White Paper, sponsored by:
Oracle

Operationalizing Agentic AI: Why the Data Layer Is the New Control Plane

WRITTEN BY:
Devin Pratt
Research Director, Data Management, IDC

PUBLISHED:
August 2026

Table of contents

IDC opinion **3**
Situation overview **4**
Real time has become the default 4
Agent infrastructure today is fragmented 4
Why the data layer is emerging as the control plane 4
Market growth follows the same direction 5
Innovate faster with AI designed for data **5**
Agent memory and the Unified Memory Core 5
Oracle AI Database Private Agent Factory and prebuilt agents 6
Select AI Agent framework 6
Autonomous AI Vector Database 6
AI Vector Search meets Apache Iceberg 7
AI Semantic Modeling 7
Additional agent-ready capabilities 7
Minimize AI data risk **7**
Deep Data Security at the source 8
Trusted Answer Search 8
Private AI Services Container 8
End data lock-in with open standards and frameworks **9**
Open Agent Spec 9
MCP servers and the Autonomous AI Database MCP Server 9
GoldenGate 26ai for event-driven AI 9
Conclusion **10**

IDC opinion

Enterprise agentic AI is no longer a research topic. Agents are in production at most enterprises, and the data platforms beneath them are struggling to keep up. IDC's February 2026 *Future Enterprise Resiliency and Spending Survey (FERS Survey)* Wave 1 (n = 903) finds that 72.1% of organizations already have AI agents running in production, with 51.6% running them across multiple business areas and 20.5% in a single area. Another 17.8% are in pilot, leaving only 1.3% with no plans for the next 12 months. IDC's April 2026 *Data Management Survey* (n = 715) confirms the spending trajectory, with 49.5% of organizations investing significantly under an established plan and 29.5% reporting several agentic AI applications already in production. The question is no longer whether to invest in agentic AI. It is whether the data layer beneath those agents can support them safely and at scale.

The bottleneck has shifted. In IDC's April 2026 *Data Management Survey,* 35.5% of organizations name data security, privacy, or regulatory and compliance constraints as the single biggest data-related roadblock to scaling generative and agentic AI, ahead of cost of preparation (34.3%) and inadequate vector and embedding infrastructure (23.2%). IDC's February 2026 *IT Data Management QuickPoll* (n = 107, 1Q26, United States) sharpens the picture. When U.S. IT leaders were asked what is actually slowing AI and agentic AI implementation, the leading answer was the security limitations of existing data platforms (43.9%), followed by industry compliance regulations such as PCI and HIPAA (39.3%) and unauthorized access to sensitive data (35.5%). The platform layer itself, not the model layer, is where enterprises are pointing.

Buyers have responded by consolidating. IDC's April 2026 *Data Management Survey* finds 96.8% of organizations actively consolidating data management capabilities into fewer products and vendors over the next 24 months. When the same survey asked where they will store and query vector embeddings for agentic AI workflows, respondents picked data warehouses or lakehouse platforms (63.2%) and integrated vector databases (DBs) (62.5%) ahead of search engines with vector capabilities (47.3%) and standalone native vector databases (30.2%). The direction is clear. Enterprises want vector search, memory, and retrieval delivered inside the platforms they already operate, not bolted on as separate point products.

Oracle's response, announced March 2026, is to architect agentic AI directly into Oracle AI Database 26ai around three pillars. The first is to innovate faster with AI designed for data. The second is to minimize AI data risk. The third is to end data lock-in with open standards and frameworks. The capabilities introduced under those pillars include Unified Memory Core, Deep Data Security, Trusted Answer Search, Private AI Services Container, Autonomous AI Vector Database, AI Vector Search with Apache Iceberg, Open Agent Spec, MCP servers, and GoldenGate 26ai. Taken together, they map directly to the buyer behavior IDC is measuring. Enterprises are moving agentic AI workloads into the platforms that already hold their operational data, and Oracle's portfolio is engineered for that move.

Key findings include the following:

- **Production agentic AI is mainstream.** 72.1% of organizations have agents running in production (source: IDC's *FERS Survey* Wave 1, February 2026; n = 903).

