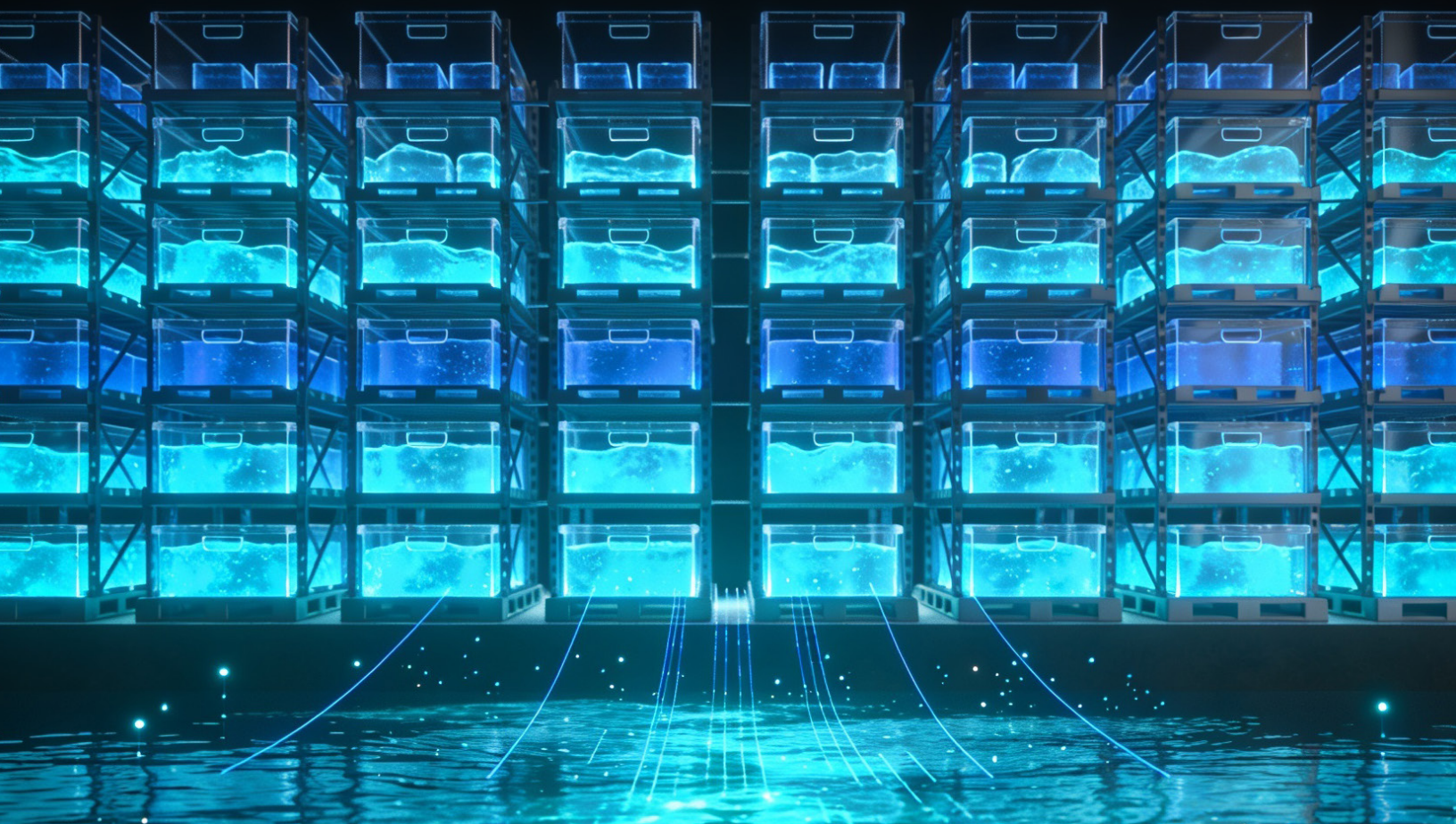


The Next Generation of Data Lakehouses

Why openness, full data-warehouse fidelity, and AI-integrated operations define the next generation of lakehouses.



