Graph Databases


Jennifer Morales (C)
Contingent Worker

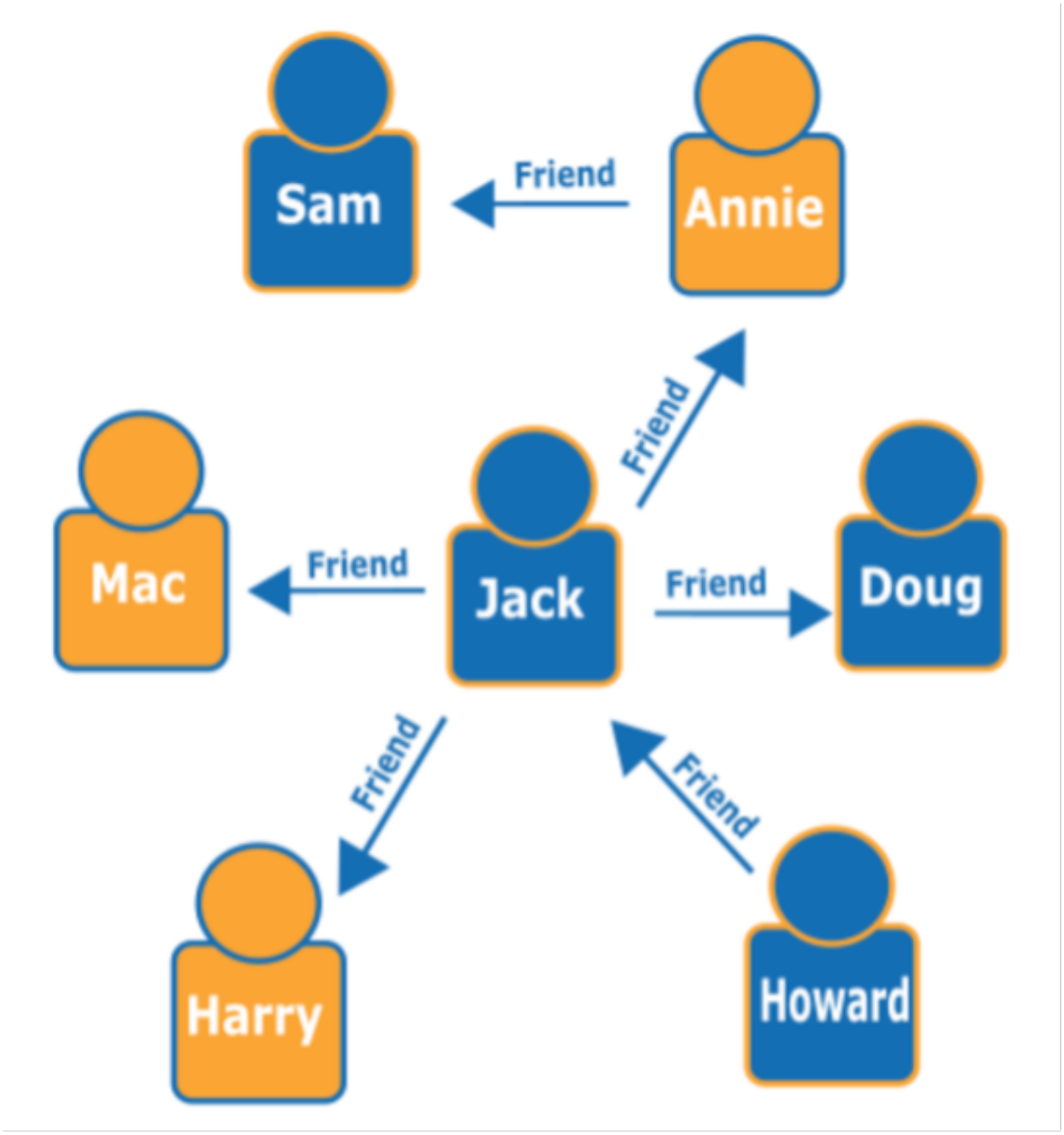
What is a Graph Database?

