ORACLE
Data Deep Dive
at AI World

Data Deep Dive at Oracle AI World 25

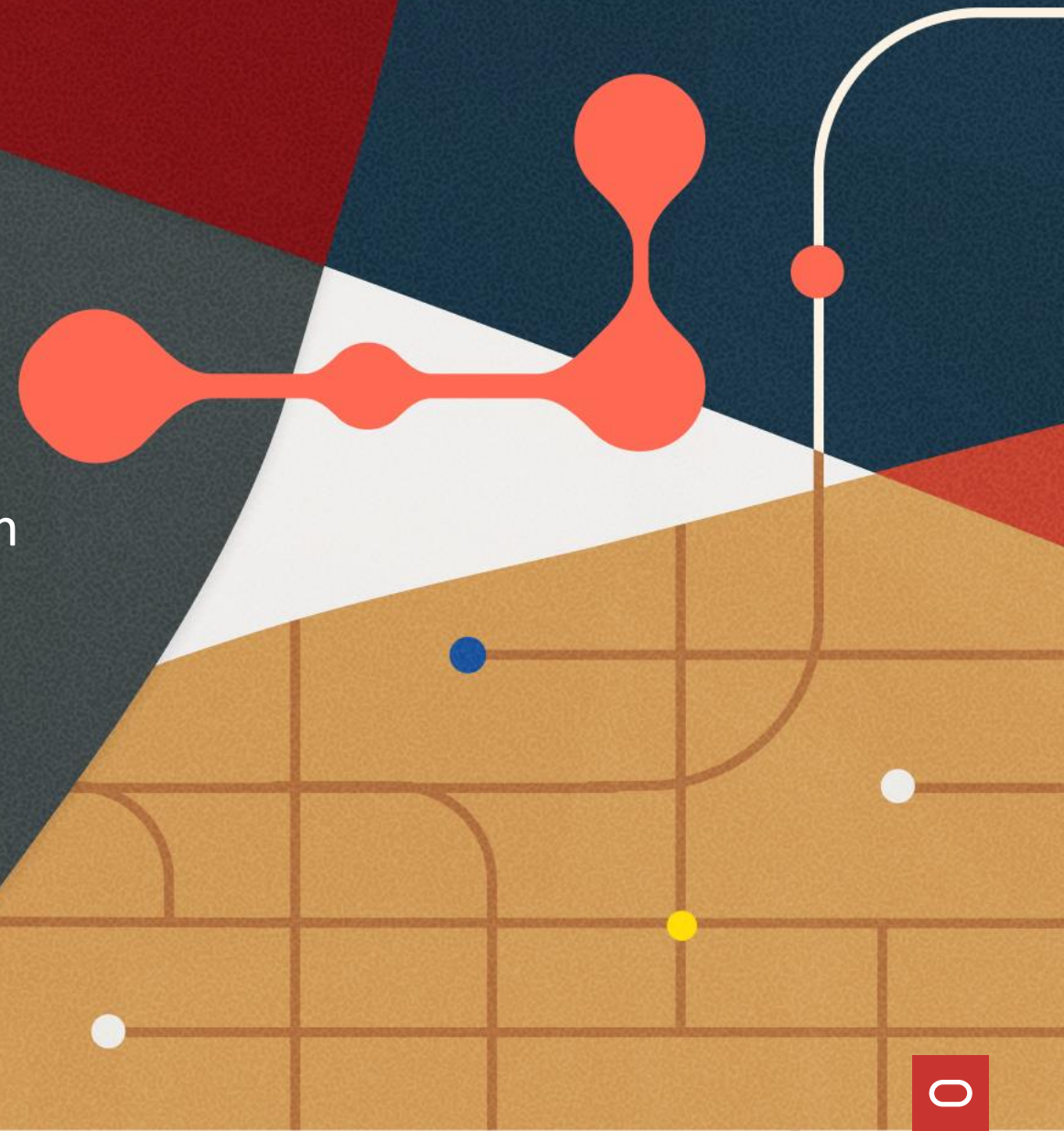
HOL2804 - Getting Started with AI Vector Search

Andy Rivenes, Sean Stacey

AI Vector Search Product Management

Oracle

October 13-16, 2025



Getting Started with AI Vector Search HOL Agenda

1 Welcome and Introduction

2 Overview: AI Vector Search

3 Introduction to the Workshop Environment

4 How AI Vector Search Works

5 Wrap Up and Next Steps

Welcome and Introduction

HOL2804 - Getting Started with AI Vector Search

Speakers

Andy Rivenes

Product Manager

Sean Stacy

Product Manager Database AI Engine

Overview: AI Vector Search

HOL2804 - Getting Started with AI Vector Search

Vector Search is a breakthrough capability for searching data by semantics

Find similar documents, images, and other unstructured data based on their semantic content, rather than their words or pixels

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

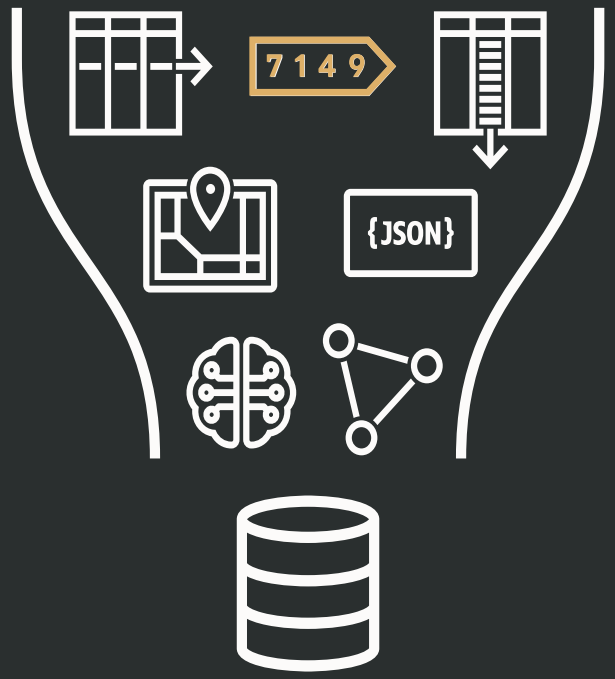
Oracle Vector Database natively stores vectors and compares vectors to find objects with similar semantic content

The main operation on vectors is the Mathematical Distance between them

Distance (Euclidean Squared)
$= ((3-2)^2+(1-6)^2+(2-2)^2+(8-3)^2)$

There are many mathematical distance formulas

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.

