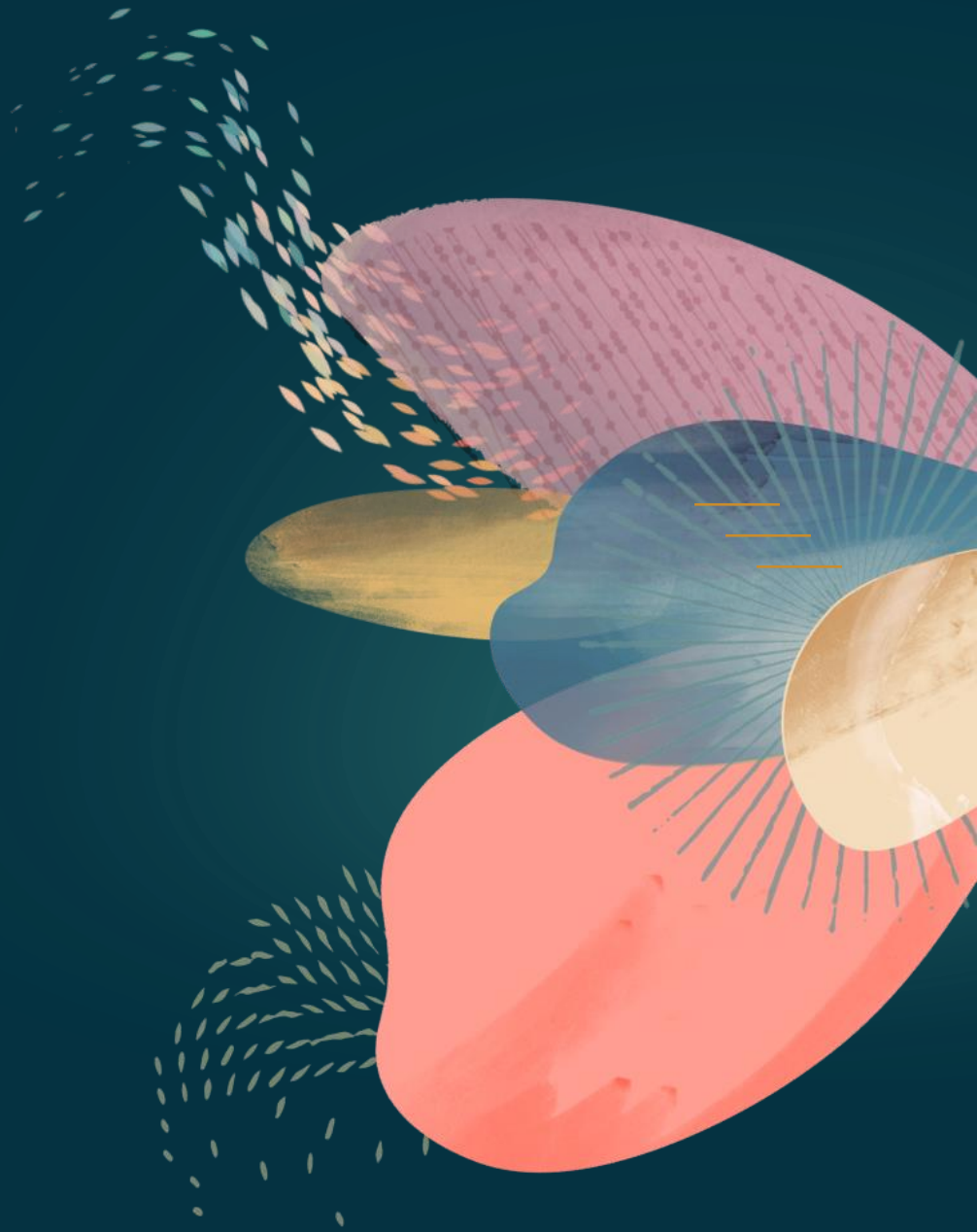
AI for Data

Data and AI Engineered to Work Together

Victor Stachura
Principal Solution Architect
Oracle North America, Database Specialist Team



Safe harbor statement

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, timing, and pricing of any features or functionality described for Oracle's products may change and remains at the sole discretion of Oracle Corporation.

Copyright © 2025, Oracle and/or its affiliates

Oracle DB | History

1977
Software Development Laboratories (SDL)
Santa Clara, California

1979
SDL becomes **Relational Software Systems, Inc.**
- First commercial RDBMS

1981
C Programming Language chosen for v3

1982
Relational Software Systems, Inc. becomes **Oracle**

V3
Portability at its best
- Atomic execution of SQL statements
- Nonblocking queries
- Mapping of data in pre-join formats

V4
- Read Consistency
- Export/Import utility
- Early Reporting capabilities

V5
- Client/Server mode
- Clustered Configuration (VAX Clusters)
- Security and Audit Features

V6
- Row level locking
- Hot backup
- First version of PL/SQL
- High Speed OLTP

V6.2
- OPS for VAX Cluster
- Foundations for HA & Scalability
- Oracle Card
- 1-2-3 Data Lens

V7
- PL/SQL Stored procedures
- Triggers
- Distributed 2-phase commit
- Shared cursors
- Cost based optimizer

V5.1
- Distributed Queries (single query access to data stored in multiple locations)
- SQL * Forms is released
- SQL * Plus is released

Oracle becomes a publicly traded company

V8
- Recovery Manager
- Partitioning
- Expanded support for Java

V8.1
- Native Internet Protocols
- Virtual Private Database
- Standby Database foundation for Data Guard

V9.1
- Real Application Clusters (RAC)
- Integrated Data Mining with OLAP
- Data Guard Broker

10g
First Database designed for enterprise grid computing

Foundation for:
- Grid Infrastructure, ASM
- Flashback Database

10g R2
Foundation for:
- Real Application Testing
- Database Vault
- Advanced Compression & TDE
- Data Guard Fast-Start Failover

11g
Foundation for:
- Active Data Guard
- Secure Files

11g R2
Foundation for:
- EBR, HCC, Data Redaction
- ACFS
- Golden Gate Replication

12c
- Multitenant Architecture
- In-Memory Column Store
- Native JSON
- SQL Pattern Matching

12c R2
- Sharding
- Snapshot-based replication
- PDB Hot Clone
- Enhanced Partitioning

