

Databricks: Growth Accelerated by the AI Wave

Data is becoming the production input of AI, extending platform value from compute to context, governance and APIs

OUTRIDGE

INVESTMENT OVERVIEW

Data is the most durable infrastructure demand created by AI

Databricks is evolving from a data engineering platform into the context, governance and execution layer for enterprise AI.

\$7bn+ annualized revenue run-rate	>80% Q2 2026 YoY growth	20,000+ global customers	\$190bn latest primary valuation
--	--------------------------------------	------------------------------------	--

Our core thesis

AI will not reduce enterprise data problems. It will turn data from an analytical input into a production resource continuously consumed by models and agents. Front-end products may take many forms, but every inference, retrieval, decision and action still depends on APIs that access organized, permissioned and interpretable data. Databricks can therefore expand from processing data faster to making enterprise data safely usable by AI.

- Growth is already visible: revenue run-rate increased from roughly \$3bn at the end of 2024 to more than \$7bn by August 2026, while growth accelerated.
- The product boundary keeps expanding from Spark engineering into SQL, governance, models, agents, AI gateways and the Lakebase transactional database.
- Customer quality is high: more than 1,000 customers spend over \$1m annually and more than 100 spend over \$10m. [1]
- Valuation already embeds high expectations: the latest primary valuation equals roughly 27x current run-rate revenue.

OUTRIDGE VIEW

A 10x to 100x increase in multimodal data is a directional scenario, not a forecast. As agents continuously read, generate and update enterprise data, governance and API call density should rise well beyond the traditional BI era.

BUSINESS MODEL

A consumption-priced platform for data and AI

Databricks does not disclose revenue by product. The mix below is an Outridge estimate based on product maturity and enterprise workload patterns.

Apps & agents	Genie / Apps / AI Agent / BI
AI & models	Mosaic AI / MLflow / Model Serving
Data & governance	Lakehouse / SQL / Unity Catalog / Lakeflow
Cloud infrastructure	AWS / Azure / Google Cloud

Estimated revenue mix midpoints sum to 100%

Engineering / ETL 40%	SQL / BI 30%	AI services 20%	Other 10%
-----------------------	--------------	-----------------	-----------

Workload	Estimated mix	Monetization and economics
Data engineering / ETL	35%-45%	DBU consumption; customers bear cloud costs in Classic
SQL / warehouse / BI	25%-35%	SQL and serverless consumption; Snowflake's core market
ML / AI services	15%-25%	Training, inference and serving; faster growth, heavier GPU cost
Streaming / governance / other	10%-20%	Streaming, Catalog, Apps and Lakebase products

OUTRIDGE VIEW	The mix is not company-reported. The key question is not a single SKU, but whether AI causes the same data to be moved, queried, trained on and called far more often.
---------------	--

BUSINESS EVOLUTION

From commercializing Spark to controlling enterprise AI

Each product expansion moves Databricks closer to the final application and its business context.



FUNDING AND VALUATION

Capital expanded the product boundary as valuation tripled in 20 months

Total equity funding is approximately \$25.55bn. Recent rounds also provided employee liquidity and acquisition capacity.

