

AMERICAS TECHNOLOGY: SOFTWARE

Thoughts on the Software Ecosystem post Databricks DAIS (SNOW, PLTR, MSFT, Apps)

We met with several industry experts at and around the Databricks' Data & AI Summit in San Francisco, 6/15-6/19. Databricks presented several architectural ideas that we believe are key to where the Software industry is going, including: a) the importance of open standards and easy data ingestion such that customers can maintain good data hygiene and a single source of truth, and drive better quality AI outcomes; b) an emerging TAM in custom agentic apps tailored to enterprise-specific use cases that sit in the white space between classic SaaS systems, in turn supported by a unified operation system/ontology layer with governance and model controls. Our conversations suggest that the value in the software stack will increasingly accrue to vendors that can enable these two architectures: **today, SNOW, PLTR & Databricks; and over the medium term, potentially MSFT, NOW and app vendors that figure out how to monetize headless.**

Gabriela Borges, CFA
+1(212)902-7839 |
gabriela.borges@gs.com
Goldman Sachs & Co. LLC

Maura Hager
+1(212)9028724 | maura.hager@gs.com
Goldman Sachs & Co. LLC

Selina Zhang
+1(212)357-9979 |
selina.zhang@gs.com
Goldman Sachs & Co. LLC

Goldman Sachs does and seeks to do business with companies covered in its research reports. As a result, the firm may have a conflict of interest that could affect the objectivity of this report. Investors should consider this report as only a single factor in making their investment decision. For more information on our research certification and other important disclosures, see the Disclosure Appendix, or go to www.gs.com/research/hedge.html. Analysts employed by non-US affiliates are not registered/qualified with FINRA in the U.S.



Key takeaways:

1. **Good sticky vs. bad sticky:** We have been using a simple litmus test to assess the durability of Software moats: “good sticky” companies are innovating at an accelerating pace on core differentiators and have products that their customers love. “Bad sticky” companies may have innovated less over the last several years, but are sticky because of the costs associated with migration, and generally have high levels of customer complaints or low levels of customer engagement and usage. We expect bad sticky moats to be eroded, in part because it is now much easier to use AI to streamline migrations. In our view, Databricks’ pace of innovation coupled with their push towards open ecosystems raises the bar on what it means to be a “good sticky” company:
 - While pace of innovation cannot be measured by the number of product releases alone, several industry experts highlighted that Databricks announced several product releases that are significant in the pain points they address, such as LTAP (which allows analytical and transactional engines to operate from a single open-format source of truth) and opensharing (an open protocol for sharing data, models, agents and skills across any cloud/vendor).
 - Today, it is clearly in Databricks’ interest to lower barriers to switching to Databricks from legacy data management platforms; however, the company also acknowledges customer freedom such that if a more innovative non-Databricks solution arrives, it will also be easy to run a multi-vendor ecosystem or switch off Databricks. This likely creates an upward bias to Databricks’ R&D and M&A intensity (Databricks plans to manage to FCF breakeven), but it also drives the company’s strategy to be aggressive in how it disrupts adjacent TAMs (like SIEM and CDP) and leverages moats like Unity Catalog into new markets (such as governance for agents). This in turn limits the potential for new entrants in the future. In the investor session, Databricks noted that it could choose to price 2x higher, but that short-term pricing optimization decisions often come at the expense of long-term competitiveness and LTV.
 - We look for similar characteristics to identify “good sticky” companies in our publicly-traded coverage: assets like **Shopify** and **Cloudflare** stand out to us: their pace of innovation is high and their pricing is oftentimes disruptive with the goal of long-term market share momentum (consider Shopify’s comments that Shopify Plus is the best deal in enterprise software, and Cloudflare’s ability to price aggressively in SASE because of the low margin cost to serve, as SASE works via forward proxy and uses the same rails as CDN does for reverse proxy).
2. **Custom vs. packaged apps and the blurring lines between infra and app software:** Databricks’ Day 1 keynote introduced tools to build a new agentic app layer: while SaaS apps have historically existed in silos, the most interesting agentic applications are likely to exist in the white space between the silos. This lends itself to custom apps for business-specific use cases that capture a greater share of the agentic software TAM, and thus the entire Software TAM, than they have in the past

(recall we have previously quantified a Software TAM uplift of 20% by 2030, implying a 2pt tailwind to industry growth, from 7% to 9%). In order to enable this new agentic app layer, customers need to build an intelligent operating system layer that pulls data from systems of record (in a headless architecture) and other enterprise data stores.