18g
- Autonomous Database announced
- User-defined sharding
- PDB Snapshot carousel
- DB In-memory support for External tables

19c
Generally Available since April 2019
- Automatic Indexing
- Hybrid Partition Tables
- Active Data Guard DML Redirect

21c
Innovation Release
- Per-PDB Data Guard
- Native JSON Type
- SQL Macros
- JavaScript Stored Logic

26ai
- AI Vector Search
- JSON Duality
- Graph
- TrueCache
- Global Distributed Database

1977 1979 1981 1982 1983 1984 1985 1986 1988 1991 1992 1997 1998 2001 2003 2005 2007 2009 2013 2016 2018 2019 2020 2023 2024 2026

Copyright © 2024, Oracle and/or its affiliates

Vector
Spatial
JSON
Text
XML
Relational
AI / ML
IoT
Analytical
Streaming
OLTP
Graph

Data Types

Workloads

Blockchain
Distributed

APEX
ORDS
SQLcl

Tools

AutoML
Enterprise Manager
SQL Developer VS Code

ORACLE AI Database 26ai

Dev Interfaces

REST

SQL
SQL First
MongoDB API

RUST
C++
NODE.JS
RUBY
C
PL/SQL
PHP
GO
Language Support

Runs Anywhere

On-Premises
Containers
Oracle Cloud
Cloud@ Customer
@Azure / @AWS / @GCP

Copyright © 2025, Oracle and/or its affiliates

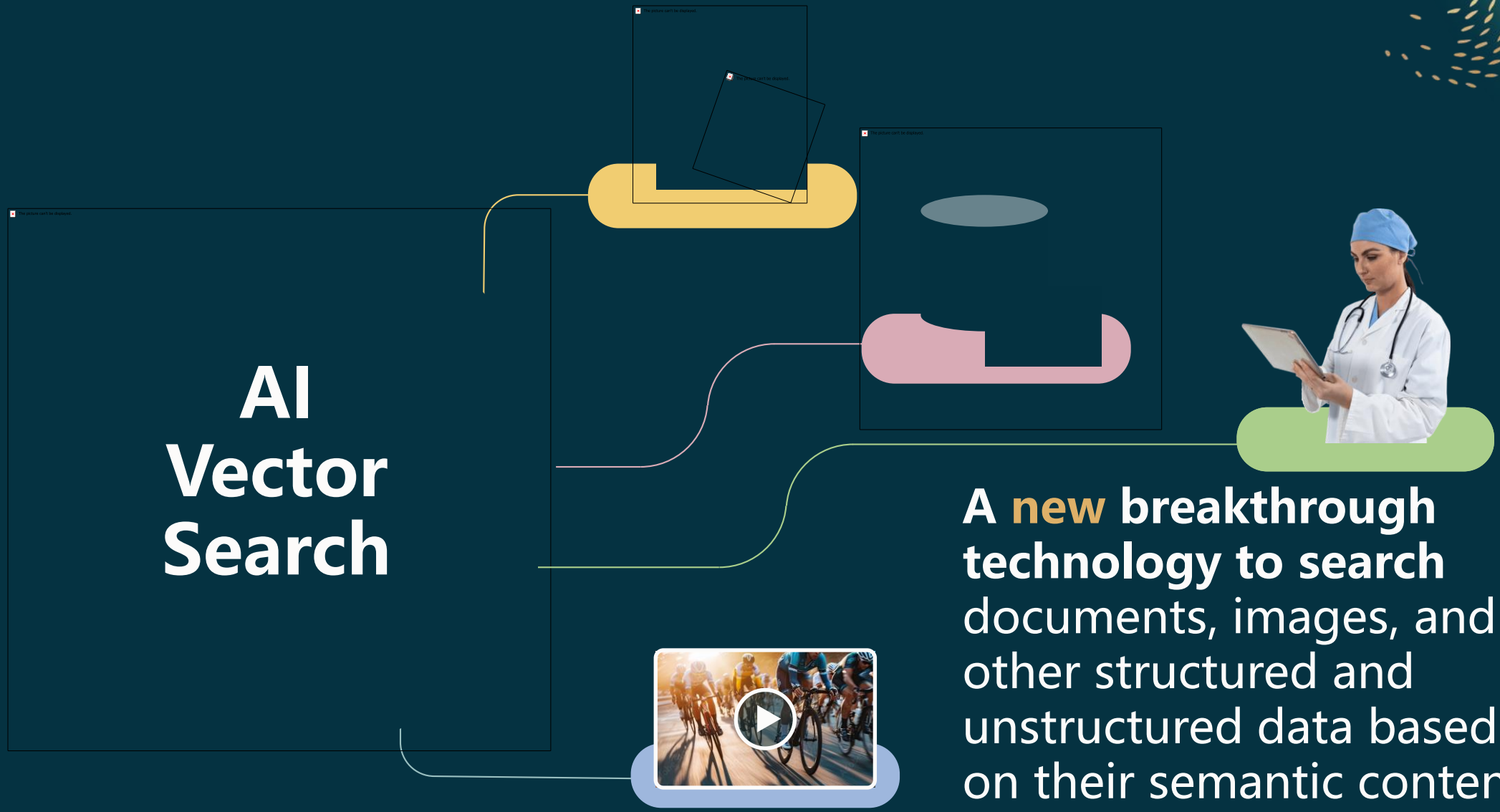
Oracle AI Database includes many AI Capabilities

Engineered to Work Together

- AI Vector Search
- APEX AI Assistant
  Chat with LLM
  Generate Text with AI
- Augmented Generative AI (RAG)
- Select AI + Select AI Agents
- AI Agents (MCP)
- Machine Learning

