happiest minds
The Mindful IT Company
Born **Digital** . Born **Agile**

ENHANCING EDUCATIONAL DATA MANAGEMENT AND DATA PRIVACY **with Graph Database**

CONTENTS

1

ABSTRACT

The Education Technology (EdTech) sector, driven by the growth in digital technology, is defined by the creation and use of large amounts of data, thus transforming traditional learning methods. With the increasing dependence on digital tools, the amount of educational data produced increases. Here are several significant challenges to effectively manage and utilize this data. Such as:

- **Data Management, Fragmentation and Interoperability**
- **Data Privacy and Security**
- **Data Analysis and Interpretation**
- **Infrastructure Challenges**
- **Digital Divide and Equity**

The presence of large amounts of data and various internet-based tutorial courses reminds about the importance of a good recommendation engine. Using this voluminous data, Graph databases can recommend things based on different search parameters and user profiles.

They assist in responding to data management issues typically encountered within EdTech digital solutions.

2

INTRODUCTION

The global EdTech (Education Technology) market size is anticipated to grow to more than $800 billion by 2033, and in India, it is expected to reach $29 billion by 2030.

Global EdTech Market

As of 2023, the worldwide EdTech market was worth more than $334 billion

In 2023, North America accounted for the largest share of the overall EdTech market globally and controlled over 37% of the revenues

EdTech is expanding as the use of digital devices, the internet, and artificial intelligence is growing

The industry is expanding at a CAGR of 13.9% from 2024 to 2033

Indian EdTech market

The Indian EdTech market was $7.5 billion in 2024.

It is anticipated to reach a CAGR of 25.8% by 2030.

India's subsegment EdTech market is growing fastest in test preparation, K-12, online certification, and skill development.

The 2019 pandemic accelerated the use of online learning platforms, virtual classrooms, and digital learning materials

With the market dominated by size and capacity, certain critical issues and challenges have to be addressed:

Data Management, Fragmentation and Interoperability

Disparate Systems: Learning information is usually dispersed across different platforms, making learning management systems and apps challenging to collate and analyse

Lack of Standardization: Poorly designed standard data formats and standards hinder the seamless exchange of information between systems

Data Privacy and Security

Sensitive Information:

- Educational data often holds sensitive information, such as student details, academic records, and behavioral data
- Protecting this information from unauthorized access and breaches is of utmost importance

Regulatory Compliance: EdTech organizations must comply with various data security regulations, including FERPA (United States) and GDPR (European Union), which are often complex and demanding

Data Analysis and Interpretation

Meaningful Insights:

- Extracting meaningful insights from vast datasets requires advanced analytical skills and tools
- Converting information into actionable strategies that maximize learning outcomes remains challenging

Data Quality: It required precise and credible data gathered for effective analysis