Enterprise Similarity Search Use-Cases

Find Similar Support Tickets

Biometric pattern recognition

Find Similar Products

Detect manufacturing anomalies

Product Recommendation

Natural language catalog search

Introduction to the Workshop Environment

HOL2804 - Getting Started with AI Vector Search

Requirements for running a Hands-on-Lab

Laptop

Oracle Account

livelabs.oracle.com/ai-world25/HOL2804

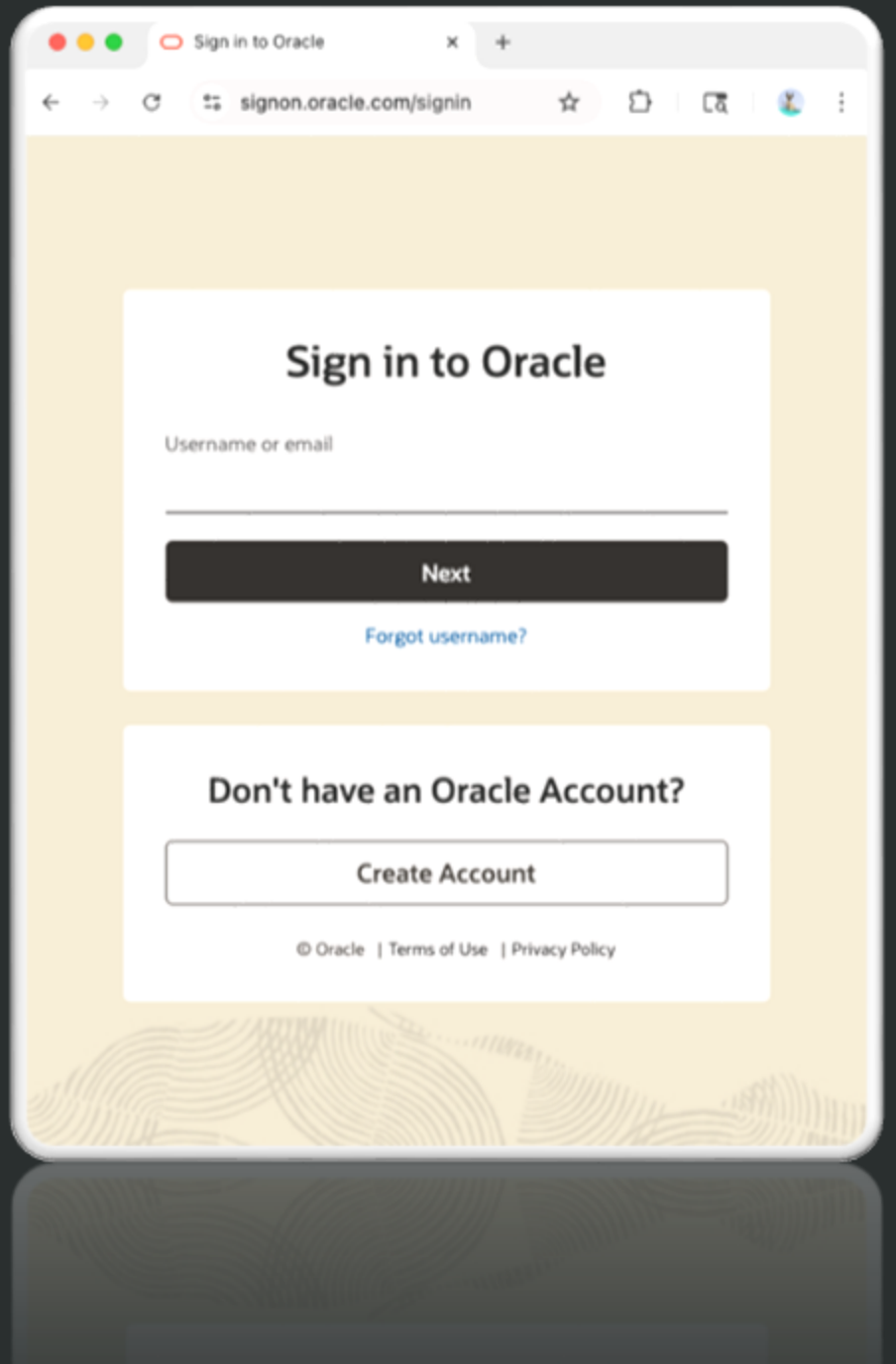
Getting Started with AI Vector Search

Let’s get started

livelabs.oracle.com/ai-world25/HOL2804

- Open a web browser
- Navigate to the provided URL
- Login with your Oracle SSO account

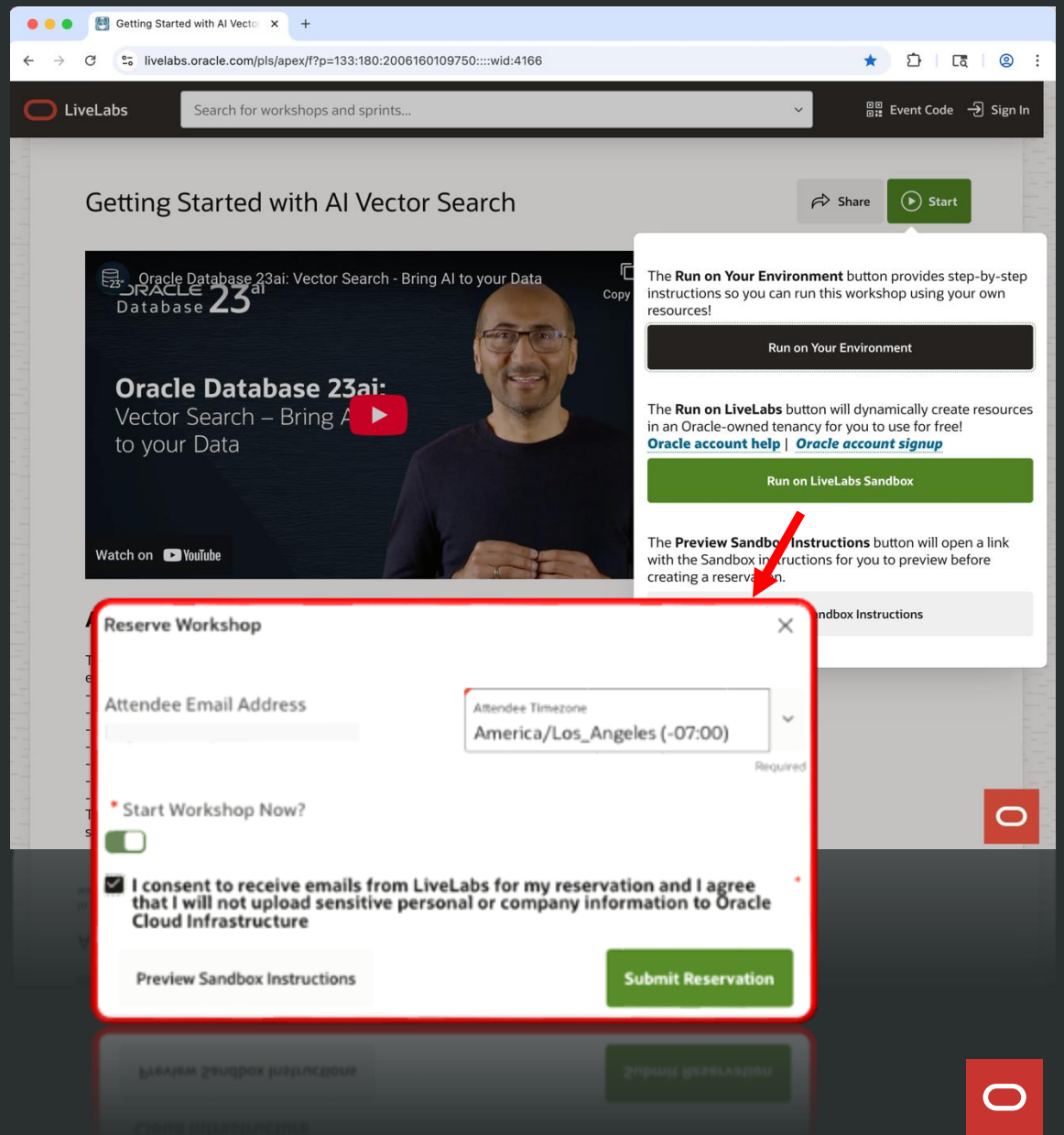
PLEASE NOTE: This is NOT your Oracle Cloud Infrastructure account



Let’s get started

livelabs.oracle.com/ai-world25/HOL2804

1. Click the **Start** button Click **Run on LiveLabs Sandbox**
2. Toggle the **Start Workshop Now** radio button
3. Check the consent box
4. Click **Submit Reservation**



