Is Seeing

- Rob noted that most of what Dakota hears from its clients is Snowflake, with Databricks also in the mix.

WHY SOME FIRMS USE BOTH

Where Databricks Tends to Win

- We see this pattern most often at larger firms, which have the infrastructure for a technical team living in Databricks and an analytics and business insights team living in Snowflake.
- Priya added that the split often comes down to what kind of governance a firm needs:
 - **Tracking what goes in (Databricks):** Unity Catalog lets teams trace what data was fed into, for example, an AI model, and what the outcome was. Built for model risk and auditing.
 - **Controlling who sees what (Snowflake):** governance is centered on what each user has access to.

ACCESSING DAKOTA MARKETPLACE DATA IN SNOWFLAKE AND DATABRICKS

Meeting Firms Where They Are

- Dakota's data is already available directly through both Snowflake and Databricks, and can be extended to whatever platform a firm runs.
- Rather than handing over one generic dataset, Dakota tailors the data to what each firm actually needs.

Private Share vs. Marketplace Listing

- A standard marketplace app comes with a predefined set of accounts, contacts, and records that may or may not fit a given firm.
- With a private share, Dakota builds a tailored dataset under a mutually agreed scope. For example, a firm that only cares about RIA accounts and contacts gets exactly that, and nothing else. This works on both Snowflake and Databricks.

Data Sharing vs. the Raw API

- **Raw API:** everything is built end to end upfront. You need developers, custom code, and a plan for every scenario behind the scenes.
- **Snowflake or Databricks share:** if your workflows already run on one of these platforms, Dakota's data drops directly into that environment, ready to plug into what you've already built.
 - The Snowflake Dakota Marketplace integration is "as easy as downloading Spotify onto your iPhone"

Want to see Dakota Marketplace data running in your Snowflake or Databricks environment?

dakota marketplace

See How It Works

KEY TAKEAWAYS

- **Same destination, different starting points.** Snowflake is a warehouse evolving toward AI. Databricks is a lakehouse pushing further into AI. Both now compete head to head.
- **Snowflake fits analytics-first firms** that want clean, structured data that is easy to access and report on.
- **Databricks fits technically advanced teams** with machine learning needs and the discipline to manage unstructured data systematically.
- **For the largest firms, using both is the ultimate unlock**, with engineering and data science in Databricks and analytics in Snowflake.
- **Dakota meets you wherever you sit in the clash.** Tailored private shares in Snowflake or Databricks skip the custom build that a raw API requires.

FREQUENTLY ASKED QUESTIONS

- **What's the difference between a data warehouse and a data lake?** A data warehouse stores structured data that is cleaned and organized before it goes in (schema-on-write), making it fast and easy to report against. A data lake stores any type of data at a massive scale, with structure applied when it is queried (schema-on-read). A lakehouse, like Databricks, combines both.
- **Should my firm use Snowflake or Databricks?** It depends on your team and use case. Snowflake is generally the better fit for analytics teams and BI-driven reporting. Databricks is generally the better fit for data science and engineering teams running machine learning on large, unstructured datasets. Many of the largest firms use both.
- **Is Dakota's data available in Snowflake and Databricks?** Yes. Dakota Marketplace data is available directly through both Snowflake and Databricks, and Dakota can extend access to whatever platform your firm runs.

- **Can I get only the data I need, rather than the full dataset?** Yes. Through a private share, Dakota tailors the dataset to your firm's needs. For example, a firm focused only on RIAs can receive just RIA accounts and contacts.
- **How is a Snowflake or Databricks share different from Dakota's API?** The API requires developers and custom code built end to end. A Snowflake or Databricks share delivers Dakota's data directly into the environment and workflows your team already uses, with no custom build required.
- **Who can I talk to about getting Dakota's data into our data platform?** [Book a demo](#) with the Dakota team, including Torben, who leads Dakota's API business, to talk through how other investment firms are connecting Dakota's data to their stack.

dakota marketplace

[See How It Works](#)