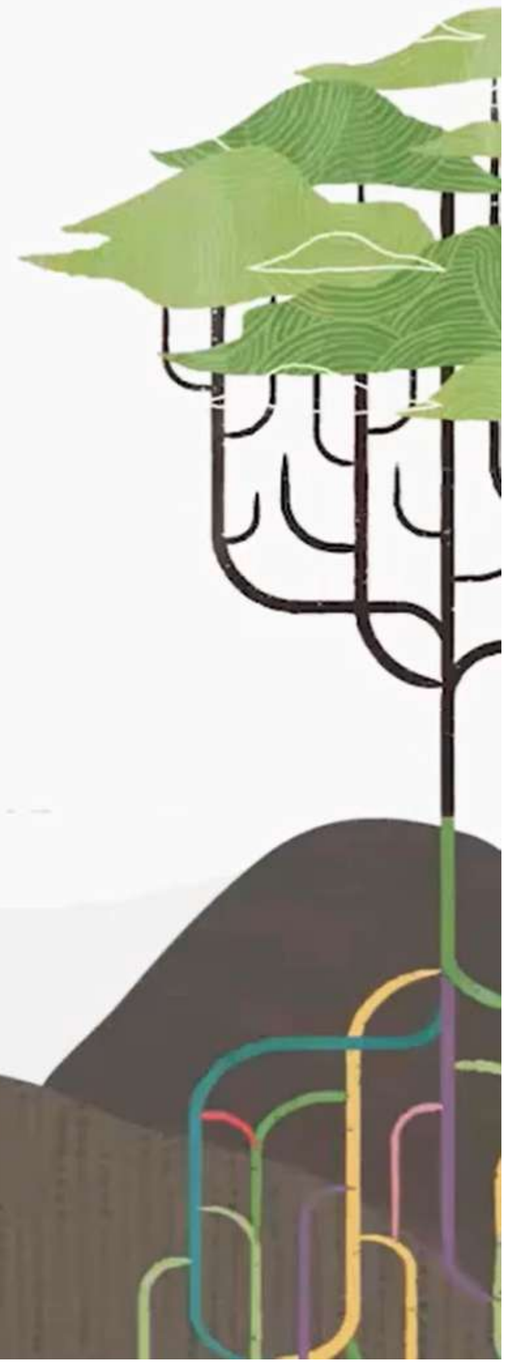
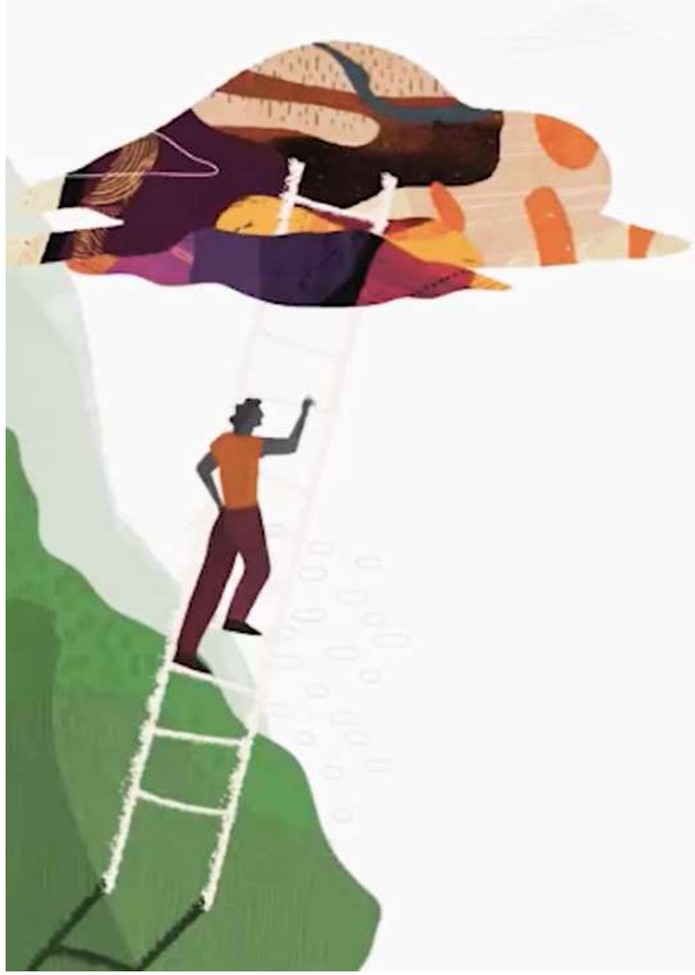

