VISCOSITY
PROFESSIONAL SERVICES

Building Gen AI-Powered Database Applications: A "Developer's" Fast Track to Production

Utah Oracle Users Group (UTOUG)
Virtual Event
Thursday, June 12, 2025

utoug
utah oracle users group

Craig Shallahamer

Oracle ACE Director
Applied AI Scientist | OraPub Founder

Oracle ACE Director
SYMPOSIUM L2

linkedin.com/in/craig-shallahamer-571a94a/
craig.shallahamer@viscosityna.com

1

Why use MySQL EE On-Prem For RAG Apps?

- **AI Ready.** MySQL EE On-Prem now has the capability to help build real AI applications... and it will only get better!
- **Scalable On-Prem**. You can develop on your desktop(s), then scale up **on-prem** as the workload grows!
- **True Database.** MySQL EE is much more than a “vector store” or “document database” because a true database “out of the box” provides high concurrency transaction management, backup and recovery, and security.
- **Super AI Secure**. You can even have your AI app working without an internet connection!

2

Next Steps URLs

HOL: RAG App with MySQL On-Premise (Private and Limited HOL Release)
https://livelabs.oracle.com/****

AI TEAM COLLABORATOR
http://64.23.163.216
username: ****
password: ****

viscosityna.com ORAPUB utoug utah oracle users group

3

Craig Shallahamer

Oracle ACE Director
Applied AI Scientist | OraPub Founder

Oracle ACE Director SYMPOSIUM

@OraPubInc
linkedin.com/in/craig-shallahamer-571a94a/
craig.shallahamer@viscosityna.com

Having worked with **Oracle technology since 1989**, Craig Shallahamer is a leader in the fields of machine learning, artificial intelligence and Oracle database performance tuning.

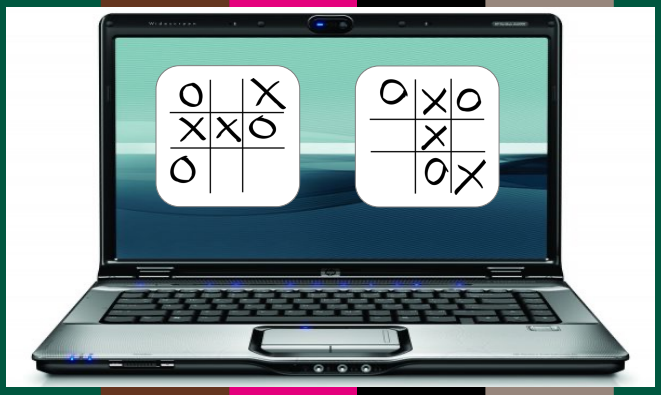
Craig has extensive experience **in constructing and teaching predictive modeling** methods, notably developing a Reinforcement Machine Learning bot in 1990. He has launched several **specialized generative AI conversational assistants**, each with distinct personalities and capabilities. As an Applied AI Scientist at Viscosity and the founder of OraPub, Craig is also the author of two acclaimed books: "Oracle Performance Firefighting" and "Forecasting Oracle Performance".

He has received recognition as an **Oracle ACE Director** for his valuable contributions to the Oracle community through his technical expertise and leadership. Craig is an active participant in Oracle user groups, **frequently presenting** at conferences and serving as a **board member and volunteer**.

ORAPUB POWERED BY VISCOSITY VISCOSITY NoCOUG

4

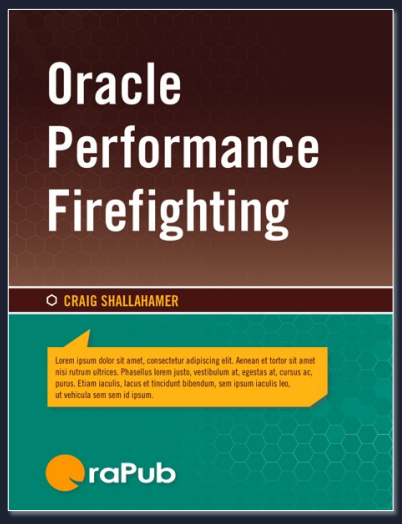
1990



I created a program where I could play the game tic-tac-toe with a computer, but also enabled the computer to play itself… and learning through that process.

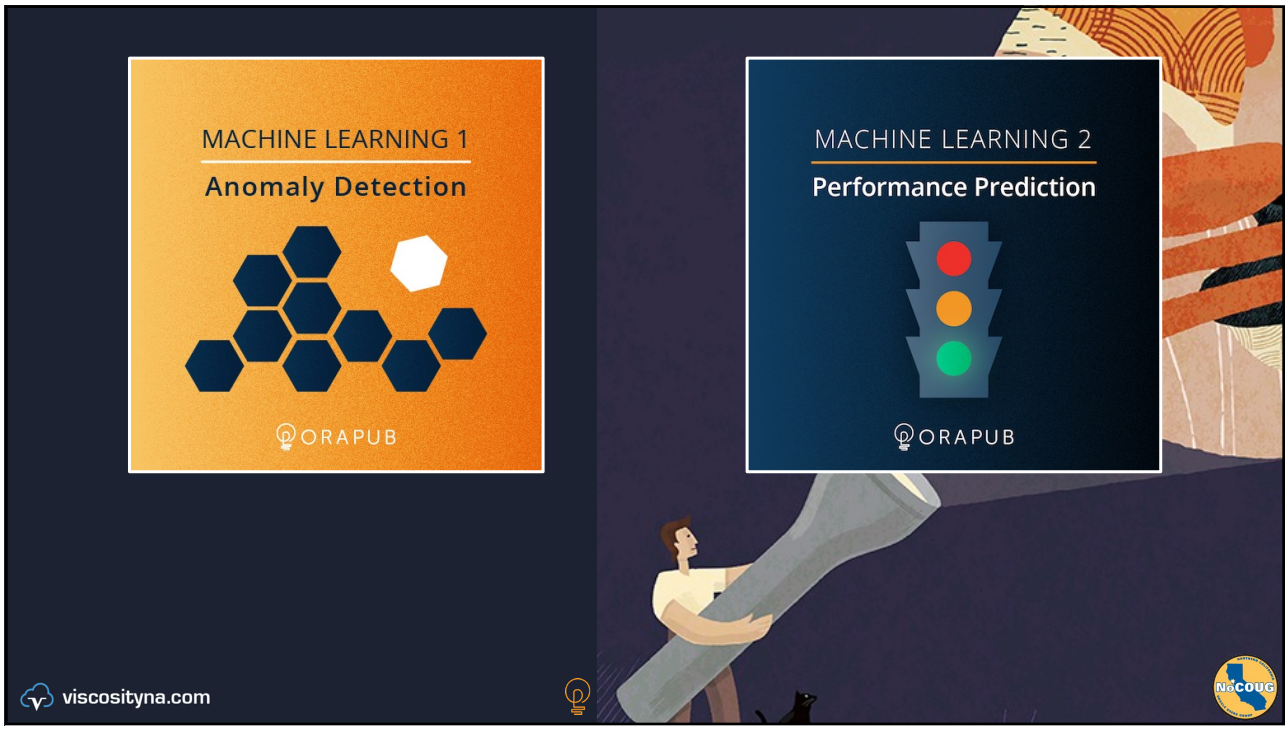
viscosityna.com ORAPUB utoug utah oracle users group

5



viscosityna.com ORAPUB utoug utah oracle users group

6



7

From a more business perspective...