Infrastructure Challenges

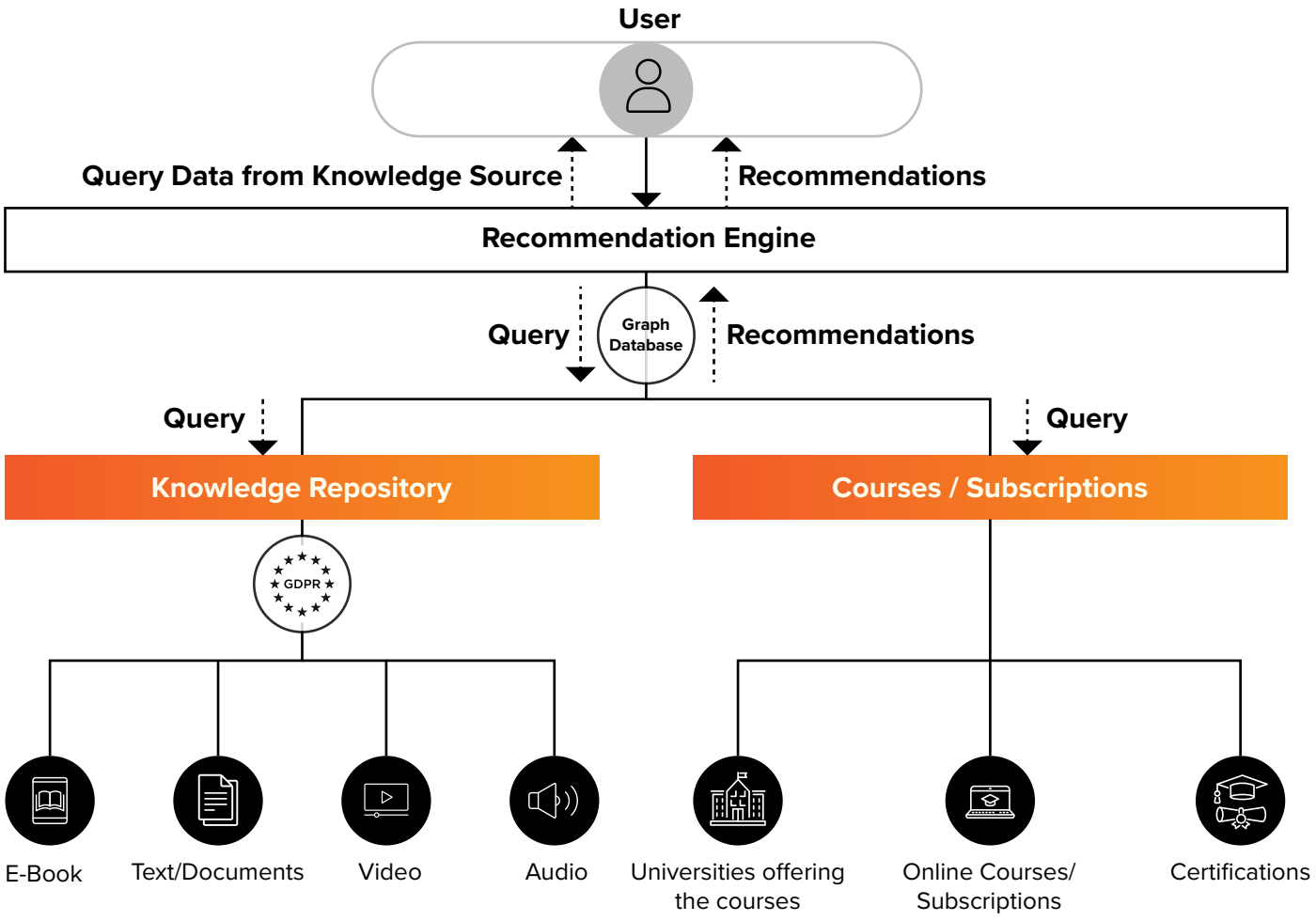
Scalability: EdTech platforms have a strong infrastructure to handle the growing volume of data

Reliability: Consistent and reliable access to technology is crucial to the data collection and utilization

Digital Divide and Equity

Unequal Access: Numerous students lack access to technology and reliable internet connections, which can result in inequities in data collection and analysis

Bias in Data: Data can reflect and perpetuate existing biases, leading to inequitable outcomes



This whitepaper discusses the data management challenges in the education industry and outlines a solution to address them.

3 BACKGROUND

To meet the above-discussed data-related issues, especially those concerning data management and performance, the graph database has some special features. We will discuss the features of graph databases.

A **Graph database (GDB)** is a NoSQL database that stores information using nodes, edges, and properties, other than tables or documents.

How Does a Graph Database Work?

A Graph database consists of:

Nodes

Represent entities in the data

Edges

Represent relationships between nodes

Properties

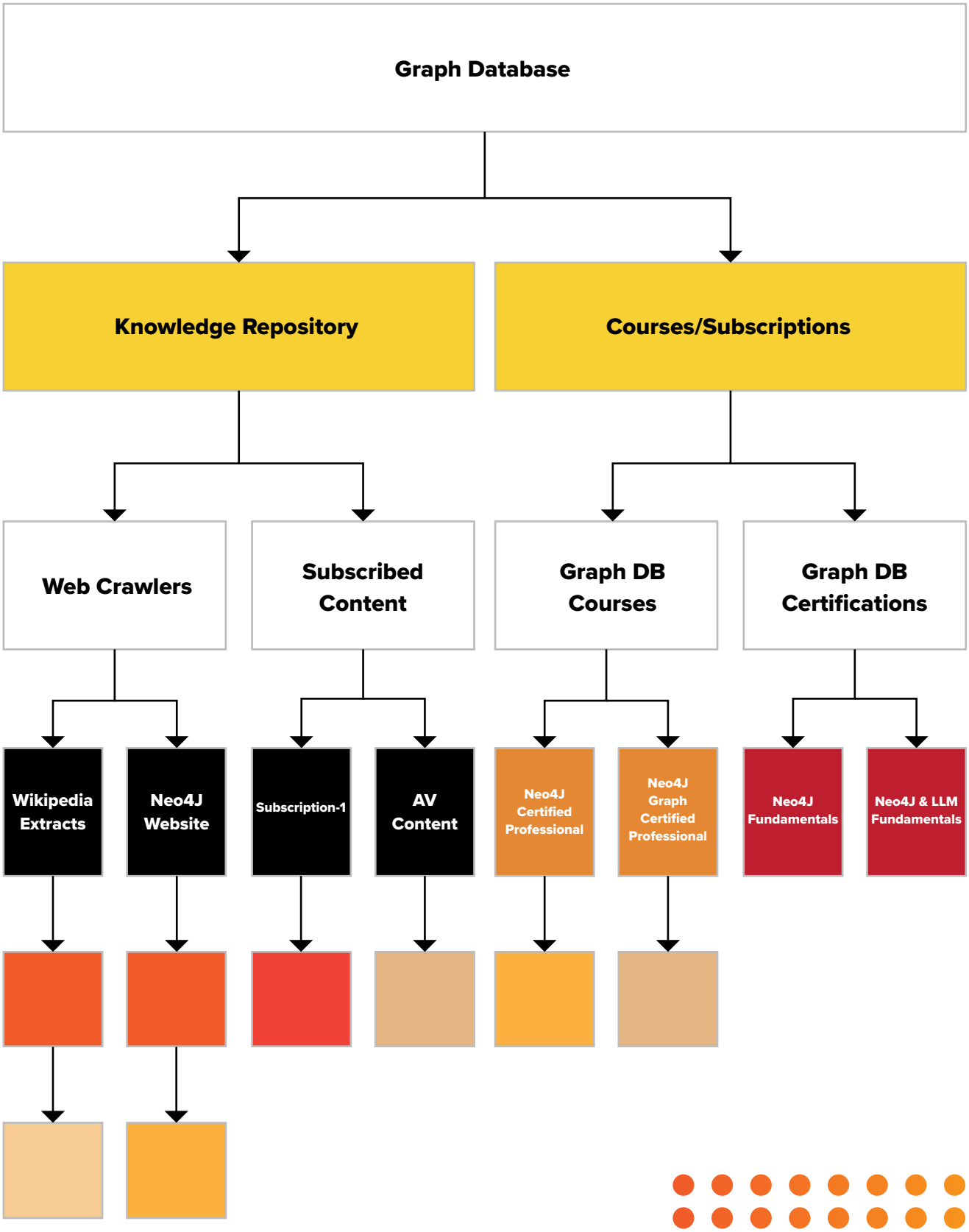
Additional information about nodes and edges

A graph database can help

To retrieve data more easily and quickly.	To discover patterns and insights across large amounts of data.

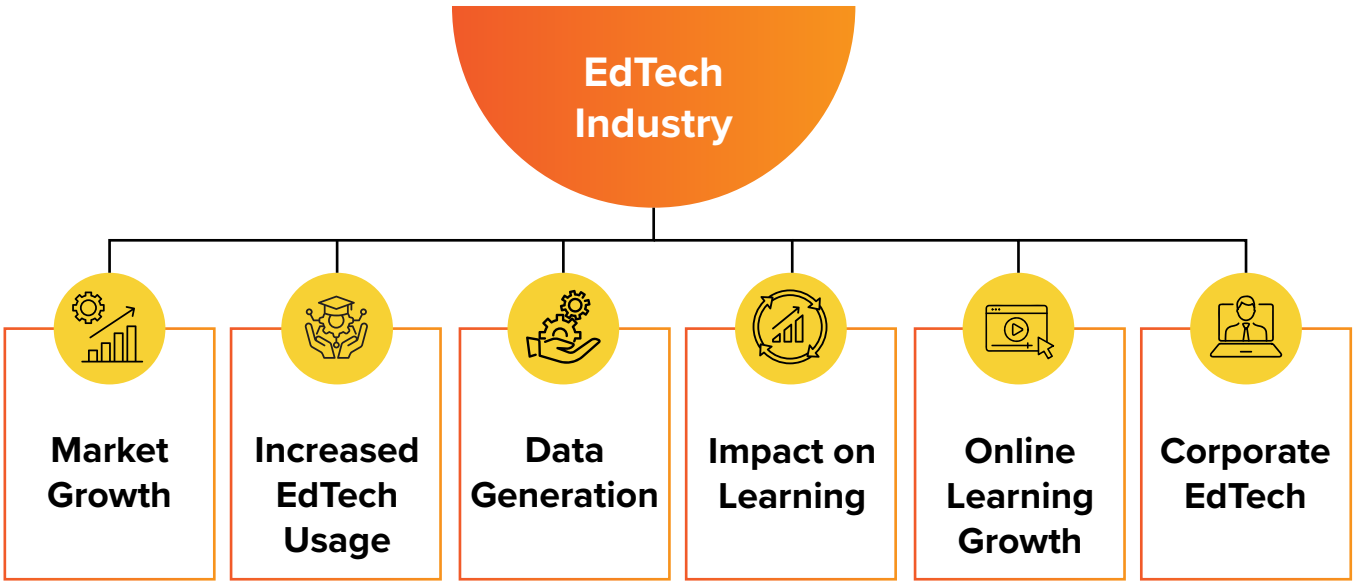
According to mathematical Graph theory concepts, a graph database enables efficient navigation among the nodes/entities by utilizing the optimal path for retrieving results. Compared to relational databases, the data is in static table structures, whereas a graph database holds data as a network of entities and relationships. This leads to enhanced performance and flexibility.