Let’s get started

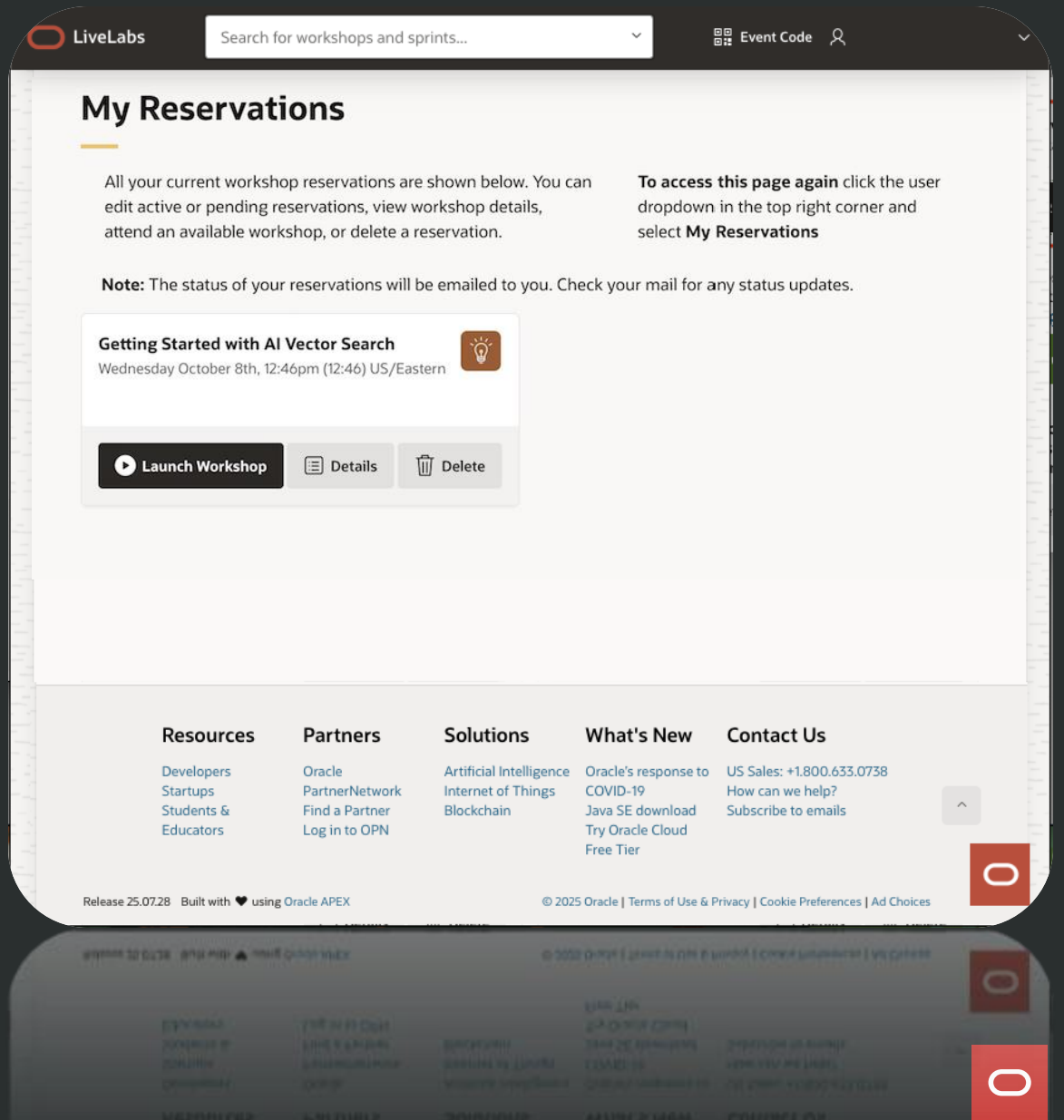
livelabs.oracle.com/ai-world25/HOL2804

Click **Launch Workshop** once available

Please note:

The environments are available within a couple of minutes.

Refresh the page.

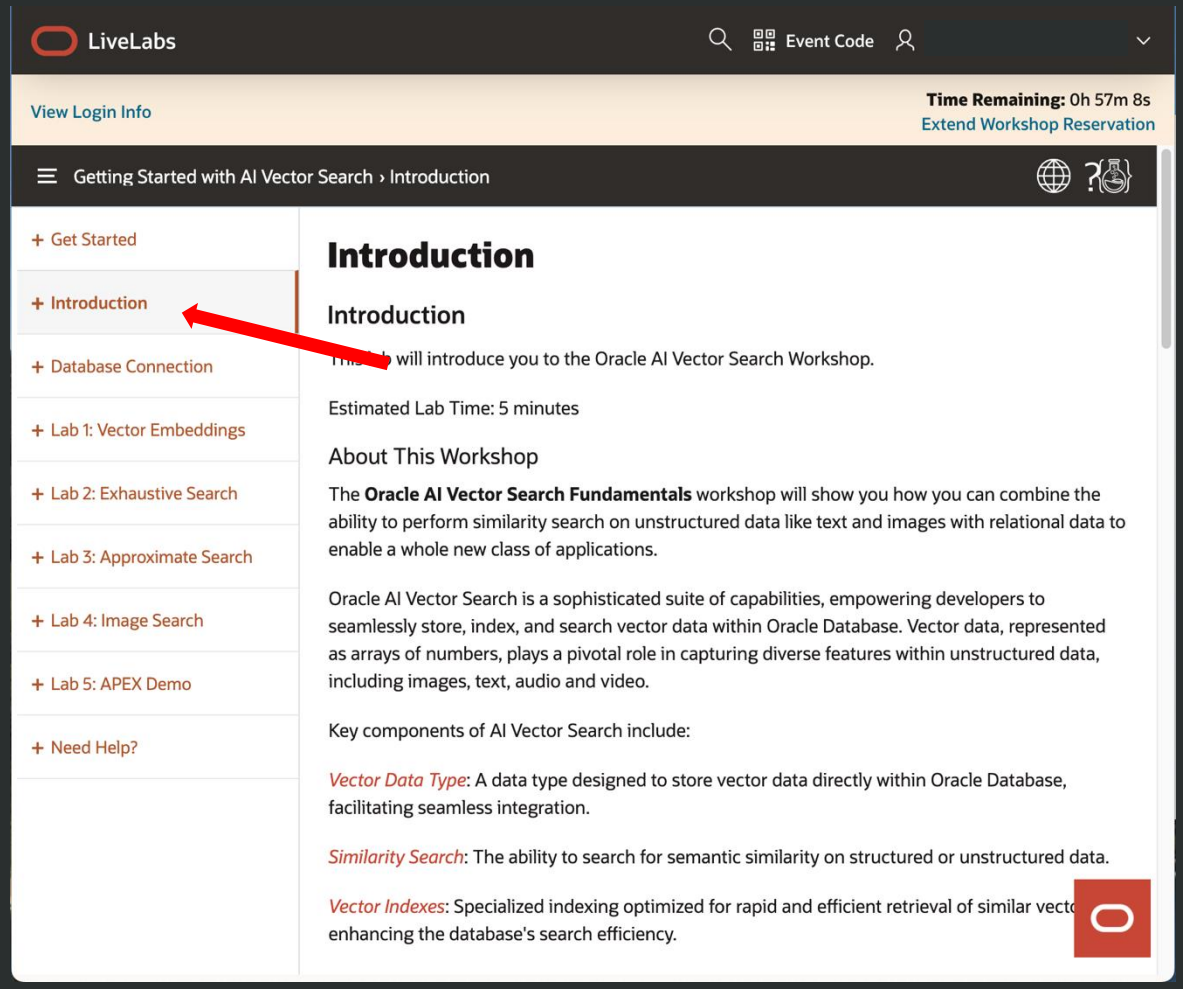


Let’s get started

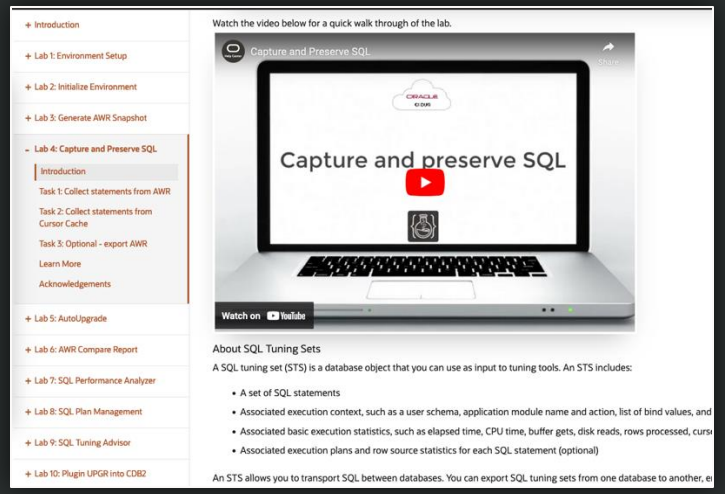
livelabs.oracle.com/ai-world25/HOL2804

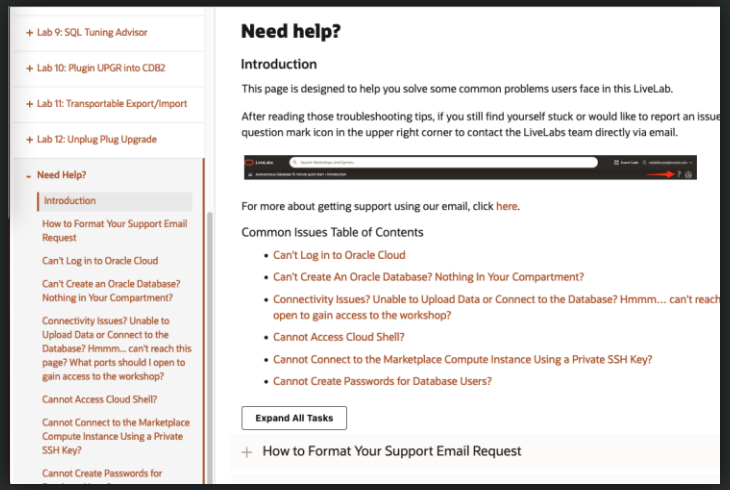
Start with the **Introduction lab** once the workshop is available

Note: Skip the Get Started lab



I need help!