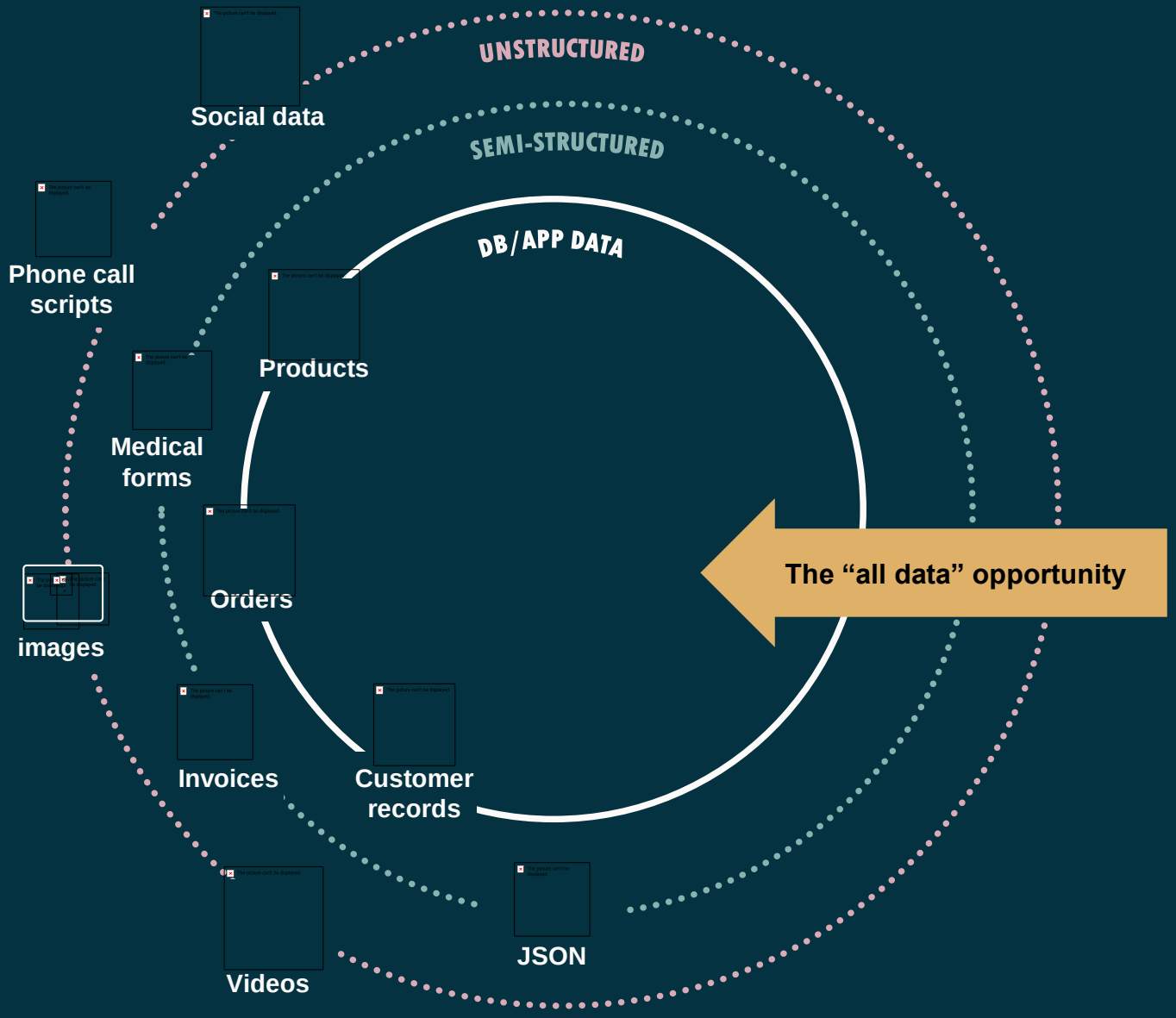
Copyright © 2026, Oracle and/or its affiliates

AI Vector Search



A new breakthrough technology to search documents, images, and other structured and unstructured data based on their semantic content, rather than their words or pixels

Copyright © 2026, Oracle and/or its affiliates



Enterprises are facing a growing need to search both unstructured and structured business data, by their semantics or meaning

Find products that match a photo or a text description

Copyright © 2026, Oracle and/or its affiliates

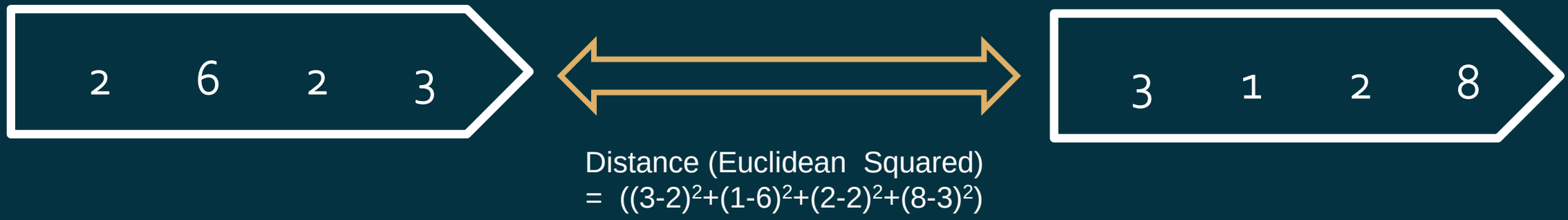
AI Vector Search works by representing the semantic content of a document, image, video, or even relational data as a sequence of numbers, called a vector

Developers create a vector for an object by just passing the object to a built-in vectorization function

Oracle Vector Database natively stores vectors and compares vectors to find objects with similar semantic content

Copyright © 2026, Oracle and/or its affiliates

The main operation on vectors is the Mathematical Distance between them

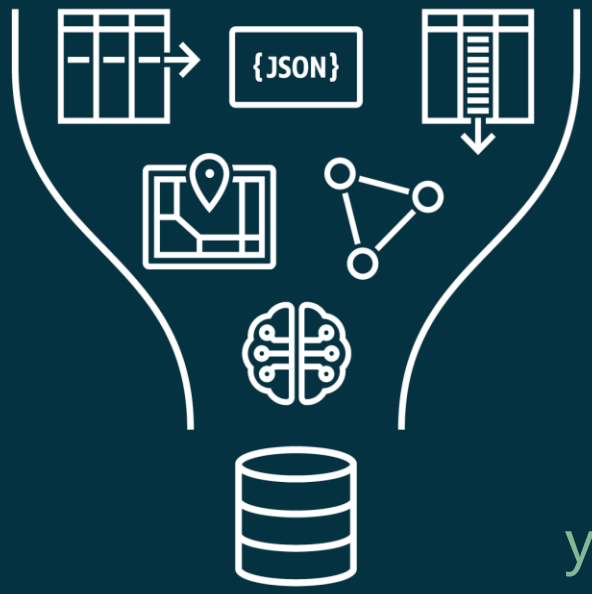


There are many mathematical distance formulas. E.g.:

1. *Cosine similarity*
2. *Euclidean Squared*
3. Manhattan distance
4. Hamming distance
5. Dot product

Copyright © 2024, Oracle and/or its affiliates

Searches on a combination of business and semantic data is more effective if both types of data are stored together



Business Data

One solution is to continuously send your business data to a vector database

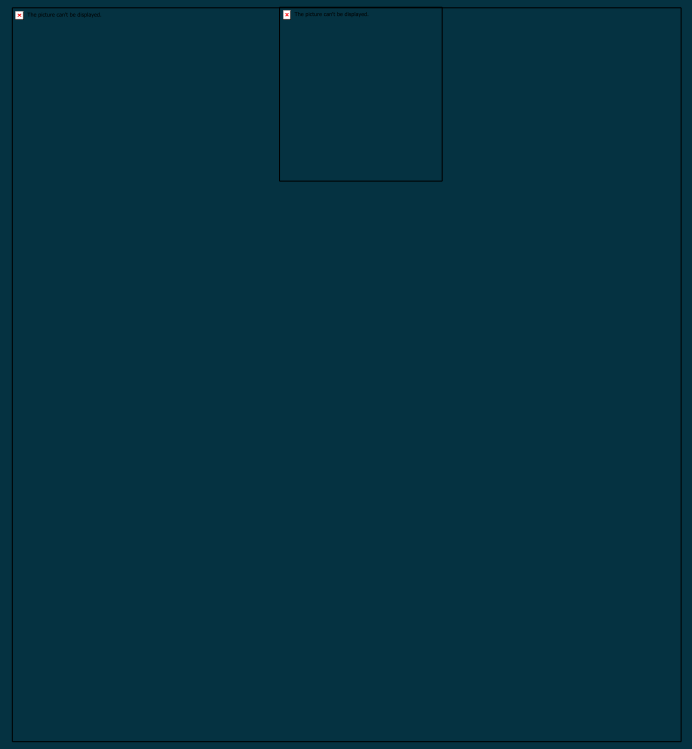
Business Database

Vector Database

However, the business data that is relevant to a question varies widely
Plus, dedicated vector databases are not good at searching or securing business data

Copyright © 2026, Oracle and/or its affiliates

Better Together: Business Data and Business Vectors



Oracle's Converged Data Architecture

Uniquely combines sophisticated business data search with vector similarity search using simple SQL

There is no need to move and synchronize data, manage multiple products, etc.

Every mission-critical feature of Oracle AI Database works transparently with AI Vector Search

Oracle Database's robust security controls ensure compliance with corporate security standards

Allowing AI Vectors to be used immediately in enterprise apps of any scale or criticality

Copyright © 2026, Oracle and/or its affiliates

Every mission-critical feature
of Oracle AI Database works transparently with AI Vectors

Allowing AI Vectors to be used immediately in enterprise apps of any scale or criticality

Copyright © 2026, Oracle and/or its affiliates

Vector Embedding Generation | Your Way

AI Vector Search offers 4 alternatives for vector embedding generation

1 Use Pre-created embeddings

Load vectors directly from external files into database into VECTOR columns or map the data as external tables

2 Use an external embedding cloud-service

Generate embeddings using external callouts via the `UTL_TO_EMBEDDING()` function in the `DBMS_VECTOR` package

3 Use an external embedding library in the mid-tier

Generate embeddings using a language -specific function calls

These embedding libraries can be from HuggingFace or use the ONNX Runtime

4 Use a database resident embedding model

Generate embeddings using the `VECTOR_EMBEDDING()` SQL function using an imported **`ONNX`** embedding model so that no data leaves the database

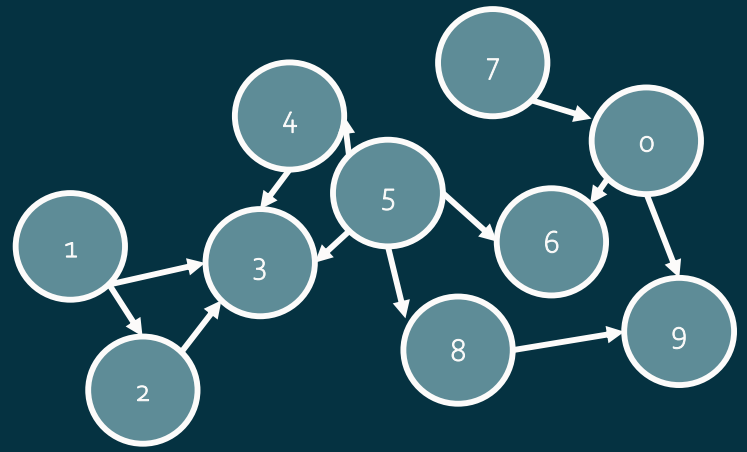
Copyright © 2026, Oracle and/or its affiliates

Vector Indexes | Neighbor Graph Vector Index

Graph-based index where vertices represent vectors and edges between vertices represent *similarity*

In-Memory only index - highly efficient for both accuracy and

```
CREATE VECTOR INDEX house_idx
ON house_for_sale(house_vector)
ORGANIZATION INMEMORY NEIGHBOR GRAPH
DISTANCE EUCLIDEAN;
```



Graph Vector Index
(e.g. **HNSW** Index)

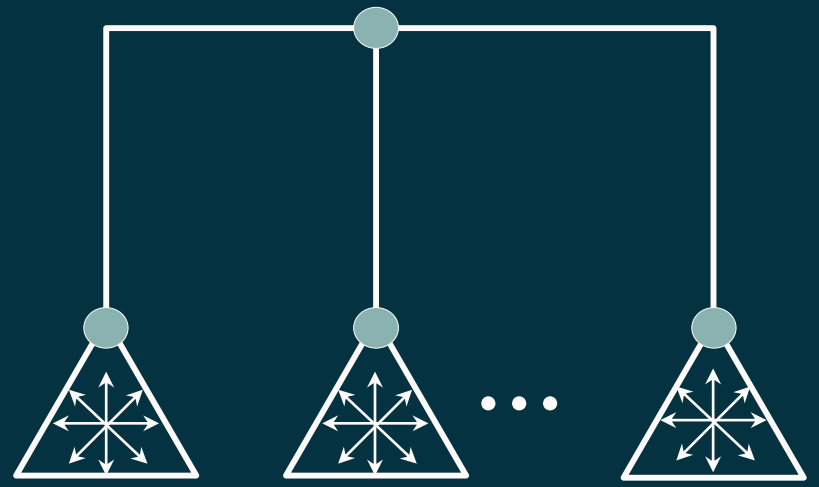
Copyright © 2026, Oracle and/or its affiliates