The diagram below illustrates how a graph database makes it easier to traverse nodes to get all the data as per the query.



Here are the recent trends in the EdTech Industry

The digitized EdTech sector is characterized by large-scale data creation, which is revolutionizing conventional educational practices. The information below gives an overview of this new environment, backed by extensive statistics.



Key Trends and Statistics

1 Market Growth

The global EdTech sector is expanding rapidly. The estimates are reaching a value of $605 billion by the year 2027, which reflects the rapid adoption of e-learning trends

This growth is driven by factors like increased internet penetration, the rise of mobile learning, and the growing demand for personalized learning experiences

2 Increased EdTech Usage

Adoption of EdTech by K-12 institutions has accelerated, resulting in a 99% increase in the scale of technology in classrooms since 2020

Moreover, a substantial number of teachers are employing EdTech regularly. According to a survey, 79% of teachers employ EdTech daily, which indicates the key position of EdTech in contemporary education

3 Data Generation

The digitization of education has led to the collection of vast amounts of data, including:

- Student performance data (grades, test scores, progress tracking)
- Learning behavior data (time spent on tasks, interaction with digital resources)
- Personalized learning data (preferences, learning styles)

4 Learning Impact

EdTech tools prove to have a positive impact on student performance. Studies show that many university students feel that EdTech solutions help improve their academic performance and the effectiveness of their learning process

For instance, more than 80% of them confirm that EdTech assists them in improving their grades, learning more, and studying better

5 Online Learning Growth

The online education market is growing at a fast pace, with expected significant revenue growth in the future. It increased due to the usage and adoption of online learning platforms

6 Corporate EdTech

The corporate sector embraces EdTech for development and training, with a steadily increasing market value. This is because of the need for continuous upskilling and on-the-job learning

Data Handling Implications

The information produced by EdTech tools is a double-edged sword

trong data analysis can support personalized learning, highlight learning gaps, and improve educational outcomes

However, it also raises questions regarding data privacy, security, and ethical considerations

