



Oracle ACE Virtual Conference

April 2, 7:00 am – 9:40 am PDT

Register for this free event:

<https://apex.oracle.com/go/oracleacevirtualconference>

Streaming Vector Data with Oracle GoldenGate 23ai



Gilson Martins
Oracle ACE Pro

Gilson Martins

CEO & Founder | OGG Admin & DBA
GGBR Solutions
April 2nd, 2025





Become an Oracle ACE

Join 500+ technical experts helping peers globally



Members get many benefits:

- Direct **access** to product development
- Global **networking** events
- Exclusive **training** and content
- Cool **swag**, **cloud credits**, and Oracle CloudWorld pass
- **Your profile** listed on oracle.com

And much more! Go to: ace.oracle.com



ACE Member Benefits



Key Benefits

Cool **swag***, Digital awards for social media, Oracle **CloudWorld pass***, & more



Exclusive Content

Exclusive **monthly virtual meetings** with product development teams + engaging guest speakers



Direct Access to Product Management

Multiple **direct communication channels** to product management and fellow ACEs



Networking

In-person & virtual networking opportunities for ACEs to **connect with product development** and each other.



Cloud Account

\$5k USD Cloud account*



Travel Support

ACE Directors are eligible for travel support to give presentations or lead workshops at conferences globally



Relational Database

X

Vector Database

Relational Database (RDBMS)

- ✓ Stores structured data in tables (rows & columns)
- ✓ Uses SQL for precise data retrieval
- ✓ Strong schema: defined data types & relationships
- ✓ Ideal for transactional systems (ERP, CRM, Finance)

Example:

Table: customers (id, name, phone)

Query: `SELECT * FROM customers WHERE id = 123;`

ID	Title	Category
1	How to replicate data across databases in real time	Replication
2	GoldenGate enables real-time data synchronization	Replication
3	Exporting a table using Oracle Data Pump	Backup
4	What is the difference between hot and cold backup?	Backup
5	Techniques for syncing databases without downtime	Replication

```
SELECT * FROM documents WHERE Category = 'Replication';
```

Returns records **1, 2, and 5** — only because they **explicitly have "Replication"** in the category column.

Vector Database (Vector DB)

- ✓ Stores vectors that represent meaning (text, image, audio)
- ✓ Used for **semantic search** and **AI applications**
- ✓ Finds "similar" items, not exact matches
- ✓ Ideal for chatbots, recommendation engines, RAG, ML

Example:

Vector: embedding of "How to replicate data in real time?"

Query: find vectors **similar to a question**, even with different words

ID	Text	Embedding (vector)
1	How to replicate data across databases in real time?	[0.12, -0.45, 0.33, ...]
2	GoldenGate enables real-time data synchronization.	[0.11, -0.44, 0.35, ...]
3	Exporting a table using Oracle Data Pump	[0.55, 0.10, -0.32, ...]
4	What is the difference between hot and cold backup?	[0.18, -0.20, 0.05, ...]
5	Techniques for syncing databases without downtime	[0.13, -0.46, 0.31, ...]

```
SELECT *  
FROM documents  
WHERE COSINE_DISTANCE(embedding, VECTOR_EMBEDDING('How to sync data')) < 0.2;
```

Returns **documents 1, 2, and 5**, because they are **semantically similar** to the query "How to sync data" — even though the words don't exactly match.

