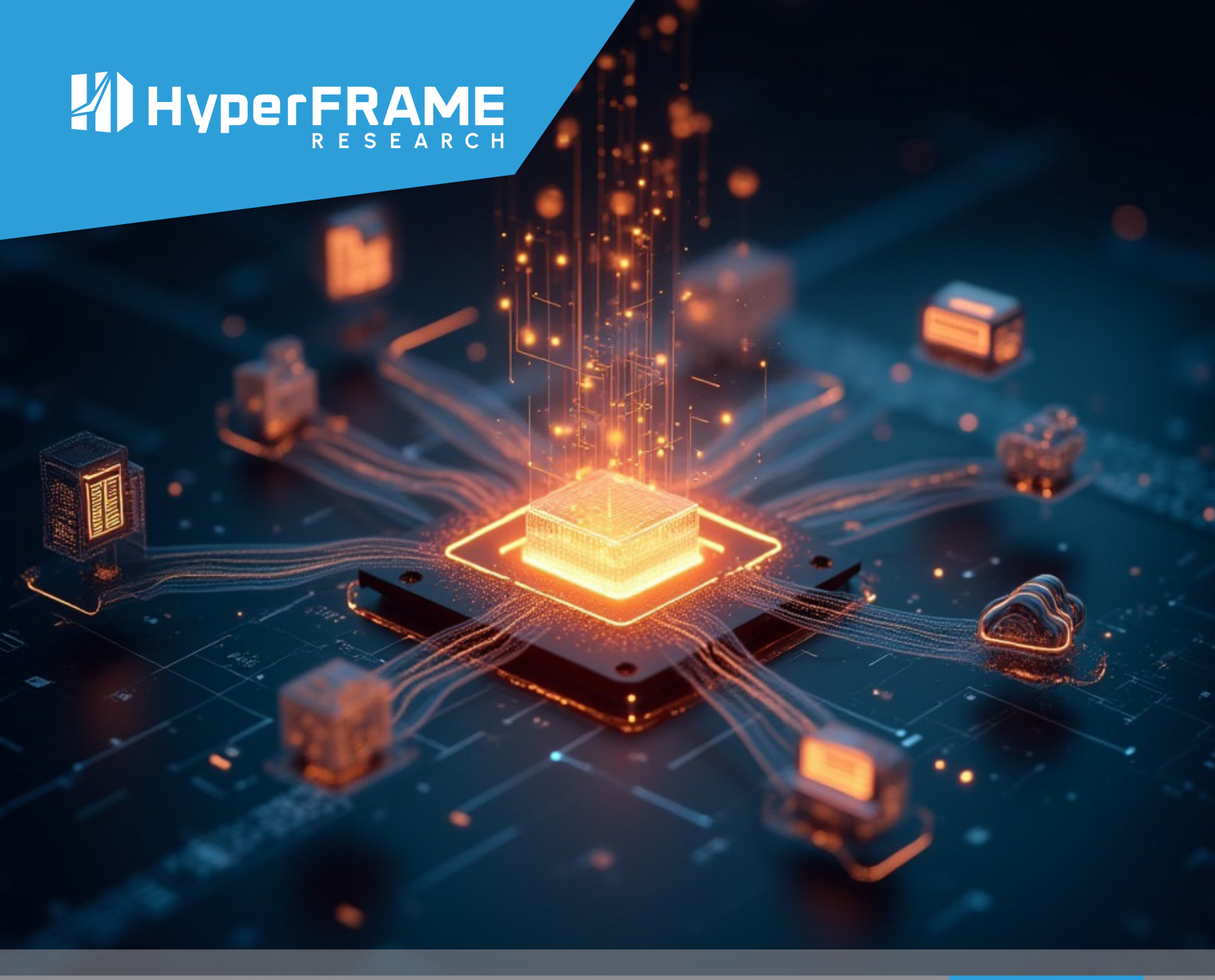
HyperFRAME RESEARCH



WHITE PAPER

Lakehouse Platforms: The New Foundation for Enterprise AI

Comparing Databricks, Snowflake, Teradata, and Cloudera on AI-Ready Data Architecture

