ORACLE
University

Oracle AI Database: AI Vector Search Training

Student Guide
D1112667GC10

Learn more from Oracle University at **education.oracle.com**

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.

1011032025

Important Note

Oracle AI Database 26ai has replaced **Oracle Database 23ai**. This change was announced at Oracle AI World in October 2025. The architecture, concepts, and features presented in this course remain fully relevant to 26ai

Table of Contents

Lesson 1: Overview of Oracle AI Vector Searchs 12
- Objectives 13
- VECTOR Data Type 14
- Vector Embeddings 16
- Similarity Search 18
- Vector Embedding Models 20
- Summary 24

Lesson 2: Why use Oracle AI Vector Search? 26
- Objectives 27
- Benefits 28
- Benefits of Oracle AI Vector Search 29
- Examples 30
- Summary 31

Lesson 3: Oracle AI Vector Search Workflow 32
- Objectives 33
- Generate Vector Embeddings 34
- Generate Embeddings Examples 36
- Example: Load ONNX Model into the DB 37
- Example: Generate the Embedding 38
- Store Vector Embeddings 39
- Vector Indexes 41
- Query Data with Similarity Searches 43
- The Complete Workflow 45

Summary 46
Lesson 4: Running Basic Queries on Vectors 48
Objectives 49
Basic Queries 50
Basic Queries: Comparison Operations 54
Demo 55
Summary 56
Lesson 5: Vector Indexes and Memory 58
Objectives 59
Vector Indexes 60
Why Do We Need Vector Indexes? 62
Vector Pool 63
In-Memory Neighbor Graph Vector Index (HNSW) 64
Neighbor Partition Vector Index (IVF) 65
Creating a Basic Vector Index 66
Important Parameters 67
Comparison 68
Using Vector Indexes 69
Monitoring Index Accuracy 70
Important Limitations 71
Best Practices 72
Memory Considerations 73
Vector Storage 75
Memory Considerations: In-Memory Neighbor Graph Indexes 76
Memory Considerations: Sample Calculation 77
Summary 78

Lesson 6: DML Operations on Vectors 80
Objectives 81
Create a Table with a Vector Column 82
VECTOR Data Type 83
Declaration Formats and Explanation 85
Vector DML 86
Vector DML Using SQL*Loader 89
Demo 90
Summary 91
Lesson 7: Vector DDL 92
Objectives 93
Tables with Different Vector Formats 94
Tables with Different Vector Formats: Example 96
Tables with Different Vector Formats: Insert Example 97
DDL Operations on Vectors 98
Prohibited Operations 101
VECTOR Data Type Restrictions 102
Demo 104
Summary 105
Lesson 8: Creating and Finding the Nearest Vectors 106
Objectives 107
Vector Constructor 108
Vector Constructor Examples 110
Vector Distance 111
Vector Distance Operand 112
Vector Distance Metrics 113

Euclidean and Euclidean Squared Distances 114
Cosine Similarity 115
Dot Product Similarity 116
Manhattan Distance 117
Hamming Similarity 118
Vector Distance Examples 119
Shorthand Operators For Distances 120
Demo 121
Summary 122
Lesson 9: Finding the Closest Vectors 124
Objectives 125
Exact Similarity Search 126
Euclidean 129
Euclidean Squared Distance 130
Approximate Similarity Search 131
Approximate Similarity Search or Exact Similarity Search? 133
Approximate Similarity Search 134
Approximate Similarity Search: HNSW 135
Approximate Similarity Search: IVF 136
Multi-Vector Similarity Search 137
Demo 140
Summary 141
Lesson 10: Narrowing Search Results 142
Objectives 143
Attribute Filtering 144
Demo 153

Summary 154
Lesson 11: Testing Other Distance Functions 156
Objectives 157
Other Distance Functions 158
Other Vector Distance Functions 159
L1_DISTANCE 160
L2_DISTANCE 161
COSINE_DISTANCE 162
INNER_PRODUCT 163
Demo 164
Summary 165
Lesson 12: Testing Other Vector Functions 166
Objectives 167
Other Vector Functions 168
Vector Constructors 170
Vector Serializers 172
Vector Serializers 173
Vector Norm 174
Vector Dimension Count 175
Vector Dimension Format 176
Demo 177
Summary 178
Lesson 13: RAG Overview 180
Agenda 181
Agenda Topic 182
Gina has a question: 183

Vector Data Workflow 184
Retrieval Augmented Generation works like this 185
RAG Workflow 189
Interact with your favorite LLM | Complete the RAG pipeline 190
Agenda Topic 191
RAG with Oracle AI Vector Search 192
RAG with Oracle AI Vector Search (cont) 194
RAG with Oracle AI Vector Search 195
Lesson 14: Using Embedding Models with Oracle AI Vector Search 196
Agenda 197
Agenda Topic 198
Embedding Models 199
VECTOR_EMBEDDING() 200
Agenda Topic 201
Create Table 202
Vectorize a Table 203
Similarity Search 204
Change Embedding Models 205
Lesson 15: Oracle Vector Search and OCI Generative AI Service (using Python) 206
Agenda 207
Agenda Topic 208
OCI Gen Service 209
Agenda Topic 210
Summary of Steps 211
Load the sources 212
Text Chunks 213

Vectorize 214
Create and Call 215
Lesson 16: RAG with Oracle AI Vector Search and OCI Gen AI Service (using PL/SQL) 216
Agenda 217
Process Overview 218
Step 1 219
Step 2 220
Step 3 221
Step 4 222
Step 4 223
Step 5: Invoke the Chain 224
Lesson 17: Oracle AI Vector Search Supporting Features 226
Agenda 227
Agenda Topic 228
Exadata: AI Storage 229
Agenda Topic 231
Refresher, some common GoldenGate use cases: 232
…still the best choice for an Enterprise Standard 233
GoldenGate Microservices 23ai 234
Distributed AI processing with vector replication 236
Generative AI with your own business data? 237
Create a real-time vector hub for GenAI 238
Actionable AI/ML from streaming pipelines 239
Summary, why GoldenGate 23ai for AI? 240
Agenda Topic 241
SQL Loader 242

Data Pump 243
Lesson 18: Select AI With Autonomous 244
Agenda 245
Select AI - Simplest way to get answers about your business 246
Select AI - Just ask your database 247
Select AI - Benefits 248
Challenges before Select AI 249
Select AI: SQL 250
Agenda Topic 251