SingleStore™

AI Readiness: The Essential Data Foundation for Enterprise Impact

How to prepare your data infrastructure for autonomous agents and real-time intelligence

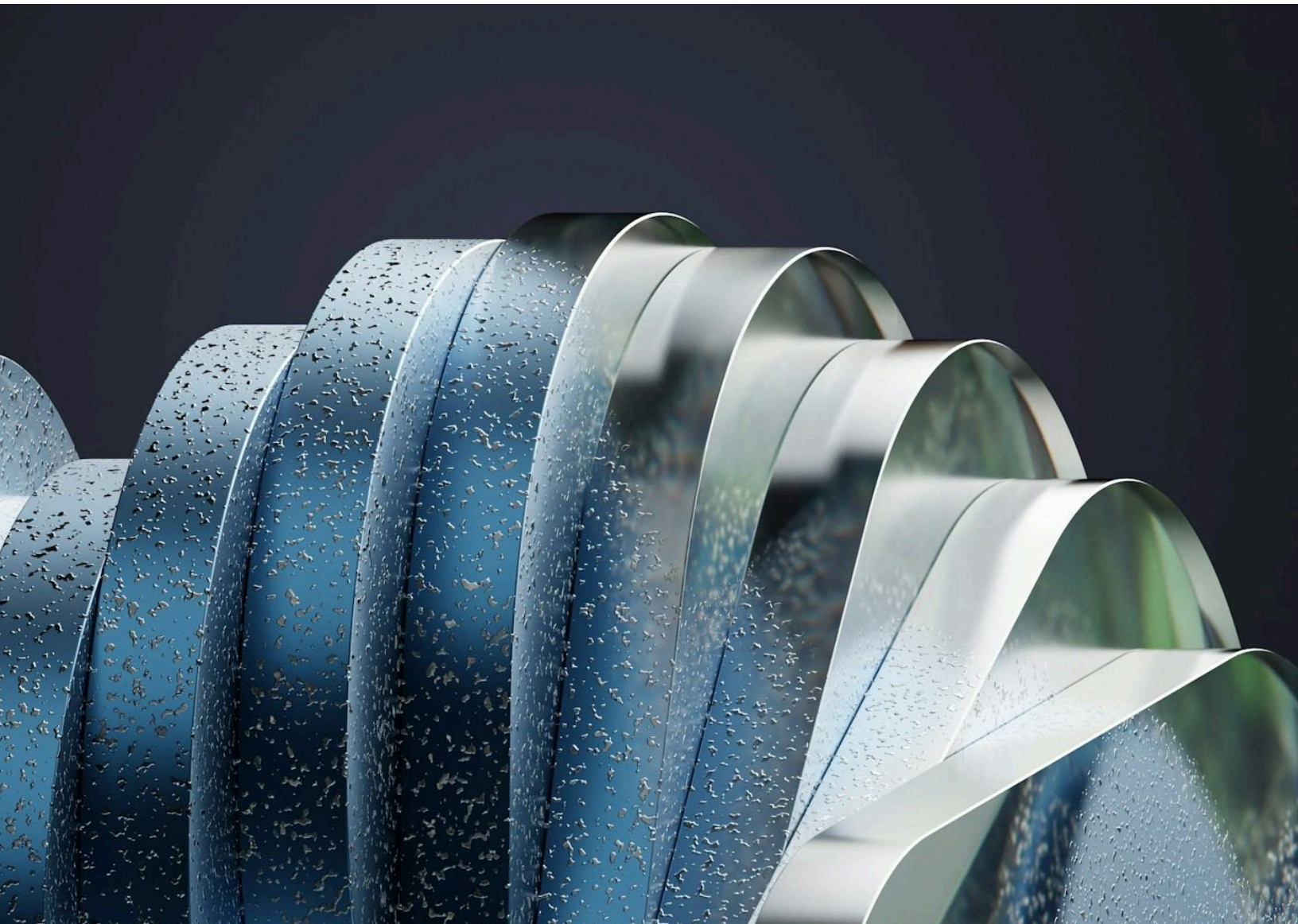


Table of contents

Foreword

In architecture and in life, the most important step in building something new is creating a solid foundation.

It's the same with enterprise AI. Without the proper foundation, your big ideas will likely be slowed by brittle infrastructure, siloed systems, and data that's simply not ready for real-time intelligence. Without that foundation, your results will be inefficient at best, and misleading or unusable at worst.

We created this eBook to ensure that doesn't happen to you.

"AI Readiness: The Essential Data Foundation for Enterprise Impact" is more than a technical guide — it's a roadmap for organizations looking to move beyond experimentation and into scalable, sustainable AI. It outlines the critical data architecture decisions required to support AI agents, accelerate time-to-value and drive impact across your business.

Inside, you'll find actionable insights and quick tips to help your teams reduce AI hallucinations, unify access to data and build a foundation that's fast, secure and future-ready. Because the formula for winning with enterprise AI is actually quite simple: it's all in the infrastructure.

Let's build that foundation together.

Raj Verma
Chief Executive Officer,
SingleStore

Summary

According to Boston Consulting Group's 2024 research, 74% of companies struggle to achieve and scale value from their AI initiatives, with only 26% having developed the necessary capabilities to move beyond proofs of concept and generate tangible value [1]. Among these companies, just 4% have developed leading AI capabilities across functions and consistently generate significant value, while an additional 22% have implemented an AI strategy, built advanced capabilities and are beginning to realize substantial gains. This gap between experimentation and enterprise-wide implementation represents a critical challenge for organizations — especially as McKinsey's 2024 research shows AI adoption has jumped to 72% across organizations, with 65% now regularly using generative AI [2].

Today's enterprises face a critical gap between AI ambition and AI reality: their data infrastructure. As organizations race to implement AI solutions, many discover their existing data architecture — built for a different era — simply cannot support the demands of modern AI applications.

The challenge is clear: Organizations struggle with fragmented data silos, performance bottlenecks and governance blind spots that prevent AI from delivering its promised value. Real-time AI applications require a performant data layer and infrastructure that delivers immediate insights and supports rapid decision-making.

1. https://www.bcg.com/press/24october2024-ai-adoption-in-2024-74-of-companies-struggle-to-achieve-and-scale-value
2. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-2024

This guide addresses the fundamental requirements for successful, scalable and responsible AI implementation. We explore how the shift toward autonomous agents, Retrieval-Augmented Generation (RAG) and real-time intelligence demands a new approach to data architecture — one that is engineered for the speed, scale simplicity and security essential for enterprise AI.

Quick wins for improving your AI data foundation:

- Reduced hallucinations through grounded, single-shot data retrieval
- Faster time-to-market for AI applications with unified data access
- Lower total cost of ownership through platform consolidation
- Improved governance and compliance with built-in security controls

By focusing on the often-overlooked data foundation, this guide helps you move beyond the AI hype cycle to create sustainable, business-impacting AI systems powered by a performant data layer for your AI agents and application stack.

The path to enterprise AI success starts with readiness. Let's go.

Chapter 1
The AI Inflection Point

The shift from buzz to business imperative

The artificial intelligence landscape has undergone a seismic shift. What began as experimental projects has rapidly evolved into a strategic, necessary business need. This transformation isn't simply about technology — it represents a fundamental change in how organizations operate, compete and deliver value.

The numbers tell a compelling story: according to recent industry research, AI adoption has accelerated dramatically, with over 70% of companies now investing in AI capabilities, up from just 40% three years ago. More significantly, the nature of these investments has changed from isolated experiments to enterprise-wide AI strategies.

This inflection point we're witnessing isn't just about increased adoption; it's about the maturation of AI from a speculative technology to a core business capability that drives measurable value. Organizations that recognize and respond to this shift will be positioned to thrive in an increasingly AI-driven business landscape.

What's changed: From experiments to enterprise-wide use

AI implementation has been remarkable in its scope and speed. Just a few years ago, most organizations were focused on narrow, isolated AI projects —smaller efforts like a chatbot for customer service, or predictive analytics for a specific business function.

Today, we're seeing a dramatic expansion in both the breadth and sophistication of AI applications:

- **Co-pilots and AI assistants:** These AI systems work alongside human employees, augmenting their capabilities rather than replacing them. From coding assistants helping developers quickly write better code to sales co-pilots providing real-time guidance during customer interactions, these tools transform productivity across the enterprise.
- **Retrieval-Augmented Generation (RAG):** This approach combines the generative capabilities of Large Language Models (LLMs) with the ability to retrieve and incorporate specific information from an organization's knowledge base. RAG systems support more accurate, contextual and trustworthy AI responses by grounding them in verified organizational data.
- **Autonomous agents:** Perhaps the most significant shift, autonomous agents can perform complex sequences of actions with minimal human intervention. These systems monitor supply chains, optimize resource allocation, detect anomalies and make decisions within defined parameters — all while continuously learning and improving.

This transition from experimental to enterprise-wide AI deployment brings unprecedented opportunities, but also introduces new challenges. As AI systems become more integrated with core business processes and handle increasingly sensitive tasks, the stakes grow higher. Organizations must ensure these systems are not just powerful but also reliable, transparent and free from hallucinations.

The hidden blocker: Data infrastructure

As enterprises race to implement AI solutions, many are encountering an unexpected obstacle: their existing data infrastructure. This hidden blocker prevents organizations from realizing the full potential of their AI investments, and creates a widening gap between AI ambition and reality.

The fundamental challenge is most traditional data architectures were not designed with modern AI requirements in mind. They were built for structured data, batch processing and human-paced analytics — not for the real-time, multi-modal, high-volume data needs AI systems demand.

