Topic 3: Retrieval Layer (Core of RAG)

MCQs (51–75)

A. Retrieval Techniques

Q51. In a RAG pipeline, the retrieval layer's primary responsibility is:
a) Generating final answers directly
b) Fetching relevant documents based on embeddings or keywords
c) Training the LLM
d) Visualizing query results
Answer: b

Q52. Which retrieval technique **only** relies on exact keyword matches?
a) Vector search
b) Hybrid search
c) Keyword search
d) Contextual reranking
Answer: c

Q53. Vector search retrieves documents by comparing:
a) Hash values
b) Embedding vectors
c) BM25 scores
d) TF-IDF indexes
Answer: b

Q54. In **hybrid search**, results are typically combined from:
a) Vector similarity + keyword relevance
b) BM25 + LLM fine-tuning
c) TF-IDF + embeddings
d) Both a and c
Answer: d

Q55. Which retrieval strategy works best for **multi-lingual queries**?
a) Pure keyword search

b) Dense vector search using multilingual embeddings
c) Hash-based document indexing
d) Sparse vector search only
Answer: b

B. LangChain & LlamaIndex Retrievers

Q56. In LangChain, which class is used to implement vector-based retrieval?
a) `VectorStoreRetriever`
b) `DocumentRetriever`
c) `LLMChainRetriever`
d) `HybridRetriever`
Answer: a

Q57. LangChain's `ParentDocumentRetriever` is used when:
a) Chunk sizes are extremely large
b) Metadata-based filtering is required
c) Only parent document summaries should be retrieved
d) The embedding model doesn't support high-dimensional vectors
Answer: c

Q58. LlamaIndex retrievers can fetch data from:
a) PDFs and APIs
b) Databases and websites
c) Vector stores
d) All of the above
Answer: d

Q59. In LlamaIndex, which retriever is optimized for **semantic search**?
a) `VectorIndexRetriever`
b) `TreeIndexRetriever`
c) `SummaryRetriever`
d) `KeywordRetriever`
Answer: a

Q60. To combine vector and keyword retrieval in LlamaIndex, you should use:
a) `HybridRetriever`

b) `SemanticKeywordRetriever`
c) `BM25Retriever`
d) `MultiModalRetriever`
Answer: a

C. Popular Vector Databases

1. Pinecone

Q61. Pinecone is best described as:
a) An in-memory graph database
b) A serverless managed vector database
c) A document-only NoSQL store
d) A cloud-based LLM provider
Answer: b

Q62. Which feature makes Pinecone suitable for production-scale RAG applications?
a) Horizontal auto-scaling
b) Built-in hybrid search
c) High availability via managed infrastructure
d) All of the above
Answer: d

---

2. Weaviate

Q63. Which of the following best describes **Weaviate**?
a) A lightweight local store
b) An open-source vector database supporting hybrid search
c) A proprietary managed-only database
d) A purely keyword-based retrieval system
Answer: b

Q64. Weaviate supports **hybrid search** by combining:
a) Vector similarity + BM25 scores

b) TF-IDF + embeddings
c) Dense + sparse retrieval
d) All of the above
Answer: d

---

3. Milvus

Q65. Milvus is optimized for:
a) Low-scale text retrieval only
b) Large-scale high-dimensional vector searches
c) Image-only retrieval
d) Database schema migrations
Answer: b

Q66. Which feature differentiates Milvus from Pinecone?
a) Open-source flexibility with on-prem deployment
b) Serverless architecture
c) Exclusive support for OpenAI embeddings
d) Lack of metadata filtering
Answer: a

---

4. Qdrant

Q67. Qdrant is particularly known for:
a) Proprietary storage engine
b) Open-source design and **fast semantic search**
c) Integrating only with Pinecone
d) Lacking hybrid search support
Answer: b

Q68. Which query language does Qdrant commonly support?
a) SQL
b) GraphQL
c) Cypher

d) XQuery
Answer: b

---

5. FAISS

Q69. FAISS, developed by Facebook AI, is primarily:
a) A distributed LLM inference library
b) An **in-memory vector similarity search library**
c) A managed cloud retrieval system
d) A PDF ingestion tool
Answer: b

Q70. FAISS works best when:
a) Queries are small, and storage is local
b) Cloud scalability is required
c) Serverless RAG pipelines are needed
d) You require built-in reranking
Answer: a

---

6. Elasticsearch / OpenSearch

Q71. Elasticsearch uses which default scoring algorithm for keyword retrieval?
a) TF-IDF
b) BM25
c) Cosine similarity
d) Euclidean distance
Answer: b

Q72. For **hybrid retrieval**, Elasticsearch combines:
a) Dense vector search + BM25 ranking
b) Pure TF-IDF + LSA
c) FAISS indexing + BM25
d) Pinecone + HuggingFace embeddings
Answer: a

---

7. ChromaDB

Q73. ChromaDB is most commonly used for:
a) Large-scale distributed deployments
b) Lightweight local vector storage for prototyping
c) Full document parsing
d) Web crawling
Answer: b

Q74. Which framework integrates seamlessly with ChromaDB?
a) LangChain
b) LlamaIndex
c) Both a and b
d) Neither
Answer: c

Q75. One limitation of ChromaDB compared to Pinecone and Milvus is:
a) Lack of open-source support
b) Limited scalability for very large datasets
c) Lack of hybrid retrieval capabilities
d) Absence of vector indexing
Answer: b