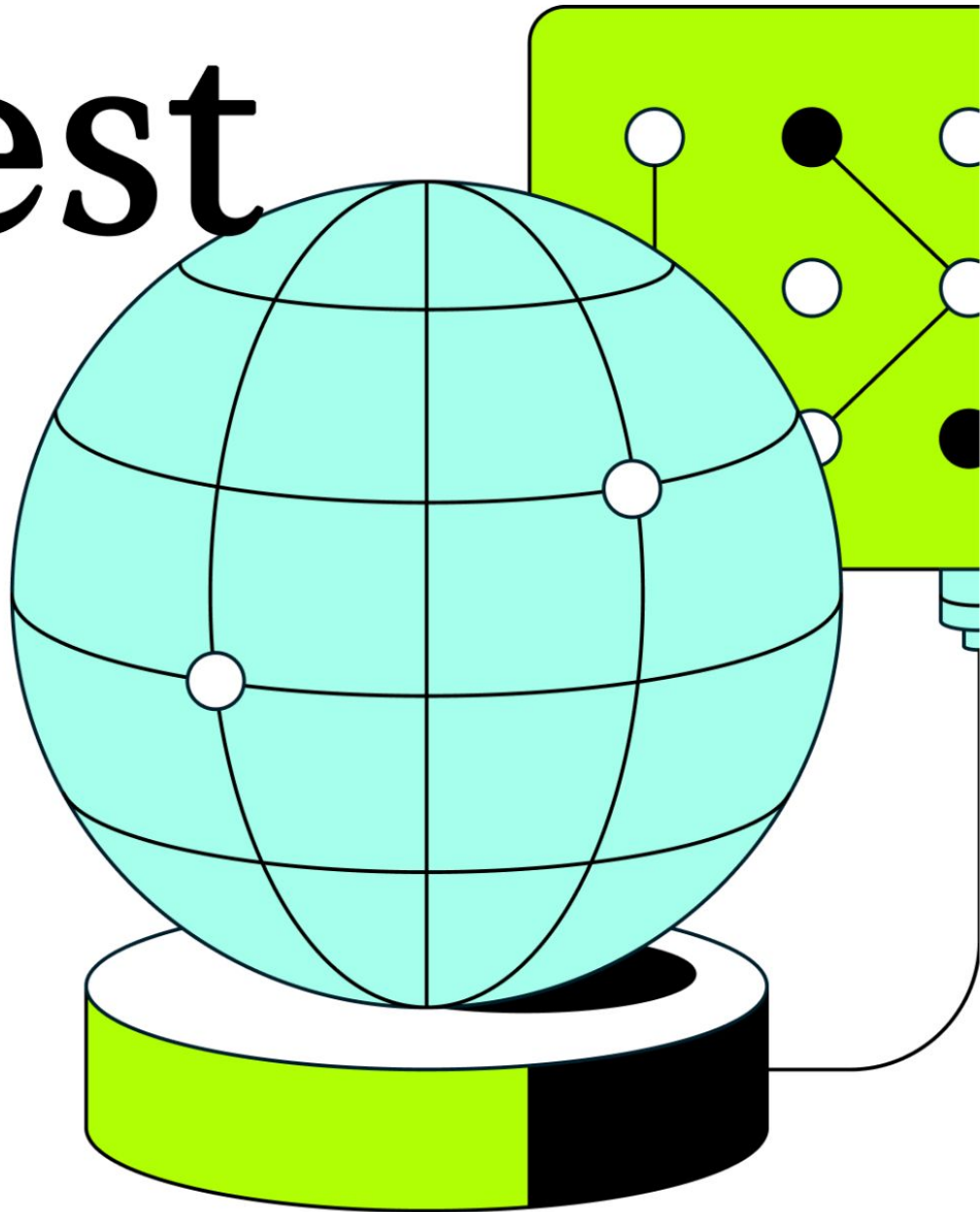


A MAGAZINE BY MONGODB



# Database Digest



ISSUE 01

LAYING THE FOUNDATION FOR FUTURE SUCCESS





© 2025 MongoDB, Inc.  
All rights reserved.

Published since 2025  
by MongoDB,  
1633 Broadway, 38th Floor  
New York, NY 10019, USA

Contributing Editors: Boris Bialek, Sebastian Rojas Arbulu  
Brand Creative: Andrew Considine, Claire Schroyer

# Letter from the Editor

Dear Reader,

Welcome to the first issue of *Database Digest*, a new publication from MongoDB, which arrives at a crucial moment in the history of technology! The rapid rise of AI is redefining how organizations operate, innovate, and compete. This shift demands an entirely new level of speed and flexibility. But as businesses race to adapt, many find themselves constrained by rigid, outdated systems that weren't built for continuous change.

This is why we created *Database Digest*—to explore the evolving role of **data** in powering modern applications. At the heart of every digital experience is a database, storing and processing the information that makes software come alive. Whether booking a dinner reservation, buying a product, managing finances, or simply browsing the Internet, a database makes it possible. Yet many enterprises remain stuck, unable to scale AI-driven applications due to legacy constraints, technical debt, and infrastructure complexity. The missing piece? A modern data foundation built for continuous evolution.

In this issue—"Laying the Foundation for Future Success"—we explore how organizations are reimagining their data infrastructure to support the next generation of intelligent applications. Each article has been carefully curated to provide you with actionable insights and forward-thinking perspectives. You'll also find relevant resources, such as guides from [MongoDB's Solution Library](#), to help you implement these concepts in your own projects.

What sets *Database Digest* apart is our commitment to diving deep. We'll bring you technical insights alongside real stories of transformation from organizations that are pushing the boundaries of what's possible with modern data architecture. You'll hear from industry leaders, technical experts, and innovative companies charting new territories in the data landscape. I hope *Database Digest* inspires you to harness the power of data to drive innovation and success in your own organization!

**Boris Bialek: Vice President and Field CTO,  
Industry Solutions at MongoDB**



# Table of Contents



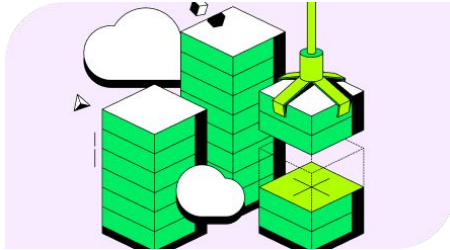
**What's new at MongoDB:** Recent product announcements

PAGE

05

PAGE

08

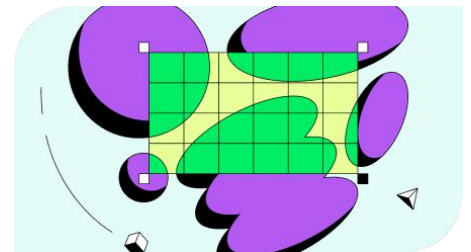


## A MODERN DATA FOUNDATION

Empower your business with strong, adaptable data infrastructures.

PAGE

19



## THE AI REVOLUTION

Leverage data to propel AI advancements and intelligent applications.

PAGE

37

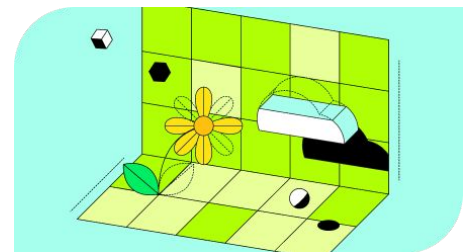


## MODERNIZING ENTERPRISE SYSTEMS

Transform legacy infrastructure with data-driven strategies for enhanced efficiency.

PAGE

44



## WHAT'S NEXT FOR DATA

Explore emerging trends and ethical considerations in the evolving data landscape.



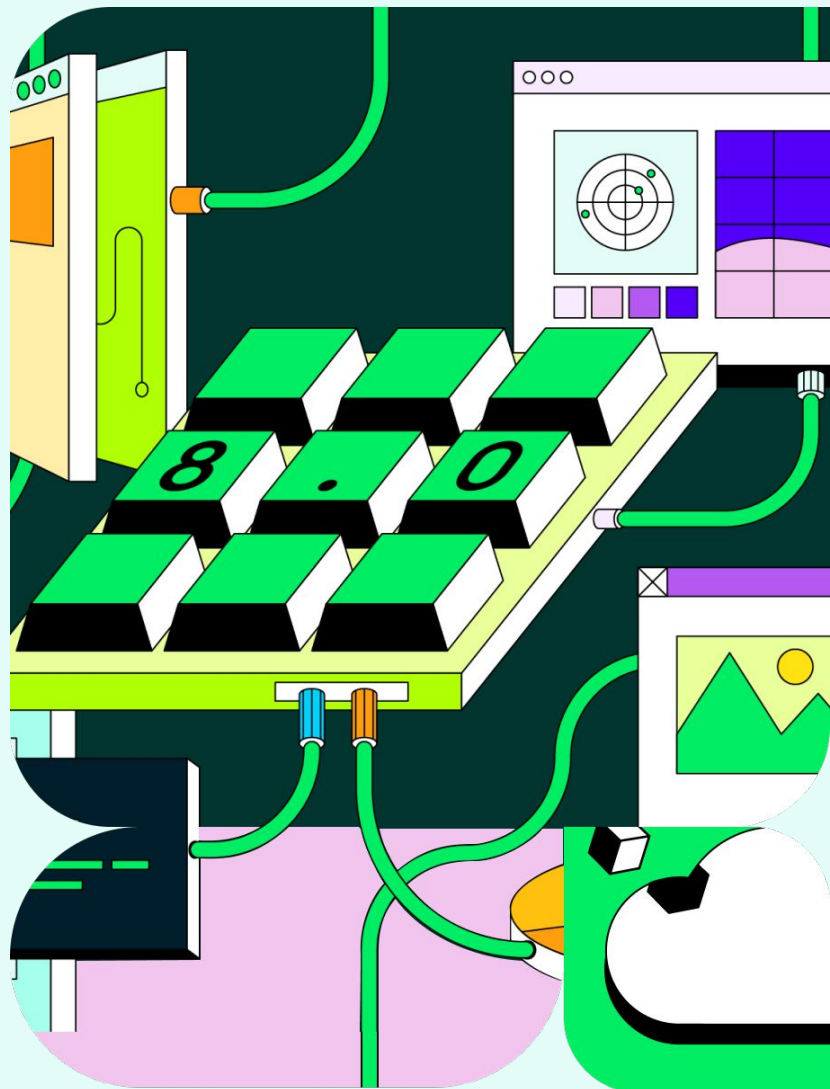


## Database Digest

What's new at MongoDB:

# Recent product announcements

In case you missed it, let's check out **what's new** at MongoDB! These updates span performance improvements, expanded AI capabilities, and strategic partnerships, all designed to empower developers and businesses to build cutting-edge applications with ease.



## MongoDB acquires Voyage AI

In February, MongoDB announced that it had acquired Voyage AI, a leader in embedding and reranking models, to enhance AI-powered search and retrieval capabilities directly within the database layer. This strategic acquisition aims to solve fundamental challenges in AI application development by making high-quality embeddings generation, vector search, and reranking native to MongoDB, eliminating the need for separate systems while improving accuracy, reducing latency, and streamlining the developer experience. The integration will happen in three phases, starting with maintaining Voyage AI's current APIs while improving scalability, followed by embedding the capabilities into MongoDB Atlas with auto-embedding and reranking services, and eventually expanding to enhanced multi-modal capabilities and instruction-tuned models.

[LEARN MORE](#)

## Atlas Architecture Center Release

The Atlas Architecture Center offers best practices and guidance for designing and operating MongoDB Atlas at enterprise scale. It's tailored for enterprise architects and platform engineers seeking a cohesive framework to align Atlas with their organization's specific needs. MongoDB's recommendations are structured around the Atlas Well-Architected Framework, which is built on five key pillars: security, operational efficiency, performance, reliability, and cost optimization.

[LEARN MORE](#)

## LangChainGo & MongoDB integration

MongoDB can now be used as a vector store with LangChainGo. With this integration, Go developers can connect LangChainGo to MongoDB Atlas and leverage Atlas Vector Search to quickly find semantically similar documents and build powerful hybrid queries. Furthermore, MongoDB Atlas offers a unified data layer for Go developers to access operational and vector data together.

[LEARN MORE](#)

## Introducing MongoDB skill badges

MongoDB Skill Badges are your new allies for professional development and career advancement. These badges validate your ability to handle real-world application development and data management challenges. Display them on your LinkedIn profile or resume to make a strong impression and gain an edge in the job market. You can learn a skill in 60-90 minutes, practice it with a hands-on lab, and earn a free badge. Explore badges on data modeling, generative AI, querying, and aggregation, and start learning a new skill today.

[GET YOUR BADGES](#)

## Private Preview: Search & Vector Search on Community edition

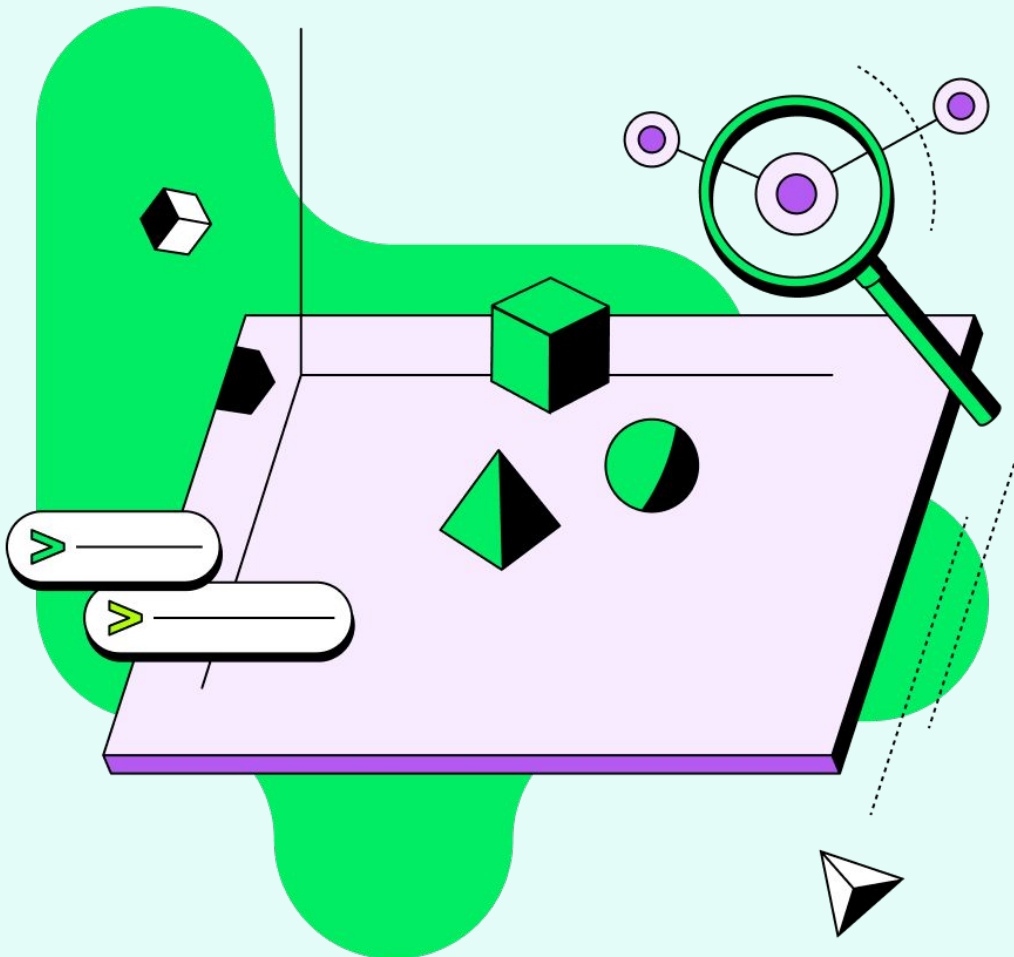
MongoDB's Search and Vector Search features are coming to the Community Edition, bringing advanced functionality to self-managed, local deployment environments. These capabilities let you build and run full text and vector search workloads locally, enhancing efficiency by scaling search separately from the core database. Available via private preview.