To better leverage data relationships, organizations need a database technology that stores relationship information as a first-class entity. That technology is a graph database.

Graph databases are purpose-built systems to store and navigate relationships. Relationships are first-class citizens in graph databases, and most of the value of graph databases is derived from these relationships.

A graph database stores nodes and relationships instead of tables, or documents. Data is stored just like you might sketch ideas on a whiteboard. Your

data is stored without restricting it to a pre-defined model, allowing a very flexible way of thinking about and using it.



People are the nodes, and the arrows are the relationships in the graph database above.

Graph-based solutions and applications are widely used across many business cases. Great examples include social media, GPS, search engines and ancestry.

Graph databases allow organizations to connect many different types of datasets that could be stored in data silos across and even outside of the organization. Graph databases provide the ability to look for data patterns and relationships.

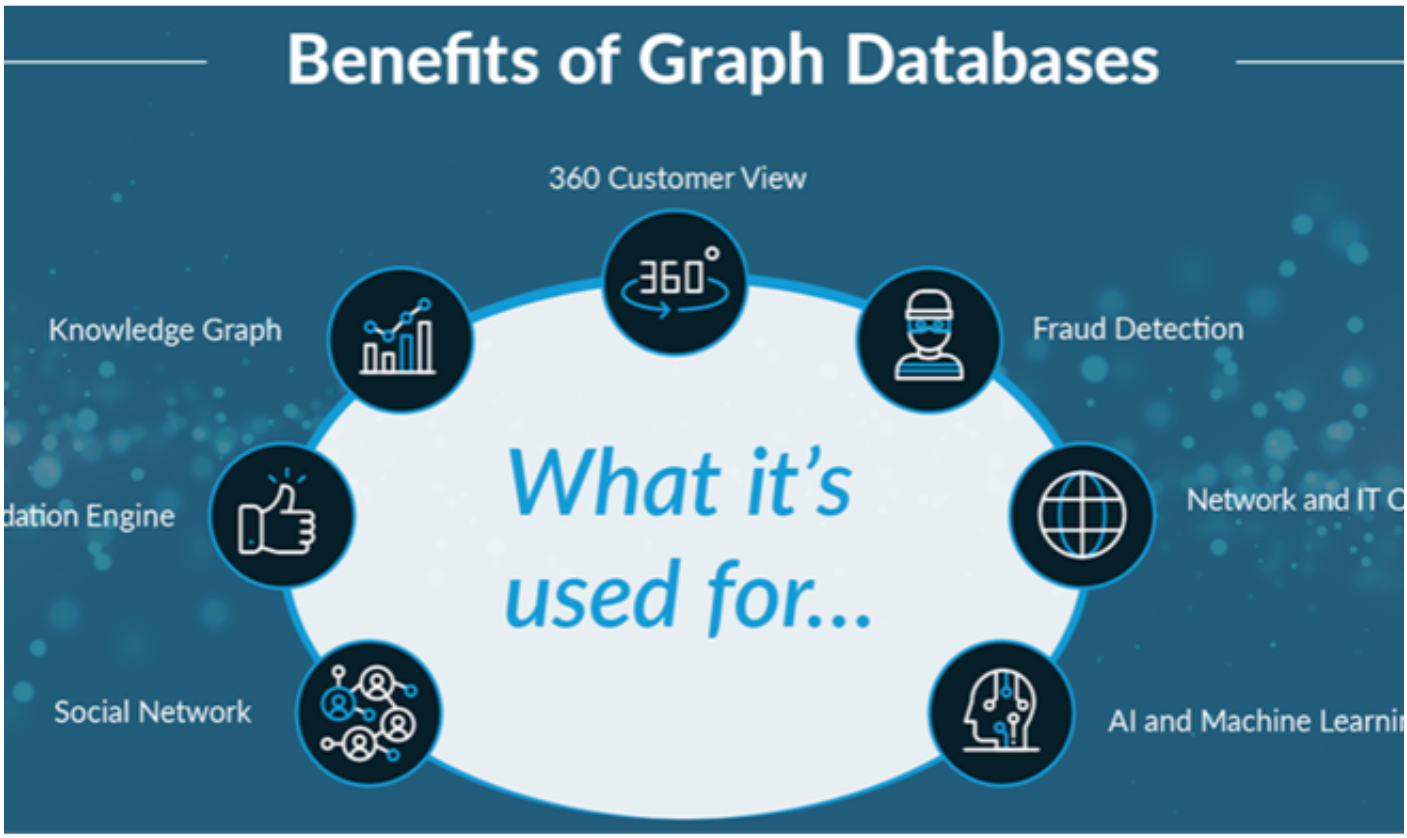
How's it possible that LinkedIn can show all your 1st, 2nd, and 3rd -degree connections, and the mutual contacts with your 2nd level contacts in real-time. The answer is - because LinkedIn organizes its entire contact network of 660+ million users with a graph!

