SteelBridge LABS

TECH STRATEGY BEST PRACTICES:

GRAPH DATABASES AS THE NEW DEVELOPMENT PARADIGM

STEELBRIDGE LABS INSIGHTS

JUNE 2019

INTRODUCTION

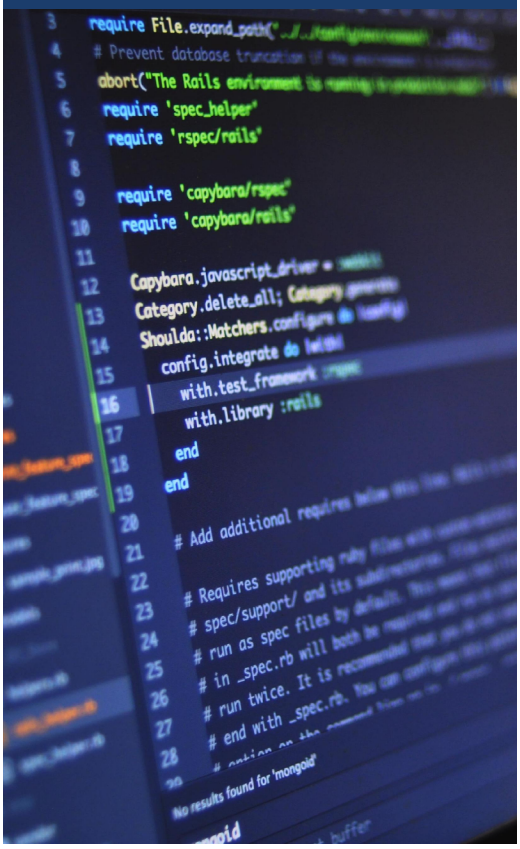
In recent years, digital storage has become more affordable with the accessibility of cloud-based technology. The core challenge, however, remains the same: utilizing massive volumes of data to gain business value. Selecting the proper database engine to perform complex data queries is a major investment for firms, and relatively new solutions with graph-based features are now being considered. Navigating the database engine landscape appears daunting, but key solutions stand out from the crowd.

www.steelbridgelabs.com | Pittsburgh, PA | hello@steelbridgelabs.com

THE DATABASE DILEMMA

"A graph database and its ecosystem of technologies can yield elegant, efficient solutions to problems in knowledge representation and reasoning."[2]

— Dr. Marko A. Rodriguez, Project Management Committee Member, Apache TinkerPop™



Choosing the Proper Database

Firms tend to spend on new core technologies where they can develop expertise, integrations, and best practices, so the transition is conducted across many projects. Once a database system has been designed and implemented, it is costly and risky to perform any migration to a different database engine.

Before a firm selects the proper database engine, burdens associated with the standard relational database solution must be considered. The relational database can create performance bottlenecks and additional maintenance when answering complex data queries.

Limitations of Traditional Models

The relational model has typically served as the primary resource for database research, since its introduction by Codd.[1] This database model is based on the simple notion of relation, along with algebra and logic. Under the concept of abstraction levels, a separation layer between the physical and logical levels was introduced. In particular, its standard query and transformation language (SQL) became a paradigmatic language for querying. However, there are limitations when employing this type of database engine, as described below.

1. **TROUBLE WITH COMPLEXITY:** The relational model is geared towards simple record-type data, where the data structure of inventories and accounting ledgers is known in advance.

2. **INFLEXIBLE SCHEMA:** It can be difficult to extend databases, and integrating different schemas is not an easy task that is far from being automatized.

3. **TABULAR DATA DIFFICULTIES:** When a database has many relationships stored in tabular data, its performance becomes slower and incurs high computational costs due to specific and difficult statements.

Although traditional relational databases are sufficient to represent and process graph-structured data, they fail to provide intuitive interfaces and efficient operations for graph queries. These limitations are addressed in new graph technology, which is becoming a viable solution for firms seeking control over their data.

[1] E. F. Codd. A Relational Model of Data for Large Shared Data Banks. Communications of the ACM, 13(6):377–387, 1970
[2] M. Rodriguez. Knowledge Representation and Reasoning with Graph Databases. Database Zone, 2011.