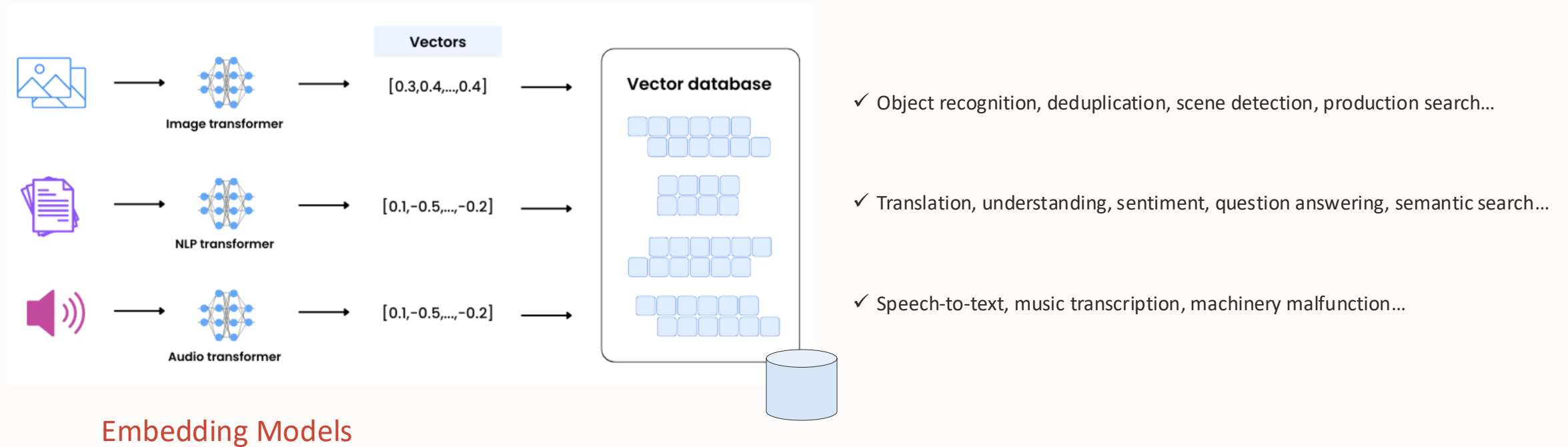
○ Key Difference

Relational.: Structured, exact match

Vector.....: Semantic, similarity-based

Vector Embedding

Vector Embedding



Vector Embedding

Step 1: Tokenization (words, word pieces, or even characters — depending on the model)

Example:

"How to replicate data in real time?"

After tokenization:

["How", "to", "replicate", "data", "in", "real", "time", "?"]

Step 2: Initial Embedding of Tokens

Example:

"replicate" → [0.34, -0.12, 0.85, ..., 0.07] # A vector of 768 dimensions

After tokenization:

```
[  
  [ 0.12, -0.34, 0.08, ..., 0.01 ], # "How"  
  [ 0.05, 0.10, -0.02, ..., 0.03 ], # "to"  
  [ 0.34, -0.12, 0.85, ..., 0.07 ], # "replicate"  
  [ 0.28, 0.03, -0.19, ..., 0.12 ], # "data"  
  [ 0.02, -0.06, 0.14, ..., -0.01 ], # "in"  
  [ 0.25, 0.18, -0.09, ..., 0.06 ], # "real"  
  [ 0.40, -0.11, 0.22, ..., 0.09 ], # "time"  
  [ 0.01, 0.00, -0.03, ..., 0.00 ] # "?"  
]
```

Step 3: Processing by the Model (Transformer)