University
Michelle Malcher
DIRECTOR, PRODUCT MANAGEMENT
ORACLE
0:11 / 8:36
1x

ORACLE
University

AI Vector Search
Oracle Database 23ai



Agenda



AI Vector Search in Oracle Database 23ai

AI Vector Search powers Gen AI pipelines

Oracle AI Vector Search

Oracle Database 23ai

- SQL support for vector generation
- Vector Data Type
- Similarity search with SQL syntax and functions
- Approximate search indexes

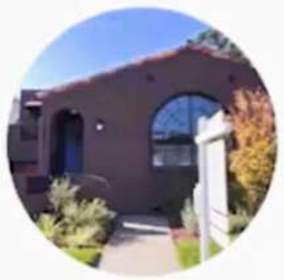
Database-Native Vector Embedding Generation

Generate vector from query image

```
VECTOR_EMBEDDING(resnet_50 USING query_image)
```

Image dataset already embedded into vectors and stored in Oracle Database 23ai

```
VECTOR_EMBEDDING(resnet_50 USING data_image)
```



Load image as BLOB

Query Image Embedding Generation

Vector embedding

23ai

Vector Search for similar matches

Top 2 results

Vector Datatype

New VECTOR datatype

```
CREATE TABLE my_images (
      id          NUMBER,
      data_image BLOB,
      image_vec   VECTOR(768, FLOAT32));
```

Optional dimension count

Optional dimension format

Dimension format can be:
`INT8, FLOAT32, and FLOAT64`

Simpler VECTOR specification

```
CREATE TABLE my_images (
      id          NUMBER,
      data_image BLOB,
      image_vec   VECTOR);
```

Why is this useful?

You can embed your data with newer ML embedding models as AI technology evolves, but your schema can stay the same, and applications don't need to be rewritten

Vector Distance Function

The key operation is vector distance computation to gauge similarity

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

Different embedding models can use different metrics, but the basic concept remains the same:

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

Distance functions supported in Oracle Database 23ai are:

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

Vector Search SQL

Vector Search is used to find Top K closest matches to a given query item

Find the top 10 positions matching a candidate's resume that are offered in a city among the candidate's preferred cities

```
SELECT …
FROM   Job_Postings
WHERE  city IN (SELECT preferred_cities FROM Applicants …)
ORDER BY vector_distance(job_desc_vectors, :resume_vector)
FETCH FIRST 10 ROWS ONLY;
```

Vector Index Syntax

Basic index creation

```
CREATE VECTOR INDEX photo_idx ON customer(photo_vec)
ORGANIZATION [INMEMORY NEIGHBOR GRAPH | NEIGHBOR PARTITIONS]
DISTANCE COSINE | EUCLIDEAN | MANHATTAN | ... WITH TARGET ACCURACY 90
```

`ORGANIZATION` for an index is based if it will fit in-memory:

- If the index data will fit in-memory, use `INMEMORY NEIGHBOR GRAPH`
- Else use `NEIGHBOR PARTITIONS`

`TARGET ACCURACY` clause is added to indicate the default accuracy the index should provide for similarity search queries