Author:
Stephanie Walter
Analyst In Residence - AI Tech Stack

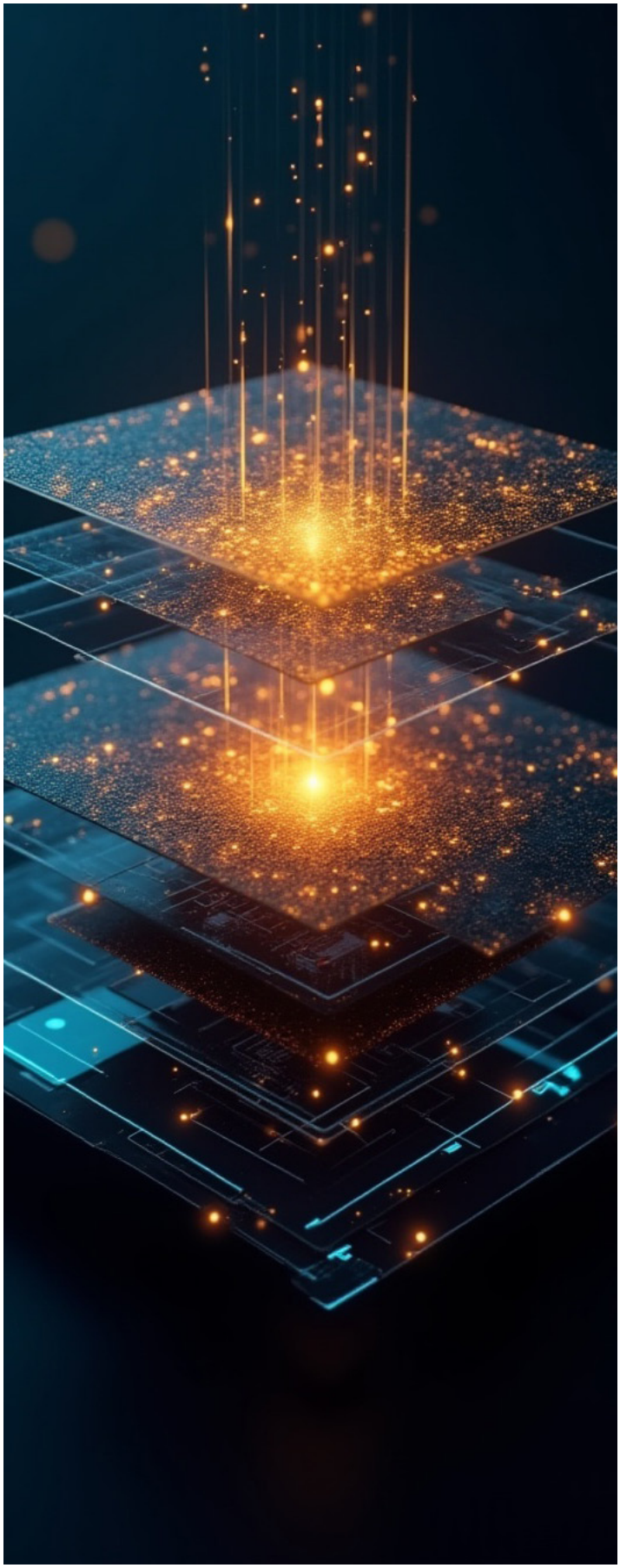
JULY 2025

Introduction

Enterprise AI has evolved from experimental prototypes to deeply embedded intelligence across business operations. AI is taking a front seat in business processes, applications, and decisions across every major industry, and the underlying enterprise data matters more than ever. However, historically, most data architectures were not built to support the scale, diversity, and speed that AI workloads require.

Data warehouses are typically used for structured data analysis. And while data warehouses are great for structured queries and reporting, they struggle with the volume and variety of data today. This is especially true for unstructured formats like text, images, and audio, which are heavily used in AI applications. Due to the limitations of data warehouses, data lakes were introduced for their unstructured and semi-structured data capabilities. Data lakes solve the volume and variety problems, but often lack the performance and governance needed for production-grade AI use cases. In addition, having both data warehouse and data lake systems can lead to data silos and complex ETL processes. Traditional data warehouses are fast and structured, but rigid. Data lakes offer flexibility, but lack governance and performance.