Context: how each word relates to the others in the sentence

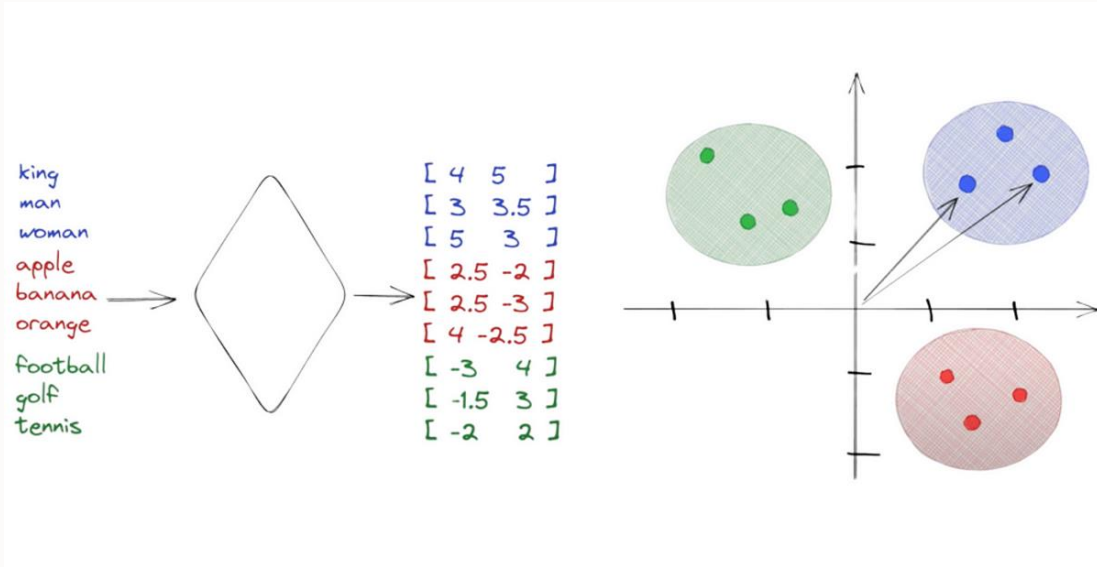
Position: the order and structure of the sentence

Semantic relationships: how token meanings interact (e.g., "replicate" and "data")

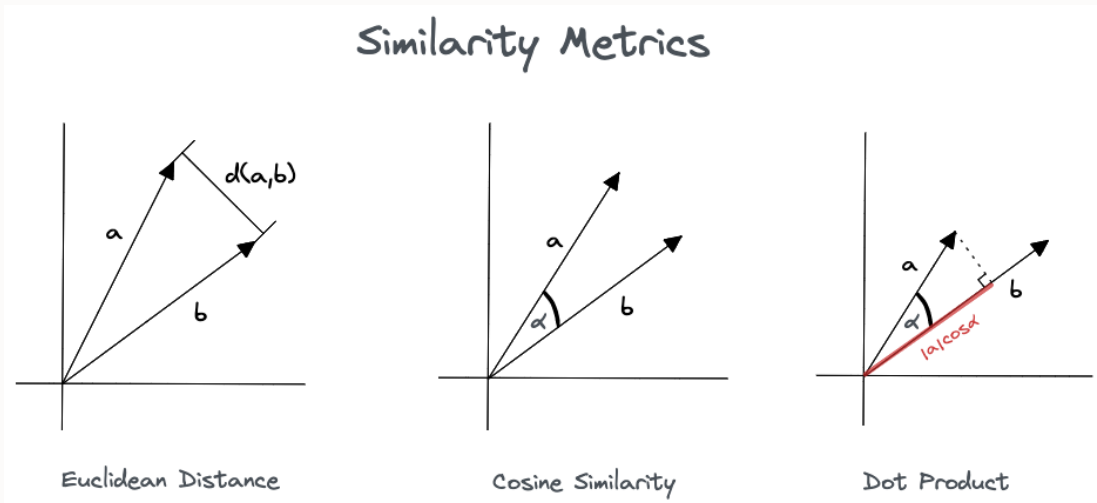
Sentence Embedding:

[0.29, -0.03, 0.41, ..., 0.07] # Represents the meaning of the FULL SENTENCE

Vector Embedding (2D Example)



Model	Embedding Dimension
all-MiniLM-L6-v2	384
all-MiniLM-L12-v2	384
all-mpnet-base-v2	768
text-embedding-ada-002 (OpenAI)	1536
bert-base-nli-mean-tokens	768



Function	Focus	Typical Range	Best For...
COSINE_DISTANCE	Direction (angle)	0 (equal) \rightarrow 1 (opposite)	Semantic comparison, RAG
L2_DISTANCE	Absolute distance	0 \rightarrow ∞	Pure geometric analysis
DOT_PRODUCT	Projection / affinity	$-\infty \rightarrow \infty$	Direct similarity ranking or affinity

