

DESIGN OF VECTOR AND DISTRIBUTED DATABASES

Discipline Status	Selective
Level of Higher Education	Second (master's)
Language of Instruction	English
Семестр	–
Discipline Volume, ECTS Credits	8,0
Forms of education for which the discipline is taught	Full-time (day)

Learning results. Upon successful completion of the course, students should: understand the fundamental principles of vector data representation, similarity search, and distributed data storage; know the architectures and functionalities of modern vector database frameworks (FAISS, ScaNN, Milvus, Qdrant) and distributed DBMS (Cassandra, CockroachDB); be able to design datalogical and physical models, configure indexing structures, and optimize performance for high-dimensional data retrieval; apply knowledge to integrate databases with ML-based systems for semantic and recommendation tasks; justify engineering decisions based on performance analysis and theoretical understanding; demonstrate initiative, independent learning, effective communication, teamwork, and professional responsibility in the development of intelligent data management systems.

Content of the academic discipline. Introduction to Vector and Distributed Databases. Fundamentals of Vector Data Representation. Designing Vector Entity Structures in Databases. Relational Models Extended for Vector Representations. Architecture and Organization of Local Vector Databases. Similarity Search in Vector Spaces. Facebook AI Similarity Search (FAISS): Architecture and Methods. Optimization and Indexing Strategies in High-Dimensional Spaces. Scalable Search and Hybrid Filtering Methods. Fundamentals of Distributed Databases. Designing Datalogical Models and Node Diagrams for Distributed Systems. Consistency, Fault Tolerance, and Replication Mechanisms. Practical Implementation of Distributed Databases Using DBMS Tools. Security, Integrity, and Access Control in Distributed Systems. Integration of Vector Databases with Machine Learning Pipelines. Emerging Trends in Vector and Distributed Databases.

Planned learning activities: The minimum amount of study hours in one ECTS credit of an academic discipline for the second (master's) level of higher education in full-time education is 8 hours per 1 ECTS credit.

Forms (methods) of learning: lectures (using visualization methods, problem-based and interactive learning, motivational techniques, information and communication technologies); practical classes (using instruction, demonstration, solving typical and applied problems, case analysis, situational tasks, discussion elements, etc.); independent work (studying theoretical material, preparing for practical work, current and final control, performing individual and homework), using information and computer technologies and distance learning technologies.

Forms of assessment of learning outcomes: laboratory work, current control work, presentation of the results of individual tasks.

Type of semester control: credit.

Educational resources:

1. Nexla. Vector Databases: Tutorial, Best Practices & Examples [Electronic resource]. – 2024. – Available at: <https://nexla.com/ai-infrastructure/vector-databases/> (accessed 07 October 2025).
2. Chandepa, S. A Beginner's Guide to Implementing Vector Databases [Electronic resource]. – 2024. – Available at: <https://zilliz.com/learn/beginner-guide-to-implementing-vector-databases> (accessed 07 October 2025).
3. Microsoft. Understanding Vector Databases [Electronic resource]. – 2024. – Available at: <https://learn.microsoft.com/en-us/data-engineering/playbook/solutions/vector-database/> (accessed 07 October 2025).
4. Jiang, R. How to Build a Dynamic Distributed Database with DistSQL [Electronic resource]. – 2022. – Available at: <https://opensource.com/article/22/9/dynamic-distributed-database-distsql> (accessed 07 October 2025).
5. Hire, A. What Is a Distributed Database? A Complete Guide [Electronic resource]. – 2025. – Available at: <https://www.pingcap.com/article/what-is-a-distributed-database-a-complete-guide/> (accessed 07 October 2025).
6. Leangroup. Distributed Database Management Systems [Electronic resource]. – 2024. – Available at: https://webweb.ams3.cdn.digitaloceanspaces.com/data/simmcdev.webweb.ai.in/MCAdigitalbook/DBMS-20240629T071926Z-001/Distributed_Database_Management_Systems_Leangroup_org.pdf (accessed 07 October 2025).
7. PingCAP. Modern Distributed Database Fundamentals [Electronic resource]. – 2023. – Available at: <https://static.pingcap.com/files/2023/08/16173906/PingCAP-eBook-Modern-Distributed-Database-Fundamentals.pdf> (accessed 07 October 2025).
8. Milvus. Milvus Open-Source Vector Database Documentation [Electronic resource]. – 2024. – Available at: https://milvus.io/docs/v2.3.0/milvus_overview.md (accessed 07 October 2025).
9. Zilliz. Chroma Vector Database Tutorial [Electronic resource]. – 2024. – Available at: <https://zilliz.com/blog/chroma-tutorial-vector-database> (accessed 07 October 2025).
10. FAISS. Facebook AI Similarity Search Library [Electronic resource]. – 2024. – Available at: <https://github.com/facebookresearch/faiss> (accessed 07 October 2025).
11. Modular learning environment. URL: <https://msn.khmnu.edu.ua/course/view.php?id=7933>
12. University electronic library. URL: <http://library.khmnu.edu.ua/>
13. KhNU Repository. URL: <https://elar.khmnu.edu.ua/home>

Teachers: Ph.D. in E. S., As. Prof. Mazurets O.V.,
Ph.D., Senior lecturer of the CS Department Molchanova M.O.