



# Supercharge your AI with OpenShift AI and Redis

Virgilio Sanz  
Octubre de 2025

**Not** investing in AI isn't an option anymore.

# Companies are using AI for a few **top use cases.**

Chatbots



AI agents



AI gateways



Search



Personalization & recommendations



Fraud & anomaly detection



# But deploying GenAI apps is hard.

Slow LLMs

**>100 ms**

for LLM responses,  
causing users to churn.

Inaccurate  
responses

**72%**

of AI/ML execs cite  
hallucinations as their top  
challenge with GenAI.

Changing tech stack

**>2000**

new GenAI tools disrupted  
the market in 2025 —and  
2026 will top that.

Redis is the **working memory layer**  
for the new AI tech stack.

# Redis leads in AI agent data storage and is the fastest -growing major database

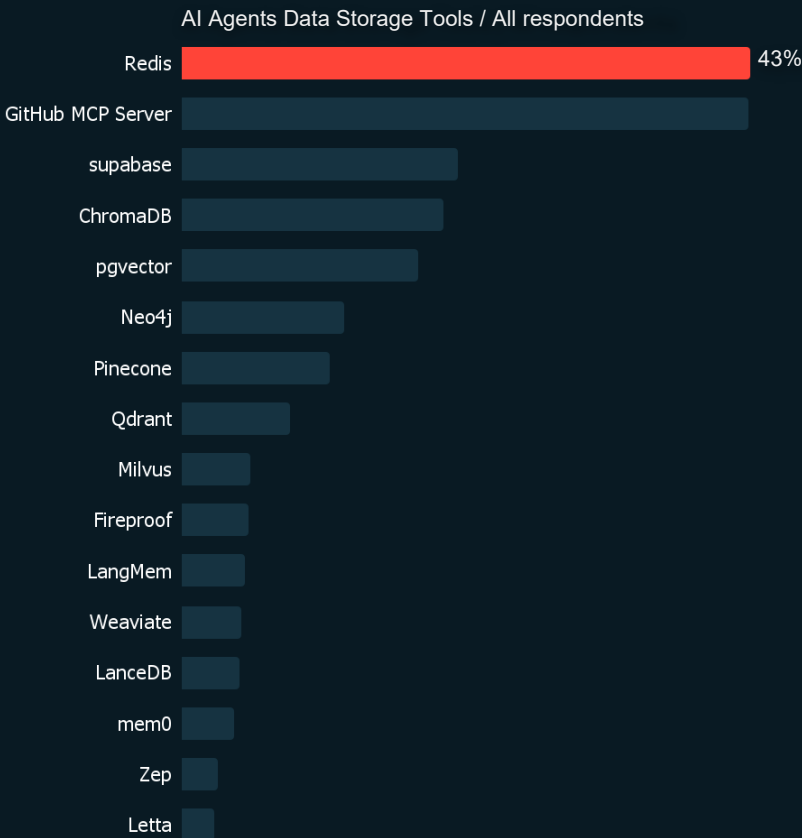
# #1

data storage choice for AI agents (43%), ahead of GitHub MCP Server

# +8%

growth in database usage year - over - year

## Top NoSQL database for AI developers



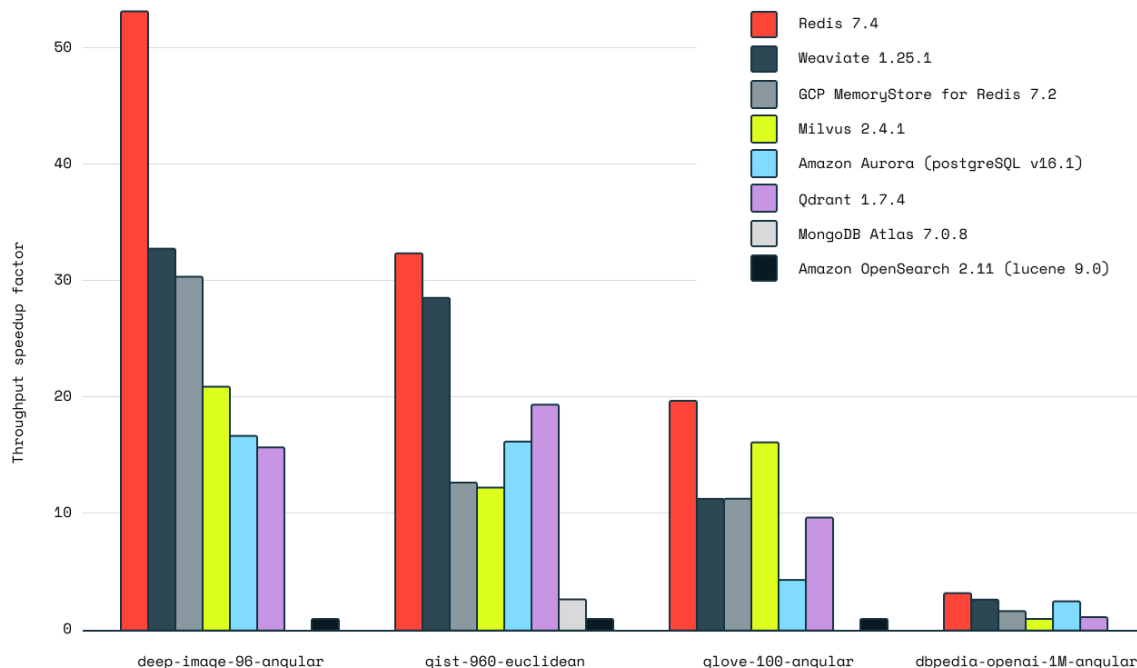
# Vector database

We offer a fully featured platform for vector and full -text search that's configurable with top search algorithms, embeddings support, and reranking algorithms.

→ **62% faster** than the second-ranked vector db

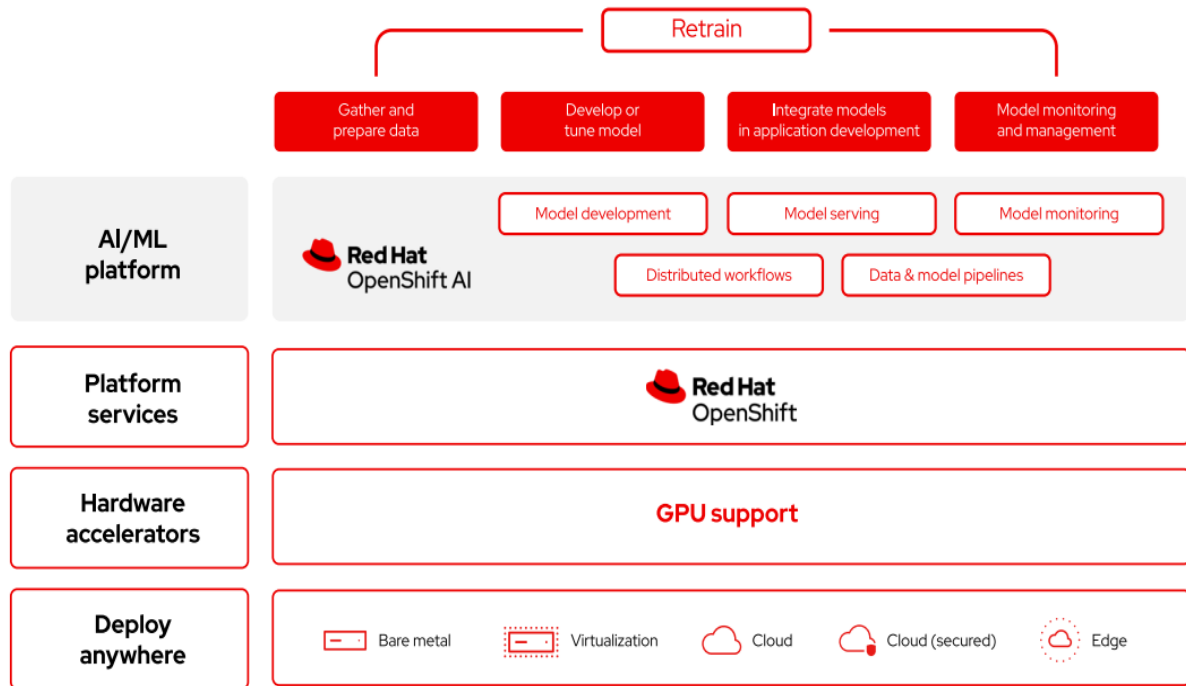
**>1 billion vectors** indexed and searched