Vector Embedding

Database / Oracle / Oracle Database / Release 23

Oracle AI Vector Search User's Guide

Prerace

- ▶ 1 What's New for Oracle AI Vector Search
- ▶ 2 Overview
- ▶ 3 Get Started
- ▶ 4 Generate Vector Embeddings
- ▶ 5 Store Vector Embeddings
- ▶ 6 Create Vector Indexes and Hybrid Vector Indexes
- ▼ 7 Use SQL Functions for Vector Operations
 - ▼ Vector Distance Functions and Operators
 - ▼ Vector Distance Metrics
 - Euclidean and Euclidean Squared Distances
 - Cosine Similarity
 - Dot Product Similarity
 - Manhattan Distance
 - Hamming Distance
 - Jaccard Similarity
 - VECTOR_DISTANCE
 - L1_DISTANCE

Vector Distance Metrics

Measuring distances in a vector space is at the heart of identifying the most relevant results for a given query vector. That process is very different from the well-known keyword filtering in the relational database world.

When working with vectors, there are several ways you can calculate distances to determine how similar, or dissimilar, two vectors are. Each distance metric is computed using different mathematical formulas. The time it takes to calculate the distance between two vectors depends on many factors, including the distance metric used as well as the format of the vectors themselves, such as the number of vector dimensions and the vector dimension formats. Generally it's best to match the distance metric you use to the one that was used to train the vector embedding model that generated the vectors.

- [Euclidean and Euclidean Squared Distances](#)

Euclidean distance reflects the distance between each of the vectors' coordinates being compared—basically the straight-line distance between two vectors. This is calculated using the Pythagorean theorem applied to the vector's coordinates $(\text{SQRT}(\text{SUM}((x_i - y_i)^2)))$.

- [Cosine Similarity](#)

One of the most widely used similarity metric, especially in natural language processing (NLP), is cosine similarity, which measures the cosine of the angle between two vectors.

- [Dot Product Similarity](#)

The dot product similarity of two vectors can be viewed as multiplying the size of each vector by the cosine of their angle. The corresponding geometrical interpretation of this definition is equivalent to multiplying the size of one of the vectors by the size of the projection of the second vector onto the first one, or vice versa.

- [Manhattan Distance](#)

This metric is calculated by summing the distance between the dimensions of the two vectors that you want to compare.

- [Hamming Distance](#)

The Hamming distance between two vectors represents the number of dimensions where they differ.

- [Jaccard Similarity](#)

The Jaccard similarity is used to determine the share of significant (non-zero) dimensions (bit's position) common between two `BINARY` vectors.

Parent topic: [Vector Distance Functions and Operators](#)

Why Oracle GoldenGate?

Create a real-time vector hub for GenAI

Business Operations

Oracle technologies

- ORACLE Cerner
- ORACLE Fusion Cloud
- ORACLE Hospitality
- ORACLE Retail
- ORACLE Micros
- ORACLE Transportation
- ORACLE JD Edwards
- ORACLE Siebel
- ORACLE E-Business Suite
- ORACLE Database
- ORACLE Exadata
- MySQL

3rd party databases

- IBM DB2
- aws
- Azure
- Google Cloud
- IBM DB2/z
- PostgreSQL
- Microsoft SQL Server
- SAP ASE
- TANDEM NonStop
- hp
- Amazon Aurora

