

# The Open Lakehouse Imperative: Delivering AI Value Without Compromise

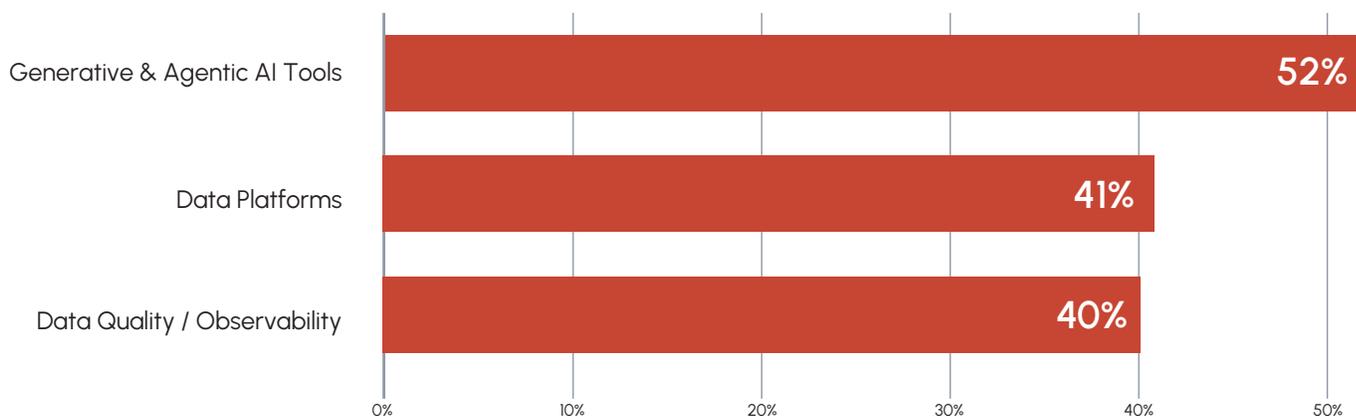


## Part 1: The AI Mandate Meets a Data Reality

Enterprises are aggressively realigning their strategies around an AI-first mandate. The directive from the board is clear: leverage AI to drive efficiency, innovation, and a durable competitive advantage. This isn't just talk. AI-first strategies are backed by significant investment, with over half of organizations (52%) prioritizing Generative and Agentic AI tools above all other technology spending (see Figure 1). Yet, this ambition is colliding with a harsh data reality.

Many AI initiatives are failing to deliver value, not because of flawed algorithms, but because of a flawed data foundation. Recent Futurum Research is unequivocal on this point: 11% of leaders directly blame poor data quality or availability for their AI project failures, making it a top reason for stalled progress<sup>1</sup>. The promise of the modern lakehouse architecture – to unify data with openness and flexibility – makes it a good candidate to provide a solid data foundation for AI. It certainly has captivated enterprises, driving a market set to reach an impressive US\$74.8 billion by 2029, according to Futurum Research<sup>2</sup>. However, a premature choice or incomplete implementation of a data lakehouse can lead to crippling performance trade-offs, new forms of vendor lock-in, and persistent data silos, creating a significant bottleneck to AI success.

**Figure 1: Top Investment Priorities for 2025 (% of Respondents).**



Source: Futurum Research, 1H 2025 Data Intelligence Decision Maker Survey.

<sup>1</sup> Futurum Research (2025). 1H 2025 Data Intelligence, Analytics, and Infrastructure Decision Maker Survey Report

<sup>2</sup> Futurum Research (2025). 1H 2025 Data Intelligence, Analytics, & Infrastructure Market Sizing & Five-Year Forecast



## Part 2: The Lakehouse Compromise: Why Modern Data Stacks Fall Short

To activate AI at scale, organizations must overcome critical hurdles inherent in many current-generation data lakehouse platforms. The market has signaled a clear preference for open, modular architectures – in fact, a massive 77% of organizations are already implementing, piloting, or planning to adopt isolated data lakehouse architectures with open table formats, according to Futurum Research. Yet this enthusiasm is often met with platforms that force a series of compromises, undermining the very goals they aim to achieve.

First is the illusion of openness. Many platforms that champion open standards subtly steer users toward proprietary catalogs and formats, creating a new, more systemic form of lock-in. This 'velvet rope' approach forces a binary choice between true cross-platform interoperability and the convenience of a single vendor's ecosystem, defeating the primary purpose of an open architecture. This is a strategic misstep for any enterprise serious about future-proofing its data stack.