## Fastest across third party benchmarks

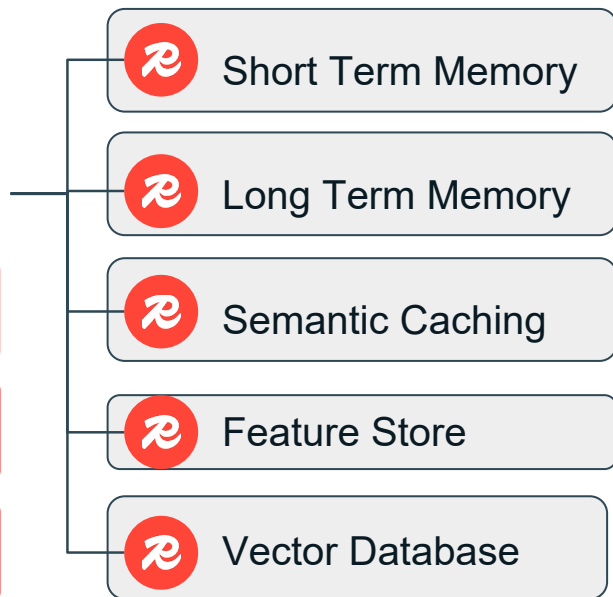


# Red Hat OpenShift AI

Red Hat's AI/ML platform



## Redis Enterprise Real Time Data Layer



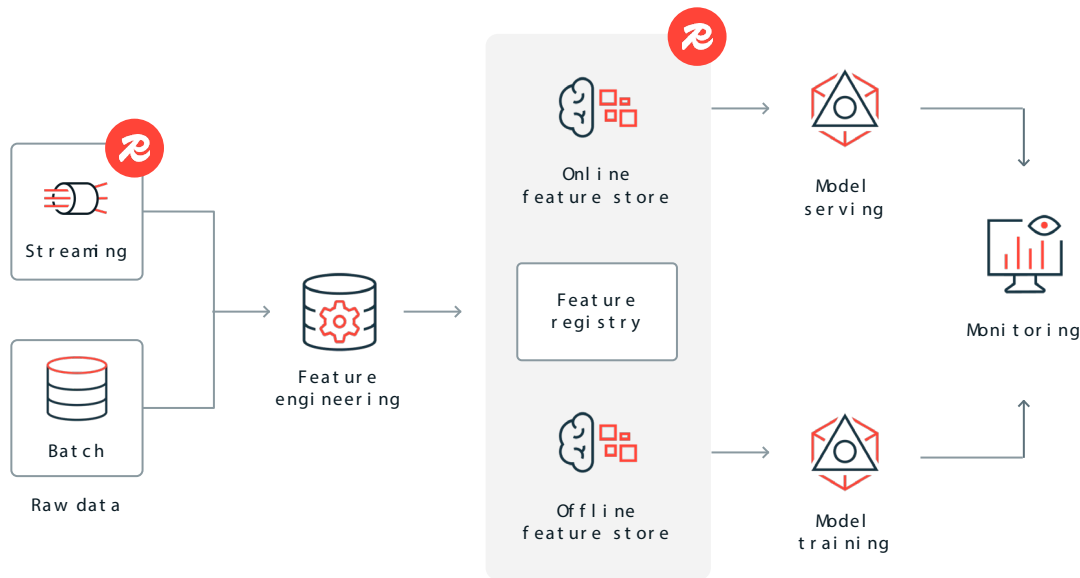


# Feature store

We store ML features for fast data retrieval to power timely predictions. Our feature store connects seamlessly with offline feature stores like Tecton and Feast at the scale companies need for instant decisions worldwide.