Why are the recommendations on Amazon always so spot-on? Well, they use a graph database - and, by the way, so do many other e-commerce giants.

Instagram, Twitter, Facebook, Amazon, and practically, all applications which must rapidly query information scattered across an exponentially growing and highly dynamic network of data, are already taking advantage of Graph Databases.

What is a Graph Database good for?

A graph database is purpose-built to handle highly connected data, and the increase in the volume and connectedness of today's data presents a tremendous opportunity for sustained competitive advantages utilizing graph databases.



As you can see in the picture above, graph databases are highly beneficial to specific use cases:

- Fraud Detection
- 360 Customer Views
- Recommendation Engines
- Network/Operations Mapping
- AI Knowledge Graphs
- Social Networks
- Supply Chain Mapping

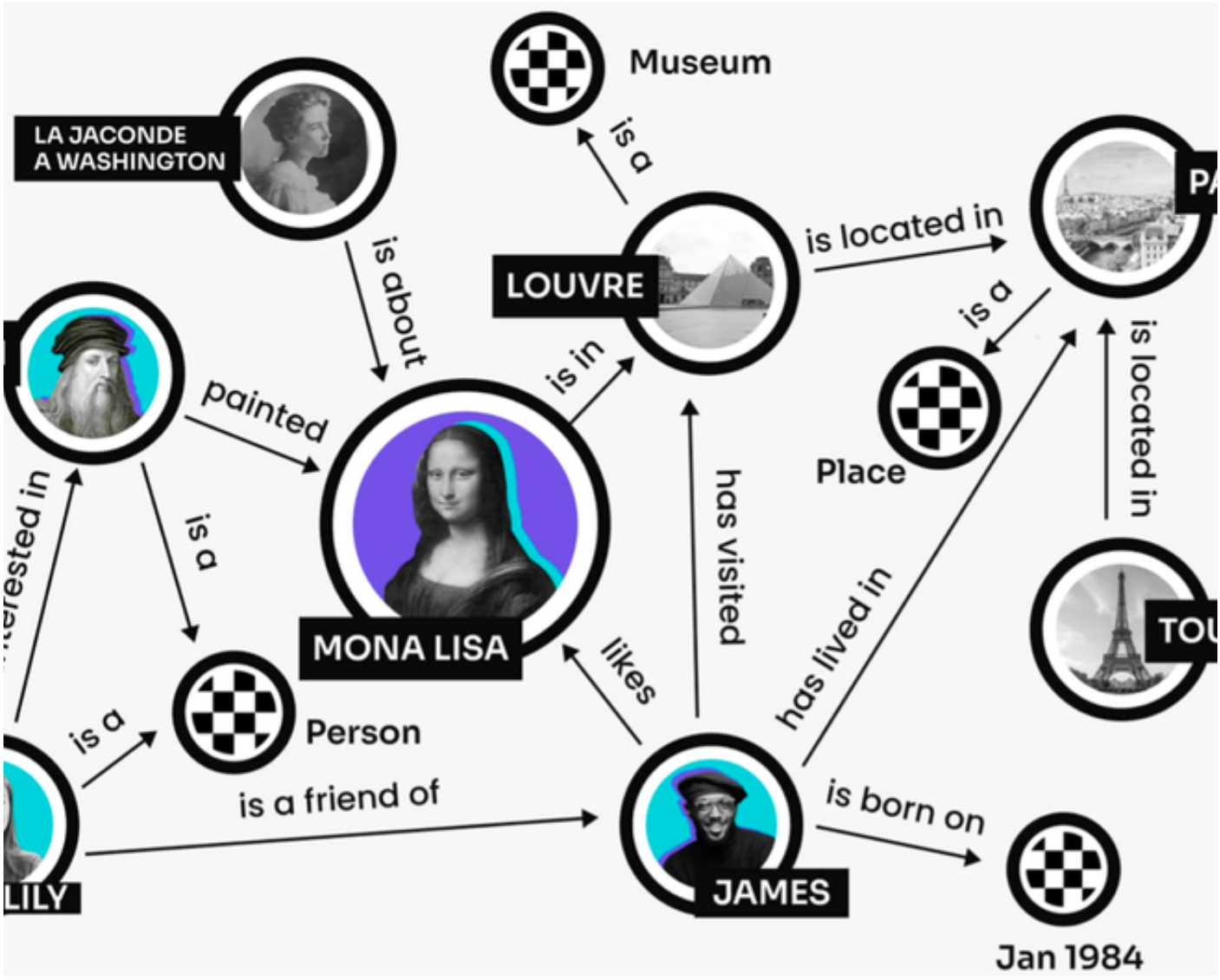
It is uniquely capable in these areas because it does the following very well:

Flexibility: The data captured can be easily changed and extended for additional attributes and objects

Search: You can run fast relationship-based searches such as "Which supplier provided the products owned by this group of customers?"

Indexing: Graph databases are naturally indexed by relationships (the strength of the underlying model), providing faster access compared to relational databases.

Example of a knowledge graph



Knowledge graph with some potential issues

Additional information on Knowledge Graphs: Knowledge Graphs

How Graph Databases work

Unlike other database management systems (DBMS), relationships take first priority in graph databases. In the graph world, connected data is equally (or more) important than individual data points.

Graph databases work by storing the *relationships* along with the data. Because related nodes are physically linked in the database, accessing those relationships is as immediate as accessing the data itself. In other words, instead of calculating the relationship as relational databases must do, graph databases simply read the relationship from storage. Satisfying queries is a simple matter of walking, or "traversing," the graph. It's as if you were on a scavenger hunt and each item tells you exactly where the next item is rather than make you rely on clues and logic to figure it out. This makes graph databases much faster than traditional databases for these types of queries.

A graph database not only stores the relationships between objects in a native way, making queries about relationships fast and easy, but allows you to include different kinds of objects and different kinds of relationships in the graph. Like other NoSQL databases, a graph database is schema-less. That means you don't have to know how the data is structured ahead of time. As you learn new characteristics to add to your database, you don't have to rework the tables. Thus, in terms of performance and flexibility, graph databases lean closer to document databases or key-value stores than they do relational or table-oriented

databases.

What Makes Graph Databases Unique

A lot of databases have similar characteristics, but graph databases have a few things that make them unique. Graph databases have two important properties:

Graph storage

Some graph databases use native graph storage that is specifically designed to store and manage graphs - from bare metal on up. Other graph technologies use relational, columnar, or object-oriented databases as their storage layer. Non-native storage is often slower than a native approach because all the graph connections must be translated into a different memory model for processing.

Graph processing

Native graph processing (a.k.a. *index-free adjacency*) is the most efficient means of processing data in a graph because connected nodes physically point to each other in the database. Non-native graph processing engines use other means to process Create, Read, Update or Delete (CRUD) operations that aren't optimized for handling connected data.

Graph Databases impacting business/market adoption:

Forrester

According to the Q4, 2020 Forrester report, graph data platforms are being widely adopted. Mission critical systems supporting customer-360, knowledge graphs, master data management, social networks. Vendors provide advanced capability with AI/ML on top of graph databases. Multiple workload types can be deployed with graph.