- **The data platform, not the model, is the bottleneck.** 43.9% of U.S. enterprises cite security limitations of existing data platforms as slowing AI and agentic AI implementation (source: IDC's *IT Data Management QuickPoll,* February 2026; n = 107; 1Q26).
- **Buyers are consolidating onto integrated platforms.** 96.8% of organizations are actively consolidating data management capabilities, and integrated vector databases (62.5%) outpoll standalone alternatives (30.2%) (source: IDC's *Data Management Survey,* April 2026; n = 715).
- **Semantic grounding is now considered table stakes.** 46.7% of U.S. IT leaders call a consistent semantic layer critical for safe and trustworthy AI, and another 38.3% call it important (source: IDC's *IT Data Management QuickPoll,* February 2026; n = 107; 1Q26).
- **Real time is the default expectation.** 91.3% of organizations require sub-200ms latency for real-time and streaming use cases (source: IDC's *Data Management Survey,* April 2026).

Situation overview

Real time has become the default

Agents do not tolerate latency. IDC's April 2026 *Data Management Survey* finds that 91.3% of organizations require end-to-end latency under 200ms for real-time and streaming use cases, with 66.0% in the 40–200ms band and 25.3% requiring ultra-real-time latency under 40ms. Real-time and near-real-time use cases together now account for 64.6% of all data-driven use cases across the surveyed enterprises. Batch has dropped to 35.4%. Architectures that cannot serve memory, retrieval, and action within those latency bands will struggle in production.

Agent infrastructure today is fragmented

Most agent stacks were assembled rather than designed. Vector stores, conversation history, knowledge graphs, and operational records typically live in different systems with different security models, different consistency guarantees, and different scalability profiles. The operational cost is visible in IDC's April 2026 *Data Management Survey.* When asked about their most challenging architectural issues, 30.6% of organizations cited security and governance, 29.7% cited data quality and consistency, 28.4% cited reliability and observability, and 26.4% cited supporting both real-time and batch workloads in the same environment. Among the workloads being consolidated, analytics and analytical platforms lead at 68.4%, followed by operational databases and data integration tied at 61.7%.

Why the data layer is emerging as the control plane

Every agent decision is, at its core, a decision about data. Which data can the agent see? Which data can it change? Which data shapes its memory? Which data triggers its actions?

Moving those decisions out of the application layer and into the data layer is the only way to guarantee they are applied uniformly across thousands of agents and across the long life cycle of an agentic workflow.

Buyers see this most clearly when describing converged workload requirements over the next 12–24 months. The single most-cited scenario was AI assistants or agents taking actions on current data and events (52.0%), ahead of GenAI with retrieval-augmented generation (RAG) (50.0%) and operational analytics on live data (49.4%). Agents are now the leading use case driving converged operational and analytical workloads, and IDC predicts that by 2029, 60% of enterprise databases will unify operational and analytical workloads within converged architectures to support real-time, AI-powered applications (see *IDC FutureScape: Worldwide Data and Analytics 2026 Predictions,* IDC #US53860225, October 2025).

Market growth follows the same direction

IDC forecasts worldwide DBMS software revenue to grow from $93.2 billion in 2024 to $168.6 billion in 2029, a 12.6% CAGR (see IDC's *Worldwide Database Management Systems Software Forecast, 2025–2029,* IDC #US53032525, June 2025). Public cloud DBMS services are on an 18.3% CAGR trajectory; on premises is essentially flat at -0.4%. Open table format adoption is rising on the same curve, from 40.6% one year ago to 52.2% today, with 63.2% projected in the next 12–18 months (source: IDC's *Office of the CDO 2025 Survey,* November 2025; n = 417). AI-driven workloads are concentrating in elastic, integrated, open environments, and the DBMS vendors best positioned to win are those that consolidate vector search, retrieval, memory, and policy enforcement into the platforms where enterprise data already lives.

Innovate faster with AI designed for data

Oracle's first pillar is about making it easier and faster to build, deploy, and operate data-centric agents and vector-based applications without first extracting, duplicating, or moving the data those agents need to reason against. The unifying idea is architectural. Memory, retrieval, and execution should run where the data lives, not in a separate tier above it.

Agent memory and the Unified Memory Core

Memory is one of the missing pieces in most enterprise agent stacks. Agents need to improve over time by remembering preferences, plans, facts, and other knowledge. Today's memory solutions are fragmented across vector stores, document databases, and external services with no shared guarantees of security, consistency, or scalability. A round trip to a separate system every time the agent wants to remember something is incompatible with how agents actually work.