AUGUST 2026

Authors:
Stephanie Walter
Practice Leader AI Stack

Ron Westefall
VP and Practice Leader for
Infrastructure and Networking

Steven Dickens
CEO and Principal Analyst

Executive Summary

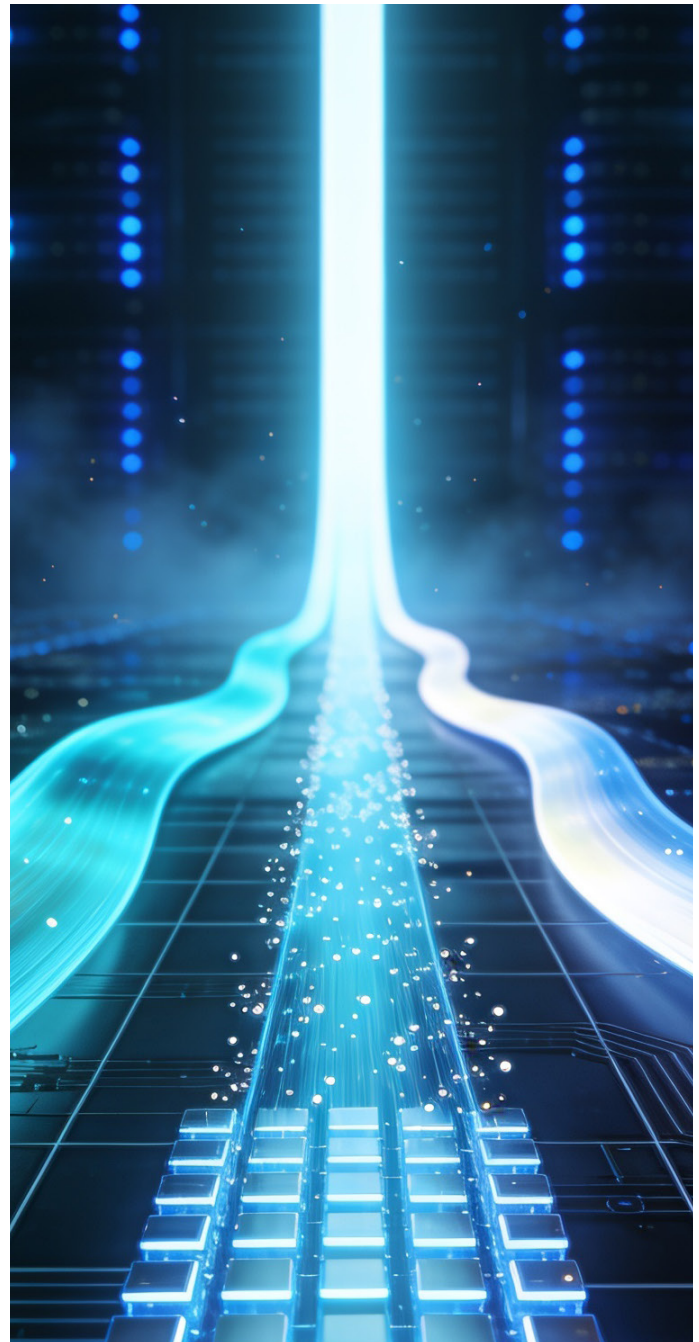
As AI moves from pilot project to enterprise production deployment, data is rapidly becoming the crucial defining factor in the success of an AI project. The question many enterprise IT leaders are struggling with is: Do we take our data to the AI or bring the AI to the data?

Against this backdrop, enterprise data, AI, and operational systems have grown together, but they have not grown in the same direction. Most organizations still manage a patchwork of cloud services, data warehouses, analytics engines, operational databases, data lakes, and AI stacks that rarely function as a unified whole. The result is familiar: Data silos. Stale data. Slow insights. Extraneous and overlapping data pipelines. Rising costs. Security compromises. And, most of all, limited trust in the systems that guide daily decisions.

The industry response was the lakehouse. First, it was an effort to merge data lakes with data warehouse performance. Many early platforms succeeded at part of that equation, but too many lakehouses still require trade-offs across openness, operational reach, enterprise performance, and AI integration. The gaps show up quickly as companies try to run AI systems in production. Enterprise architects and data leaders are struggling with flexibility without governance, openness without performance, and AI applications that never reach the last mile of operations.

The next generation must close these gaps. A true AI lakehouse needs to deliver four things at once. It must be open and interoperable. It must access data wherever it resides. It must deliver warehouse-grade, enterprise-class performance with the reliability, security, and governance that enterprises expect. And it must integrate AI and operational systems in a way that supports real-time intelligence. Not dashboards. Decisions.

For CIOs, CDOs, architects, and AI leaders, this shift marks a practical turning point. The AI lakehouse is no longer just an analytics solution. It is the strategic control plane for enterprise-scale intelligence.



The winning platforms will be the ones that erase the boundaries between analytical data, operational data, and AI-driven action.



