ANALYSIS OF AI DATABASES

Arina CARAUS

Department of Software and Automation Engineering, CR-212, Faculty CIM,
Technical University of Moldova, Chişinău, Republic of Moldova

Corresponding author: Arina Caraus, arinacaraus@gmail.com

Scientific advisor/coordinator: **Dorian SARANCIUC**, university lecturer,
Department of Software and Automation Engineering

Summary. *Nowadays, data is money. Business operations and success rely on the ability to get really important information from databases.That way they can strategize to get changes in real time, allowing them to optimize processes, reduce delays and manage assets efficiently. The solution used by enterprises is artificial intelligence databases, which can analyze vast streams of data that can be leveraged to open up deep business opportunities. In this article, AI databases(DB) are analyzed through the business point of view with the goal to let people understand where these technologies can be used and where not. Firstly, the key characteristics of AI DB are enumerated. Secondly, the workflow of the different databases are described, to bring a more open view on the theme and to let people carry out a rational analysis of the opportunities offered by both databases . The decision of choosing the database should be made depending on the specific goals and needs of the business, not on hype.*

Key words: data, vector database, business, embeddings

Introduction

The AI database is like a behind-the-scenes hero, quietly but effectively tackling the complex data storage and manipulation requirements of artificial intelligence and machine learning. It also works with massive datasets, convoluted structures, and complicated queries to power sophisticated AI operations [1].

Artificial intelligent databases aim to optimize computing resources and databases. They can simultaneously assimilate, explore, analyze and visualize complex, fast-moving data in milliseconds [2]. They allow handling unstructured data in a professional way, especially when managing vector embeddings. Embeddings are completely suitable for storage, but can be computationally heavy. This is one reason why Milvus databases use GPU acceleration - it increases performance and maintains the workflow of AI to run smoothly.

Key characteristics of AI databases:

- *vector storage* – is one of the most efficient ways to represent and query of high-dimensional data, like area embeddings for ML models;
- *scalability* – possibility to handle a large amount of data having a horizontal scaling without compromising on performance;
- *complex query support* – can be done different complex queries, mostly important for searches, ranking, and pattern recognition*;*
- *real-time* – possibility to analyze data in real time and to get insights from changes in milliseconds, which can optimize the business process and to react really fast to changes;
- *integration with ML Frameworks* - unstructured data can be converted as vector embeddings , with an ML model, in an AI Database;

- *flexibility* - can be used different data types and structures and the dynamic schema, which allow seamless changes to data structures without disrupting database operations;
- *parallel processing* - it is the key for more dynamic simulation and modeling of the real-world data.

How does an AI Database work?

AI databases offer the possibility to process real-time streaming data. So , they can react to changes as fast as possible [3].

One example of AI Database is Google's AutoML. This example is a new approach to ML models. It is redirected to generate new neural network architectures based on particular data sets. The model is tested and iterated, as many times as is needed, for generating the new neural net architecture. The development of this machine-generated architecture will help non-experts to create neural nets fitted for their needs [4].

AI databases offer fast data entry, the possibility of parallel processing between database modes, which can be associated with multitasking at the person level, and a common space for software engineers, data scientists, database administrators to go more quickly through them and test models and to be able to use results directly to analyze.
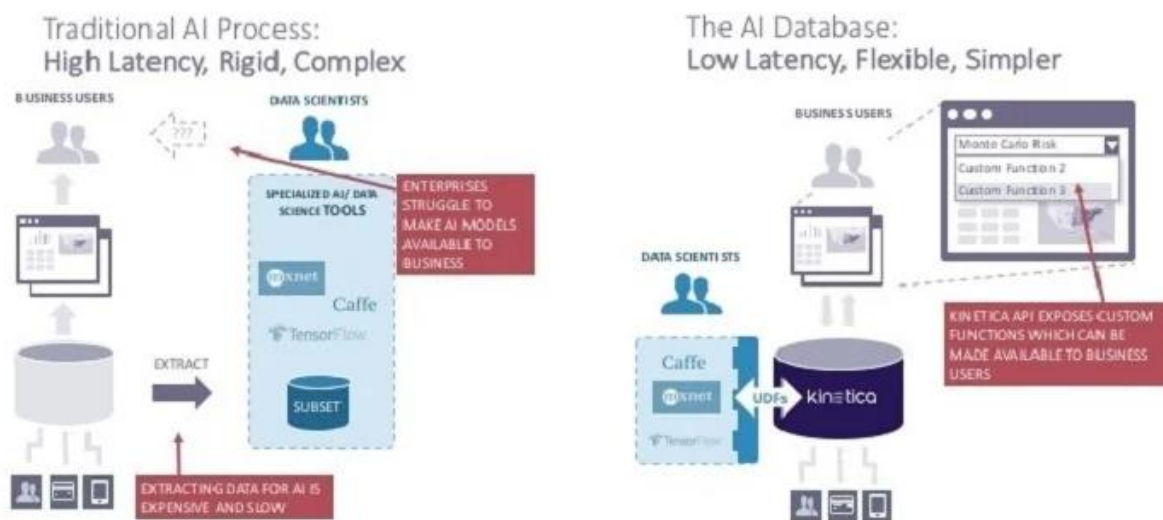

