RAAVI
An International Journal of Multidisciplinary Research
ISSN:3108-253X
Special Issue : February,2026

Essential DBMS Tools and Techniques for Supporting Artificial Intelligence Development

Yogita Dattatray Solankar[1]

solankaryogita@gmail.com

Pooja Dattatray Solankar[2]

poojasolankar09@gmail.com

[1,2]*VidyaPratishthan's Arts, Science & Commerce College, Baramati*

ABSTRACT

The rapid evolution of Artificial Intelligence (AI) necessitates robust Database Management Systems (DBMS) to handle vast datasets efficiently. AI applications such as machine learning (ML), natural language processing (NLP), and computer vision rely on structured and unstructured data, demanding scalable, high-performance database solutions. This paper explores essential DBMS tools and techniques that support AI development, including relational databases, NoSQL databases, in-memory databases, and distributed data processing frameworks. Additionally, we examine optimization strategies, data integration methods, and emerging trends in AI-driven database technologies. The study concludes with recommendations for selecting appropriate DBMS solutions based on AI use cases, performance needs, and scalability requirements.

Keywords: Artificial Intelligence (AI), Database Management Systems (DBMS), Machine Learning (ML), NoSQL, Data Integration, Big Data

1. INTRODUCTION

Artificial Intelligence (AI) depends heavily on structured and unstructured data for training models, decision-making, and automation. Database Management Systems (DBMS)

Relish Publications,
Amrapali Apartment, Canal Road, Baramati (MS), India.
Pin code- 413102

Website: www.relishpublications.com

Email: raavipublications@gmail.com

WhatsApp: 9665841996

ISSN:3108-253X
Special Issue : February,2026
RAAVI
An International Journal of Multidisciplinary Research

serve as the backbone for data storage, retrieval, and processing in AI applications. Traditional relational databases (RDBMS) and modern NoSQL systems provide different trade-offs in scalability, consistency, and flexibility, making their selection critical in AI development [1].

This paper investigates key DBMS technologies essential for AI development, including:

- Relational databases (RDBMS) for structured data
- NoSQL databases (document, key-value, graph) for unstructured/semi-structured data
- In-memory databases for real-time AI processing
- Distributed databases (e.g., Hadoop, Spark) for big data analytics

We also discuss emerging techniques such as vector databases for AI-driven search and knowledge graphs for semantic reasoning.

2. RELATIONAL DATABASES IN AI DEVELOPMENT

Relational databases (RDBMS) such as MySQL, PostgreSQL, and Oracle remain foundational for structured AI data. AI applications requiring ACID (Atomicity, Consistency, Isolation, Durability) compliance benefit from SQL-based querying for:

- Feature Engineering: Structured data preparation for ML models [2]
- Join Operations: Combining multiple datasets for training
- Transaction Management: Ensuring data integrity in AI-driven automation

However, scaling RDBMS for big data can be challenging due to performance bottlenecks in distributed systems.

3. NOSQL DATABASES FOR UNSTRUCTURED AI DATA

NoSQL databases excel in handling heterogeneous, unstructured datasets common in AI:

Relish Publications,
Amrapali Apartment, Canal
Road, Baramati (MS), India.
Pin code- 413102

Website: www.relishpublications.com

Email: raavipublications@gmail.com

WhatsApp: 9665841996

3.1 Document Stores (MongoDB, Couchbase)

Store JSON-like documents for NLP and recommendation systems

Flexible schemas accommodate dynamic AI training data

3.2 Key-Value Stores (Redis, DynamoDB)

High-speed read/write operations for real-time AI decision-making

Used in session management and caching for AI-driven applications

3.3 Graph Databases (Neo4j, ArangoDB)

Optimized for relationship-heavy AI models (social networks, fraud detection)

Enable knowledge graphs for semantic AI applications [3]