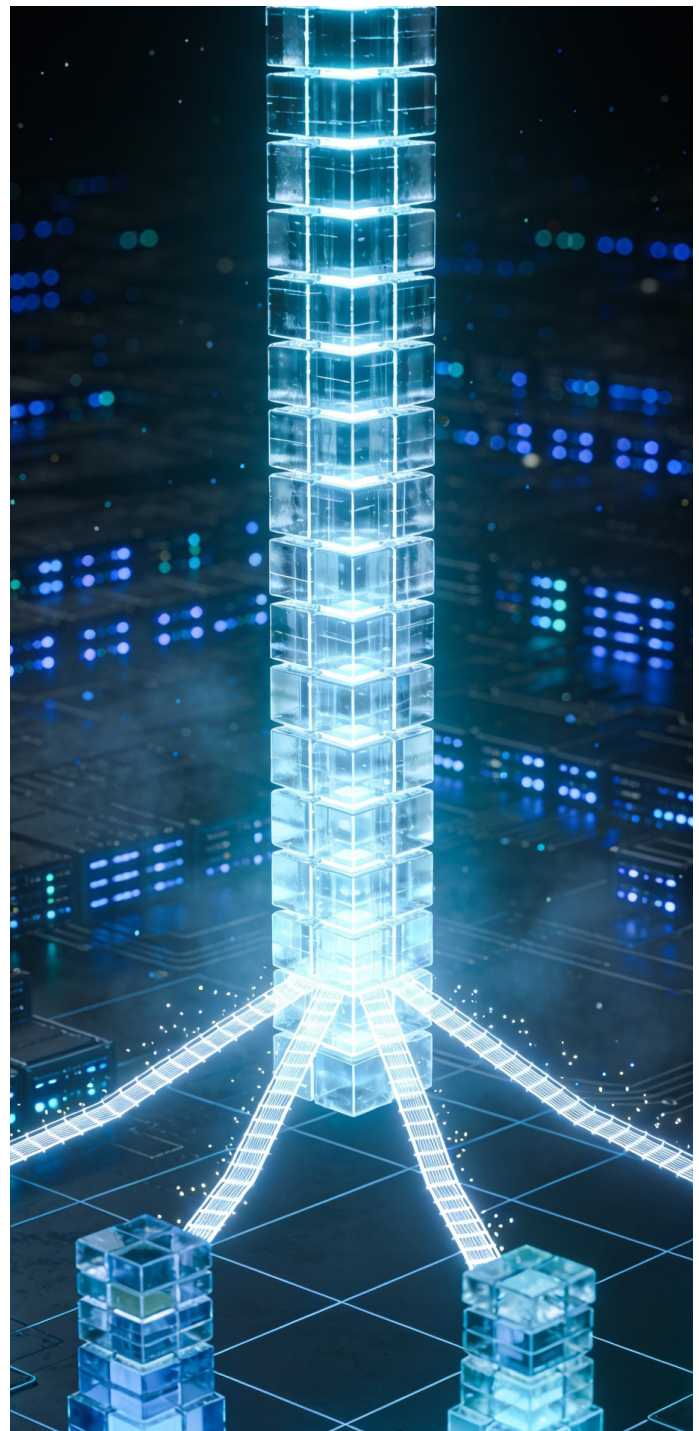
The Lakehouse Landscape: A Platform in Transition

A modern lakehouse sits in the middle of a rapidly changing data world. It blends the elasticity of a data lake with the structure and performance of a warehouse. It matters now because AI workloads are pushing demand for unified, governed access to every class of data wherever it resides. Companies do not want separate stacks for AI, analytics, machine learning, operations, and governance. They want one platform that can carry all of it.

Analytics platforms are evolving into operational intelligence layers. The shift is clear. Generative models need a broad context. Reinforcement-style agents need live operational data. Automated, AI-powered decision systems need consistent governance. This pressure is pulling the lakehouse into the center of the enterprise data architecture.

Underpinning the four things a true AI lakehouse needs to deliver are three forces that stand out. AI workloads require integrated data access across structured, semi-structured, and unstructured sources. Hybrid and multi-cloud environments are now the default architecture for large companies. And enterprises want a unified data platform that spans AI, analytics, and real-time operations, without creating new silos. This is where the market is heading.

Each major vendor is reshaping its platform around these pressures. AWS is aligning Redshift, Athena, and Lake Formation into a single path for analytics and data sharing. Azure combines Synapse, Fabric, and Databricks into a broader ecosystem for data and AI. Google Cloud is pulling BigQuery, BigLake, and Vertex AI into a tight loop that connects data, governance, and model operations. Snowflake is expanding beyond its data warehouse strength to handle streaming, AI-driven features, and transactional workloads through its Unistore and Hybrid Tables initiatives. Databricks continues to extend Delta and Unity Catalog across an open ecosystem. Oracle evolves its Autonomous AI Database portfolio with Autonomous AI Lakehouse, adding support for Apache Iceberg open tables, multi-cloud catalog federation, and direct analysis of real-time operational data. We see a platform class in transition.



The lakehouse is no longer just an analytical foundation. It is becoming the decision backbone that feeds and powers enterprise AI.

Open and Interoperable by Design

Openness has moved from a preference to a requirement. Enterprises want the freedom to run AI and analytics on platforms that fit their needs without locking the data itself into a single vendor's design. Multi-cloud made this inevitable. AI simply made it urgent. When models depend on broad, high-quality context, organizations cannot afford data silos or proprietary formats that slow collaboration across teams and clouds.

