

# Enhancing AI Application Development with Oracle Database 23ai's AI Vector Search and Exadata X10M

## AUTHOR

**Steven Dickens**  
Chief Technology Advisor | The Futurum Group

**JUNE 2024**

IN PARTNERSHIP WITH

**ORACLE®**



## Executive Summary

The rapid advancement of AI technologies is set to revolutionize various industries, leading to an unprecedented influx of AI-driven applications within enterprises over the next few years. This surge will place significant demands on developers, who must continually adapt to the evolving landscape. The emerging multimodal AI applications, which integrate various data types and sources, will further compound these challenges, necessitating sophisticated and scalable database solutions to support their development and deployment.

This research brief delves into the transformative impact of advanced database solutions that simplify AI application development while enhancing system scalability and performance. As enterprises brace for the wave of AI applications, developers will be tasked with managing increasingly complex data environments. Advanced AI capabilities, such as vector search, provide developers with powerful tools to efficiently handle the diverse workloads with voluminous amounts of data. These tools enable developers to focus on innovation, leveraging familiar yet powerful resources to streamline the integration of complex data types.

For IT management, the key to successfully navigating this AI-driven future lies in robust and scalable infrastructure. High-performance systems are essential to support the surge of AI applications without sacrificing speed or reliability. As enterprises expand their AI initiatives, maintaining optimal performance levels will be critical. Advanced infrastructure solutions ensure that businesses can scale effectively, accommodating the growing demands of AI without compromising operational efficiency.

By addressing the dual needs of developers and IT management, advanced database solutions facilitate the seamless integration of AI into enterprise operations. These solutions offer a balanced approach, providing developers with the tools needed to innovate while ensuring that IT infrastructure can scale and perform under increasing loads. This synergy is essential for enterprises to remain competitive and responsive in an increasingly AI-driven business environment.

In conclusion, the anticipated influx of AI applications will reshape enterprise landscapes, demanding more from developers and IT management alike. Advanced database solutions will play a pivotal role in this transformation, enabling enterprises to harness the full potential of AI. By providing the necessary tools and infrastructure, these solutions empower businesses to navigate the complexities of AI integration, ensuring sustained innovation and operational excellence in a rapidly evolving technological landscape.





## Introduction

Oracle Database 23ai can play a strategic role in enhancing an enterprise's AI strategy through its new AI Vector Search capabilities. This new functionality aligns with Oracle's vision for technological innovation, meeting the needs of both developers and IT management. By integrating AI Vector Search within a converged database and making it available through SQL, Oracle simplifies the development process, allowing developers to leverage their existing SQL knowledge. This approach eliminates the need for standalone vector databases and the associated management complexities.

The technical synergy between Oracle's database technologies and the cutting-edge hardware features of Exadata X10M platforms, powered by 4<sup>th</sup> generation AMD EPYC processors in both database servers and storage servers, enables more parallel searches, faster results, and timely update of vector indexes. This synergy will be discussed in detail, highlighting how more cores in database servers and storage servers enhance performance and scalability.

## Oracle 23ai - Solution Overview

Oracle Database 23ai, with its AI Vector Search capabilities, enables efficient application development. The integration of vector data types, vector indexes, and vector search within the database allows developers to manage and query vector data using standard SQL, a familiar tool for most developers. This approach simplifies the management of vector data by eliminating the need for a separate vector database and the complexities associated with data replication, consistency management, integration, protection, and security they introduce.

Exadata's architecture supports AI Vector Search use cases through scale-out configurations, exceptional memory bandwidth, and smart storage solutions. These features enhance the performance and scalability of AI applications. Exadata's scale-up/scale-out capabilities and the use of AMD EPYC processors with 96 processor cores per socket support complex AI operations and seamless scalability, making it a well-suited platform for enterprise AI applications. The extreme scalability of Exadata X10M platforms, with the ability to support thousands of AMD EPYC processor cores and petabytes of storage, means that neither developers nor IT organizations should be concerned about hitting performance roadblocks as the demand grows for data-intensive GenAI applications that leverage their private business data.

# Technical Advantages for Developers and IT Management

Oracle's AI Vector Search capabilities offer significant advantages for developers and IT management. Developers can leverage these capabilities without needing to be data scientists. Key benefits include:

- No need to specify data type and vector size.
- Simplified searches without low-level parameter specifications.
- Integration of vector search results with SQL queries of business data.
- Direct application integration using standard SQL or AI frameworks like LangChain.

These features make AI Vector Search accessible to a broader range of programmers, not just those with deep expertise in vector databases. Oracle has taken a hard problem and made it simple by making vector search and integration with business data easy.

