Deloitte. | databricks

THE AGENTIC AI BLUEPRINT FOR DATA QUALITY MANAGEMENT:

A MULTI-AGENT APPROACH TO ENTERPRISE DATA QUALITY



A Deloitte and Databricks roadmap

High quality data is the bedrock of effective artificial intelligence (AI). And yet, organizations often struggle with fragmented systems, manual-intensive processes and inconsistent data standards. To unlock real-time insights and advanced analytics, organizations need a scalable, proactive, and automated approach to data quality management (DQM).

This paper provides a practical blueprint for modernizing DQM with Deloitte and Databricks by deploying orchestrated AI agents to automate, unify and scale DQM operations across the full data lifecycle. ***With an agent-based architecture, organizations can:***

- Dynamically harmonize data across functions
- Aggregate and reconcile metrics in real time
- Continually improve the health of enterprise data assets

This approach can empower organizations to shift from reactive DQM "firefighting" to a proactive, intelligent and insights-driven paradigm, setting a stronger foundation for trustworthy analytics and innovation. And with Deloitte and Databricks providing the technology foundation, organizations can detect and resolve data quality issues faster, achieve broader and deeper coverage, and establish standardized, business-ready DQM health metrics.

Gaps in traditional approaches to data quality management

Traditional DQM approaches are often reactive and resource-heavy. They cannot keep pace with the scale, pace and complexity of today's enterprise data environments.

Effective DQM consists of six foundational pillars:

1. *Accuracy*
2. *Completeness*
3. *Consistency*
4. *Timeliness*
5. *Uniqueness*
6. *Validity*

Without robust measurement and continuous stewardship of these dimensions, analytic outcomes may become unreliable, decisions are delayed and operational risk accumulates. Yet, achieving excellence across all six pillars is elusive with conventional methods, especially when organizations are constrained by labor-intensive tools and disconnected workflows.

A new paradigm: Multi-agent systems for enterprise data

Agentic AI is redefining how organizations approach data management—shifting from siloed, manual interventions to a coordinated ecosystem of specialist agents, each designed to tackle a specific aspect of the modern data landscape. In this model, intelligent agents for data ingestion, lineage, observability, metadata and quality assurance work collaboratively under the guidance of an orchestration layer. This orchestrator plans, assigns and monitors workflows, helping data move through the pipeline efficiently while meeting organizational policies and compliance requirements.

What is agentic AI in data management?

Agentic AI in data management refers to a coordinated system of modular AI agents, each specializing in a distinct component of the data lifecycle. Together with a powerful orchestration layer, these agents enable scalable, automated and resilient data operations that can adapt to dynamic business needs.

Core capabilities of agentic AI

Some of agentic AI's core capabilities include:

Planning and reasoning: The orchestrator agent dynamically allocates resources, designs efficient workflows and adapts to data changes.

Specialized tool use: Individual agents leverage leading tools for their assigned tasks, such as ingestion frameworks, DQM engines, or lineage trackers.

Evaluation and telemetry: Continuous monitoring, validation and performance benchmarking allow agents to self-optimize and alert human operators to anomalies.

Guardrails and controls: Automated enforcement of policies, security rules, and business logic is installed across all touchpoints.

Human-in-the-loop: Critical exceptions, ambiguous cases and key policy controls are escalated to (or confirmed with) human specialists, confirming trust and alignment with business objectives.



Deep dive: Agentic AI for data quality

While agentic AI brings advancement across all domains of data management, its impact is particularly profound in DQM. Specialized DQM agents can continuously monitor, assess and remediate data, creating a system that is not only automated but also adaptive and transparent.

The multi-agent framework supporting the end-to-end DQM lifecycle

Effective DQM should behave more like a system than a workflow. A multi-agent framework enables enterprise-scale data quality by decomposing the DQM lifecycle into autonomous, stateful agents that operate continuously and coordinate through orchestration rather than static workflows.

Each agent specializes in a core function—profiling, rule inference and validation, cleansing, remediation, observability or reporting—while reasoning over shared context such as metadata, lineage, historical quality signals and business criticality.

Then, an orchestration layer governs agent interaction, prioritizes actions based on impact and risk, and determines when automation is appropriate versus when human intervention is required. This architecture replaces brittle, tool-centric pipelines with an adaptive control plane for data quality capable of learning from outcomes, responding to change in real time and scaling trust without scaling operational overhead.