This is why open table formats like Apache Iceberg, Delta, and Hudi have gained traction. They aim to deliver consistent metadata, schema evolution, ACID guarantees, and scalable table management. Just as important, they allow enterprises to read, write, and stream data with a wide ecosystem of tools. The industry shift toward shared metadata catalogs reinforces this trend. Companies want their governance rules, lineage, quality checks, and access controls to extend across clouds and analytics engines. It is a natural response to the complexity of modern AI pipelines.

The requirements are clear. Enterprises expect the ability to discover, ingest, query, and stream data directly from open table formats. They want interoperable catalogs that work across vendors and tools. They want governance that travels with the data rather than getting trapped inside a proprietary platform. And they expect multi-cloud architectures to work with the same level of fidelity that they see on a single cloud.

From HyperFRAME's perspective, openness and interoperability now serve as the price of admission for AI scalability. Any platform that cannot operate across formats, clouds, and tools will fall behind as AI workloads become more distributed and more operational. Enterprises want choice. They want durability. And they want an ecosystem that stays open long after individual services change.

Enterprise-Grade Performance Without Compromise

Early lakehouses favored flexibility over performance. That trade-off worked for exploratory analysis but not for enterprise workloads. Today, the bar is higher. Companies expect real-time access to operational data, consistent SQL performance, predictable concurrency, and strong governance controls layered across analytics and AI systems. They also expect cost efficiency. Growing model workloads have made that point sharp.

Performance parity with cloud data warehouses is becoming the new benchmark. Enterprises want Atomicity, Consistency, Isolation, and Durability (ACID) alongside fault tolerance and governance embedded into every tier of the platform. They want BI tools to run with predictable latency. They want streaming pipelines that deliver fresh data to AI systems. And they want all of this without recreating the complexity that lakehouses were supposed to solve.



Enterprise-grade performance must also extend to data wherever it resides. Querying across clouds, platforms, and on-premises environments cannot introduce unacceptable latency or require every dataset to be copied into a single system. Achieving zero-copy, cross-environment data querying depends on decoupling compute from storage and supporting open table formats. By shifting the control plane to push execution logic down to where the data physically resides, modern data architectures eliminate the high egress costs and latency penalties of centralized ETL pipelines while maintaining unified security and governance.

In our view, this is where the market is converging. Performance cannot be an afterthought. Governance cannot be optional. Real-time access to operational and analytical data must work at scale. The lakehouse that succeeds will be the one that delivers open data structures with the speed, predictability, and control that enterprises already expect from their mission-critical systems.

AI-Driven Operational Integration

AI is shifting from analysis to action. That shift sounds simple, but it is the hardest part of the enterprise AI journey. Many organizations are prioritizing adopting models for their business through RAG. Many can run proofs of concept. The real challenge is putting AI into the flow of everyday operations. This

is where the current lakehouse generation shows its limits. Too many platforms remain analytic islands. They store data. They process queries. They support dashboards and offline training jobs. What they do not do is connect AI directly to business systems that need real-time decisions to take action.

A modern lakehouse must close that gap. It needs to function as a unified intelligence platform that spans operational data, analytical data, streaming pipelines, and AI system outputs. This means a lakehouse can no longer rely on batch updates or delayed ingestion. Operational AI demands live signals. It needs to capture data from core transactional systems. It needs event streams that reflect business activity as it happens. And it needs catalogs that understand operational metadata and analytic metadata within the same structure. Governance must travel across both worlds.

Agentic AI frameworks reinforce these requirements. As autonomous and semi-autonomous systems grow, they depend on strong feedback loops. They need low-latency access to operational data to make accurate decisions and then push those decisions back into transactional systems for actionable responses.

The next competitive frontier is operational AI. The lakehouse will not succeed solely as an analytical platform. It must serve as the engine that takes real-time decisions into execution across the business. This is where enterprises will see ROI from AI. This is where platforms will differentiate. And this is where the next generation lakehouse will win.



A lakehouse that cannot support this pattern will limit the reach of AI in the enterprise.

Comparative Analysis: The Lakehouse Leaders

To compare the major platforms, HyperFRAME uses four criteria: openness and interoperability; access to data wherever it resides; enterprise-grade performance; and integration across operational systems and AI workloads.

Based on our analysis, AWS brings the broadest infrastructure ecosystem. Its analytics and AI services cover nearly every workload type. The challenge is consistency across the data model. Redshift, Athena, EMR, and Lake Formation work together, but they still feel like a collection of services rather than a unified platform. It's mostly up to the customer to stitch them together. Openness is improving with Iceberg and Delta support, yet multi-cloud reach remains partial.

From our research, Azure stands out for enterprise integration. Fabric and Synapse bridge data engineering, analytics, AI, and business applications through Microsoft 365 and Power BI. This makes Azure attractive for organizations with heavy Microsoft investments. Openness is improving, although metadata gravity often remains tied to Azure services.

Through our evaluation, Google Cloud has a strong AI native design and a growing commitment to open formats. BigQuery provides high concurrency and scalable performance. BigLake and Omni strengthen multi-cloud governance. Vertex AI ties models tightly to the data platform. Governance depth is still evolving, but the architecture is built for next-generation workloads.