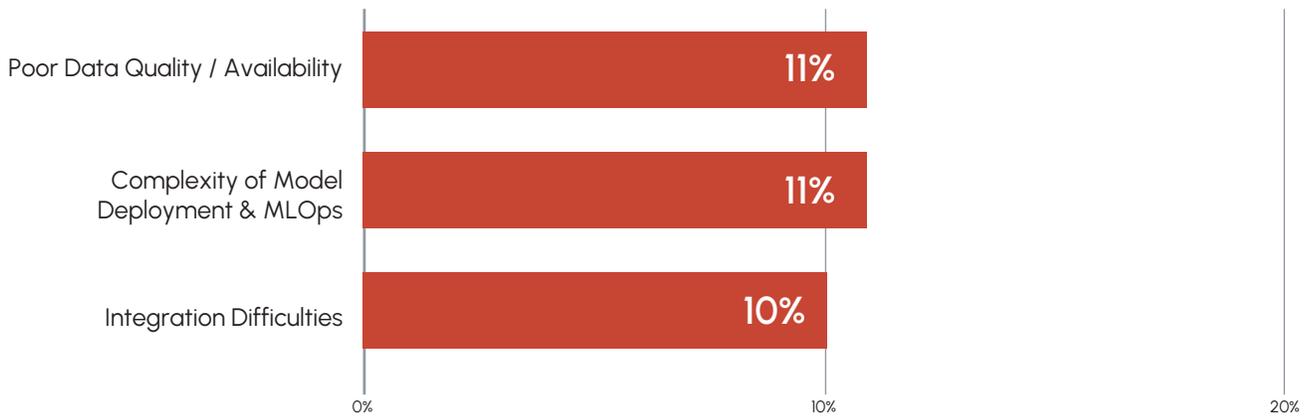
Next are the persistent performance gaps. Adopting open table formats such as Apache Iceberg grants architectural flexibility but can come at the cost of enterprise-grade performance, security, and concurrency. Foundational database capabilities – such as fine-grained updates, mature security models, and the indexing needed for high-performance querying – are not native to object storage. This forces data teams into an unwelcome trade-off between the adaptability of open formats and the reliability of a mission-critical data warehouse.

A fragmented, multi-cloud reality compounds this challenge. Enterprise data is not monolithic; it spans a sprawling estate of distributed assets across multiple environments. Futurum Research confirms this complexity: organizations report deploying data solutions across on-premises environments (31%), private cloud (36%), public cloud (52%), and hybrid architectures (47%)<sup>3</sup>. These figures reflect overlapping strategies rather than exclusive choices – most enterprises operate across several of these environments simultaneously. In this context, a 'rip and replace' strategy is not just impractical; it is a non-starter. Any data platform tied to a single cloud – or unable to integrate with existing investments in platforms such as Snowflake and Databricks – fails to reflect the operational reality of the modern enterprise.