- **Domain specific Generative AI** Q&A Bots, Conversational Assistants, and AI Teams with personality in multiple domains. Such as Oracle Performance Tuning, Biblical Texts, Real estate investing platform, City Zoning Codes.
- **Internet marketing**, focusing on identifying "hot leads" for very specific promotions.
- **IT Operations** when many Oracle systems must be **monitored** and Operations possibly **alerted**; both anomaly detection and performance prediction.

viscosityna.com ORAPUB utoug utah oracle users group

8

Shane

Shane AI: The Oracle Tuning AI Assistant

Leveraging **Oracle Cloud Infrastructure**, AI Shane is a **conversational AI** assistant with a vast curated knowledge base of **all things Oracle performance**. Shane also has a very distinct stereotypic **surfer persona**. His objective is to help anyone **improve** their **Oracle system performance** optimization skills.

AI Shane leverages the **Oracle Generative AI** services within **Oracle Cloud Infrastructure** (OCI).

This includes **LLMs, vector** and traditional capabilities within Oracle Databases.

Specifically, AI Shane can use the Oracle Vector Store, Oracle GenAI Services, Oracle Autonomous DB, Oracle MySQL, Oracle MySQL Heatwave and OCI Compute.

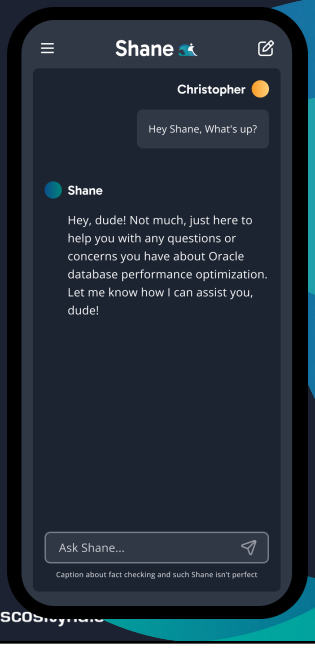
ORAPUB
POWERED BY VISCOSITY

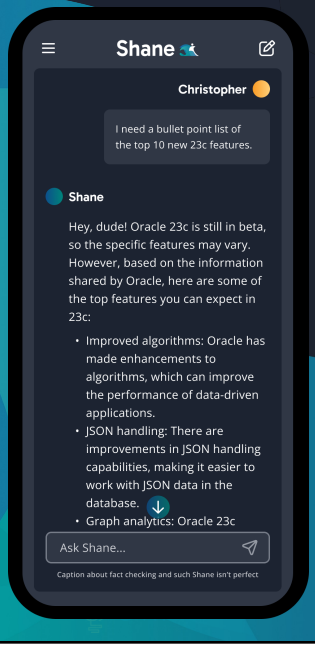
VISCOSITY

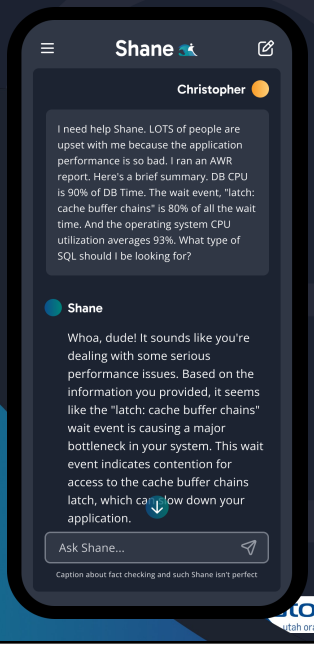
9

ORAPUB
POWERED BY VISCOSITY

VISCOSITY





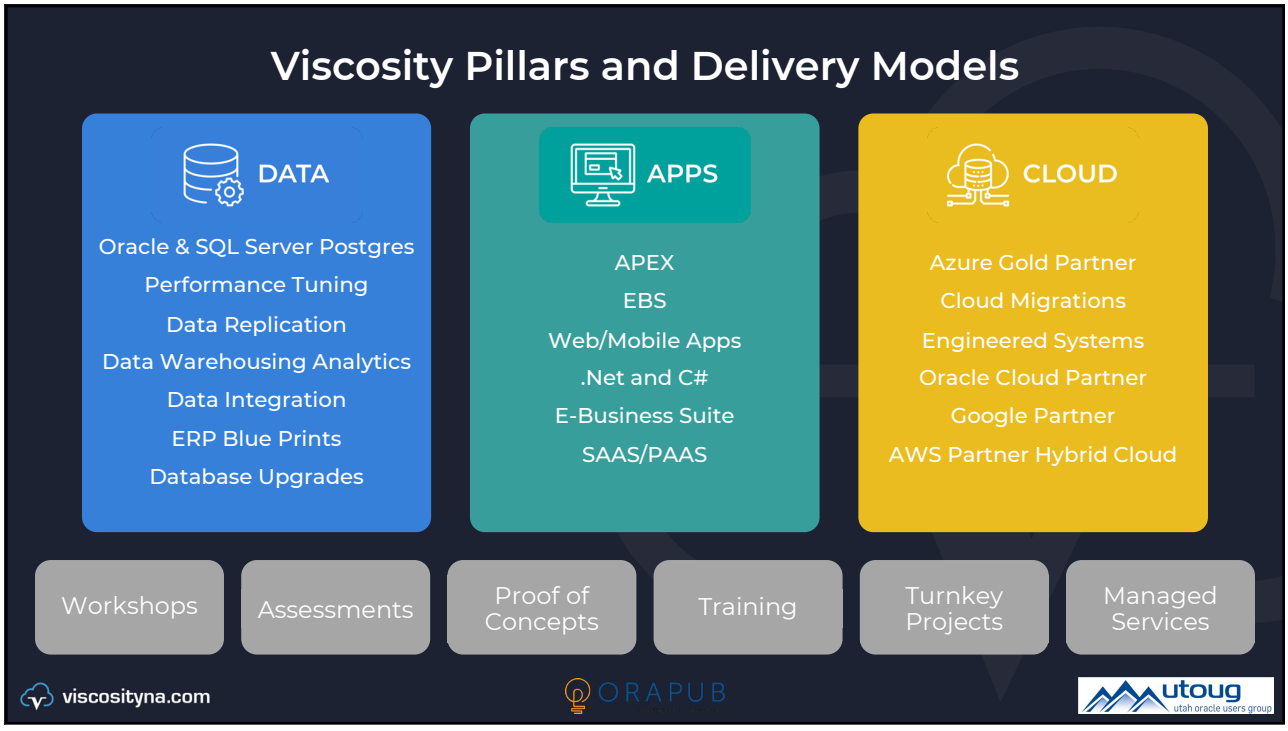


viscosity.com

10


ORAPUB
POWERED BY VISCOSITY
VISCOSITY
https://shane.adalo.com
viscosityna.com
utoug
utah oracle users group