Oracle's Unified Memory Core treats memory as a native database capability. Oracle AI Database 26ai provides built-in support for every layer of agent memory inside a single converged platform. Short-term (working) memory holds conversation snapshots, vector-

summarized history, and context cards. Long-term experiential memory captures generalizations, instructions, and plans stored as JSON and retrieved through vector or hybrid search. Long-term factual memory lives in knowledge graphs and structured records, retrieved through hybrid semantic and relational search. The Unified Memory Core is exposed through a Python library and through the DBMS_AGENT_MEMORY PL/SQL package, with adapters that allow open source memory libraries such as Mem0 and Letta to transparently use Oracle for storage and retrieval. One system of record for agents, memory, and enterprise data is simpler to secure and operate than a stitched-together stack.

Oracle AI Database Private Agent Factory and prebuilt agents

Oracle AI Database Private Agent Factory is a no-code, containerized platform that lets business users, domain experts, and engineers build and deploy data-centric agents and workflow automations. What distinguishes Oracle AI Database Private Agent Factory is Oracle's choice to place agent development closer to the data platform rather than treating it as a separate orchestration layer. Three prebuilt data-centric agents ship with the factory. The Knowledge Agent augments AI Vector Search and LLM capabilities with enterprise data. The Data Analysis Agent interacts with enterprise databases, understands schema structure, and extracts semantic insights automatically. The Deep Research Agent decomposes a user's question into sub-questions and produces an in-depth analysis. Oracle's worked example, SupremoBite, illustrates the pattern. A hierarchical multiagent food-delivery assistant handles a customer refund through four cooperating agents, with memory, structured records, business policy, and transactional updates all running inside a single governed platform.

Select AI Agent framework

For developers who prefer to define agents in code rather than in a visual designer, Oracle's Select AI Agent framework provides a declarative model in PL/SQL or Python built around four primitives. Tools are callable functions such as vector search. Tasks are logical units of work that use one or more tools to produce structured output. Agents are stateful workflows composed of multiple tasks. Teams are collections of agents that work together to complete a multistep workflow. Oracle also ships prebuilt agents for code assistance, natural language–to-SQL, and RAG workflows and a Data Science Agent that handles data profiling, feature engineering, and machine learning (ML) model building inside the database.

Autonomous AI Vector Database

For developers and data scientists who want a purpose-built, fully managed vector database, Oracle's Autonomous AI Vector Database provides self-driving, self-securing, and self-repairing operations behind an API-first interface. The service is available through a Python SDK, REST API, and PL/SQL and SQL interfaces, with a Vector Database Console for quick-start and

management. A Free Tier on Oracle Cloud supports exploration with two ECPUs and 20GB of storage, and a one-click upgrade path takes customers to the full Autonomous AI Database. The product aligns with the integrated vector buyer preference documented previously in this paper.

AI Vector Search meets Apache Iceberg

Apache Iceberg has become a leading open table format for analytical data. Among formats in production use today, Delta Lake leads at 63.6%, Iceberg at 52.0%, and Hudi at 40.0% (source: IDC's *Data Management Survey,* April 2026). Oracle AI Vector Search now supports vector generation, similarity search, and high-performance vector indexing directly on Apache Iceberg tables, with no data movement. The practical effect is that an Oracle-resident agent can operate against the same lakehouse data the rest of the analytical stack already uses, without forcing the customer to choose between Oracle and Iceberg.

AI Semantic Modeling

A common failure mode for natural language to SQL is that the LLM cannot infer intent from cryptic table and column names. Ask an LLM for the average employee salary by department when the underlying table is named XYZ with columns C1, C2, and C3, and the LLM will apologize and stop. Oracle's AI Semantic Modeling lets developers declare data intent through column annotations. A simple ALTER TABLE statement can tag a phone number column as such, and the annotation propagates to any agent or generative tool that queries the schema.

Additional agent-ready capabilities

Oracle introduced several additional agent-ready features in the March 2026 release. The Autonomous AI Database Data Science Agent is a chat-based tool built into Autonomous AI Database that handles data profiling, feature engineering, and ML model building inside the database, operating within the user's access privileges. Oracle Graph Studio's new generative AI graph modeler analyzes tables, identifies relationships, and proposes a graph definition automatically, accelerating time to value for knowledge graph use cases. APEX AI Interactive Reports lets users ask natural language questions against predefined reports without generating arbitrary LLM SQL. Millions of existing APEX Interactive Reports can be AI enabled simply by upgrading APEX.