Vector Indexes – Neighbor Partition Vector Index

Partition-based index with vectors clustered into table partitions based on *similarity*

Efficient scale-out index for unlimited data size

```
CREATE VECTOR INDEX house_idx
ON house_for_sale(house_vector)
ORGANIZATION NEIGHBOR PARTITIONS
DISTANCE EUCLIDEAN;
```



Partition Vector Index

(e.g. **IVF_FLAT** index)

Copyright © 2026, Oracle and/or its affiliates

Hybrid Vector Index | Overview

There are two approaches for searching document data:

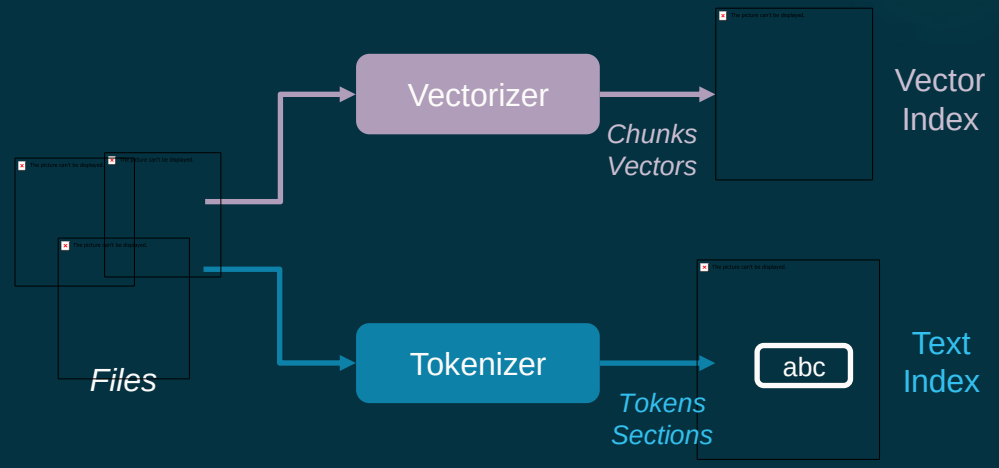
- Keyword Search: Deterministic and explainable, but does not fully capture document semantics
- AI Vector Search: Powerful semantic search technique, but neither deterministic nor explainable

Many real-world document searches require both techniques to be combined:

- E.g., Find homes **with " Cape Cod architecture"** that **contain** ”**Water View**” in the description

  Requires semantic match (”Like Cape Code Arch”) plus keyword match (“containing Water View in description”)

A Hybrid Vector Index combines keyword search with vector search for improved accuracy



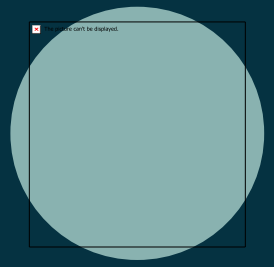
Hybrid Vector Index

```
CREATE HYBRID VECTOR INDEX
house_desc_idx ON
house_for_sale(house_desc)
PARAMETERS('MODEL embed-model');
```

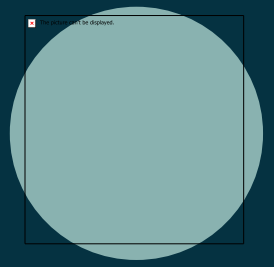
ONNX model that was imported into the database earlier

Copyright © 2026, Oracle and/or its affiliates

Enterprises are already using Oracle AI Vector Search to create innovative new solutions



Visual Search for Products



Real-time offer management

Bio-Tech pattern recognition

Detect Product Defects

Copyright © 2026, Oracle and/or its affiliates

Oracle AI Database includes many AI Capabilities

Engineered to Work Together

- AI Vector Search
- Creating vectors with ONNX framework
- Augmented Generative AI
- Select AI
- AI Agents (MCP)
- over 30 ML algorithms

Copyright © 2026, Oracle and/or its affiliates

Comparing Generative AI and Agentic AI

Capability	Generative AI	Agentic AI
Main objective	Creating. Content creation (text, image, audio, code, SQL) using deep learning models, e.g., LLMs.	Doing. Perform actions and make decisions to achieve specific goals with minimal human intervention. Automate complex tasks.
Degree of autonomy	Requires explicit instructions. Waits for instructions. Works on narrowly defined tasks.	Can operate independently with minimal human guidance on multi-step tasks. Operate proactively.
Decisioning	Limited to generating outputs within prompt constraints.	Can assess situations, reason through problems, plan steps, and adapt strategies to achieve goals and solve problems. Uses “perceive, reason, act, and learn” approach to complete objectives.
Mode of interaction	One-time response to specific prompts. Chatbot conversations can extend memory across invocations. Can adapt to feedback provided in subsequent prompts.	Can handle complex, multi-step scenarios to achieve goals. Can respond/adjust behavior in real-time based on environment feedback or data with iterative task completion and solution evaluation.
Real-world integration	Primarily performs operations within digital environments.	Can interface with external systems, tools, and APIs to accomplish tasks.