11


Viscosity Pillars and Delivery Models
DATA
Oracle & SQL Server Postgres
Performance Tuning
Data Replication
Data Warehousing Analytics
Data Integration
ERP Blue Prints
Database Upgrades
APPS
APEX
EBS
Web/Mobile Apps
.Net and C#
E-Business Suite
SAAS/PAAS
CLOUD
Azure Gold Partner
Cloud Migrations
Engineered Systems
Oracle Cloud Partner
Google Partner
AWS Partner Hybrid Cloud
Workshops
Assessments
Proof of Concepts
Training
Turnkey Projects
Managed Services
viscosityna.com
ORAPUB
utoug
utah oracle users group

12


ACE Program
Viscosity's Oracle ACEs
The Oracle ACE Program
The Oracle ACE Program recognizes and rewards individuals for their contributions to the Oracle community.
Charles Kim
CEO | Co-Founder
@racdba
ACE Director
Rich Niemiec
Chief Innovation Officer
@richniemiec
ACE Director
Craig Shallahamer
Applied AI Scientist
@orapub
ACE Director
Sean Scott
Principal Consultant
@oraclesean
ACE Director
Gary Gordhamer
Principal Consultant
@ggordham
ACE Director
Julio Ayapan
Senior DBA
@jayapangt
ACE Associate
viscosityna.com
@ViscosityNA

13


We've written over 25 technical books!
Linux and Solaris Recipes for Oracle DBAs
Oracle Database Performance Tuning
Oracle Database 12c Release 2 Performance Tuning Tips and Techniques
Linux Recipes for Oracle DBAs
Oracle Exadata EXPERT'S HANDBOOK
PDB Me to Oracle Cloud
Upgrading Oracle Databases
Quick Start Guide to Oracle Query Tuning: Tips for DBAs and Developers
Forecasting Oracle Performance
Oracle Performance Firefighting
Oracle on Docker
viscosityna.com
@ViscosityNA

14

Topics for today

This presentation is about using
AI-Powered MySQL Enterprise Edition On-Premise

- What does the app do?
- How does the app work?
- Cool things about the app
- Next steps, including the HOL

15

What does the app do?

Ask it ANY movie related question.

If the movie is in the Sakila database
you'll get a great answer!

If not… you will still get a great answer!!

16

Most people think of a large language model (LLM) as one big thing, that can be **used to generate** another thing.

viscosityna.com ORAPUB utoug utah oracle users group

17

But, there is another way to look at this.

An LLM provides **two** fundamental qualities.

It understands **language** and it contains **content:** knowledge, data, context.

For our web app, we want to use **our data** that's in MySQL EE and use the **LLM language capabilities**.

viscosityna.com ORAPUB utoug utah oracle users group

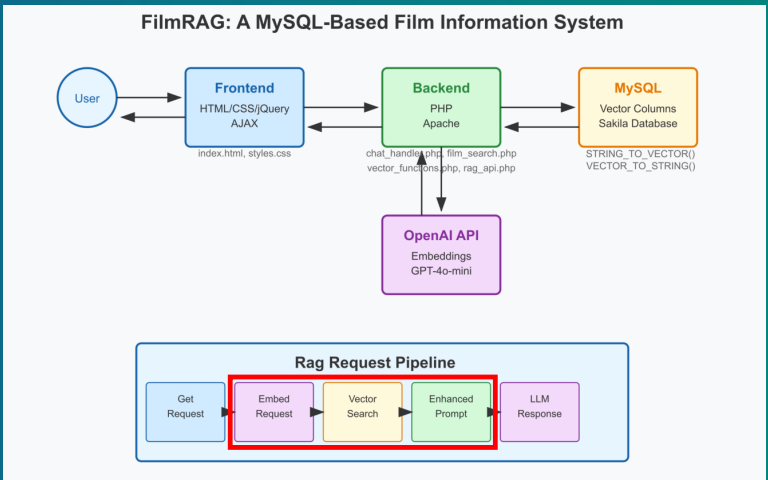
18

How does the app work?

What is RAG?

Retrieval Augmented Generation

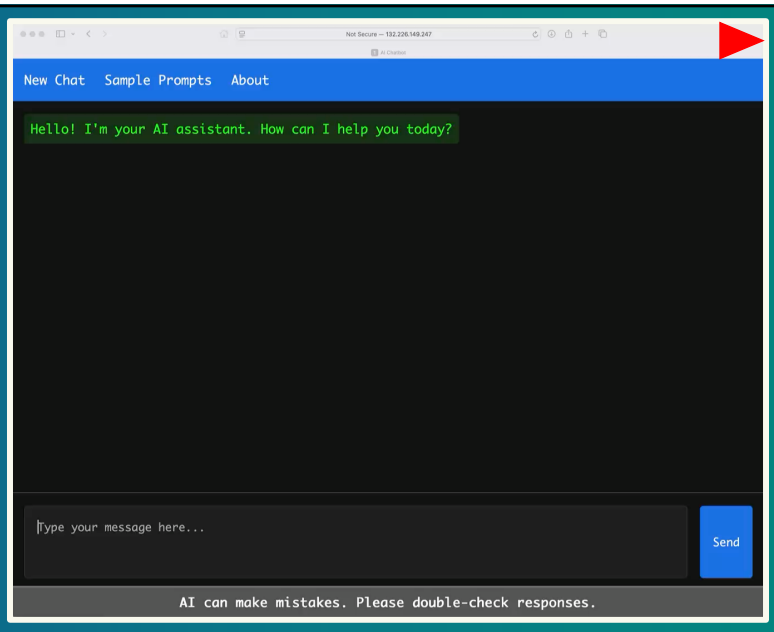
We “augment” the prompt with data from MySQL EE before we give it to the LLM.



