Graph Steps Onto the Main Stage of Data and Analytics: A Gartner Trend Insight Report

Published 14 December 2020 - ID G00734700 - 25 min read

By Analysts Mark Beyer

Initiatives: Data and Analytics Strategies; Analytics, BI and Data Science Solutions; Data Management Solutions

Graph technology has played a small but powerful role in advanced analytics for more than two decades. Now that digital business has introduced expanded tools, platforms and techniques, data and analytics leaders can also leverage graph for data science, metadata management and data management.

Overview

Opportunities and Challenges

- Business users revise their analysis in response to newly identified needs and gradually move away from linear or isolated analysis to include evermore inputs and complex calculations, often exceeding the capabilities of the tools and platforms they are accustomed to using.

- Graph technology completes most of the highly complex processes transparently as functions of the tools or platforms, enabling users to have a more interactive experience with both their data and the analytic mathematics.

- The differences between graph and relational concepts and technologies demand new skills from both implementers and practitioners.

- Case studies and usage guidelines can help to identify use cases for graph as distinguished from relational-based processes.

What You Need to Know

- Executives and more traditional business analysts will realize pervasive and significant (but indirect) benefits from using graph technology, as their direct reports and application teams are able to explore available datasets at a far more granular level of detail.

- The most advanced users of data for analytics (data scientists, data engineers, data analysts) can place advanced solutions into production for citizen users with a correct use of graph technology.
Graph exploration of data and the resulting analysis require a new set of skills focused on significant understanding of data abstraction, advanced statistical models and analytical models’ use of semantics.

The peer community who use and benefit from graph technology can draw on case studies to help navigate the rapidly expanding array of practices, tools and platforms.

Strategic Planning Assumption
By 2023, graph technologies will facilitate rapid contextualization for decision making in 30% of organizations worldwide.

Insight From the Experts
To Graph or Not to Graph? That Is Not the Question — You Will Graph

Mark Beyer, Distinguished VP Analyst

Most analysis begins with a hypothesis — sometimes emerging from some limited data discovery perspective, other times from a business agenda driving a demand for targeted analysis.

Traditionally, analysts design models based upon a theory or a specific business agenda and then fill the model up with data. A slightly better analyst will develop trending and insight analysis that identifies when behavior is “as expected” and even allows inspection of exceptions to find outliers and errors. Yet even for all the power of an outlier, the very term indicates that something is beyond normal — even if true. Graph adds a new dimension.

Graph comes almost from the opposite direction.

Graph assumes that any noun defined in data (person, place, object, thing) can be an entity that is always a participant in multiple relationships all at once. These relationships change through time, each leaves a sort of “imprint” on the other in terms of relationships that take place in clusters and groups. Each relationship can become stronger and reinforced by more relationships around each entity — and this can happen in parallel or in sequence. It is even possible to see how relationships give information about seemingly unrelated events when the graph is viewed through time and even three-dimensional space.

Graph assumes there is a proximity between entities and even the relationships that bind them together. That proximity can be measured to show how valid each relationship is in both intensity (think of how many times the same relationship is reinforced or how many other relationships simply cannot pull the first two entities apart) and distance (often referred to as “hops” in the graph). In other words, graph relationships in data reflect the real world in that all things can be somehow related and connected — even logical concepts like currency with trade.
With a graph solution, users are no longer limited to only the dominant data interpretations — all possible data interpretations can be explored.

Graph capabilities have existed for compute environments and data management for decades. Increased experience and maturity, innovative platforms, and composite artificial intelligence (AI) functions have brought graph to the fore.

For the technologist, graph offers the opportunity to shrink development cycles for data preparation, insight discovery and productionization of achieving a data product as output. Graph also improves data quality and data governance. Data quality continues its redefinition from validation, verification and rule-based analysis, moving even further into data linking and enhancing the current trend for reinforced trust as a result of graph-based metadata analytics. The best part is that new solution designs leverage existing and well-entrenched technology (preserving large portions of those investments), but simultaneously reduce future costs for both existing tech stacks and introduce a lower cost “future” tech stack.