3rd party vector sources

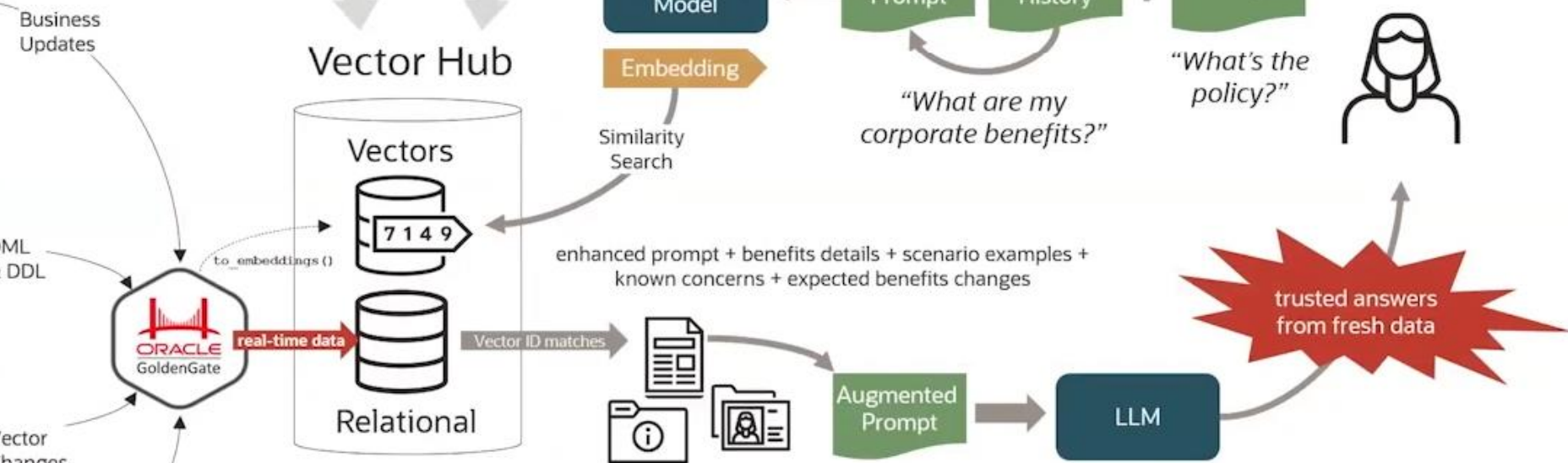
- PostgreSQL
- AlloyDB
- Amazon RDS
- Amazon Aurora
- pgvector

3rd party messaging, NoSQL

- mongoDB
- JMS
- CoAP (B=H)
- MQTT
- kafka connect