THE EMERGING SOLUTION

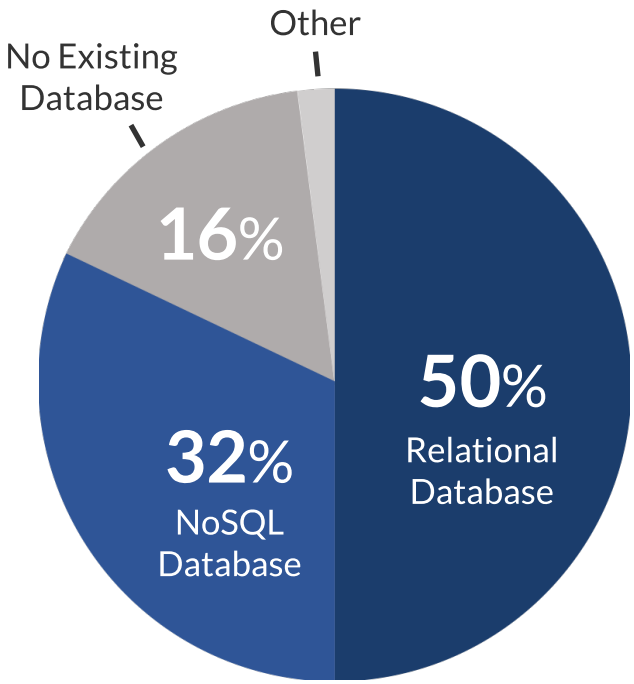
How do Graph Databases Differ?

Recently, storage and processing of graph-structured data have attracted significant interest from both industry and academia, leading to the development of many graph analytics systems and graph databases. The design of graph-oriented databases takes advantage of the simple structure of relationships by eradicating the abstraction layer. Data and relationships are effectively stored together as a graph and make up the interconnected nodes (vertices) via relationships (edges) that form the network of data. Graph databases explicitly store relationships to allow simple and fast retrieval of complex hierarchical structures, thus providing an intuitive abstraction to model objects and highly connected structures.

Making the Complex Simple

Graph models treat relationships as first-class source of knowledge and enable efficient traversals by native graph storage and index-free adjacency access. In this sense, graph models reduce complexity and eliminate the extra effort involved in transforming the data from the model to storage.

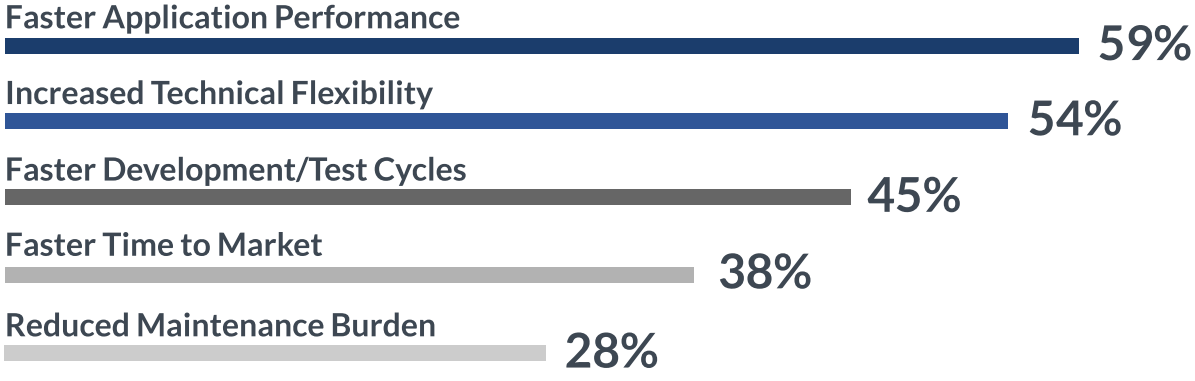
Representing highly-connected data in the relational model results in a large amount of many interfacing relations, which can produce complex, join-heavy SQL statements for graph queries. By focusing on the relationships among the entities, graph databases offer efficient processing of graph operations like reachability queries and pattern matching. Now, graph databases can provide efficient graph-centric operations, such as retrieving the neighbors of a vertex. Unlike in the relational model, this performance is not affected by the data size, such as the number of vertices in the graph.



BREAKDOWN OF DATABASE BY USER

Advantages of the Graph Model

The ability to map extremely complex data relationships is vital. Graph modelling technology opens the door to a range of business uses, from fraud detection and real-time recommendation engines. The following figure highlights the benefits that IBM survey respondents witnessed following implementation of a graph database.[3]



[3] IBM. The State of Graph Databases - Worldwide Adoption and Use Case Characteristics, 2017.