Graph does not replace relational data logic; in fact, it builds upon it. All data starts out with some type of integrity or relative reference (“four” is not informative, but “four” plus “apple” is data). Modern relational database management systems were founded on the principle of a statistical motif or primitive actually named “join.” Join specifically states that the left side of logical argument is exactly equal to the right side, or it is not. It is a binary concept, and when that is the only argument, then relational constructs can be optimized to scale on computing technology, which is also binary.

There are more perfectly binary primitives besides join. MapReduce itself is two stacked primitives and the basis for Hadoop compute clusters, for the exploration of data when join was false. Graph is the next highest level primitive above Reduce. Graph is the next evolution of data management and analytics — it just needed enough compute power and a better data construct to make it worth our efforts. Graph and cloud go together just like compute clusters and MapReduce went together.

So graph is a powerful compute engine and powerful data management approach, but it also enhances almost everything else we do with data. In this research, you will begin to see how graph analytics, graph databases and graph theory enhance data and metadata usage. When graph technology is combined with machine learning (ML) and solutions like data profiling, the data fabric begins to emerge. A continuous, virtuous cycle emerges between data capture, data users and data analysis that becomes a major driver to the intelligent composable business.

Kind regards,

Mark A. Beyer

Executive Overview
Graph is a powerful engine to support AI and data science practices (see Evidence: Figure 2). Graph enables significantly improved metadata capabilities, enhanced decisions using combinations of prescriptive and predictive analytics, and faster discovery of both internal and external data that is valid for existing use cases. The result is faster delivery of an advanced infrastructure that more ably supports citizen user roles — even including enhanced visualization and display capabilities.

This report will help senior managers and data and analytics leaders identify appropriate opportunities to utilize graph technology in their data management, governance, analytics and data science needs. We break the discussion down into three key areas (see also Figure 1).

- **Enhanced experience**: There is no point in adopting any new technology trend without benefit to operations, users or financials. Graph enhances visuals for analyzing data, provides additional insights for decisions, and enables more effective analysis by citizen users.

- **Accelerated utilization**: Improved data use and accelerated analytics come directly from graph’s capability to reduce time to delivery for analysis, provide a vital input to guided data and insight discovery, and enhance metadata capabilities to improve knowledge management.

- **Technology base**: Graph-enabled data management, coding languages, prebuilt analytic models and more are built into many data and analytics tools. Additionally, many components of the underlying infrastructure can be leveraged to accelerate existing solutions.

**Figure 1: Nature and Benefits of Deploying Graph**

**Nature and Benefits of Deploying Graph**

- **Enhanced Experience**
  - Visuals
  - Decisions
- **Technology Base**
  - Delivery
  - Discovery
- **Accelerated Utilization**
  - Metadata/Knowledge

**Definition**
Graph is the utilization of a blend of advanced statistical modeling from graph theory, graph-enabled data structures that focus on relationships in preference to entities or events, and a new generation of user tools that leverage processing and visualization that specifically support the mathematics and data structures involved.

In order to benefit most from graph and target your use of its capabilities, it is imperative that you understand four different aspects, each of which forms the basis for the related research that follows:

- **The business case for deeper analytics** — Here, we help you identify appropriate categories of when to apply this type of solution, as opposed to leveraging existing (and already deployed) technology stacks.

- **Tools are accelerating graph adoption** — It is generally understood that graph technology enables a broad diversity of analysis and requires specific data processes and data management. But new tools and features or functions in existing tools and platforms now manage some of the most advanced deployment and production needs to utilize graph.

- **When to use graph technology** — It would seem that once the use cases are understood, using graph technology becomes a simple step. But that is not the case. Sometimes, graph merely validates that existing solutions already work and should not be replaced. We discuss the balance of keeping the existing (proven and efficient) solutions in place rather than replacing them. We also provide guidance on how to determine when the technology is ready for your use case and when your use case is ready for graph.

- **Real-world applications of graph** — We don’t stop with guidance and technology discussions. Eventually, it’s time to actually deploy. Learn from these case studies and organizations that changed their operations, their decisions and even their implementation scenarios in the real world.

**Research Highlights**

This special report will address the technology involved, portray use cases and identify emerging best practices in graph technology. It will also describe how graph databases and graph analytics can lower costs while scaling the output from smaller data scientist, analyst and data engineer teams into production solutions. We start with the advice to help you identify appropriate use cases. We suggest you start right at the top, then quickly move among the more specific topics.