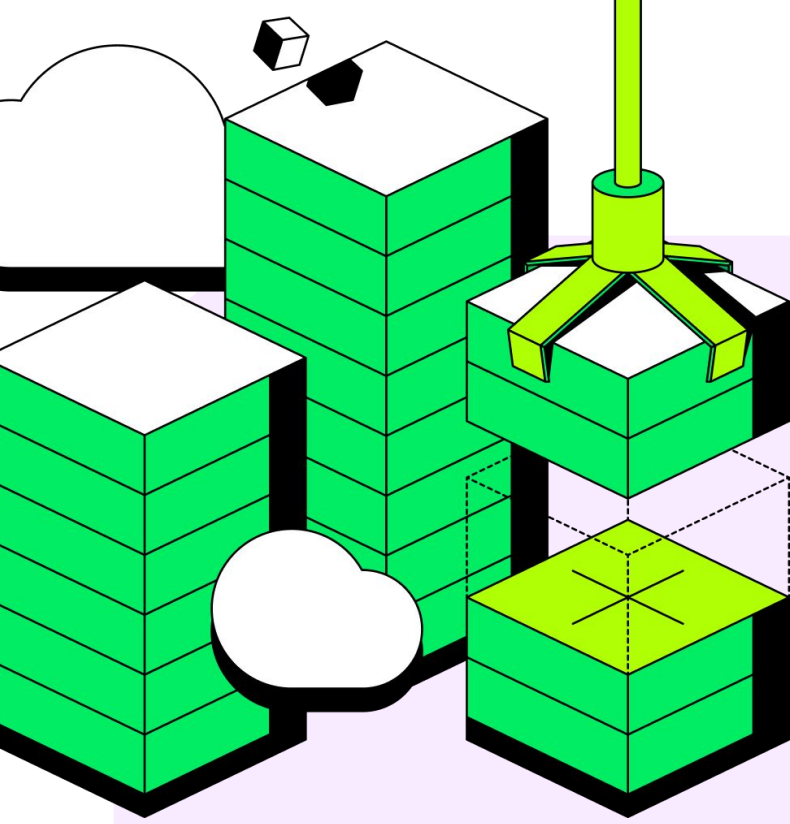
[SIGN UP FOR PRIVATE PREVIEW](#)

## Atlas Stream Processing AWS Private Link Support for AWS MSK

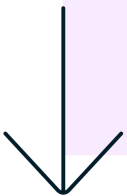
Atlas Stream Processing now supports using Private Link for AWS MSK. This allows for private, secure communication between Stream Processors and Kafka clusters. Security-conscious customers can rest assured knowing their data is being processed by Atlas Stream Processing doesn't traverse public networks.

[LEARN MORE](#)





# A Modern Data Foundation



# A modern data foundation

We live in a world where data is constantly being generated, from our smartphones to our smart homes. But how do we make sense of all this information? Businesses are realizing that this data holds immense potential, but only if they can effectively manage and use it. The pace of technological change has created both opportunities and challenges—from managing exponentially growing data volumes to meeting ever-increasing customer expectations for real-time, personalized experiences. Organizations find themselves at a crossroads: adapt how they store and use data to meet these demands or risk falling behind.

This digital transformation isn't just about adopting new technologies; it's about fundamentally rethinking how we store, process, and leverage data to drive business value. Traditional systems, designed in an era of simpler data requirements, often struggle to meet the demands of modern applications. Companies need solutions that can handle diverse data types, scale seamlessly, and provide the agility to adapt to changing business needs.

The business world is entering the "[data ubiquity](#)" era, where data flows naturally through organizations. This widespread availability and interconnectedness of data creates a fertile ground for advanced technologies. One such technology that thrives in this environment is generative AI, powering personalized experiences and optimized processes. Furthering this, Agentic AI, with its autonomous agents, is emerging as a powerful force, enabling systems to independently learn, adapt, and execute tasks based on vast datasets. Within this data-driven reality, enterprises are seeking solutions that keep pace with innovation, and they require a foundation that fully fosters

agility, scalability, and the ability to handle diverse data.

Consider [Swisscom](#), Switzerland's leading provider of mobile, internet, and TV services. The company chose a fully managed cloud database to apply generative AI to its extensive library of expert content for the banking industry. This enables Swisscom to extract maximum value from its content, producing bespoke summaries for the hundreds of users within the 30 banks subscribing to its services. By converting unstructured data into vector embeddings and using vector search to find segments of relevant information, it can feed LLMs up-to-date contextual, domain-specific data.

Swisscom can then produce a condensed summary within seconds, transforming the speed, accuracy, and relevance of search requests. In the fast-moving financial services industry, with constant breakthroughs in gen AI capabilities, this approach empowered Swisscom to move quickly from idea to production, enabling faster speed to market. Many others, like Lombard Odier, Paychex, and more, are doing so as well.

This transformation isn't just about having more data—it's about having the right foundation to make that data work for you. Companies that adapt their data architecture to meet these new demands will be better equipped to thrive in this data-driven future.

The crucial question is: **how do you build this foundation for the future?** The answer lies in understanding the fundamental building blocks of a modern data architecture, and how specific solutions can empower this new era of data-driven innovation. To guide us, let's explore key database capabilities, essential to this architecture.

# The cornerstone of AI success

The rise of AI, and the speed at which the market has been changing have fundamentally shifted the importance of adaptability. However, software can only adapt as fast as its foundation allows.

At the center of this software foundation is the database. Databases store and process the data that brings software to life. Whether booking a dinner reservation, buying a product, managing your finances, scheduling a service appointment, or just doing something as simple as browsing the Internet, a database makes it possible.

However, traditional relational databases were built around the idea of structure. And “structure” is the anti-pattern for what the new AI world requires: **adaptability**.

Modern databases that are capable of reacting and adapting in real-time and flexible enough to scale with ease are essential to support the continuous evolution required for AI. But it’s more than that.

To thrive, AI applications need a database that can:



Handle All  
Forms of Data  
Seamlessly



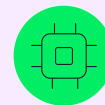
Scale  
Without  
Constraints



React  
Instantly



Provide Intelligent,  
Contextual  
Search



Embed  
Domain-  
Specific AI



Secure Data  
at Every  
Stage



## Handle all forms of data seamlessly

- Consolidate structured and unstructured data—text, video, audio, time series, vectors, event streams—into a single system. AI depends on fluid, instantly accessible data, not fragmented silos.

For example, a fashion retailer can unify purchase history, website clicks, social media sentiment, and in-store foot traffic to predict trends, personalize marketing, and optimize inventory in real time.



## Scale without constraints

- Adapt to unpredictable AI workloads, massive data volumes, and low-latency operations—without constant re-architecting or hitting performance walls as demand spikes.

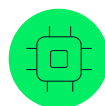
For example, a healthcare system can seamlessly process patient data streams from thousands of devices, run diagnostic AI models, and deliver critical alerts to medical staff in milliseconds—all while scaling to accommodate sudden increases in emergency room demand.



## React instantly

- Power real-time decisions for dynamic pricing, fraud detection, and adaptive user experiences. When market conditions shift or anomalies appear, AI systems can't wait for batch updates.

Real-time processing enables immediate action on critical events, allowing businesses to capitalize on opportunities and mitigate risks the moment they arise.



## Embed domain-specific AI

- Enhance accuracy with industry-specific AI models, leveraging vector embeddings and contextual search to reduce hallucinations and deliver more relevant, reliable, and trustworthy results for your applications.

For example, a manufacturing company can deploy AI that understands specialized machinery terminology, recognizes anomalies in equipment sensor data, and provides accurate maintenance recommendations—significantly reducing downtime and extending asset life cycles.



## Provide intelligent, contextual search

- Retrieve the right information at the right time, grounding AI outputs in accurate, up-to-date data—whether that's surfacing relevant products or powering intelligent agents.

Context-aware information retrieval elevates user experiences by understanding relationships between data points, allowing organizations to deliver precise, relevant results that drive better decisions and outcomes.



## Secure data at every stage

- Protect sensitive information at rest, in motion, and even in use—enabling real-time threat detection and compliance without sacrificing speed or functionality.

Comprehensive security controls embedded throughout the data lifecycle allow organizations to maintain regulatory compliance and protect valuable assets while still leveraging their data for innovation and competitive advantage.

The most successful companies aren't solving these problems by bolting on more tools or throwing resources at the issue—they're rethinking the database they build and deploy software on. To move fast and win, businesses must break free from rigid architectures, fragmented systems, and outdated development cycles. The real challenge isn't just adopting AI—it's ensuring the entire technology stack, from applications to the database, is **built for continuous change**.

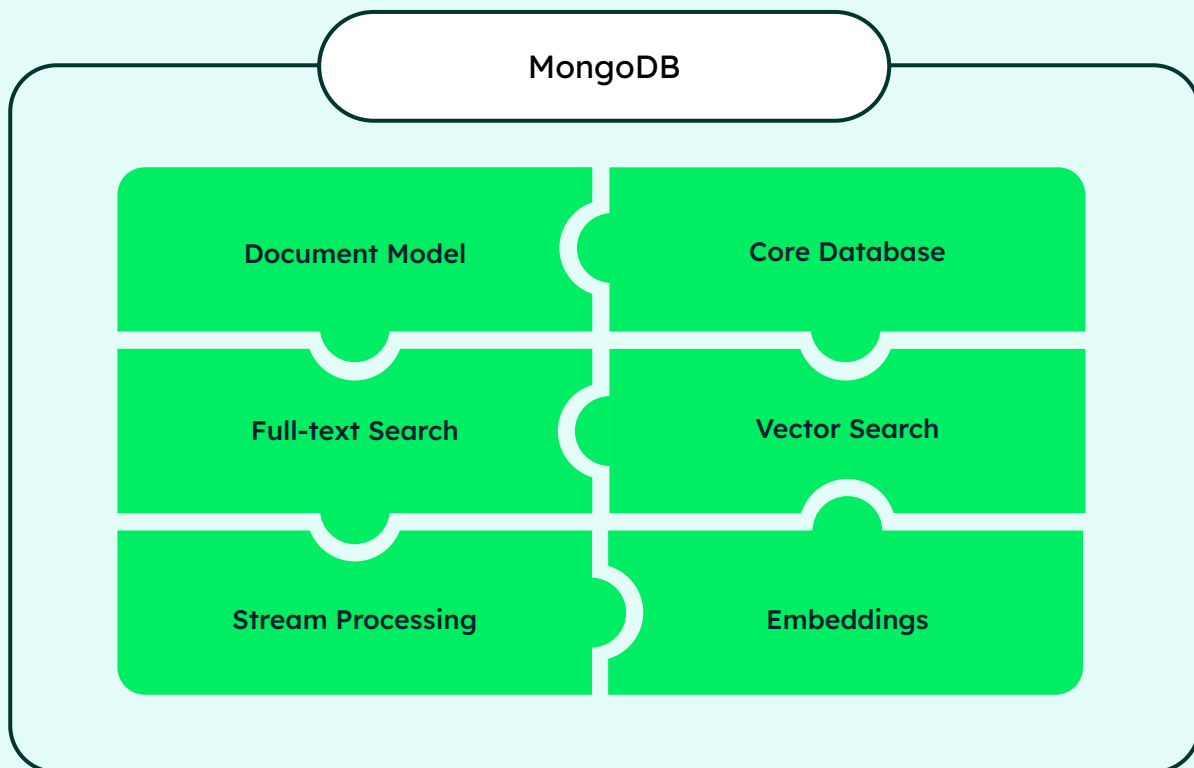
# The Document Model:

## A flexible foundation for modern applications

So, what exactly is the document model? Imagine it as a highly adaptable container for your data. Unlike rigid tables, the document model lets you store information in flexible 'documents', much like JSON files. This approach makes it a superset of other data models. For instance, simple key-value pairs can be stored directly within a document. Relational data, often found in tables, can be more naturally represented by keeping related information together in a single document. And because documents closely mirror objects in programming languages, developers find them intuitive to use.

Even complex data structures, like graphs and geospatial information, can be efficiently stored and queried within these flexible documents. This versatility is what makes the document model a powerful foundation for modern applications.

Due to their rich data modeling capabilities, document databases are general-purpose databases that can store data for a variety of use cases.





## Why Documents Matter

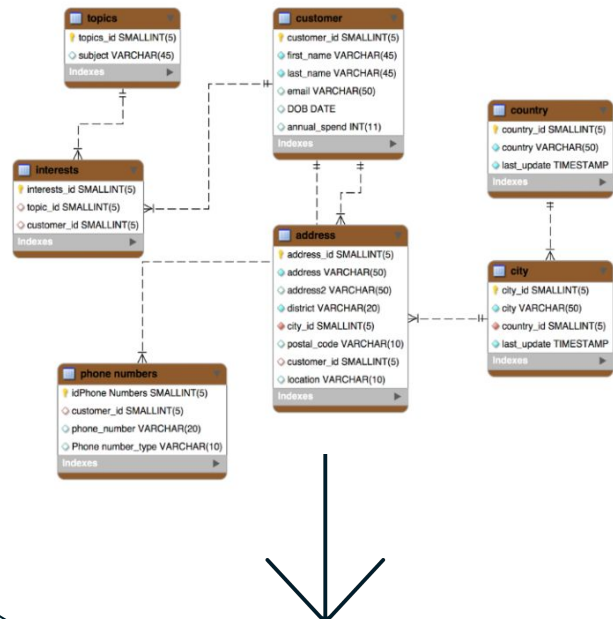
By storing related data together in flexible, JSON-like documents, this approach:

- Aligns naturally with how developers think and code
- Eliminates complex joins and transforms
- Accommodates evolving data requirements without disruptive schema changes
- Supports diverse data types from simple values to complex nested structures

The document model's ability to handle diverse data in a unified way makes it ideal for applications with changing data structures, offering a stark contrast to rigid relational models that require extensive schema migrations. The document model was specifically designed to adapt to evolving data structures and application needs. The pioneers at MongoDB aimed to leverage this model to create a database that was built for change.

