

Azure Databricks – Getting Started with Vector Search

Wilson Mok

May 21, 2025





Wilson Mok

Sr. Data architect & Consultant

 /wilson-mok

 /wilson-mok

 @the-analytics-lab



My experiences includes:

- Avanade
- CAE
- Air Canada



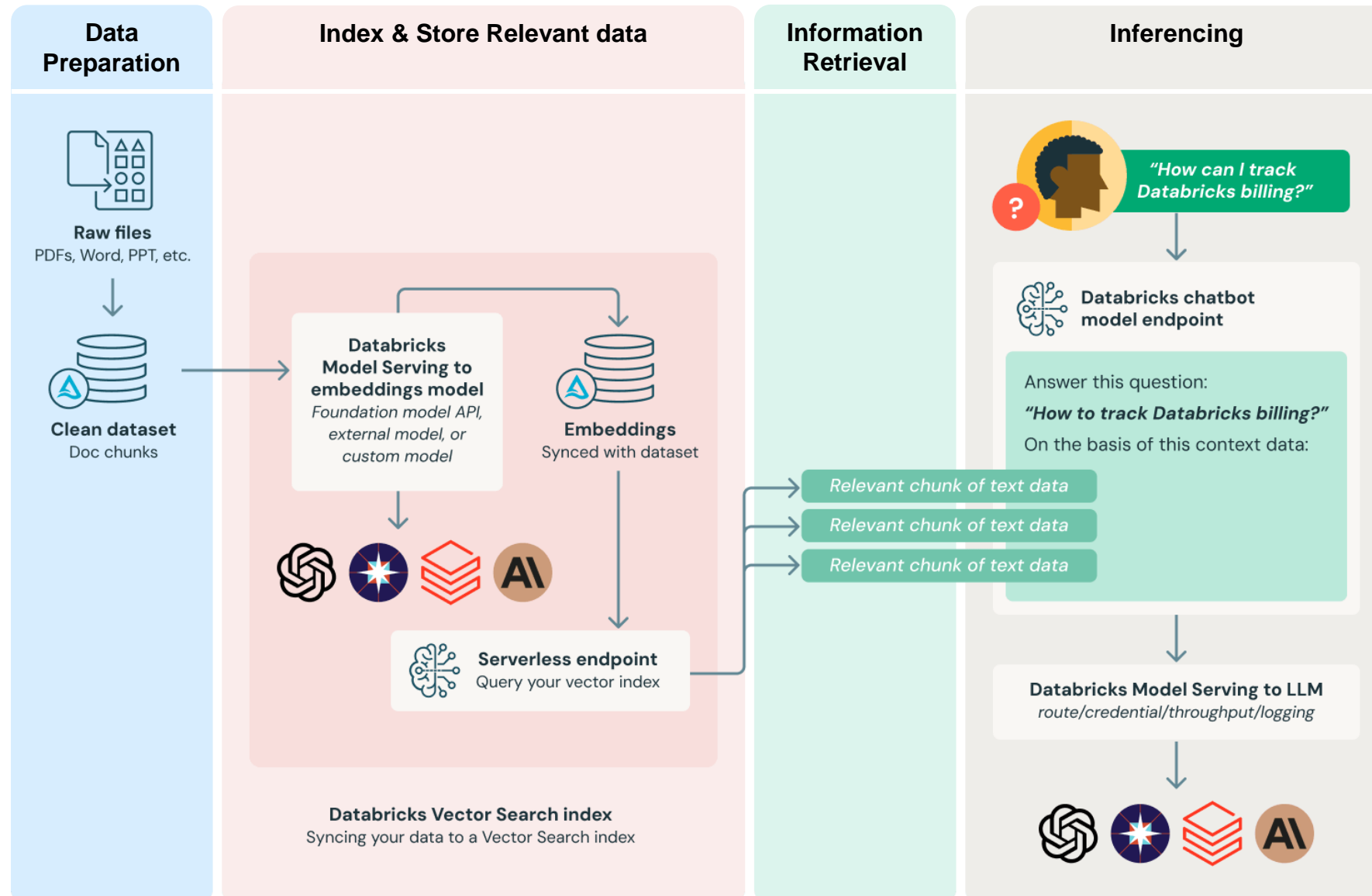
Expected Audience and Session Structure

- This session, we will focus on:
 - How is Databrick's Vector Search fits into the Gen AI ecosystem?
 - Demo: Create a RAG workflow to enable semantic search for our internal documentation/knowledge base.
- Experienced with Databricks, Python and RAG would be beneficial.

Gen AI has changed the way we interact with our data. Our users expects to “Talk to our data”, not just a search box!



You Don't Need Another Stack – Databricks provides a Unified Platform from Data Preparation to Inferencing



Credit: <https://www.databricks.com/glossary/retrieval-augmented-generation-rag>

Q & A

Summary & Takeaways

1. Prepared our Semi-structure Markdown data using Medallion Architecture.
 - Bronze: stored the Markdown as a string.
 - Silver: extracted the Markdown metadata.
 - Gold: Chunked the Markdown content based on embedding model's requirement.
 2. Using Mosaic AI Vector Search to create embedding from our Gold table.
 - Using 'databricks-bge-large-en' embedding model.
 3. Executed Similarity Queries to demonstrate the retrieval is returning meaningful results.
-
- The code is in my Github repo.
 - [code]

Thank you



LinkedIn