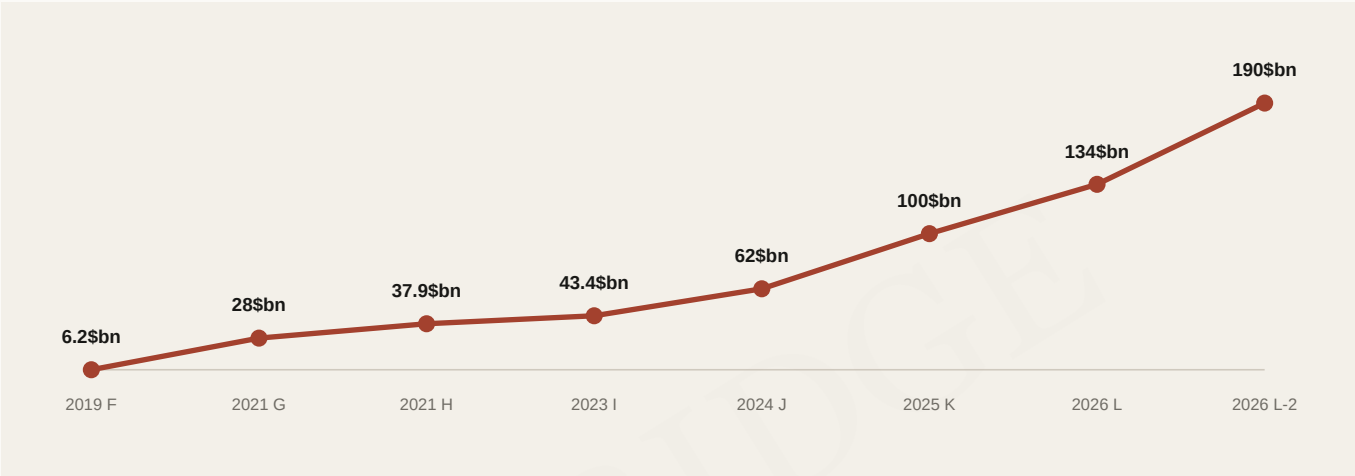


Figure 1: Post-money valuation of major funding rounds. Sources: Forge and company announcements. [1][4]

Round	Capital raised	Post-money value	Price / share
2023 Series I	\$0.68bn	\$43.4bn	\$73.50
2024 Series J	\$10.18bn	\$62.0bn	\$92.50
2025 Series K	\$1.13bn	\$100bn	\$150.00
2026 Series L	\$5.0bn	\$134bn	\$190.00
2026 L-2	\$5.0bn	\$190bn	\$253.00
Secondary reference	Forge Price	about \$202bn implied	\$268.57

OUTRIDGE VIEW Primary valuations reflect preferred shares, while secondary quotes may represent common shares or SPV interests. Rights, fees and liquidity are not directly comparable.

CUSTOMERS AND NEEDS

Customers are not analyzing one table - they need complex data to keep working

The typical buyer is a large enterprise data platform, AI/ML or technology leader rather than a single business team.

20,000+ global customers	60%+ of the Fortune 500	1,000+ customers spending >\$1m	100+ customers spending >\$10m
------------------------------------	-----------------------------------	---	--

Customer type	Core problem	Why Databricks
Financials / pharma	Sensitive data, many permissions, strict audit	Unified governance, lineage and model evaluation
Retail / platforms	Massive behavior data and real-time decisions	Streaming, features and personalization
Manufacturing / auto	Sensors, images and predictive maintenance	Large-scale unstructured and time-series data
Gaming / media	High-volume events and recommendations	Spark, real-time analytics and ML
AI-native companies	Multiple models, RAG, agents and cost control	Data, models and permissions on one platform

Poor fit

- Early-stage companies with small datasets and simple systems.
- Organizations without data engineering or AI teams and no plan to build a unified platform.
- Needs that can be met more cheaply by PostgreSQL, BigQuery and lightweight BI.

OUTRIDGE VIEW	Databricks is not selling a replacement for Excel. It turns data scattered across clouds, databases, files and applications into a governed asset that people and agents can jointly consume.
----------------------	---

USE CASE

E-commerce: from clickstream to real-time recommendations and shopping agents

Databricks connects customer, product, order and content data to support real-time operations and personalization.