Self-Service

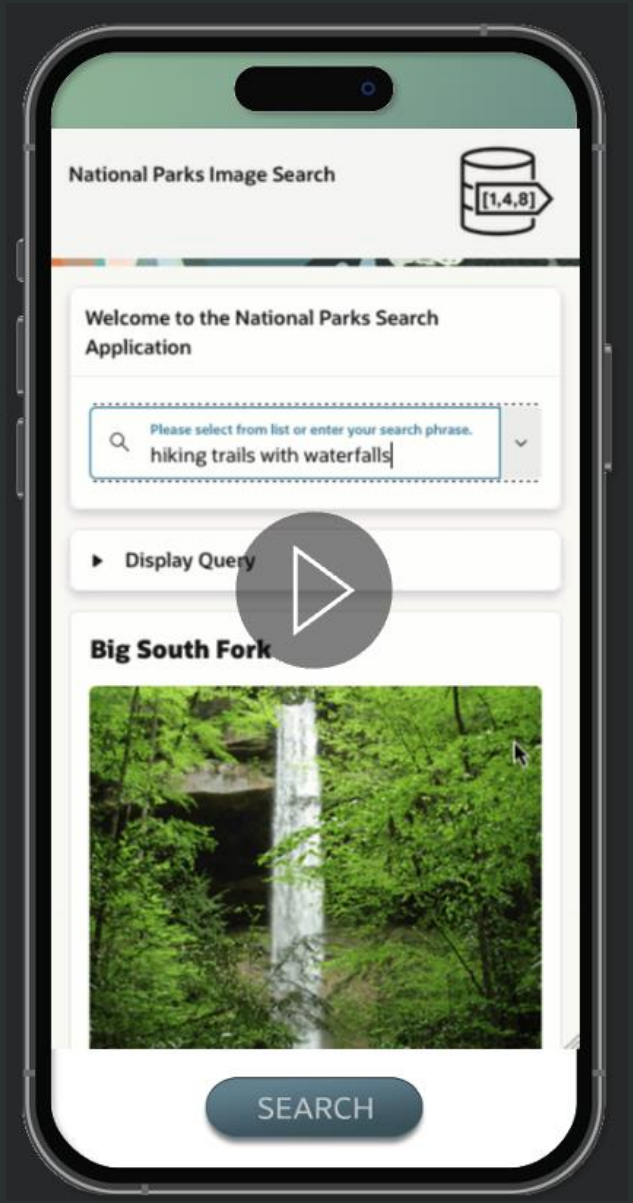
Check the workshop for helpful videos and details

Get help — Self-Service

Find the ‘Need help?’ section in the sidebar of the workshop

Ask us!

AI Vector Search In Action



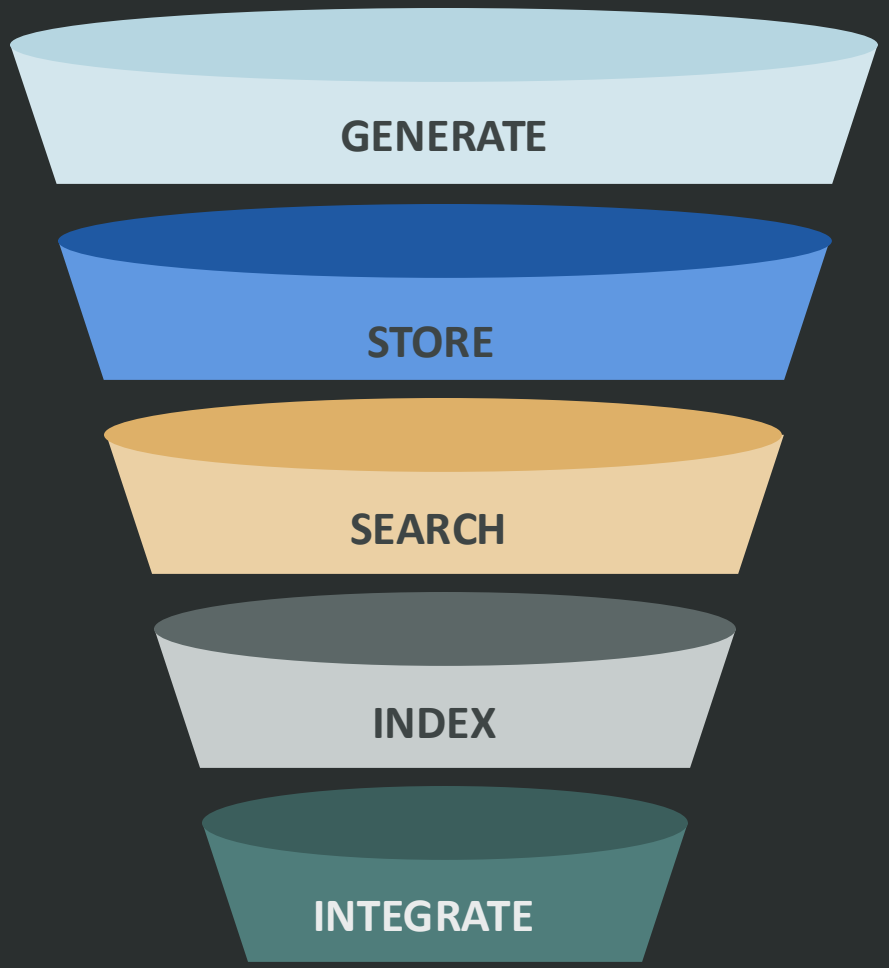
Let's take a look at vector search in action:

Imagine a National Parks **app that helps users find nearby parks featuring similar views or activities to their search criteria**

How AI Vector Search Works

HOL2804 - Getting Started with AI Vector Search

AI Vector Search Highlights



Generate vector embeddings from unstructured data

Store vectors in table columns using new VECTOR type

Perform AI Vector Search on VECTOR columns using SQL

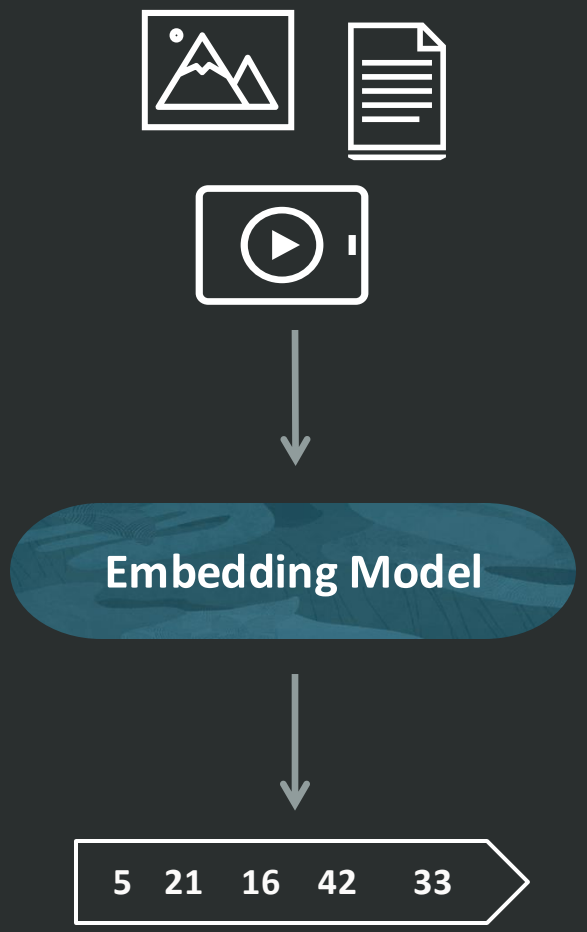
Build approximate vector indexes on VECTOR columns

Integrate with Mission-Critical Enterprise Capabilities

Vector Embeddings

Lab 1 Review

What is Vector Generation or Embedding?



