

Survey of Vector Database Management Systems

James Jie Pan · Jianguo Wang · Guoliang Li

Presenter: Lim Yuan Jee

National Cheng Kung University

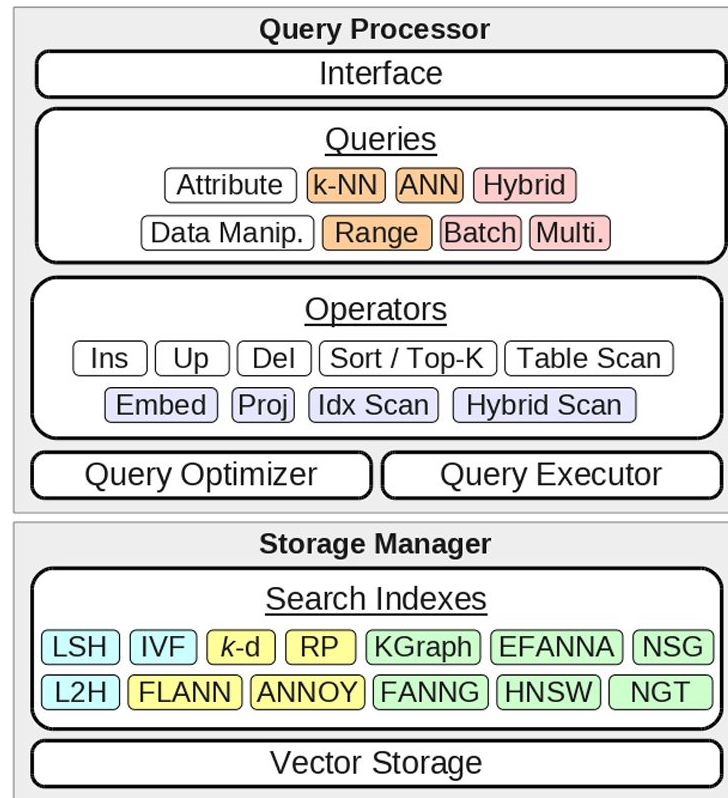


Discussion topics

- Vector Database Management Systems (VDBMSs)
- Challenges
- Query Processing
 - Queries
 - Operators
- Indexing Techniques
 - Key Partitioning Strategies
 - Primary Index Structures
- Query Optimisation
- Types of VDBMSs

Vector Database Management Systems (VDBMSs)

- Database system designed to manage and query large collections of high-dimensional vectors
- Support embedding-based retrieval (EBR) and similarity search for:
 - Large Language Model (LLM)
 - Data-intensive tasks
- Addresses several key challenges for above applications
- Incorporate various techniques for query processing, storage and indexing, and query optimization and execution

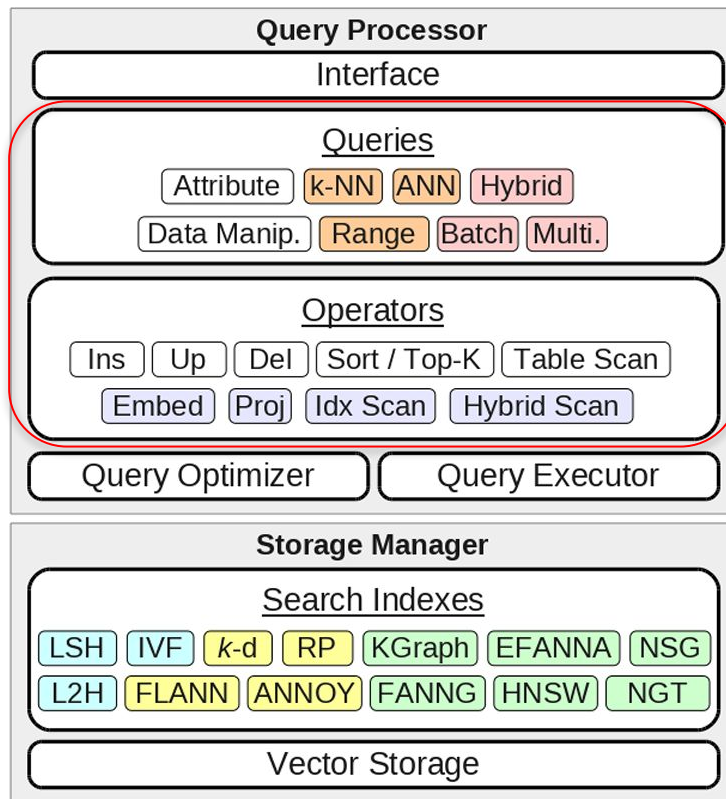


Architecture of a VDBMS.