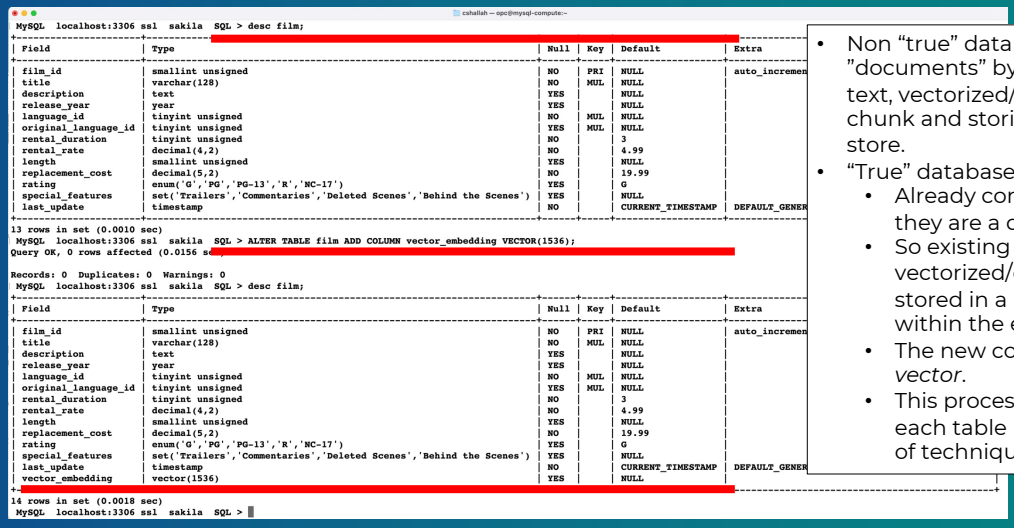
How does the app work?

So, what’s going on under the covers?

It’s a RAG app, so there is “ingestion” and “request”.



RAG Ingestion



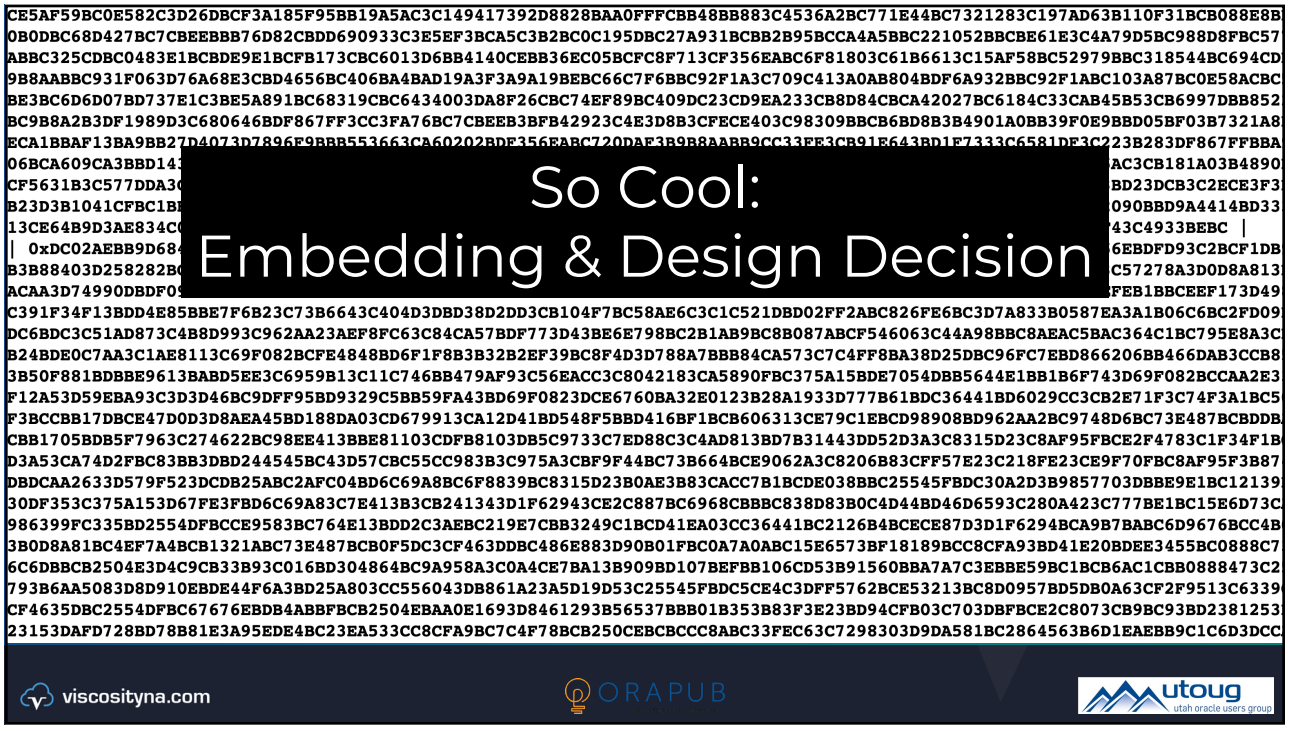
```
MySQL  localhost:3306 ssl  sakila  SQL > desc film;
+----------------------+-----------------------------------------------------------------+------+-----+-------------------+-----------------------------+
| Field                | Type                                                            | Null | Key | Default           | Extra                       |
+----------------------+-----------------------------------------------------------------+------+-----+-------------------+-----------------------------+
| film_id              | smallint unsigned                                               | NO   | PRI | NULL              | auto_increment              |
| title                | varchar(128)                                                    | NO   | MUL | NULL              |                             |
| description          | text                                                            | YES  |     | NULL              |                             |
| release_year         | year                                                            | YES  |     | NULL              |                             |
| language_id          | tinyint unsigned                                                | NO   | MUL | NULL              |                             |
| original_language_id | tinyint unsigned                                                | YES  | MUL | NULL              |                             |
| rental_duration      | tinyint unsigned                                                | NO   |     | 3                 |                             |
| rental_rate          | decimal(4,2)                                                    | NO   |     | 4.99              |                             |
| length               | smallint unsigned                                               | YES  |     | NULL              |                             |
| replacement_cost     | decimal(5,2)                                                    | NO   |     | 19.99             |                             |
| rating               | enum('G','PG','PG-13','R','NC-17')                              | YES  |     | G                 |                             |
| special_features     | set('Trailers','Commentaries','Deleted Scenes','Behind the Scenes') | YES  |     | NULL              |                             |
| last_update          | timestamp                                                       | NO   |     | CURRENT_TIMESTAMP | DEFAULT_GENER               |
+----------------------+-----------------------------------------------------------------+------+-----+-------------------+-----------------------------+
13 rows in set (0.0010 sec)
MySQL  localhost:3306 ssl  sakila  SQL > ALTER TABLE film ADD COLUMN vector_embedding VECTOR(1536);
Query OK, 0 rows affected (0.0156 se

Records: 0  Duplicates: 0  Warnings: 0
MySQL  localhost:3306 ssl  sakila  SQL > desc film;
+----------------------+-----------------------------------------------------------------+------+-----+-------------------+-----------------------------+
| Field                | Type                                                            | Null | Key | Default           | Extra                       |
+----------------------+-----------------------------------------------------------------+------+-----+-------------------+-----------------------------+
| film_id              | smallint unsigned                                               | NO   | PRI | NULL              | auto_increment              |
| title                | varchar(128)                                                    | NO   | MUL | NULL              |                             |
| description          | text                                                            | YES  |     | NULL              |                             |
| release_year         | year                                                            | YES  |     | NULL              |                             |
| language_id          | tinyint unsigned                                                | NO   | MUL | NULL              |                             |
| original_language_id | tinyint unsigned                                                | YES  | MUL | NULL              |                             |
| rental_duration      | tinyint unsigned                                                | NO   |     | 3                 |                             |
| rental_rate          | decimal(4,2)                                                    | NO   |     | 4.99              |                             |
| length               | smallint unsigned                                               | YES  |     | NULL              |                             |
| replacement_cost     | decimal(5,2)                                                    | NO   |     | 19.99             |                             |
| rating               | enum('G','PG','PG-13','R','NC-17')                              | YES  |     | G                 |                             |
| special_features     | set('Trailers','Commentaries','Deleted Scenes','Behind the Scenes') | YES  |     | NULL              |                             |
| last_update          | timestamp                                                       | NO   |     | CURRENT_TIMESTAMP | DEFAULT_GENER               |
| vector_embedding     | vector(1536)                                                    | YES  |     | NULL              |                             |
+----------------------+-----------------------------------------------------------------+------+-----+-------------------+-----------------------------+
14 rows in set (0.0018 sec)
MySQL  localhost:3306 ssl  sakila  SQL >
```