## Tabular (Relational) Data Model

Related data split across multiple records and tables



## Document Data Model

Related data contained in a single, rich document

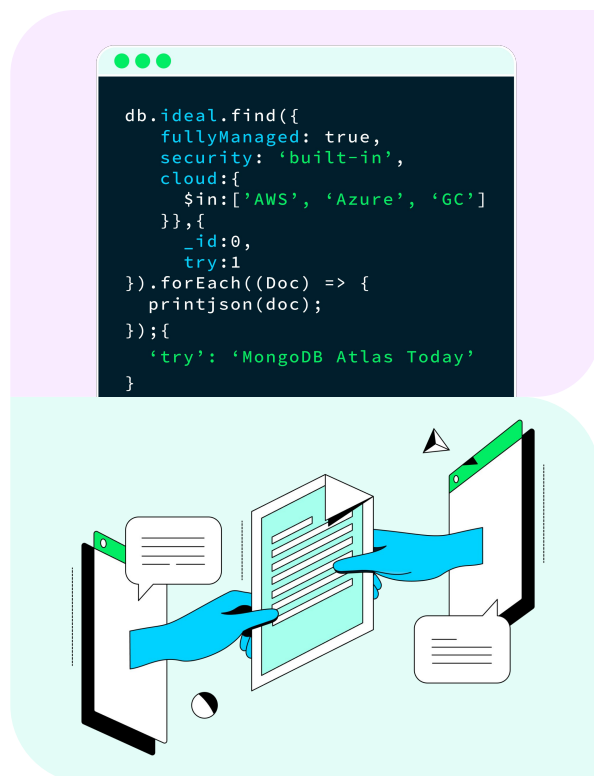
```
{
  "_id" : ObjectId("5ad88534e3632e1a35a58d00"),
  "name" : {
    "first" : "John",
    "last" : "Doe" },
  "address" : [
    { "location" : "work",
      "address" : {
        "street" : "16 Hatfields",
        "city" : "London",
        "postal_code" : "SE1 8DJ"},
      "geo" : { "type" : "Point", "coord" : [
        -0.109081, 51.5065752] } } ],
    { ... }
  ],
  "dob" : ISODate("1977-04-01T05:00:00Z"),
  "retirement_fund" : NumberDecimal("1292815.75")
}
```

# MongoDB's journey: Powering enterprise transformation through innovation

Founded in 2007, MongoDB has revolutionized data storage by utilizing flexible, JSON-like documents rather than the rigid tables and columns of traditional relational databases. MongoDB's document model allows related data to be embedded within a single document or collection. This approach organizes information naturally, reflecting how applications actually utilize data, rather than forcing it into predefined spreadsheet cells. This approach, using the familiar JSON/BSON format, empowers developers to create modern applications far faster and with significantly less complexity. Because JSON is so widely used in web development, developers can quickly pick up MongoDB and start building without a steep learning curve. This reduces development time and allows teams to focus on innovation rather than wrestling with complex database schemas.

MongoDB's evolution has been marked by continuous innovation and expansion since its founding. When the first version shipped in August 2009, MongoDB validated a new approach to database design—built on a JSON-like document data model with an elastic and distributed systems foundation. Those early releases quickly gained adoption across startups and enterprises alike.

With product-developer fit established, MongoDB's engineering team shifted focus toward transforming the system from a niche NoSQL database into the industry's first comprehensive, unified, intelligent data platform. MongoDB Atlas now supports everything from operational and transactional workloads with integrated full-text search to real-time analytics and mobile computing at the network edge, all accessible through a unified API.



## “Focus on innovation rather than wrestling with complex database schemas”

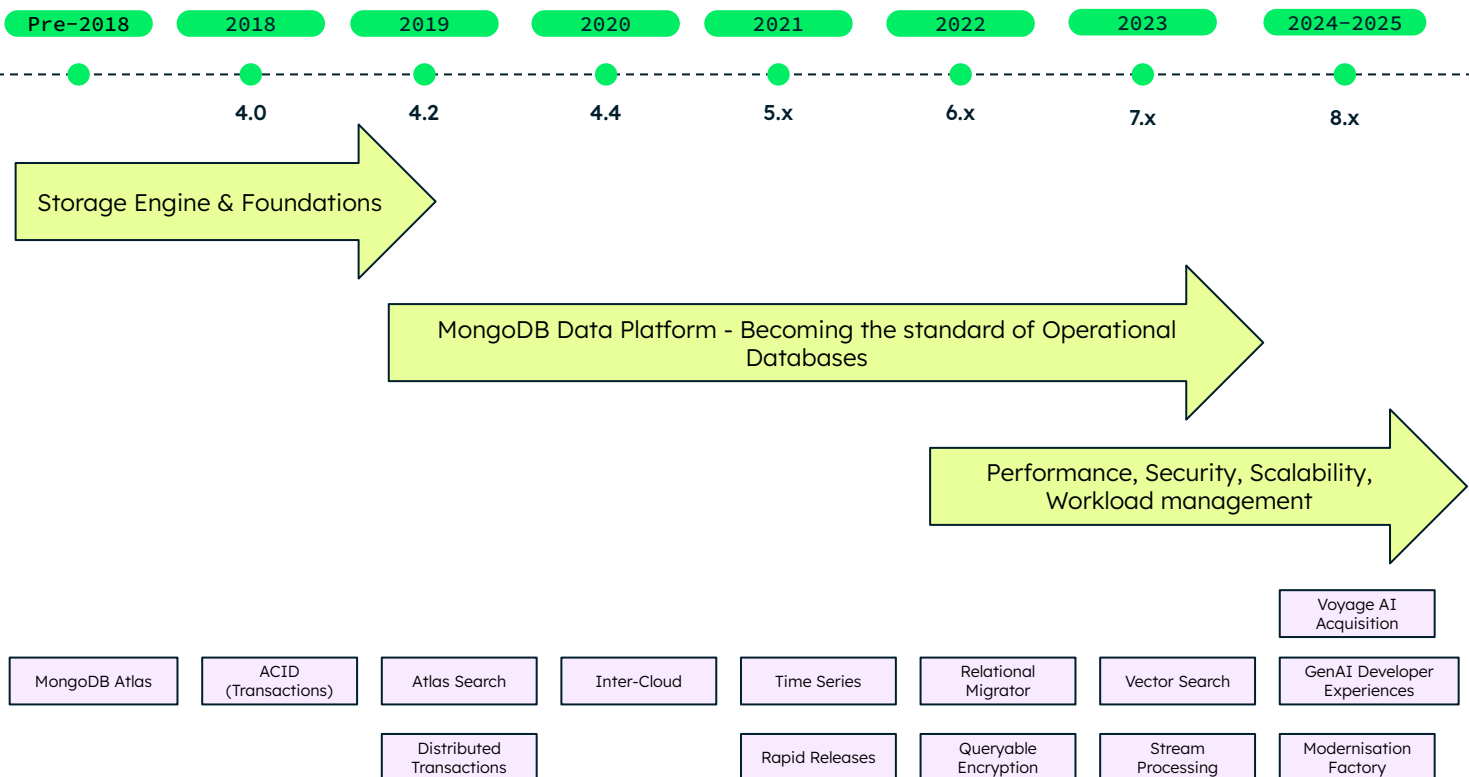
The acquisition of WiredTiger in 2015 and the launch of MongoDB Atlas in 2016 established MongoDB's foundation for enterprise-scale operations. From 2018 onward, MongoDB evolved into a comprehensive suite of database services and tools, becoming a market leader in operational databases with the introduction of critical enterprise features such as ACID transactions, distributed transactions, and multi-cloud capabilities.

As of 2025, MongoDB continues to advance with version 8.0 marking a significant milestone in our enterprise journey. This latest release sets a new industry standard by delivering remarkable performance improvements across key operations—with up to 36% faster reads, 56% faster bulk inserts, and an impressive 200% improvement in time series aggregations. The release also introduces expanded queryable encryption capabilities, including support for range queries, providing industry-leading security without sacrificing functionality.

Beyond performance enhancements, MongoDB 8.0 brings significant advancements in AI integration and scalability, with features focused on generative AI capabilities, vector search, and enhanced developer experiences.

MongoDB is also committed to assisting enterprises in achieving their transformation goals through the Modernization Factory, which specializes in modernizing legacy applications at scale, and Atlas Stream Processing, which facilitates real-time data processing needs. And the recent acquisition of Voyage AI, a leader in embedding and reranking models, further strengthens MongoDB AI's capabilities by bringing state-of-the-art search and retrieval directly into the database layer.

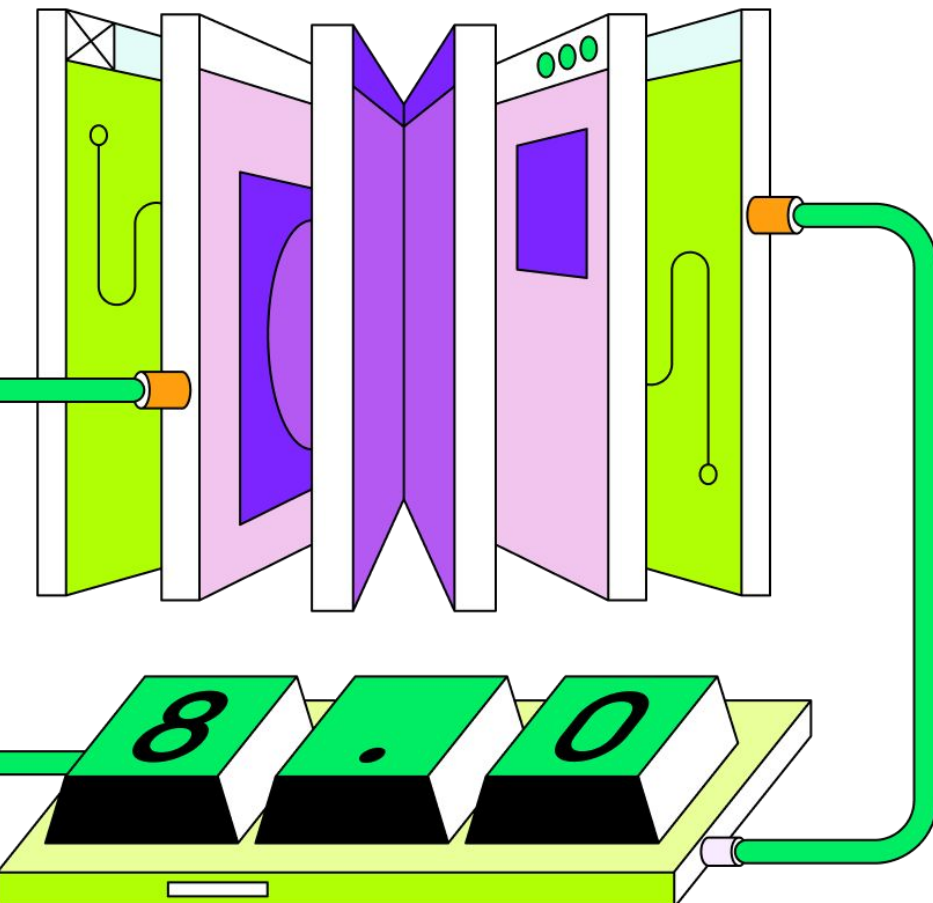
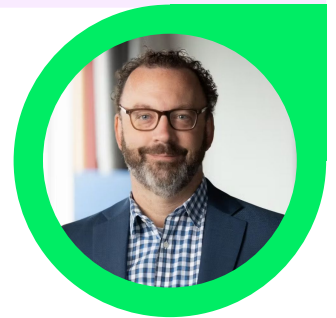
## The Evolution of MongoDB



“

Developers have long loved building with MongoDB, so we've ensured that 8.0 kept the bar extremely high for developer usability. MongoDB 8.0 was also built to exceed our customers' most stringent security, resiliency, availability, and performance requirements, and is the most impressive version of MongoDB yet. MongoDB 8.0 gives customers the strongest possible foundation for building a wide range of applications, now and in the future.

**Jim Scharf, Chief Technology Officer,  
MongoDB**



MongoDB's enterprise journey is particularly notable in its consistent focus on performance, security, and scalability. The evolution from basic database operations to sophisticated workload management capabilities shows MongoDB's deep understanding of enterprise needs. Recent partnerships with major cloud providers like AWS, Microsoft, and Google Cloud have further expanded MongoDB's reach, offering features such as the Atlas flex tier for dynamic scaling and predictable pricing.

# The MongoDB AI Applications Program (MAAP)

In 2024, MongoDB introduced the [MongoDB AI Applications Program \(MAAP\)](#) to help organizations rapidly build and deploy applications with advanced AI capabilities. MAAP provides reference architectures, an end-to-end technology stack that includes integrations with leading technology providers, professional services, and a coordinated support system to make it easier for organizations to create transformative AI-powered applications.

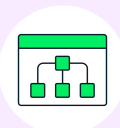
Developed with input from customer experiences, the program addresses practical implementation challenges. For example, [CentralReach](#) is using MAAP along with MongoDB, Anthropic, and LlamaIndex to enhance their platform for autism and intellectual and developmental disability care.

By harnessing AI-driven retrieval and agentic systems powered by a flexible, scalable document database, CentralReach is transforming how therapists, caregivers, and providers deliver care—faster, smarter, and more personalized than ever before.



## MAAP Ecosystem

Technology and expertise from the world's leading AI and tech organizations for a cohesive, supported development experience.



## Composable Architectures

Pre-designed architectures serve as accelerated frameworks for fully customizable applications to accommodate ever-evolving generative AI use cases.



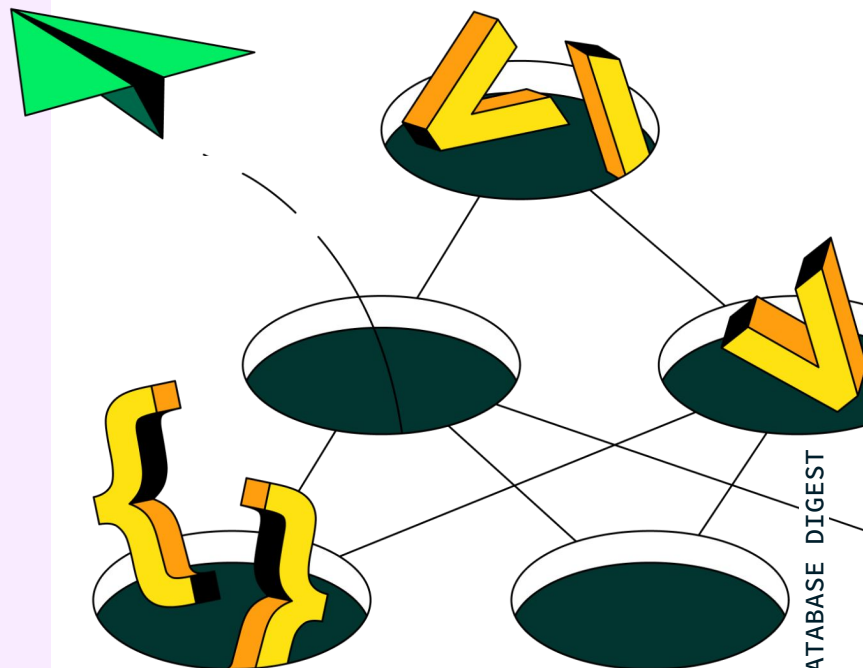
## Expert Services

Strategic guidance on roadmaps, help with data integration into advanced AI technologies, and even development of production-ready applications.



## Education and Enablement

Resources to quickly—and repeatedly—build modern AI applications faster while helping to expand an organization's in-house AI expertise.