Vector Search

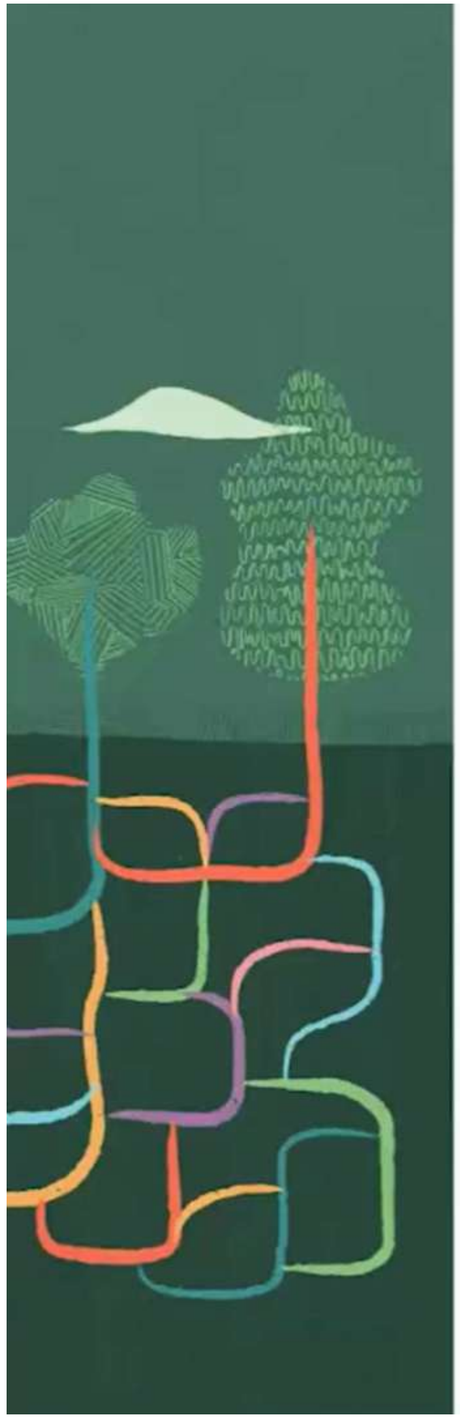
Specifying approximate searches

A new `APPROXIMATE` keyword in the Row Limiting (`FETCH`) clause indicates that the user wants to perform a similarity search using a Vector Index

Find the top 5 customers by similarity with a search photo vector:

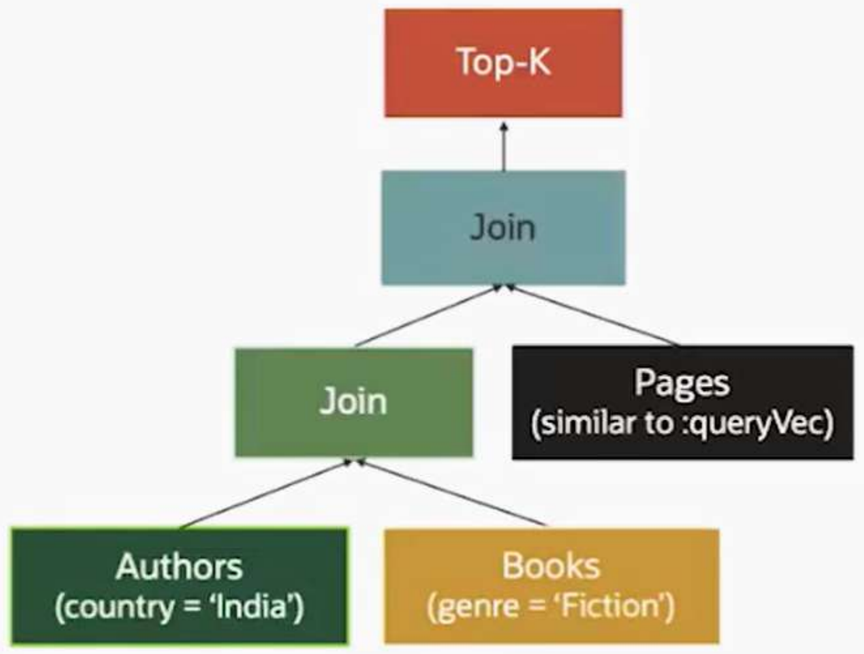
```
SELECT id, name, photo
FROM   Customers
ORDER BY VECTOR_DISTANCE(photo_vec, :QUERY_VEC)
FETCH APPROXIMATE FIRST 5 ROWS ONLY;
```