Our analysis shows that Snowflake continues to differentiate in data warehouse performance. Elasticity, predictable throughput, and ease-of-use remain its core advantages. It is expanding into AI and streaming with Cortex AI and external Iceberg tables, but real-time operational integration remains limited, as a result AI agents act on stale data. Offerings run in hyperscaler clouds, but not all features are consistent across providers.

Based on HyperFRAME's analysis, Databricks has the strongest open ecosystem. Delta, MLflow, Unity Catalog, and Mosaic AI create a broad platform for data science and machine learning. However, the operational side of the enterprise remains a issue as its Lakebase Postgres is a nascent offering. Most workloads still live within the analytics environment rather than core business systems. The path toward operational integration is widening but not complete.



From HyperFRAME’s perspective, Oracle is unifying operational and AI data workflows more directly. Its Autonomous AI Lakehouse with open table format support, catalog of catalogs, and true multi-cloud deployments allow it to run the same lakehouse architecture on OCI, AWS, Azure, Google Cloud, and on-premises. With Autonomous AI Database powered by Exadata, Oracle delivers what’s considered the best

performance. Federation across catalogs provides visibility across data platforms and hybrid environments. And connecting operational databases to the lakehouse strengthens real-time AI integration.

The table below summarizes each vendor’s capabilities:

Vendor	Openness, Interoperability, and Data Access	Enterprise-grade Performance	AI and Operational Integration	Notable Strengths	Notable Shortcomings
AWS	Supports Iceberg and Delta. Broad tool ecosystem (S3, Glue, Athena). Partial multi-cloud.	Mature and elastic but fragmented across independent services. Hybrid query via Redshift/Athena.	Integrates Bedrock and SageMaker with Aurora, Redshift, Kinesis, DynamoDB.	Largest infrastructure portfolio. Strong streaming and operational data paths.	High architectural complexity; managing disparate services can create glue overhead.
Azure	Interoperability improving Synapse/Fabric. Metadata often tied to MS ecosystem.	Strong for BI and SQL workloads. Fabric scalability is improving.	Deep integration with Microsoft Copilot and the application ecosystem.	Enterprise-ready integration with productivity (365) and app layers.	Ecosystem lock-in; best features often require full commitment to the Microsoft stack.
Google Cloud	Supports the Apache Iceberg open table format across its lakehouse portfolio, with BigQuery Omni providing cross-cloud analytics.	BigQuery provides high concurrency for structured/unstructured data.	Vertex AI is tightly integrated with BigQuery, BigLake, and AlloyDB.	AI-native platform design. Strong convergence of data, governance, and model ops.	Market share and support; smaller third-party ecosystem compared to AWS.
Snowflake	Growing Iceberg support. Footprint across major cloud providers.	Leading warehouse performance, elasticity, and predictable throughput.	Cortex AI for managed model inference. Snowflake Postgres extends the platform into transactional workloads.	Ease of use. Core warehouse architecture consistent across major clouds; some features and regional variations remain.	Storage and egress costs; can become expensive at scale; historically a walled garden
Databricks	Open by default (Delta/Unity Catalog). Broad data science ecosystem available in major cloud providers.	Strong across ML, data engineering, and large-scale analytics, with mature SQL and BI performance through Photon and serverless warehouses.	Mosaic AI orchestrates workflows. Operational integration centered on analytics.	Strongest open data science ecosystem. Unified governance across ML	Operational system-of-record integration remains less native than database-centric platforms.
Oracle	Native Iceberg support. Multi-cloud (OCI, AWS, Azure, GCP, On-prem).	Autonomous AI Lakehouse provides self-tuning, warehouse-grade performance.	Direct integration with operational DBs, GoldenGate CDC, and AI services.	Strong heritage in mission critical transactional systems and database engineering. Unified security and governance across all data.	Ecosystem breadth and maturity continue to evolve relative to hyperscaler-native platforms.

From these comparisons, the distinctions are clear: all major vendors are converging on the unified lakehouse model. Databricks leads in open data science and machine learning. Snowflake remains the benchmark for cloud-based warehouse workloads while attempting to plug the operational gap. Hyperscalers deliver unmatched breadth and are increasing architectural cohesion through services like Zero-ETL. To its

competitive advantage, Oracle offers a highly mature and robust alignment between operational systems and AI due to its converged database heritage, which positions Oracle as a competitively advantageous platform for agentic AI deployments and a compelling choice for enterprises with large existing operational footprints.

Oracle Autonomous AI Lakehouse: Stand-out Architecture and Approach

Overall, we believe that Oracle's competitive advantage in the AI lakehouse market stems from its heritage in transactional data and enterprise-grade database technology. This gives Oracle a strong foundation for bringing AI to the data, as opposed to moving data into separate environments before it can be used.