Figure 1. The workflow of an AI Database

In order to redeem the business potential through data, it is advisable to allocate investments towards engaging qualified experts. Data Scientists will choose the most suitable database for the project, from the options below:

1. *Relational Database* - This type of database is used to handle structured data organized into rows and columns, which form tables, with predefined formats. AI Similarity Search (FAISS), IVFFLAT, or Hierarchical Navigable Small Words have incorporated vector search indexes to enhance their projects and facilitate straightforward vector searches.

2. *Vector Database* – This type is designed to handle data represented in multi-dimensional vector space. It is suitable for AI and ML applications, where data takes the form of vectors as image embeddings, video embeddings, text embeddings, or other types of vectors.

3. *Other DataBases*: Search Engine database and NoSql. They have recently incorporated basic vector capabilities, making it more realistic to handle vector-related tasks.

The difference between the traditional database (relational database) and the vector database is huge [5]. Traditional databases work with discrete, scalar data types such as strings and numbers, organizing them into tables, whereas vector databases work with complex, high-dimensional data.

Table 1

Difference between Traditional Database and Vector Database

Feature	Traditional Databases (E.g. MySQL, MongoDB)	Vector Databases
Data Structure	The main form is tabular or specific data structures	Designed for multi-dimensional data
Optimization for High Dimensionality	May require optimization for high-dimensional operations	Rafted specifically for high-dimensional spaces immune to the curse of dimensionality
Role in AI & Machine Learning	General-purpose storage, might need customization for specific AI tasks	Indispensable for complex AI tasks due to their intrinsic design for high-dimensionality
Primary operations	Proficiency in CRUD operations	Efficient similarity searches and high-dimensional data operations
Indexing Techniques	Standard indexing techniques	Use advanced techniques like ANN algorithms for high-dimensional data.

Depending on the type of their data, developers should choose the ideal database.

They should take into consideration limitations of the traditional databases:
• *lack of flexibility* – it is not suitable for unstructured data and it has a rigid schema structure. It makes it challenging to adapt to changing data requirements;
• *performance bottlenecks* - it can face challenges with a large amount of data and complex queries;
• *scalability challenges* - scaling traditional relational databases horizontally (across multiple servers) is a complex and costly endeavor, which can limit their capacity to handle large amount of data;

Limitations of the vector database:
• *complexity* – for the solution development it is necessary to have expertise in this area;
• *data structure* - they are not suitable for structured data;
• *implementation challenges* – it may have challenges considering the compatibility with other existent systems.

DATA TYPE	TRADITIONAL DATABASES	VECTOR DATABASES
Text (Short, Structured)	✔	
Numbers	✔	✔
Dates/Time	✔	
Images		✔
Audio Files		✔
Videos		✔
Text (Long, Unstructured)		✔
Sensor Data		✔