- Non “true” databases ingest ”documents” by chopping up the text, vectorized/embedded each chunk and storing that in a vector store.
- “True” databases like MySQL EE,
  - Already contain data… because they are a database!
  - So existing column(s) are vectorized/embedded and stored in a new column(s) within the existing table.
  - The new column has the type, *vector*.
  - This process is performed for each table row using a variety of techniques.

viscosityna.com ORAPUB utoug utah oracle users group

21



```
CE5AF59BC0E582C3D26DBCF3A185F95BB19A5AC3C149417392D8828BAA0FFFCBB48BB883C4536A2BC771E44BC7321283C197AD63B110F31BCB088E8B
0B0DBC68D427BC7CBEEBBB76D82CBDD690933C3E5EF3BCA5C3B2BC0C195DBC27A931BCBB2B95BCCA4A5BBC221052BBCBE61E3C4A79D5BC988D8FBC57
ABBC325CDBC0483E1BCBDE9E1BCFB173CBC6013D6BB4140CEBB36EC05BCFC8F713CF356EABC6F81803C61B6613C15AF58BC52979BBC318544BC694CD
9B8AABBC931F063D76A68E3CBD4656BC406BA4BAD19A3F3A9A19BEBC66C7F6BBC92F1A3C709C413A0AB804BDF6A932BBC92F1ABC103A87BC0E58ACBC
BE3BC6D6D07BD737E1C3BE5A891BC68319CBC6434003DA8F26CBC74EF89BC409DC23CD9EA233CB8D84CBCA42027BC6184C33CAB45B53CB6997DBB852
BC9B8A2B3DF1989D3C680646BDF867FF3CC3FA76BC7CBEEB3BFB42923C4E3D8B3CFECE403C98309BBCB6BD8B3B4901A0BB39F0E9BBD05BF03B7321A8
ECA1BBAF13BA9BB27D4073D7896E9BBB553663CA60202BDF356EABC720DAF3B9B8AABB9CC33FE3CB91E643BD1E7333C6581DE3C223B283DF867FFBBA
06BCA609CA3BBD14                                                                                                AC3CB181A03B4890
CF5631B3C577DDA3C                                                                                               BD23DCB3C2ECE3F3
B23D3B1041CFBC1B                                                                                                090BBD9A4414BD33
13CE64B9D3AE834C0                                                                                               43C4933BEBC |
| 0xDC02AEBB9D684                                                                                               6EBDFD93C2BCF1DB
B3B88403D258282B0                                                                                               C57278A3D0D8A813
ACAA3D74990DBDF09                                                                                               FEB1BBCEEF173D49
C391F34F13BDD4E85BBE7F6B23C73B6643C404D3DBD38D2DD3CB104F7BC58AE6C3C1C521DBD02FF2ABC826FE6BC3D7A833B0587EA3A1B06C6BC2FD09
DC6BDC3C51AD873C4B8D993C962AA23AEF8FC63C84CA57BDF773D43BE6E798BC2B1AB9BC8B087ABCF546063C44A98BBC8AEAC5BAC364C1BC795E8A3C
B24BDE0C7AA3C1AE8113C69F082BCFE4848BD6F1F8B3B32B2EF39BC8F4D3D788A7BBB84CA573C7C4FF8BA38D25DBC96FC7EBD866206BB466DAB3CCB8
3B50F881BDBBE9613BABD5EE3C6959B13C11C746BB479AF93C56EACC3C8042183CA5890FBC375A15BDE7054DBB5644E1BB1B6F743D69F082BCCAA2E3
F12A53D59EBA93C3D3D46BC9DFF95BD9329C5BB59FA43BD69F0823DCE6760BA32E0123B28A1933D777B61BDC36441BD6029CC3CB2E71F3C74F3A1BC5
F3BCCBB17DBCE47D0D3D8AEA45BD188DA03CD679913CA12D41BD548F5BBD416BF1BCB606313CE79C1EBCD98908BD962AA2BC9748D6BC73E487BCBDDB
CBB1705BDB5F7963C274622BC98EE413BBE81103CDFB8103DB5C9733C7ED88C3C4AD813BD7B31443DD52D3A3C8315D23C8AF95FBCE2F4783C1F34F1B
D3A53CA74D2FBC83BB3DBD244545BC43D57CBC55CC983B3C975A3CBF9F44BC73B664BCE9062A3C8206B83CFF57E23C218FE23CE9F70FBC8AF95F3B87
DBDCAA2633D579F523DCDB25ABC2AFC04BD6C69A8BC6F8839BC8315D23B0AE3B83CACC7B1BCDE038BBC25545FBDC30A2D3B9857703DBBE9E1BC12139
30DF353C375A153D67FE3FBD6C69A83C7E413B3CB241343D1F62943CE2C887BC6968CBBBC838D83B0C4D44BD46D6593C280A423C777BE1BC15E6D73C
986399FC335BD2554DFBCCE9583BC764E13BDD2C3AEBC219E7CBB3249C1BCD41EA03CC36441BC2126B4BCECE87D3D1F6294BCA9B7BABC6D9676BCC4B
3B0D8A81BC4EF7A4BCB1321ABC73E487BCB0F5DC3CF463DDBC486E883D90B01FBC0A7A0ABC15E6573BF18189BCC8CFA93BD41E20BDEE3455BC0888C7
6C6DBBCB2504E3D4C9CB33B93C016BD304864BC9A958A3C0A4CE7BA13B909BD107BEFBB106CD53B91560BBA7A7C3EBBE59BC1BCB6AC1CBB0888473C2
793B6AA5083D8D910EBDE44F6A3BD25A803CC556043DB861A23A5D19D53C25545FBDC5CE4C3DFF5762BCE53213BC8D0957BD5DB0A63CF2F9513C6339
CF4635DBC2554DFBC67676EBDB4ABBFBCB2504EBAA0E1693D8461293B56537BBB01B353B83F3E23BD94CFB03C703DBFBCE2C8073CB9BC93BD2381253
23153DAFD728BD78B81E3A95EDE4BC23EA533CC8CFA9BC7C4F78BCB250CEBCBCCC8ABC33FEC63C7298303D9DA581BC2864563B6D1EAEBB9C1C6D3DCC
```

