

Summary and Highlights: Vector Databases for Recommendation Systems and RAG

Congratulations! You have completed this lesson. At this point in the course, you know:

- That Chroma DB operations form the foundation for building intelligent, vector-based applications
- That collections are a way to organize your data within Chroma DB
- That you define your embedding model using Chroma DB's `embedding_functions`
- That you create a collection using the `create_collection` method
- That metadata helps you keep track of the purpose and contents of your collections
- That you connect to an existing collection by using the `get_collection` method
- That you alter an existing collection by using the `modify` method
- That you use the `add` method to insert documents into a collection
- That you get data from a collection using the `get` method
- That you update existing data in a collection by using the `update` method
- That you delete data from a collection using the `delete` method
- That you use the HNSW space configuration parameter to specify the distance function when performing approximate nearest neighbor searches in Chroma DB
- How to analyze data with similarity search
- That RAG enhances LLM response quality by retrieving relevant external information, helping the model generate more accurate and well-supported outputs
- That vector databases are the foundation that makes Retrieval-Augmented Generation work
- That the tasks vector databases can perform in a RAG pipeline include: embedding source documents and user prompts, storing embeddings, retrieving most relevant matches, and providing the retrieved content for prompt augmentation