

Homework 7: Vector Databases

Correct answers are labeled red bold.

Some necessary explanations are labelled blue.

Q1: Which of the following is a primary use case for vector databases in cloud computing?

- A. Storing transactional data.
- B. Storing and querying high-dimensional vectors for similarity search.**
- C. Managing relational data.
- D. Performing batch processing of structured data.

Q2: Which of the following is a key characteristic of a vector database?

- A. Support for SQL queries.
- B. Handling ACID transactions.
- C. Optimized for similarity searches in high-dimensional spaces.**
- D. Designed for data warehousing.

Q3: Which of the following Faiss Index types is most suitable for brute-force search in a small dataset?

- A. IndexIVFFlat.
- B. IndexFlatL2.**
- C. IndexIVFPQ.
- D. IndexHNSW.

Q4: In Faiss, which index is designed to handle very large datasets by partitioning the data into smaller subsets, each indexed independently?

- A. IndexIVFFlat.**
- B. IndexFlatIP.
- C. IndexHNSW.
- D. IndexLSH.

Q5: In Faiss, an _____ is constructed by dividing the set of vectors into k Voronoi partitions.

- A. Locality Sensitive Hashing Index
- B. Inverted File Index**
- C. Hierarchical Navigable Small Worlds Index
- D. Flat Index

Q6: In the Inverted File Product Quantization Index (IVFPQ), the _____ algorithm is run on vectors from all the partitions:

- A. K-means clustering

- B. Principal Component Analysis
- C. Product Quantization**
- D. Singular Value Decomposition

Q7: In the Inverted File Product Quantization Index (IVFPQ), sub-vectors are quantized into a finite number of _____, each represented by _____.

- A. clusters; a centroid**
- B. bits; a hash code
- C. partitions; a mean value
- D. blocks; an eigenvector