3.4 Wide-Column Stores (Cassandra, HBase)

Designed for horizontal scalability in big data AI applications

Efficient for time-series data in predictive analytics

4. IN-MEMORY DATABASES FOR HIGH-PERFORMANCE AI

In-memory databases (e.g., Redis, MemSQL, SAP HANA) minimize disk I/O, accelerating AI real-time analytics and model inference. Benefits include:

Low-Latency Querying: Essential for autonomous systems and chatbots

Stream Processing: AI applications requiring immediate data analysis (e.g., financial fraud detection)

5. DISTRIBUTED DATA PROCESSING FRAMEWORKS

5.1 Apache Hadoop

Distributed storage (HDFS) for large-scale AI datasets

Map Reduce for batch processing in ML pipelines

Relish Publications,
Amrapali Apartment, Canal
Road, Baramati (MS), India.
Pin code- 413102

Website: www.relishpublications.com

Email: raavipublications@gmail.com

WhatsApp: 9665841996

5.2 Apache Spark

In-memory processing for iterative ML algorithms

Structured streaming for real-time AI model updates

5.3 Delta Lake & Data Lakes

Enables ACID transactions on big data (useful for AI training datasets)

Unified storage for structured and unstructured AI data [4]

6. EMERGING DBMS TECHNIQUES FOR AI

6.1 Vector Databases (e.g., Pinecone, Milvus)

Optimized for similarity search in AI embeddings (recommender systems, NLP)

Efficiently store and retrieve high-dimensional vectors

6.2 Knowledge Graphs

Enhance AI reasoning with semantic relationships (e.g., Google Knowledge Graph)

Improve explainability in AI decision-making processes

6.3 Automated Database Tuning for AI

AI-driven DBMS optimization (e.g., Oracle Autonomous Database)

Self-tuning indexes and query optimization for ML workloads

7. RECOMMENDATIONS FOR DBMS SELECTION IN AI PROJECTS

AI Use Case	Recommended DBMS	Key Benefits
Structured Data (ML)	**PostgreSQL, MySQL**	**ACID compliance, SQL joins**
Unstructured Data (NLP)	MongoDB, Elasticsearch	Schema flexibility, full-text search
Real-Time AI (DL)	Redis, Apache Kafka	Law-latency event streaming

Relish Publications,
Amrapali Apartment, Canal
Road, Baramati (MS), India.
Pin code- 413102

Website: www.relishpublications.com

Email: raavipublications@gmail.com

WhatsApp: 9665841996

Big Data (AI Analytics)	Hadoop, Spark, Delta Lake	Scalability, batch/stream processing
Semantic AI (KG)	Neo4j, Amazon Neptune	Relationship-based querying

8. CONCLUSION

The success of AI systems depends on efficient data management. While relational databases remain essential for structured data, NoSQL and distributed databases provide scalability for AI's growing data demands. Emerging technologies like vector databases and knowledge graphs further enhance AI capabilities. Future research should focus on integrating AI-driven autonomous database tuning with ML pipelines for self-optimizing DBMS architectures.

9. REFERENCES

[1] C. Coronel and S. Morris, Database Systems: Design, Implementation, & Management, 13th ed. Cengage Learning, 2018.

[2] A. Ng, "Machine Learning Yearning," DeepLearning.ai, 2018.

[3] J. Pokorný, "Graph Databases: Their Power and Limitations," Computer Science and Information Systems, vol. 12, no. 2, pp. 95–114, 2015.

[4] M. Armbrust et al., "Delta Lake: High-Performance ACID Table Storage over Cloud Object Stores," Proc. VLDB Endow., vol. 13, no. 12, 2020.

[5] L. George, HBase: The Definitive Guide, O'Reilly Media, 2011.

[6] S. Chaudhuri and U. Dayal, "An Overview of Data Warehousing and OLAP Technology," ACM Sigmod Record, vol. 26, no. 1, pp. 65–74, 1997.

❑❑❑

Relish Publications,
Amrapali Apartment, Canal Road, Baramati (MS), India.
Pin code- 413102

Website: www.relishpublications.com

Email: raavipublications@gmail.com

WhatsApp: 9665841996