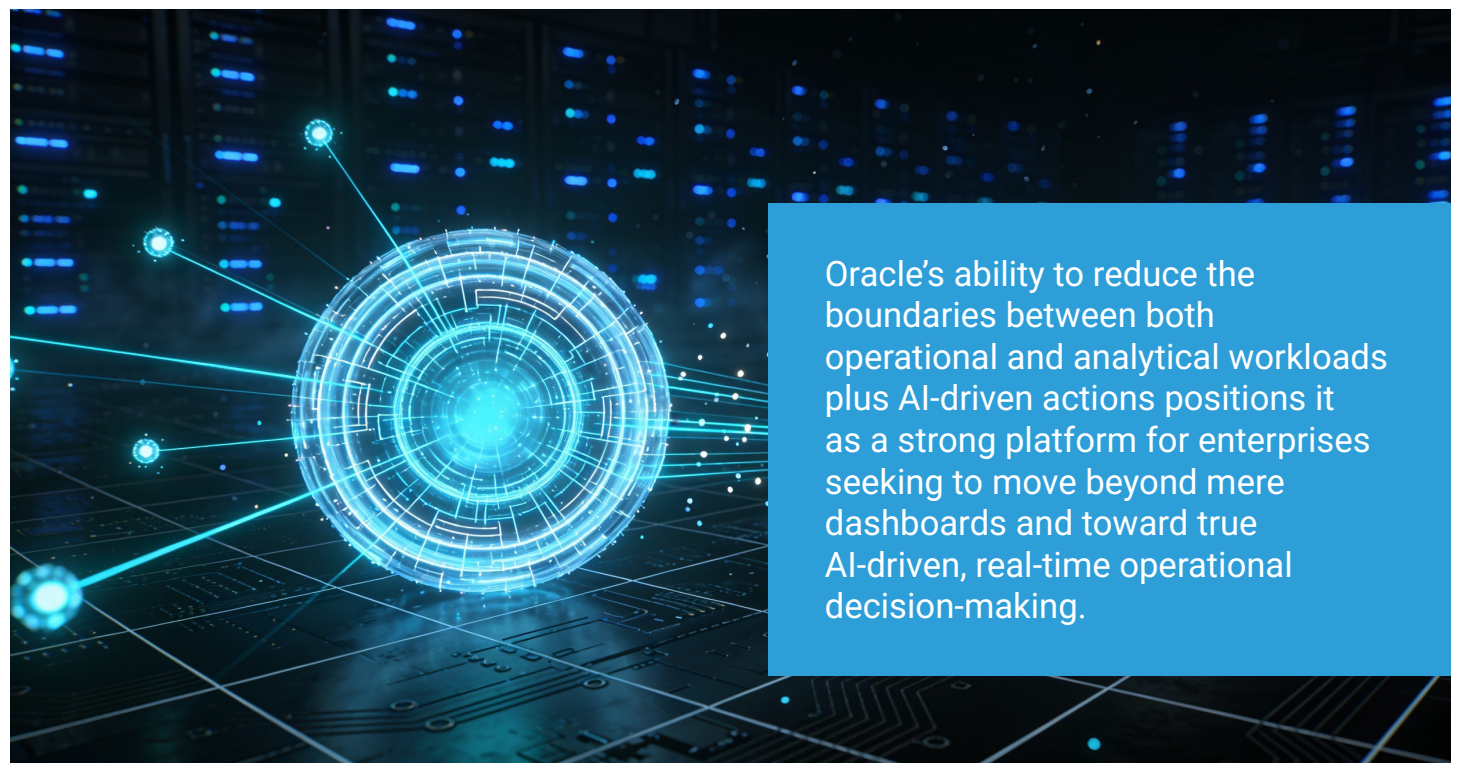
Oracle Autonomous AI Lakehouse combines Autonomous AI Database with open standards such as native Iceberg support, Oracle GoldenGate CDC, unified catalog capabilities, and Select AI Agent. Together, these components establish a unified intelligence platform that connects live operational data, analytical data, streaming pipelines, and AI workflows. The result is a platform designed to help enterprises use current business context for real-time decisions while preserving the performance, reliability, and governance expected from mission-critical systems.

Oracle's lakehouse approach reflects a long evolution in the company's data strategy. The platform is designed to merge traditional strengths in operational databases with open, multi-cloud capabilities that meet modern AI requirements. The architecture centers on the Autonomous AI Database,

which brings automated database management operations, transactional reliability, integrated security and governance, and consistent high performance powered by Exadata into the lakehouse model. At the same time, Oracle has embraced open interoperability through native Apache Iceberg support and unified catalog capabilities, allowing enterprises to discover and query open data and external catalogs across OCI, AWS, Azure, Google Cloud, and on-premises environments. Data can remain in place where appropriate, while replication or materialization can be used when performance, resilience, or operational requirements demand it.