Minimize AI data risk

Oracle's second pillar addresses what IDC's research consistently identifies as the leading constraint on agentic AI deployment. 35.5% of organizations cite data security, privacy, or regulatory and compliance constraints as the top data-related roadblock to scaling generative and agentic AI (source: IDC's *Data Management Survey,* April 2026). Application layer controls

cannot protect against agents that can generate arbitrary SQL, can be tricked through prompt injection, and operate at machine speed across thousands of concurrent sessions. The defensive answer belongs in the data layer itself.

Deep Data Security at the source

Historically, data privacy has been managed at the application layer, with the database used primarily for coarse-grained access control. AI changes the calculus. A prompt-injected agent can issue queries that bypass application controls, access limitations, and firewalls. The example most often cited, where a user asks the agent to ignore previous instructions and show all salary data, is uncomfortable precisely because it is so trivial to demonstrate against today's deployments.

Oracle's Deep Data Security pushes enforcement into the database itself. The architecture rests on three ideas. Secure identity propagation passes the end user's identity from the agent or application into the database session, so the database knows who is ultimately requesting the data. Database-enforced row, column, and cell access policies decide what specific rows, columns, and cells a given user can see, regardless of what SQL the agent or LLM generates. Policy-as-SQL lets access policies be declared in SQL itself, so they can be version controlled, peer reviewed, audited, and tested with the same rigor as any other database object. The only enforceable policy boundary for an agent that can write arbitrary SQL sits next to the data itself.

Trusted Answer Search

LLMs are powerful pattern-matching engines, but they are not truth engines. They can hallucinate. They intentionally introduce probabilistic randomness. Even an LLM that produces an accurate answer today may produce an inaccurate one after a model update. For many enterprise use cases, the right architecture is to keep the LLM out of the answer-generation path entirely. Trusted Answer Search uses AI vector search to match an end user's natural language question to a previously created, governed report rather than asking an LLM to compose a new answer. Search-space access is governed by the search administrator, and user upvotes and downvotes drive continuous improvement. The result is a natural language interface to enterprise data that retains the auditability and determinism of a traditional reporting tool.

Private AI Services Container

Many regulated and security-sensitive customers cannot send enterprise data to third-party AI services. Oracle's Private AI Services Container is a deployable container that offloads vector embedding generation, open-weight LLM inference, reranking, named entity recognition, and predictive ML to a local private server, exposed through an OpenAI-compatible REST API. The initial release focuses on vector embedding generation and ships with several leading text and image embedding models, with the ability to load additional models of the customer's choice.

Deployment options include on premises, air-gapped, private cloud, public cloud, and the Oracle Database Operator for Kubernetes.

End data lock-in with open standards and frameworks

Oracle's third pillar is openness. The message is that enterprises should retain the ability to substitute models, frameworks, and deployment platforms without rebuilding their agentic AI stack. Model Context Protocol (MCP) has had the fastest protocol adoption curve of any AI interface to date. Agent2Agent (A2A) is emerging for interagent communication. The agent specification layer is the next frontier.

Open Agent Spec

The agent framework ecosystem is fragmenting. Each framework reinvents how to define data access, reasoning, memory, and workflow. MCP and A2A describe how agents access tools and how agents talk to one another but say nothing about the internal structure of an agent. Open Agent Spec is an open source, framework-agnostic configuration language for defining agents and agentic workflows. The intent is to do for agents what ONNX did for ML models. Adapters allow an agent defined in one framework to execute in another, providing customers with insurance against framework-level lock-in.

MCP servers and the Autonomous AI Database MCP Server

Oracle has built MCP support in two places. The SQLcl-powered MCP server lets AI models list connections, connect, run queries, and disconnect against any supported Oracle Database, with passwords encrypted in local wallets and the LLM never seeing credentials. The Autonomous AI Database MCP Server is a fully managed multitenant MCP server embedded directly in Autonomous AI Database Serverless and Dedicated Region in both 26ai and 19c, with dynamic tool discovery, integration with Oracle Database Identity and OCI IAM, and support for external OIDC and OAuth 2.1. The point of both is to make the Oracle database addressable through the same standard interface the rest of the agent ecosystem is converging on.

GoldenGate 26ai for event-driven AI

Agents are most valuable when they can react to business events in real time. The velocity and volume of those events, often millions of subsecond transactions in a typical enterprise, are incompatible with traditional batch or microbatch data movement. GoldenGate 26ai delivers trusted, exactly-once business events to AI platforms in real time, transforming the agentic AI

conversation from better answers to automated action. Plugged into an agentic mesh with AsyncAPI automations, GoldenGate becomes the sensory layer that lets reactive AI agents sense and respond to business events as they happen, across Oracle technologies and third-party databases, vector sources, and messaging systems.