The process of converting unstructured data (text, images, etc.) into vectors

Vectors are generated using deep-learning embedding models

Vector embedding models can be obtained commercially, from Open Source, or can be created

There may be no *BEST* embedding model, and different data may need different models

- No one size fits all
- Evaluate a model based on your data

Vector Embedding Generation | Your Way

AI Vector Search offers 4 alternatives for vector embedding generation

1

Use Pre-created embeddings

2

Use an external embedding cloud-service

3

Use an external embedding library in the mid-tier

4

Use a database resident embedding model

Vector Embedding Generation | Your Way

Load vectors directly from external files

1 Use Pre-created embeddings

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

```
CREATE TABLE ext_vec_table (ID    number,
                            VECT  vector)
   ORGANIZATION EXTERNAL (
      TYPE ORACLE_LOADER
      DEFAULT DIRECTORY DATA_PUMP_DIR
      ACCESS PARAMETERS (
          RECORDS DELIMITED BY NEWLINE
          FIELDS TERMINATED BY ":"
          MISSING FIELD VALUES ARE NULL )
     LOCATION ('vecdata.csv') )
REJECT LIMIT UNLIMITED;
```

Vector Embedding Generation | Your Way

Generate vector embeddings using a callout to a cloud service

2 Use an external embedding cloud-service

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

```
-- Initialize parameters to specify provider
var params CLOB;
DBMS_VECTOR.CREATE_CREDENTIAL(“OPENAI_CRED”, auth_params);
params := '
{
  "provider": ”openai",
  "credential_name": ”OPENAI_CRED",
  "url": https://api.openai.example.com/embeddings”,
  "model": "embed-model"
}'
-- Generate vectors from park descriptions
UPDATE parks
SET desc_vector = DBMS_VECTOR.UTL_TO_EMBEDDING(description, json(params));
```

Vector Embedding Generation | Your Way

Generate vector embeddings using an embedding model on the mid-tier

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

```python
import openai

# Set your OpenAI API key (replace with your actual key)
openai.api_key = 'YOUR_OPENAI_API_KEY'

# Define your text
text = "Creating vector embeddings in Python is straightforward."

# Generate embedding using the OpenAI API
response = openai.embeddings.create(
    input=text,
    model="text-embedding-ada-002"  # Or another suitable model
)

embedding = response.data[0].embedding
print(embedding)
print(f"Embedding dimension: {len(embedding)}")
```

Vector Embedding Generation | Your Way

Generate vector embeddings inside the database

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

```
-- import onnx embedding model
DBMS_VECTOR.load_onnx_model(
  directory => <database directory>
  file_name => 'all_MiniLM_L12_v2.onnx'
  model_name => 'minilm_l12_v2',
  metadata =>  <source>
);


-- generate vectors from park descriptions
UPDATE parks
SET desc_vector = VECTOR_EMBEDDING(minilm_l12_v2 USING description AS data);
```

VECTOR Datatype to store and process vectors

New VECTOR datatype

```
CREATE TABLE parks(
      park_id VARCHAR2,
      description VARCHAR2,
      desc_vector VECTOR(768, FLOAT32));
```

Optional
# of dimensions

Optional
format

Format for dimension values can be FLOAT32, FLOAT64, and INT8

Additional formats coming ...

Vector Operations

Insert

`TO_VECTOR()` converts a string representing an array of vector dimensions into a native VECTOR

```
CREATE TABLE parks(
     park_id          VARCHAR2,
     description      VARCHAR2,
     desc_vector      VECTOR(3, 'FLOAT32'));

INSERT INTO parks VALUES
(1, 'Valley Forge'  TO_VECTOR('[1.1, 2.2, 3.3]');
```

Fetch

`FROM_VECTOR()` converts a vector into a CLOB or VARCHAR2 – default behavior for pre-23ai clients

```
SELECT desc_vector
FROM   parks;
```

↓

```
SELECT FROM_VECTOR(desc_vector)
FROM parks;
```

↓

```
      '[1.1, 2.2, 3.3]'
```


Exhaustive Search

Lab 2 Review

Exhaustive Similarity Search for top-K matches will be 100% accurate but slow as data volumes grow

Vector Search SQL | Distance Function

- The main operation on vectors is to find how similar they are

```
VECTOR_DISTANCE(VECTOR1, VECTOR2, <distance metric>)
```

The Distance between two vectors is smaller for entities that are more similar

Distance functions supported are:

COSINE (Default), EUCLIDEAN, EUCLIDEAN_SQUARED, HAMMING, MANHATTAN, DOT

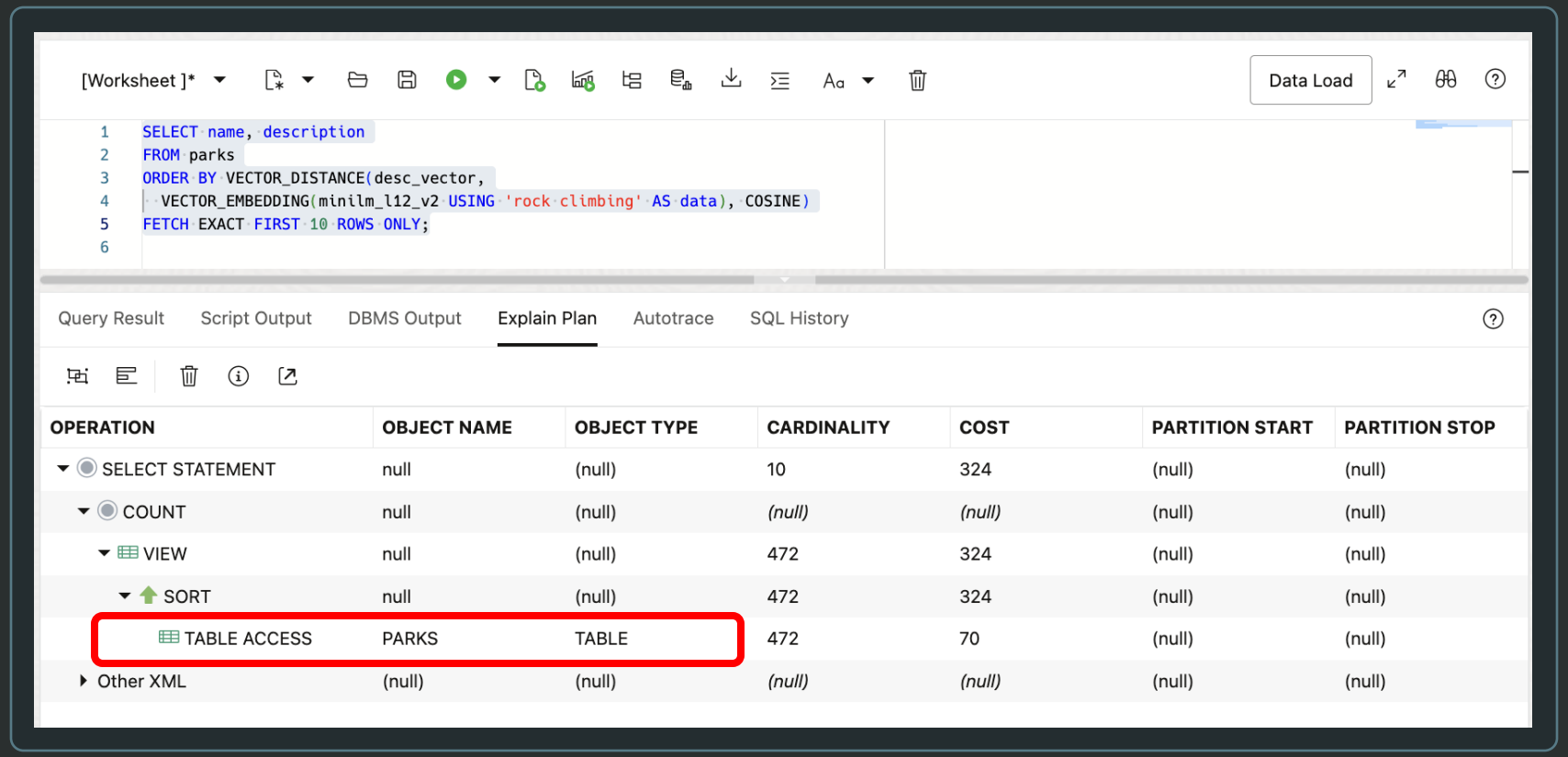
Vector Search SQL | Specifying an Exhaustive Similarity Search

Parks DESCRIPTION Search Example

Find the top 10 matching descriptions

```
SELECT …
FROM   parks
ORDER BY VECTOR_DISTANCE(desc_vector,
  VECTOR_EMBEDDING(minilm_l12_v2 USING 'rock climbing' AS data), COSINE)
FETCH EXACT FIRST 10 ROWS ONLY;
```


Exhaustive Similarity Search | Execution Plan



OPERATION	OBJECT NAME	OBJECT TYPE	CARDINALITY	COST	PARTITION START	PARTITION STOP
SELECT STATEMENT	null	(null)	10	324	(null)	(null)
COUNT	null	(null)	*(null)*	*(null)*	(null)	(null)
VIEW	null	(null)	472	324	(null)	(null)
SORT	null	(null)	472	324	(null)	(null)
TABLE ACCESS	PARKS	TABLE	472	70	(null)	(null)
Other XML	(null)	(null)	*(null)*	*(null)*	(null)	(null)