Key limitations of legacy data infrastructure include:

- **Performance bottlenecks:** AI applications, particularly those involving real-time decision-making or customer interactions, require sub-second query response times. Many existing data systems struggle to deliver this level of performance, especially at scale.
- **Data silos and fragmentation:** Enterprise data typically resides in numerous disconnected systems, making it difficult to provide AI applications with the comprehensive view they need to generate accurate insights and recommendations.
- **Inability to handle diverse data types:** Modern AI requires seamless integration of structured, semi-structured and unstructured data, including text, images and vectors. Legacy systems often are one-trick ponies: excelling at one data type, but struggling with others.
- **Governance and security gaps:** As AI systems access and process sensitive information, robust governance and security mechanisms become essential. Many existing data architectures lack the fine-grained controls needed for responsible AI deployment.
- **Scalability challenges:** AI workloads can be unpredictable and resource-intensive. Data infrastructure must be able to scale dynamically to accommodate these demands — without compromising performance or requiring constant manual intervention.

These limitations create a critical disconnect between AI aspirations and capabilities. Organizations may successfully develop promising AI models in controlled environments, only to find they cannot effectively deploy them in production due to shortcomings in their underlying data infrastructure.

Addressing this hidden blocker requires a fundamental reassessment of data architecture. Building an AI-ready data foundation isn't just about incremental improvements to existing systems — it demands a new approach that prioritizes speed, integration, flexibility and governance.

Chapter 2
What does it mean to be AI-ready?

Defining AI readiness in the enterprise context

AI readiness represents an organization's capability to effectively develop, deploy and maintain artificial intelligence solutions. In an enterprise, this extends far beyond having access to the latest AI models or hiring data scientists. True AI readiness encompasses a holistic set of capabilities across technology, processes and people.

At its core, enterprise AI readiness means having:

- **A robust data foundation:** The ability to access, process and analyze relevant data at the speed and scale required for AI applications — a performant data layer built to power your AI agents and application stack.
- **Technical infrastructure:** Computing resources, development environments and deployment pipelines that support the full AI lifecycle.
- **Governance frameworks:** Policies and procedures that ensure AI systems operate ethically, securely, and in compliance with relevant regulations.
- **Organizational alignment:** Clear business objectives for AI initiatives and executive sponsorship to drive adoption.
- **Talent and skills:** The necessary expertise to develop, implement and maintain AI solutions, either in-house or through partnerships.
- **Change management capabilities:** The ability to effectively integrate AI into existing workflows and help employees adapt to new ways of working.

AI readiness isn't binary — it exists on a spectrum, with organizations at different stages of maturity across these dimensions. What's crucial is that enterprises assess their current state honestly, and develop a clear roadmap for addressing gaps.

The most AI-ready organizations view artificial intelligence not as a standalone technology initiative but as a transformative capability that touches every aspect of their business. They recognize becoming AI-ready is an ongoing journey rather than a destination, requiring continuous evolution as technologies advance, business needs change and data volume grows.

The green vs. brown data estate paradigm

When assessing AI readiness, it's helpful to consider the concept of "green" versus "brown" data estates — a paradigm that illuminates the fundamental differences between AI-optimized and legacy data architectures.

Brown data estates are characterized by legacy systems built for traditional analytics and reporting, batch-oriented processing with significant latency and siloed data repositories requiring complex integration. They typically have rigid schemas that struggle with

unstructured data, limited scalability and point-to-point integrations that create brittle dependencies. In these environments, governance is often an afterthought rather than a built-in feature.

Brown data estates were designed for a different era of data management and struggle to meet the demands of modern AI applications. Organizations with predominantly brown data estates often find themselves spending more time on data preparation and infrastructure maintenance than on actual AI innovation.

Green data estates, by contrast, feature modern architectures engineered for speed, scale, simplicity and security. They provide real-time data processing capabilities, unified data platforms that eliminate silos and flexible data models that accommodate diverse data types. These environments offer elastic scalability, API-first approaches and built-in governance and security features.

Green data estates provide the foundation for responsive and responsible AI deployment. They enable organizations to move quickly from concept to production, iterate based on feedback and scale successful initiatives across the enterprise.

Most organizations have a mix of brown and green elements in their data infrastructure. The journey to AI readiness often involves thoughtfully (and strategically) transforming brown components to green where they create the most significant bottlenecks for AI initiatives. This doesn't necessarily mean complete overhauls and replacements of existing systems — in many cases, it's about augmenting legacy infrastructure with modern capabilities that bridge the gap.

Common signs you — and your data infrastructure — are not ready

Organizations often discover their lack of AI readiness only after encountering significant problems during implementation. Recognizing these warning signs early can help you address underlying issues before they derail your AI initiatives. Here are the most common indicators that your data foundation isn't yet AI-ready:

- **AI hallucinations and inaccuracies:** When AI systems generate plausible but incorrect information, it often points to problems with the underlying data. These hallucinations may stem from outdated information, training data that doesn't reflect your specific business context, inability to ground AI responses in authoritative sources, lack of effective feedback mechanisms or absence of single-shot retrieval capabilities.

- **Performance bottlenecks:** AI applications require responsive data access, and slow performance renders even the most sophisticated models ineffective. Warning signs include query latency exceeding acceptable thresholds, batch processing windows that can't accommodate real-time decision making, inability to handle peak loads, performance degradation as data volumes grow and systems that can't deliver the responsiveness required for real-time AI.
- **Compliance and governance blind spots:** As AI systems access and process sensitive information, governance weaknesses become increasingly problematic. Watch for inability to track data lineage, lack of transparency into AI decision-making, insufficient access controls and challenges demonstrating compliance with regulations.
- **Integration challenges:** AI rarely exists in isolation — it needs to connect with existing systems and processes. Problems include difficulty incorporating AI insights into workflows, manual processes required to move data between systems, inconsistent data definitions and disjointed architectures requiring multiple specialized databases.
- **Scalability limitations:** Many AI initiatives start small but quickly grow in scope and complexity. Warning signs include proof-of-concepts that can't transition to production, solutions that work for one department but can't be extended enterprise-wide, disproportionate infrastructure costs and unsustainable maintenance overhead.

These signs don't just indicate technical problems — they have direct business consequences. AI hallucinations damage customer trust and brand reputation. Performance issues frustrate users and reduce adoption. Compliance blind spots lead to regulatory penalties and legal exposure. And integration challenges prevent AI from delivering its promised value.

Recognizing these warning signs is the first step toward addressing the underlying data foundation issues that limit your AI readiness.

Chapter 3
Laying the right foundation

Why real-time, unified data is the key enabler

At the heart of AI readiness lies a critical capability: real-time, unified data access. This isn't just a technical nice-to-have — it's the essential foundation that enables AI to deliver transformative business value.

Today's most impactful AI applications operate in real time, making decisions and providing insights as events unfold. The difference between real-time and batch processing isn't merely one of speed — it fundamentally changes how your AI acts and responds.

Equally important is the unification of data across sources, types and time horizons — a single platform for all data types. This unification provides AI systems with the comprehensive context they need to generate accurate, relevant and trustworthy outputs.

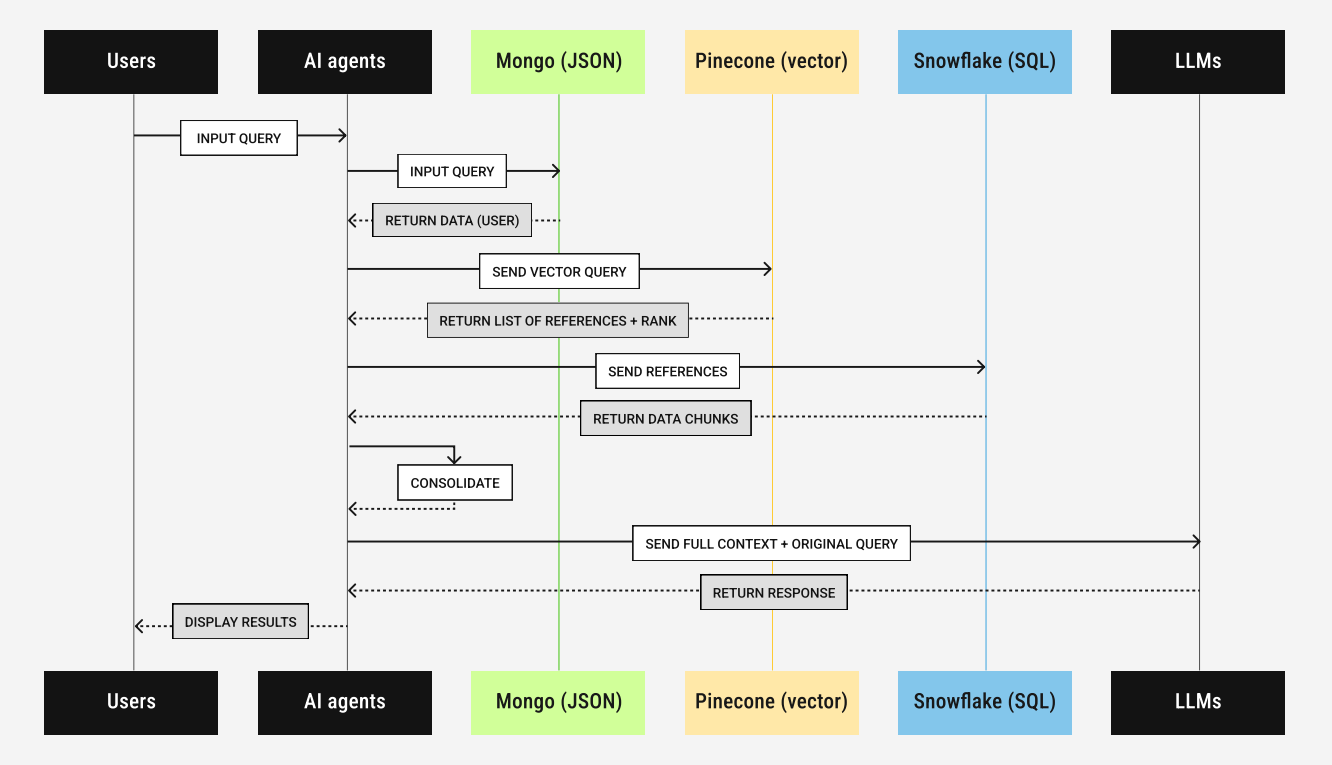
The power of single-shot data retrieval

Single-shot data retrieval represents a fundamental shift in how AI applications interact with data. Rather than the traditional approach of extracting data from operational systems, transforming it through complex pipelines and loading it into separate analytical environments, single-shot retrieval enables direct access to the most current data in a single operation.