# Customer Story

## CentralReach

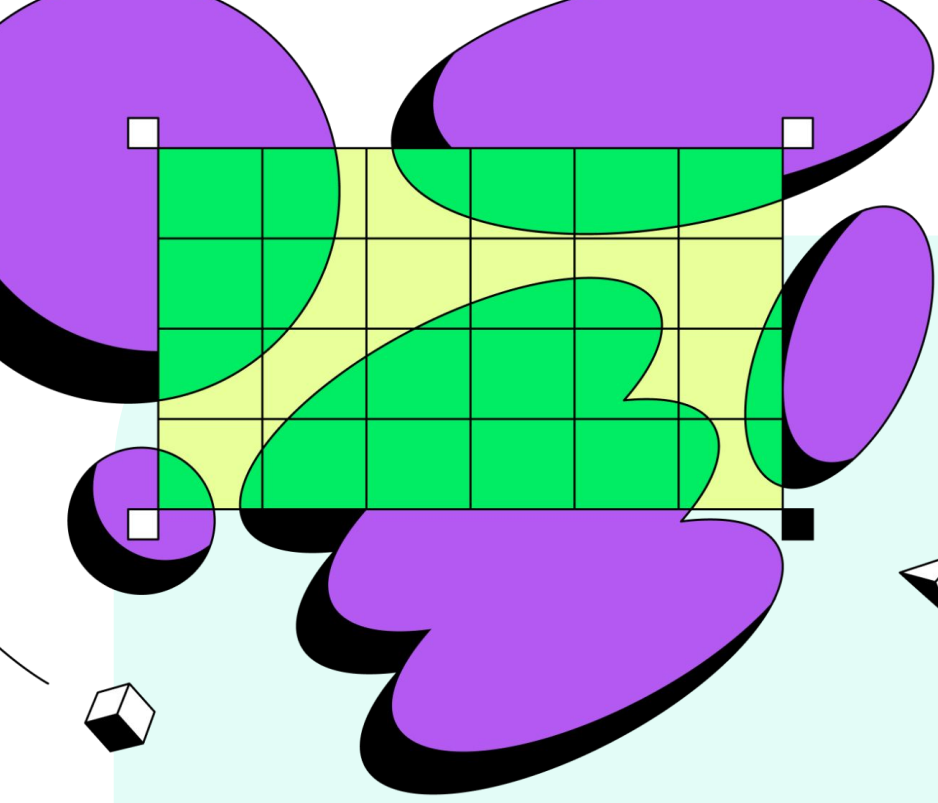
*AI-boosted autism and IDD care:  
CentralReach and MongoDB transform how  
care is delivered*

CentralReach is leveraging MongoDB and the MongoDB AI Applications Program (MAAP) to enhance its AI-driven autism and IDD care platform. With over four billion clinical data points, CentralReach uses MongoDB's flexible data model to integrate information from various providers into a unified Care360 initiative, enabling precise, individualized care. MongoDB's AI tools, including vector search and semantic enrichment, help clinicians efficiently query and analyze unstructured clinical data, reducing administrative burdens and improving patient outcomes.



To further enhance AI-driven retrieval and reasoning, CentralReach collaborates with MAAP partner LlamaIndex for advanced document ingestion, extraction, and indexing. This integration enables their AI system to process vast amounts of clinical documentation, making critical insights more accessible. By working with MongoDB's MAAP network, CentralReach is accelerating its AI capabilities to transform autism and IDD care, ensuring more efficient clinical operations and better support for both providers and patients.

[READ MORE](#)[MAAP ECOSYSTEM](#)



# The AI Revolution







As we delve deeper into the era of data ubiquity, artificial intelligence has become a transformative force in how organizations interact with, analyze, and extract value from their data. Consider, for example, the ability of [AI to diagnose car trouble simply by analyzing engine sounds.](#) The convergence of AI capabilities with robust database infrastructures is unlocking new possibilities in data accessibility, insight generation, and automated decision-making.

From basic chatbots to advanced agentic systems, the evolution of AI technologies is democratizing access to data-driven insights across organizations. This transformation goes beyond simply adopting new tools—it's about fundamentally reimagining how data assets can drive innovation and deliver value at scale.

Through practical implementations, customer success stories, and technical deep dives, we'll explore how the combination of a unified, intelligent data platform and AI is empowering organizations to build responsive, scalable systems that are prepared to meet the challenges of tomorrow.

A woman with curly hair, wearing a light green shirt, is smiling and looking at a tablet device. She is sitting on a couch, and a snake plant is visible in the background. The image is partially obscured by a large white circular graphic element.

Prepared to  
meet the  
challenges of  
tomorrow.



Before diving into complex AI systems, it's essential to understand the **fundamental components** that make them work and how we visualize the space itself. At its core, implementing AI in your organization requires a well-structured stack of technologies working in harmony to deliver intelligent capabilities.

*NOTE: This definition refers to the AI stack from the perspective of the generative AI landscape. The term "AI stack," as used here, is interchangeable with "gen AI" or "generative AI stack."*

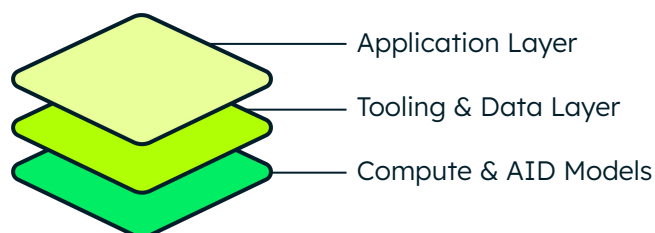
An AI stack refers to a comprehensive combination of tools, libraries, and solutions leveraged to create applications with generative AI capabilities. The components of the AI stack include:

- Programming languages
- Model providers
- LLM frameworks
- Vector databases
- Operational databases
- Monitoring and evaluation tools
- Deployment solutions

While we previously explored the detailed components of an AI stack, we can also examine it from a higher-level perspective through three fundamental layers. At the foundation sits the compute and AI models layer, where companies like NVIDIA provide the high-performance computing infrastructure essential for AI operations. Above this sits the tooling and data layer, where MongoDB serves as a critical bridge between AI models and data, enabling vector storage and real-time processing. The application layer caps this structure, where all

these capabilities converge to create powerful end-user solutions. This simplified architectural view provides a clear framework for understanding how modern AI applications are built and deployed.

### The 3 Layers of AI-Powered Applications



MongoDB serves as the foundation within this stack, managing and storing both vector and operational data. It offers robust data management and search capabilities, supports the real-time processing needs of modern AI applications, handles unstructured data efficiently, and integrates seamlessly with LLM orchestrators and frameworks.

[LEARN MORE](#)

[GET STARTED](#)

By starting with a robust, foundational layer, organizations are able to integrate AI-powered architectural patterns such as retrieval-augmented generation (RAG) and advanced agentic systems.

# Going from LLMs to Agentic Systems

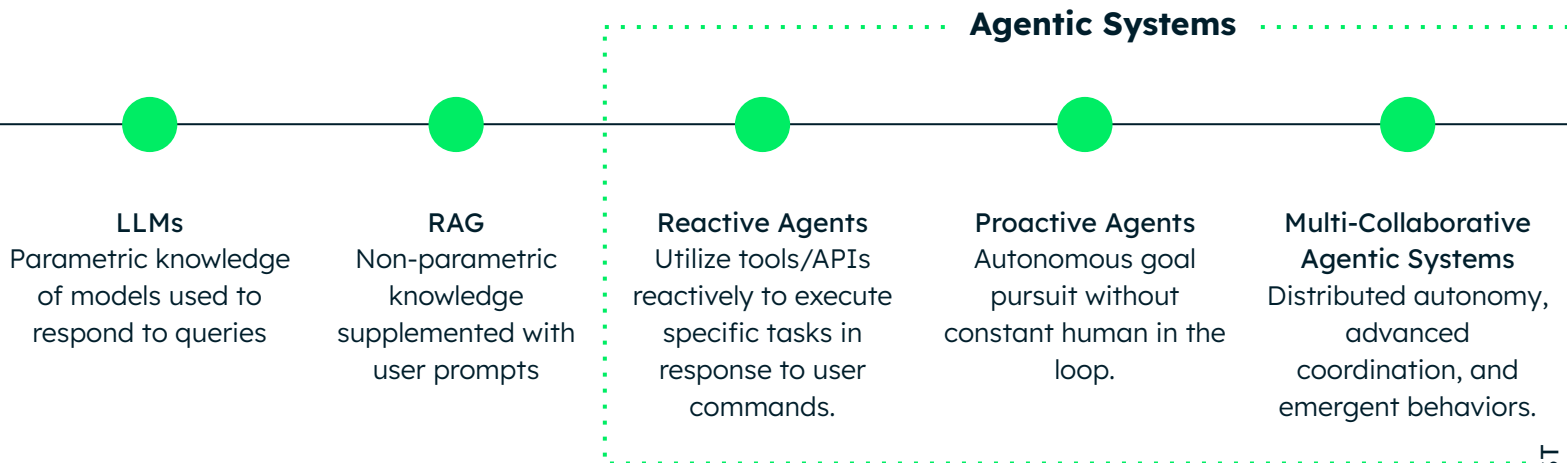
The proliferation of AI applications has occurred in a remarkably short time span. This progression reflects not just technological advancement but a fundamental shift in how we interact with and leverage artificial intelligence.

The journey began with LLMs, popularized by the release of ChatGPT. These initial applications provided access to the vast parametric knowledge contained within large language models, enabling powerful semantic search capabilities through simple chat interfaces. While groundbreaking, this was just the first step.

The next phase introduced retrieval-augmented generation (RAG), which represented a significant leap forward.

Rather than relying solely on an LLM's built-in knowledge, RAG systems supplement queries with relevant, real-time information from external sources. This advancement allowed for more accurate, contextual, and up-to-date responses, making AI systems more practical for real-world applications.

Today, we're entering the era of Agentic Systems and AI Agents driven by sophisticated LLMs with enhanced capabilities in tool usage, reasoning, and planning. Agentic systems comprise reactive agents that respond to explicit commands, proactive agents that autonomously pursue goals, and multi-collaborative agents with distributed coordination capabilities. Improvements are still ongoing in all these areas.



The evolution from simple LLM-based AI applications to autonomous agentic systems represents not only technological progress but also a fundamental shift in how AI can serve as a proactive partner in solving complex business challenges.

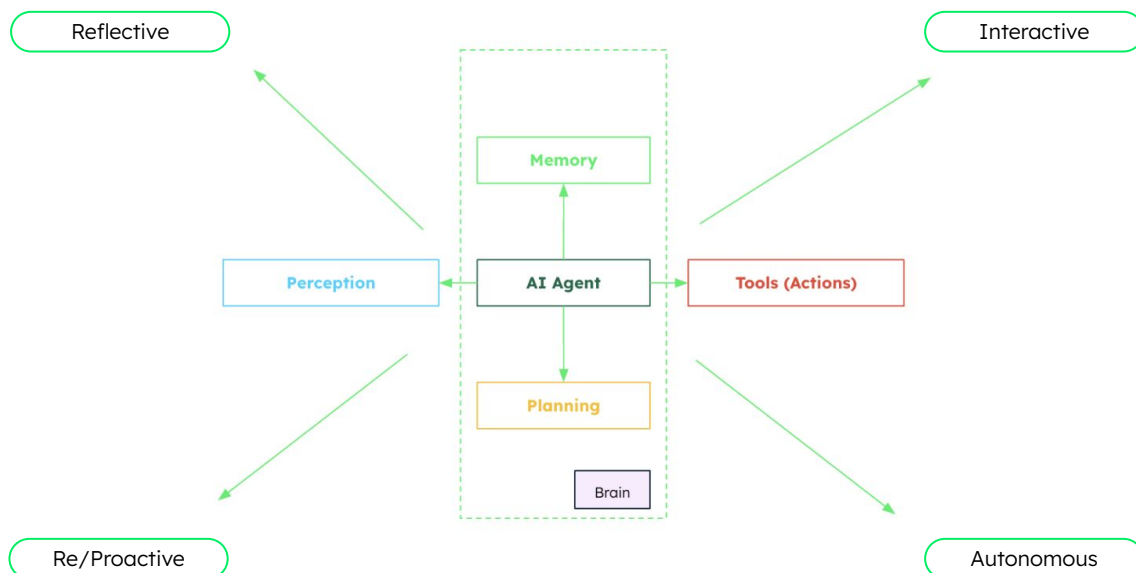
[WATCH WEBINAR](#)

# Demystifying AI agents: A guide for beginners

The recent shift from chatbots powered by large language models (LLMs)—to what the field now defines as agentic systems or agentic AI—can be summarized with an old saying: “Less talk, more action.” AI agents are an advanced step in artificial intelligence designed to perform tasks autonomously by leveraging data and machine learning. Unlike traditional AI systems that require explicit instructions for each action, AI agents can analyze their environment, set objectives, and make decisions to achieve those goals independently. This capability is made possible through the integration of LLMs, which enable agents to understand and process complex information.

The architecture of AI agents typically includes several key components: perception, which allows them to gather data from their surroundings; planning, which helps them devise strategies to meet their objectives; action, enabling them to execute tasks; and memory, which stores past experiences to inform future decisions. By combining these elements, AI agents can adapt to new situations, collaborate with other agents or humans, and handle intricate tasks across various domains.

## Key components and characteristics of an AI agent



As they continue to evolve, AI agents are poised to transform industries by enhancing automation and enabling more intelligent and responsive systems.

[LEARN MORE](#)

The increasing complexity of agentic AI highlights the need for a scalable and flexible database capable of handling dynamic, data-driven decision-making. Recognizing this, MongoDB has positioned itself at the heart of agentic AI—not as a passive data store, but as an active enabler of intelligent systems.

# MongoDB's role in agentic AI

As organizations scale their AI agent deployments, they face several critical challenges: complex data integration across disparate sources, maintaining high concurrency with low latency, and ensuring robust data governance and security. MongoDB provides a versatile set of tools and capabilities to help address these challenges, offering a strong foundation for agentic systems.

## Data related challenges of running Agents at scale



### Complex Data Integration and Handling

AI agents often rely on disparate data sources and formats (structured logs, unstructured text, sensor streams) making it difficult to unify data for real-time decision-making.



### High Concurrency and Low Latency

Agents in production must handle a large volume of requests and respond quickly which can overwhelm databases that aren't built for high throughput and performance.



### Data Governance and Security

Sensitive data like user logs, sensor readings and PII must be stored and accessed securely all while maintaining compliance (e.g. GDPR, SOC 2).

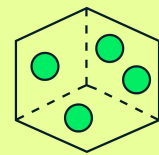
At the core of MongoDB's agentic AI solution is a flexible architecture that supports the entire agent workflow—from profile management to memory handling. The document data model naturally accommodates agent profiles, instructions, and both short-term and long-term memory storage, while powerful features like Vector Search and time series collections provide the building blocks for sophisticated AI applications. This integrated approach eliminates the need for multiple specialized databases, reducing complexity while ensuring secure, scalable, and efficient agent operations.