Step-by-step: How agentic AI drives data quality

1 *Data profiling*

- Event-driven and scheduled profiling triggered by schema or ingestion changes
- Column-level statistics, entropy, cardinality and distribution analysis
- Schema and semantic drift detection using embeddings
- Machine learning (ML)-based anomaly detection and predictive baselining
- Cross-dataset correlation and stateful profiling history

2 *Rule generation and validation*

- Adaptive rule inference from profiling and metadata
- Executable, versioned rule policies with lifecycle tracking
- Inline and asynchronous validation enforcement
- Rule effectiveness scoring and downstream impact simulation
- Explainable validation failures using large language model (LLM) reasoning

3 *Data cleansing and standardization*

- Orchestrated, multi-step cleansing workflows
- Probabilistic deduplication and record linkage
- Confidence based auto-correction with escalation thresholds
- Learning loops from steward feedback
- Full remediation lineage and audit trails

4 *Remediation and root cause analysis*

- Stateful issue lifecycle management
- Business-impact driven prioritization
- Lineage- and observability-based root cause analysis
- Automated task routing and service level agreement (SLA) tracking
- Preventative pattern detection and upstream fix recommendations

5 *Monitoring, measurement and reporting*

- Continuous quality state evaluation
- Enterprise data quality index (DQI) computation
- Freshness, volume and pipeline reliability monitoring
- Trend-based degradation detection
- Dashboard, application programming interface (API) and conversational access

6 *Governance and stewardship enablement*

- Metadata enrichment and sensitive data classification
- Policy enforcement via orchestration guardrails
- Human-in-the-loop approvals for high-risk actions
- Explainability, traceability and compliance-ready auditing
- Databricks Unity Catalog as a core to maintaining the context layer of learned knowledge

The Deloitte and Databricks difference

Drawing on the extensive real-world deployments and collaborations through our alliance with Databricks, Deloitte leverages agentic AI to deliver next-generation DQM for complex, regulated enterprises. This alliance relationship can provide robust frameworks, cloud-native scalability and rapid innovation—allowing clients to bridge the gap between vision and action.

Benefits of the Deloitte and Databricks collaboration include:

- Orchestrated, multi-agent systems for high-volume, high-velocity data environments
- Accelerated time-to-insight and error reduction through automated telemetry and guardrails
- Effective integration across leading enterprise data and AI platforms

Agent Bricks by Databricks supports DQM by shifting from traditional manual checks to an agentic AI framework that automates the entire quality lifecycle. This framework leverages intelligent agents to monitor, detect and resolve issues with minimal human intervention.

The framework also provides two primary models to address a range of organizational needs:

- **Agent Bricks:** A low-code solution for rapid assembly and deployment of data quality agents, ideal for standard automation tasks with minimal configuration.
- **Agent Framework:** A customizable workflow that allows developers to build advanced agents with specialized logic, real-time response scoring and deep system integration.

> *Deloitte helps integrate these Databricks capabilities with strategic, customized solutions tailored for each client and their practical DQM needs.*

How Agent Bricks by Databricks transforms DQM

Contextual reasoning:

Unlike static rules, these agents use a semantic understanding of business data to identify nuances and anomalies that traditional systems might miss.

Automated monitoring and detection:

Agents learn historical data patterns and seasonal behaviors, such as weekend volume dips, to detect anomalies across entire schemas without manual threshold setting.

Intelligent prioritization:

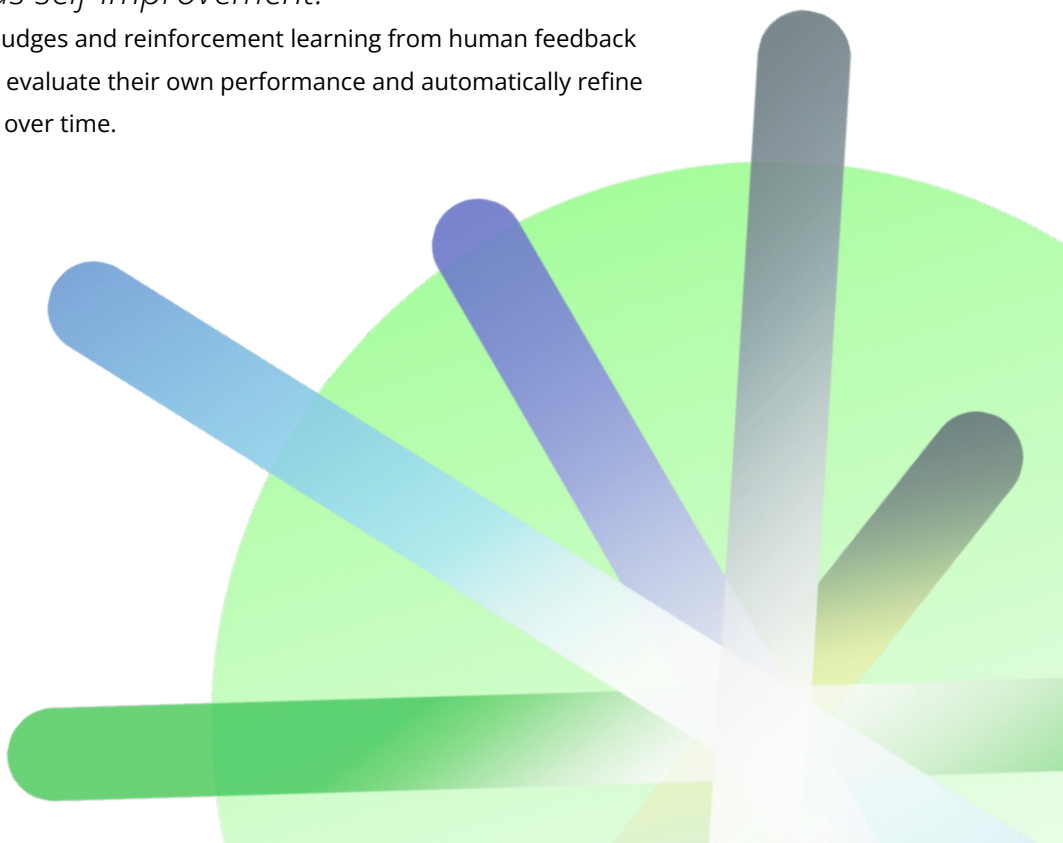
Using Unity Catalog lineage, the system identifies which datasets are most critical and prioritizes the resolution of quality issues that impact high-value downstream applications.

Root cause analysis (RCA):

The framework traces issues back to upstream Lakeflow jobs or pipelines, allowing teams to quickly investigate specific failures or code changes.

Continuous self-improvement:

Through LLM judges and reinforcement learning from human feedback (RLHF), agents evaluate their own performance and automatically refine their accuracy over time.



The Deloitte and Databricks model in action

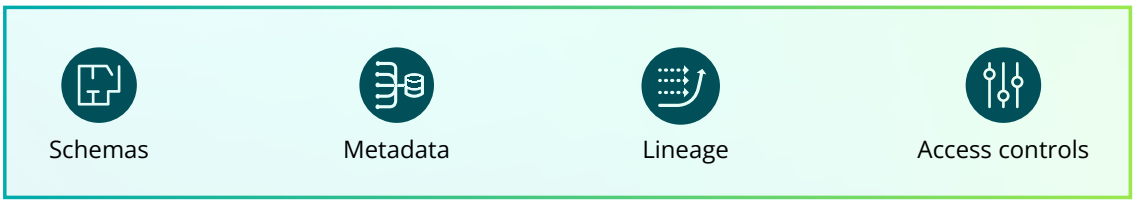
Deloitte configures customized DQM logic, maps industry-specific requirements, and manages governance, compliance and change management.

Databricks delivers the core agent technology, infrastructure and AI orchestration layer. The Agent Bricks Agent Framework turns governed data quality into something teams can run at scale.

At its core, these capabilities provide a secure, governed foundation for creating, deploying and coordinating intelligent data quality agents on Databricks. Agent Bricks speeds up adoption of common agent interaction patterns, such as extracting insights from unstructured data, and enabling natural-language interaction over structured data using AI/BI Genie.

This makes it easier for users and systems to engage with data quality signals. Teams implement more advanced data quality workflows—multi-step validation, remediation orchestration and cross-dataset reasoning—as code-first agents within the same framework. This approach helps extend capabilities while maintaining a consistent, enterprise-grade execution and governance model.

Agents can then run over a shared, governed data foundation built on the Lakehouse. ***Lakeflow prepares and orchestrates data and quality signals at scale, while Unity Catalog centralizes governance across Databricks-native and federated data sources, including:***



This approach breaks down data silos and helps agents reason against consistent enterprise definitions. When agents need semantic understanding—such as aligning concepts across datasets or interpreting unstructured content—they use **Databricks Vector Search** to retrieve relevant definitions, guidance, and logic. Agents also use Databricks Foundation Models, including frontier LLMs, to support reasoning across data quality workflows.

