

Oracle AI Vector Search Deep Dive

Student Guide
D1109960GC10



Copyright © 2025, Oracle and/or its affiliates.

Disclaimer

This document contains proprietary information and is protected by copyright and other intellectual property laws. The document may not be modified or altered in any way. Except where your use constitutes "fair use" under copyright law, you may not use, share, download, upload, copy, print, display, perform, reproduce, publish, license, post, transmit, or distribute this document in whole or in part without the express authorization of Oracle.

The information contained in this document is subject to change without notice and is not warranted to be error-free. If you find any errors, please report them to us in writing.

Restricted Rights Notice

If this documentation is delivered to the United States Government or anyone using the documentation on behalf of the United States Government, the following notice is applicable:

U.S. GOVERNMENT END USERS: Oracle programs (including any operating system, integrated software, any programs embedded, installed or activated on delivered hardware, and modifications of such programs) and Oracle computer documentation or other Oracle data delivered to or accessed by U.S. Government end users are "commercial computer software" or "commercial computer software documentation" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, the use, reproduction, duplication, release, display, disclosure, modification, preparation of derivative works, and/or adaptation of i) Oracle programs (including any operating system, integrated software, any programs embedded, installed or activated on delivered hardware, and modifications of such programs), ii) Oracle computer documentation and/or iii) other Oracle data, is subject to the rights and limitations specified in the license contained in the applicable contract. The terms governing the U.S. Government's use of Oracle cloud services are defined by the applicable contract for such services. No other rights are granted to the U.S. Government.

Trademark Notice

Oracle®, Java, MySQL, and NetSuite are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

Intel and Intel Inside are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. AMD, Epyc, and the AMD logo are trademarks or registered trademarks of Advanced Micro Devices. UNIX is a registered trademark of The Open Group.

Third-Party Content, Products, and Services Disclaimer

This documentation may provide access to or information about content, products, and services from third parties. Oracle Corporation and its affiliates are not responsible for and expressly disclaim all warranties of any kind with respect to third-party content, products, and services unless otherwise set forth in an applicable agreement between you and Oracle. Oracle Corporation and its affiliates will not be responsible for any loss, costs, or damages incurred due to your access to or use of third-party content, products, or services, except as set forth in an applicable agreement between you and Oracle.

1005162025

Table of Contents

Course Overview	7
Agenda	8
Audience	9
Learning Objectives	10
Refresher: AI Vector Search Fundamentals	13
Objectives	14
Refresher: Vector & Vector Embeddings	15
Vector Data Type	16
Vector Embeddings	17
Generate Vector Embeddings	19
Vector Data Type	20
Refresher: Exact Similarity Search	21
Similarity Search	22
Exact Similarity Search	24
Euclidean and Euclidean Squared Distances	25
Cosine Similarity	26
Dot Product Similarity	27
Manhattan Distance	28
Hamming Similarity	29
Vector Distance Operand	30
Refresher: Approximate & Multi-vector Similarity Search	31
Approximate Similarity Search	32
Multi-Vector Similarity Search	33

Comparison	34
Refresher: Vector Index	36
Vector Indexes	37
Why Do We Need Vector Indexes?	38
Vector Pool	39
In-Memory Neighbor Graph Vector Index (HNSW)	40
Neighbor Partition Vector Index (IVF)	41
Creating a Basic Vector Index	42
Important Parameters	43
Comparison	44
Using Vector Indexes	45
Monitoring Index Accuracy	46
Important Limitations	47
Best Practices	48
RAG Overview	49
Agenda	50
Agenda Topic	51
Gina has a question:	52
Vector Data Workflow	53
Retrieval Augmented Generation works like this	54
Retrieval Augmented Generation works like this	55
RAG Workflow	58
Interact with your favorite LLM Complete the RAG pipeline	59
Agenda Topic	60
RAG with Oracle AI Vector Search	61
Using Embedding Models with Oracle AI Vector Search	65

Agenda	66
Agenda Topic	67
Embedding Models	68
VECTOR_EMBEDDING()	69
Agenda Topic	70
Create Table	71
Vectorize a Table	72
Similarity Search	73
Change Embedding Models	74
Oracle Vector Search and OCI Generative AI Service (using Python)	75
Agenda	76
Agenda Topic	77
OCI Gen Service	78
Agenda Topic	79
Summary of Steps	80
Load the sources	81
Text Chunks	82
Vectorize	83
Create and Call	84
RAG with Oracle AI Vector Search and OCI Gen AI Service (using PL/SQL)	85
Agenda	86
Process Overview	87
Step 1 - Text Extraction and Preparation	88
Step 2 - Embedding Models and Vectors	89
Step 3 - Similarity Search and Response Generation	90
Step 4 - Build the Prompt	91

Step 5: Invoke the Chain	93
Oracle AI Vector Search Supporting Features	95
Agenda	96
Agenda Topic	97
Exadata: AI Storage	98
Agenda Topic	100
Refresher, some common GoldenGate use cases:	101
...still the best choice for an Enterprise Standard	102
GoldenGate Microservices 23ai	103
Oracle GoldenGate	104
Distributed AI processing with vector replication	105
Generative AI with your own business data?	106
Create a real-time vector hub for GenAI	107
Actionable AI/ML from streaming pipelines	108
Summary, why GoldenGate 23ai for AI?	109
Agenda Topic	110
SQL Loader	111
Data Pump	112
Summary	113
Agenda	114
Learning Outcomes	115