- **As best we understand, there are four key ways to build this OS layer:** a) **Databricks'** new Ontology product, that includes functionality like OntoRank to prioritize data sources based on context; b) **Palantir**, which in turn leverages Databricks and Snowflake as building blocks, and moves the burden of engineering entirely from the customer to Palantir's FDE team; c) **hyperscalers**; recall Satya Nadella's June 14 post on frontier ecosystems, that emphasizes that every enterprise should have its own version of frontier intelligence that marries frontier models with their specific enterprise data; d) select AI point products, such as **ServiceNow's** agent control tower or privately-held **Dataiku**. We suspect the line of demarcation between custom and packaged will likely hinge on organizational complexity: more complex orgs and apps will likely be custom built; less complex including SMBs will likely leverage SaaS off the shelf offerings like Salesforce Agentforce or ServiceNow Control Tower. Given how early the agentic Software TAM is, we expect a rising tide to lift many boats; however, our conversations over the last 3 weeks (including at the SNOW conference) make us incrementally positive on custom apps as a driver of underlying consumption for Databricks, Snowflake and Palantir.
 - **Databricks is also directly addressing the cost of maintaining a custom app, in the context of the cost of building already having collapsed with coding tools**. Recall that the cost of building an app is only one part of the equation; an organization also has to maintain their custom app in production, which ServiceNow has previously quantified as having a 5-10x TCO vs. SaaS. Many of the innovations that Databricks announced directly lower this maintenance TCO. For example, Unity Catalog acts as a universal governance layer embedded in ontology; ontology leverages a single source of truth; and LTAP collapses the need for otherwise brittle pipelines and ETL (extract, transform, load) processes between different data sources.
 - **On competition vs. Palantir**, Databricks noted a generally complementary relationship, in part because of Palantir's strong FDE motion and Palantir pulling through usage of Databricks. Based on our customer conversations, the biggest question is whether you are willing to trade off outsourcing the heavy lift on building the ontology for the risk of longer term lock in with Palantir, as Palantir's pricing model is recurring, but the value provided is weighted up front to its original ontology build. To this end, Databricks noted it will sometimes engage with Palantir customers upon their renewals who look to rebuild agentic workflows with Databricks.
- 3. New CDP and SIEM apps:** Databricks described its evolution to date in four chapters: 1) Lakehouse; 2) next-gen Database products; 3) Agentic AI; and 4) agentic apps. Databricks has announced two Chapter 4 apps to date: SIEM (at RSA in March)

and CDP (customer data platform at DAIS last week), and sees natural expansion opportunities around these two vectors of security and marketing. Our conversations suggest that customers were already using Databricks as the cornerstone of their custom SIEM and CDP strategies; these announcements commercialize what more sophisticated customers were already doing. We believe these product offerings are likely to be compelling because they directly solve problems around data ingestion and aggregation; however, our customer conversations suggest it will likely take time for Databricks to build out its domain specific experience.

- For example, in **Security**, in addition to solving the key cost problem around data ingestion, Databricks may also need to introduce more sophisticated event streaming and correlation functionality, and build out its go to market. Recall that Cloudflare took several years to evolve its Security go to market before seeing more success in the 2025 time frame; and that Datadog is still in the process of crystallizing its security strategy after mixed traction in cloud security (per our customer conversations).
- We think milestones in **CDP** may come faster, as switching costs in marketing are generally lower than in Security, and customers can map good CDP functionality directly to more effective revenue generation.

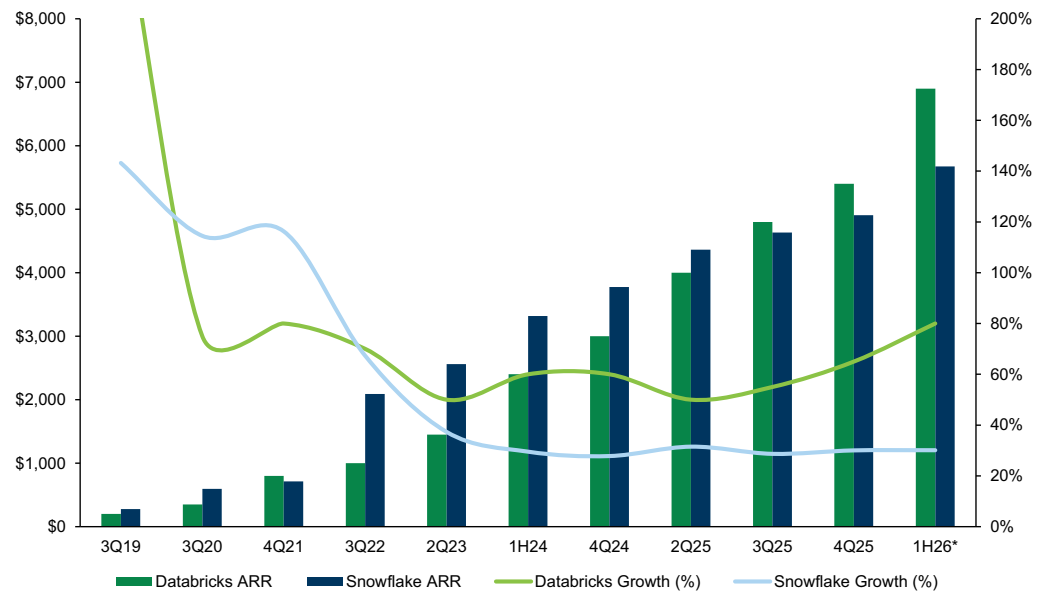
4. When describing its product innovation, Databricks noted that while the rest of the industry is focused on agentic, it sees an opportunity to further advance the database technologies that in turn will support agentic. All of the new software being written today needs a database, and Databricks can provide database technology that takes outsized share in this greenfield TAM. Examples include LTAP, which solves a 40-year problem in database engineering with a unified data stack that spans both OLTP and OLAP databases; and Reyden, a new compute engine that powers real-time lakehouse queries. Databricks' product expansion aims to collapse a fragmented database and data infrastructure market, where enterprises have used separate tools for transactional databases, ETL, warehouses, BI, governance, and AI; into an integrated platform built on open formats, shared governance, and a common data layer. With the combination of product-market-fit and new innovation, Databricks' core revenue growth (excluding token pass through) has accelerated for the last 5 quarters. Key product announcements:

- **OLTP/LTAP via Lakebase:** LTAP (Lake Transactional/Analytical Processing) represents a strategic move beyond Databricks' historical OLAP core into OLTP. With LTAP, Databricks now positions its platform as a single governed foundation spanning transactional, analytical, streaming, and operational data, structurally reducing reliance on ETL, replicas, and CDC pipelines that have long separated application databases from analytics systems. Lakebase is the architectural unlock: a serverless Postgres layer on open object storage that enables low latency operational workloads while maintaining shared, open data formats with Lakehouse. This expands Databricks' TAM while improving its competitive positioning to capture operational workloads that have historically sat outside its core competencies.

- **Postgres for agents:** Databricks framed Postgres as increasingly central in an agent-driven world, where workloads operate at machine speed and require cost efficient, deterministic, and inherently branchable environments. Lakebase introduces primitives such as near-instant database branching, enabling agents to safely fork datasets, experiment, write changes, and revert to prior states as needed. In our view, the emphasis on Postgres is critical, as agentic applications require not just read access to enterprise context, but governed, low latency transactional systems where agents can act safely, at scale, and with repeatability.
- **Data engineering: Lakeflow:** New innovations in Lakeflow focus on minimizing upstream complexity through a unified, declarative framework for ingestion, transformation, and orchestration built natively on Spark. Lakeflow abstracts pipeline creation via no code and declarative constructs, while maintaining open, non proprietary execution under the hood. The innovation shifts toward operationalizing data engineering at scale: automated orchestration across >50 integrations, and tight integration with agents via Genie Code and Genie Ops. These innovations are important in alleviating data engineers' ongoing maintenance burden, which has historically consumed a disproportionate share of data engineering resources.
- **Data science/AI:** Databricks' core lakehouse and data science positioning remains central as it provides the enterprise context layer for AI. The introduction of Genie Ontology and OntoRank create a structured graph across enterprise data, documents, and usage patterns, improving agent accuracy by 30% and reducing runtime by 50%. The broader implication is that Databricks is capturing incremental spend in AI infrastructure, semantic layers, and enterprise knowledge systems, beyond traditional ML workloads.
- **Data warehousing:** Databricks continues to expand beyond its roots in engineering and ML into a more comprehensive lakehouse architecture, anchored by its next generation compute engine, Reyden, which is designed to support both complex analytical queries and highly interactive workloads. The launch of Genie One, an agentic coworker, provides conversational analytics on data stored in Databricks, embedding analytics into natural language and automated decision flows.
- **Real-time analytics:** Databricks launched Lakehouse//RT, a real-time Lakehouse powered by a new compute engine built for concurrency and latency demands of modern agentic enterprises (Reyden, discussed above). Historically, enterprises needed a distinct real-time serving layer, which introduced duplication, latency, governance tax, and higher costs; Lakehouse//RT eliminates this layer by querying live data in place under a single governance model. On standard analytical benchmarks, Databricks found sub-100ms latency at 12,000 queries per second, and noted customers have seen up to 16x better performance than existing specialized real-time serving stacks.

Exhibit 1: Databricks core growth has accelerated over the last 5 quarters as a function of its product market fit with AI; Snowflake is also starting to show acceleration, albeit to a lesser extent

*Databricks expects ARR to grow 80% in 1H27, or ~65% when excluding token pass through



Source: Company data, Goldman Sachs Global Investment Research