01 Ingest

Clickstream, search, orders, ads, service chats, images and model logs

02 Unify

Customer IDs, SKUs, channels, inventory, prices and promotion definitions

03 Analyze

GMV, conversion, repeat purchase, fulfillment, ad ROI and unit economics

04 Model

Recommendations, search ranking, demand forecasts, churn and fraud

05 Agent

Use catalog, inventory and preferences for shopping and after-sales service

06 Govern

Consent, sensitive data, merchant isolation, audit and deletion

The call chain behind one AI shopping request

Identity

>

Catalog

>

Inventory

>

Recommender

>

Conversion

OUTRIDGE VIEW

The hard part is not generating a product sentence. It is reading the right customer, product, inventory and price in milliseconds, then writing every exposure, choice and purchase back to a unified data system.

COMPETITIVE LANDSCAPE

Competition is shifting from products to the entire enterprise data stack

Snowflake is the closest product competitor. Microsoft, AWS and Google are the longer-term platform risks.

Group	Control point	Advantage	Threat to Databricks
Snowflake	SQL warehouse and data sharing	Simple, stable, low operations	Expanding into engineering, AI and open lakehouse
Microsoft	Azure, Fabric and Power BI	Enterprise channel and bundling	The strongest integrated platform competitor
AWS	S3, Redshift and SageMaker	Infrastructure scale and pricing power	Can native-build the middleware layer
Google	BigQuery, Vertex AI and Gemini	Analytics and model capabilities	Complete end-to-end AI stack
Open source	Spark, Iceberg and Trino	Open, composable, low license cost	Compresses compute and format premiums

Partnership and competition happen at the same time

- Databricks runs on all three hyperscalers and is both a valuable workload and a co-selling partner.
- Cloud vendors also control infrastructure pricing, customer contracts and native products, which can compress Databricks' bargaining power.
- The durable defense is not one feature, but multi-cloud neutrality, open formats, developer ecosystems and embedded workflows.

OUTRIDGE VIEW

Databricks meets Snowflake most often in direct deals. The larger long-term variable is whether hyperscalers can bundle data, models, productivity distribution and procurement contracts.

DATABRICKS VS. SNOWFLAKE

One moved from engineering into warehousing; the other from warehousing into AI

Their boundaries are converging, but organizational DNA, user mindshare and strongest workloads remain different.

Dimension	Databricks	Snowflake
Traditional strength	Data engineering, Spark, Python and ML/AI	SQL, BI, data warehousing and sharing
Typical user	Engineers, data scientists and AI teams	Analysts, warehouse and business data teams
Product experience	Flexible and broad, but more complex	Fully managed, easier to adopt and operate
Data architecture	Cloud object storage plus Delta / Iceberg	Snowflake storage plus Iceberg support
AI position	Models, RAG and agents across the lifecycle	Fast path from structured data into Cortex AI
Core moat	Developer ecosystem, workflows and governance	Ease of use, sharing network and reliability

Scale and valuation snapshot



OUTRIDGE VIEW Databricks has broader optionality, but large models will not automatically erase Snowflake's strengths. AI lowers the SQL and BI barrier and may make Snowflake easier for non-technical users to consume.

HYPERSCALER COMPETITION

The largest structural risk is a partner becoming the full-stack platform

Databricks buys compute and uses cloud distribution while competing with the same vendors' native data products.

Microsoft Fabric + OneLake Power BI + Copilot The strongest enterprise entry point. Azure and Office contracts can bundle the stack, even as Azure Databricks remains a deep partnership.	AWS S3 + Redshift + Glue SageMaker Controls storage and compute pricing and offers the broadest cloud-native substitute portfolio.	Google BigQuery + Dataplex Vertex AI + Gemini Strong analytics and models for customers that prefer to remain entirely inside GCP.
---	--	--

Databricks' response

- Remain multi-cloud neutral and reduce customer dependence on any single vendor.
- Use Delta, Iceberg, Spark and MLflow to preserve an open developer entry point.
- Move upward into Unity Catalog, AI Gateway and business semantics to control how agents access enterprise data.
- Use large cloud commitments to secure discounts while managing utilization and take-or-pay risk.

