



MongoDB Atlas Now Delivers Industry-Leading Context Retrieval with Precision Accuracy

August 13, 2026

Automated Embeddings on Atlas powered by Voyage AI, the Atlas Embedding and Reranking API, voyage-code-4, and vector search in Atlas Stream Processing give AI applications and agents real-time context from live operational data for precision retrieval

SAN FRANCISCO, Aug. 13, 2026 /PRNewswire/ -- Retrieval accuracy determines what an AI application or Agent decides and what it costs to host. Today, **MongoDB, Inc.** (NASDAQ: MDB) announced at MongoDB.local Build Fest a set of capabilities that add benchmark-leading retrieval directly into MongoDB's intelligent data platform to solve this exact challenge. Today's capabilities start with a new model in voyage-code-4, purpose-built for agentic code retrieval. That model, along with the full set of Voyage AI embedding and reranking models, is now available through a standalone API to power any application. MongoDB's Voyage AI embedding models hold the top spot on the [Retrieval Embedding Benchmark](#) (RTEB).



Historically, developers powering advanced retrieval use cases have had to pull data out of operational systems, manage complex and brittle embedding pipelines, and keep vector stores in sync. With Automated Embedding in MongoDB Atlas, powered by Voyage AI models, developers easily insert and retrieve data. Atlas handles embedding and indexing automatically behind the scenes, with no synchronization or embedding pipelines to manage.

The Financial Times (FT) was dealing with this exact complexity as they worked to make its journalism easy to find for readers. Search had splintered across the FT's teams and products, so it consolidated onto the MongoDB platform. With Automated Embeddings in MongoDB Atlas powered by Voyage AI, the FT is improving the accuracy of its AI-powered semantic search while holding down retrieval costs.

"Our job is to make the FT's journalism fast and easy to reach, however our readers come to it. With Automated Embedding and Voyage AI models on Atlas, we've improved retrieval accuracy while keeping costs in check across more than 100,000 searches a day, and being able to test and balance models lets us tune quality against cost as we go. With less infrastructure to run, the team can spend more of its time on the reading experience our subscribers rely on," said Elitsa Pavlova, Principal Engineer CM Platform, at Financial Times.

Eve, a legal AI platform, was also running into complexity and cost issues. They are now using the Atlas Embedding and Reranking API to surface the most relevant material as it works across the life of a case.

"In legal AI, retrieval quality is foundational—the right evidence has to surface at the right moment. MongoDB's Atlas Embedding and Reranking API gives us a promising way to improve relevance directly in the RAG layer, while simplifying the infrastructure needed to build and evolve these experiences," said Urvesh Patel, Staff AI Engineer at Eve.

MongoDB is the memory and context layer that agents are built on

On MongoDB, high-precision retrieval runs in the same platform as real-time operational data, not bolted on as a separate system that adds cost and synchronization overhead. An agent searches live data for context, including records written seconds earlier, so what it retrieves reflects the current state of the business rather than a stale copy.

Four new capabilities give teams accuracy in production while reducing complexity and cost:

- **Fresh and accurate context for agents:** AI applications and agents can only make good decisions with accurate, reliable, up-to-date context. Automated Embeddings in MongoDB Atlas powered by Voyage AI keep that context current. Users set a Voyage AI embedding model for their search index, and Atlas embeds new documents as they are written and re-embeds existing ones when they change, with nothing to run or schedule.
- **One endpoint for benchmark-leading embeddings:** The Atlas Embedding and Reranking API gives any application, even ones running outside MongoDB, direct access to MongoDB's top-ranked embedding and reranking models through Atlas.
- **Accurate code retrieval with lower cost:** General-purpose embedding models were not built for codebases, and it shows in retrieval accuracy. voyage-code-4 is a retrieval model tuned specifically for coding agents, matching information to code with higher precision and lower cost than previous models.
- **Retrieval for streaming data:** Agents acting on live events should not work from worse context than agents reading from data at rest. Vector search in Atlas Stream Processing brings that same retrieval accuracy to data in motion.

"Too many organizations are running AI in production with an operational database, a vector store, a search engine, and embedding and reranking models, all from different vendors, bolted together instead of built for it," said Jim Scharf, Chief Technology Officer, at MongoDB. "That's where stale data and errors creep in, and it's usually where teams spend their time babysitting instead of building. Agents raise the bar. They need to retrieve live context continuously and cannot wait on overnight batch jobs. MongoDB was built as an operational platform from the start, so retrieval and memory run on the same live data, nothing to sync, and agents act on what's happening instantly."

These new capabilities are generally available today; get started for free at mongodb.com/atlas.

MongoDB also announced at Build Fest new integrations that put MongoDB inside the AI tools developers build with—across Claude, Claude Code, ChatGPT, Codex, Grok Build, and Devin—along with the MongoDB Atlas Managed MCP Server.

About MongoDB

Headquartered in New York, MongoDB's mission is to empower innovators to create, transform, and disrupt industries with software. MongoDB's unified database platform was built to power the next generation of applications, and MongoDB is the most widely available, globally distributed database on the market. With integrated capabilities for operational data, search, real-time analytics, and AI-powered data retrieval, MongoDB helps organizations everywhere move faster, innovate more efficiently, and simplify complex architectures. Millions of developers and more than 67,000 customers across industries—including ~75% of the Fortune 100—rely on MongoDB for their most important applications. To learn more, visit mongodb.com.

Forward-Looking Statements

This press release includes certain "forward-looking statements" within the meaning of Section 27A of the Securities Act of 1933, as amended, or the Securities Act, and Section 21E of the Securities Exchange Act of 1934, as amended, including new capabilities announced at MongoDB.local Build Fest 2026. These forward-looking statements reflect our current views about our plans, intentions, expectations, strategies and prospects, which are based on the information currently available to us and on assumptions we have made. Actual results may differ materially from those described in the forward-looking statements and are subject to a variety of assumptions, uncertainties, risks and factors that are beyond our control including those risks detailed under the caption "Risk Factors" and elsewhere in our Securities and Exchange Commission filings and reports. Except as required by law, we undertake no duty or obligation to update any forward-looking statements contained in this release as a result of new information, future events, changes in expectations or otherwise.

Contacts

Investors: ir@mongodb.com

Media: press@mongodb.com

 View original content to download multimedia: <https://www.prnewswire.com/news-releases/mongodb-atlas-now-delivers-industry-leading-context-retrieval-with-precision-accuracy-302850848.html>

SOURCE MongoDB, Inc.