So Cool:
Embedding & Design Decision

viscosityna.com ORAPUB utoug utah oracle users group

22

What is embedding?

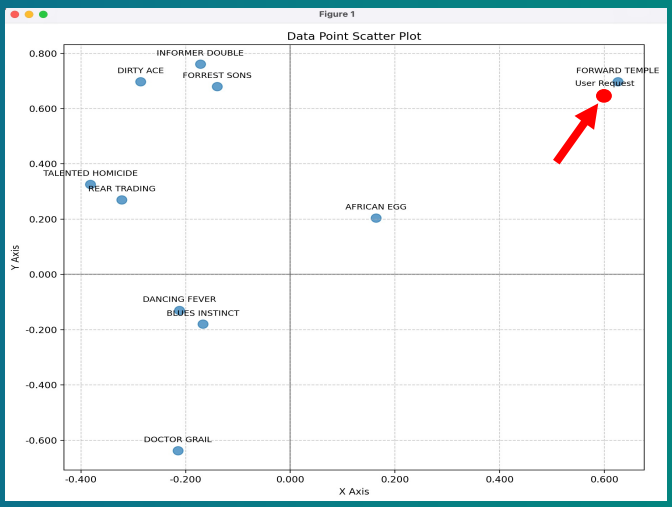
User Request: *I'm interested in watching a movie, set in a deep African forest that centers around life in the community temple.*

Imagine for each film, taking all the information about then film, then reducing it a point; coordinates

Then, imagine taking a user's request, then reducing it to a point.

Plot all the points.

Locate the closest 3 films to the User Request. This is the context the LLM will reference when responding to the User Request.



What is embedding?

Transforming text, pictures, graphics, videos, music, etc. to vectors (array of 1536 numbers) enables thematic searching not rigid keyword searches and *regex*.

This distance calculation is based on 1536 dimensions, not just 2 or 3 dimensions.

So Cool: **Embedding** & Design Decision

```
[opc@mysql-compute rag_service]$ pwd
/var/www/html/my-web-app/rag_service
[opc@mysql-compute rag_service]$ ls
db_config.php     generate_embeddings.php  vector_functions.php
film_search.php   rag_api.php
[opc@mysql-compute rag_service]$ php generate_embeddings.php
```


So Cool: **Embedding** & Design Decision

generate_embeddings.php

```
20  function generate_film_embeddings() {
26      try {
53          // Process each film
54          $processed = 0;
55          foreach ($films as $film) {
56              echo "Processing film: {$film['title']}\n";
57
58              // Create text representation with enhanced data
59              $text = create_film_text($film);
60
61              // Generate embedding
62              $embedding = generate_embedding($text, $openai_api_key);
63              if (!$embedding) {
64                  echo "  Error generating embedding\n";
65                  continue;
66              }
67
68              // Convert embedding array to proper format for STRING_TO_VECTOR
69              $vector_str = "[" . implode(",", $embedding) . "]";
70
71              // Store embedding directly using STRING_TO_VECTOR MySQL function
72              $stmt = $conn->prepare("
73                  UPDATE film
74                  SET vector_embedding = STRING_TO_VECTOR(?)
75                  WHERE film_id = ?
76              ");
77
78              $stmt->bind_param("si", $vector_str, $film['film_id']);
```

1. Every row in the film table.
2. Create a text string related to the film. See next slide.
3. Vectorize/embed the text.
4. Update the vector_embedding column in the film table.

So Cool: Embedding & **Design Decision**

generate_embeddings.php

app-home > my-web-app > rag_service > generate_embeddings.php

```php
/**
 * Create enhanced text representation of a film for embedding
 * Now includes more details from film_list
 */
function create_film_text($film) {
    $text = "Film Title: {$film['title']}\n";
    $text .= "Description: {$film['description']}\n";
    $text .= "Release Year: {$film['release_year']}\n";
    $text .= "Rating: {$film['rating']}\n";
    $text .= "Length: {$film['length']} minutes\n";

    // Add price information
    if (isset($film['price'])) {
        $text .= "Rental Price: \${$film['price']}\n";
    }

    // Add category information
    if (isset($film['category'])) {
        $text .= "Categories: {$film['category']}\n";
    }
```

Each film's information (row) is formatted into a text string. Then the text will embedded/vectorized.

The result is each MySQL EE **row** will contain the existing application columns and also the new vectorized/embedded text column.

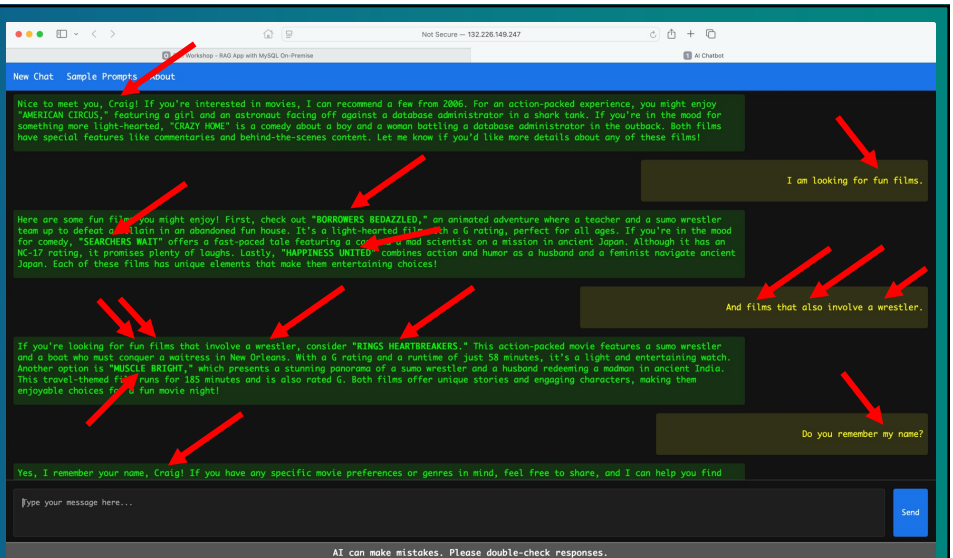
Why? The similarity search will be between the users' embedded/vectorized request and the "text" embeddings/vectors.

