

Vectara and Confluent Deliver Scalable RAG Solutions on Real-time Data

Empowering Real-Time Data with AI-Driven RAG Capabilities

This strategic partnership enables Confluent customers to seamlessly integrate **Retrieval-Augmented Generation (RAG)** into their data workflows—unlocking new levels of efficiency, intelligence, and scalability.

The Power of Partnership

Together, Vectara and Confluent are helping enterprises unlock the full potential of their data by combining:

- Vectara's simplified RAG platform
- Confluent's real-time data streaming infrastructure

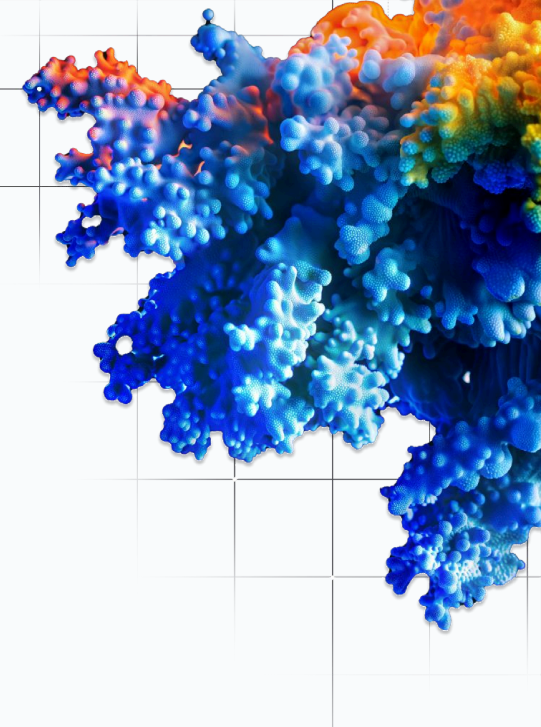
This collaboration enhances **Total Cost of Ownership (TCO)** and makes cutting-edge AI capabilities more accessible for organizations using Confluent's platform.

Real-Time, Trusted AI for Smarter Decisions

Confluent and Vectara together empower enterprises to make decisions at the speed of data. While Confluent streams continuous real-time events—from customer interactions to market fluctuations—Vectara retrieves and synthesizes these signals into accurate, domain-specific insights using advanced Retrieval Augmented Generation processes focused on high-accuracy and low hallucination. The result is AI that delivers explainable, up-to-the-second insights that business leaders can trust and act on with confidence.

Enterprise-Grade AI: Secure, Compliant, and Risk-Aware



Security and trust are at the core of the Confluent-Vectara partnership. Confluent ensures secure data streaming through real-time pipelines with enterprise-grade compliance and governance. Meanwhile, Vectara grounds AI outputs in enterprise data without ever training on the underlying sensitive information, dramatically reducing the risk of data leakage or hallucinated outputs. Together, they offer a future-proof solution that meets the highest standards of enterprise security and regulatory adherence.



Vectara and Confluent Deliver Scalable RAG Solutions on Real-time Data

Faster AI Deployment, Better Customer Experiences

From data ingestion to intelligent response, Confluent and Vectara eliminate friction in deploying AI solutions. Confluent simplifies real-time data integration across sources, while Vectara’s fully managed Retrieval-Augmented Generation (RAG) platform accelerates AI deployment from months to days. This unified approach enables organizations to deliver real-time, hyper-personalized experiences—supported by Vectara’s hybrid semantic-keyword search, citation transparency, and multilingual capabilities—ensuring every interaction is relevant, responsive, and reliably AI-powered.

 CONFLUENT	 vectara	Combined Outcome
Real-time data pipelines with Kafka Streams for immediate data ingestion	Real-time semantic and hybrid search ensuring fast retrieval	AI-driven responses grounded in timely, real-time enterprise data
Enterprise-grade event-streaming infrastructure	Best-in-class, multilingual semantic retrieval and accurate RAG generation	AI systems capable of supporting global, real-time use cases without latency or accuracy trade-offs
Streamlined and scalable data orchestration	Serverless, managed RAG platform eliminating the need for in-house ML operations	Significantly reduced operational complexity and faster deployments at enterprise scale

Conclusion

From data ingestion to intelligent response, Confluent and Vectara eliminate friction in deploying AI solutions. Confluent simplifies real-time data integration across sources, while Vectara’s fully managed Retrieval-Augmented Generation (RAG) platform accelerates AI deployment from months to days. This unified approach enables organizations to deliver real-time, hyper-personalized experiences—supported by Vectara’s hybrid semantic-keyword search, citation transparency, and multilingual capabilities—ensuring every interaction is relevant, responsive, and reliably AI-powered.