See also:
What is Agentic AI and How Does it Work?
Generative AI Vs. Agentic AI: The Key Differences Everyone Needs To Know

Copyright © 2026, Oracle and/or its affiliates

Oracle AI Database improves Generative AI by **augmenting LLM prompts** with **private database content** that is found using any combination of data and AI Vector Search

Enables LLMs to use business data to produce better and more contextually relevant answers to user questions while keeping business data secure

Called: Retrieval Augmented Generation (RAG)

Copyright © 2026, Oracle and/or its affiliates

AI Vector Search in Oracle AI Database Powers Complete Gen AI Pipeline

Retrieval Augmented-Generation (RAG) with your enterprise data

Vectorize Question
An end-user's human language question is encoded as a vector

1

AI Vector

2

Find Related Data
AI Vector Search finds private database data that matches the user's vector

The issue is with the firmware controlling the fan. Apply OS update 42 while plugged in, in a cool air-conditioned environment to prevent overheating

User

Retrieval Augmented Generation (RAG)

Product Info

Support Tickets

Ask LLM
The combination is sent to an LLM to provide an informed answer to the question

4

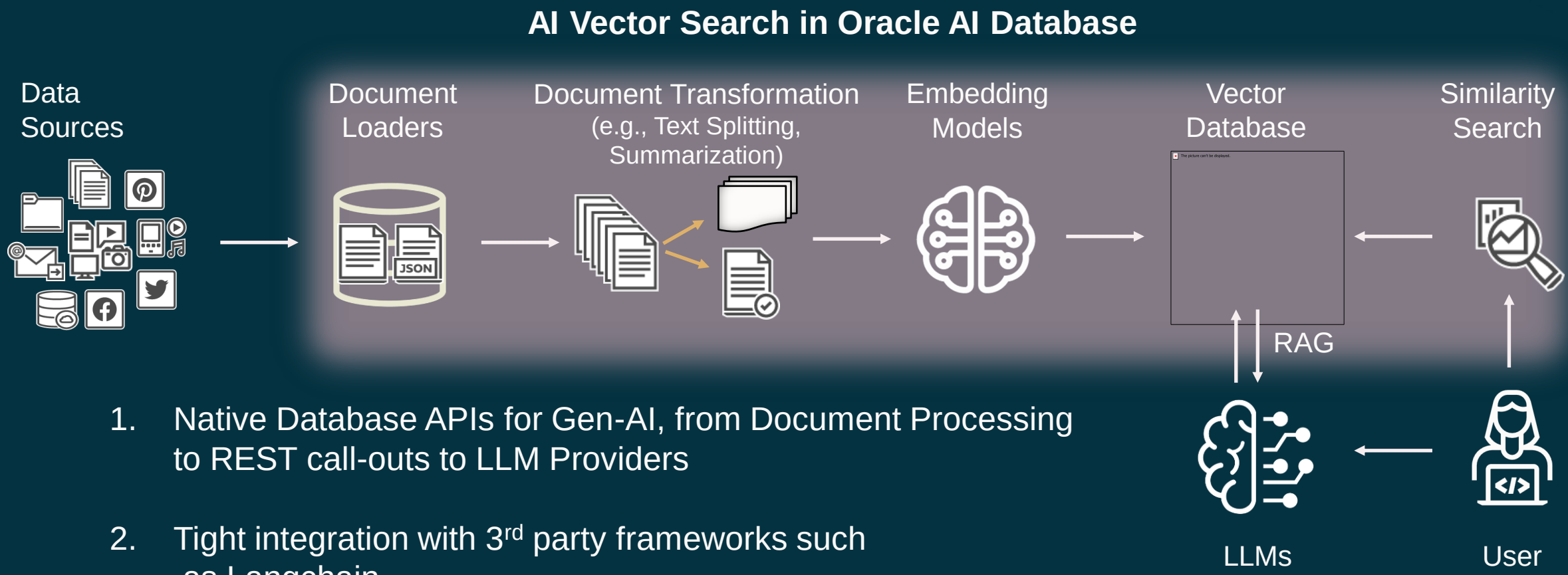
GenAI

3

Augment Prompt
The user's question is augmented with this private data

Copyright © 2026, Oracle and/or its affiliates

AI Vector Search powers Complete Gen AI pipeline



1. Native Database APIs for Gen-AI, from Document Processing to REST call-outs to LLM Providers

2. Tight integration with 3rd party frameworks such as Langchain

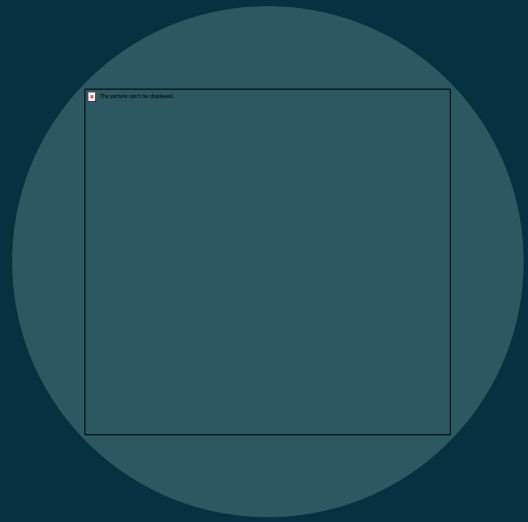
Copyright © 2026, Oracle and/or its affiliates

Agentic AI Overview

AI agents are the building blocks of autonomous AI systems

LLMs are extremely powerful but one-shot LLM access can suffer from:

- Poor multi-step reasoning
- Lack of adaptive capabilities
- Hallucination



AI applications using LLMs require agentic modules for:

- Detailed business context from multiple sources
- Long-term memory for user interactions
- Handling dynamic requests with variable outcomes
- Evaluation and adjustment of outputs, etc.