Enter the lakehouse to bridge the gap. Lakehouses aim to deliver a unified, scalable, and inherently AI-native environment. The goal of the lakehouse is to combine the best attributes of data lakes and data warehouses. The lakehouse is a single platform that can handle diverse data types, support both analytical and AI workloads, and maintain data quality and governance. It is architected to address the complexities of modern data landscapes and provides the foundation for enterprise AI adoption. The lakehouse has moved from an emerging architecture to a strategic foundation for AI-driven enterprises.



Defining the AI-Ready Lakehouse

Not all lakehouses are created equal. At their best, lakehouses provide a unified repository for all data types, a decoupled compute layer, native support for open table formats like Apache Iceberg or Delta Lake, and strong governance. However, what truly distinguishes an AI-ready lakehouse is its capacity to meet AI-specific requirements. A lakehouse should:

- Handle both structured and unstructured formats like images, audio, and documents.
- Support real-time streaming data.
- Integrate easily with model development, deployment, and monitoring.
- Enable cross-functional collaboration between data engineers, data scientists, and analysts in a unified environment.

The most significant evolution is support for agentic AI patterns that enable autonomous decision-making and complex workflow orchestration. These systems require platforms capable of supporting validator-critic loops, cascading memory agents, and domain mesh architectures.

Multimodal AI capabilities have become essential for enterprise adoption. The integration of text, image, audio, and video processing capabilities within lakehouse architectures enables more sophisticated AI applications and better contextual understanding across diverse data types. The incorporation of vector databases has also become critical for AI-ready architectures as generative AI applications rely more and more on vector databases for grounding AI foundation models.

This white paper takes a closer look at how Databricks, Snowflake, Teradata, and Cloudera are building toward this vision. The updated analysis incorporates the latest insights on the industry's leading lakehouse platform vendors, providing enterprise architects with the information needed to navigate this rapidly evolving landscape.

Vendor Landscape: Comparing Lakehouse Strategies

Databricks has arguably been the most vocal champion of the lakehouse concept, having coined the term and contributed Delta Lake to the open source community. Its platform is built on Apache Spark and integrates with Delta Lake for ACID-compliant table operations. Databricks enhances its offering with MLflow for model lifecycle management, Unity Catalog for cross-platform governance, and the MosaicML foundation model stack for generative AI capabilities. With strong real-time streaming through Structured Streaming and a focus on open standards, Databricks provides a comprehensive, cloud-optimized environment for data and AI teams.

Databricks has unveiled its most significant platform evolution with the introduction of Agent Bricks and Lakebase, announced at their 2025 Data + AI Summit. These innovations represent a fundamental shift toward comprehensive data intelligence platforms. Agent Bricks introduces a revolutionary framework for building enterprise AI agents with automated optimization. Rather than managing the complexity of agent development, teams can focus on defining their agent's purpose and providing strategic guidance through natural language feedback.

Lakebase represents a new class of operational database for AI applications and provides a fully-managed, Postgres-compatible database built for AI. It enables seamless data loading and transformation from over 300 sources. Lakebase stores operational data in low-cost data lakes with continuous autoscaling of compute to support agent workloads. Additionally, MLflow 3.0 has been redesigned specifically for generative AI with enhanced monitoring and observability, while AI Functions now delivers enhanced performance to make SQL-based AI operations significantly more accessible and scalable.

Snowflake, known historically for its managed data warehouse, has shifted toward a lakehouse model. The addition of support for Apache Iceberg, launching its Polaris catalog, and Snowflake Horizon shows a commitment to openness and interoperability. Snowflake also brings native generative AI tools through Cortex and Snowflake Intelligence, allowing users to work with LLMs directly within its SQL-centric interface. While Snowflake prioritizes simplicity and governance in a managed environment, it is also expanding support for unstructured data and streaming use cases.

Cortex AISQL brings generative AI directly into customers' queries. This lets teams analyze multimodal data and build flexible AI-powered pipelines with familiar SQL syntax. Data Science Agent automates ML pipeline development from concept to production, utilizing Anthropic's Claude to break down machine learning workflow issues into clear steps, including data analysis, preparation, feature engineering, and training. This automation helps data science teams reclaim valuable hours typically spent on experimentation or debugging.

The Crunchy Data acquisition for approximately $250 million brings enterprise-ready Postgres capabilities to the AI Data Cloud. This strategic move enables Snowflake to provide early access to a PostgreSQL database service, simplifying the process of moving data into Snowflake and supporting the growing demand for operational workload convergence.