Integrated evaluation, monitoring, identity and governance enable safe scaling across the full agent lifecycle. Before deployment, MLflow-based evaluation pipelines validate agent behavior and model outputs, using automated quality checks to assess accuracy, consistency and alignment with defined guidelines.

In production, MLflow monitoring captures live telemetry from agent interactions and model inferences and tracks performance trends and outcome quality over time. The Lakehouse stores these signals, and Lakehouse Monitoring surfaces them to support continuous oversight and controlled iteration of agents. Finally, Unity Catalog enforces identity and access controls so agents operate only within approved data, model and usage boundaries—helping scale data quality automation while meeting enterprise compliance and trust requirements.

Measurable enterprise outcomes

Agentic AI is changing data quality fast. It moves organizations from one-off fixes to measurable enterprise outcomes. In Deloitte benchmarks and client work, we see AI-driven automation strengthen data integrity and coverage, and reduce operational risk and cost in the process

Additional benefits include:

Reduction in operational data quality issues

By orchestrating data quality activities through intelligent agents, organizations can realize a tangible reduction in recurring data errors and exceptions. Automation applies quality rules consistently, detects anomalies in real time, and routes remediation efficiently so teams spend less time firefighting and more time resolving issues proactively. Representative benchmarks have shown up to a 20% reduction in operational data quality issues, translating into significantly fewer service disruptions and downstream process inefficiencies.

Productivity gains and cost savings

Agentic AI delivers productivity gains by automating previously labor-intensive tasks such as data profiling, monitoring, validation and remediation. This acceleration of core data quality processes can help skilled resources to focus on high-value data stewardship and governance initiatives.

Expanding coverage & accelerating benefits

As AI agents scale across more data domains and use cases, the coverage and efficacy of enterprise data quality programs increase substantially. Organizations can now enforce data quality standards consistently across hundreds of systems and domains, promoting compliance and resilience. This can result in more reliable analytics, faster delivery of new products and insights, and a stronger compliance posture bolstered by data that business collaborators trust.

Enabling a cycle of improvement

Agentic AI empowers organizations to continuously learn and adapt. As data quality agents orchestrate interventions and capture new issues, these insights feed back into the knowledge base, enabling ongoing tuning of rules and models. This cycle propels sustained improvements, sharper risk detection and greater business agility over time.

Phased approach: From foundation to scalable excellence

Implementing a modern DQM strategy—especially when leveraging agentic AI—demands a clear phased plan and focus on organizational readiness. ***Based on field experience and client-tested methodologies, a successful DQM transformation can be mapped across four progressive phases:***

1 *Define/Pilot*

This initial phase centers on articulating the vision and piloting targeted use cases.
Key activities include:

- Identifying high-impact data assets and critical data quality issues
- Deploying AI agents for automating targeted DQM rules and tasks, demonstrating tangible early wins
- Testing integration with existing data platforms and tools to ensure feasibility and value

2 *Build*

Once they validate concepts and establish momentum, organizations can formalize processes and expand coverage by:

- Standardizing DQM rule execution, reporting and remediation workflows
- Expanding the library of automated rules, leveraging AI agents to orchestrate rule execution and enhance decision-making
- Deepening integration with core data systems, governance platforms and stewardship processes

3 *Scale*

The focus in this phase shifts to expanding both the breadth of data sets covered and the automation of DQM operations by:

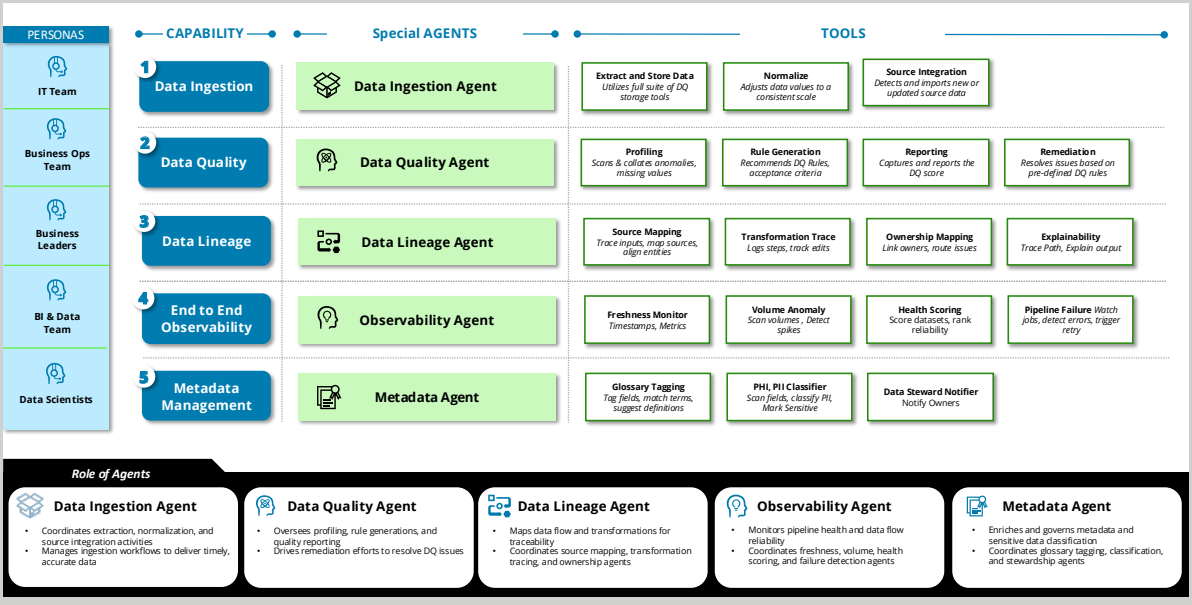
- Rolling out AI agent-powered DQM management across additional domains, business units and geographies
- Embedding continuous monitoring, exception management and feedback loops to accelerate learning
- Quantifying value through improved data quality metrics, operational efficiency and business outcomes

4 *Sustain and transform*

Long-term transformation hinges on continuous improvement, proactive stewardship and future-proofing by:

- Institutionalizing DQM as a core enterprise capability, driven by AI agent-based orchestration
- Advancing self-service capabilities, automating root-cause analysis and supporting cross-functional data quality collaboration
- Regularly reviewing and refining frameworks, integrating the latest advancements in agentic AI and dynamic rule automation

Example: *Multi-agent approach to enable scalable data quality capability*



Operating model enablers: Creating the environment for success

A robust implementation is supported by a set of operating model enablers:

Core DQM team:

Build a multidisciplinary data quality function responsible for strategy, stewardship and operational execution.

Stewardship and change management:

Drive buy-in through clear roles, transparent communication and targeted upskilling, ensuring business and technical teams are aligned and empowered.

Tooling and integration:

Effectively connect new AI-driven DQM capabilities with existing data platforms, analytics tools and process automation solutions to maximize return on investment and minimize disruption.

Federated relationships:

Foster strong collaboration across sectors, functions and regions—sharing best practices, aligning on standards and accelerating adoption through strategic governance.



Strategic coverage expansion and automation

The Deloitte and Databricks roadmap is designed for progressive coverage expansion and increased automation—from piloting in select domains to enterprise-wide sustained transformation. At each step, the use of AI agents to automate rule execution and reporting not only scales data quality management, but makes it adaptive and intelligence-driven.

Successful deployment of an agentic AI–enabled data quality framework requires more than technology enablement. It demands foundational organizational alignment, governance clarity and responsible AI controls embedded by design.

It is important for organizations to prioritize the following six factors for their agentic AI-driven DQM transformation:

1 *Leadership sponsorship and governance alignment*

Sustained executive sponsorship is critical to prioritize investment, align cross-functional stakeholders and ensure enterprise-wide adoption.

A clearly defined accountability model, anchored in a federated governance structure, should delineate ownership across:

- Executive sponsors (strategic direction and oversight)
- DQM core team (framework design, standards and orchestration)
- Domain data stewards (business rule ownership and quality thresholds)
- Platform and ML engineers (agent deployment, monitoring and scalability)
- Value realization analysts (KPI tracking and measurable impact)

Shared accountability across business and technology functions is essential to embed DQM into operating rhythms, rather than positioning it as a parallel initiative.

2 *Federated domain ownership*

Agentic AI systems must operate within clearly defined domain boundaries. Cross-functional domain ownership ensures business rules are defined, validated and continuously maintained in alignment with enterprise standards. This federated model strengthens trust, accelerates remediation and improves decision velocity while maintaining consistency at scale.