Copyright © 2026, Oracle and/or its affiliates

Oracle AI Database includes many AI Capabilities

Engineered to Work Together

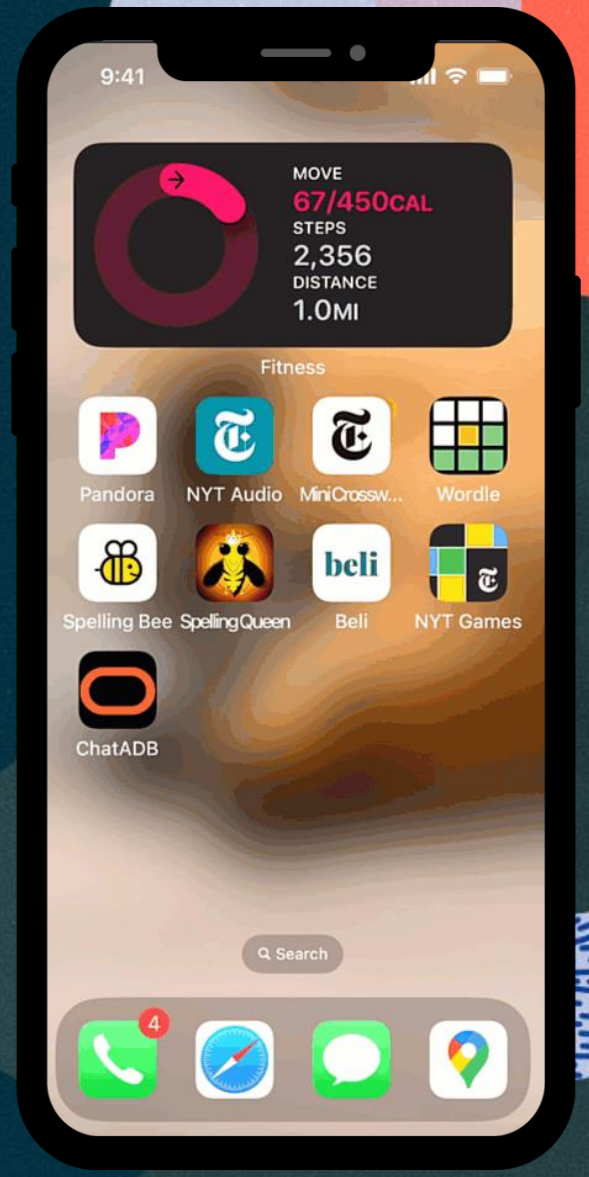
- AI Vector Search
- Creating vectors with ONNX framework
- Augmented Generative AI
- Select AI
- AI Agents (MCP)
- over 30 ML algorithms

Copyright © 2026, Oracle and/or its affiliates

Select AI Goal

Reduce the “distance” between business users and their data

Get answers from structured and unstructured sources using natural language without the complexity of underlying platform or implementation details



Copyright © 2025, Oracle and/or its affiliates | Oracle Confidential

Select AI example use cases

SQL Query	Ask natural language questions about your database data	Assist database developers in writing SQL queries against their application data	
	Assist database developers to understand SQL queries – step by step	Return structured query results in text to present to users	
Chatbot	Generate content with simple or complex custom prompts easily from your database – e.g., email generation, sentiment analysis	Ask questions and get more relevant and accurate responses by using content from your trusted, private documents	RAG
SDG	Create “realistic” data in database tables to support, e.g., testing/debugging applications and interfaces	Build interactive and autonomous agents that perform tasks and use tools	Agents
Summarize	Generate a summary of long text with choice of output style and processing method	Translate text from one language to another – simplify app-dev and assist in translating LLM results to the desired language	Translate

Copyright © 2025, Oracle and/or its affiliates | Oracle Confidential

AI Profile

Configure how you want Select AI to behave for SQL generation

Choose your AI provider

Choose your LLM

Work with your full schema or specific tables and views

Create profile →

- Profile name
- AI provider
- Credential
- Model
- Object list

```
BEGIN
  DMBS_CLOUD_AI.create_profile(
    profile_name => 'openai_gpt4',
    attributes   =>
     '{"provider": "openai",
       "credential_name": "OPENAI_CRED",
       "model": "gpt-4o",
       "object_list": [{"owner": "MOVIESTREAM", "name": "movies"},
                       {"owner": "MOVIESTREAM", "name": "streams"},
                       {"owner": "MOVIESTREAM", "name": "actors"},
                       {"owner": "MOVIESTREAM", "name": "genre"},
                       {"owner": "MOVIESTREAM", "name": "customer_segment"}]}'
    );

  dbms_cloud_ai.set_profile(profile_name => 'openai_gpt4’);
END;
```

Set profile →

Copyright © 2025, Oracle and/or its affiliates | Oracle Confidential

Packages to support in DB AI

Layer	Packages
SELECT AI (natural language SQL)	`DBMS_CLOUD_AI`
Agent Framework (reasoning, tools)	`DBMS_CLOUD_AI_AGENT`
Tools	`DBMS_CLOUD_AI_TOOLS`
Vector / RAG	`DBMS_VECTOR` , `DBMS_CLOUD_AI_INDEX`
Credentials	`DBMS_CLOUD`

List AI profiles and their attributes

```
select * from user_cloud_ai_profiles;
select * from user_cloud_ai_profile_attributes;
```

Copyright © 2025, Oracle and/or its affiliates | Oracle Confidential

Oracle AI Database includes many AI Capabilities

Engineered to Work Together

- AI Vector Search
- Creating vectors with ONNX framework
- Augmented Generative AI
- Select AI Agent
- AI Agents (MCP)
- over 30 ML algorithms

Copyright © 2026, Oracle and/or its affiliates

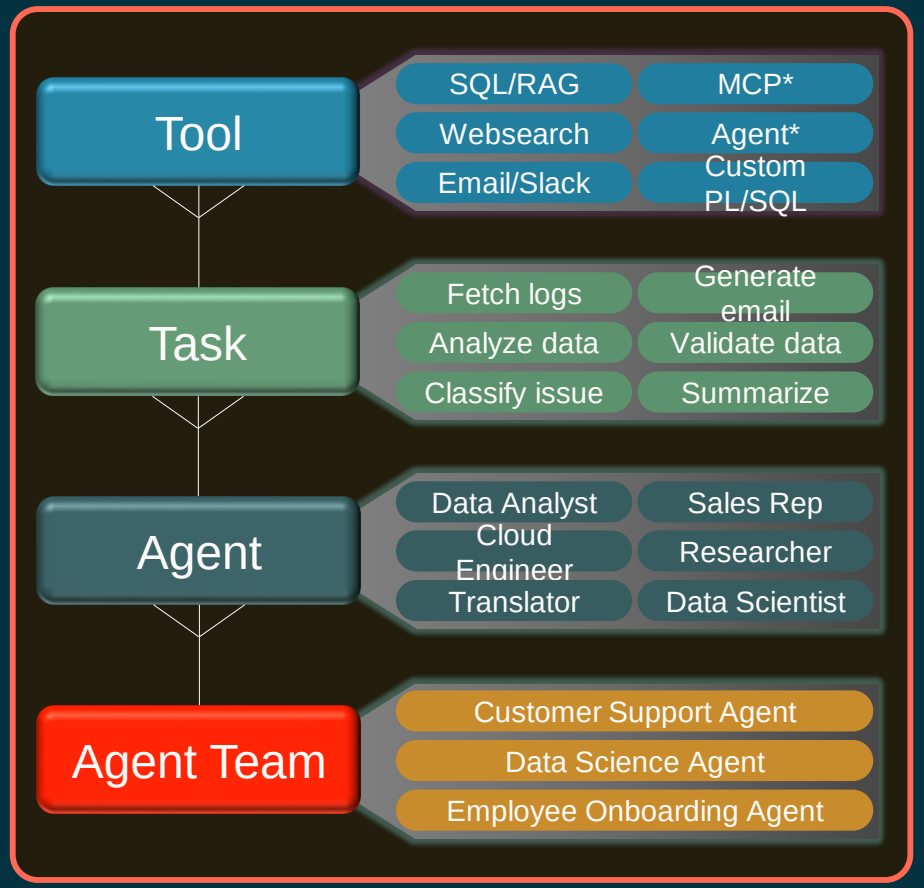
Select AI Agent – an autonomous agent framework