Figure 1: Forrester Wave™: Graph Data Platforms, Q4 2020

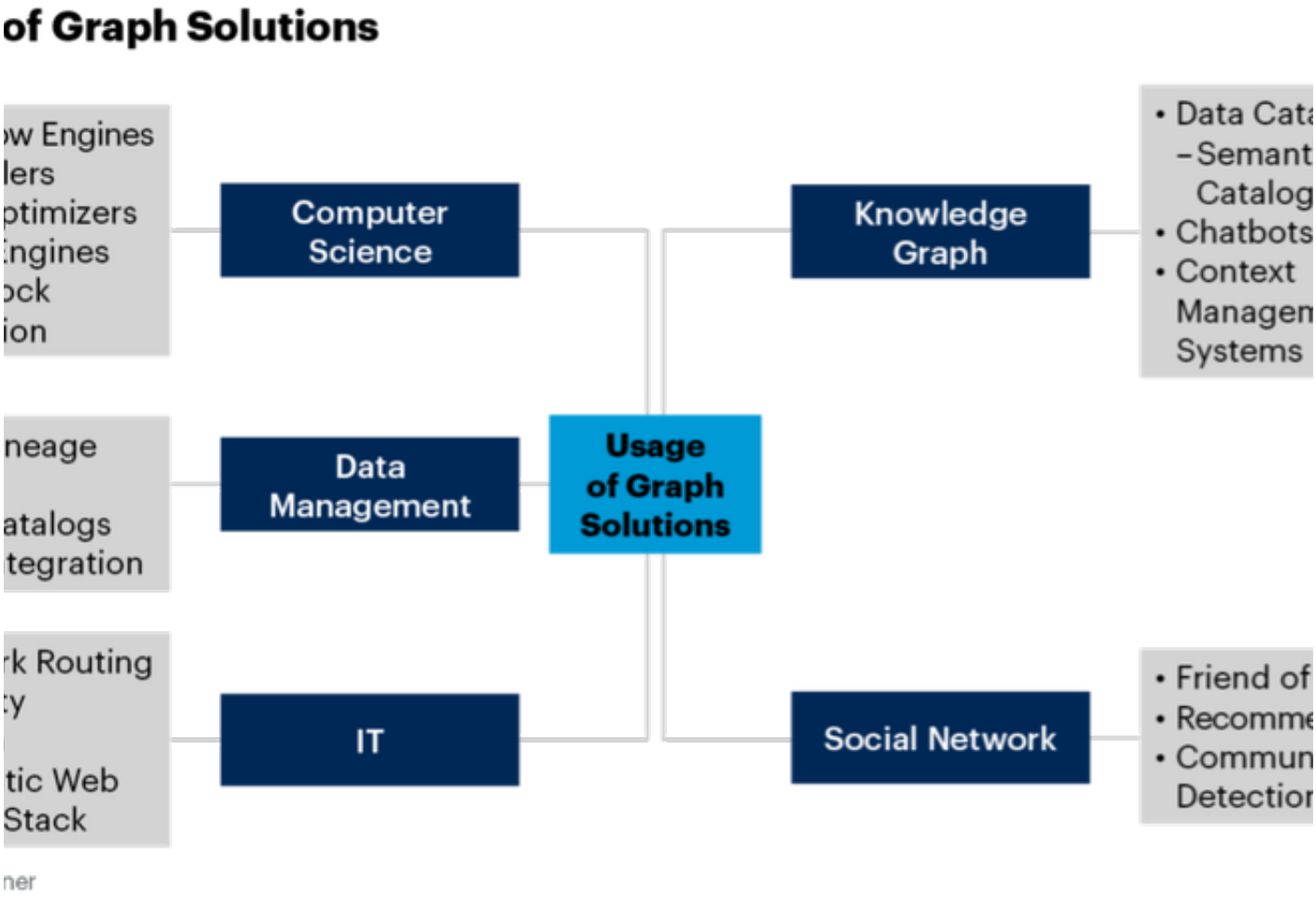
Market leaders: Neo4j, AWS Neptune, Microsoft Cosmos, Oracle, TigerGraph

Gartner

AI and ML using graph databases is becoming a common use case.

