

Lowering Total Cost of Ownership (TCO) for Modern NoSQL Databases in AI-Driven Enterprises



Contents

EXECUTIVE SUMMARY	3
INTRODUCTION: WHY NOSQL MATTERS FOR AI-DRIVEN ENTERPRISES	3
MAIN CHALLENGES ENTERPRISES FACE IN DATABASE COSTS	3
HOW LOWERING TCO FREES FUNDS FOR AI WORKLOADS	4
COMPARATIVE ANALYSIS: COUCHBASE VS. MONGODB VS. AWS DYNAMODB	4
THE ADVANTAGES OF COUCHBASE'S DATABASE PLATFORM	5
CONCLUSION AND RECOMMENDATIONS	6



EXECUTIVE SUMMARY

This whitepaper explores how enterprises can cut database TCO while supporting AI applications and agents. It focuses on deployment, licensing, scaling, and management for global, enterprise-grade applications. By examining key cost drivers and comparing leading solutions – Couchbase, MongoDB, and AWS DynamoDB – we highlight how Couchbase’s platform delivers superior efficiency, flexibility, and performance. Lowering TCO not only streamlines operations, but also reallocates funds to fuel AI workloads, such as traditional and semantic caching, enabling CIOs and architects to build scalable, adaptive systems without compromising innovation.

INTRODUCTION: WHY NOSQL MATTERS FOR AI-DRIVEN ENTERPRISES

CIOs and enterprise architects managing global deployments face increasing pressure to balance control and convenience as AI adoption accelerates. Modern applications demand databases that can handle unstructured data, support vector embeddings for semantic search, and process information in real time for intelligent agents. Meanwhile, rising AI compute costs are forcing organizations to seek more cost-effective infrastructure. NoSQL document-oriented and key-value stores offer the agility and scalability needed to meet these demands. They combine flexible data models, distributed performance, and deployment options that span self-managed and cloud-managed environments.

MAIN CHALLENGES ENTERPRISES FACE IN DATABASE COSTS

Enterprises deploying enterprise-grade, global applications face multifaceted cost challenges that can hinder scalability and innovation. These include:

- **Training and running AI models** requires significant compute and storage resources. If the database layer has an inefficient use of nodes, slower queries, or redundant systems, those costs compound. Every inefficiency in storage or retrieval increases the price tag of already expensive GPU and cloud resources.
- Many enterprises maintain **multiple databases** for key-value, search, analytics, mobile, text, time series, and now vector workloads. This overlap not only inflates licensing fees, but also adds integration and operational overhead. A database that consolidates functions reduces duplication, simplifies architecture, and lowers both direct and hidden costs.
- When the database can process queries faster and with lower latency (especially under AI workloads like vector search or semantic caching), it reduces the number of nodes and servers needed. **Performance** directly translates into cost savings by lowering hardware footprints and cloud consumption.



- Beyond licenses and hardware, enterprises pay for administration, monitoring, downtime, compliance audits, and multi-region replication. These **“hidden” operational costs** often exceed the visible line items on the invoice. Poor database design or limited automation drives up labor costs and makes scaling globally far more expensive than it appears at first glance.

These challenges often lead to data sprawl, inefficient resource utilization, and budgets strained by siloed systems, making TCO optimization essential for sustainable growth.

HOW LOWERING TCO FREES FUNDS FOR AI WORKLOADS

Reducing database TCO creates a direct pathway to funding AI initiatives by reallocating savings from infrastructure and operations. Key mechanisms include:

- **Efficiency gains** – Consolidating multiple database functions (e.g., key-value, document, and search) into a single platform reduces hardware footprints and licensing fees, yielding **savings of up to 50% in some cases**.
- **Resource reallocation** – Lower operations costs free IT teams to focus on AI development, while budget savings cover compute-intensive tasks like model training on GPUs.
- **AI-specific optimizations** – Implementing traditional caching (e.g., in-memory stores for frequent queries) and semantic caching (storing context-aware responses to reduce redundant AI calls) can dramatically reduce inference costs. For instance, semantic caching in vector-enabled databases minimizes API calls to large language models, lowering latency and expenses.
- **Case for investment** – Enterprises can redirect TCO savings to AI agents for real-time decision-making or adaptive applications, enhancing return on investment (ROI). This shift supports global scalability, where cost-efficient replication funds edge AI deployments.