OUTRIDGE VIEW

Microsoft may be the most dangerous long-term competitor. It does not need every product to be superior if Azure, Fabric, Power BI, Excel and Copilot are cheap and convenient as a bundle.

AI OPPORTUNITY: DATA VOLUME

AI first expands data turnover, not the number of models

Text, voice, images, video and agent action logs will all enter enterprise data systems.

Multimodal data growth scenarios - directional, not forecasts

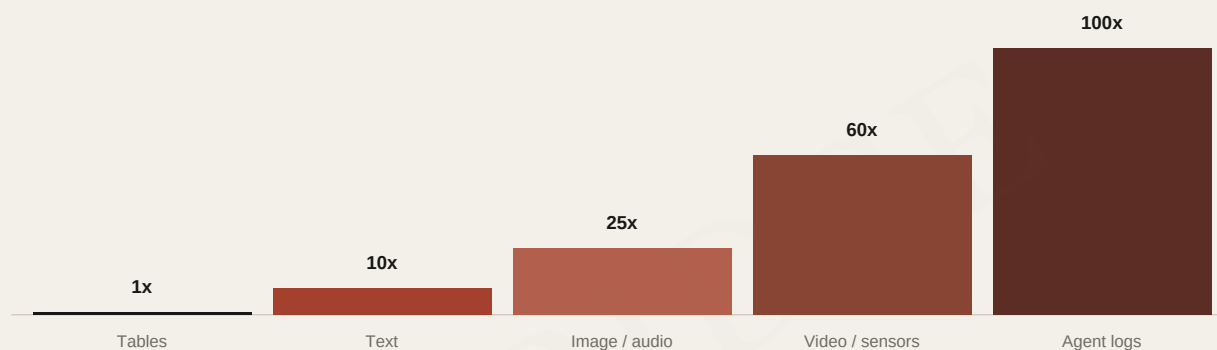


Figure 2: Outridge scenarios. Multiples illustrate relative magnitude rather than a precise market forecast.

Why data can grow faster than user count

- One user request can trigger retrieval, model routing, tool calls, evaluation and multiple write-backs.
- Images, audio and video are much larger than tables and also require embeddings, labels and intermediate features.
- When agents move from answering to acting, every step creates state, permission and audit data.
- Enterprises must retain not only the answer, but versions, sources, cost, quality and accountability.

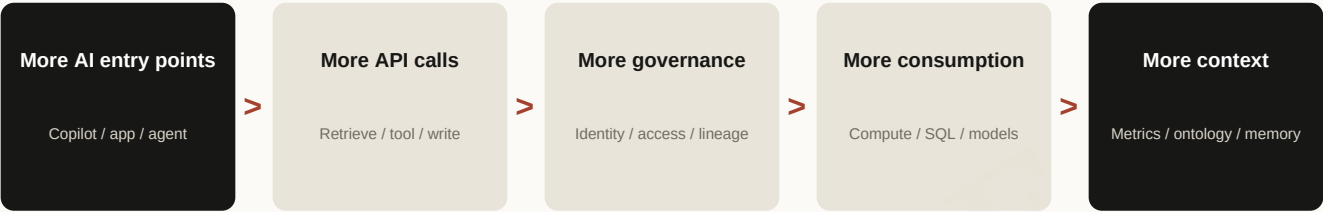
OUTRIDGE VIEW

The AI opportunity is not storage price multiplied by data volume. It is the same data being called more frequently by more models under more complex permissions. Call density may matter more than capacity growth.

AI OPPORTUNITY: API VALUE

Front ends may differ, but API and context value will expand

End users may not see the infrastructure, yet every agent must find data, understand meaning and obtain authorization.



Traditional BI era	Agent era	Value migration
People open dashboards	Agents continuously read and write	From seats to calls
Fixed tables and metrics	Dynamic multimodal context	From warehouse to semantics and APIs
Weekly or monthly refresh	Seconds or event-driven	From batch to real time
People own judgment	Systems need traceable decisions	Governance and audit become more valuable

OUTRIDGE VIEW The scarce asset in the AI era is not only the model, but enterprise context that machines can consume safely. Control over data APIs, permissions and business semantics can create infrastructure value across many front ends.

