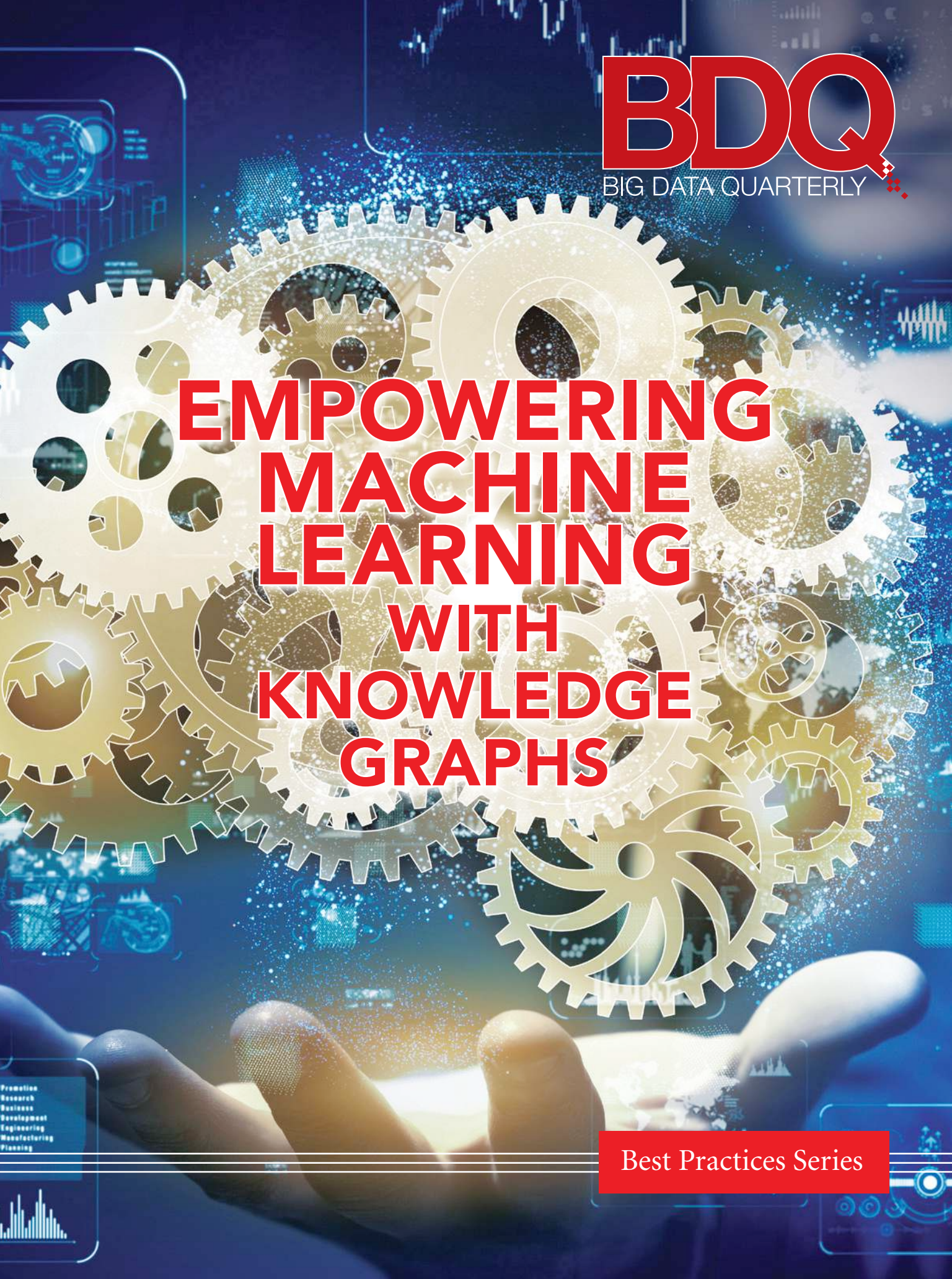
BDQ
BIG DATA QUARTERLY
EMPOWERING MACHINE LEARNING WITH KNOWLEDGE GRAPHS
Best Practices Series

BDQ
BIG DATA QUARTERLY

IMPROVING DATA PROCESSES WITH KNOWLEDGE GRAPHS

Rapid data collection is creating a tsunami of information inside organizations, leaving data managers searching for the right tools to uncover insights. Knowledge graphs have emerged as a solution that can connect relevant data for specific business purposes.

Knowledge graphs link together data of any variety, structure, or format in business terms via uniform data models. Organizations can then join and traverse all of their data, semantically tagged with unique machine-readable identifiers, making the platform ideal for intelligent systems, machine learning analytics, interoperability, and an array of other benefits influential for AI applications.