Real-time ingestion is a foundational element. GoldenGate Change Data Capture (CDC) moves changes from operational systems into Iceberg tables with minimal lag, making current business activity available to analytics and AI workloads. This allows the lakehouse to respond to change as it happens. Live operational signals can feed AI models, and the outputs of these models can drive actions across agents and/or flow back to transactional systems while minimizing the latency and complexity of separate data pipelines.

Performance is equally important. Early lakehouses often forced enterprises to choose between open data and warehouse-grade performance. Oracle's approach is designed to avoid that trade-off. Autonomous AI Database brings established capabilities for concurrency, security, resilience, and workload management, while Data Lake Accelerator scales out intensive scans and



Oracle's ability to reduce the boundaries between both operational and analytical workloads plus AI-driven actions positions it as a strong platform for enterprises seeking to move beyond mere dashboards and toward true AI-driven, real-time operational decision-making.

Lake Cache accelerates access to frequently used Iceberg data. This gives enterprises a way to run analytical and AI workloads on open data without sacrificing the performance, reliability, and control they expect from core business systems.

Metadata unification is another design pillar. The Autonomous AI Lakehouse Data Catalog creates a catalog of catalogs across hybrid, multi-cloud environments, and data platforms. External tables provide an additional path to query data in place. These capabilities help teams discover and query data, enforce governance, evaluate lineage, and audit usage regardless of where the data resides. They also simplify cross-functional AI workloads by helping teams to govern and trace the data used across both analytical and operational systems.

Oracle's advantage is not simply the ability to catalog and query live data. It is the ability to bring AI closer to the operational data and enterprise controls that already govern core business processes. GoldenGate CDC can provide current operational signals, while Autonomous AI Database adds mature transaction management, security, identity, auditing, and governance. This gives enterprises a stronger foundation for building AI systems that use trusted operational context rather than relying only on virtualized access to raw or intermediate data states. The result is a more reliable path from live data to decision-ready intelligence and, where appropriate, controlled business action.

AI workflow integration is handled through Select AI Agent, and the broader AI services inside Oracle's data platform. Models and agents can query structured, semi-structured, and unstructured data through a unified interface, reducing unnecessary data movement while preserving security and control. Where enterprise policy permits, agents can use approved tools to generate recommendations, initiate

workflows, and support action in operational systems. These actions can be governed through enterprise identities, role-based access controls, audit trails, and approval workflows for high-impact decisions. This creates a path from governed data to controlled, real-time decision-making.

A typical workflow illustrates the approach. An operational database streams updates through GoldenGate into Iceberg tables stored on OCI, AWS, Azure, or Google Cloud. The AI engine combines this live data with analytical history and external context to generate predictions or recommendations. Where enterprise policy permits, the result is written back to an operational system that drives business action. The process is automated where appropriate, governed throughout, and anchored in standard formats, creating a direct connection between current operational activity and enterprise decision-making.

This architecture supports multi-cloud and hybrid deployment while preserving performance, reliability, and governance. It gives organizations the flexibility to use open data formats and external platforms without losing the operational control that complex AI workloads require. The catalog of catalogs reduces data silos. Real-time ingestion reduces latency. Autonomous AI Database provides the performance foundation. And Select AI Agent creates a path from governed data to controlled action.

In practice, the design reflects a pragmatic balance between openness, automation, and operational AI integration. Autonomous AI Lakehouse serves as a unified intelligence layer rather than a siloed analytics component. Its advantage is not only faster access to operational data. It is the ability to make operational context trusted, decision-ready, and actionable inside the same enterprise control plane.





Recommendations for Enterprise Data and AI Leaders

Most organizations are already on the path toward an AI lakehouse, even if they have not labeled it that way. The shift shows up in the desire for a single data platform, a unified catalog, and a more direct link between AI, analytics, and operational systems. Still, the transition is complex. Clear criteria can help enterprise leaders evaluate platforms and avoid architectural dead ends. Here are call to action recommendations for decision makers to prioritize:

- **A practical checklist starts with openness.** Any platform should support Iceberg, Delta, or Hudi natively and provide transparent interoperability with other catalogs. It should also discover and query data across clouds, platforms, and on-premises environments without requiring every dataset to be copied into one platform. Next is warehouse-grade SQL performance. Enterprises should expect high concurrency, predictable latency, strong indexing, full ACID reliability, and strong governance. Streaming and CDC ingestion are essential for operational AI, so real-time integration must work without custom engineering. AI frameworks should connect directly to the lakehouse, not through separate pipelines.
- **Pilot a cross-domain AI use case that touches both analytics and operations.** A supply chain prediction feeding an order management system. A customer signal flowing into a service workflow. A risk model updating policy decisions. These examples expose gaps early and force the platform to deliver on its promise.
- **Establish an open, governed catalog at the start.** Metadata sprawl becomes costly as AI workloads multiply. Enterprises that centralize governance early avoid painful refactoring later. We also advise considering interoperability over short term optimization. A highly tuned proprietary stack could limit long term choice and complicates agentic AI adoption.
- **Organizational alignment matters as much as technology.** DataOps, AIOps, CloudOps, and platform teams need a shared view of pipelines, catalogs, and operational patterns. AI without operational readiness will stall. The lakehouse should become the anchor for this alignment.