AI LEARNING HUB

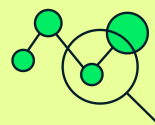
## Why MongoDB is well suited for Agentic AI



### Document Data Model



### Vector Search



### Time Series Collections



### Fine Grained Security & Access Controls



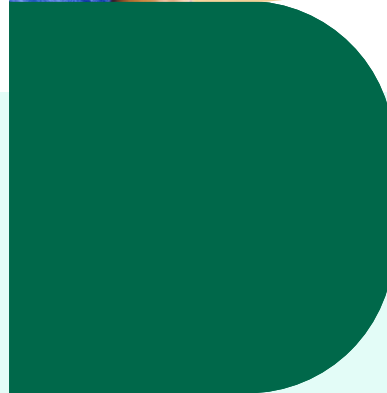
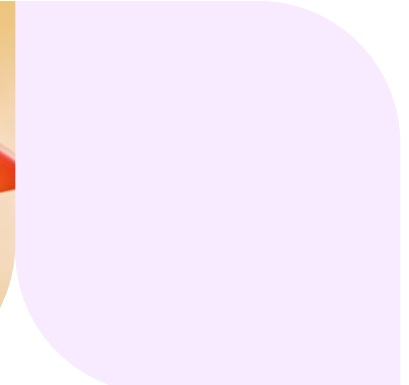
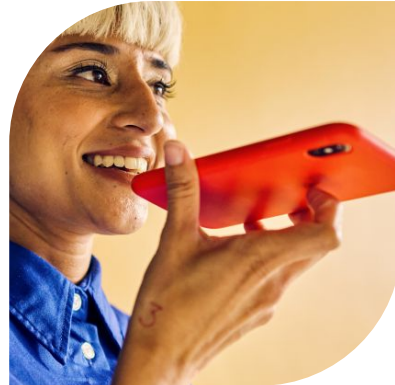
# Customer Story



Questflow

Questflow is revolutionizing the future of work through its decentralized autonomous AI agent network, leveraging MongoDB Atlas to orchestrate multiple AI agents that can collaborate with humans in a "1+N" scenario. By utilizing MongoDB's flexible document model and vector search capabilities, Questflow enables startups to create dynamic AI agents that can handle various tasks from data analysis to content creation, while maintaining context and learning from interactions. The integration with MongoDB Atlas has been particularly crucial for Questflow's development of personalized AI agents, allowing them to efficiently store and process unstructured and AI data—including text, images, and audio—which demonstrates how MongoDB's architecture can support sophisticated agentic systems at scale.

[LEARN MORE](#)

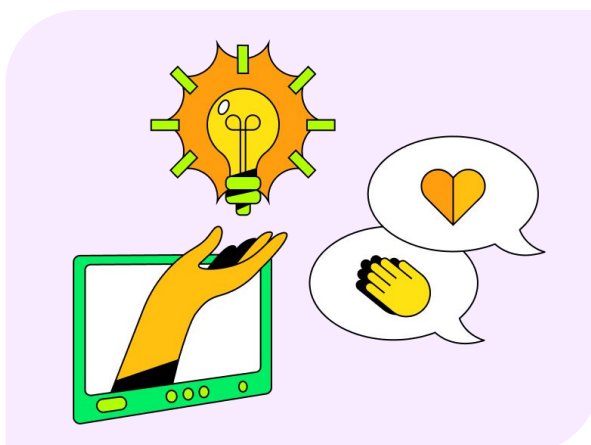
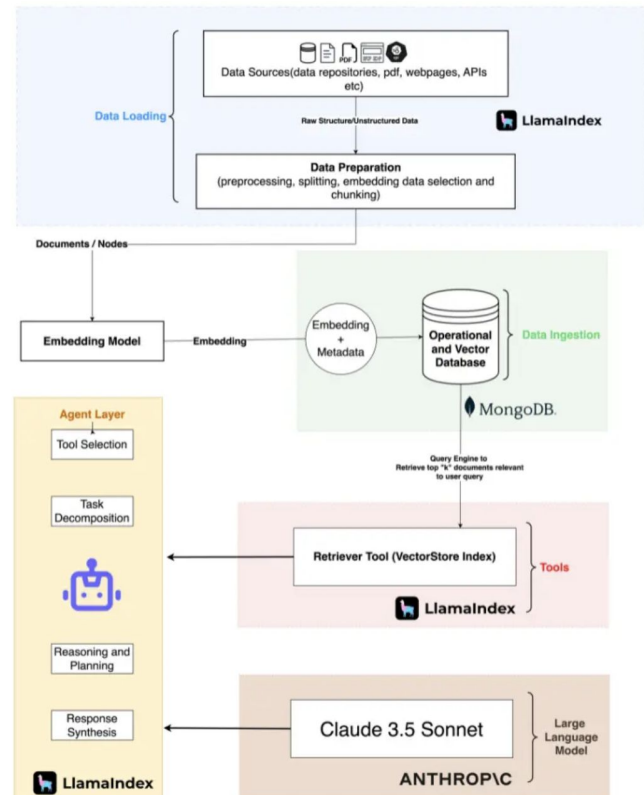


# How to implement agentic RAG using Claude 3.5 Sonnet, LlamaIndex, and MongoDB

Agentic RAG represents the next evolution in AI workflows, moving beyond simple information retrieval to create more dynamic and intelligent systems. Through our partnership with Anthropic and integration with LlamaIndex, MongoDB enables a sophisticated approach to RAG that leverages active reasoning and iterative refinement. For example, an AI assistant planning a trip would iteratively research hotels, flights, and weather conditions, rather than just pulling a single set of results as with traditional RAG. This dynamic approach leverages AI to build comprehensive solutions through multiple rounds of research and reasoning.

This advanced implementation became even more powerful with the June 2024 release of Claude 3.5 Sonnet, Anthropic's multimodal model that sets new standards in reasoning, knowledge comprehension, and coding capabilities. As illustrated in the architecture diagram, this system integrates data loading and preparation through LlamaIndex, vector storage and retrieval via MongoDB, and sophisticated reasoning through Claude 3.5 Sonnet, creating a comprehensive framework for building intelligent, agentic applications

*NOTE: This tutorial is also applicable to the latest version of Claude 3.7.*

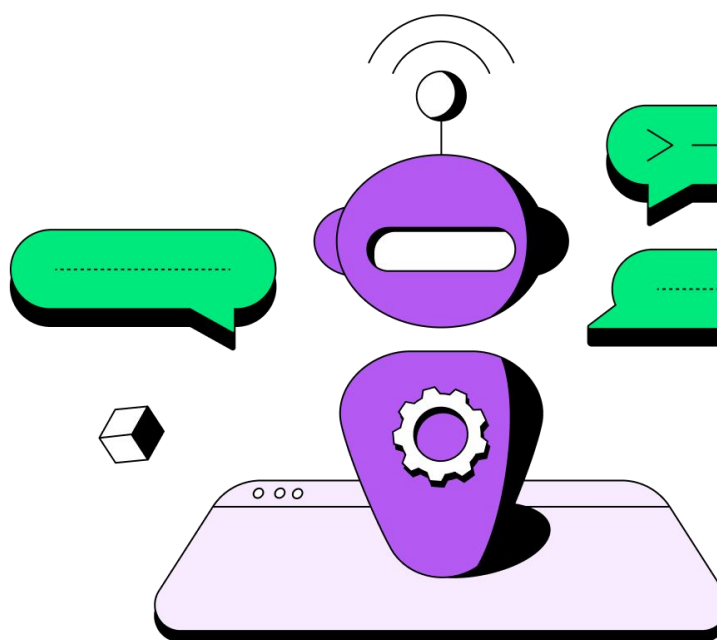


By combining MongoDB's vector storage capabilities with LlamaIndex's data processing and Claude 3.5 Sonnet's reasoning abilities, organizations can build powerful AI applications that go beyond simple question-answering to deliver highly intelligent, context-aware solutions.

[TRY DEMO](#)



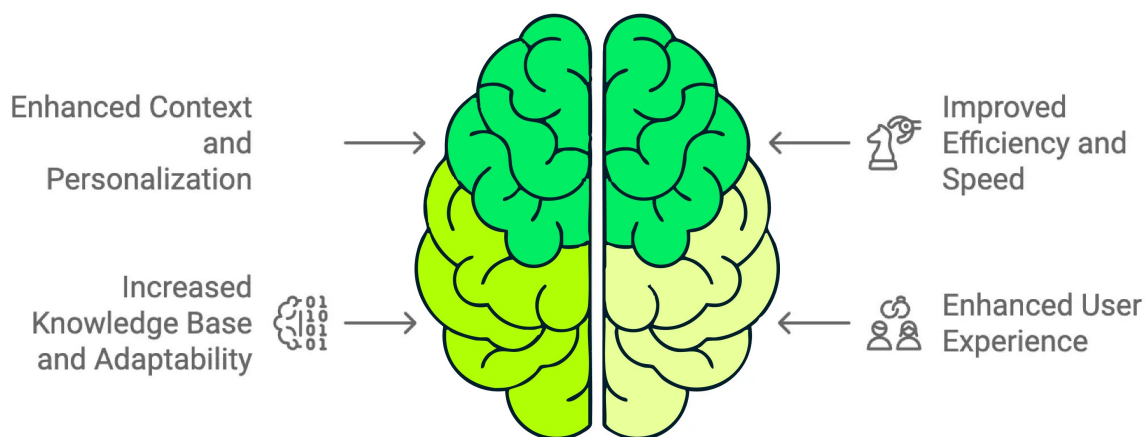
# How to implement working memory in agentic systems for real-time applications



Memory is the cornerstone on which all forms of intelligence emerge and evolve. For computational systems in the modern AI application landscape, such as LLM-powered chatbots, AI agents, and agentic systems, memory is the foundation for their reliability, performance, and applicability, determining their capacity to maintain context, learn from interactions, and exhibit consistent, intelligent behavior. Working memory in agentic systems acts as a dynamic computational component that manages both transient and temporary information through integrated system resources.

As illustrated in the illustration below, this implementation enhances AI systems across four crucial dimensions: context and personalization, efficiency and speed, knowledge base adaptability, and user experience. By leveraging technologies like MongoDB for data storage and retrieval, organizations can build AI systems that not only process information in real time but also learn and adapt from each interaction, creating more intelligent and responsive applications.

## Enhancing AI with Working Memory



[LEARN MORE](#)

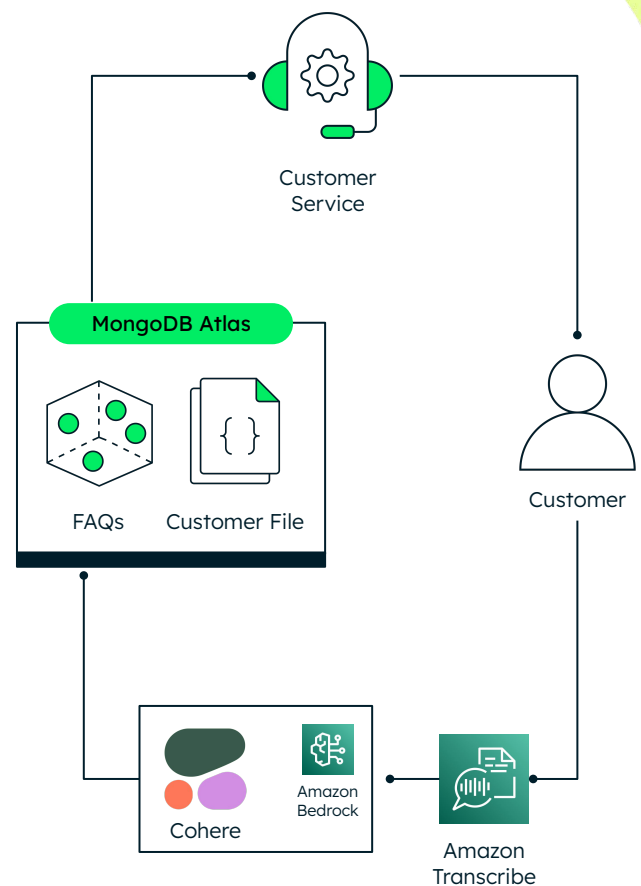
These technical implementations pave the way for real-world applications across various industries.

# AI and digital transformation in action

The implementation of AI and MongoDB across industries is revolutionizing how organizations operate and deliver value to their customers. From enhancing customer service in call centers to transforming healthcare outcomes in breast cancer diagnosis, these real-world applications demonstrate the powerful impact of combining AI capabilities with robust data infrastructure to create more efficient, personalized, and intelligent solutions. Let's explore some of the most impactful use cases where AI and MongoDB are driving digital transformation across different sectors.

## AI-powered call centers: A new era of customer service

By integrating MongoDB Atlas with AI technologies like Amazon Transcribe and Cohere, insurance companies are revolutionizing their call center operations, leading to significant improvements in customer satisfaction and business performance. This transformation enables call center agents to quickly access relevant information through vector search of past interactions, while real-time speech-to-text conversion and semantic search capabilities allow for faster, more accurate responses to customer inquiries. The result is a more efficient system that not only improves customer experience—with insurance companies seeing up to [65% increase in total shareholder return](#)—but also empowers agents with intelligent tools for better decision-making and service delivery.



**System architecture and modules**

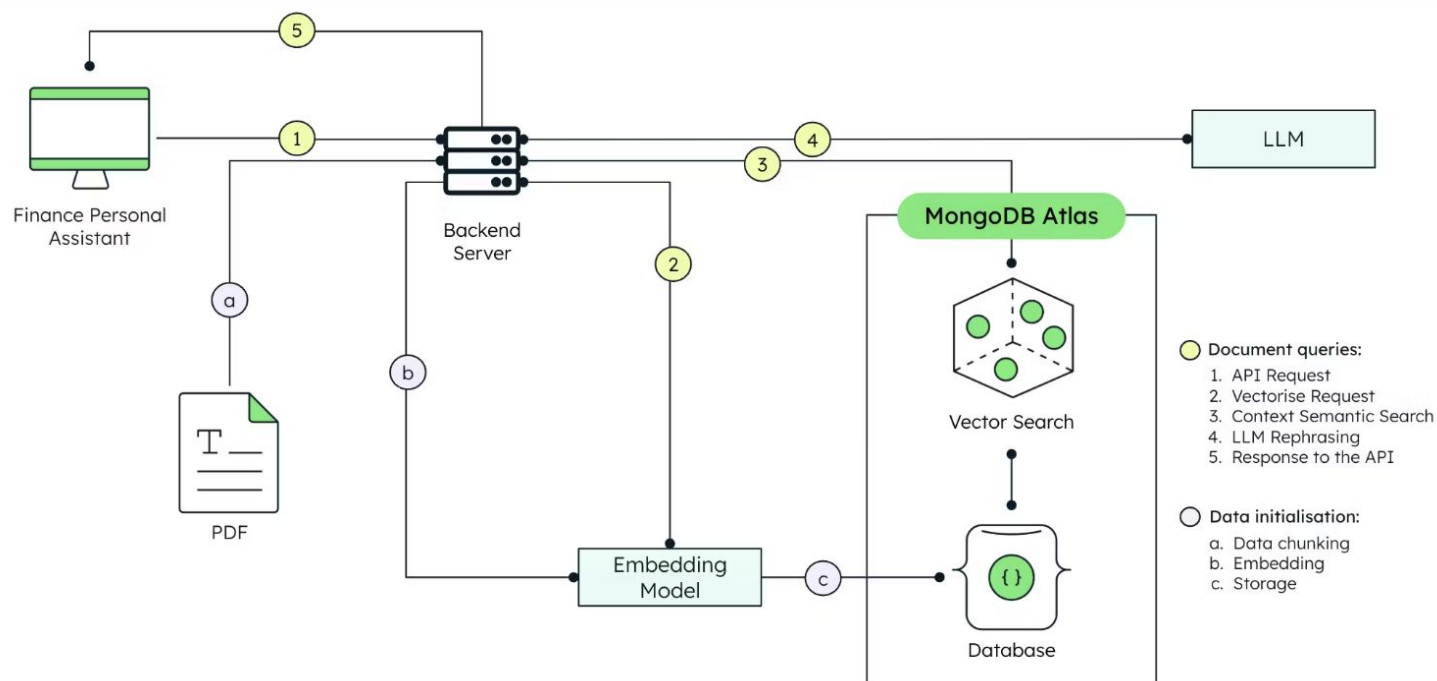
[LEARN MORE](#)





## Better digital banking experiences with AI and MongoDB

Interactive banking is entering a new era powered by MongoDB and AI technologies, addressing the growing demand for more responsive digital solutions in financial services. By leveraging MongoDB's flexible document model and vector search capabilities, banks can create intelligent chatbots and virtual assistants that provide instant, contextual responses to customer queries about complex banking topics like overdrafts, fees, and transactions. This integration, as illustrated in the architecture diagram, enables a seamless self-service experience where customers can get immediate answers without leaving their banking app, leading to improved satisfaction and reduced support calls—particularly important given that 44% of younger consumers report difficulty accessing human support when needed.

