Information Management article — The Unparalleled Utility of Graph Databases

Many aspects of data management—particularly concerning big data—hinge upon the utility of graph databases. When deployed with additional semantic technologies such as ontologies, taxonomies and vocabularies, there are few analytic feats an RDF graph cannot achieve. In most instances, end users are largely unaware of the degree of complexity that semantic graphs account for when linking and contextualizing disparate data elements for unified results.

Graph databases initially gained prominence with use cases involving social media and facets of sentiment analysis; this technology gained credence by provisioning '360 degree views' of customer and product information in MDM systems. Other commonly found uses of graph databases include applications of time-sensitive data such as recommender engines for e-commerce, fraud detection for finance, search engine augmentation, and ERP optimization.

But as valuable and as proven as these individual uses cases are they only hint at, and do not truly attest to, the full array of possibilities of databases powered by semantic graphs.

Today,semantic graphs are greatly expanding the utility of data lakes in a sustainable manner. The enhanced analytic capabilities of semantic graphs are integral to cognitive computing analytics, as well as to analysis of integrated unstructured and structured data.

Comprehensive enterprise data

The true potential of semantic graphs is realized in linking the entire information assets of an organization for comprehensive analytics of overall internal data as well as public data and third-party data. The underlying architecture for such an undertaking commonly involves Hadoop or some other other data lake to account for issues of scale; ontologies are required for a semantically consistent model, and terminology systems are needed to clarify terms and definitions.

Once this architecture is in place, an RDF graph links the data via the semantic model to distinguish points of relevance between all data throughout the enterprise. Laymen end users in the business can access all of their company's data when performing analytics.

Best of all, semantic graphs are responsible for denoting just how a particular node relates to another node to inform analytics with a critical element of context. End users need not understand those relationships themselves to benefit from the propensity of semantic graphs to link and identify relationships between any type of data in a comprehensive semantic model.

Integrating multiple data structures

The comprehensive nature of semantic graph-based analytics partly pertains to its incorporation of structured data alongside unstructured and semi-structured data.

Once sources are loaded, semantic graphs enable expedient data integration based on their ability to link different data elements. Typically, such integration efforts involve a combination of internal, proprietary structured data with external big data that is either structured or semi-structured.

As long as that data is incorporated into the overarching semantic model, an RDF graph can determine which sources are relevant to a particular query and how the individual nodes

relate to one another. This facet of semantic graphs enables a ready incorporation of different types of data with one another, while accelerating conventional integration processes that required lengthy periods of IT remodeling data when sources or business requirements changed.

Cognitive analytics

A unique advantage that semantic graph databases provide is the ability to organize and link data similar to how humans process information.The semantic view of data is a human one which is much more flexible than the longstanding relational models and tables that are simply too rigid for big data. The variety that big data encompasses mandates technologies that can enable users to incorporate agile methods to keep pace with shifting business practices and requirements.

Semantic graphs are suited for such endeavors since they link and contextualize data with an expeditiousness that is commensurate with big data applications. Certain cognitive solutions are able to utilize this human-like aspect of graph-based analytics to provide rationale or explanations for analytics results.

Most importantly, understanding the context of data and how it relates to business objectives improves the type of questions asked of it and the results of analytics performed. There is a confluence of semantic technologies responsible for gleaning that context—semantic graphs are arguably at the forefront of them.

Contextualized linking

The ability to link, form relationships between, and understand the correlation between different data types that is the crux of the value that graph databases facilitate will prove invaluable as data management continues to increase the benefit it derives from big data technologies.

Part of that usefulness has impacted certain data preparation processes, as semantic graphs require data to adhere to an RDF format. Nonetheless, with tools such as Spark, text analytics, and mapping procedures used to accomplish such purposes, the manual transformation procedures that threatened to consume the time of data scientists and IT is not needed to leverage graph analytics.

Even more significantly, graph databases can do the difficult work of integrating and contextualizing data for a true self-service analytics experience that continually pushes data management further into the hands of the business.