The technology is gaining the attention of research firms and consultancies. In 2018 and 2019, knowledge graphs appeared on Gartner's Hype Cycle for Emerging Technologies, acknowledged for their hearty connections to pertinent data. According to Gartner, "These ecosystems developed as digitalization morphed traditional value chains, enabling more seamless, dynamic connections to a variety of agents and entities across geographies and industries. In the future these will include decentralized autonomous organizations (DAOs), which operate independently of humans and rely on smart contracts."

Knowledge graphs are already seen in technologies people interact with routinely. In 2012, Google introduced its own platform based on the technology. The knowledge graph is used by Google Search to help users discover information more quickly and easily. The knowledge graph has millions of entries that describe real-world entities such as people, places, and things. Knowledge graphs are also the backbone of LinkedIn, Facebook, Wikipedia, and several other platforms that involve deep search.

Meanwhile, AI continues to advance beyond Silicon Valley's wildest dreams. However, according to PwC's third annual "AI Predictions" report, only 4% of executives surveyed plan to deploy AI enterprise-wide in 2020. A year ago, nearly 20% hoped to do the same. In addition, 90% of executives surveyed also believe that AI offers more opportunities than risks, and nearly half are expecting AI to disrupt either their geographical markets, the sectors in which they operate, or both.

AI is proving to be more of a challenge than previously thought. However, knowledge graphs can be the answer, enabling executives to facilitate AI initiatives.

Knowledge graphs can hasten preparation for advanced data science models; they imbue data with machine-readable, unique identifiers for integrating data and facilitating intelligent system interoperability. In turn, machine learning capabilities can also increase—as well as improve—the knowledge contained within these graphs.

Best Practices Series

KNOWLEDGE GRAPHS, AI, AND MACHINE LEARNING

Knowledge graphs can become the foundation for AI analytics. They expedite lengthy data preparation measures for data scientists engaged in advanced, predictive analytics.

By enriching data with machine-readable global identifiers, they're suitable for interoperability and feeding systems of machine intelligence. They offer a foundation of knowledge that can inform machine learning inputs and are, in turn, improved by the outputs of machine learning applied to real-world business problems—once they're input into the graph.

According to Gartner, the application of graph processing and graph DBMSs will grow at 100% annually through 2022 to continuously accelerate data preparation and enable more complex and adaptive data science.

AI and graph technologies can improve real-time decision support, search, chatbots, NLP, automated customer services, recommendation systems, product management, digital marketing, rules engines, goal setting, workflows, integration, automated schema mapping, data discovery, anomaly detection, security threat analytics, master data management, and next-best-action prediction.

CREATING A KNOWLEDGE GRAPH

With a knowledge graph:

- The model can be as large, wide, and deep as you want.
- Semantic standards can be used to describe the structure of the information in the graph to support reasoning and inference.
- Both high-quality complete data and sparse and incomplete data can be captured and made usable.

To develop a knowledge graph, the most frequently used approaches include curating it similar to Cyc and editing it via the crowd as with Wikidata. Data can then be extracted from large-scale, semi-structured knowledgebases such as Wikipedia, similar to the approach of DBpedia and YAGO, or they can be created by information extraction methods for unstructured or semi-structured information. This can lead to knowledge graphs such as Knowledge Vault.

When formal semantics are used to express and interpret the data of a knowledge graph, there are a number of representation and modeling instruments.

- **Classes.** Most often entity description contains a classification of the entity with respect to a class hierarchy. For instance, when dealing with general business information, there could be classes such as person, organization, and location. Persons and organizations can have a common superclass agent. Location usually has numerous sub-classes, e.g., country, populated place, city. The notion of class is borrowed from object-oriented design, where each entity should belong to exactly one class.
- **Relationship types.** The relationships between entities are usually tagged with types, which provide information about the nature of the relationship, e.g., friend, relative, competitor. Relation types can also have formal definitions, e.g., that parent-of is the inverse relation of child-of, they both are special cases of relative-of, which is a symmetric relationship, or define that sub-region and subsidiary are transitive relationships.
- **Categories.** An entity can be associated with categories which describe some aspect of its semantics; often, the categories are described and ordered into taxonomy.
- **Free text descriptions.** It is possible to add "human-friendly text" to further clarify design intentions for the entity and improve search.
- **Ontologies**. They serve as a formal definition between the developers of the knowledge graph and its users. A user could be another human being or a software application that wants to use the data in a reliable and precise way. It ensures a shared understanding of the data and its meanings.

Though knowledge graphs are finally getting their time to shine, the technique has been in the making for years. The manufacturing of the knowledge graph depends on three basics: people that envision it, data diversity, and a good product to build it.