This approach delivers several key benefits:

- Immediate access to fresh, operational data without ETL delays
- Reduced complexity by eliminating intermediate data movement
- Lower latency through direct querying of source systems
- Simplified architecture with fewer components and integration points
- Improved data consistency by avoiding synchronization challenges

For AI applications, single-shot retrieval means working with the most current information available, reducing the risk of making recommendations based on outdated data. It enables truly interactive AI experiences where each interaction can consider the latest context — including actions the user may have taken moments earlier.



How a modern data platform unlocks AI capabilities

A modern, unified data platform serves as the foundation that enables the most advanced AI capabilities. By providing real-time access to comprehensive, contextual data, these platforms unlock the full potential of technologies like RAG, autonomous agents and context-aware AI applications.

Powering RAG

RAG systems enhance LLMs by retrieving relevant information from an organization's knowledge base before generating responses. This approach combines the flexibility of generative AI with the accuracy and specificity of enterprise data. A modern data platform enables effective RAG by providing:

- High-performance vector search capabilities to find semantically relevant information
- Real-time access to the latest data, ensuring responses reflect current reality
- Unified storage for both structured data and document embeddings
- Ability to maintain context across multiple interactions
- Fine-grained access controls to ensure information security

With these capabilities, RAG systems deliver responses that are not only conversational and helpful but also accurate, up-to-date and grounded in your organization's specific knowledge. This dramatically reduces hallucinations and increases trust in AI outputs.

Enabling autonomous agents

Autonomous agents represent the next frontier in AI, capable of performing complex sequences of actions to accomplish goals with minimal human intervention. These agents require a performant data foundation that provides:

- Comprehensive visibility across business domains to inform decision making
- Real-time awareness of changing conditions to adapt plans dynamically
- Historical context to learn from past outcomes and improve over time
- Ability to reason across different data types and sources
- Secure, controlled access to operational systems for taking actions

A modern data platform enables agents to operate effectively by providing this comprehensive, real-time view of the business environment. This allows them to make informed decisions, prioritize actions based on current conditions and continuously learn from outcomes.

Delivering context-rich AI experiences

Perhaps the most significant advantage of a modern data platform is its ability to provide rich context for AI applications. Context-aware AI understands not just the immediate query or interaction but the broader situation including:

- The user and their role
- What actions have they taken recently
- How the interaction relates to broader business processes
- What constraints or policies influence the response
- Relevant historical patterns

By maintaining this context across interactions, AI systems provide more personalized, relevant and helpful responses — while also anticipating needs, making proactive suggestions and adapting their behavior based on changing circumstances.

Chapter 4
Defining AI use cases

Exploring business possibilities through AI

As organizations move beyond initial AI experimentation, identifying applications with genuine business impact becomes a critical challenge. This section explores AI use cases across key industries to inspire your thinking, and illustrate how a robust data foundation enables enterprise transformation. These examples demonstrate how organizations might leverage AI to address specific business challenges, highlighting the data requirements that make these solutions possible — and the outcomes they deliver.

Financial Services

Real-time fraud detection and prevention

Business challenge

Financial institutions face increasingly sophisticated fraud attempts that traditional rule-based systems struggle to detect. Fraudsters continuously advance their techniques, requiring even more advanced detection capabilities. Meanwhile, false positives create significant customer friction and operational costs.

AI solution

AI-powered fraud detection systems analyze transactions in real time, considering hundreds of factors simultaneously — including transaction history, location data, device information and behavioral patterns — to identify suspicious activities with greater accuracy than rule-based approaches.

Data requirements

Real-time processing of transaction data as it occurs

Unified view of customer accounts, historical transactions and authentication events

Vector operations to identify similar patterns across fraud attempts

Ability to combine structured transaction data with unstructured information

Sub-millisecond query performance to prevent fraudulent transactions before completion

Expected outcomes

Reduction in fraud losses through earlier detection

Decreased false positives, improving customer experience

Ability to adapt to new fraud patterns without manual rule updates

Lower operational costs for fraud investigation teams

Retail

Supply chain optimization

Business challenge

Retailers face unprecedented complexity in managing global supply chains with multiple suppliers, distribution centers, transportation methods and fulfillment channels. Traditional supply chain management approaches struggle with demand volatility, inventory optimization across locations and rapid response to disruptions.

AI solution

AI-powered supply chain optimization systems analyze data across the entire supply network to forecast demand with greater accuracy, optimize inventory placement, predict potential disruptions and recommend mitigation strategies. These systems might continuously adjust inventory levels, routing and fulfillment strategies based on real-time conditions.

Data requirements

Unified view of inventory across all locations and channels

Real-time sales data and demand signals

Supplier performance and lead time information

External data sources (weather, events, economic indicators)

Transportation network status and capacity

Historical patterns with seasonal decomposition

Expected outcomes

Reduced inventory carrying costs through more precise stocking

Improved product availability and fewer stockouts

Enhanced resilience to supply chain disruptions

Lower transportation costs through optimized routing

Decreased waste from perishable or seasonal items

Improved sustainability through more efficient resource utilization

Healthcare

Clinical decision support

Business challenge

Healthcare providers face increasing complexity in clinical decision-making due to the growing volume of medical research, treatment options and patient data. Clinicians struggle to stay current with the latest evidence while considering each patient's unique circumstances.

AI solution:

AI-powered clinical decision support systems analyze patient data — including medical history, lab results, medications and genetic information — alongside the latest medical research to provide contextually relevant guidance to clinicians. These systems identify potential diagnoses, recommend appropriate tests, suggest treatment options and flag potential medication interactions.

Data requirements

Unified patient records integrating clinical, laboratory and medication data

Real-time access to the latest patient information

Vector search capabilities to find relevant medical literature and similar cases

Ability to process both structured clinical data and unstructured notes

Strict security and compliance controls for protected health information

Expected outcomes

Improved clinical outcomes through evidence-based care

Reduced medical errors and adverse events

More efficient clinical workflows

Decreased variation in care delivery

Cybersecurity

Advanced threat detection and response

Business challenge

Organizations face increasingly sophisticated cyber threats that traditional security tools struggle to detect. Attack techniques evolve rapidly, security teams are overwhelmed with alerts and the time to detect and respond to breaches remains too long — increasing potential damage.

AI solution

AI-powered security systems continuously monitor network traffic, user behavior, application activities and system logs to identify anomalous patterns indicative of threats. These systems correlate events across multiple sources, prioritize alerts based on risk, identify attack patterns similar to previous incidents and recommend or automate response actions.

Data requirements

Real-time processing of security telemetry from diverse sources

Historical baseline data of normal behavior patterns

Threat intelligence feeds and known attack signatures

User and entity behavior analytics

Vector similarity search to identify related incidents

Ability to process massive volumes of log and event data

Expected outcomes

Reduced time to detect security incidents

Decreased false positive alerts requiring investigation

Improved identification of sophisticated attacks that evade rule-based detection

Faster incident response through automated triage and recommended actions

Enhanced threat hunting capabilities

Better utilization of limited security analyst resources

Additional industry applications

Beyond the previously detailed use cases, AI applications powered by a performant data foundation have the potential to transform other industries, including:

Telecommunications

- Network optimization and predictive maintenance
- Customer experience personalization and churn prediction
- Spectrum efficiency and dynamic resource allocation

Pharmaceuticals

- Accelerated drug discovery through molecular simulation
- Clinical trial optimization and patient matching
- Adverse event prediction and pharmacovigilance

Energy

- Smart grid optimization and demand forecasting
- Predictive maintenance for generation and distribution assets
- Renewable energy integration and storage optimization

Manufacturing

- Predictive maintenance and asset optimization
- Quality control through computer vision
- Production planning and scheduling optimization

For more detailed use cases across these and other industries, **see Appendix B.**

Common data foundation requirements

No matter the industry application, several common requirements emerge for the data foundation and layer needed to support effective AI implementation:

Unified data access	Eliminating silos to provide AI applications with comprehensive context
Real-time processing	Enabling immediate analysis and response based on the latest information
Support for diverse data types	Handling structured, semi-structured and unstructured data in a single platform
Vector operations	Providing high-performance similarity search for modern AI applications
Enterprise-grade security and governance	Ensuring appropriate data protection, especially for sensitive information
Scalable performance	Maintaining responsiveness as data volumes and query complexity grow

These requirements highlight why a unified data platform like SingleStore deliver the performance needed for enterprise AI across industries. By addressing these fundamental data capabilities, organizations are better positioned to move from AI experimentation to implementation.