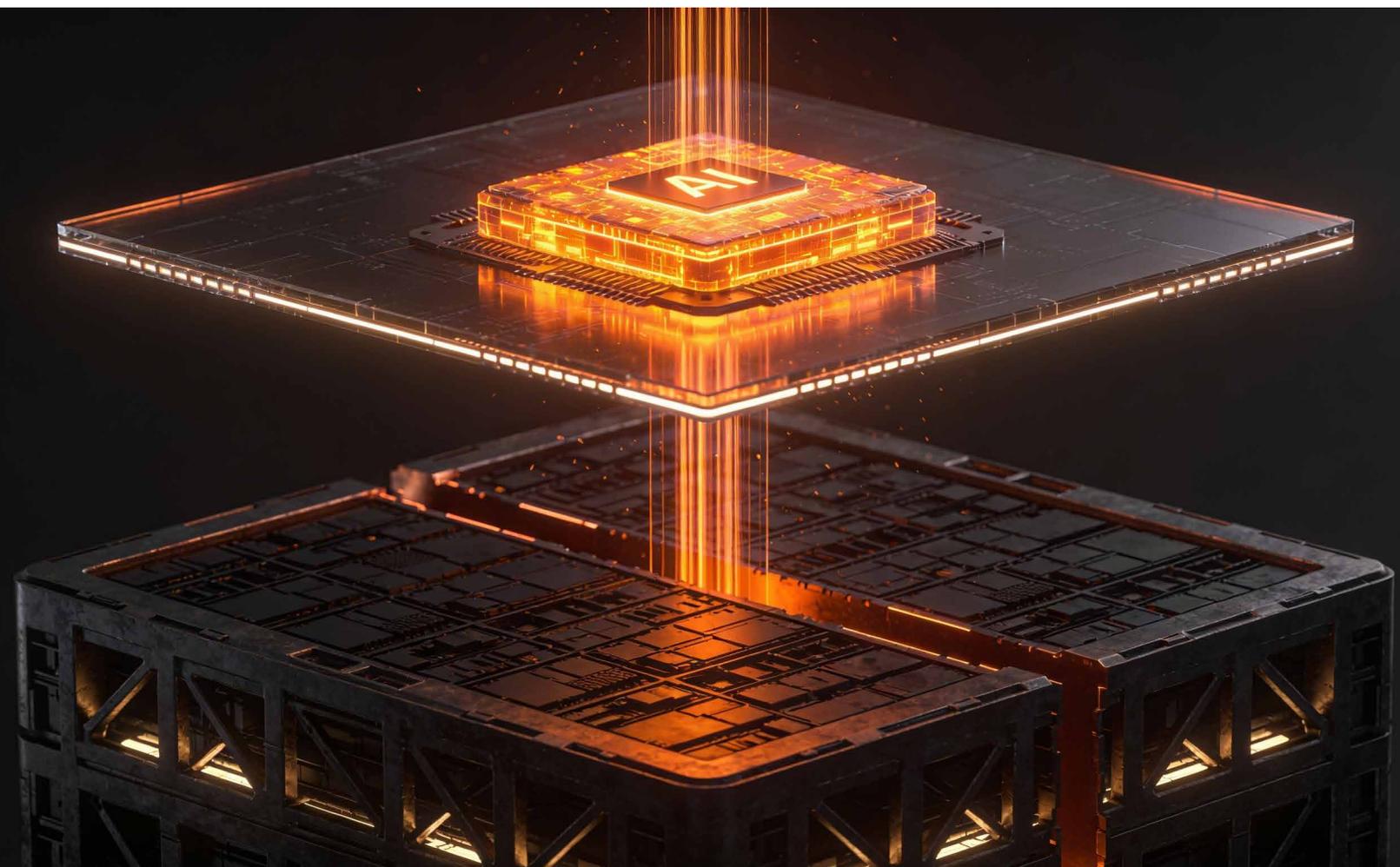
<sup>3</sup> Futurum Research (2025). IH 2025 Data Intelligence, Analytics, and Infrastructure Decision Maker Survey Report