THE CURRENT LANDSCAPE

Market Players

Most existing graph database solutions, as seen below, employ the property graph model, offering a directed, edge-labeled multi-graph with an arbitrary number of key-value pairs attached to vertices and edges. Although most vendors have proprietary APIs and languages, open-source efforts are currently focused on unifying the graph processing space. The most-prominent example of this project is Apache TinkerPop™.

NEO4J

Neo4j's Graph Platform is specifically optimized to map, analyze, store, and traverse networks of connected data by intuitively mapping data points and the connections between them for the toughest enterprise challenges.

TITANDB

Titan is a transactional, scalable graph database optimized for storing and querying graphs containing hundreds of billions of vertices and edges distributed across a multi-machine cluster for concurrent users.

ORIENTDB

OrientDB Enterprise Edition 3.0 expands from an unparalleled legacy of powerful multi-model concepts to unify graph-document concepts, adding new APIs, and major OrientDB Studio enhancements.

Insights Highlight: Apache TinkerPop™

Although most vendors have proprietary APIs and languages, open-source efforts are currently focused on unifying the graph processing space. The most-prominent example of this project is Apache TinkerPop™.

The Apache TinkerPop™ stack provides a collection of tools and libraries for storage, querying, and analysis of graph-structured data. For example, the Gremlin Structure API and the Gremlin query language lie at the core of TinkerPop™ stack. While Gremlin Structure API provides a common set of interfaces for the property graph model, the query language defines a procedure structured around the API. Both of these can be considered analogues to JDBC and SQL for graph databases, providing a standardized, unified way of querying and processing data modelled as a property graph.

Strength in Numbers

In addition to the current market players offering graph-based solutions, a strong community surrounding this database model exists. There are significant advantages to the existence of a strong community based around a technology, particularly with databases. A database model with many tech-focused users makes it easier to find and hire developers that are familiar with the product. This database ecosystem enables firms and teams to uncover and share best practices, solutions documentation, and code samples, all of which mitigates risk as firms embark on new projects. The technical network associated with graph-based databases also helps organizations retain key technical talent.

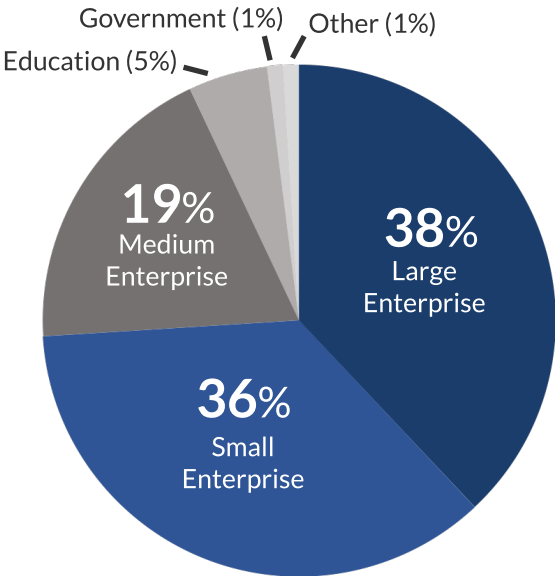
Graph-based database systems tend to provide rich query models where simple and complex relationships are integrated. Integrations with other solutions and technologies are needed in order for the database to make direct and indirect inferences about the data in the system. This strong community encourages other technology vendors to develop additional integrations and continue to participate in the ecosystem.