For IT management, running AI Vector Search on Exadata X10M platforms optimize overall system efficiency, reduces latency, and provides high performance, scalability, and security. Exadata platforms can transparently offload critical AI Vector Search capabilities to smart storage servers, with a single search being automatically parallelized across all of them. As a result, Exadata X10M platforms support multi-user environments with the ability to offload thousands of concurrent vector searches to storage servers and accelerate queries by up to 30x. This frees database server cores to support more users running diverse workloads.

Several development scenarios illustrate how developers can easily access AI Vector Search capabilities:

**1. SQL Developers:** AI is nothing without data. By integrating vector search with SQL queries, developers can enrich their applications with AI-driven insights. Given that most developers already know SQL, the fact that vector search is done within the Oracle Database that holds their private business data means that combining vector search with other database queries can be done inside the database instead of using application code – reducing development time and maintenance requirements.

**2. Embedded AI in Applications:** Developers can add Retrieval-Augmented Generation (RAG) functionality to existing applications using SQL or REST APIs to increase the accuracy and relevance of Large Language Model (LLM) results. Running semantic search on private business data and combining those results with users' initial questions provide LLMs with additional context and use-case-specific information that it would not have been trained on.

**3. AI Pipelines:** Creating AI pipelines using LangChain lets developers leverage Oracle's AI Vector Search capabilities within their workflows. This popular method of developing new GenAI workloads lets organizations quickly start developing new applications while continuing to use existing business data.

These scenarios underscore the practical value and versatility of Oracle's AI Vector Search in real-world applications.

# Recommendations

To effectively integrate AI Vector Search from Oracle Database 23ai into existing IT infrastructures, organizations should consider the following guidelines:

**Data Migration:** Plan for data migration strategies to ensure seamless integration with existing systems. This involves assessing current data structures, identifying data that needs to be migrated, and establishing a clear migration path. Utilizing Oracle's migration tools can facilitate this process, ensuring data integrity and minimizing downtime.

**Cost-Efficiency:** Evaluate the cost-efficiency of transitioning to Oracle's solutions. If you're already an Oracle Database user, the long-term benefits of simplified management and enhanced performance provided by Oracle Database 23ai and Exadata X10M platforms can reduce operational costs by eliminating the need for standalone vector databases and associated management complexities. If you're not an Oracle Database user, consider the financial benefits available from simplified application development and efficient operations provided by running AI Vector Search in your data center or Oracle's cloud. Perform a cost-benefit analysis to understand the financial impact and potential ROI.

**System Compatibility:** Ensure compatibility with existing IT infrastructures to minimize disruptions and optimize performance. This involves checking hardware and software requirements, verifying network compatibility, and ensuring that all system components can seamlessly integrate with Oracle Database 23ai. If you want to start using AI Vector Search but your source databases are running on Oracle Database 19c, Oracle GoldenGate 23ai can easily replicate real-time data updates to an Oracle Database 23ai instance where vector embeddings and searches can take place. And, if you're using Exadata, you can easily run both on the same system. Engaging with Oracle's consulting services can provide insights and assistance in achieving optimal compatibility.

**Security and Compliance:** Implement robust security measures to protect sensitive data and ensure compliance with industry regulations. Oracle Database 23ai offers advanced security features, including encryption, access controls, auditing capabilities, and a new in-database SQL firewall. Leverage these features to safeguard data and maintain compliance with regulatory standards.