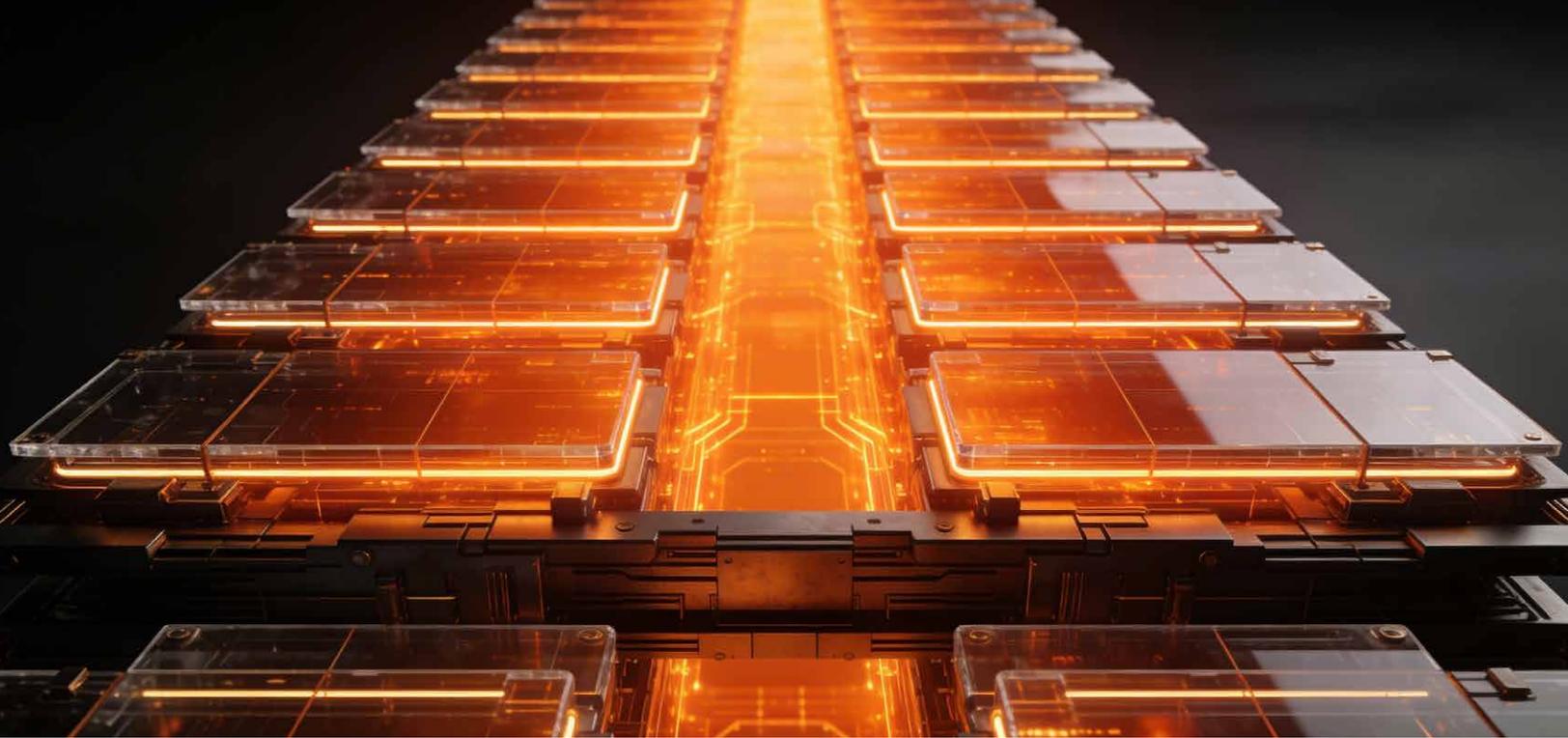
Finally, there is the 'last mile' problem. Generating a powerful insight within an analytical environment, such as a data lakehouse, is only half the battle. A critical gap remains in operationalizing these AI-driven insights, with 10% of data leaders citing 'Difficulties Integrating Models with Existing Systems/Workflows' as a key reason for AI project failure (see Figure 2). Too often, there is no straightforward path to embed analytics outcomes into the core transactional business processes where they can generate tangible value, leaving potential ROI stranded in dashboards and reports.

**Figure 2: Top Data-Specific Factors Contributing to AI Project Failure (% of Respondents).**



Source: Futurum Research, 1H 2025 Data Intelligence Decision Maker Survey.





## Part 3: A Truly Open, No-Compromise Data Platform

A new architectural approach is required, one that delivers on the original promise of the lakehouse without forcing these debilitating compromises. This new paradigm is built on three core principles that directly address the shortcomings revealed in the market data: openness and interoperability, run everywhere with full fidelity, and bridge the gap between analytics and operations.

It must be founded on real openness and interoperability. A truly open platform must not only embrace open standards such as Apache Iceberg but also integrate seamlessly with a diverse and evolving ecosystem. Architecturally, this demands a 'catalog of catalogs' approach that unifies data discovery and access across disparate sources wherever they reside, without forcing the centralization of all metadata into yet another proprietary silo. This directly addresses the 13% of IT leaders who cite 'Integration Complexity' as a top point of dissatisfaction with their current data stack.

The platform must run everywhere with full fidelity. It must meet data where it resides, offering a consistent, high-performance, and fully managed experience across all major public clouds (Oracle Cloud Infrastructure (OCI), AWS, Microsoft Azure, and Google Cloud) as well as in private, on-premises environments. This is the only way to build a cohesive data strategy that honors data sovereignty requirements and reflects the distributed nature of enterprise data.