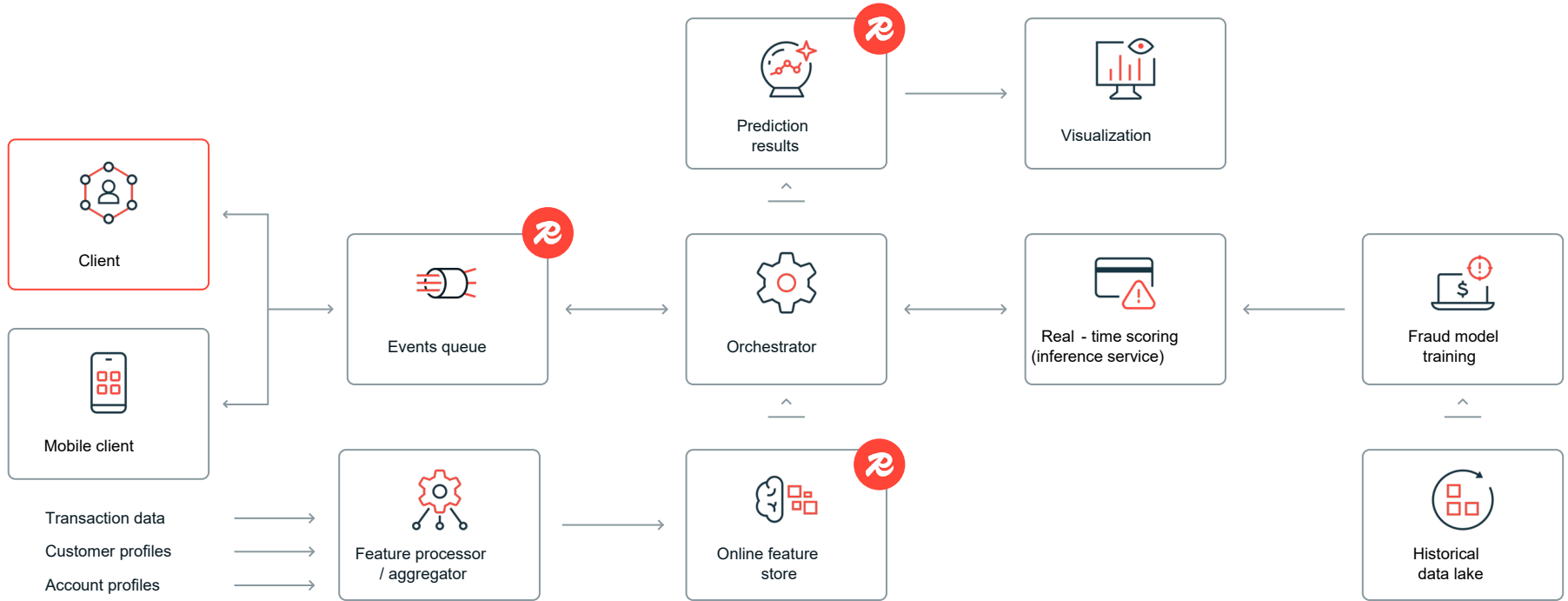
→ **90% improvement** in response times serving ML models

**<1 millisecond** per Read operation



# Anomaly & fraud detection

Stop the bad guys in their tracks.



## ● AGENTS

AI agents do the work for you. Redis makes them better.

83% of tech leaders currently or plan to use agents

### What you're trying to do

Build AI agents or agentic systems that complete tasks autonomously.

Let LLMs plan and execute the best way to fulfill a user request.

Use LLMs to evaluate output from GenAI or other systems before proceeding.

### How we help

Short-term memory speeds up agents and processes to provide more accurate answers faster.

Long-term memory and session management recall relevant interactions to provide better outputs.

Semantic caching stores similar questions and answers to speed up responses and cut down costly LLM calls.

We integrate with top GenAI ecosystem tools so you can build more reliable systems faster.

Our leading vector database retrieves info faster, boosting agents that need multiple data retrieval calls.

### How it works for businesses like yours



Rox automates customer research for sales teams and customizes messaging for each customer. Rox uses Redis for a fast and user -friendly front end.



Kore.ai provides a platform for chatbots and AI agents. They chose Redis for our vector search speed and ecosystem integrations.



Relevance AI needed to streamline and scale its AI agents to help workers work faster. Redis brought their query speeds down from 2 seconds to 10 ms.

# Our customers build GenAI apps on Redis every day.



Thank You



Know More