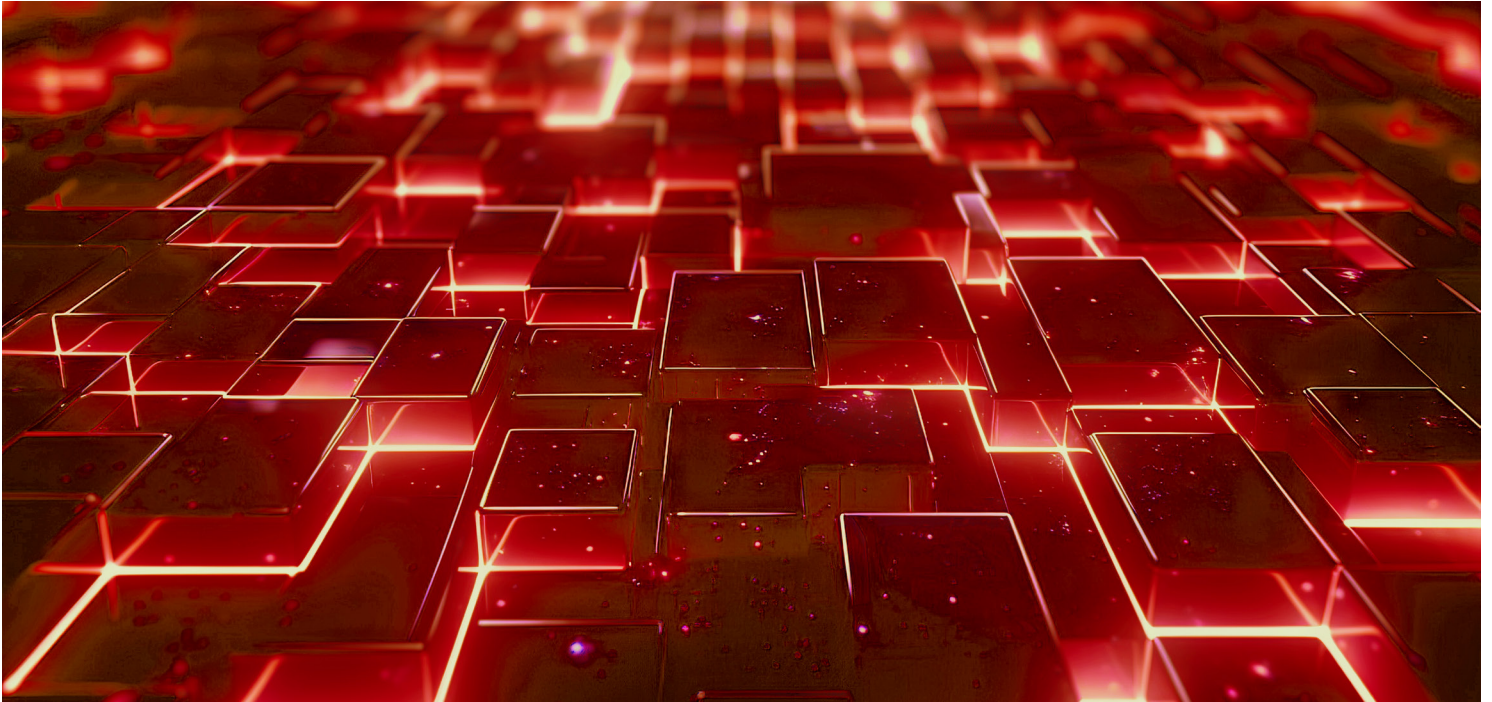
**Performance Monitoring and Optimization:** Continuously monitor system performance and optimize configurations to ensure efficient operation. Oracle provides tools for performance monitoring and tuning, enabling IT management to identify bottlenecks and make necessary adjustments. Regular performance reviews can help maintain high system efficiency and reliability.

For developers, leveraging Oracle's advanced features for AI-driven projects is crucial. Here are specific recommendations to maximize the benefits of AI Vector Search:

**Utilize Familiar Tools:** Take advantage of the familiar SQL environment to manage and query vector data. This eliminates the learning curve associated with new databases and allows developers to leverage their existing SQL skills. By integrating AI Vector Search with SQL queries of business data, developers using AI Vector Search can seamlessly incorporate AI-driven insights into their applications.

**Embed AI into Applications:** Use SQL or REST APIs to add RAG into existing applications. This approach allows developers to enhance application functionality with AI capabilities without overhauling their entire suite of applications and data management systems. The integration of AI Vector Search with REST APIs provides a flexible and scalable solution for embedding AI into diverse applications.





**Leverage AI Frameworks:** Utilize AI frameworks such as LangChain to create AI pipelines. These frameworks enable developers to build sophisticated AI models and workflows, leveraging Oracle's AI capabilities. By incorporating AI Vector Search into these frameworks, developers can streamline AI development and deployment processes.

**Optimize Query Performance:** Implement best practices for optimizing query performance, such as indexing vector data and utilizing parallel processing capabilities. Oracle Database 23ai offers tools and features that can enhance query performance, ensuring fast and efficient retrieval of vector data. Furthermore, developers can reduce the complexity of their applications by letting Oracle Database combine vector search results with regular business queries using a few lines of SQL instead of implementing multiple queries and merging results in application code. Regularly review and optimize query configurations to maintain high performance.

**Collaborate with IT Management:** Work closely with IT management to ensure that the organization's infrastructure capabilities and modernization plans align with the needs of AI projects. Organizations can maximize the benefits of Oracle's AI solutions by aligning development efforts with IT infrastructure modernization efforts such as the implementation of Exadata X10M platforms to efficiently support OLTP, analytics, and mixed workloads.