AI CHALLENGE: ABSTRACTION

The same AI capability can also flatten platform differentiation

As models understand modalities, generate pipelines and select tools, users may care less about the underlying platform.

Capability at risk	How AI abstracts it	Impact on Databricks
SQL and dashboards	Natural language creates queries, explanations and charts	BI interfaces and development aids commoditize
Cleaning and mapping	Models infer field meaning and align schemas	Some engineering labor and tool premiums decline
Pipeline development	Agents generate, test and repair ETL	Complexity no longer guarantees platform stickiness
Multimodal interfaces	Text, images and audio become model inputs	Formats and connectors disappear behind the front end
Multi-platform routing	Agents select the cheapest storage and compute	Engines may become replaceable resources

Layers that remain difficult for models to replace

- Deterministic storage, transactions, consistency and distributed computing at scale.
- Real identities, access controls, privacy policies and regulatory audit.
- Production reliability, recovery, cost controls and accountability boundaries.
- Accumulated business definitions, data lineage and organizational workflows.

OUTRIDGE VIEW The largest risk is not that an LLM becomes a database. It is that LLMs make databases and compute engines invisible, interchangeable components. Databricks must own governance, context and the API control layer.

COST STRUCTURE AND LEVERAGE

Asset-light does not mean infrastructure-cost risk disappears

RAM, NAND and GPU inflation first enters hyperscaler capex, then flows through serverless costs and cloud commitments.

Business type	Who bears cloud cost	Marginal cost	Operating leverage
Classic DBU	Customer	Low	High
Governance / Catalog	Mostly control plane	Low	Very high
Serverless SQL / ETL	Databricks	Scales with usage	Medium
Model training / serving	Databricks or customer	GPU / HBM-linked	Medium-low
Professional services	Databricks labor	People scale with projects	Low

Key economic variables

- The company does not disclose an overall serverless utilization rate; Classic resources belong to customers and do not form a company-wide metric.
- Multi-tenant pools, autoscaling, Photon, Arm and procurement discounts can improve margin per task.
- Multi-billion-dollar commitments to all three clouds provide discounts but create minimum-spend and utilization risk. [10]
- Existing contracts and competition limit the ability to pass through all upstream cost increases immediately.
- Higher DRAM and enterprise SSD prices pressure serverless margins but do not require Databricks to build data centers.

OUTRIDGE VIEW

Investors should ask for serverless revenue mix, product-level gross margin, cloud-commitment utilization and net revenue retention. These determine how much profit can emerge from the \$7bn revenue base.

OUTRIDGE VIEW

The more invisible AI becomes, the more important data and APIs become

Databricks' long-term role is not a better data tool, but the control plane through which enterprise AI accesses the real world.

OUTRIDGE CONCLUSION

The largest demand increase in the AI era is data. Front-end forms may be completely different, but API value keeps expanding. At the same time, models can erase differences across modalities, interfaces and tools, making the platform invisible to users. Databricks' opportunity and risk come from the same question: can it become hidden infrastructure without becoming replaceable infrastructure?

PUBLIC COMPARABLE

Snowflake: Q2 FY27 product revenue +37% YoY; approximately 21x EV/Revenue as of Sep. 18, 2026.

Investment logic

- Growth durability: multimodal, real-time and agent workloads expand processing and governance demand.
- Platform optionality: SQL, ML, agents, Lakebase and AI Gateway widen the addressable market.
- Moat: developer ecosystem, data gravity, governance, business semantics and production workflows.
- Risks: cloud bundling, AI abstraction, serverless margins and valuation expectations.