Crucially, it must bridge the gap between analytics and operations. The architecture must tear down the traditional wall separating these two heretofore separate data paradigms. This means creating a fluid, bi-directional data flow that allows AI insights that combine transactional and warehoused or LLM data to be directly and easily integrated into the core business applications that run the enterprise. The goal is to finally solve the 'last mile' problem and close the loop between insight and action.



## Part 4: Solution Spotlight: Oracle Autonomous AI Lakehouse – An Open Data Foundation for Enterprise AI

Oracle Autonomous AI Lakehouse is engineered to deliver on this new paradigm, providing a data platform without lock-in and without compromise, designed for the realities of the modern, AI-driven enterprise.

It is built on a proven, enterprise-class database. At its core, Autonomous AI Lakehouse combines the proven maturity of the Oracle Autonomous AI Database with native support for Apache Iceberg tables. This is not simply a bolt-on capability; it infuses Iceberg data with the full power of a converged database, including built-in tools such as Select AI for natural language queries and agentic frameworks, JSON-Relational Duality Views for developer agility, and AI Vector Search for semantic retrieval.

The platform is centered on genuine openness. It fully embraces Apache Iceberg and is built around the Autonomous AI Database Catalog, a true 'catalog of catalogs' that provides plug-and-play SQL access to data across Databricks Unity, AWS Glue, and Snowflake Horizon. This is accomplished without requiring disruptive data movement or costly duplication, enabling unified analytics across the entire data estate.

Autonomous AI Lakehouse is a cloud native design, and supports both multi-cloud and hybrid deployments. It offers a consistent, fully managed experience across all major cloud providers and on-premises via Oracle Exadata Cloud@Customer or OCI Dedicated Region. This uniquely allows organizations to build a modern lakehouse architecture that spans their entire data estate, honoring the fact that data has gravity and should be processed where it lives.

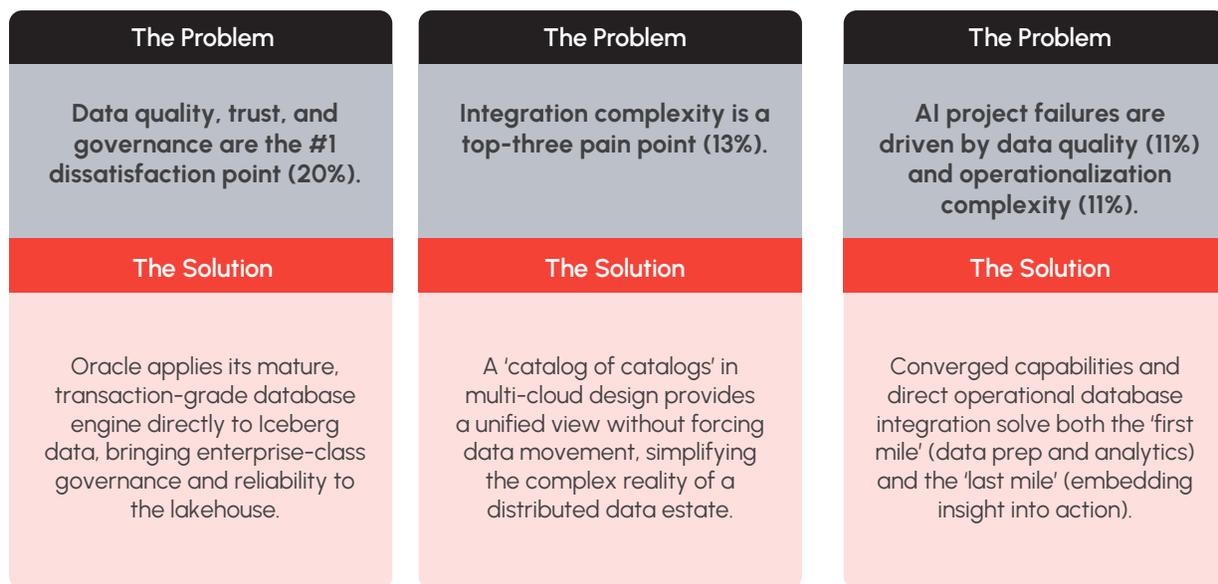
Finally, it coexists with existing investments. Recognizing that enterprises have significant investments in their current data stacks, the platform is designed to be fully interoperable with platforms such as Snowflake and Databricks. This allows organizations to extend and enhance their data capabilities without the disruption and risk associated with a 'rip and replace' migration, providing a pragmatic path to modernization.