**The Business Case for Deeper Analytics**

Graph technology enables “deeper” analytics, but what does that mean? When data is first captured, it follows a design and intent that is part of the original business process. Analyzing how a business
process is designed requires metrics, performance thresholds and trend analysis. But what happens when two business processes collide? Most traditional business intelligence was designed for such collisions and then deployed to answer theoretical questions by analysts. Deeper analytics are those that ignore predisposition to any conclusion. If the data goes together, if the business processes overlap, then they are related somehow. Graph use cases don’t assume any of the connections and instead let any possibility rise to the top. Sometimes, those data errors were not errors, but were actually modifications to the business operations. Deeper analytics try to explore any and all connections, and make sense of them even when they “shouldn’t be there.”

The following Gartner publications will help you build an appropriate use case for graph technology in your organization.

Related Research

- **What Is ‘Graph’? — An Elementary Version for the Uninitiated** answers the question we hear most often from organizations trying to understand it. Data and analytics leaders are faced with the sudden interest in graph theory, graph analytics and graph data management. Analysts, data scientists and data management professionals are attempting to leverage graph technology in various and often confusing conversations. The result is often a confused justification with even more obscure benefits. This research parses the conversation into three main topics to allow data and analytics leaders to make sense of what their direct reports are telling them.

- **Understanding When Graph Technologies Are Best for Your Business Use Case** is important to avoid graph being the “new, best answer to everything.” Organizations are seeking to use their underleveraged data to uncover insights and patterns otherwise hidden in complex connected data. What is the best way to determine when graph analytics are the right capability to address specific use cases? The most important aspect of any business process is understanding the relationships of actions or assets as they relate to each other inside that process. Graphs can be used to analyze all sorts of relationships across all kinds of systems, even beyond process or beyond the confines of individual operational models. Here, we sort out the most common business questions that can benefit from applying graph capabilities.

- **Connecting the Dots: Why Graph Analytics Are Key to Understanding Human and Machine Misbehavior** is a 25- to 30-minute set of slides to educate yourself and your teams on how graph analytics can enhance your use of data. We start by defining graph analytics, then quickly move on to use cases and how to get some benefits. All of that is great, but the best part is that we provide guidance on how to sort out graph use cases with practical tips for recognizing when to use the graph approach.

- **How to Build Knowledge Graphs That Enable AI-Driven Enterprise Applications** puts together graph and metadata concepts in a meaningful way. Increasingly, knowledge graphs are powering AI applications. However, for scalable implementations that can solve enterprise data integration challenges, data and analytics leaders must take an agile approach to knowledge graph development. Consider that, by 2024, companies that use graphs and semantic approaches for natural language
technology projects will have 75% less AI technical debt than those that do not. In this note, we lay out the benefits and approach to putting graph, knowledge graphs and AI together.

- **How Augmented Data Management Capabilities Are Impacting MDM and Data Governance** goes over the intense relationship between metadata, master data management (MDM) and data governance that is still largely deployed through manual efforts. But it does not have to stay that way. Augmented data management capabilities offer data and analytics leaders new opportunities to extend the business value of MDM. It is critical that existing MDM strategies are revamped to reflect the impact of bringing application data into existing MDM and data governance processes. The struggle to accelerate the generation of insights needed for effective decision making in the digital realm have provided impetus to apply graph, AI/ML and similar augmented data management capabilities.

**Tools Are Accelerating Graph Adoption**

Recent advances in product features and functions, along with new entries into the various data management and analytics markets, have automated or platform enabled many of the most complex functions required. Embedded features also mean you can embed the technology into your analytics development and leverage these new capabilities. At the same time, data scientists (both experienced and recently graduated from university) who are familiar with advanced statistical models will find that their favorite programming languages are supported and that the functions and logic are strikingly familiar. When you combine familiar operations and functions used by experts with technology that faithfully reproduces their work, the barriers to putting appropriate graph solutions into production begin to fall away.