3 *Responsible AI and model risk management*

Given the autonomous nature of agentic systems, guardrails must be embedded to manage model risk and bias. ***Key considerations include:***

- Continuous model evaluation and performance monitoring utilizing Databricks AI Gateway
- Bias detection and mitigation strategies
- Lineage-based explainability to trace decisions back to source data and rule logic
- Transparent audit trails and logging mechanisms

Explainability is critical in regulated environments, where the traceability from AI decisions to the underlying data elements they are pulled from must be demonstrable.

4 *Security, privacy and access controls*

Agentic DQM frameworks must operate within strict enterprise security standards. ***These include:***

- Role-Based Access Control (RBAC) and least-privilege principles
- Secure handling of sensitive and personally identifiable information (PII) data
- Environment segregation across development, testing and production
- Comprehensive logging for auditability and compliance

Security architecture should be designed in tandem with AI orchestration layers to avoid downstream rework.

5 *Human oversight and decision governance*

Human oversight remains a non-negotiable control point.
Clearly defined human-in-the-loop checkpoints should be established for:

- High-impact rule changes
- Material data corrections
- Model retraining or drift remediation
- Exception handling and escalations

This ensures that AI autonomy enhances—not replaces—enterprise accountability.

6 *Adoption and change enablement*

Adoption must be intentionally driven. Embedding the DQM framework into ongoing AI and data transformation programs accelerates scale and prevents fragmentation. Success depends on integrating agentic capabilities into existing workflows, KPIs and performance management structures, which Deloitte and Databricks can help drive at scale.

Looking ahead to the future of data quality management

Early results from agentic AI adoption demonstrate quantifiable business value, including reduced operational issues, lower costs, broader coverage and self-reinforcing improvement—all positioning data quality as a catalyst for enterprise growth and resilience.

Enterprises that intend to compete on AI and data-driven decisioning are approaching a structural inflection point. Traditional data quality capabilities built around manual controls, static rule sets and periodic assessments cannot sustain the scale, speed and autonomy required by modern analytics and AI workloads. These approaches remain reactive by design, addressing quality issues only after trust has been eroded.

Implementing a multi-agent–powered data quality framework, such as the combined approach that Deloitte and Databricks bring to clients, represents a necessary shift for the future. This includes embedding autonomous, stateful intelligence across the data lifecycle to continuously evaluate fitness-for-use, adapt to change, and intervene before quality degradation impacts business outcomes.

For CDOs and CIOs, adopting an agentic data quality framework is also becoming a prerequisite for AI readiness and for leveraging data as a true strategic asset. Multi-agent systems orchestrate profiling, validation, remediation and monitoring as an integrated control plane that learns from outcomes, enforces adaptive governance and maintains explainability at scale.

Organizations that invest in this capability can position themselves to move beyond managing data defects, and toward engineering trust as a core enterprise competency. This can enable faster insights, safer AI adoption and a sustained competitive advantage. In this future state, data quality is no longer a downstream function—rather, it is an intelligent, always-on foundation for enterprise decision making.

Authors

Bill Murphy
AI & Data
Managing Director
Deloitte Consulting LLP
bimurphy@deloitte.com

Maria Thomas
AI & Data
Associate Vice President
Deloitte Consulting LLP
marithomas@deloitte.com

Lynell Amanna
AI & Data
Lead Analytics and Insights Engineer
Deloitte Consulting LLP
lamanna@deloitte.com

Chinmay Khaitan
Finance Transformation
Senior Consultant
Deloitte Consulting LLP
ckhaitan@deloitte.com

Youngjoon Lim
Engineering as a Service
Software Engineer
Deloitte Consulting LLP
younglim@deloitte.com

As used in this document, "Deloitte" means Deloitte Consul ting LLP, a subsidiary of Deloitte LLP. Please see www.deloitte.com/us/about for a detailed description of our legal structure. Certain services may not be available to attest clients under the rules and regulations of public accounting.

This publication contains general information only and Deloitte is not, by means of this publication, rendering accounting, business, financial, investment, legal, tax, or other professional advice or services. This publication is not a substitute for such professional advice or services, nor should it be used as a basis for any decision or action that may affect your business. Before making any decision or taking any action that may affect your business, you should consult a qualified professional advisor. Deloitte shall not be responsible for any loss sustained by any person who relies on this publication.

Copyright © 2026 Deloitte Development LLC. All rights reserved.