There are many examples of applications that people interact with every day that have knowledge graphs working behind the scenes, unknown to them. The most successful companies in the world are implementing new techniques and using the best tools available to build a data fabric that helps them identify data and make it available where it is needed. Knowledge graphs can be a foundational technology for this success.

—Stephanie Simone

Knowledge Graph Solutions for the Enterprise

TOP MARKET ANALYSTS such as Gartner and Forrester have identified Knowledge Graphs as a key new technology in the enterprise for delivering insights that drive revenue and growth.

> **"Graph databases and Knowledge Graphs are now viewed as a must-have by enterprises serious about leveraging AI and predictive analytics within their organization," said Dr. Jans Aasman, CEO of Franz Inc. "We are working with organizations across a broad range of industries to deploy large-scale, high-performance Knowledge Graphs that serve as the foundation for AI-driven applications for personalized medicine, predictive call centers, digital twins for IoT, predictive supply chain management and domain-specific Q&A applications—just to name a few."**

Enterprises trying to modernize have a great need for data science and comprehensive overviews of individual customers and/or products. Despite decades-long efforts in IT integration, most enterprises still lack a single source of truth on these entities of interest. Regardless of extensive hiring of expensive data scientists, it is common knowledge that these scientists waste 90% of their time on preparation of data. Enterprises find that current solutions (ETL, Data Lakes, Data Warehouses, Property Graphs) exacerbate the problem and don't really deliver the needed insights. Their challenge is that every new analytic requires a new and expensive effort by the IT department to gather semi-usable data for the data scientists, wasting millions annually. Franz's Knowledge Graph Solution uniquely solves these Sisyphean integration efforts by unifying data silos into an Entity-Event Knowledge Graph with unlimited scalability through AllegroGraph's unique FedShard[tm] horizontal scalability feature.

ARCHITECTING KNOWLEDGE GRAPHS

To support this burgeoning Knowledge Graph solutions market, a successful analytic system will have to fuse and integrate data, not just in representation, but in context (ontologies, metadata, domain knowledge, terminology systems), and time (temporal relationships between components of data). The rich functional and contextual integration of multi-modal, integrated large-scale analytics, predictive modeling, and artificial intelligence is what distinguishes AllegroGraph as a modern, scalable, enterprise Knowledge Graph platform. AllegroGraph's Knowledge Graph technology approach encapsulates a novel **Entity-Event Model** natively integrated with domain ontologies and metadata, and dynamic ways of setting the analytics focus on all entities in the system (patient, person, devices, transactions, events, operations, etc.) as prime objects that can be the focus of an analytic (AI, ML, DL) process.

AllegroGraph is the only Knowledge Graph platform that natively supports this new Entity-Event Data Modeling approach. Entity-Event Knowledge Graphs are about moving away from statistical averages and broad-based patterns in order to connect the many dots, from different contexts and throughout time, to support and recommend industry-specific solutions that can take into account all the subtle differences and nuisances of entities and their relevant interactions to deliver insights and drive growth. This is what AllegroGraph's Entity-Event Data Modeling brings to the table.

ALLEGROGRAPH—PROVEN SCALABILITY FOR ENTERPRISE KNOWLEDGE GRAPHS

AllegroGraph v7 provides a patented solution for distributing semantic graph data that is a hybrid of horizontal sharding with vertical federation of industry specific knowledge bases. In AllegroGraph's FedShard[tm] approach every shard can be federated with a knowledge base that lives on the same machine as the shard which makes it several orders of magnitude easier to perform ad-hoc queries, semantic reasoning, and feature extraction for machine learning and advanced analytics over the entire entity population. This capability would be impossible with labeled Property Graph Databases because of their inability to process data in complex knowledge bases and taxonomy systems.

AllegroGraph significantly enhances the document database model with its native support for JSON and JSON-LD. AllegroGraph leverages JSON-LD to swiftly integrate with web-based applications. Organizations can therefore link specific information in their internal Knowledge Graphs (e.g., pertaining to customers or products) to web applications for timely action such as recommendations. JSON-LD provides a way to create a network of standards-based, machine-readable data across websites.

KNOWLEDGE GRAPH CONSULTING

Franz provides a variety of services as part of its Knowledge Graph solution, from architectural consulting and technical seminars to training. AllegroGraph provides the necessary power and flexibility to address high-security data environments such as HIPAA, privacy rules for banks, and security models for policing and intelligence. If you really want to develop your corporate Knowledge Graph and address complex AI problems, you need a data system that goes beyond just data. You have to create a system that can link to anything outside your own predefined parameters—and that can learn from previous experiences. That is where a Graph Database, like AllegroGraph, comes into the picture.

Franz, Inc.
franz.com