27

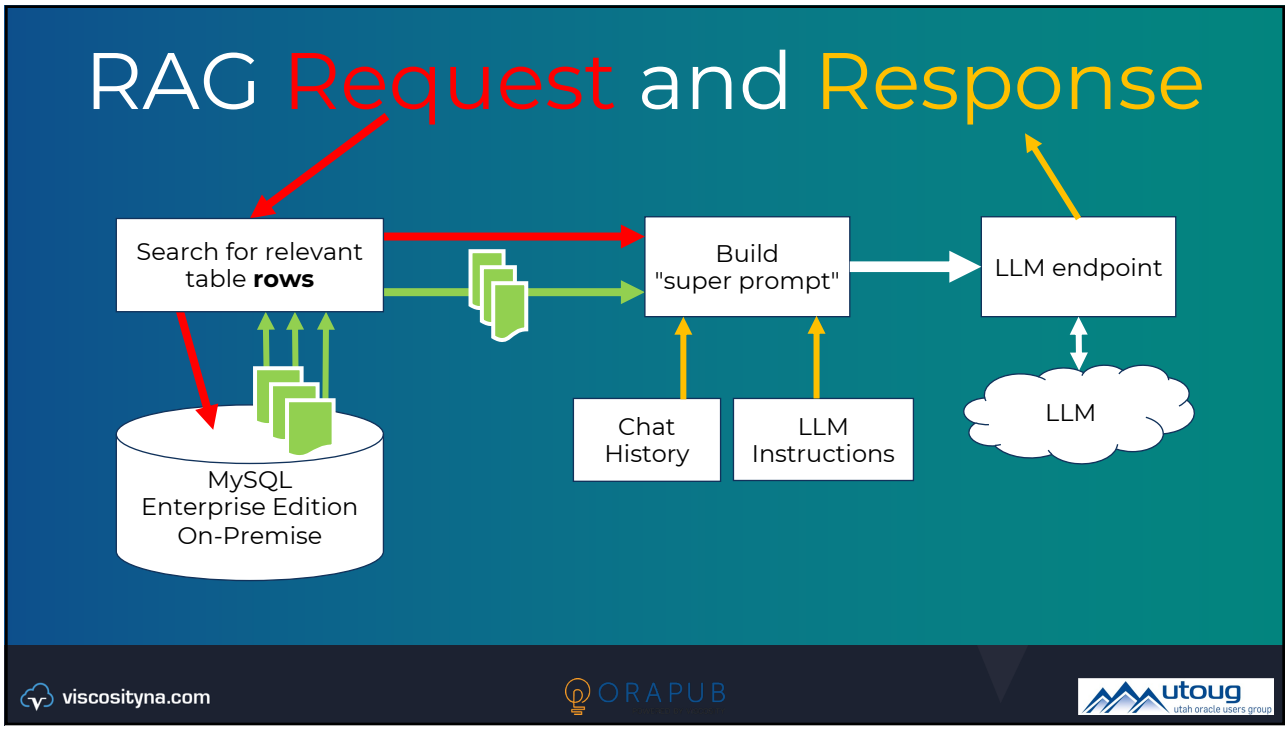
And… It just works!

Notice how we are taking advantage of LLM language power, our context and conversation history (RAG).

Without Linear Algebra, matrixes and vectors, current day AI would simply not exist.



28



29

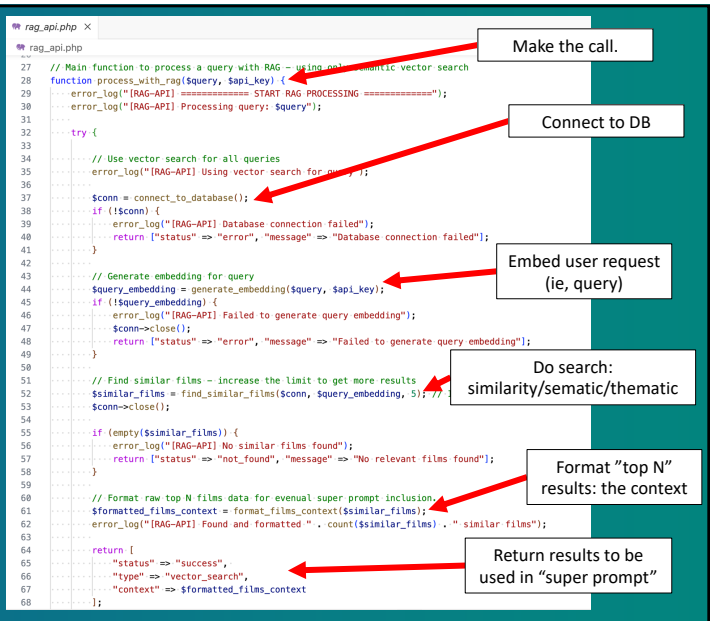
RAG Request & Context Build

We create the "super prompt" so the LLM can **respond using our data**.

Part of the super prompt contains details about the films that are "similar" to the user's request.

"Similar" means relatively **close in vector space.**

This **distance between** the user's request vector (embedding) and a film's vector (embedding) are computed, then the top N closest films data is returned, formatted and return to used in super prompt.



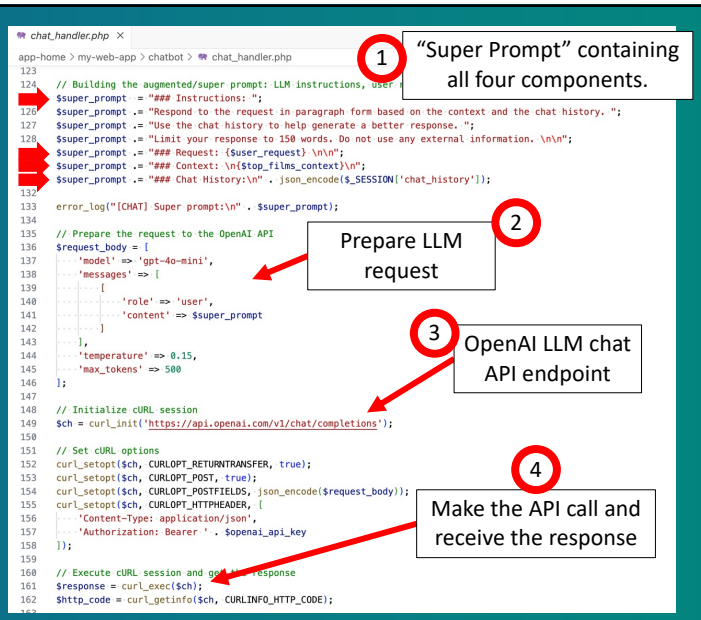
30

RAG
Response Send To LLM

We are giving the LLM: 1) instructions, 2) user request, 3) context (films) and 4) chat history.