Conclusion

Enterprise agentic AI will be limited not by the quality of the underlying models but by the quality of the data infrastructure that surrounds them. Memory must be durable and governed. Security must be enforced at the source. Real-time context must be guaranteed. Openness must be preserved so that enterprises retain the right to evolve their stack. And the platform must be one that buyers already trust to run their business. The buyer signal is consistent. 72.1% of enterprises have agents in production today (source: IDC's *FERS Survey* Wave 1, February 2026). 96.8% are consolidating data management capabilities. 35.5% identify data security and compliance as the biggest barrier to scaling agentic AI (source: IDC's *Data Management Survey,* April 2026).

Oracle is well positioned for this moment. Its response is technically coherent and strategically aligned with the market's direction, and the capabilities introduced in March 2026 sit atop an installed base that already runs much of the world's most valuable transactional data. Oracle was named a Leader in *IDC MarketScape: Worldwide Analytical Databases 2025–2026 Vendor Assessment* (IDC #US52973925, November 2025), where IDC specifically cites broad in-database analytics capabilities, including machine learning, graph, spatial, JSON, and vector, as a differentiator that reduces data movement and governance complexity. The converged database posture aligns with the integration preference IDC has been measuring, and the installed base in the verticals most advanced on agentic AI gives Oracle a receptive audience for the new capabilities. The risks worth watching are real but manageable. The credibility of Oracle's open standards posture depends on the ecosystem around MCP, Open Agent Spec, and Apache Iceberg continuing to mature. Competitive pressure on autonomous database operations remains strong, with databases such as IBM's Db2 Genius Hub recently introduced.

For data and AI leaders evaluating agentic AI architectures, IDC recommends asking three questions of any candidate platform. First, where does your agent memory live, and who governs it? Memory that lives in a separate tier from the data the agent reasons against will fail under production pressure. Oracle's Memory Core is one architectural answer, placing memory inside the database itself. Second, where are your data access policies enforced when an LLM generates the SQL? Application layer controls are insufficient against agents that can write arbitrary queries. Oracle's Deep Data Security addresses this directly, with enforcement at the row, column, and cell. Third, how do you preserve the option to substitute frameworks, models, and deployment platforms over the next five years? Platforms that lock you into a single agent framework or a single vector store will be expensive to evolve. Oracle's support for Open Agent Spec, MCP, Apache Iceberg, and multicloud deployment is designed to keep those options open. Platforms that can answer all three at the data layer rather than the application layer are best positioned to move agentic AI from pilot to production at scale.

About the IDC analyst



Devin Pratt
Research Director, Data Management

Devin Pratt is a research director within IDC's AI, Automation, Data, and Analytics (AAD&A) practice, where he oversees the database management tools and technologies software market. His primary research centers on the evolution of database management tools and technologies, covering both current and future capabilities across various relational, non-relational, and dynamic database systems designed for operational or analytic data workloads. In addition, Mr. Pratt's research encompasses related technologies used for data modeling, database development, optimization, and maintenance. By analyzing market trends, buyer behaviors, and the business value of these technologies, Mr. Pratt's research helps vendors enhance their products and refine marketing strategies while also guiding end users in selecting data management solutions for challenges such as cloud migration and AI-driven data initiatives.

More about Devin Pratt

IDC Custom Solutions

IDC Custom Solutions produced this publication. The opinion, analysis, and research results presented herein are drawn from more detailed research and analysis that IDC independently conducted and published, unless specific vendor sponsorship is noted. IDC Custom Solutions makes IDC content available in a wide range of formats for distribution by various companies.

This IDC material is licensed for external use, and in no way does the use or publication of IDC research indicate IDC's endorsement of the sponsor's or licensee's products or strategies.

IDC Research, Inc.
One Beacon Street, Suite 33100, Boston, MA 02108, USA
T +1 508 872 8200

blogs.idc.com | LinkedIn @IDC | www.idc.com

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications, and consumer technology markets. With more than 1,300 analysts worldwide, IDC offers global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries. IDC's analysis and insight help IT professionals, business executives, and the investment community make fact-based technology decisions and achieve their key business objectives.

©2026 IDC. Reproduction is forbidden unless authorized. All rights reserved. CCPA