Approximate Search

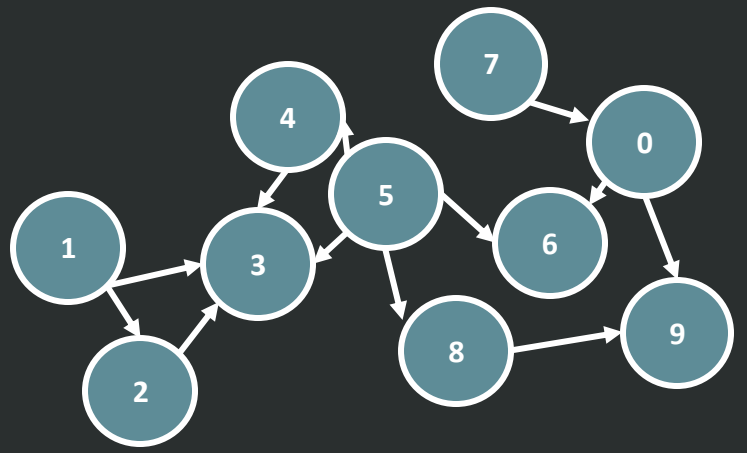
Lab 3 Review

An exhaustive search for top-K matches will be 100% accurate but slow as data volumes grow

New vector indexes trade-off some search accuracy for up to100x speed up

Vector Indexes | Neighbor Graph Vector Index

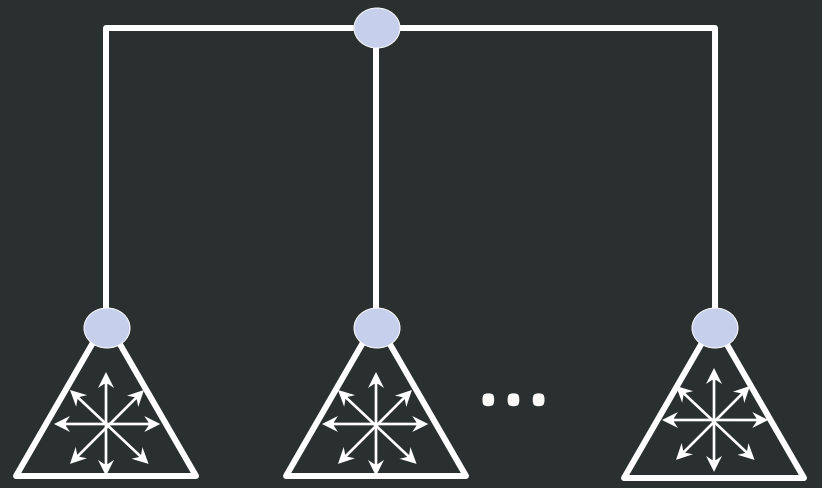
- Graph-based index where vertices represent vectors and edges between vertices represent similarity
- In-Memory only index - highly efficient for both accuracy and speed



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

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



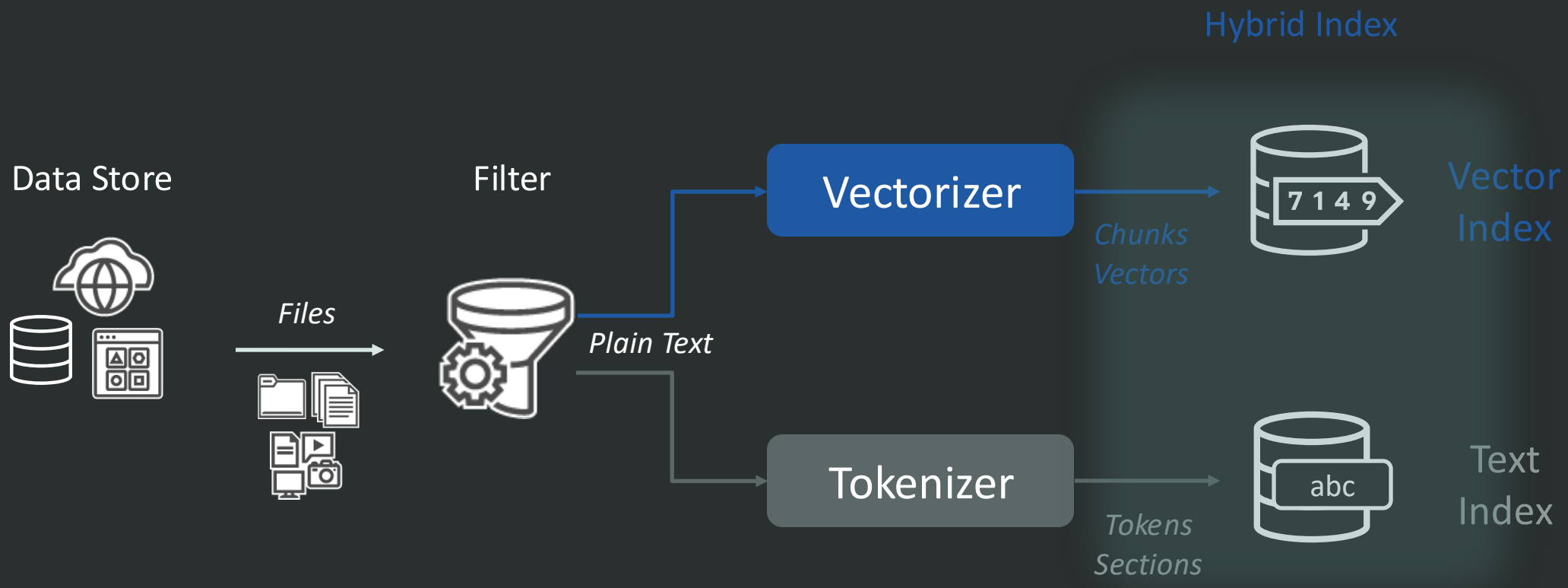
Partition Vector Index (e.g. **IVF_FLAT** index)

Vector Index | Hybrid Vector Index

Indexing Pipeline

Hybrid Vector Index is a single index for searching by **Similarity** and by **Keywords**

- Combines the power of AI Vector Search with Text Search



Vector Index Creation

Basic index creation syntax:

```
CREATE VECTOR INDEX parks_hnsw_idx ON parks(desc_vector)
ORGANIZATION INMEMORY NEIGHBOR GRAPH
DISTANCE COSINE WITH TARGET ACCURACY 95;
```

Choosing the ORGANIZATION for an index is simple:

- If the index data will fit in memory, it is best to use an INMEMORY NEIGHBOR GRAPH index

The DISTANCE clause is optional (the default is **COSINE**)

As a rule of thumb, the distance function should be chosen based on the embedding model used to generate the vectors

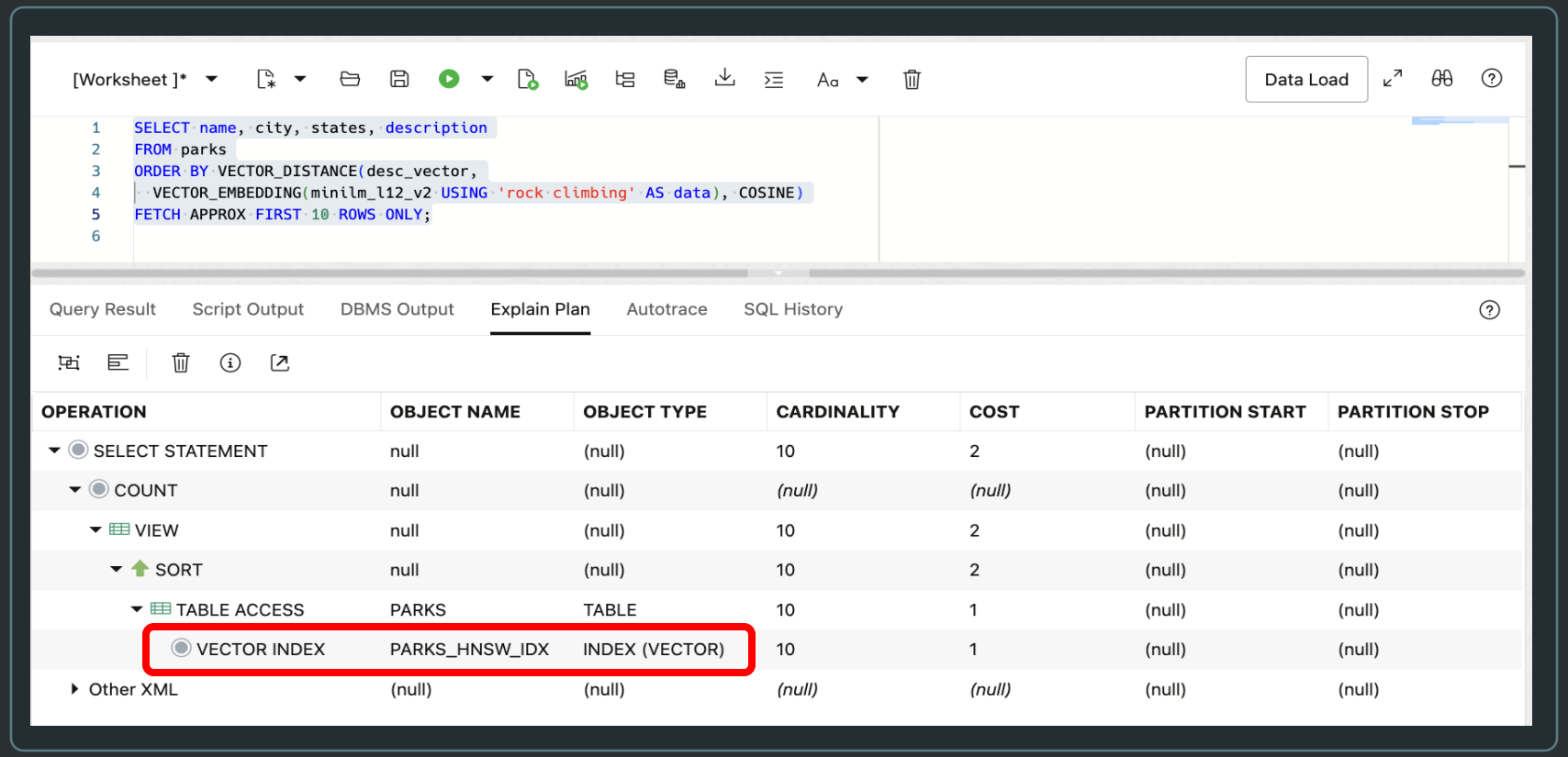
Vector Search SQL | Specifying an Approximate Similarity Search