Teradata is targeting enterprise AI through a lakehouse strategy that builds on its established reputation for high-performance analytics. VantageCloud Lake delivers elastic compute on object storage, while ClearScape Analytics supports advanced in-database ML, including time series, geospatial, and LLM integration.

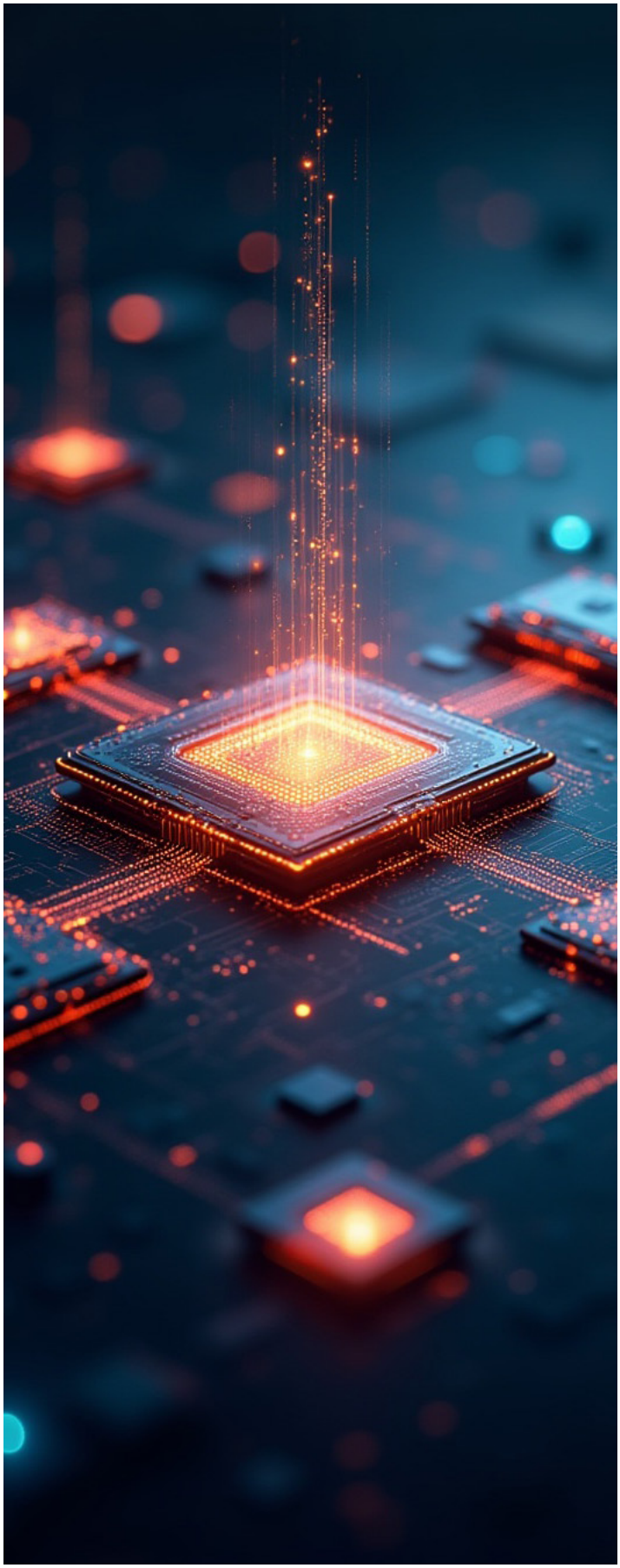
Teradata has launched Teradata AI Factory, an integrated solution for delivering cloud-based AI/ML capabilities to secure, on-premises deployments. The solution addresses enterprises' growing demand for data sovereignty and predictable AI costs. The platform also includes Teradata's Enterprise Vector Store for integration of structured and unstructured data, AI Workbench for self-service access to analytics libraries and functions, and native RAG processing capabilities.

AI Factory unifies critical components, including data pipelines, algorithm execution, and software infrastructure, into a single, scalable system design for rapid innovation through private, trusted AI. This can significantly accelerate AI development, including native RAG pipelines. Teradata's integration with NVIDIA for generative AI and its enterprise vector store enables responsive AI applications on structured and unstructured data. Its emphasis on governance, resilience, and interoperability positions it well for regulated industries.