Chapter 5
Breaking through AI barriers

From obstacles to opportunities

The journey to AI readiness is rarely straightforward. Organizations face numerous challenges when implementing AI at scale, from technical limitations to organizational resistance. However, these barriers don't need to be roadblocks — with the right approach, they become opportunities for transformation.

This chapter provides practical strategies to help you navigate common AI barriers. Instead of theoretical discussions, we focus on actionable approaches that organizations of all sizes and technical maturity can apply — including those with complex legacy environments.

Practical implementation paths for complex environments

The legacy integration blueprint

Large enterprises typically operate dozens — or even hundreds — of databases and data stores accumulated over decades. Completely replacing these systems is rarely feasible due to cost, risk and operational disruption. Instead, consider this practical implementation path:

1 Implement automated data quality monitoring

Rather than ripping and replacing existing systems, implement a high-performance database layer that sits between your AI applications and existing data sources:

- Configure real-time data pipelines that synchronize critical data from legacy systems
- Prioritize data sources based on their importance for AI use cases
- Start with a subset of data needed for initial AI applications, then expand incrementally
- Use Change Data Capture (CDC) to maintain real-time synchronization with source systems

2 Create a unified semantic layer

Develop a semantic layer that provides a consistent view across disparate data sources:

- Define standardized data models that harmonize inconsistent definitions
- Create virtual views that combine data from multiple sources
- Implement metadata management to track data lineage and transformations
- Establish governance controls that respect source system restrictions

3 Implement progressive data migration

As AI applications prove their value, gradually migrate data from legacy systems:

- Begin with read-only synchronization to minimize risk
- Transition applications one-by-one to read from the unified layer
- Eventually shift write operations for modernized applications
- Decommission legacy systems as their functionality is fully migrated

The hybrid architecture blueprint

For organizations that need to maintain certain legacy systems indefinitely, a hybrid architecture provides a pragmatic path to AI readiness:

1 Implement automated data quality monitoring

Categorize your data assets based on AI relevance and system constraints:

- **Fast-moving data:** Operational data that requires real-time access (transactions, customer interactions, sensor data)
- **Reference data:** Relatively static information used for context and enrichment
- **Historical data:** Long-term archives needed for training and analysis
- **Specialized data:** Information that must remain in purpose-built systems

2 Implement a tiered data architecture

Create a multi-tier architecture that optimizes each data category:

- Move fast-moving data to a high-performance database for real-time AI access
- Synchronize reference data for low-latency lookups
- Implement federation capabilities for historical and specialized data
- Create a unified query layer that spans all tiers

3 Optimize data movement

Minimize unnecessary data duplication through intelligent data movement:

- Use materialized views for frequently accessed subsets of legacy data
- Implement intelligent caching for reference data
- Create data pipelines that prioritize business-critical information
- Leverage compression and efficient storage formats

Overcoming data quality and governance challenges

AI systems are only as good as the data that powers them — because let's be clear: AI without data isn't really AI. Poor data quality leads to inaccurate insights, while inadequate governance creates compliance risks. Here are practical strategies to address these challenges:

Data quality enhancement blueprint

1 Implement automated data quality monitoring

Deploy continuous monitoring of data quality metrics:

- Define key quality dimensions (completeness, accuracy, consistency, timeliness)
- Establish baseline metrics for critical data elements
- Implement automated checks at ingestion points
- Create alerts for quality degradation

2 Develop data quality improvement pipelines

Create automated processes to enhance data quality:

- Implement standardization rules for common data elements
- Deploy entity resolution to identify and merge duplicates
- Use machine learning to detect anomalies and outliers
- Create feedback loops that capture and incorporate corrections

3 Prioritize quality efforts based on AI impact

Focus quality initiatives on data elements most critical for AI success:

- Identify the data that drives key AI features and predictions
- Assess quality gaps in high-impact data elements
- Implement targeted enhancement for priority areas
- Monitor quality impact on AI performance

Governance-first AI blueprint

1 Establish AI-specific governance policies

Develop governance frameworks tailored to AI requirements:

- Define appropriate use policies for different data categories
- Establish clear ownership and stewardship responsibilities
- Create transparency requirements for AI decision-making
- Implement audit mechanisms for AI data usage

2 Implement technical governance controls

Deploy technical safeguards that enforce governance policies:

- Role-based access controls aligned with data sensitivity
- Automated policy enforcement at query time
- Comprehensive audit logging of all AI data access
- Data lineage tracking throughout AI pipelines

Scaling AI from pilot to production

Many organizations successfully develop AI pilots but struggle to scale them to production. This "last mile" challenge often derails promising initiatives. Here's a practical blueprint for successful scaling:

1 Design for production from the start

Build pilot projects with production scaling in mind:

- Use representative data volumes and varieties
- Implement proper security and governance controls from day one
- Design for operational monitoring and management
- Consider performance requirements at scale

2 Create a standardized deployment pipeline

Establish consistent processes for moving AI from development to production:

- Implement CI/CD pipelines for model deployment
- Create standardized testing protocols for accuracy and performance
- Develop automated monitoring for model drift and performance
- Establish clear rollback procedures

3 Build operational support capabilities

Develop the operational foundation needed for production AI:

- Implement comprehensive monitoring and alerting
- Create runbooks for common operational scenarios
- Establish SLAs and performance baselines
- Develop incident response procedures

The data foundation imperative

As we've explored the various implementation paths and strategies for overcoming AI barriers, a common theme emerges: the critical importance of a performant, unified data foundation. Organizations that successfully scale AI beyond pilot projects invariably build their initiatives on a data architecture specifically designed for AI workloads.

The ideal data foundation for enterprise AI should provide:

- **Real-time** access to operational data
- **Unified storage** for structured, semi-structured and vector data
- **High-performance query** capabilities for complex, multi-modal operations
- **Scalability** to handle growing data volumes and user demands
- **Comprehensive governance and security** controls
- **Simplified architecture** that reduces integration complexity

In the next chapter, we'll explore how SingleStore's unified data platform addresses these requirements, providing the essential, performant foundation for enterprise AI readiness.

Chapter 6
The SingleStore Advantage

Addressing AI barriers with a unified data platform

As we saw in the previous chapter, organizations face significant challenges when implementing AI at scale. SingleStore's unified data platform directly addresses these barriers, providing the essential data foundation for enterprise AI readiness. By combining transactional (OLTP), analytical (OLAP) and vector capabilities in a single system, SingleStore eliminates the complexity, performance limitations and integration challenges that often derail AI initiatives.

Why speed, scale and simplicity make you AI ready

As organizations pursue AI readiness, three fundamental capabilities emerge as critical differentiators between success and struggle: speed, scale and simplicity. These capabilities form the foundation upon which effective enterprise AI is built, and they represent the areas where SingleStore delivers exceptional value.

Speed: The currency of modern AI

In the world of AI, speed isn't just a nice-to-have — it's essential for delivering experiences that meet user expectations and drive business value:

- **Real-time interaction:** Users expect AI systems to respond instantly, whether they're customer-facing applications or internal tools. Delays of even a few seconds significantly reduce engagement and adoption.
- **Contextual awareness:** AI systems need to incorporate the latest information to provide relevant responses. Processing delays mean working with outdated context, leading to less accurate or useful outputs.
- **Operational agility:** AI-driven automation requires real-time decision-making based on current conditions. Latency in data processing directly impacts the effectiveness of these automated processes.
- **Competitive advantage:** In markets where multiple players are deploying AI, speed becomes a key differentiator. The organization that can deliver faster, more responsive AI experiences gains significant advantage.

SingleStore's architecture is designed specifically for the performance requirements of modern enterprise AI applications. Its in-memory processing, columnar storage and distributed query execution deliver sub-millisecond response times — even for the most complex queries across massive datasets. This performance foundation enables AI applications that respond instantly, incorporate the latest information and deliver exceptional user experiences.

Scale: The backbone of enterprise AI

As AI moves from experimentation to enterprise-wide deployment, scalability becomes increasingly critical:

- **Data volume growth:** AI applications generate and consume enormous amounts of data, from training datasets and operational logs to user interactions.
- **Query complexity:** Advanced AI applications often require complex queries that combine different types of operations and data sources.

- **Concurrent users:** Enterprise AI systems must support thousands — or millions — of simultaneous users without performance degradation.
- **Unpredictable workloads:** AI usage patterns can be highly variable, requiring infrastructure that can scale dynamically to meet changing demands.

SingleStore's distributed architecture provides the scalability that enterprise AI requires. Its automatic sharding, parallel query execution and elastic resource management enable organizations to start with modest implementations and scale seamlessly as adoption increases, maintaining consistent performance throughout.

Simplicity: The enabler of AI innovation

Perhaps the most underappreciated aspect of AI readiness is simplicity — the ability to reduce complexity in data architecture:

- **Unified data access:** AI applications need comprehensive context from across the organization, which traditional fragmented architectures make difficult to provide.
- **Reduced integration complexity:** Every integration point in a data architecture introduces potential failure points, latency and maintenance overhead.
- **Operational efficiency:** Complex, multi-system architectures require specialized skills and significant resources to maintain.
- **Development agility:** Simplified architectures enable faster development cycles and more rapid innovation.

SingleStore's unified approach dramatically simplifies data architecture by eliminating the need for separate specialized systems. By providing a single platform for all data types — transactional, analytical and vector — it reduces complexity, lowers costs and accelerates AI development and deployment.