Exact search may be performed if there is no index or based on the SQL Optimizer costs



Similarity Search Over Joins

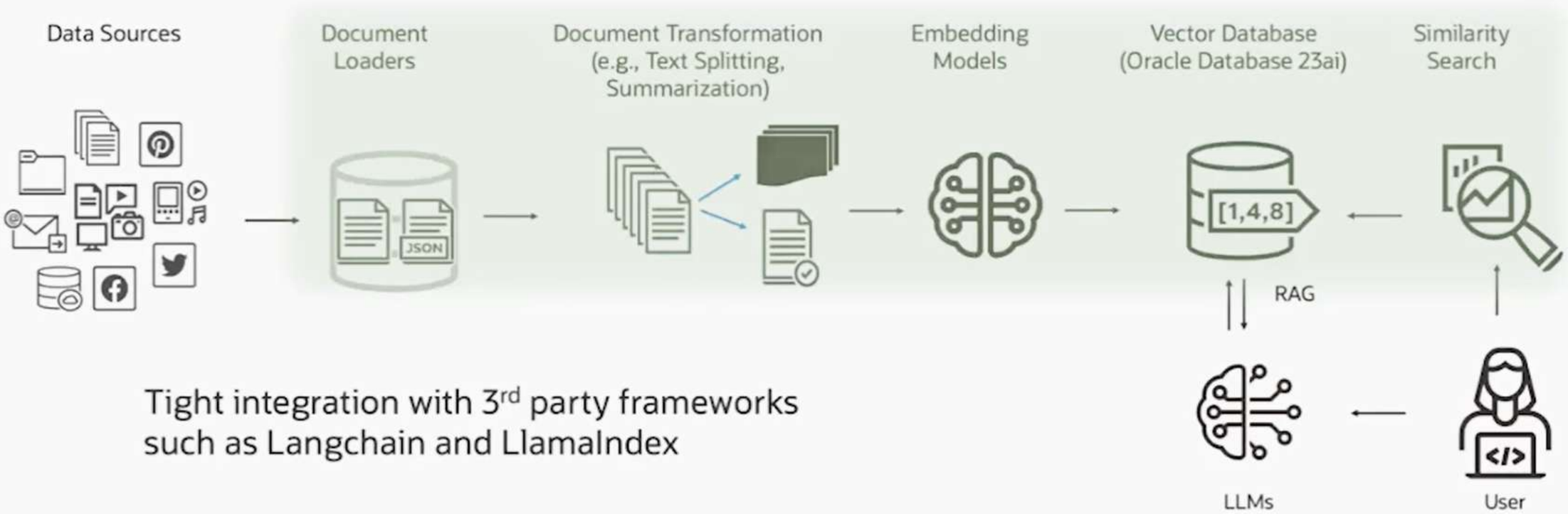
- Join with normalized enterprise data
- 3 Tables: Authors, Books, and Pages
- Pages table has each page individually embedded into vector column pageVec
- Return the top 5 books containing text similar to this query text
- Genre is 'Fiction'
- Author comes from 'India'




```
SELECT pageID FROM Authors, Books, Pages
WHERE  Authors.authorID     = Books.authorID
AND    Books.bookID         = Pages.bookID
AND    Books.genre          = 'Fiction'
AND    Author.country       = 'India'
ORDER BY vector_distance(pageVec, :queryVec)
FETCH APPROX FIRST 5 ROWS ONLY;
```

AI Vector Search powers Gen AI pipelines

AI Vector Search in Oracle Database 23ai



Tight integration with 3rd party frameworks such as Langchain and LlamaIndex

Simplest way to get answers
about your business

Key Takeaways

Simplest way to get answers about your business

Key Takeaways

- AI Vector Search is seamlessly integrated with Oracle Database 23ai
- Simple to combine relational and AI vector Search in single query
- Efficient orchestration of Gen-AI pipelines