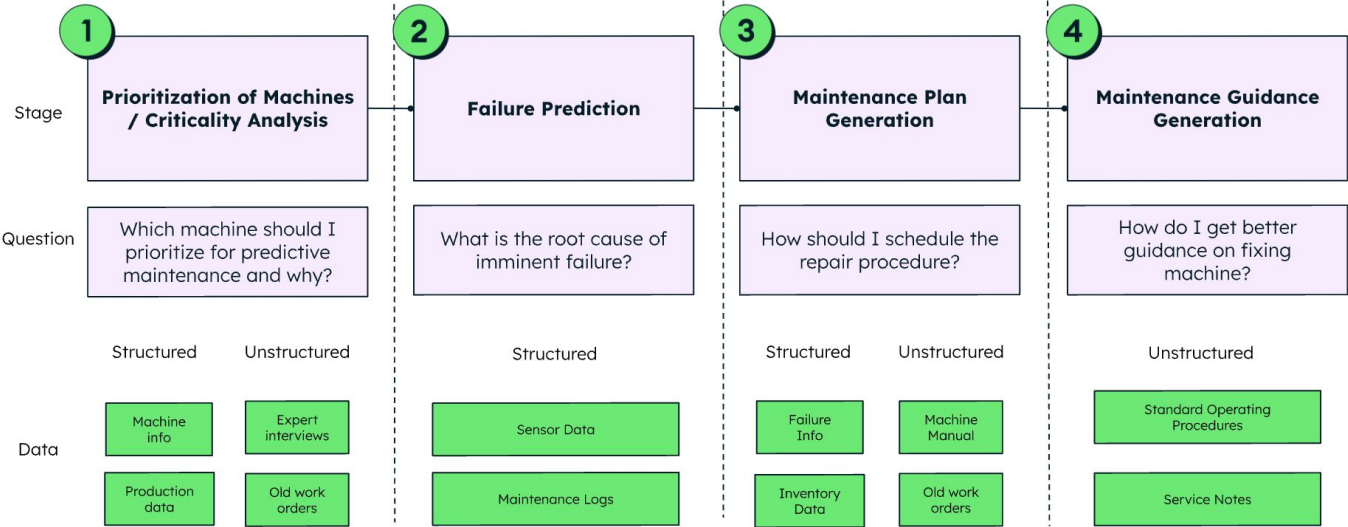


Digital banking solution architecture

LEARN MORE

# Generative AI in predictive maintenance applications

MongoDB and AI technologies are revolutionizing predictive maintenance through a comprehensive four-stage approach that covers machine prioritization, failure prediction, maintenance planning, and guidance generation. By combining MongoDB Atlas with AI technologies, organizations can process both structured data (like sensor readings and maintenance logs) and unstructured data (such as expert interviews and service notes) to create an intelligent maintenance system that significantly reduces downtime and maintenance costs. This integration enables manufacturers to achieve remarkable operational improvements, including a 15-20% reduction in downtime, 5-20% increase in labor productivity, and 30-60% reduction in maintenance costs through AI-powered analytics that help predict failures, generate repair plans, and optimize maintenance schedules.



Four stages of predictive maintenance workflow

LEARN MORE



# Customer Story



In the maritime industry, where ships operate 24/7 moving 90% of global trade, Ceto AI is revolutionizing predictive maintenance by leveraging MongoDB Atlas and Vector Search to analyze vast amounts of sensor data from ship machinery. By implementing AI and machine learning models with MongoDB's flexible architecture, Ceto can detect patterns and anomalies in real time, helping prevent costly machinery breakdowns across maritime fleets. This implementation demonstrates how MongoDB's scalable solution enables companies to rapidly deploy AI models and process massive amounts of data points to drive predictive maintenance solutions in mission-critical industries.

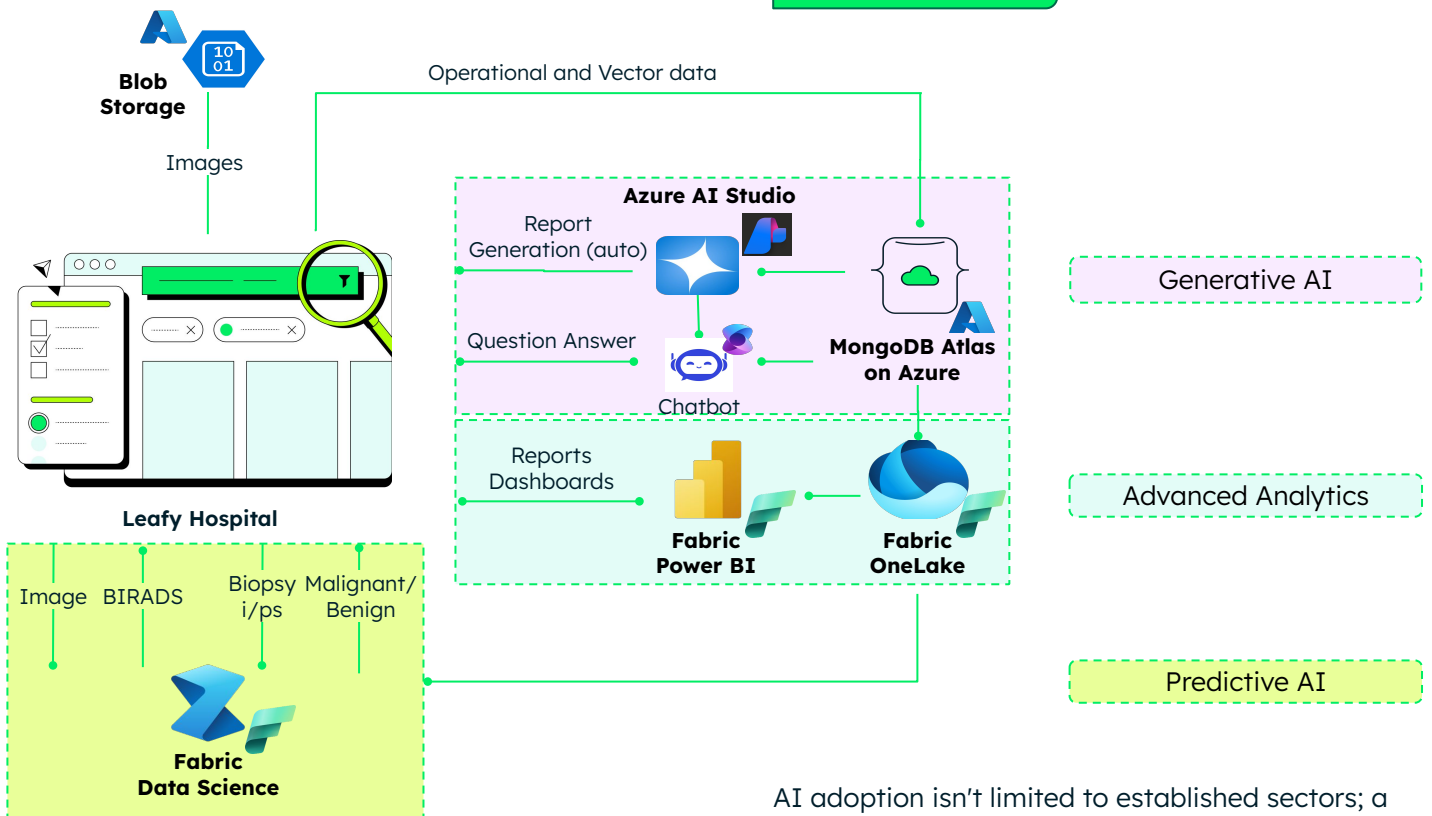
[LEARN MORE](#)[WATCH VIDEO](#)



## Transforming breast cancer diagnosis and treatment with AI

MongoDB and Microsoft have partnered to create an innovative healthcare solution called "Leafy Hospital" that revolutionizes breast cancer diagnosis and treatment through a comprehensive, three-layer AI approach. The solution combines predictive AI for early detection using deep learning models, generative AI for workflow automation through RAG-based chatbots, and advanced analytics powered by MongoDB Atlas and Microsoft Fabric. This integration enables healthcare providers to streamline diagnostic processes, enhance clinical documentation, and make data-driven decisions while maintaining secure handling of sensitive patient information—all supported by MongoDB's unified, intelligent data platform that seamlessly bridges operational, metadata, and AI data components.

LEARN MORE



Leafy hospital solution architecture

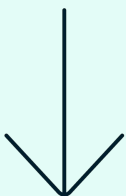
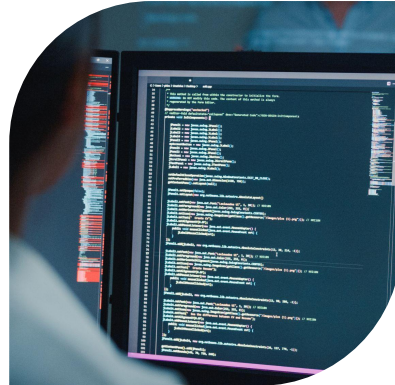
AI adoption isn't limited to established sectors; a diverse array of startups is harnessing its power, with MongoDB playing a pivotal role in their innovative endeavors.



# Customer Stories

## Startups leading the way: AI use cases with MongoDB

Innovative startups are at the forefront of AI adoption, leveraging MongoDB's capabilities to build cutting-edge solutions that transform their industries. Let's explore how emerging companies like **Arc53**, **Hanabi Technologies**, and **Devrev** are using MongoDB to power their AI initiatives and create breakthrough applications that showcase MongoDB's versatility in supporting next-generation AI technologies.



**Arc53's** open-source project, DocsGPT, showcases how MongoDB Atlas can power sophisticated AI documentation assistants, serving nearly 20,000 users including the UK government. The company recently migrated from Elasticsearch to MongoDB Atlas Vector Search for its vector database needs, citing MongoDB's superior handling of dynamic workloads and rapidly changing embeddings in real-time generative AI applications. This migration demonstrates MongoDB's ability to unify source data, metadata, and vector embeddings in a single data layer, making it easier for developers to build and scale AI applications while improving user experience by up to 50% through optimized embedding models.

[LEARN MORE](#)



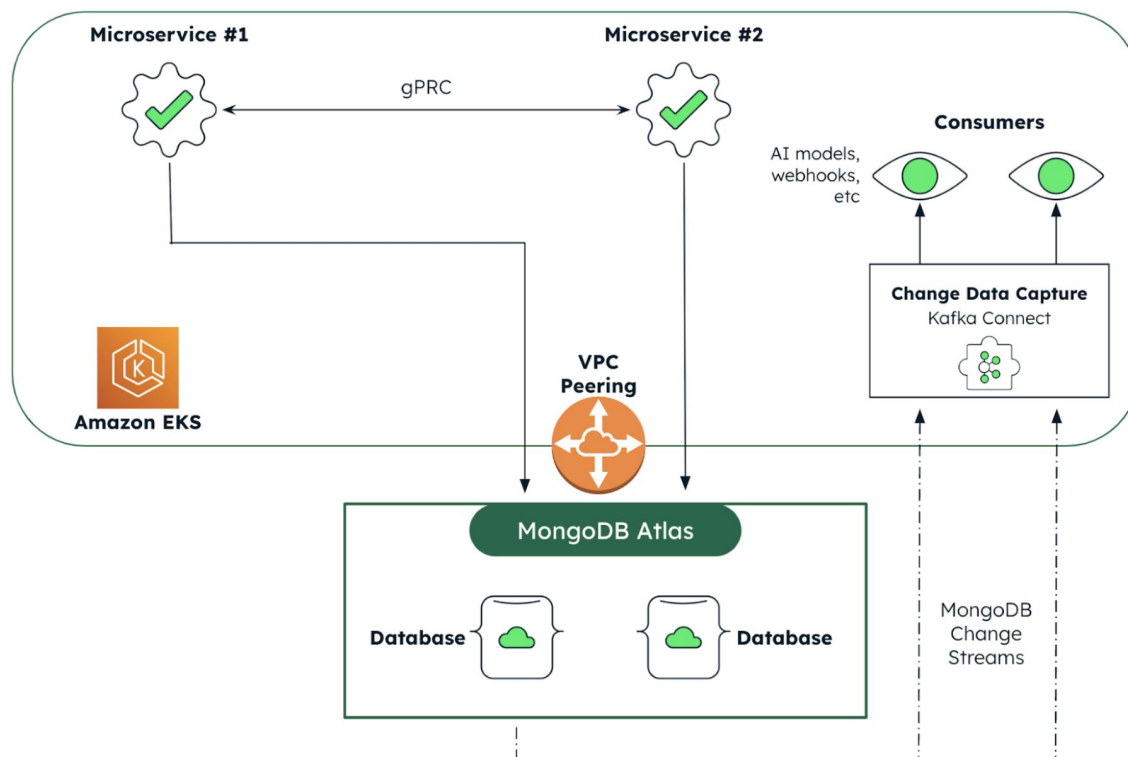
**Hanabi Technologies** has developed Hana, an innovative AI assistant that goes beyond traditional chatbots by proactively helping teams collaborate within Google Workspace, serving over 600 organizations and 17,000 installations. Using MongoDB Atlas as the data layer, Hana leverages the flexible document model and vector database capabilities to enable sophisticated features like retrieval-augmented generation (RAG) across Google Docs while handling massive amounts of data securely and scalably. The combination of MongoDB's capabilities with Hanabi's vision has created an AI assistant that truly functions as a team member, automatically managing tasks like scheduling meetings across time zones and providing intelligent reminders without requiring direct commands.

[LEARN MORE](#)



**DevRev** is redefining CRM for SaaS companies through its OneCRM platform, which serves over 4,500 customers by unifying product management, support, and software development teams on a single AI-native platform. Using MongoDB Atlas as its primary database, DevRev has built a reactive, event-driven microservices architecture that enables it to handle over 4,000 pull requests per month and achieve 3-4x higher development velocity compared with alternative databases. Their innovative approach combines customizable LLMs with MongoDB's flexible document model and change streams to power real-time AI features across their distributed system, creating a truly customer-centric platform that connects developers directly with end users.