Parks DESCRIPTION Search Example

Find the top 10 matching descriptions

```
SELECT …
FROM   parks
ORDER BY VECTOR_DISTANCE(desc_vector,
  VECTOR_EMBEDDING(minilm_l12_v2 USING 'rock climbing' AS data), COSINE)
FETCH APPROX FIRST 10 ROWS ONLY;
```


Approximate Similarity Search | Execution Plan



OPERATION	OBJECT NAME	OBJECT TYPE	CARDINALITY	COST	PARTITION START	PARTITION STOP
SELECT STATEMENT	null	(null)	10	2	(null)	(null)
COUNT	null	(null)	*(null)*	*(null)*	(null)	(null)
VIEW	null	(null)	10	2	(null)	(null)
SORT	null	(null)	10	2	(null)	(null)
TABLE ACCESS	PARKS	TABLE	10	1	(null)	(null)
VECTOR INDEX	PARKS_HNSW_IDX	INDEX (VECTOR)	10	1	(null)	(null)
Other XML	(null)	(null)	*(null)*	*(null)*	(null)	(null)

Image Search

Lab 4 Review

AI Vector Search works on all types of unstructured data

- AI Vector Search can search on all types of unstructured data
- Previous labs searched on TEXT based data using text searches
- The Image Search lab searched IMAGE based vector embeddings using text phrases
- Depending on the embedding model, the search could also have used an image to search for an image of a waterfall, rock climbing, etc.

Vector Search SQL | Image Based Similarity Search

```
SELECT …
FROM   park_images
ORDER BY VECTOR_DISTANCE(image_vector,
VECTOR_EMBEDDING(clip_vit_txt USING
'Civil War' AS data), COSINE)
FETCH EXACT FIRST 10 ROWS ONLY;
```

Vector Search SQL | Combine value-based and semantic search

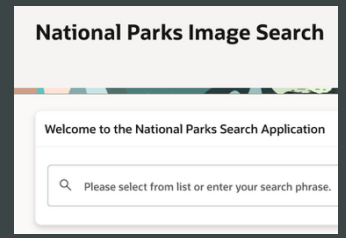
Value-based Attribute Filters can be combined seamlessly with Vector Search in SQL

Find the top 10 matching parks with waterfalls in the Western U.S.

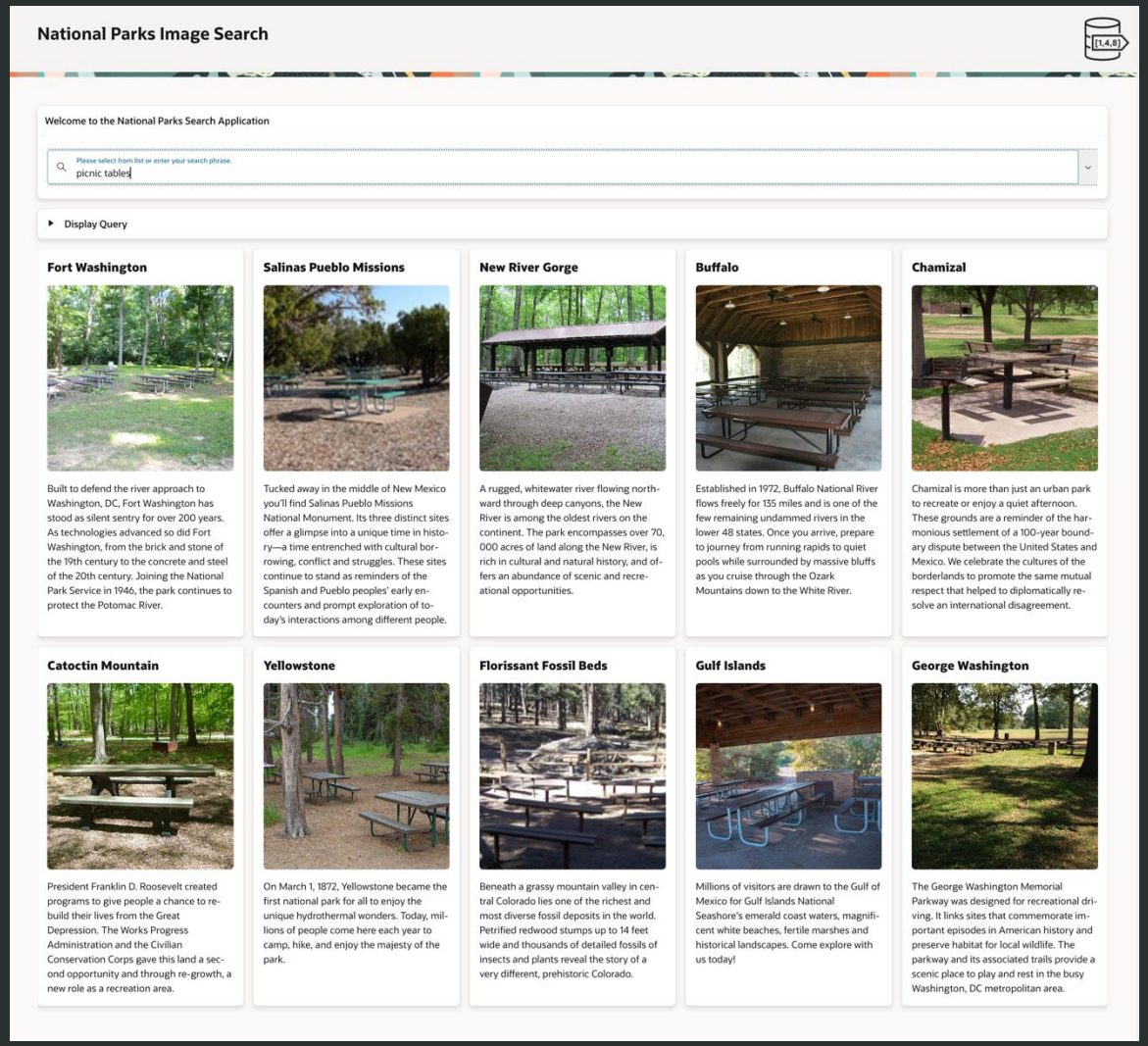
```
SELECT …
FROM   park_images pi, parks p
WHERE  pi.park_code = p.park_code AND p.states IN
('CA','OR','NV','WA','AZ','CO')
ORDER BY VECTOR_DISTANCE(pi.image_vector,
VECTOR_EMBEDDING(clip_vit_txt USING
'waterfall' AS data), COSINE)
FETCH EXACT FIRST 10 ROWS ONLY;
```