► Challenges in Vector Data Management

- Five key obstacles
 - Vague Search Criteria
 - Expensive Comparisons
 - Large Vector Sizes
 - Lack of Structure
 - Incompatibility with attributes

Query Processing



- Similarity Scores

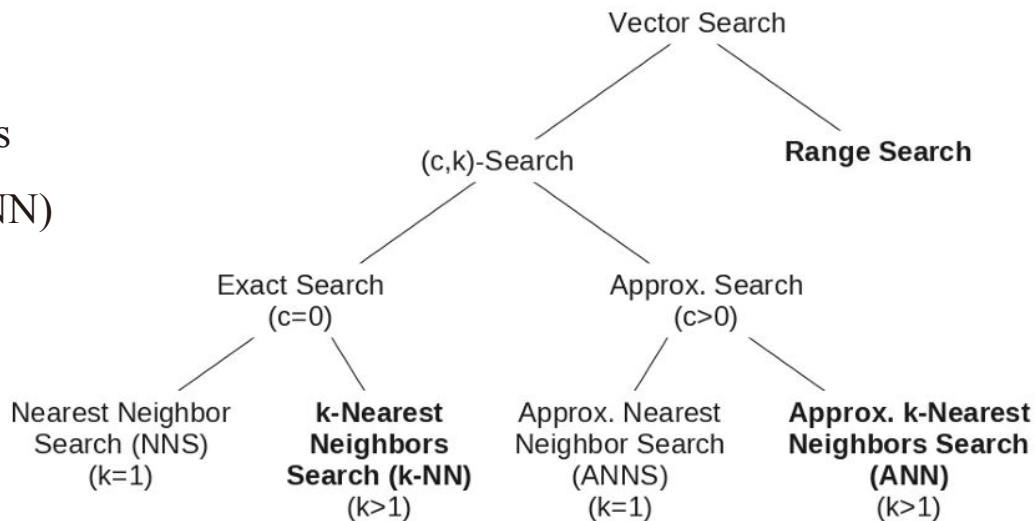
- Inner Product
- Cosine Similarity
- Euclidean Distance
- Hamming Distance
- Minkowski Distance
- Mahalanobis Distance

Type	Score	Metric	Complexity	Range
Sim.	Inner Prod.	✗	$O(D)$	\mathbb{R}
	Cosine	✗	$O(D)$	$[-1, 1]$
Dist.	Minkowski	✓	$O(D)$	\mathbb{R}^+
	Mahalanobis	✓	$O(D^2 + O(1))$	\mathbb{R}^+
	Hamming	✓	$O(D)$	\mathbb{N}

- Similarity Score function: $f : \mathbb{R}^D \times \mathbb{R}^D \rightarrow \mathbb{R}$

Queries (cont.)

- Search Queries
 - (c, k)-Search Queries
 - Exact k-NN queries
 - Approx. k-NN (ANN)
 - Range Search Queries

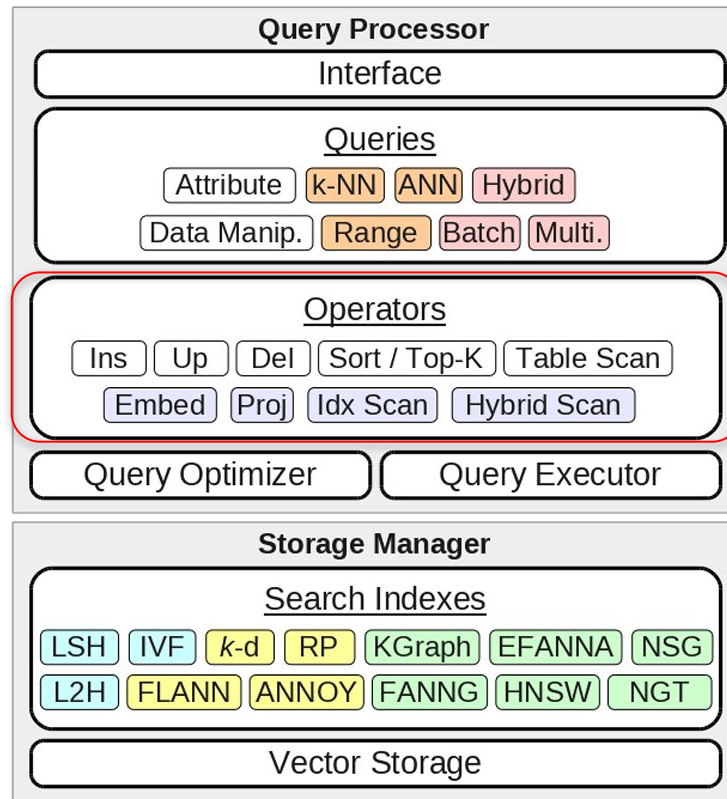


Queries (cont.)