LEARN MORE



Innovative startups are leveraging MongoDB to build cutting-edge AI solutions. You too can be part of this transformation. If you're an early-stage startup (Series A or earlier) building AI-powered products or services, join our AI Innovators Program to receive free Atlas credits, technical expertise, and connections to the broader AI ecosystem that can help accelerate your journey to success.

**JOIN INNOVATORS PROGRAM**

To further amplify the impact of AI innovation, a robust partner ecosystem plays a crucial role in providing diverse solutions and expertise.

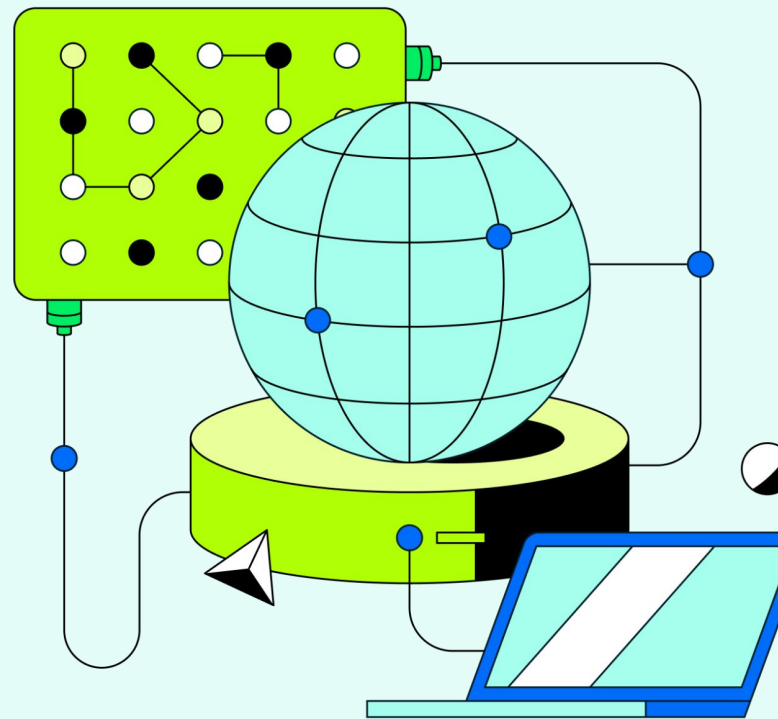
### Partner Ecosystem ISVs/Partners



**Partner Highlight:** *Fireworks AI and MongoDB: The Fastest AI Apps with the Best Models, Powered By Your Data*

## Fireworks AI

Fireworks AI and MongoDB are now partnering to make innovating with generative AI faster, more efficient, and more secure. Fireworks AI was founded in late 2022 by industry veterans from Meta's PyTorch team, where they focused on performance optimization, improving the developer experience, and running AI apps at scale. It's this expertise that Fireworks AI brings to its production AI platform, curating and optimizing the industry's leading open models. Benchmarking by the company shows gen AI models running on Fireworks AI deliver up to 4x faster inference speeds than alternative platforms, with up to 8x higher throughput and scale.

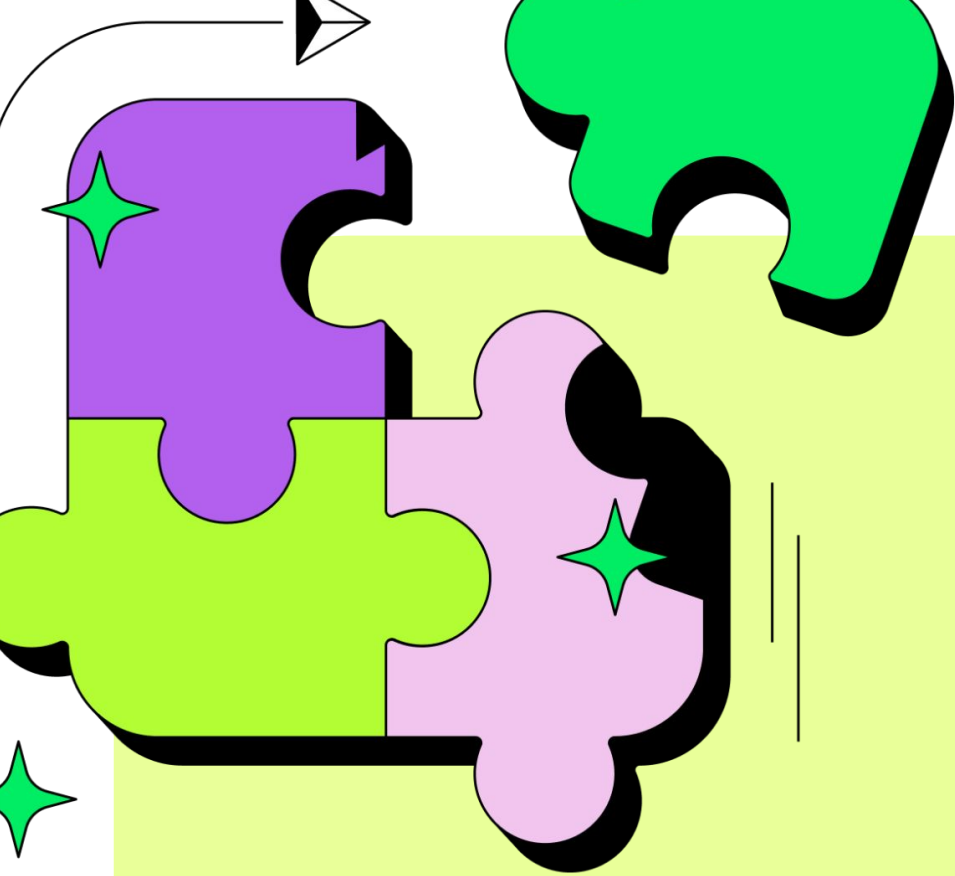


Models are one part of the application stack. But for developers to unlock the power of gen AI, they also need to bring enterprise data to those models. That's why Fireworks AI has partnered with MongoDB, addressing one of the toughest challenges to adopting AI. With MongoDB Atlas, developers can securely unify operational data, unstructured data, and vector embeddings to safely build consistent, correct, and differentiated AI applications and experiences. Fireworks AI and MongoDB provide a solution for developers who want to leverage highly curated and optimized open-source models, and combine these with their organization's own proprietary data — and to do it all with unparalleled speed and security.

Fireworks AI is one of the MongoDB AI Applications Program (MAAP) launch partners. MAAP helps organizations rapidly build and deploy modern applications enriched with AI technology at enterprise scale, including an [end-to-end technology stack](#) that includes integrations with leading technology providers such as Fireworks AI.

**LEARN MORE**





# Modernizing Enterprise Systems



# T

The future belongs to organizations that evolve. Legacy applications, while foundational, can hinder progress. Modernizing these systems is more than a tech upgrade; it's a strategic move toward growth, efficiency, and enhanced competitiveness. Legacy systems consume up to 80% of IT budgets for maintenance, while modernization can cut these costs in half, freeing resources for innovation.

MongoDB is deeply committed to enabling organizations to modernize their systems and applications effectively. The document model and MongoDB's distributed architecture are designed to handle any type of data at scale, which makes it the ideal foundation for AI-powered applications. With Modernization Factory, MongoDB helps to modernize applications up to 60 times faster, empowering them to achieve scalability, cost efficiency and enhance customer experiences.

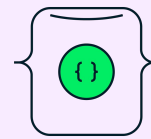
## The ideal foundation for AI-powered applications

# Application modernization and data integration

## Building a modern data architecture: essential considerations

A truly modern data architecture must embody flexibility across multiple dimensions to support today's enterprise needs. From offering both relational and document models to providing deployment options spanning cloud-native, cloud-agnostic, and on-premise environments, organizations require solutions that adapt to their specific contexts. Exceptional non-functional requirements including high availability, robust disaster recovery across multiple availability zones, and active-active configurations ensure systems remain resilient under any circumstance. This technical flexibility, combined with the support of a vibrant developer community, creates the foundation upon which successful modernization initiatives can be built.

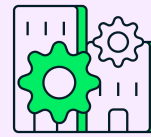
This modern approach enables organizations to break free from legacy data silos, seamlessly integrate AI capabilities, and create a unified data foundation that supports both current needs and future innovations.



Relational or Document



Cloud Native



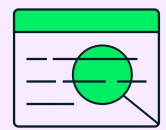
Available as Managed, Service, and on Premise



Exceptional NFRs; Availability, RPO, RTO, active-active



Cloud Agnostic



Open Source



Multi AZ, multi-region, multi-cloud HA/DR



Developer Community



# Customer Story



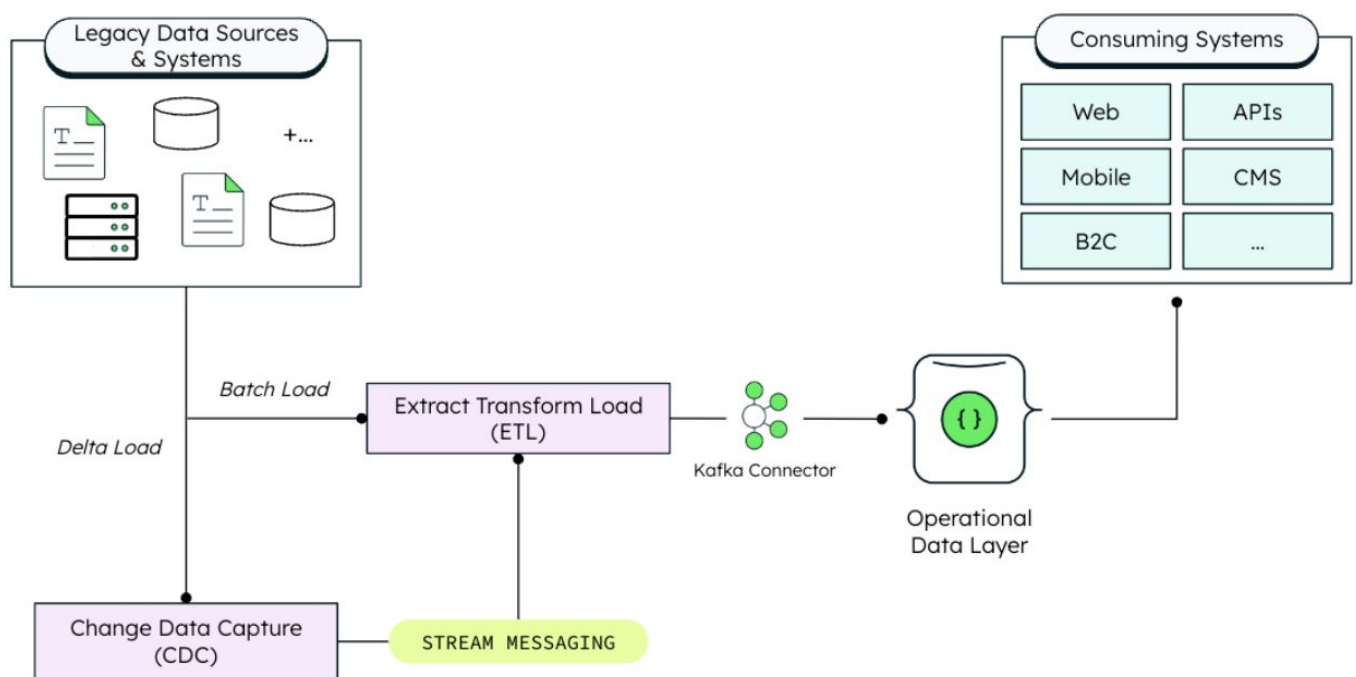
By leveraging MongoDB, Bendigo and Adelaide Bank achieved a 90% reduction in both time and cost to modernize its core banking applications. Bendigo Bank streamlined its transformation by implementing MongoDB's repeatable modernization framework and gen AI-powered [Relational Migrator](#) tool.

READ MORE



## Implementing an operational data layer

An operational data layer (ODL) is an architectural pattern that centrally integrates and organizes siloed enterprise data, making it available to consuming applications. It enables a range of board-level strategic initiatives such as legacy modernization and data as a service, and use cases such as single view, real-time analytics, and mainframe offload. As shown in the architecture diagram, data is integrated into the ODL through both batch and delta loading processes, with change data capture (CDC) and stream messaging ensuring real-time synchronization between legacy systems and the operational layer. The ODL matures through several phases—from simple read offloading to becoming a full system of record—allowing organizations to gradually modernize their infrastructure while maintaining business continuity. Successful implementations illustrate this progression—for example, [Nationwide](#) leveraged MongoDB Atlas to modernize a mainframe system, while [Toyota Financial Services](#), reduced infrastructure management time by 40%.



Conceptual model of an Operational Data Layer

### 3 Pitfalls to avoid when migrating PostgreSQL to MongoDB

When migrating from PostgreSQL to MongoDB, avoiding common pitfalls is crucial to ensure a smooth transition and lay the foundation for future success. First, it is essential to address technical debt in the legacy relational database, such as SQL antipatterns or missing constraints, which can complicate migration. Second, aligning the MongoDB schema design with the application's primary use case ensures optimal data access patterns and performance. Lastly, leveraging MongoDB schema design patterns—like embedding or referencing—can help tailor the data model to specific use cases while reducing rework.

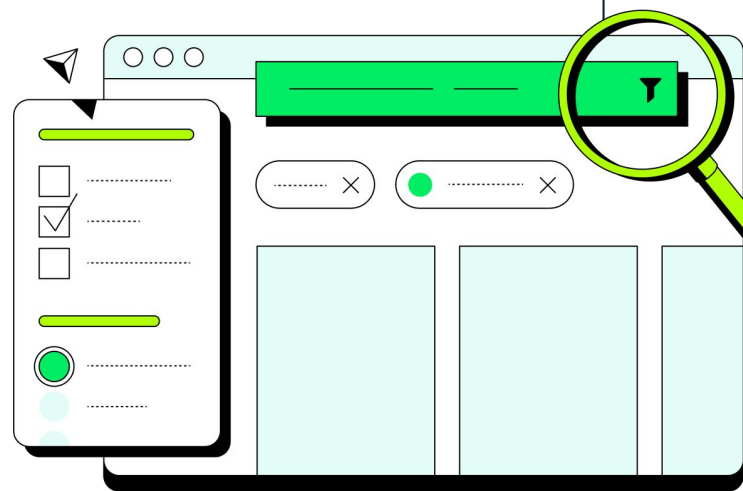
Tools like [MongoDB Relational Migrator](#) simplify this process by streamlining schema mapping and migration execution. This tool streamlines migration by connecting to the relational database, analyzing its schema, and suggesting optimized mappings for MongoDB. Addressing these pitfalls ensures a smoother and more optimized transition.

LEARN MORE

### Process vs. payload fields: Efficiently querying semi-structured MongoDB documents

In the journey toward modernization, understanding how to efficiently structure and query data is crucial for building a solid data foundation. MongoDB's approach to distinguishing between "process" and "payload" fields provides an elegant solution for handling semi-structured data—where process fields enable efficient querying through strategic indexing while payload fields store complex data that only needs to be retrieved, not searched. This intelligent data organization, combined with MongoDB's flexible document model and advanced features like partial indexing and covered queries, enables organizations to optimize their data architecture for both current operational needs and future scalability. Just as organizing a dresser with specific drawers for different types of clothing makes items easier to find, this structured approach to data management creates a more efficient, future-proof foundation for modern applications.