**Related Research**

- **What Is Artificial Intelligence? Seeing Through the Hype and Focusing on Business Value** brings everything “down to earth” to help with anticipating both benefits and delivery. Superheated rhetoric surrounding the potential benefits of AI is inflating expectations among business leaders. Data and analytics leaders must enable conversations focused on real business problems and use cases. Gartner’s 2019 CIO Survey points to the fact that, although 92% of respondents indicate that they either have AI on their radar, or have initiated projects, only 19% have projects currently deployed. This research offers approaches that data and analytics leaders can use to manage business leaders’ expectations of AI, set realistic goals and increase the success rate of analytics projects.

- **Cool Vendors in Graph Technologies** are just that — new solutions that have incorporated graph either as the primary engine or an embedded capability. The rising role of context and situational awareness for delivering insights, and AI’s growing role in evaluating implied and explicit data connections, have pulled graph technologies into the spotlight. AI and context are the demand. Increasingly, multifaceted, contextual and heterogeneous data drives digital business. Graph technologies provide the detection and analysis of complex patterns of relationships between data elements, including relationships that are recursive, circular or of unknown path length. Graph enables the flexible combination of multiple internal and external data sources to explore context and to create a richer situation awareness.
Cool Vendors in Data Management brings forth how even the most traditional data management concepts can be and have been taking advantage of graph. Data and analytics leaders responsible for data management are under pressure to deliver projects faster and at a lower cost. This has led to a rise in demand for augmented data management in various offerings like active metadata, AI/ML algorithms and data fabric designs utilizing semantic knowledge graphs. From graph databases to metadata analytics to graph analytics platforms, these vendors bring graph capabilities to the enterprise.

Cool Vendors in Analytics and Data Science is a straightforward explanation of how new vendors bring together the most advanced analytics solutions using graph. Augmented analytics is the future of analytics. Yet despite organizations adopting new augmented capabilities, they often struggle to provide a collaborative environment in which sufficient numbers of people can communicate and exchange the value of analytics. Organizations using search, catalogs or other data asset identification approaches could benefit from adding a graph layer to see how their data is connected.

Magic Quadrant for Metadata Management Solutions explains how “active metadata” is dependent on continuous graph analytics over your data utilization and usage in the organization. It is expected that having graph analytics on every conceivable type of metadata will provide the necessary information for introducing ML capabilities into various data management tools, including data integration, data quality, data preparation and even database management systems (DBMSs). Existing or passive metadata can be “activated” by analyzing the overall data experience that users and an organization have with data, and evaluating how it differs from expectations inherent in system designs.

Critical Capabilities for Cloud Database Management Systems for Analytical Use Cases is a listing and rating of leading vendor solutions in the market, including how they utilize graph. Analytics continues to be a top priority for many organizations. This, combined with the trend toward adoption of the cloud, has meant that the ability of cloud DBMS solutions to address analytical use cases has become very important. An increasing demand for supporting data science and deep learning has become a core capability for all databases. Databases must now support the exploration of how new data values form variants and relationships using search, graph, predictive modeling and other capabilities to uncover information models.

Working With Semistructured and Unstructured Datasets explains that one of the first things you have to do with unstructured data is parse it into structures — this results in massive amounts of complementary and conflicting metadata outputs that can feed a knowledge graph. Unstructured data volume within an organization is increasing quickly, and organizations are rapidly falling behind in leveraging it primarily due to a lack of awareness of the benefits of using it. We lay out multiple techniques for working with less structured data, many of which are based on graph analytics, graph stores and knowledge graph development and support. After exploring applied use of the tools and graph approaches, a quick reference list of platforms and tools summarizes examples from the market.
When to Use Graph Technology

A business case is a driver and defines usage, but utilization is another matter. While case studies and business benefits (or demands) may ask the question, utilization requires balancing your existing infrastructure and what it can accomplish with the new capabilities of graph technology. We start off with a midlevel infographic that provides a high-level overview of your data environment, but then goes into some needed details about how your modern data management environment will work (whether automated or manually managed). We then present an end-user utilization model for data scientists, and an applied approach to using graph for event and time series analysis. We return to the discussion of how graph enables the emerging data fabric, which begins to convert it from a theoretical model to a pragmatic and practical approach for modernizing your data management infrastructure.