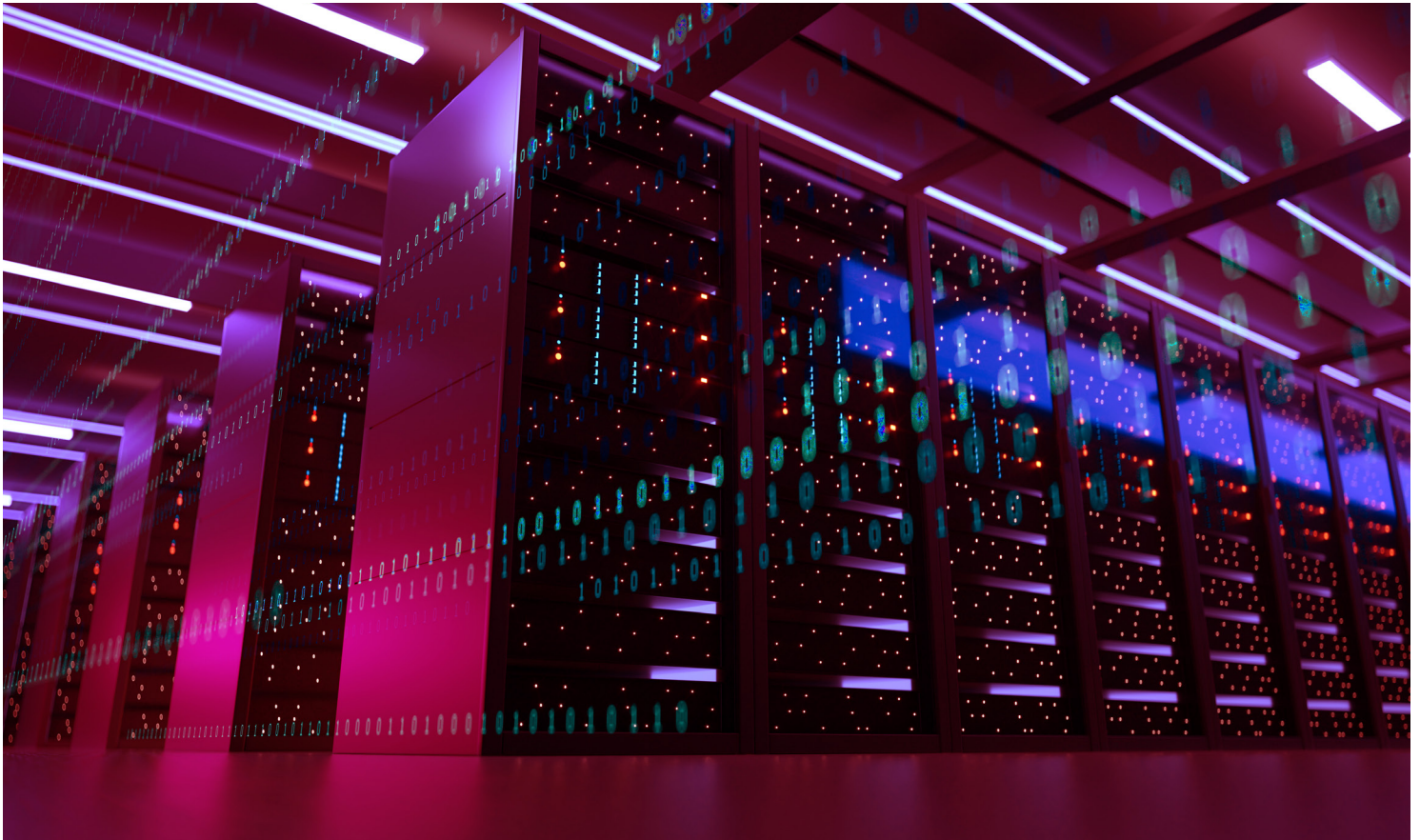
## Strategic Advice for Developers

- 1. Integrate Incrementally:** Start by integrating AI Vector Search capabilities into specific parts of your application, gradually expanding its use as you become more familiar with the technology. This incremental approach reduces risk and allows for iterative improvements.
- 2. Leverage Oracle Documentation and Support:** Utilize Oracle's comprehensive documentation and support services to troubleshoot issues and optimize your use of AI Vector Search. Oracle's resources can provide valuable insights and assistance, ensuring that you fully leverage the capabilities of Oracle Database 23ai.