Valuation scenario	Revenue at listing	Revenue multiple	Implied value
Bear case	\$12bn	15x	\$180bn
Base case	\$15bn	20x	\$300bn
Bull case	\$18bn	25x	\$450bn

OUTRIDGE VIEW

Snowflake's roughly 21x revenue valuation supports a 20x Base case reference. A 25x Bull case requires Databricks to remain materially faster at listing while expanding serverless margins and free cash flow.

APPENDIX

Methodology, data conventions and principal sources

All conclusions should be read in light of private-company disclosure limits, share-class differences and third-party data constraints.

Methodology

- Revenue refers to company-disclosed annualized revenue run-rate, not audited annual revenue or contracted ARR.
- Valuation multiples use equity value as an approximation because net cash, debt and fully diluted shares are not fully disclosed.
- Funding and share prices come from Forge and company announcements; secondary prices are indications, not executable quotes.
- Databricks does not disclose product revenue. The business-model mix is an Outridge estimate, not a company reporting segment.
- The 10x to 100x multimodal scenarios express direction and magnitude, not a market forecast.
- Snowflake data comes from SEC filings; company metrics are not perfectly comparable.

Principal sources

- [1] Databricks / MGX, \$5bn financing, \$190bn valuation and >\$7bn revenue run-rate, Aug. 13, 2026. publicnow.com/view/C384BD5053AA7476D5A32E5059BA4F30DAC77D7F
- [2] Databricks website, 20,000+ customers and Fortune 500 penetration, accessed Sep. 2026. databricks.com
- [3] Databricks Docs, platform, lakehouse, ETL, ML, Lakebase and governance documentation, updated Sep. 11, 2026. docs.databricks.com/aws/en/introduction
- [4] Forge, funding history, share prices, \$25.55bn total funding and \$268.57 Forge Price, Sep. 18, 2026. forgeglobal.com/databricks_stock
- [5] Snowflake SEC Q2 FY2027, \$1.55bn revenue, product revenue +37% and \$9.0bn RPO, Sep. 2, 2026. sec.gov/Archives/edgar/data/1640147/000164014726000033/fy2027q2earnings.htm
- [6] Nasdaq Private Market, Databricks \$264.01 indicative price, Sep. 4, 2026. nasdaqprivatemarket.com/company/databricks
- [7] Apache Spark History, Spark began in 2009 and entered Apache in 2013. spark.apache.org/history
- [8] Databricks / TechCrunch / Axios, MosaicML, Tabular and Neon acquisition materials, 2023-2025.
- [9] Databricks Docs, Classic and Serverless costs, resource management and billing. docs.databricks.com/aws/en/compute/serverless
- [10] TechCrunch, Databricks' multi-billion-dollar commitments to the three hyperscalers, Aug. 13, 2026. techcrunch.com/2026/08/13/databricks-wanted-to-raise-1b-investors-wanted-15b-it-settled-on-5b-at-a-190b-valuation
- [11] TrendForce, AI server demand and Q3 2026 DRAM / enterprise SSD pricing, Jul. 3, 2026. trendforce.com/presscenter/news/20260703-13134.html
- [12] Li et al., Unveiling Challenges for LLMs in Enterprise Data Engineering, 2025. arxiv.org/abs/2504.10950
- [13] Bloomberg Law, CEO says 2026 is not an ideal IPO year, Jun. 4, 2026. news.bloomberglaw.com/capital-markets/databricks-ceo-plans-to-avoid-ipo-during-year-of-huge-offerings
- [14] Snowflake valuation reference: approximately \$120bn EV and 21.0x EV/Revenue as of Sep. 18, 2026. multiples.vc/public-comps/snowflake-valuation-multiples

Disclaimer

This report is for internal Outridge research and discussion only. It is not a securities offering, investment advice or a guarantee of transaction pricing. Third-party data and private-company information may use different definitions. Readers should complete legal, financial, tax, technical and commercial diligence before investing. Contact: insight@outridgecap.com.