Cloudera has re-emerged as a leader in hybrid and open lakehouse platforms. By combining open-source components like Apache Iceberg, Hive, and NiFi, Cloudera offers a flexible architecture that spans public cloud and on-premises deployments. SDX (Share Data Experience) brings unified metadata, governance, and security across environments. Cloudera's recent enhancements in real-time stream processing, AI model inference, and low-code application development tools make it especially appealing to enterprises maintaining hybrid or edge footprints.

Cloudera Data Visualization now extends AI capabilities to on-premises environments, providing secure and integrated AI capabilities native to the Cloudera platform. The platform enables users to unlock the value of on-premises data through intuitive visualization and natural language querying, supporting enhanced collaboration and decision-making without compromising data security or governance.

The company has enhanced its AI Inference Service with embedded NVIDIA NIM microservices, boosting LLM performance speeds using NVIDIA accelerated computing. This service protects sensitive data by providing secure development and deployment within enterprise control, enabling efficient development of AI-driven chatbots, virtual assistants, and agentic applications.



Key Capabilities for AI Workflows

The table below outlines a comparison of how each platform supports critical capabilities needed for enterprise AI.

Capability	databricks	snowflake	teradata.	CLOUDERA
Table Format	Delta Lake	Iceberg	Proprietary + Iceberg	Iceberg
Agentic AI	Agent Bricks (Auto-optimized)	Intelligence + Data Science Agent	AI Factory with NVIDIA	Hybrid AI Inference
Operational Database	Lakebase (Postgres)	Postgres via Crunchy Data	Native Integration	Limited Support
Vector Database	Native Vector Storage	Native Vector Capabilities	Enterprise Vector Store	Integrated Support
Multimodal Processing	AI Functions	Cortex AISQL	ClearScape Analytics	Enhanced Visualization
Real-Time Streaming	Structured Streaming	Snowpipe & Streaming	Via Partners	NiFi, Kafka
ML/AI Integration	MosaicML, MLflow, AI Functions	Cortex AI, Intelligence	ClearScape, NVIDIA NeMo	Predictive Builder, BYO Models
Governance	Unity Catalog	Horizon Catalog	Vantage Metadata Services	SDX
Deployment Model	Cloud, Limited on-premises	Fully Cloud	Multi-Cloud	Strong Hybrid & on-premises
End-User Access	Notebooks + SQL + BI Genie	SQL + Natural Language	SQL + REST APIs	SQL + Visual UI

Each vendor offers a different blend of openness, control, scalability, and AI integration. The right fit often depends on the enterprise's existing stack, AI maturity, and regulatory needs.

Market Dynamics and Future Outlook

The shift from AI experimentation to production deployment is accelerating. This represents a move beyond "pilot purgatory" toward scalable AI solutions that require robust data foundations and governance frameworks. The trend toward operational and analytical convergence represents a key architectural evolution. Platforms increasingly support both transactional and analytical workloads within unified environments, as demonstrated by Databricks' Lakebase and Snowflake's Postgres integration through Crunchy Data acquisition.

The growing importance of data democratization through natural language interfaces is evident across all major platforms. Snowflake Intelligence, Databricks' AI/ BI Genie, and Cloudera's AI-powered visualization tools all focus on empowering business users with direct access to data and AI tools without technical dependencies. These technologies get better data and decision-making insights into the hands of a broader range of stakeholders.

Strategic Considerations for Enterprise Architects

Choosing the right lakehouse platform for AI isn't just about technical features. It's a strategic decision. Architects need to consider the trade-offs between open source flexibility and fully managed services. They must evaluate whether the platform can support both the development and deployment sides of AI, including vector search, prompt engineering, and real-time inference.

Governance is another critical area. AI use cases often involve sensitive or proprietary data. Platforms must provide strong access controls, lineage tracking, and compliance features. Integration with semantic layers, feature stores, and business glossaries can also drive consistent and explainable AI.

Lastly, enterprises should look at ecosystem maturity. Does the vendor have a vibrant partner network? Are tools and models easily portable? Is there sufficient support for cloud-native and hybrid deployments? A broad ecosystem allows businesses to access a wider range of specialized solutions and to benefit from collaborative innovation.