- Digital transformation of complex, interrelated data buried across data silos in an organization are best addressed by knowledge graph (KG) or graph-based approaches.
- Analyze knowledge graph (KG) platforms for use cases that demand usage of KGs. Do not attempt to build KG platforms from the ground up or use property graphs for KG solutions.

- Graphs are being used to solve complex, real-world problems —such as dataset tracing and drug development — beyond classical graph and academic-centric algorithms.



Potential Vertex applications

- Graph based analytics to find relationships between tax rules/regulations between jurisdictions and/or customer data.
- Potential data insights/connections to drive additional product offerings,

customer tax rule administration/guidance, drive Vertex tax consulting opportunities, discover and understand relationships between products to make customer recommendations.
- Labeling rules to classify them for improved analysis and assignment. The ability to use AI and ML with graph database data to learn relationships between nodes and edges.
- The ability to link tax rules in audit trails for policies.

Vertex could offer knowledge graphs with nodes of tax rules and a multitude of financial data sources to uncover a comprehensive view of tax implications for customers.

Vertex could also investigate using a graph database and a knowledge graph to research and track competitor's technology and business interests.

Vertex could utilize graph databases and knowledge graphs to make connections between regulations and provide options for financial offerings.

Graph databases are increasingly adopting AI/ML integrations to further the capabilities and value proposition. AI use cases could take advantage of search feature capabilities when connections are being evaluated between nodes. The ability to visualize the knowledge graphs provides understanding on how the ML algorithms explain the AI findings.

Graph databases and knowledge graphs are a data fabric component in an enterprise data management solution.

Vertex projects in progress with Graph databases:

TRM (Tax Research Modernization) will be using a graph database to store and represent tax rules and constraints.

TRMUI-018: Semantic Data And Knowledge Management | Solution Option 1 Knowledge Graph

TRMUI-021: Vocabulary And Ontology Editor

Competition status / Competitive advantage:

Avalara

Research did not discover any known public announcements that discuss utilizing this technology.

Sovos

Research did not discover any known public announcements that discuss utilizing this technology.

Thompson Reuters

Thompson Reuters has multiple offerings that utilize graph databases and knowledge graphs. They have a Risk Management platform named the Data Fusion platform. The DF platform offers multiple Thompson Reuters Knowledge Graphs targeting the Financial Services market. These knowledge graphs are content feeds that show relationships between entities that affect financial services.

Stripe

Research did not discover any known public announcements that discuss utilizing this technology. However, a 3rd party company called Workload offers a service that integrates Stripe data with Google Knowledge Graph Search API based a no-code offering.

Wolters Kluwer

Wolters Kluwer Governance, Risk and Compliance center provides services for “driving business transformation and new product innovation”. Graph databases, AI and ML are at the core of these services. Regulatory Obligation Management provides content enrichment, smart authoring, and obligation graphs. Wolters Kluwer has presented at “The Knowledge Graph Conference”.

Cloud Marketplace (Leaders-see Forrester Wave for complete list)

AWS:

Neo4j: Neo4j is a AWS partner. Neo4j is a true graph database.

Amazon Neptune: Tool in graph database service. Supports popular graph model property graph and W3C’s resource description framework (RDF). Support their query languages, Apache TinkerPop, Gremlin and SPARQL. Query to build queries that efficiently navigate highly connected datasets.

DynamoDB -Nosql db is a key value store but can support graphs.

MongoDB offers graphing with $graphLookup stage.

Azure:

Azure Cosmos DB: Provides graph database service via the Gremlin API. Independently scalable graph engine.

Conclusion:

Graph databases turn data into knowledge, which is visualized through

knowledge graphs.

Graph databases could provide advantageous to Vertex in the ability to connect diverse areas of data sources.