SingleStore's implementation blueprints

SingleStore directly addresses the implementation challenges outlined in Chapter 5 with specific capabilities designed for enterprise AI readiness:

Legacy integration with SingleStore

For organizations with complex legacy environments, SingleStore provides:

1 Performance layer capabilities

- Real-time data pipelines with built-in CDC support
- High-performance caching of frequently accessed data
- Materialized views that automatically stay in sync with source systems
- Flexible deployment options that work with existing infrastructure

2 Semantic layer enablement

- Universal table format that harmonizes diverse data models
- SQL compatibility for seamless integration with existing tools
- Comprehensive metadata management
- Fine-grained security controls that respect existing permissions

3 Progressive migration support

- Hybrid deployment options that minimize disruption
- Incremental migration capabilities
- Compatibility with existing applications and tools
- Comprehensive monitoring during transition

Hybrid architecture with SingleStore

SingleStore's flexibility makes it ideal for hybrid architectures:

1. **Data estate segmentation:**
   - High-performance storage for fast-moving operational data
   - Efficient synchronization of reference data
   - Federation capabilities for historical archives
   - API-based integration with specialized systems
2. **Tiered architecture implementation:**
   - Unified query layer across all data tiers
   - Intelligent caching of frequently accessed data
   - Automated data placement based on access patterns
   - Seamless integration with existing data lakes and warehouses

Key platform capabilities

SingleStore's unified data platform provides a comprehensive set of capabilities specifically designed to support enterprise AI readiness. These capabilities address the fundamental requirements for successful AI deployment — while eliminating the complexity and performance limitations of traditional data architectures.

Universal Storage for all data types

Modern AI applications require access to diverse data types, from structured records to unstructured content. SingleStore provides Universal Storage that handles:

- **Structured data:** Traditional rows and columns with the transactional integrity required for operational systems
- **Semi-structured data:** JSON, XML and other flexible formats that don't fit rigid schemas
- **Unstructured text:** Documents, emails, chat logs and other text content that powers many AI applications
- **Vector embeddings:** Numerical representations that capture semantic meaning for AI models

Universal Storage eliminates the need for specialized systems for different data types, reducing complexity and enabling AI applications to access relevant information through a single interface. Organizations can store customer records, product catalogs, support tickets and AI embeddings in the same platform, laying comprehensive foundation for context-rich AI.

Unified transactional and analytical processing

Traditional data architectures separate transactional systems (OLTP) from analytical systems (OLAP), creating latency and complexity. SingleStore unifies these workloads with:

- **ACID-compliant transactions:** Support for the transactional integrity required for operational applications
- **Columnar storage:** High-performance analytical processing for complex queries across large datasets
- **Real-time data ingestion:** Ability to immediately incorporate new information without batch delays
- **Concurrent workload management:** Intelligent handling of mixed workloads without performance degradation

This unified approach eliminates the need for complex data pipelines between operational and analytical systems, so AI applications access work the latest data while maintaining historical context. Organizations can build AI systems that incorporate real-time events while analyzing historical patterns, all through a single platform.

Vector search and similarity operations

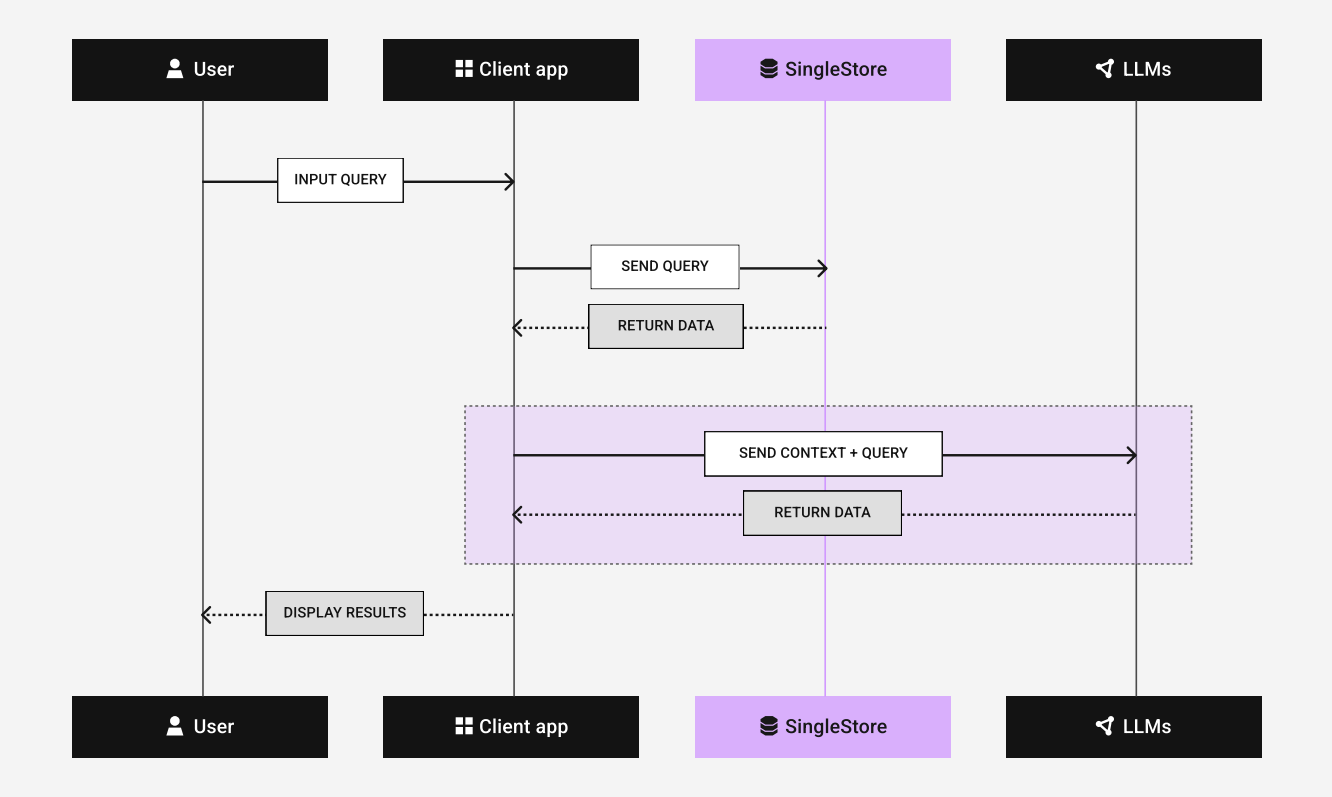
MAI applications increasingly rely on vector representations to understand semantic relationships and similarities. SingleStore provides native vector capabilities including:

- **Vector storage:** Efficient storage for embedding vectors from language models, image processors and other AI systems
- **Similarity search:** High-performance nearest neighbor search to find semantically related content
- **Hybrid queries:** Ability to combine vector similarity with traditional filters and joins
- **Multiple distance metrics:** Support for various similarity measures to match specific AI requirements

These vector capabilities are fully integrated with SingleStore's other features, enabling AI applications to combine semantic understanding with structured data operations. This integration is particularly valuable for RAG applications that need to find relevant information based on semantic similarity while filtering based on metadata attributes.

Single-shot retrieval

SingleStore's architecture supports single-shot retrieval — the ability to access all relevant data in a single query without complex joins across multiple systems. This critical capability:



- **Reduces latency:** Eliminates the need to query multiple systems sequentially
- **Improves accuracy:** Provides AI applications with comprehensive context for better decisions
- **Simplifies development:** Reduces the complexity of data access patterns
- **Enhances reliability:** Eliminates failure points in cross-system integration

Single-shot retrieval is particularly valuable for RAG applications, where it enables AI systems to ground their responses in comprehensive, up-to-date information from an organization's repository, dramatically reducing hallucinations and improving accuracy.

Comprehensive security and governance

Enterprise AI requires robust security and governance to ensure responsible use. SingleStore provides:

- **Role-Based Access Control (RBAC):** Fine-grained permissions that limit data access based on user roles
- **Row-level security:** Ability to control access at the individual record level based on attributes
- **Audit logging:** Comprehensive tracking of all data access and modifications
- **Data lineage:** Visibility into how data flows through the system and is used by applications
- **Encryption:** Protection for data at rest and in transit with flexible key management

These capabilities enable organizations to implement AI with appropriate controls, ensuring sensitive information is protected while still being available for legitimate AI use cases. This governance foundation builds trust with users, regulators and keeps you on the right side of compliance.