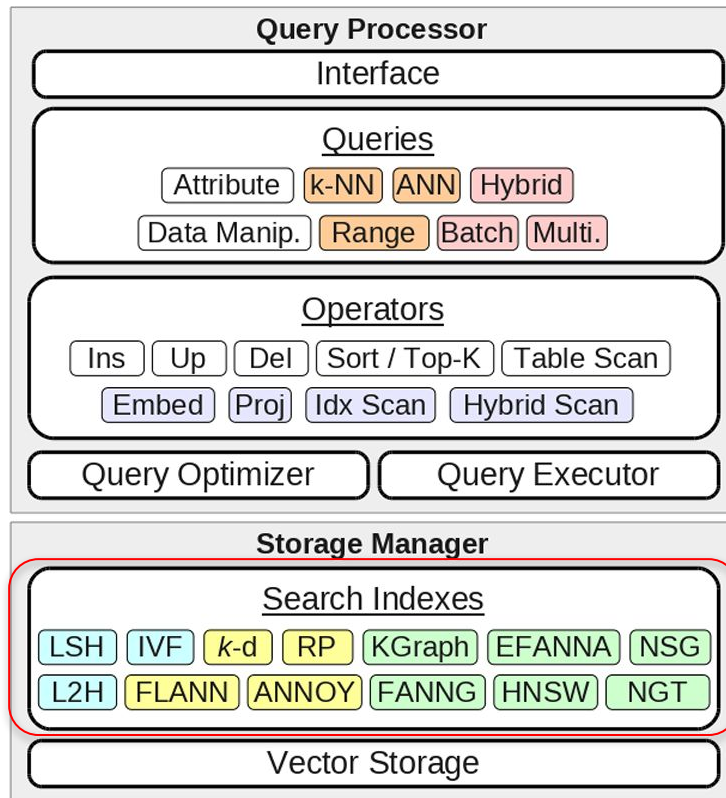
- Query Variants
 - Predicated (Hybrid) Queries
 - Batched Queries
 - Multi-Vector Queries
 - Multi-Query, Single Feature (MQSF)
 - Multi-Query, Multi-Feature (MQMF)
 - Single-Query, Multi-Feature (SQMF)

Operators

- Projection
- Index-based Operators



Vector Indexing



Key Partitioning Strategies

- Randomization
 - Locality Sensitive Hashing (LSH)
- Learned Partitioning
 - k-Means Clustering
 - Product Quantization
- Navigable Partitioning
 - Graph-Based Index

Primary Index Structures

- Table Based Indexes
 - LSH based indexes (E2LSH)
 - Quantization based indexes (IVFADC, SPANN)
- Tree Based Indexes
 - Includes k-d trees, FLANN, RPTree, ANNOY
 - Can suffer from the curse of dimensionality
- Graph Based Indexes
 - k-Nearest Neighbour Graphs(KNNG)
 - Monotonic Search networks (FANNG, NSG)
 - Small-world Graphs (HNSW)

Query Optimization

- Optimizing Hybrid Queries
 - Block-First Scan
 - Filters out vector that do not satisfy the attribute predicate **BEFORE** performing the vector search
 - Online - AnalyticDB-V and Milvus - bitmasks from traditional attribute filtering
 - Offline - add edges for graph-based index - Milvus pre partition based on predicates
 - Visit-First Scan
 - Incorporate traditional attribute filtering mechanism of the scan operator.

Query Optimization (cont.)

- Plan Enumeration Methods
 - Predefined Plans
 - Some VDBMS for specific workload predefined query plans
 - simplifies optimization
 - inflexible
 - Automatic Enumeration
 - Systems built on RDBMSs leverages underlying relational optimizer for plan enumeration.
 - Extend relational language to distance function & vector index scans
 - pgvector & PASE (PostgreSQL)

Query Optimization (cont.)

- Plan Selection Strategies
 - Rule-Based Selection
 - Handcrafted rules (limited # of plans)
 - Cost-Based Selection
 - lowest estimated cost
 - AnalyticDB-V and Milvus utilize cost models
 - Considered factors: distance calculations, memory and disk retrievals, and predicate selectivity

Types of VDBMS

- Native VDBMS
 - Mostly Vector Workloads
 - Designed to handle mostly queries that involves accessing vector data
 - Fast and efficient vector search
 - EuclidesDB, Vald, Vearch, Pinecone, Chroma
 - Mostly Mixed Workloads
 - Involve both vector and non vector data
 - Ability to handle mixed queries
 - Milvus, Quadrant, NucliaDB, Weaviate

Types of VDBMS (cont.)

- Extended VDBMS
 - NoSQL Systems
 - Incorporates vector into existing NoSQL
 - Leverages schemaless storage and distributed architecture
 - Vespa, Cassandra, Databricks
 - Relational Systems
 - Singlestore, PASE, pgvector, AnalyticDB-V, Clickhouse
- Search Engines
- Libraries

Types of VDBMS (cont.)