Truly, a "super prompt".

There are many nuanced strategies to design, structure and build the final "super prompt".

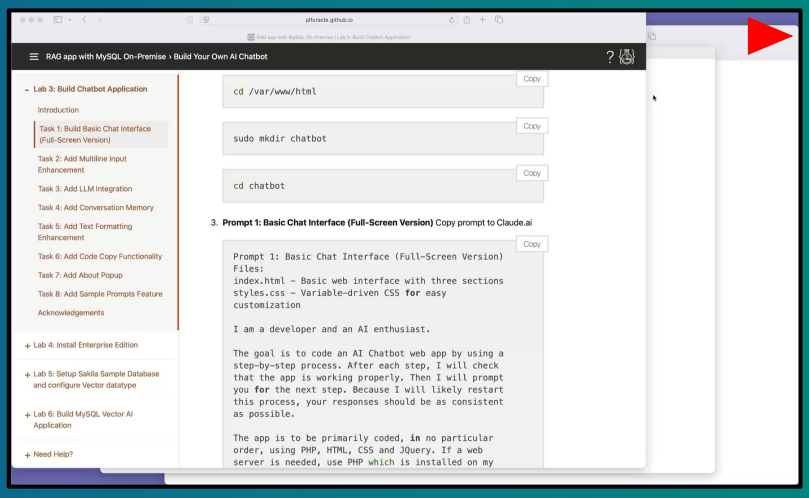


Cool things about the app.

100% Claude.ai collaboration... hard work.
Uses my preference of vendors.
Embedding design decision is important.

So Cool: 100% Claude.ai Collaboration

The initial Chatbot was 100% prompt "copy and paste". The final RAG app was an intense and fun collaboration with AI.



So Cool: Uses **my** vendor preferences

- **OpenAI** for RAG LLM:
  - Always works with coding assistants,
  - Offer small and fast LLMs.
- **Claude.ai** web app interface:
  - Coding LLM (eg, Sonnet) just works,
  - I like the AI persona and Anthropic.

- Cohere. Inexpensive, great models, but coding assistants don't work well with Cohere.
- Grok. Expensive and don't need a scientific truth-seeking answer.
- Llama. Did not want **the LLM on my desktop (super secure)**, build an infrastructure or use an API service.
- Gemini. Have not tried it yet… but will.
- Deepseek. Do not need a "reasoning" model with a long response time.

Why use MySQL EE On-Prem For RAG Apps?

- **AI Ready.** MySQL EE On-Prem now has the capability to help build real AI applications… and it will only get better!
- **Scalable On-Prem**. You can develop on your desktop(s), then scale up **on-prem** as the workload grows!
- **True Database.** MySQL EE is much more than a "vector store" or "document database" because a true database "out of the box" provides high concurrency transaction management, backup and recovery, and security.
- **Super AI Secure**. You can even have your AI app working without an internet connection!

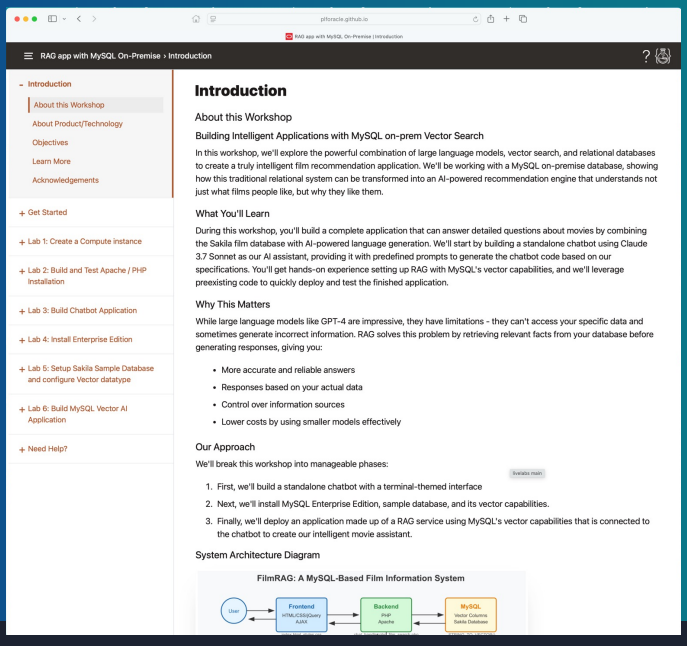
Why use MySQL EE On-Prem For RAG Apps?

Because, if you have of lots data in a database and:

- Extracting and analyzing the data takes multiple SQL statements, requires crazy syntax or takes forever.
- You desire a more collaborative analysis environment.
- Your data is proprietary….period.
- Deep research or casual questions specific to your proprietary data is the norm.

Next Steps…

- Get MySQL Enterprise Edition on-prem/desktop/OCI/AWS/… Go!
- Advanced RAG. There are a ton of techniques and technologies to made RAG much better. Goto: www.orapub.com
- Multiple AIs working together collaborating with you. Not a "reasoning" model, but a true collaboration. Goto: www.orapub.com or the Viscosity table for a URL.
- Cursor AI. I love this coding AI assistant!!! Goto: www.cursor.ai
- ….and…



Next Steps…
Hands On Lab

Step-By-Step you can build a simple Q&A web app using only prompts that you copy/paste into Claude.ai
-AND-
MySQL EE On-Prem Q&A chatbot web app

Thank You!

viscosityna.com ORAPUB utoug utah oracle users group

39

Next Steps URLs

HOL: RAG App with MySQL On-Premise (Private and Limited HOL Release)
https://livelabs.oracle.com/pls/apex/dbpm/r/livelabs/view-workshop?wid=4158

AI TEAM COLLABORATOR
http://64.23.163.216
username: nocoug
password: vna123

viscosityna.com ORAPUB utoug utah oracle users group

40

VISCOSITY
PROFESSIONAL SERVICES

Building Gen AI-Powered Database Applications: A "Developer's" Fast Track to Production

Utah Oracle Users Group (UTOUG)
Virtual Event
Thursday, June 12, 2025

utoug
utah oracle users group

Craig Shallahamer

Oracle ACE Director
Applied AI Scientist | OraPub Founder

Oracle ACE Director
SYMPOSIUM L2

linkedin.com/in/craig-shallahamer-571a94a/
craig.shallahamer@viscosityna.com

viscosityna.com

ORAPUB

41