APEX Demo

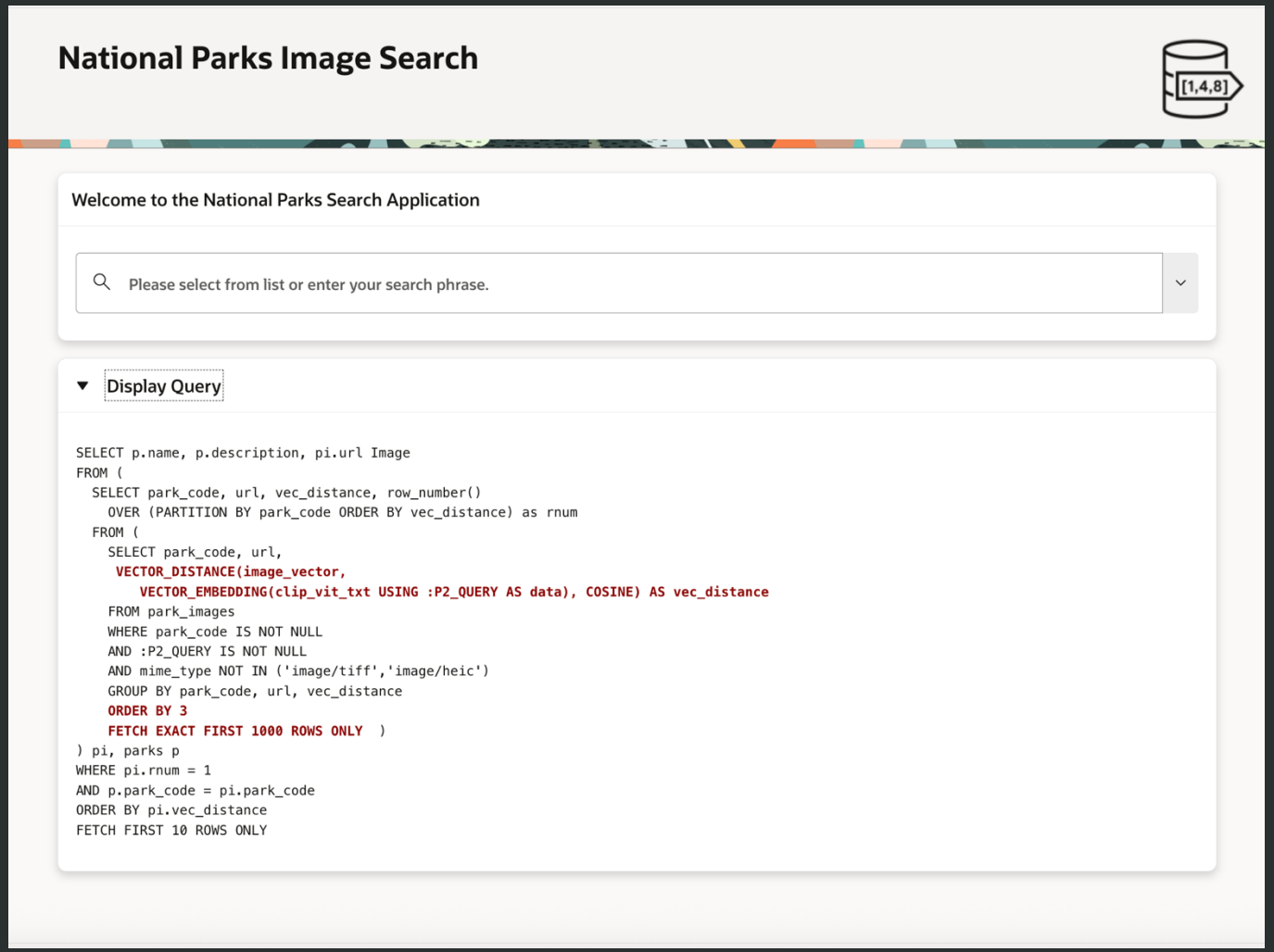
Lab 5 Review



National Parks Image Search with AI Vector Search



Query Used for the National Parks Image Search



Integrate Vectors

HOL2804 - Getting Started with AI Vector Search

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

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

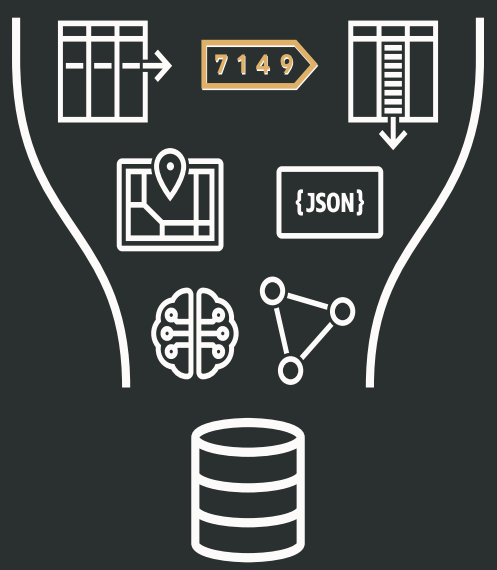
- Real-Application Cluster
- Parallel SQL
- Transactions
- Security
- Analytics
- Disaster Recovery

Wrap Up and Next Steps

HOL2804 - Getting Started with AI Vector Search

Key Takeaways

Oracle brings AI vector search to your business data accelerating the development of AI-enabled applications



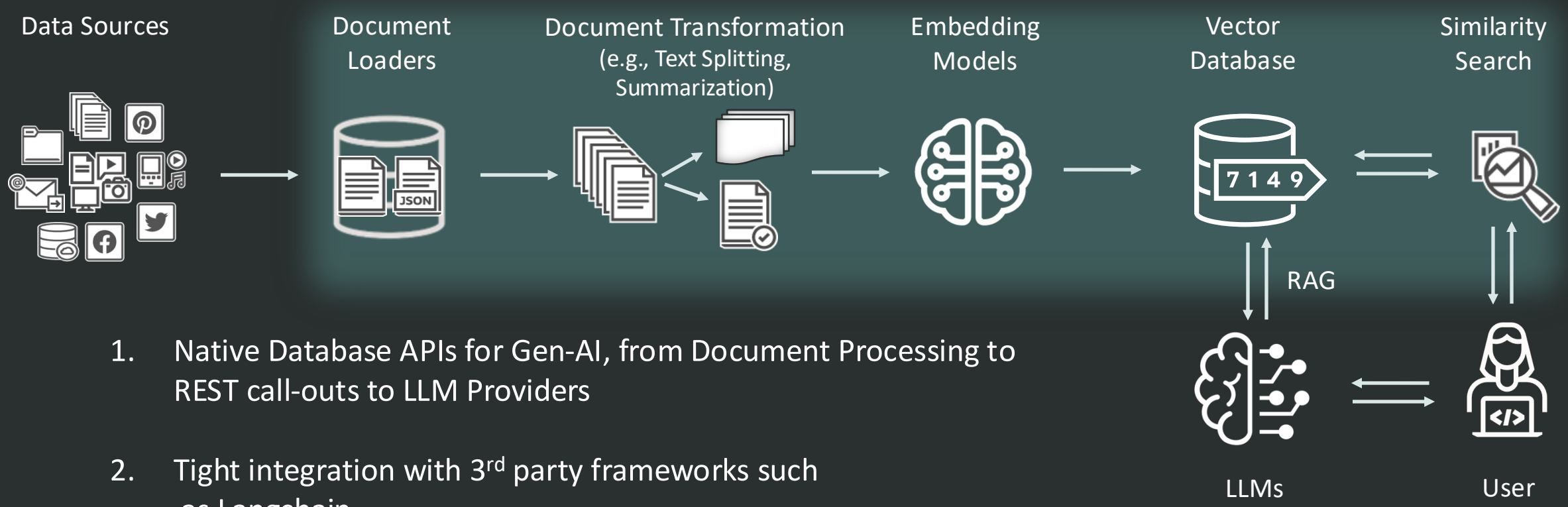
NEW Vector data type for storing vectors

NEW SQL syntax and functions express similarity search with ease

NEW vector indexes for high-performance

AI Vector Search powers Complete Gen AI pipeline

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

To learn more and try Oracle AI Vector Search yourself

AI Vector Search
on Oracle.com

oracle.com/database/ai-vector-search/

Oracle
Vector Search Blog

blogs.oracle.com/database/category/db-vector-search

Try AI Vector Search
on Oracle LiveLabs

Livelabs.oracle.com

Oracle LiveLabs

Showcasing how Oracle's solutions solve your business problems — available 24/7 for free!

15 million

people have already visited LiveLabs

1300+

free workshops available

2200+

events run using LiveLabs workshops

Your feedback is important.

Scan this QR Code or use the Mobile App to share your thoughts on this session.


ORACLE