SingleStore for AI scaling

SingleStore provides the essential foundation for successfully scaling AI from pilot to production:

Performance at scale

- Linear scalability for increasing data volumes
- Consistent sub-millisecond query response times
- Support for thousands of concurrent users
- Ability to handle complex, multi-modal queries

Simplified operations

- Unified platform eliminates integration challenges
- Automated sharding and data distribution
- Built-in monitoring and management tools
- Cloud-native deployment options

Enterprise-grade reliability

- Smart grid optimization and demand forecasting
- Predictive maintenance for generation and distribution assets
- Renewable energy integration and storage optimization

Measuring ROI with SingleStore

Organizations leveraging SingleStore as the data platform for their AI initiatives typically see value in across important dimensions:

Performance improvements

- 10-100x faster query performance for AI applications
- Real-time data availability versus batch processing delays
- Ability to handle complex, multi-modal queries in milliseconds

Cost efficiencies

- 30-50% reduction in infrastructure costs through platform consolidation
- Decreased development time for AI applications
- Reduced operational overhead for data management

Business outcomes

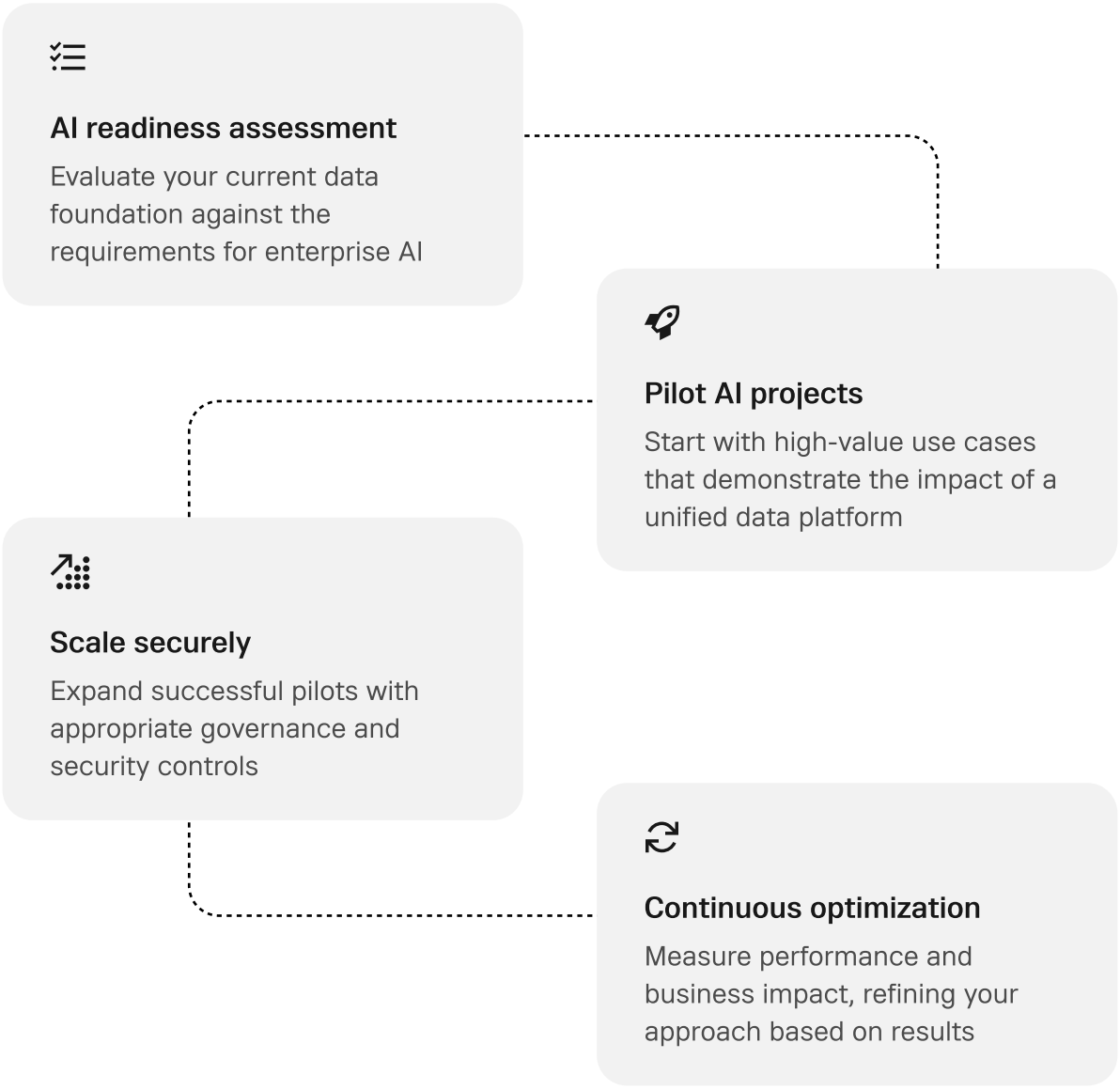
- Faster time-to-market for AI-powered products and services
- Improved customer experiences through real-time personalization
- Enhanced decision-making with comprehensive, up-to-date data
- Competitive differentiation through AI capabilities

By providing a unified data platform engineered for speed, scale and simplicity, SingleStore empowers organizations to overcome the common barriers to AI adoption and unlock the full potential of AI for their business.

Chapter 7
The path to readiness

A practical roadmap

Your journey to AI readiness — from production to deployment — follows a clear progression:



This roadmap provides a practical framework for organizations at any stage of their AI journey. And by focusing on incremental progress and measurable outcomes, you can build momentum while managing risk.

AI readiness assessment

Before embarking on your AI readiness journey, it's essential to understand your current state. A comprehensive assessment should evaluate:

- **Data architecture maturity:** How well does your current data infrastructure support AI requirements? Consider factors like real-time capabilities, support for diverse data types and integration complexity.
- **Data quality and governance:** Are your data quality processes and governance frameworks adequate for AI applications? Evaluate completeness, accuracy, consistency and appropriate controls.
- **Technical capabilities:** Do you have the necessary technical components for AI development and deployment? Assess your vector processing capabilities, model management tools and operational monitoring.

Organizational readiness. Is your organization prepared to develop, deploy and maintain AI applications? Consider skills, processes and cultural factors.

Pilot AI projects

With your assessment complete, the next step is to implement targeted pilot projects that demonstrate value while building capabilities. Effective pilots should:

- **Address clear business needs:** Focus on use cases with well-defined business outcomes and measurable value.
- **Start with manageable scope:** Choose projects that can be implemented with reasonable effort and timeframes.
- **Build on existing strengths:** Leverage areas where your data foundation is already relatively strong.
- **Address key weaknesses:** Include elements that help you develop capabilities in priority improvement areas.
- **Demonstrate the value of unified data:** Showcase how bringing together different data types and sources enhances AI capabilities.

Successful pilots build organizational confidence, establish success metrics for larger-scale projects and and provide valuable learning experiences for your teams.

Scale securely and efficiently

As pilot projects demonstrate success, the focus shifts to scaling AI capabilities across the enterprise. This phase involves:

- **Standardizing development and deployment processes:** Create consistent approaches for AI development, testing and production deployment.
- **Implementing comprehensive governance:** Establish policies, procedures and technical controls that ensure responsible AI use.
- **Building operational capabilities:** Develop the monitoring, management and support processes needed for production AI.
- **Expanding data integration:** Bring additional data sources into your unified platform to support new use cases.
- **Developing reusable components:** Create shared services, libraries and patterns to accelerate new AI initiatives.

This phase transforms AI from isolated projects to enterprise capabilities that can be leveraged across the organization.

Continuous optimization

AI readiness is an ongoing, iterative initiative. Continuous optimization involves:

- **Measuring and improving performance:** Regularly assess technical performance metrics like query response times, throughput and resource utilization.
- **Tracking business impact:** Measure how AI initiatives affect key business metrics and return on investment.
- **Refining data processes:** Continuously improve data quality, integration and governance based on operational experience.
- **Adopting emerging capabilities:** Stay current with evolving AI technologies and incorporate them into your architecture.
- **Expanding use cases:** Identify new opportunities to apply AI capabilities to business challenges.

This ongoing optimization strengthens your AI capabilities — so they can continue to deliver value as technologies evolve and business needs change.

Checklist: Is your enterprise ready for AI?

Use this checklist to assess your organization's progress toward AI readiness:

Data foundation

- We have a unified view of data across operational and analytical systems
- Our data architecture supports real-time processing for AI applications
- We can efficiently store and process diverse data types, including vectors
- Our data quality processes ensure AI systems have reliable inputs
- We have appropriate governance controls for AI data usage

Technical capabilities

- We can deploy and manage AI models in production environments
- Our infrastructure scales to meet AI performance requirements
- We have monitoring capabilities for AI applications and models
- We can integrate AI capabilities with operational systems
- We have appropriate security controls for AI systems

Organizational readiness

- We have clear business objectives for our AI initiatives
- Our teams have the necessary skills for AI development and management
- We have established processes for AI project selection and prioritization
- We have mechanisms to measure and communicate AI business value
- Our culture supports responsible AI innovation and adoption

This checklist provides a quick reference for assessing your current state and identifying areas for further development on your AI readiness journey.

Chapter 8
Future-proofing your enterprise

The evolving AI landscape

The pace of AI innovation continues to accelerate, with each breakthrough creating new possibilities and raising the bar for enterprise capabilities. Organizations that merely implement today's AI technologies without considering tomorrow's evolution risk falling behind competitors who take a more forward-looking approach.

This chapter explores emerging AI trends, their implications for enterprise data architecture and practical steps you can take today to ensure your organization remains competitive as AI capabilities evolve. By building a future-proof foundation, you can position your organization to rapidly adopt new AI innovations while maximizing the value of your current investments.

The rise of agentic AI

Perhaps the most significant trend reshaping the AI landscape is the emergence of agentic AI — autonomous systems that can perform complex sequences of actions with minimal human intervention. Unlike traditional AI that responds to specific prompts or performs predefined tasks, agentic AI plans, reasons and executes multi-step processes to achieve broader objectives.

Key capabilities of agentic AI

Agentic AI systems combine several advanced capabilities:

- **Autonomous reasoning:** The ability to break down complex goals into actionable steps, evaluate alternatives and make decisions based on available information.
- **Tool utilization:** The capacity to select and use appropriate tools or APIs to accomplish specific tasks, from searching databases to calling external services.
- **Memory and context management:** Maintaining awareness of past actions, current state and future goals across extended interactions and processes.
- **Adaptive learning:** Improving performance over time based on outcomes and feedback, refining strategies and approaches.
- **Multi-modal interaction:** Working across different data types and formats, from text and images to structured data and code.