Recommendations

To ensure that lakehouse software meets the demands of an organization, enterprise architects should consider the following recommendations:

- **Match platform capabilities to AI use cases**. For advanced generative AI and LLMs, platforms with native model hosting and vector search, like Databricks and Teradata, may be better suited. For embedded analytics and SQL-centric workflows, Snowflake or Cloudera may offer simpler integration.
- **Favor open table formats like Iceberg or Delta** to ensure long-term interoperability and avoid vendor lock-in. These formats support features like ACID transactions, schema evolution, and time travel, which are critical for AI reproducibility and data governance.
- **Build with governance in mind**. Select platforms that offer unified catalogs, access control, and auditing across environments. This is especially important in industries with strict compliance requirements.
- **Invest in platforms that support full-lifecycle AI**. From ingestion to inference, platforms should offer native or pluggable support for streaming, model training, experiment tracking, deployment, and monitoring.
- **Plan for hybrid and multi-cloud scenarios**. Cloudera and Teradata provide strong options for enterprises with on-premises or edge needs, while Databricks and Snowflake focus more on cloud-native scalability.
- **Evaluate the vendor's AI roadmap and community**. A strong open source ecosystem, active feature development, and third-party integration options are signals of a platform that will remain relevant as AI demands evolve.

Looking Ahead

The data lakehouse has moved from an emerging architecture to a strategic foundation for AI-driven enterprises. While no platform checks every box for every use case, the four vendors profiled here offer compelling, differentiated paths toward an AI-ready future. By understanding the trade-offs and strengths of each, enterprise architects can make informed decisions that support both short-term AI wins and long-term scalability. A well-architected lakehouse doesn't just support analytics; it enables the next generation of intelligent applications, data products, and decisions. The opportunity is here. The foundation matters more than ever.

These platforms are no longer simple data storage and analytics solutions; they have evolved into complete AI-powered ecosystems that support autonomous agents, multimodal processing, and real-time decision-making. The convergence of operational and analytical workloads, combined with natural language interfaces and vector database integration, enables business users to interact with data in ways previously reserved for technical specialists.

The market dynamics clearly indicate that enterprises are moving from experimental AI to production-grade, autonomous systems that can reason across all their data assets. The choice of a lakehouse platform has become a strategic business decision that will determine competitive advantage in the AI-driven economy. Organizations that choose platforms with strong agentic AI capabilities, operational-analytical convergence, and comprehensive governance frameworks will be best positioned to capitalize on the next wave of AI innovation.

ABOUT HYPERFRAME RESEARCH:

HyperFRAME Research delivers indepth research and insights across the global technology landscape, spanning everything from hyperscale public cloud to the mainframe and everything in between. We offer strategic advisory services, custom research reports, tailored consulting engagements, digital events, go to market planning, message testing, and lead generation programs.

Our industry analysts specialize in rigorous qualitative and quantitative assessments of technology solutions, business challenges, market forces, and end user demands across industry sectors. HyperFRAME Research collaborates closely with your Analyst Relations, Product, and Marketing teams to build and amplify your thought leadership, positioning your expertise to enhance brand and product recognition. Through content that engages readers, viewers, and listeners alike, we ensure your voice resonates across channels.

CONTACT HYPERFRAME RESEARCH:

Steven Dickens
CEO & Principal Analyst | HyperFRAME Research

Email Address:

steven.dickens@hyperframeresearch.com

Telephone Number:

+1 845 505 1678

X: - @StevenDickens3

LinkedIn: Steven Dickens

BlueSky: Steven Dickens

CONTRIBUTORS

Steven Dickens CEO & Principal Analyst
HyperFRAME Research

Stephanie Walter
Analyst In Residence - AI Tech Stack

INQUIRIES

Contact us if you would like to discuss this report and HyperFRAME Research will respond promptly.

CITATIONS

This paper can be cited by accredited press and analysts, but must be cited in-context, displaying author's name, author's title, and "HyperFRAME Research." Non-press and non-analysts must receive prior written permission by HyperFRAME Research for any citations.

LICENSING

DISCLOSURES

HyperFRAME Research provides research, analysis, advising, and consulting to many high-tech companies, including those mentioned in this paper. No employees at the firm hold any equity positions with any companies cited in this document.


HyperFRAME
RESEARCH