LEARN MORE





## Microservices and composable APIs

Microservices and composable APIs provide a transformative approach to making data more accessible and reusable, especially when supported by MongoDB. By decoupling application functionalities, microservices enable independent scaling, development, and maintenance of components, reducing downtime and enhancing agility. MongoDB's flexible document model and robust API capabilities, such as the unified [Query API](#), empower developers to seamlessly connect services, share data across components, and adapt to evolving needs without rethinking underlying schemas. With MongoDB Atlas, organizations can leverage a fully managed, scalable database service that supports real-time data access and composable architectures, laying a strong foundation for innovation and future growth.

WATCH VIDEO

## Event-driven architecture

Event-driven architectures enable real-time data access and insights by facilitating seamless communication between systems through events, making them ideal for modern applications. MongoDB supports this approach with features like Change Streams, Triggers, and integrations with tools like Kafka, allowing organizations to process and react to data changes instantly. For example, MongoDB's flexible document model and time series collections enable businesses to handle diverse data types and workloads, while Triggers automate actions in response to database changes. This architecture empowers organizations to unlock actionable insights, optimize operations, and build scalable solutions, laying a strong foundation for future growth.

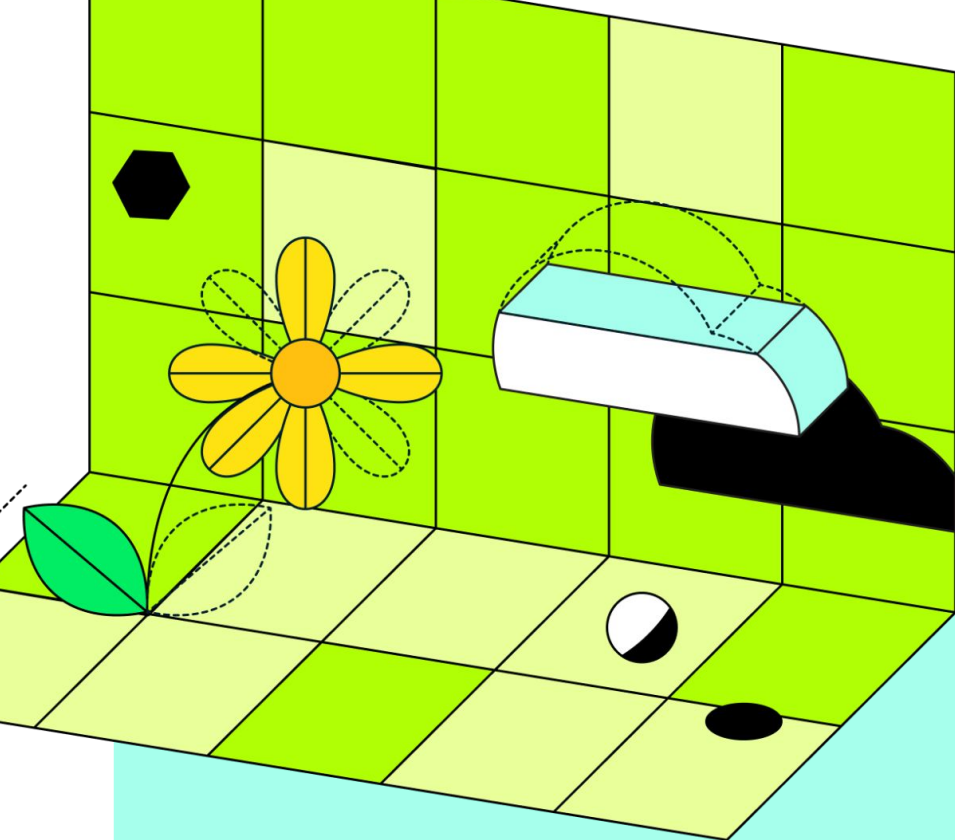
WATCH WEBINAR

## Why MongoDB for modernization

Just as a well-designed building starts with a flexible foundation that can support both current needs and future expansions, modern data architecture requires a foundation that can adapt to evolving business demands. In today's digital landscape, organizations are often held back by rigid, legacy systems that are like buildings with fixed walls—difficult and expensive to modify as needs change.


## MongoDB's document model breaks free from these constraints

by naturally aligning with how developers think about and structure data, allowing organizations to build and modify their applications with the agility of working with building blocks rather than fixed structures. This flexibility, combined with cloud-native capabilities and advanced features like Vector Search, enables organizations to create a future-proof architecture that can seamlessly integrate emerging technologies like AI while reducing the complexity of maintaining multiple specialized systems. The success of companies like Bendigo Bank, which achieved a 90% reduction in modernization time and cost, shows how this approach can accelerate the journey from legacy systems to a modern, adaptable foundation that grows with your business.



# What's Next for Data





# F

## Future perspectives

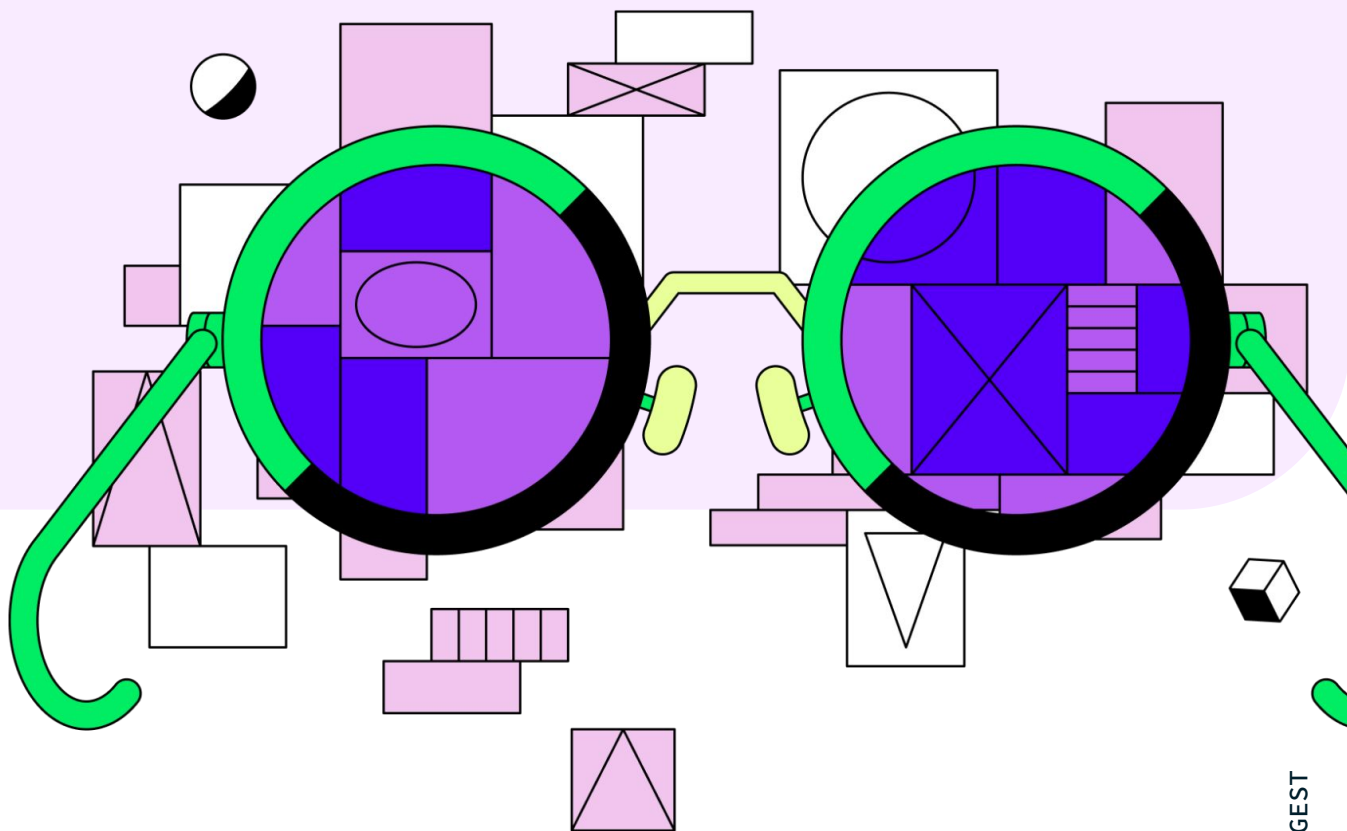
As organizations strive to lay the foundation for future success, embracing modern data architectures is essential for driving digital transformation. Emerging trends such as data mesh and cloud-first strategies enable businesses to decentralize data management, facilitating real-time access and empowering teams to make informed decisions swiftly. Additionally, the integration of artificial intelligence and machine learning enhances data processing capabilities, allowing organizations to harness insights more effectively. By prioritizing a robust data foundation, companies can adapt to evolving market demands and unlock new opportunities for growth in an increasingly data-driven landscape.

Adapt to evolving  
market demands  
and unlock new  
opportunities for  
growth

# Expert insights: the future of data-driven industries

We spoke with industry leaders to explore how data and database technologies are shaping the future of their sectors.

In this series, two experts share their vision for the evolving role of data in their fields, offering valuable insights on the trends driving innovation and transformation.



## Financial Services Industry (FSI)



Financial services are inherently a digital business, and the future lies in fully maximising advanced forms of AI, such as Autonomous Agents—where tapping into both structured data like transactions and customer profiles, and unstructured data from formats such as video, audio, and documents, along with multi-modal financial data types like geospatial insights, knowledge graphs, and vectors, to drive smarter, more intuitive, and fully digital financial solutions.

**Wei You Pan, Global Director,**  
Financial Industry Solutions, MongoDB

With 20+ years of experience leading digital transformation initiatives in the financial sector, Wei You Pan helps global financial institutions modernize their data architectures to drive agility, security, and innovation.





## Manufacturing & Motion

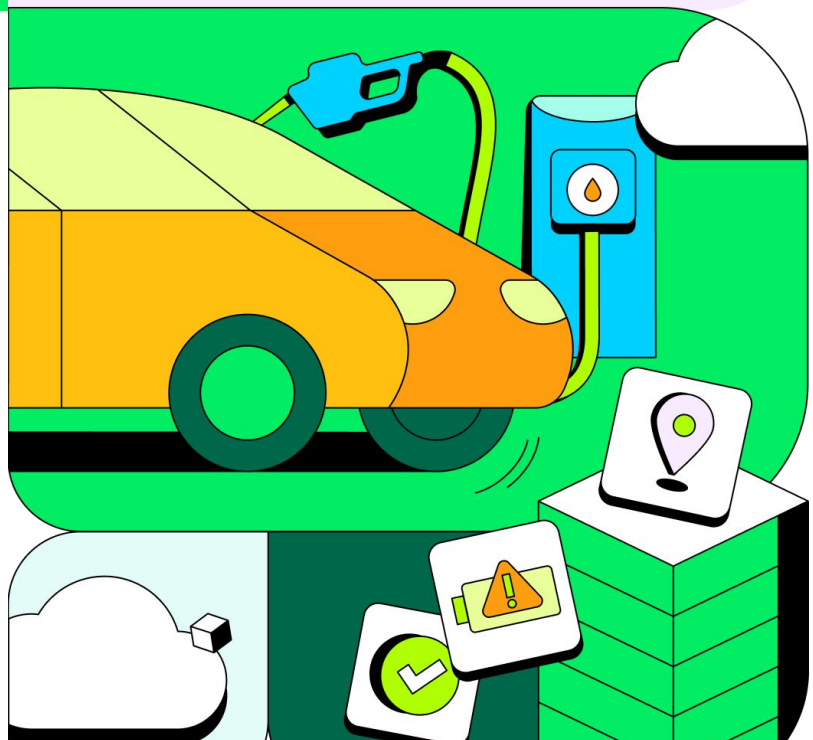
With over 20 years of experience in manufacturing and engineering technology, Raphael has held multiple CTO roles, specializing in Industry 4.0, IIoT, and data-driven digital transformation. His background spans from designing precision tools to leading innovations in automation and smart manufacturing.



“

The future of manufacturing is no longer just about machines and production lines; it's about how data, AI, and automation converge to create a smarter, more agile, and resilient ECO System. The manufacturing industry is speeding up the data revolution towards Industry 5.0, driven by the need for real-time insights, predictive analytics, and AI-driven automation. Cloud-based databases, IoT sensors, and digital twins enhance efficiency, reduce downtime, and optimize supply chains. Generative AI starts playing a critical role for accelerating product design, automating complex workflows, streamlining quality control, and enabling intelligent decision-making.

**Raphael Schor, Industry Principal,  
Manufacturing & Mobility, MongoDB**





# Foundation for tomorrow: Key takeaways

As we conclude this inaugural issue of *Database Digest*, the journey through modern data architecture has revealed how organizations are laying the foundations for future success in an increasingly data-driven world. From MongoDB's evolution into a comprehensive enterprise platform to the transformative impact of AI agents and vector search capabilities, we've witnessed how businesses across industries are reimagining their data infrastructure to drive innovation and create more intelligent, responsive systems. The convergence of flexible document models, advanced AI integration, and cloud-native capabilities is enabling companies to build applications that were never before possible, as demonstrated by success stories from Swisscom, Bendigo Bank, and Questflow.

The rise of agentic AI systems and sophisticated RAG implementations marks a significant shift in how organizations interact with and extract value from their data, while the modernization of legacy systems through processes like the Modernization Factory is helping enterprises reduce costs and accelerate their digital transformation journey. Industry leaders across manufacturing, financial services, and healthcare sectors emphasize how this transformation extends beyond traditional boundaries, as organizations evolve toward more integrated, AI-driven operations that combine real-time analytics, IoT sensors, and digital twins to enhance efficiency and innovation.

Looking ahead, the success of modern enterprises will increasingly depend on their ability to unify their data foundation, leverage AI capabilities, and embrace event-driven architectures that enable real-time insights and decision-making. The future belongs to organizations that can build flexible, scalable data architectures by prioritizing data governance, embracing cloud-native technologies, and fostering a culture of continuous learning. The journey to digital transformation is about creating a foundation that empowers organizations to adapt, innovate, and thrive—delivering tangible value today and tomorrow.

## Additional resources

Ready to dive deeper? Whether you're looking to innovate with AI, kickstart your next project, or learn from industry leaders, we've curated these essential resources to help you build your foundation for success:

### **Innovate with AI: The Future Enterprise**

A comprehensive 200+ page e-book exploring how leading industries are transforming their businesses with AI and MongoDB Atlas.

### **Solutions Library**

Browse our collection of ready-to-use solutions and detailed use cases to accelerate your development journey.

### **Customer Success Stories**

See the leading organizations creating applications never before possible with.

### **AI Learning Hub**

A curated list of resources for developers across skill levels. Use it to guide your efforts in building AI applications in MongoDB.

### **MongoDB Skill Badges**

MongoDB Skill Badges are free, focused credentials designed to help you quickly learn and validate specific MongoDB skills.

## Want to learn more?

**CONTACT US**