Enterprise applications of agentic AI

Business implications of agentic AI are profound, with applications emerging across industries:

- **Autonomous business processes:** Agents that manage entire business processes like order fulfillment, claims processing or customer onboarding with minimal human intervention.
- **Intelligent assistants:** Advanced systems support knowledge workers by researching information, drafting documents, analyzing data and making recommendations.
- **Supply chain optimization:** Agents continuously monitor and adjust supply chain operations, responding to disruptions and optimizing resource allocation in real time.
- **Customer engagement:** Sophisticated agents manage customer relationships across channels, handling complex inquiries and proactively addressing needs.
- **Research and development:** Systems accelerate innovation by exploring solution spaces, designing experiments and analyzing results.

Data foundation requirements for agentic AI

Agentic AI places unprecedented demands on enterprise data infrastructure:

- **Comprehensive data access:** Agents need visibility across organizational data to make informed decisions and take appropriate actions.
- **Real-time responsiveness:** The ability to access and process data instantly is essential for agents to respond to changing conditions.
- **Transactional capabilities:** Agents need to not only read data, but also write back results and track changes.
- **Fine-grained security:** As agents gain broader access, sophisticated security controls become essential to prevent misuse.
- **Operational monitoring:** Comprehensive visibility into agent actions and outcomes is necessary for governance and improvement.

SingleStore's unified architecture provides the foundation for these requirements, enabling organizations to deploy agentic AI with the performance, context and control needed for enterprise applications.

The evolution of AI reasoning

Beyond agentic capabilities, AI systems are developing increasingly sophisticated reasoning abilities that will transform how they interact with data and support decision-making.

From pattern recognition to causal reasoning

Early AI systems excel at identifying patterns in data but struggle to understand causality — the relationship between cause and effect. The next wave of AI is beginning to bridge this gap:

- **Causal inference:** Identifying not just correlations but actual causal relationships in data, leading to more accurate predictions and interventions.
- **Counterfactual reasoning:** Evaluating "what if" scenarios by modeling how outcomes would change under different conditions.
- **Uncertainty quantification:** Explicitly representing confidence levels and areas of uncertainty, rather than providing point predictions.
- **Explainable decisions:** Providing clear rationales for recommendations that humans can understand and evaluate.

Implications for data architecture

These advanced reasoning capabilities create new requirements for enterprise data:

- **Temporal data management:** Capturing and preserving time-based relationships to support causal analysis.
- **Semantic context:** Maintaining rich metadata that describes the meaning and relationships of data elements.
- **Provenance tracking:** Recording the origin and transformations of data to evaluate reliability and potential biases.
- **Simulation support:** Enabling rapid testing of scenarios and hypotheses through efficient data manipulation.

Organizations building these capabilities into their data architecture today are better positioned to leverage advanced reasoning as it matures.

Multi-modal AI and the convergence of data types

Another significant trend is the convergence of AI capabilities across different data modalities — text, images, audio, video and structured data. This convergence is breaking down traditional boundaries between data types and creating new possibilities for comprehensive analysis and insight.

Emerging multi-modal capabilities

Multi-modal AI systems can:

- **Process diverse inputs:** Accepting and analyzing different data types simultaneously, from documents and images to sensor readings and transaction records.
- **Generate integrated outputs:** Creating responses that combine multiple formats, like explanations with supporting visualizations.
- **Perform cross-modal reasoning:** Drawing connections between information in different formats, like linking textual descriptions to visual patterns.
- **Translate between modalities:** Converting information from one format to another, like generating text descriptions of visual data.

Data architecture implications

Supporting multi-modal AI requires rethinking traditional data architecture:

- **Universal storage:** The ability to efficiently store and retrieve diverse data types in their native formats.
- **Vector representations:** Support for embedding representations that capture semantic meaning across modalities.
- **Unified processing:** Capabilities to query and analyze different data modalities through consistent interfaces.
- **Performance optimization:** Efficient handling of the computational demands of multi-modal processing.

SingleStore's Universal Storage provides a foundation for these requirements, empowering organizations to unify diverse data types in a single platform optimized for AI workloads.

Preparing for the AI-driven enterprise

As AI capabilities continue to evolve, organizations must prepare for a future where AI is embedded throughout their operations. This transition requires not just technological changes, but shifts in organizational structure, processes and culture.

The AI-native organization

Forward-thinking enterprises are moving toward an AI-native model where:

- **AI augments every role:** From frontline workers to executives, everyone has AI tools that enhance their capabilities and productivity.
- **Processes are AI-optimized:** Business processes are redesigned to leverage AI capabilities, often becoming more fluid and adaptive.
- **Data and AI are unified:** The distinction between data infrastructure and AI systems blurs as intelligence is embedded throughout the technology stack.
- **Continuous learning is the norm:** Organizations systematically capture outcomes and feedback to improve AI systems over time.

Building the foundation today

While the fully AI-native organization may still be emerging, there are concrete steps you can take today to prepare:

1. Implement a unified data platform

Create a data foundation that supports evolving AI requirements:

- Consolidate critical data onto a platform like SingleStore that unifies transactional, analytical and vector workloads
- Implement real-time data pipelines to keep information current
- Develop a semantic layer that provides consistent meaning across data sources
- Build governance frameworks that scale with expanding AI usage

2. Develop AI enablement capabilities

Create the organizational capabilities needed to leverage AI effectively:

- Establish AI centers of excellence to provide expertise and best practices
- Implement AI development platforms, accelerating creation and deployment
- Create training programs to build AI literacy across the organization
- Develop ethical guidelines and governance processes for responsible AI use

3. Adopt an experimental mindset

Embrace approaches that allow rapid learning and adaptation:

- Implement agile methodologies for AI development and deployment
- Create safe spaces for experimentation with emerging AI capabilities
- Develop metrics that capture both immediate impact and learning value
- Build feedback loops to systematically improve AI systems over time

4. Plan for AI integration

Prepare for deeper integration of AI into core systems and processes:

- Assess business processes for AI enhancement opportunities
- Develop API strategies that enable AI systems to interact with operational applications
- Create data architectures to support real-time decision-making
- Build monitoring capabilities that ensure AI systems operate as intended

5. Establish an AI governance framework

Create structures to ensure responsible AI development and use:

- Define clear policies for AI development, deployment and use
- Establish review processes for high-risk AI applications
- Implement monitoring for bias, fairness and ethical concerns
- Create transparency mechanisms for AI decision-making
- Develop incident response procedures for AI issues

The competitive advantage of future-proof architecture

Organizations that build a future-proof data architecture gain significant competitive advantages as AI continues to evolve:

- **Faster AI adoption:** The ability to quickly implement new AI capabilities as they emerge without fundamental architectural changes.
- **Greater operational agility:** The capacity to rapidly develop and deploy AI solutions in response to changing market conditions.
- **Enhanced customer experiences:** The capability to deliver personalized, contextual interactions across all customer touchpoints.
- **Improved operational efficiency:** The means to continuously optimize processes through AI-driven insights and automation.
- **Reduced technical debt:** The avoidance of point solutions and architectural compromises that limit future flexibility.

By investing in a data foundation engineered for the future of AI, organizations position themselves to not just keep pace with AI evolution but to lead in their industries through superior AI capabilities.

A vision for the future

The coming decade will see AI transform from a set of specialized tools to a pervasive capability embedded throughout the enterprise.

Organizations that prepare for this transformation will find themselves able to:

- Make decisions with greater speed, precision and confidence
- Personalize products and services to individual customer needs
- Optimize operations continuously in response to changing conditions
- Innovate more rapidly by augmenting human creativity with AI capabilities
- Adapt more effectively to market disruptions and competitive pressures

This future isn't distant or theoretical — it's emerging now through the efforts of organizations that recognize the strategic importance of AI, and are building the foundations to leverage it fully.

The journey to this AI-driven future begins with the decisions you make today about your data architecture. By implementing a unified, real-time data platform like SingleStore, you create the essential foundation for not just current AI applications but for the transformative capabilities that will define competitive advantage in the years ahead.

Start today

Take the next step in your AI readiness journey:

- **Assess your readiness:** Complete our AI readiness assessment to identify your organization's strengths and opportunities
- **See SingleStore in action:** Schedule a personalized demo focused on your specific AI use cases
- **Start a proof of concept:** Test SingleStore's performance within your own data environment and AI applications
- **Join our community:** Connect with other organizations on the AI readiness journey

Enterprise AI success starts with the right data foundation. SingleStore is ready to help you build it.

About SingleStore

SingleStore delivers the performance you need for enterprise AI. Our unified platform handles transactional, analytical and vector workloads in a single system, providing speed, scale and simplicity.

Glossary of terms

Agentic AI
AI systems capable of performing complex sequences of actions autonomously to achieve goals, often using tools and APIs to interact with other systems.

ACID compliance
A set of properties (Atomicity, Consistency, Isolation, Durability) that guarantee database transactions are processed reliably.

Batch processing
Processing data in large groups at scheduled intervals rather than in real-time.

Change Data Capture (CDC) A technique for tracking and capturing changes to data in a database.

Columnar storage
A database storage architecture that organizes data by columns rather than rows, optimized for analytical queries.

Data lineage
The documentation of data's origins, movements, transformations and destinations throughout its lifecycle.

Embeddings
Vector representations of data (text, images, etc.) that capture semantic meaning in a form usable by machine learning models.