Related Research

- **Infographic: An Intelligent Composable Business Demands a Data Fabric** is a view of your data environment from the messy metadata side outward to create a dynamic data mesh or data fabric. This concise infographic lays out the approach and benefits for you. An adaptive data fabric utilizing active metadata reduces manual data analysis, integration design and data sharing tasks up to 70% while supporting adaptive data governance. By shifting your data strategy from one driven by human awareness of gap analysis and implementing the solution to one that is machine guided, you recover innovation time. Your human experts can switch from a scenario in which they do 80% data work (with only 20% human innovation) to one that is geared toward 24%, 35% or even 40% human innovation. That doubles your human output.

- **Emerging Technologies: Data Fabric Is the Future of Data Management.** Semantic modeling via graph technology and metadata-driven augmented data integration are two emerging, yet critical, technology components of the data fabric design, because they support the connections among the other components and enable the initial design to evolve over time. Data fabric is an important emerging trend that requires a combination of multiple established and emerging data and analytics technologies. Product leaders must ultimately offer complete data fabric solutions via a combination of their own products and strategic partnerships.

- **2021 Planning Guide for Data Analytics and Artificial Intelligence** lays out the phases of planning your advanced analytics environment. Data and analytics technical professionals must embrace emerging trends, harness technologies and foster innovation to create enterprise-grade analytic architectures. As organizations look beyond building data lakes to collectively store data, the emphasis will be more on finding relationships across a diverse dataset by using graph techniques and metadata management, which will form the foundation of modern data and analytics. The combination of knowledge graphs, data fabrics, natural language processing (NLP), explainable AI and analytics on all types of content (X analytics) will help provide richer context for ML and AI.

- **Pursue Citizen Data Science to Expand Analytics Use Cases** explains one of the key benefits of enabling citizen users with discovery tools — their use of graph functionality. Citizen data science scales machine learning and data science efforts across the organization. Use cases for graph analytics reflect capabilities that model relationships between entities such as organizations, people
and transactions. Another feature of graph analytics is the ability to analyze relations, influence or proximity between multistructured data. Citizen data scientists will often have access to a visual workbench to explore data networks.

- **Time Series Database Architectures and Use Cases** goes over how the relative nature of time in data is highly complex and variable, and is often resolved internally using graph capabilities. Time series databases solve two fundamental problems centered on read/write asymmetry: quickly ingesting and persisting large volumes of data, and reading data at an exponentially higher rate. The real-time nature of time series data makes it necessary to expose new data in queries as fast as possible, although not every point matters in the series. In the case of time series data, it is often an optimal approach to perform graph processing on it.

- **Demystifying the Data Fabric** is a detailed description of the different types of technology and functionality needed to start your data fabric journey. The data fabric takes data from a source to a destination in the most optimal manner. It constantly monitors the data pipelines to suggest, and eventually take, alternative routes if they are faster or less expensive — just like an autonomous car. The purpose of augmented data integration is to make the process of integrating data sources and delivering data easier and more optimal in terms of performance, effort required, accuracy and cost. One approach that promises to deliver these capabilities is the data fabric design concept.

- **How to Activate Metadata to Enable a Composable Data Fabric** brings the metadata aspects home to your data fabric, and how metadata itself is a graph. Existing data management solutions contain an abundance of metadata that most organizations find difficult to identify, much less analyze. Data and analytics leaders can reuse existing metadata to build the data fabric for the enterprise and simultaneously get real value from enterprise metadata. Graph analysis can be used to discover and evaluate significant data objects or clusters and overlay them with use-case analysis. Once use cases and data objects become clearly defined, then the entities participating in the fabric need to be resolved (both actors and assets).

**Real-World Applications of Graph**

Theory is great. Implementation is better. Practices become models for delivery. Real-world cases of using graph to enhance everything from supply chain, through food supply and even into manufacturing and disease management already exist in the world today. The following research and case studies make it clear that graph technology is no longer only for the “laboratory,” but can go into production.

**Related Research**

- **Innovation Insight for Digital Supply Chain Twin** takes you through the digital supply chain twin. The digital supply chain twin is a key capability in any supply chain decision-making system. From a data model perspective, a graph model is a good way to do this because it can manage these relationships. The resulting digital supply chain twin is transformational. The vast majority of the technologies that are working in and around the area of digital supply chain twins are using a graph model to represent the physical supply chain.
Case Study: Answering Critical Business Questions With Graph Analytics (Jaguar Land Rover) demonstrates how graph analytics can give the business a connected view of supply and demand, enabling efficient answers to critical business questions. Data and analytics leaders struggle to advance a shared understanding of data across business verticals and functions. At JLR, the use of graph techniques creates a connected view of business, from supply to manufacturing and demand.