- DATASTAX
- RabbitMQ
- elastic
- cassandra
- kafka
- splunk
- Google Analytics
- ActiveMQ
- Subclipse

Real-time Data

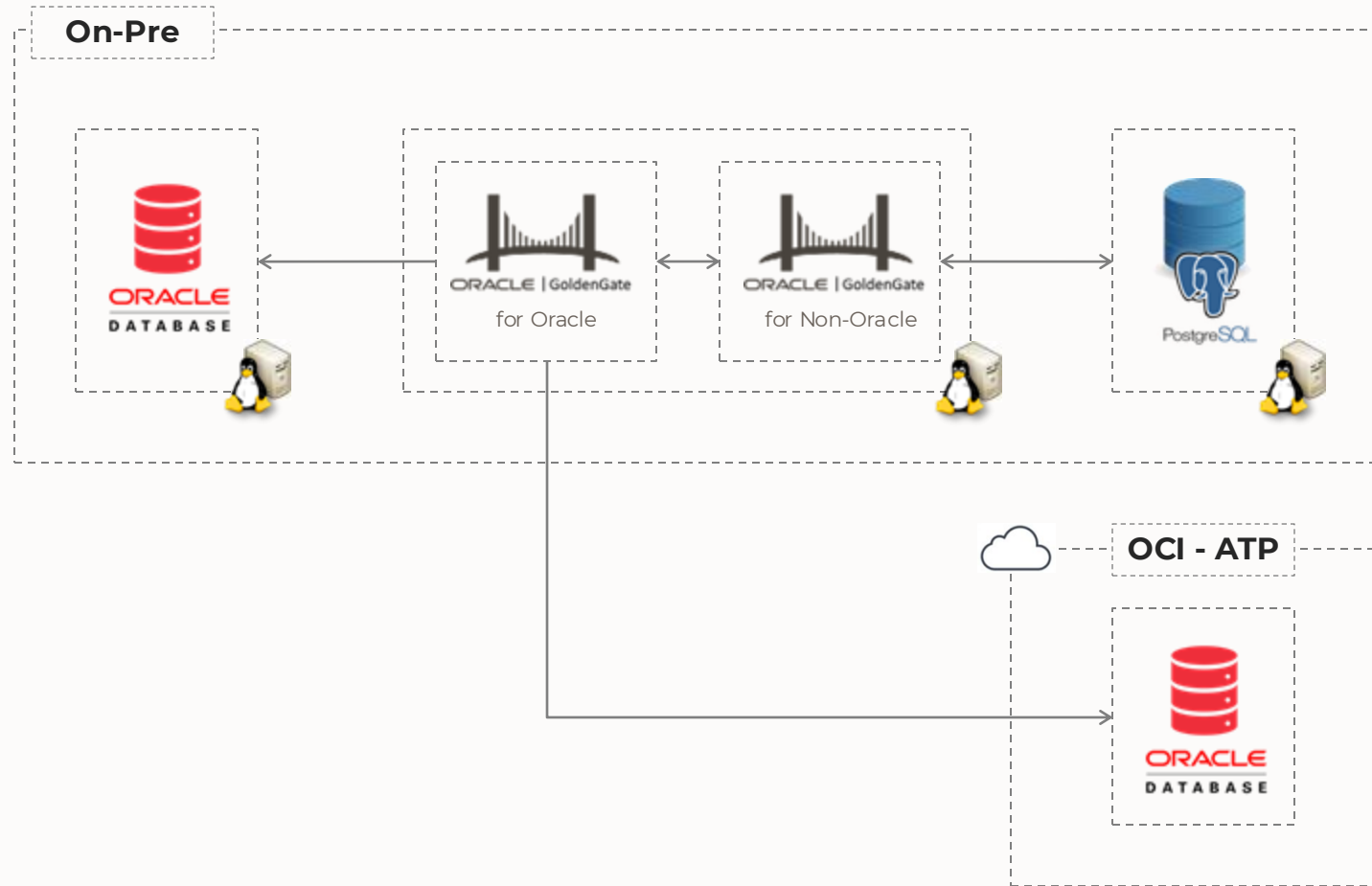
Retrieval Augmented Generation (RAG)