Insights based on Vertex tax rules, customer data as well as 3rd party data could provide interesting and valuable information.

Insights could benefit Vertex internal and external resources as well as Vertex customers.

According to Gartner and Forrester research organizations, graph databases are widely adopted across multiple industries.

Are Graph Databases Ready for Adoption?

Graph Databases are common in the marketplace as well as being utilized by solutions, by Vertex competitors.

How Graph Databases work (technical):

Unlike other database management systems (DBMS), relationships take first priority in graph databases. In the graph world, connected data is equally (or more) important than individual data points.

This connections-first approach to data means relationships and connections are persisted (and not just temporarily calculated) through every part of the data lifecycle: from idea, to design in a logical model, to implementation in a physical model, to operation using a query language and to persistence within a scalable, reliable database system.

Unlike other database systems, this approach means your application doesn't have to infer data connections using things like foreign keys or out-of-band

processing, like MapReduce.

The result: Your data models are simpler and yet more expressive than the ones you'd produce with relational databases or NoSQL (Not only SQL) stores.

Graph databases are based on a branch of mathematics called graph theory. Graph theory is based on networks of points connected by lines. A graph is a set of vertices (nodes, nouns, points) and of edges (lines) that connect the vertices. A graph provides an underlying architecture that is a highly connective data model. Graph databases can answer complex or unknown questions within the data set. Graph databases make the paths between data within the dataset easier to traverse (less complex query plan).

Graph Algorithms: There are different types of algorithms utilized in analyzing connected data within a graph database. Different libraries can be loaded on a graph database platform to take advantage of different algorithms. Different algorithms show how and what data is connected. For example, in neo4j the neo4j graph data science library (GDSL) has the following implementations for these types of algorithms: Path Finding, Centrality, Community Detection, Similarity, Topological link predictions, Node Embeddings, Node Classifications, Link Prediction.

An Ontology is how knowledge is represented in a domain model-how entities relate to each other; types, properties, relationships. Taxonomy is how content is described and structured, categories and subcategories. A knowledge graph is created using algorithms along with an ontology and taxonomy for the set of data being analyzed.

Graph databases have three key technical advantages:

Performance

For intensive data relationship handling, graph databases improve performance by several orders of magnitude. With traditional databases, relationship queries

will come to a grinding halt as the number and depth of relationships increase. In contrast, graph database performance stays constant even as your data grows year over year.

Flexibility

With graph databases, a company can move at the speed of business because the structure and schema of a graph model flexes as applications and industries change. Rather than exhaustively modeling a domain ahead of time, data teams can add to the existing graph structure without endangering current functionality.

Agility

Developing with graph databases aligns perfectly with today's agile, test-driven development practices, allowing your graph database to evolve in step with the rest of the application and any changing business requirements. Modern graph databases are equipped for frictionless development and graceful systems maintenance.

Source articles for this page:

- Graph Databases (Technology) - Bloor Research
- Making Sense of Data with RDF* vs. LPG - OpenCredo
- Graph database use cases (10 examples) - Profium
- The Benefits of Graph Databases [Infographic]
- Why Graph Databases?
- https://Neo4j.com/developer/graph-data-science/graph-algorithms
- What is a graph database? A better way to store connected data
- What is a Graph Database? - Developer Guides
- Forrester: The Forrester Wave: Graph Data Platforms, Q4 2020
- Gartner: Graph Technology Applications and Use Cases : Published 16 March 2021 - ID G00736066
- Gartner: Building Knowledge Graphs: Published 7 July 2022 - ID G00763598

- Gartner: Graph Technology Applications and Use Cases : Published 16 March 2021 - ID G00736066
- Gartner: Graph Technology Applications and Use Cases : Published 16 March 2021 - ID G00736066
- Gartner: Graph Technology Applications and Use Cases : Published 16 March 2021 - ID G00736066
- A New Age of Data — What is A Graph and How Can It Help Me?

Ideation & Emerging Technology

Edit **Contact Us**