# Part 5: Unlocking Exadata Performance and AI for All Your Data

Oracle Autonomous AI Lakehouse moves beyond basic interoperability by infusing open data with decades of enterprise-grade innovation, effectively eliminating the compromises that hold back lakehouse adoption. The connection between enterprise pain points and Oracle's architectural response is illustrated in Figure 3.

**Figure 3: Connecting the Solution to the Problem**

A Futurum survey<sup>4</sup> of data professionals reveals a clear set of frustrations holding enterprises back. Oracle's architecture is purpose-built to address them:



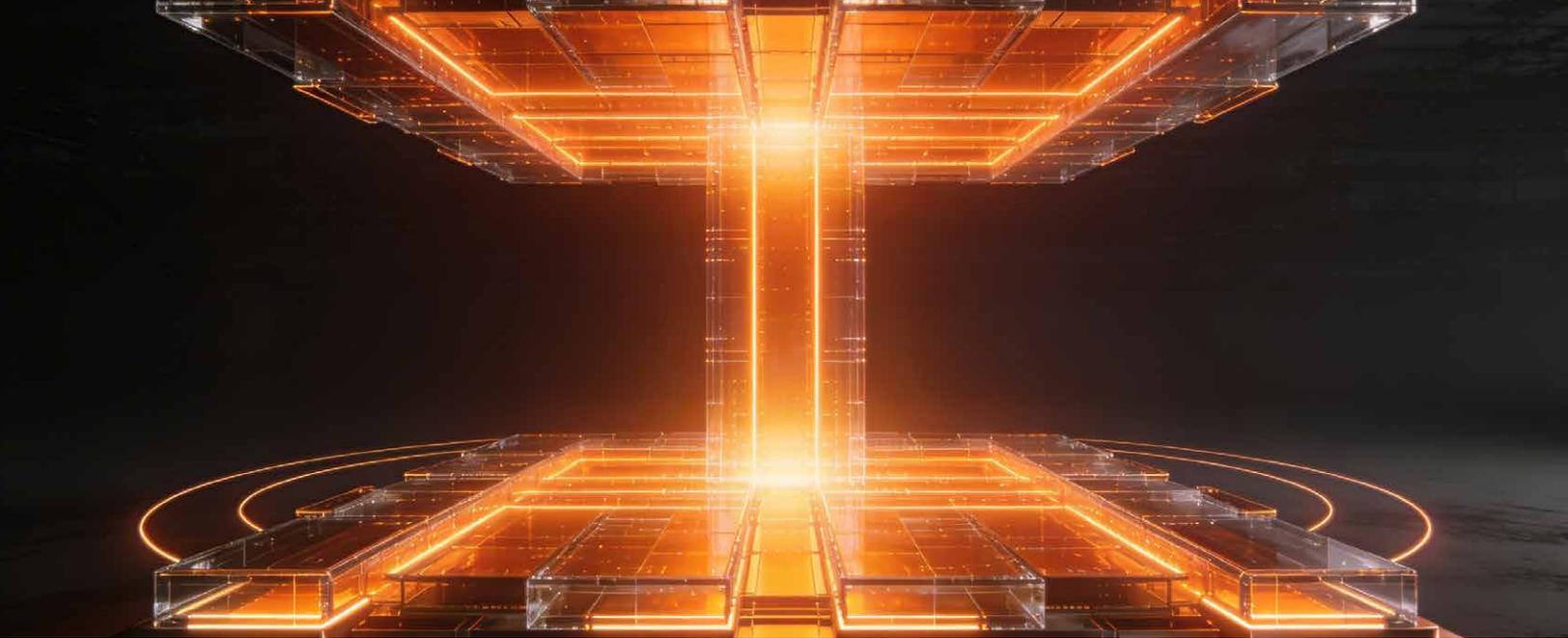
Source: Futurum Research, 1H 2025 Data Intelligence Decision Maker Survey.

Oracle Autonomous AI Lakehouse delivers a lakehouse without architectural or functional tradeoffs. The Data Lake Accelerator dynamically scales out compute resources for massive, parallel scans of multi-petabyte datasets. For interactive workloads, Exadata-based Iceberg table caching delivers up to 5x faster query performance on frequently accessed data. This closes the performance gap that often plagues open-format analytics.