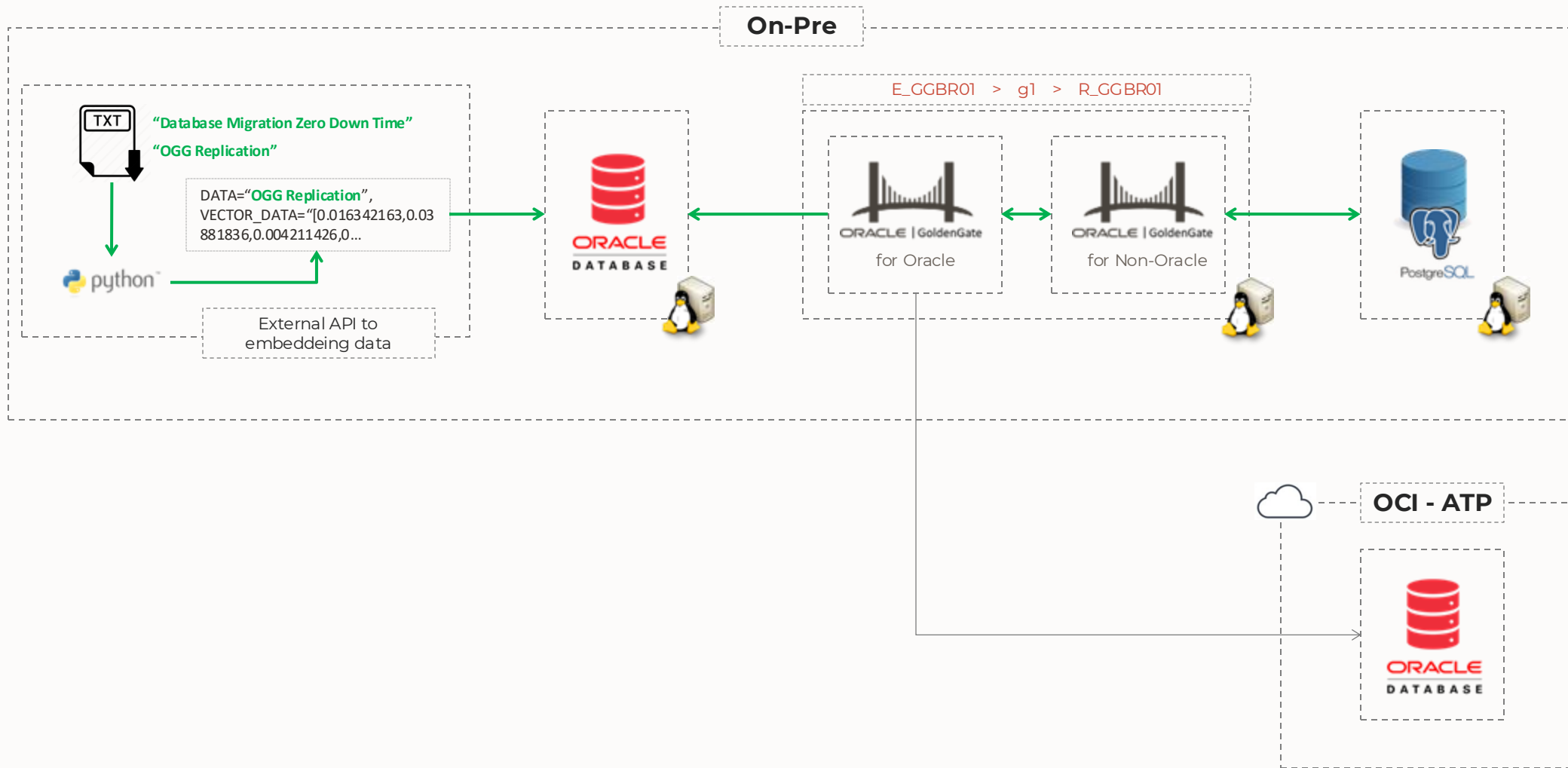
real-time, private content from any GoldenGate source