Figure 2. Data Types and Their Ideal Databases

The vector databases store and manage the vector data – which serve as numerical codes that encapsulate the key characteristics of the object, for example: images in a streaming app [6]. By analyzing and extracting crucial features, each image is converted into a vector embedding through an embedding model.
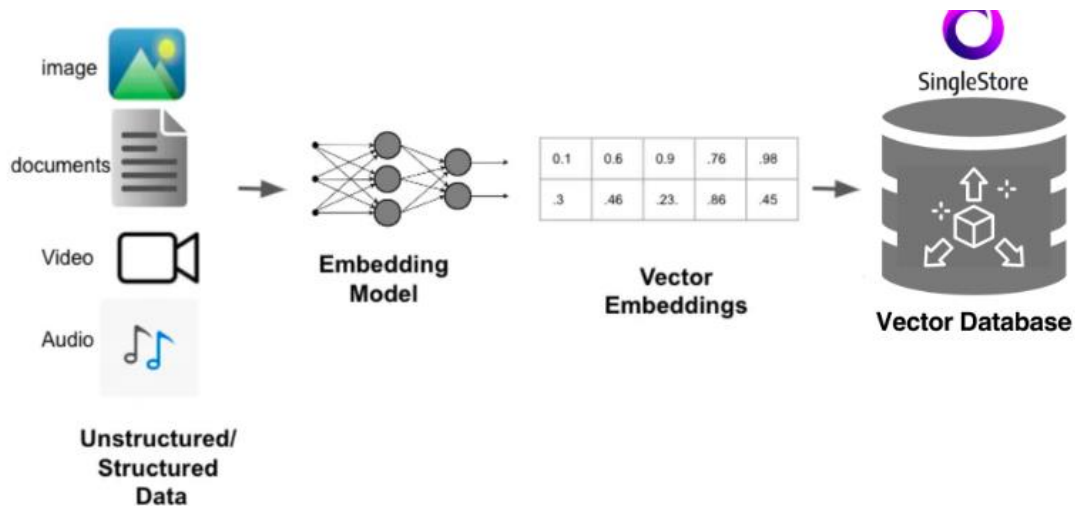
How does a vector database work?

Figure 3. The workflow of a vector database

Initially, all structured or unstructured data is processed through an embedding model. This model is often a complex neural network, translating data into high-dimensional numerical vectors and effectively encoding the data's characteristics into vector embeddings – which are stored into a vector database [7].

At the moment we are writing a query, the vector database executes operations (such as similarity searches) to find and retrieve the vectors most similar to the query, efficiently handling complex queries and providing relevant results for the user. This entire process allows a fast and accurate management of large and varied data types in applications that require high-speed lookup and fetch functions.

Where is it used in business?

The usage of the AI Database is the possibility to optimize compute and DB's resources. As Amit Vij, President and Co-founder of Kinetica ,,"The world is now instrumented in a way that is always on, always tracking, always monitoring, always listening, and always watching." [8]

It is really important to be informed about the business situation and to get real-time insights. As a result, in different areas there are integrated sensors that create a large amount of data streams. For the assimilation of the streamed data current systems are expensive and have a high compute cost. Here it is needed to have a strategic approach. The problem should be solved from the root. The companies need to focus on real-time decisions and determination of the context mechanism.

It is valuable to see the relationships between different data sets when parsing sensors and machine data in high-stakes scenarios.

Conclusion

People should choose the database depending on their goals. AI Databases is the best solution for analyzing the unstructured data. It offers the possibility to get real-time insights, which allows us to react fast to different challenges. Best solution is a vectorized database, as Vij said "A vectorised database lets organizations marry up their streaming inventory data with their

fleet data, for instance, and pair that with consumers in different regions to provide real-time replenishment. "

For modern businesses, patching gaps in existing data architectures is not merely advantageous, but crucial for survival in a competitive landscape.

Bibliography:
[1] MIT-Making-Sense-of-Sensor-Data. [pdf] [accessed 09.01.2024].
[2] AI Databases: What They Are and Why Your Business Should Care. [online] [accessed 10.01.2024] Available: https://www.pcmag.com/news/ai-databases-what-they-are-and-why-your-business-should-care
[3] Using Machine Learning to Explore Neural Network Architecture.[online] [accessed 12.01.2024] Available:https://blog.research.google/2017/05/using-machine-learning-to-explore.html
[4] AI Database.[online] [accessed 12.01.2024] Available: https://zilliz.com/glossary/ai-database
[5] What is an AI Database.[online] [accessed 13.01.2024] Available: https://redis.com/nosql/ai-databases/
[6] Understanding Vector Databases în Generative AI Evolution.[online] [accessed 14.01.2024] Available: https://medium.com/@hello_21549/understanding-vector-databases-in-generative-ai-evolution-7014190a682a
[7] Revolutionizing Computation With Hyperdimensional Computing. [online] [accessed 14.01.2024] Available: https://techandtasty.com/revolutionizing-computation-with-hyperdimensional-computing/
[8] Vector Database [online][accessed 14.01.2024] Available: https://www.gigaspaces.com/data-terms/vector-database