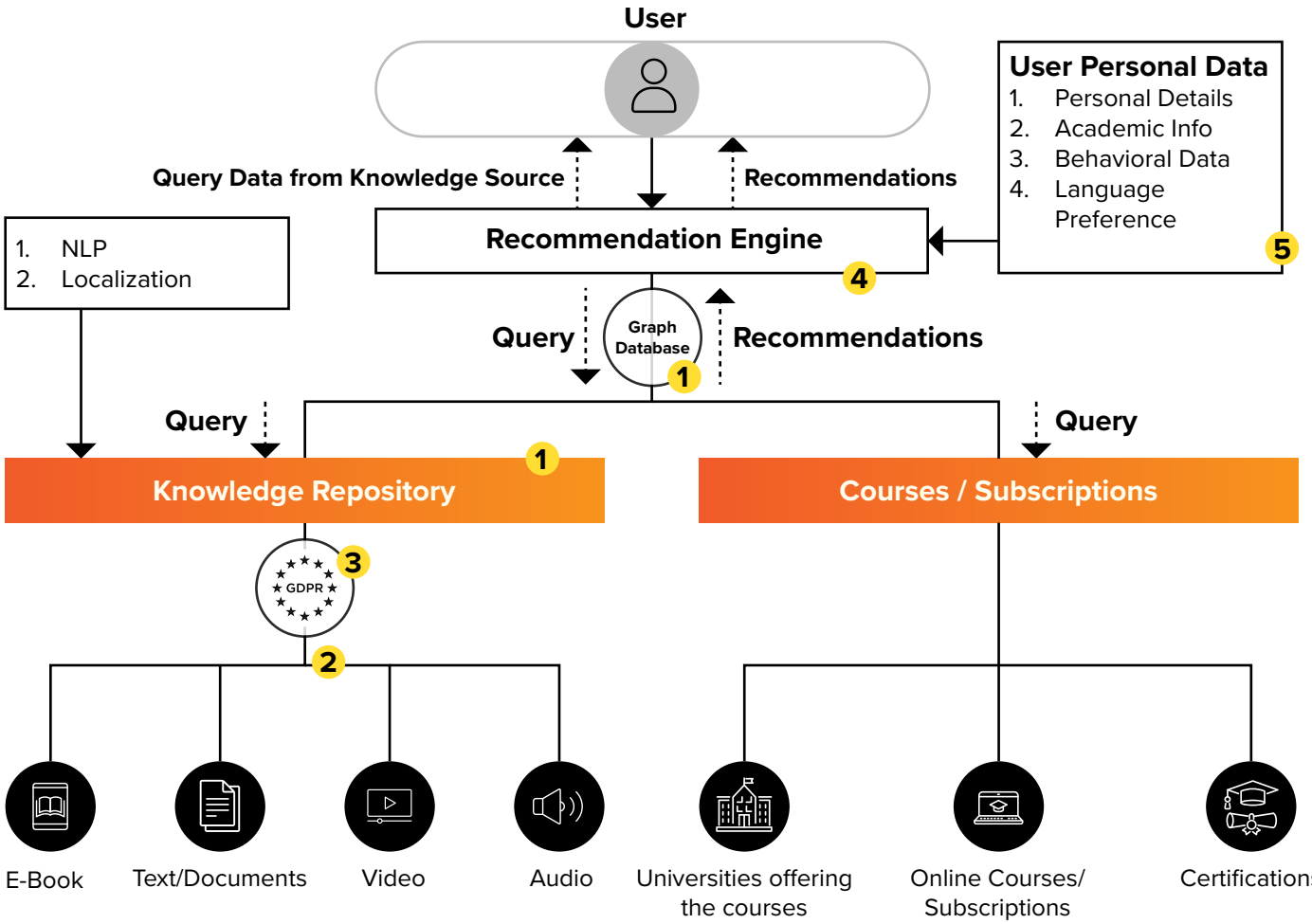
In short, the digitized educational technology sector is revolutionizing by integrating technology and data analysis. Rapid expansion of this market, adoption of education technology tools, and generation of large volumes of data define the future of learning.

4

PROPOSED SOLUTION

Using the graph database, the solution addresses the following key challenges:

1. **Data Management, Fragmentation and Interoperability**
2. **Data Privacy and Security**
3. **Data Analysis and Interpretation**
4. **Infrastructure Challenges**
5. **Digital Divide and Equity**



Data Management, Fragmentation and Interoperability

Graph database technology solves challenges in data management by bringing data into a flexible and consolidated framework, thereby providing end-to-end insights into the data. One of the benefits of graph databases is that they can extract all the discrete data from various sources

The Elasticsearch techniques are utilized by the APIs to retrieve the metadata and load the data into the graph database. Data collected is then transformed into standard formats for interoperability

Data Privacy and Security

Compliance is the foundation for global visibility and accessibility of content. Neo4j's Privacy Shield provides a comprehensive solution for meeting the EU's GDPR. It brings your data together on all your systems and gives you a single view

The Neo4j Privacy Shield allows individuals to view and manage personal data while allowing internal teams to respond to private enquiries on time. It also enables privacy managers to track data flows, identify possible breaches, and prove compliance to regulators

Data Analysis and Interpretation

The graph topology allows billions of data points to be processed in seconds, uncovering hidden relationships that enhance predictive accuracy. As data traverses through the nodes, insights begin to emerge. The graph database also uses machine learning algorithms to generate various useful insights.