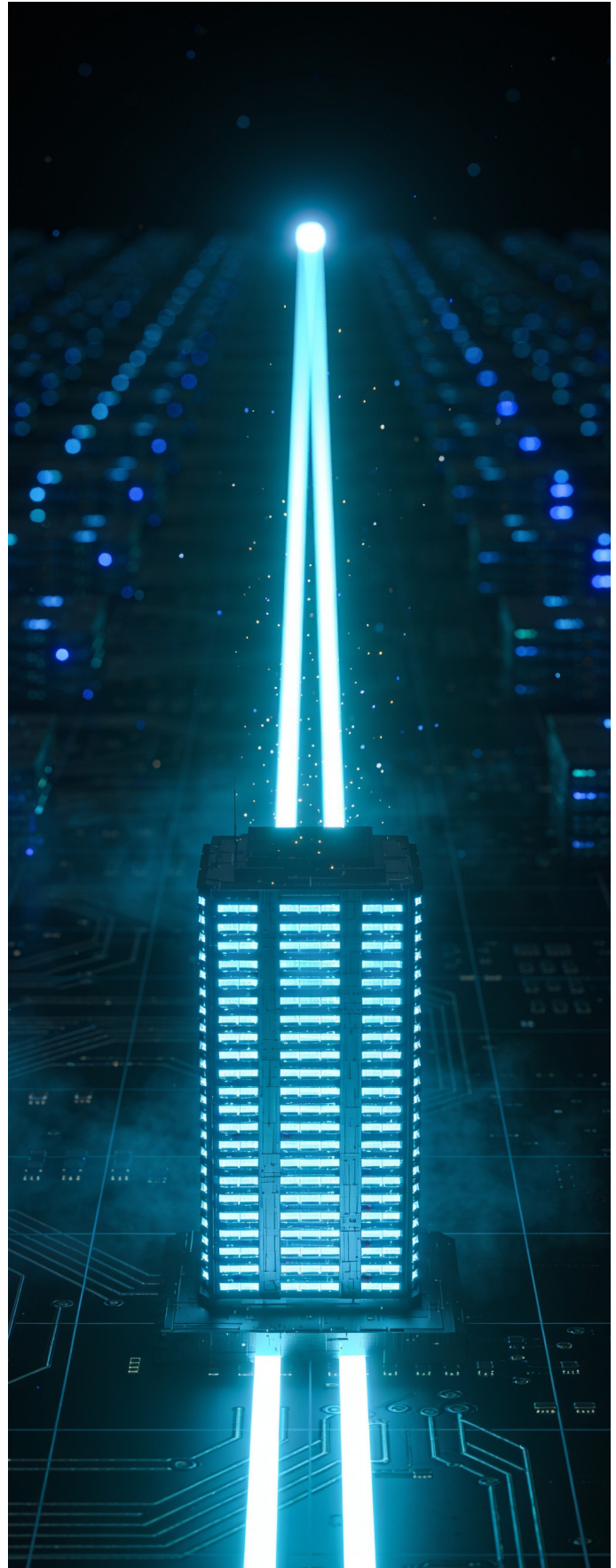
Decision makers must assess risk. Migration difficulty, workforce skills, platform lock-in, data privacy, and security all influence the sustainability of the chosen architecture. The best lakehouse is fast and open. It also lowers operational risk and provides a durable path for future AI systems. That is the value enterprises should measure.

Looking Ahead

The success of modern AI projects depends on governed access to data wherever it resides and its frictionless curation and presentation to LLM and agentic systems. This transition is supported by the emergence of fully transactional open table formats, edge-to-cloud streaming, and orchestration hubs that manage not just metadata, but the coordination of AI agents and hybrid-cloud policies.

This shift transforms data management into intelligence orchestration, where static pipelines become adaptive, and the lakehouse acts as the control plane for autonomous business workflows. By bridging the gap between analytical and operational workloads, the lakehouse provides the essential foundation of trust and scale required for enterprises to transition from predictive analytics to agentic AI.

The AI lakehouse has transcended its role as a mere data platform to become the central architecture for the autonomous enterprise. As the market moves toward an open, coordinated intelligence platform, organizations must focus their planning on this integrated direction.





ABOUT HYPERFRAME RESEARCH:

HyperFRAME Research delivers in-depth 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

Ron Westfall
VP and Practice Leader for
Infrastructure and Networking

Stephanie Walter
Practice Leader, AI Stack

Steven Dickens
CEO & Principal Analyst

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

This document, including any supporting materials, is owned by HyperFRAME Research. This publication may not be reproduced, distributed, or shared in any form without the prior written permission of HyperFRAME Research.

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.