COMPARATIVE ANALYSIS: COUCHBASE VS. MONGODB VS. AWS DYNAMODB

This section provides a detailed comparison of leading database solutions across the main challenge areas, drawing on **performance benchmarks and TCO studies**. Couchbase emerges as a leader due to its multi-model capabilities, memory-first architecture, and flexible deployment options, which deliver better price-performance for AI workloads.



TCO Factors	Couchbase	MongoDB	AWS DynamoDB
Deployment Needs	Supports both self-managed and cloud for hybrid control.	Offers both, but hybrid setups are harder.	Fully managed only, no self-managed option.
Licensing Costs	Usage-based with clear pricing and fewer add-ons.	Cluster or usage-based, extra costs for features.	Pay-per-request or capacity; bills can spike.
Performance, Caching, and Scaling Costs	Scales efficiently with fewer nodes using memory-first design.	Scales via sharding but often needs more nodes.	Auto-scales easily but throughput costs rise fast.
Management Overhead	Integrated services and self-healing reduce ops work.	Needs more manual tuning; Atlas eases some burden.	Fully managed but limited customization and extra monitoring costs.
Global Application Support	Cross data center replication (XDCR) enables low-cost, multi-region replication.	Global clusters work but cross-region costs are high.	Global tables replicate but add egress and consistency fees.
TCO for AI Workloads	Lower TCO with native caching and vector support.	Solid AI support but higher node count inflates cost.	Cheap for key-value AI, but lacks multi-model features.

THE ADVANTAGES OF COUCHBASE'S DATABASE PLATFORM

Couchbase is the optimal choice for CIOs and architects building AI-centric, global applications:

- **Multipurpose flexibility** – Combines document, key-value, graph traversal, mobile, analytics, and vector capabilities in one platform, reducing the need for multiple databases and slashing TCO through consolidation.
- **Performance and efficiency** – Memory-first architecture delivers sub-millisecond latencies for AI queries, outperforming competitors in benchmarks while using fewer resources – enabling cost savings of 50% or more.
- **AI-ready features** – Built-in vector search and semantic caching support AI agents efficiently, minimizing compute costs and accelerating development.
- **Deployment versatility** – Seamless self-managed and cloud-managed options provide the control needed for enterprise-grade global apps, without vendor lock-in.



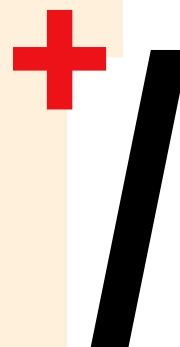
- **Proven TCO reductions** – Customers report dramatic savings, with independent studies showing Couchbase’s edge in speed, functionality, and cost over MongoDB and DynamoDB.
- **Future-proofing** – Enhances innovation by freeing budgets for AI workloads, ensuring scalability as enterprises expand globally.

CONCLUSION AND RECOMMENDATIONS

Lowering TCO in NoSQL databases is not just a cost-cutting exercise – it’s a strategic enabler for AI-driven transformation. By addressing deployment, licensing, scaling, and management challenges, enterprises can redirect resources to high-value AI applications. Couchbase’s platform offers unmatched advantages in this space, providing the performance, flexibility, and efficiency needed for tomorrow’s global architectures.

Next steps, contact the Couchbase team to:

1. Discuss how Couchbase Capella™ DBaaS can lower your TCO.
2. Help create a pilot semantic caching or vector search PoC with an AI workload.
3. Understand scaling costs compared to MongoDB and DynamoDB with your real data.





Modern customer experiences need a flexible database platform that can power applications spanning from cloud to edge and everything in between. Couchbase's mission is to simplify how developers and architects develop, deploy and run modern applications wherever they are. We have reimaged the database with our fast, flexible and affordable cloud database platform Capella, allowing organizations to quickly build applications that deliver premium experiences to their customers – all with best-in-class price performance. More than 30% of the Fortune 100 trust Couchbase to power their modern applications.

For more information, visit www.couchbase.com and follow us on X (formerly Twitter) @couchbase.

© 2025 Couchbase. All rights reserved.