ETL (Extract, Transform, Load)
The process of extracting data from source systems, transforming it to fit operational needs and loading it into a destination system.

Hallucination
When AI systems generate content that is factually incorrect or fabricated, despite appearing plausible.

OLAP (Online Analytical Processing)
Systems optimized for complex queries across large datasets for analysis and reporting.

OLTP (Online Transaction Processing)
Systems designed to manage transaction-oriented applications with high throughput and concurrency.

RAG (Retrieval-Augmented Generation)
An approach that enhances large language models by retrieving relevant information from a knowledge base before generating responses.

Row-level security
A database feature that restricts which rows a user can access based on their identity or role.

Semantic layer
An abstraction layer that simplifies complex data models and presents data in business terms, rather than technical ones.

Single-shot retrieval
The ability to access all relevant data in a single query operation without complex joins across multiple systems.

Vector database
A database designed to store and query vector embeddings efficiently, often using similarity search.

Vector search
The process of finding items in a database that are semantically similar to a query, based on vector representations.

Appendix A

Technical details on single-shot ata retrieval and unified OLTP/OLAP/vector stack

Single-shot retrieval: Technical deep dive

Single-shot retrieval represents a paradigm shift in how applications access and process data. This approach eliminates the traditional separation between operational and analytical systems, enabling applications to retrieve all necessary data in a single query operation.

Technical architecture

The technical implementation of single-shot retrieval involves several key components:

- **Unified storage engine:** A storage layer that can efficiently handle both row-oriented (transactional) and column-oriented (analytical) data formats within the same system.
- **Distributed query processing:** A query execution engine that parallelizes complex operations across multiple nodes while maintaining transactional consistency.
- **Hybrid indexing strategies:** Specialized indexing mechanisms that support both point lookups (common in transactional workloads) and range scans (common in analytical workloads).
- **Vector embedding integration:** Native support for vector operations within the same query execution framework as traditional SQL operations.
- **Memory-optimized architecture:** In-memory processing capabilities that minimize I/O operations and enable sub-millisecond response times.

Performance characteristics

Single-shot retrieval delivers several performance advantages:

- **Latency reduction:** By eliminating data movement between systems, single-shot retrieval reduces query latency by 10-100x compared to traditional architectures.
- **Throughput improvement:** The ability to process complex queries in a single operation increases overall system throughput — particularly for workloads that combine transactional and analytical operations.
- **Resource efficiency:** Consolidating workloads on a unified platform reduces the total infrastructure footprint, often by 30-50% compared to maintaining separate specialized systems.
- **Scalability characteristics:** Single-shot architectures can scale linearly with data volume and query complexity, maintaining consistent performance as demands increase.

Implementation considerations

Organizations implementing single-shot retrieval should consider:

- **Data modeling strategies:** How to structure data to optimize for both transactional and analytical access patterns.
- **Query optimization techniques:** Approaches for ensuring efficient execution of complex hybrid queries.
- **Caching mechanisms:** Strategies for caching frequently accessed data to further improve performance.
- **Consistency requirements:** How to balance consistency guarantees with performance needs for different workloads.

Unified OLTP/OLAP/vector stack: Technical architecture

A unified OLTP/OLAP/vector stack integrates capabilities that traditionally required separate specialized systems. This section explores the technical architecture that makes this unification possible.

Storage architecture

The foundation of a unified stack is a storage architecture that efficiently supports diverse workloads:

- **Universal table format:** A flexible table structure that stores both row-oriented and column-oriented data, with dynamic optimization based on access patterns.
- **Multi-tiered storage:** Intelligent data placement across memory, SSD and disk tiers based on access frequency and performance requirements.
- **Compression strategies:** Workload-aware compression techniques that balance storage efficiency with query performance.
- **Vector storage optimizations:** Specialized storage formats for high-dimensional vectors that enable efficient similarity search operations.

Query processing engine

The query processing engine in a unified stack must efficiently handle diverse operation types:

- **Hybrid execution mode:** An execution engine that dynamically switches between row-oriented and column-oriented processing based on query characteristics.
- **Vectorized execution:** SIMD (Single Instruction, Multiple Data) processing capabilities that accelerate analytical and vector operations.
- **Distributed join algorithms:** Specialized algorithms for efficiently joining data across nodes in a distributed environment.
- **Vector search acceleration:** Optimized algorithms for approximate nearest neighbor search, including techniques like HNSW (Hierarchical Navigable Small World) and IVF (Inverted File Index).

Transaction management

Maintaining transactional integrity while supporting analytical workloads requires sophisticated transaction management:

- **Data modeling strategies:** Techniques for maintaining ACID properties across a distributed environment.
- **Isolation level flexibility:** Support for different transaction isolation levels to balance consistency and performance for different workloads.
- **Write optimization:** Techniques like log-structured merge trees that optimize write performance while maintaining read efficiency.

Performance optimization techniques

Several advanced techniques enable unified stacks to deliver high performance across workload types:

- **Query compilation:** Just-in-time compilation of queries into native code to reduce interpretation overhead.
- **Adaptive execution:** Dynamic query plan adjustment based on runtime statistics and resource availability.
- **Workload management:** Intelligent resource allocation that prioritizes critical workloads while ensuring fair sharing of system resources.
- **Predictive caching:** Machine learning-based approaches to predict and pre-load data likely to be needed by upcoming queries.

Appendix B
Detailed industry use cases

Financial Services

Intelligent wealth management

Business challenge

Wealth management firms struggle to provide personalized investment advice at scale. Traditional approaches either rely on standardized portfolios that don't address individual needs, or require expensive human advisors for customization.

AI solution

AI-powered wealth management platforms deliver personalized investment recommendations by analyzing individual financial situations, goals, risk tolerance and market conditions. These systems continuously monitor portfolios and suggest adjustments based on changing market conditions or client circumstances.

Data requirements

Unified view of client financial data, including accounts, transactions and holdings

Real-time market data integration for timely investment decisions

Historical performance data for strategy backtesting

Ability to process both structured financial data and unstructured information like news and research reports

Vector search capabilities to identify relevant investment opportunities based on client preferences

Expected outcomes

Increased client satisfaction through more personalized service

Enhanced advisor productivity by automating routine analysis

Ability to serve previously unprofitable client segments cost-effectively

Improved investment performance through timely adjustments

Financial Services

Conversational banking with RAG

Business challenge

Banks struggle to provide consistent, accurate information across customer service channels. Traditional chatbots are limited to predefined scripts and frequently fail to address complex or unusual customer queries, leading to frustration and escalations.

AI solution

RAG enables conversational banking assistants that access the bank's knowledge base — including product documentation, policies, transaction histories and account information — to provide accurate, contextual responses to customer inquiries.

Data requirements

Unified storage for both structured account data and unstructured knowledge base content

Vector search capabilities to find relevant information based on query intent

Real-time access to account information and transaction history

Ability to maintain conversation context across multiple interactions

Comprehensive security controls to protect sensitive financial information

Expected outcomes

Improved customer satisfaction through more accurate, helpful responses

Lower training costs for customer service reps

Reduced call center volume as more inquiries are resolved through digital channels

Consistent information delivery across all customer touchpoints

Retail

Dynamic pricing optimization

Business challenge

Setting optimal prices is increasingly complex in competitive retail environments. Static pricing strategies fail to capture the value of products as market conditions change, while manual adjustments can't keep pace with the volume of SKUs and competitive price movements.

AI solution

AI-powered dynamic pricing systems continuously analyze market conditions, competitor prices, inventory levels and customer behavior to recommend optimal price points. These systems identify price elasticity by product and customer segment, enabling more sophisticated pricing strategies.

Data requirements

- Real-time competitive pricing data
- Current inventory levels and sell-through rates
- Historical price performance and elasticity metrics
- Customer segment behavior and price sensitivity
- Ability to process high volumes of price changes across large catalogs

Expected outcomes

- Increased profit margins through more optimal pricing
- Improved competitive positioning
- Faster response to market changes
- More effective promotion and clearance strategies

Retail

Omnichannel customer experience

Business challenge

Customers expect consistent, personalized experiences regardless of how they interact with retailers. Traditional siloed systems create disconnected experiences as customers move between online, mobile and in-store shopping.

AI solution

AI-powered omnichannel platforms could create seamless customer journeys by maintaining context across touchpoints, predicting customer needs and enabling personalized interactions regardless of channel. These systems might recognize customers across devices and locations, providing consistent recommendations and service.

Data requirements

Unified customer profiles accessible across all channels

Real-time synchronization of customer interactions and transactions

Contextual awareness of the customer's current journey stage

Integration of online browsing behavior with in-store activities

High-performance queries to enable responsive experiences

Expected outcomes

Increased customer satisfaction and loyalty

More effective cross-channel promotions

Higher conversion rates through contextually relevant experiences

Improved associate productivity with AI-assisted customer insights

Healthcare

Predictive patient risk stratification

Business challenge

Healthcare organizations struggle to identify patients at risk for adverse events, readmissions or disease progression early enough for effective intervention. Traditional risk models often rely on limited data points and fail to capture complex patterns.

AI solution

AI-powered risk stratification systems could analyze comprehensive patient data to identify subtle patterns indicating elevated risk. These systems might continuously monitor patient status, detect early warning signs and recommend preventive interventions tailored to each patient's specific risk factors.

Data requirements

Longitudinal patient records with comprehensive clinical history

Real-time integration of monitoring data and vital signs

Social determinants of health and behavioral information

Ability to process both structured clinical data and unstructured notes

High-performance queries to enable timely interventions

Expected outcomes

Reduced hospital readmissions through