# Strategic Advice for IT Management

- 1. Invest in Training:** Provide training for your development team on the new features and capabilities of AI Vector Search. Ensuring that your team is well versed in the latest tools and techniques can enhance productivity and innovation.
- 2. Establish Best Practices:** Develop and implement best practices for using AI Vector Search within your organization. This includes guidelines for data management, security, and performance optimization. Establishing best practices can streamline development processes and ensure consistent quality across projects.
- 3. Plan for Growth:** Enterprise workloads are always expanding, and the integration of GenAI capabilities such as semantic search and RAG will accelerate this expansion. Exadata X10M platforms powered by AMD EPYC processors provide a scalable platform to support existing mission-critical OLTP and analytics workloads as well as the current wave of GenAI-induced growth.

By following these recommendations, both developers and IT management can effectively integrate AI Vector Search into their IT infrastructure, leveraging its advanced features to drive innovation and efficiency in AI application development. This strategic approach ensures that organizations can harness the full potential of Oracle's solutions, positioning them for success in the evolving AI landscape.



# Looking Ahead

Oracle continues to focus on simplifying AI application development, making advanced vector technology accessible and easy to use for developers. Oracle Database 23ai's AI Vector Search capabilities demonstrate Oracle's commitment to reducing complexity and enhancing usability for developers at all levels. By integrating vector capabilities directly within the SQL framework, Oracle enables developers to leverage their existing skills and tools, eliminating the need to learn new, specialized databases.

Oracle's declarative approach to vector management means that developers need not concern themselves with the intricate details of vector data structures. Instead, they can focus on building applications that utilize AI insights seamlessly. This ease of use extends to the integration with AI frameworks such as LangChain, allowing developers to incorporate vector search into their AI pipelines with minimal effort.

For developers, the ability to use standard SQL to manage and query vector data simplifies the development process and accelerates time-to-market for AI applications. The familiar SQL environment, combined with Oracle's robust vector search capabilities, ensures that developers can quickly and efficiently build AI-driven applications without needing specialized knowledge in vector databases.

IT management also benefits from this approach since it reduces the need to manage additional data stores and simplifies overall system architecture. Oracle's Exadata X10M platforms use AMD EPYC processors to provide the scalable high performance needed to handle the demands of GenAI applications that leverage semantic search and RAG, enabling organizations to leverage AI technologies effectively and efficiently.

In conclusion, Oracle Database 23ai's AI Vector Search and Oracle Exadata X10M platforms provide a powerful, user-friendly solution for AI application development. By making vector technology accessible through SQL and simplifying the management of vector data, Oracle empowers developers to create innovative AI applications with ease. This approach not only enhances system efficiency but also drives innovation, positioning enterprises to thrive in the AI era. Oracle's continued focus on usability and developer-friendly solutions keeps its customers at the forefront of AI-driven advancements, ready to meet the challenges and opportunities of the future.



# Important Information About this Report

## CONTRIBUTORS

### Steven Dickens

Chief Technology Advisor | The Futurum Group

## PUBLISHER

### Daniel Newman

CEO | The Futurum Group

## INQUIRIES

Contact us if you would like to discuss this report and The Futurum Group will respond promptly.

## CITATIONS

This paper can be cited by accredited press and analysts, but must be cited in-context, displaying author's name, author's title, and "The Futurum Group." Non-press and non-analysts must receive prior written permission by The Futurum Group for any citations.

## LICENSING

This document, including any supporting materials, is owned by The Futurum Group. This publication may not be reproduced, distributed, or shared in any form without the prior written permission of The Futurum Group.

## DISCLOSURES

The Futurum Group provides research, analysis, advising, and consulting to many high-tech companies, including those mentioned in this paper. No employees at the firm hold any equity positions with any companies cited in this document.

# ORACLE®

## ABOUT ORACLE

Oracle is a leading global technology company specializing in cloud computing, database management, and enterprise software solutions. With a rich history of innovation, Oracle empowers businesses of all sizes to streamline their operations, optimize data management, and harness the power of cutting-edge technologies, ensuring they remain at the forefront of the digital age. [Learn more here.](#)



## ABOUT THE FUTURUM GROUP

[The Futurum Group](#) is an independent research, analysis, and advisory firm, focused on digital innovation and market-disrupting technologies and trends. Every day our analysts, researchers, and advisors help business leaders from around the world anticipate tectonic shifts in their industries and leverage disruptive innovation to either gain or maintain a competitive advantage in their markets.



## CONTACT INFORMATION

The Futurum Group LLC | [futurumgroup.com](http://futurumgroup.com) | (833) 722-5337

© 2024 The Futurum Group. All rights reserved.