Infrastructure Challenges

Some key infrastructure problems associated with graph databases are scalability issues in dealing with large data, optimization of complex queries for complex relationships, distributed data across multiple nodes, and an increased learning curve due to specialized query languages such as Cypher. There may also be performance overhead when dealing with data that does not heavily depend on relationships.

While modern graph databases are evolving to address most of these challenges, they still present unique infrastructural challenges compared to traditional relational databases. The level of these challenges has been significantly reduced with increased compatibility.

Digital Divide and Equity

Internet-retrieved content is of diverse types in different geographies. Localizing and standardizing content by user language preference is easily handled in a graph database, as it offers interoperability support

The effectiveness of a graph database is based essentially on its ability to define relationships and trace towards leaf nodes. In addition, the recommendation system is accessing user-subscribed content, educational courses, and certification information, and thus provides extra assistance to users in finalizing their search

In conclusion, the graph database manages complexities across data points efficiently and handles vast amounts of data, from billions of nodes to trillions of edges. This characteristic enables it to handle highly connected data sets and respond rapidly to dramatic volume data growth. Nevertheless, real capacity depends on underlying hardware and database implementation.

Most important graph database data volume management issues:

Scalability
Graph databases are designed to scale out over multiple servers, supporting the handling of large datasets without any degradation in performance.

Relationship-Oriented
Unlike relational databases, graph databases persist relationships directly between data points, allowing for rapid traversal of intricate connections.

Performance Advantages
The relationship-based model facilitates rapid execution of queries, even for large and interrelated datasets.

Other applications involve scaling the existing solution to other learning platforms and using Generative AI to develop large language models (LLMs).

Use Cases

Below are a few interesting use cases from the EdTech industry, using graph databases:

Use Case 1

Knowledge Graphs: Enabling Intelligent Apps

READ NOW

Use Case 2

How graph databases can enhance learning

READ NOW

References

Graph Database https://en.wikipedia.org/wiki/Graph_database
Neo4j Graph database https://neo4j.com/
Amazon graph database ref. https://aws.amazon.com/nosql/graph
Elsevier Makes Science Research Accessible in Milliseconds with Neo4j https://neo4j.com/customer-stories/elsevier/
Oracle Spatial and Graph https://en.wikipedia.org/wiki/Oracle_Spatial_and_Graph
Graph Query Language (GQL) https://en.wikipedia.org/wiki/Graph_Query_Language

About the Author

Raghunandan S
Director Engineering, PDES

Raghunandan has over 22 years of experience in delivering engineering software and embedded solutions across various business domains, including transport/mobility, industrial products, building technologies, life sciences, energy utilities, and retail.

He successfully delivered enterprise and business-critical applications and services, process automation, legacy technology migration, integration programs/projects, application rollouts, walkthroughs, and end-user training.

Raghu holds an M. Tech in Software Systems from BITS Pilani and a B. Tech in Electrical and Electronics Engineering from Mysore University. He is also certified as a PMP from PMI, a Scrum Master from SEI, and a PSP/TSP certification from SEI.

He has previously worked with Cognizant Softvision Technologies, Infosys Technologies, and L&T Technology Services.

About Happiest Minds

Happiest Minds Technologies Limited (NSE: HAPPSTMNDS), a Mindful IT Company, enables digital transformation for enterprises and technology providers by delivering seamless customer experiences, business efficiency and actionable insights. We do this by leveraging a spectrum of disruptive technologies such as: artificial intelligence, blockchain, cloud, digital process automation, internet of things, robotics/drones, security, virtual/ augmented reality, etc. Positioned as 'Born Digital. Born Agile', our capabilities span Product & Digital Engineering Services (PDES), Generative AI Business Services (GBS) and Infrastructure Management & Security Services (IMSS). We deliver these services across industry groups: Banking, Financial Services & Insurance (BFSI), EdTech, Healthcare & Life Sciences, Hi-Tech and Media & Entertainment, Industrial, Manufacturing, Energy & Utilities, and Retail, CPG & Logistics. The company has been recognized for its excellence in Corporate Governance practices by Golden Peacock and ICSI. A Great Place to Work Certified™ company, Happiest Minds is headquartered in Bengaluru, India with operations in the U.S., UK, Canada, Australia, and the Middle East.

For more information, write to us at
business@happiestminds.com

happiest minds
The Mindful IT Company
Born **Digital** . Born **Agile**

www.happiestminds.com