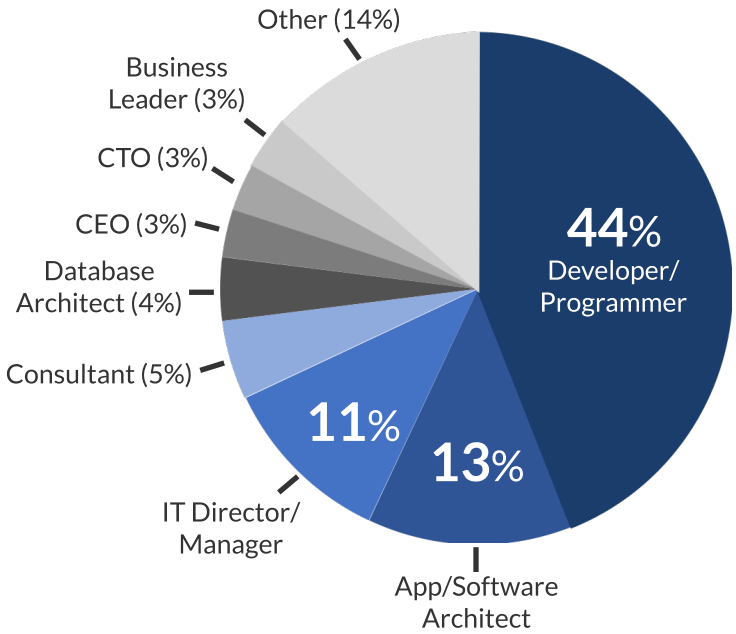
THE COMMERCIAL REACH

Graph Database's Growing Popularity

Global research conducted by IBM, in partnership with Tech Validate, reveals the growing potential that 1,365 technology developers, entrepreneurs, and leaders see for the graph-based database.[4] This survey uncovered how firms of different sizes in various industries consume graphs to address their most pressing problems.

Respondents represented a range of industries, with the majority in technical industries: computer services (42%) and computer software (22%). The featured pie charts demonstrate the range of firms that are taking advantage of all that graph-based databases have to offer their firms.





1. USAGE BY COMPANY SIZE

Although the greatest segment currently using graph databases is large enterprises, smaller enterprises nearly equal their adoption rates. This statistic reveals graph databases are beneficial to firms of varying size and, it can be assumed, varying levels of data complexity.

2. USAGE BY ROLE

It is no surprise that developers and programmers are the greatest user segment of graph databases. However, the breakdown of this usage reveals that various positions within an organization come into contact with the database to increase its value to the business.

Commercial Support

Overall, the graph community entertains a wide range of database solutions and best practices with solid commercial support and resource availability. However, both current users and future consumers should consider the health of their company or project when evaluating the range of database solutions. It is vital that their product not only continues to function properly, but also evolves and provides new features that suit customer needs. Having a strong, experienced support organization capable of providing services globally to a firm's customers is a major consideration before selecting and implementing a graph-based database.

[4] IBM. The State of Graph Databases - Worldwide Adoption and Use Case Characteristics, 2017.

THE FINAL TAKEAWAY

Key Conclusions

Cloud applications feature data that is constantly changing, large in size, and highly connected. A graph database is best at solving the various business problems that need to make use of this data. Unlike the traditional relational model, the graph-based model is the most capable technology for extracting value from it, so key business decisions can be quickly made.

As the technology landscape evolves, organizations increasingly find the need to evaluate new databases to support changing applications and business requirements. Key criteria when evaluating these technologies is the data archetypal analytics, popularity, query model, consistency, and APIs, as well as the commercial support and community strength available with a given solution. Our team encourages technology decision makers to evaluate these considerations for their needs and their firm's needs. Varying trade-off strategies demonstrate that graph database systems highly depend on the specific application domain and the required functionalities.

The media hype surrounding non-relational databases and the commensurate lack of clarity in the market means it is even more important for organizations to understand the differences between available solutions. Given the complexity of the application, it can be best to seek an experienced strategic partner to navigate the graph-based database landscape and best practices.

How SteelBridge Labs Can Help

Software Engineering

Reporting Solutions

Technical Expertise

At SteelBridge Labs, our Technology Solutions experts have worked day in and day out with graph databases for more than 5 years. Our team of Neo4j-certified engineers support successful selection and deployment of graph databases within complex enterprises of various sizes.

Our expertise, however, goes beyond graphs. We equally understand the relational database model, and are uniquely positioned to help design and convert your data from one model to the other using an efficient methodology.

We have successfully built finance platforms and leading data management and process-oriented software for firms with millions of dollars in assets under management (AUM). Startups and world-leading institutions have come to rely on our services and expertise.

For more information regarding our services, please contact us at hello@steelbridgelabs.com.

SteelBridge LABS

DREAM. CREATE. GROW.

SteelBridge Laboratories is an early-stage FinTech incubator. We partner with exceptional founders and management teams to equip them to create great software companies. The Lab provides a safe space for innovation to tackle real world problems. Entrepreneurs, coders, investors, and technologists looking to take an idea from its infancy and grow it from the ground up now have a place to call home.

MIAMI, FL

PITTSBURGH, PA

646.737.7960

HELLO@STEELBRIDGELABS.COM