DEMO

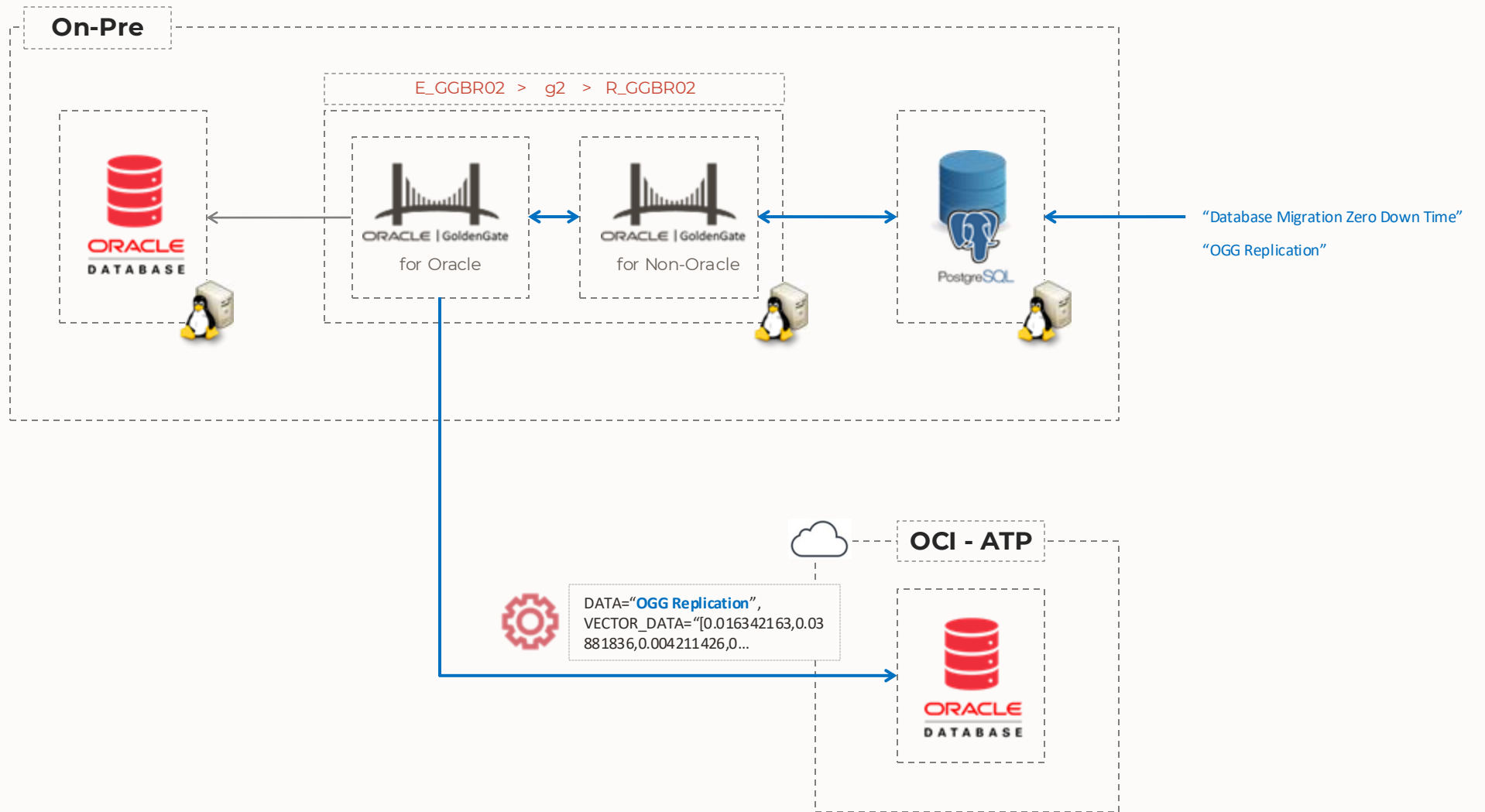
Replication Environment



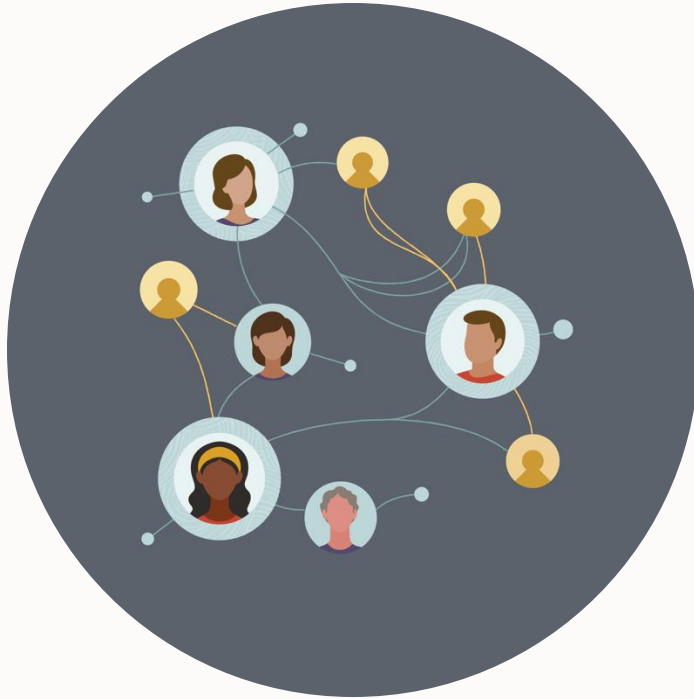
Replication 01



Replication 02



Learn more & connect with the Oracle ACE Program



Program Details

ace.oracle.com

Nomination

ace.oracle.com/nominate

ACEs in Action Blog

blogs.oracle.com/ace

Twitter

@oracleace

Facebook

The Oracle ACE Program

Linkedin

bit.ly/OracleACEs

Email

ACEprogram_ww@oracle.com

Gilson Martins
GGBR Solutions



GGBR
SOLUTIONS