Notably, the platform applies converged capabilities directly on open formats. The full power of Oracle's industry-leading converged database is now extended to data stored in Iceberg tables without requiring the underlying data to be moved. This means users can run sophisticated graph analytics, perform high-speed vector searches for RAG applications, and analyze complex spatial and JSON data, all within their existing lakehouse environment.

Most importantly, it enables AI-powered operations. In a move that truly sets the company apart, Oracle is extending these catalog capabilities to every Oracle Database. This means an on-premises operational database running a critical ERP system can directly query Iceberg data in a cloud-based lakehouse. This elegantly solves the 'last mile' problem, enabling AI insights to be seamlessly and instantly operationalized within the core transactional workflows that drive the business. No other platform bridges the analytical-operational divide so innately.

<sup>4</sup> Futurum Research (2025). 1H 2025 Data Intelligence, Analytics, and Infrastructure Decision Maker Survey Report



## Part 6: Conclusion and Recommendations

Achieving sustained success with AI depends far less on the novelty of the AI models and far more on the quality, accessibility, and performance of the underlying data foundation. The initial promise of the lakehouse was to provide this foundation, but the first wave of open-source solutions has forced enterprises into a series of difficult compromises. As a result, data practitioners' single most significant point of dissatisfaction with the current data stack remains data quality, trust, and governance (20%).

How to address this dissatisfaction? First, enterprise decision-makers must critically evaluate 'open' claims and avoid solutions that simply create new silos. Second, a pragmatic approach that respects existing investments and meets the multi-cloud reality is essential. And third, brings enterprise-class qualities to the data foundation.

Compared to other options, Oracle Autonomous AI Lakehouse is rather unique in that it provides architectural equivalency regardless of where it's deployed. For example, it runs on Exadata on OCI inside of AWS, Azure, and Google Cloud and in Cloud@Customer hybrid deployments. The software and hardware are the same no matter where you run it. That cannot be said for the likes of Snowflake and Databricks, for example, which rely on whatever infrastructure options are available to them across all of the hyperscalers. Customers seeking a same:same experience from hybrid cloud deployments to the hyperscaler of their choice cannot find a more optimal solution than Autonomous AI Lakehouse.

Additionally, the levels of AI-powered automation with respect to auto-indexing, auto-tuning, auto-scaling, auto-patching, and automatic threat detection and remediation offer organizations a level of capabilities such that they can focus on other priorities beyond managing their lakehouse environment. Particularly for organizations that are cost-conscious or lack data management practitioners, having an Autonomous AI Lakehouse is analogous to having an autonomous AI-powered robot or taxi handle tasks that enable you to increase productivity in more business-critical, strategic areas.

Oracle Autonomous AI Lakehouse provides that pragmatic approach and a clear, actionable path forward. By combining true, multi-vendor openness with the proven strengths of the Oracle Autonomous AI Database and Oracle Exadata in performance, security, and operational integration, Oracle delivers on the original promise of the lakehouse. It offers a unified, high-performance data platform for a multi-cloud world, ready to power the next generation of enterprise AI without compromise.

5 Futurum Research (2025). IH 2025 Data Intelligence, Analytics, and Infrastructure Decision Maker Survey Report

# Important Information About This Report

## AUTHORS

### Brad Shimmin

Vice President & Practice Lead  
Data Intelligence, Analytics, & Infrastructure

## PUBLISHER

### Futurum Research

## INQUIRIES

Contact us if you would like to discuss this report and The Futurum Group 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 "The Futurum Group." Non-press and non-analysts must receive prior written permission by The Futurum Group for any citations.

## LICENSING

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

## DISCLOSURES

The Futurum Group 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.

This Competitive Assessment report was commissioned by Oracle.

# Futurum<sup>®</sup>

## ABOUT THE FUTURUM GROUP

The Futurum Group is an independent research, analysis, and advisory firm, focused on digital innovation and market-disrupting technologies and trends. Every day our analysts, researchers, and advisors help business leaders from around the world anticipate tectonic shifts in their industries and leverage disruptive innovation to either gain or maintain a competitive advantage in their markets.

# Futurum<sup>®</sup>

CONTACT INFORMATION: The Futurum Group LLC | [futurumgroup.com](https://www.futurumgroup.com) | (833) 722-5337