A simple framework to build, deploy, and manage AI agents

Developers define agents declaratively using PL/SQL, and coming soon, Python

Core components

- **Tool** – a software feature or integration that enables an AI agent to perform a specific function or interact with external systems as part of an automated workflow
- **Task** – a specific action or unit of work assigned to an AI agent to perform as part of an automated workflow
- **Agent** – an actor with a clearly-defined role that performs assigned tasks as part of an overall workflow
- **Agent Team** – one or more agents that may collaborate to complete tasks within an automated workflow



Copyright © 2026, Oracle and/or its affiliates

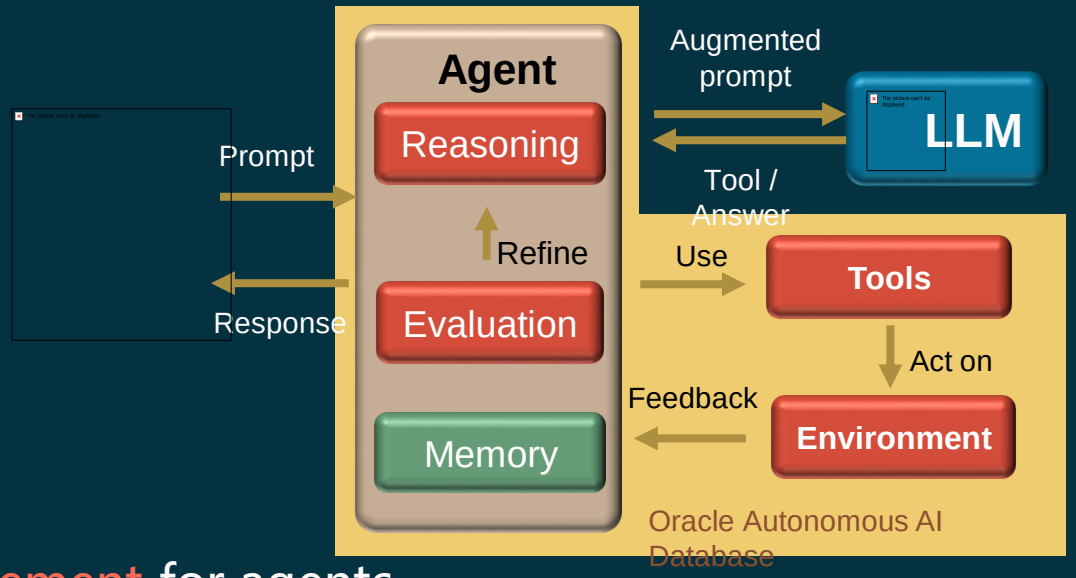
Select AI Agent – an autonomous agent framework
DBMS_CLOUD_AI_AGENT package

Simple, automated framework to build, deploy, and manage AI agents

Built-in ReAct agentic pattern

Tool options

- Build custom tools using PL/SQL
- Invoke external tools using REST
- Use cloud functions such as OCI Functions and AWS Lambda functions
- Use Select AI-provided pre-built tools



Automatic short-term and long-term memory management for agents

Multi-turn chat using conversations

Auto-scalable in Autonomous AI Database

Open Agent Specification-format compatible agent import and export

Copyright © 2026, Oracle and/or its affiliates

Pre-built Autonomous AI Database-focused AI agents using Select AI Agent

https://github.com/oracle-devrel/oracle-autonomous-database-samples/autonomous-ai-agents

Agent	Description
NL2SQL Data Retrieval Agent	Query databases using natural language by automatically interpreting questions, handling ambiguity, and generating charts or visualizations from your data
Database Inspect Agent	Natural language-driven assistant that helps you explore, understand, and manage Oracle database objects—such as tables, views, triggers, functions, procedures, packages, and schemas—using simple conversational queries
OCI Autonomous AI Database Provisioning and Lifecycle Agent	Manage Autonomous AI Database in OCI through natural, conversational interactions
OCI Network Load Balancer Agent	Manage, configure, and monitor OCI Network Load Balancers—including their listeners, backend sets, and health checks—using natural language commands
OCI Object Storage Agent	Manage buckets, objects, lifecycle policies, retention rules, and data replication in OCI Object Storage using natural language commands
OCI Vault Agent	Secure, conversational management of secrets in OCI, supporting operations like creation, inspection, rotation, and scheduled deletion of sensitive credentials
Cloud Repository Connector Agent	Interact with GitHub, AWS CodeCommit, and Azure repositories using a common toolset for repository initialization and management, file management, and database object and metadata/DDL export
Jira Insight Agent	Connect ADB to Atlassian Jira Cloud APIs for conversational operations such as issue search and insight generation, assignee lookup, comments/changelog/worklog retrieval, project and user lookup, and board discovery

Copyright © 2026, Oracle and/or its affiliates

Select AI capabilities by database type

Capability	ADB 26ai	ADB 19c	ODB 23.26.1	ODB 19.30
Chat	✓	✓	✓	✓
NL2SQL	✓	✓	✓	✓
Feedback	✓		✓	
Auto Object Selection	✓		✓	
RAG	✓		✓	
SDG	✓	✓	✓	✓
AI agents	✓	✓	✓	✓
Summarization	✓	✓	✓	✓
Translation	✓	✓	✓	✓

Copyright © 2026, Oracle and/or its affiliates