Case Study: Data and Analytics Monetization With Knowledge Graphs and AI (Turku City Data) is a public-sector-applied graph use case. Turku City Data has a best-practice solution: A knowledge graph framework for organizing data, exploring business problems, and building reusable data products and analytic solutions. The POLE framework (people, objects, locations and events) has emerged in the public sector as public safety, social services and other government agencies modeled their data using this approach. The interrelationships between each of the four constructs of POLE are best captured and exploited using graph theory and methods including knowledge graphs, graph databases and graph algorithms.

COVID-19 Demands Urgent Use of Graph Data Management and Analytics explains not only how graph can be applied to healthcare emergencies, but any complex social crisis. Data and analytics leaders in all areas of healthcare and supporting sectors must improve their response to infectious disease management crises by using graph approaches. Graph technology can help them meet the challenges posed by complex relationships between manufacturers, suppliers, distributors and caregivers. Graph technologies introduce the capability to overlay multiple components of supply, manufacturing, equipment utilization and staff resource management on top of each other to aid a shift from a service model to a targeted response model.

3 Best Practices to Reset Healthcare Organizations’ Data and Analytics Strategy illuminates three best-practice approaches that CIOs must implement to thrive in the digital healthcare ecosystem and economy. Utilizing graph analytics, implementing metadata and knowledge graphs for fast data identification and use case analytics, and even deploying graph databases all add to your data management and analytics infrastructure in the complex world of healthcare data management and analytics. This research connects technology to outcomes in a best practices overview.

Gartner Associates Supporting This Trend

- Mark Beyer
- Jim Hare
- Pieter den Hamer
Directed graph

Pairs of data exist, and one or many defined relationships exist, but they have a directionality. A table is made of wood, a chair is made of wood. A chair is not a table. It is best to think of a directed graph as a triple of subject<-predicate->object. An object can be part of many subjects for the same or many reasons (predicate), but all have direction. In the example, to find all things made of wood, the subject is material = wood, the predicate is manufactured, the object in one pair is table, another pair is chair. At the same time, it could be that material = wood, and chair and table are objects, but a new predicate = biodegradable exists.

Undirected graph (simple, ordered and unordered pair)

Data is arranged in pairs, showing that they have some relationship. If ordered, then the first value is driving the relationship to the second value. If unordered, then the relationship can be in either direction. Each pair stands alone in a single relationship.

Undirected multigraph

Similar to “simple.” Permits multiple different joins or relationships between any two members of a pair.

Evidence

- Gartner inquiries comparing interest in “graph” in October 2020 are 250% of the same month in 2018. The greatest increase was approximately 150% in the period March through September 2020 compared with the same period in 2019.

- Google Trends indicates that “graph” is appearing at almost three times the rate of “analytics.”

- Graph is most commonly associated with AI, DBMSs, data integration and metadata.
Figure 2: Artificial Intelligence Adoption by Type of Technique

Artificial Intelligence Adoption by Type of Technique

0% 50% 100%

Machine Learning (n = 607)
Rule-Based Systems (n = 607)
Optimization Techniques (n = 606)
Natural Language Processing (n = 607)
Graph Techniques (n = 605)

Q. What is the adoption status of the following AI solutions in your organization?
Source: 2020 Gartner AI in Organizations Survey
Note: Bases vary, based on all respondents, “excluding “don’t know.”

Recommended by the Author

Use Gartner’s Reference Model to Deliver Intelligent Composable Business Applications
Smart Data Sharing Requires Mapping Use Cases to Architectures and Vendor Solutions
Magic Quadrant for Metadata Management Solutions
2021 Planning Guide for Application Platforms, Architecture and Integration
The 3 Pillars of Citizen-Driven Data Management
Top 10 Trends in Data and Analytics, 2020
Magic Quadrant for Data Integration Tools
Hype Cycle for Data Management, 2020
2021 Planning Guide for Data Management
Data Fabrics Add Augmented Intelligence to Modernize Your Data Integration