

DB Engines article – Master Data Management Remains Siloed Without Semantics

As Big Data continues to grow, companies are increasing their commitments to extract every last drop of value from the information they gather. Since siloed data has limited value, organizations are also embracing a more comprehensive strategy of managing data across multiple data domains. But this requires a shift away from traditional Master Data Management (MDM) to broader and more flexible data governance environments that enable enterprises to quickly synthesize diverse data sources that can result in transformative business decisions.

As a result, there has been growing interest in using graph databases, rather than relational databases, to store master data. This is an eminently sensible switch, as graph databases offer a 360-degree view of master data and can answer questions about data relationships in real time, providing new, actionable insights from existing data. In addition, semantic graph databases can serve as the glue to connect new or different data sources.

As Forrester analysts Michele Goetz says, “MDM is not a data integration tool... But, what if you changed the underlying repository to a graph? It immediately changes the mindset and strategy of MDM from systems to views (*It becomes*) much more intuitive, analytic and intelligent about our master data.” Put simply, semantic graphs have the power to unlock the potential behind MDM.

Any graph database architecture will provide noticeable benefits for MDM processes. But the best graph database match for MDM usage is a semantic graph database. MDM is, after all,

grounded in semantics – the relationships between different classes of data. Let's discuss the benefits that graph databases bring to MDM and the specific use cases that make semantic graph databases the obvious choice for MDM projects. First we'll look at the foundation and functions of MDM, and explore how they're complemented by semantic structures.

Know Your Data

Master Data Management (MDM) is not a type of technology—it's an approach aimed at identifying the organization's most valuable data, and integrating that data in a way that it provides a comprehensive, authoritative view of the business.

This trustworthy data is then used to inform users towards better decisions and more efficient business processes, through deep connections across all information silos and data storehouses. This may sound like a simple task in principle, but in practice, it's more challenging, due in part to enterprises' tendencies to maintain multiple solutions across business divisions. And within that assortment of data storage, there will be one or (usually) more of the following data types:

Structured: Information that can be easily broken into defined sets suitable for storing in a traditional (relational) database format. This data is organized into "tables" that are comprised of rows and columns, with row representing a single entry (instance) with a key that can be linked to rows in other tables. Each column ("attribute") represents a value that further describes the row.

Unstructured/semistructured: This type of data doesn't neatly fit into the traditional relational database format. It is typically comprised of video clips, text messages, social media interactions, customer care transcripts or recordings, images, information collected from sensors, and other types of "freeform" collateral.

Transactional: This type of data contains information related to sales, billing, payments, claims, contracts, customer care, bug fixes, etc. and tends to be structured. It often is of analytic value for a limited period of time.

Non-Transactional: Customer information like names, locations and preferences are examples of non-transaction data. This type of data tends to retain value over a long period of time, and makes up much of the information businesses work with in MDM.

Metadata is data that describes data, typically contained in configuration files, XML documents, database column descriptions, catalogs, and spreadsheets.

Hierarchical data stores information about data relationships, and its functionality falls into the same domain as MDM, as it is intended to reveal the connections across disparate data sets.

Master Data is typically organized by people, things, places, and (abstract) concepts, which are then further divided by domains, entity or subject types. As an example, in “People”, a company would store information about executives, employees, customers, contractors, etc. “Customers” might be further divided by loyalty status, locations, and other types of demographic data.

So, how do we make all of that data work together as a team? With semantics to find all of the relationships that exist across all of that data.

Powering Data with Semantic Graphs

Again, it’s important to remember that MDM is really not a data integration tool, but a process to manage data models. You need a foundation to store and organize the data. As noted above, there are across-the board benefits to using semantic graph databases as the underlying structure, but there are

critical specific use cases where using a semantic graph database is essential.

For instance, if you choose to use MDM over multiple internal and external databases, you'll want to deploy a semantic graph database. Property graphs simply won't do the job. Semantic graph (Triplestore) databases enable data architects to model the data in as a set of triples with rules to easily integrate the data. Triples are easy to define as they use a semantics data model (subject, predicate, and object) to link two entities (people, places, or things) – the relationship between them then forms the “triple”.

Subject: BobJones	Predicate: buyswidgets
Object: AcmeWidgets	

When linked together, triples form a graph that would, using the example above, provide us with further information about our customers' widget buying habits and preferences.

Trustworthy Data

The point of MDM is to present an authoritative view of the company's most important data. Semantic Graph databases are the tool of choice here, because writing a triple is so straightforward that it is easy to integrate databases with no loss of “truth”.

If you try to use property graphs for MDM you run into the problems because it doesn't have any standards for naming of nodes and links between the nodes. To collaborate with other enterprise MDM hubs, you'd have to develop a common naming scheme to ensure the identifiers line up. Semantic graph databases are architected to share datasets and standardized to provide seamless connectivity with other semantic graph databases. So it is fair to say that property graph databases would have to reinvent the naming standards for semantic graphs.

Deriving More Value From MDM

To use MDM to query over complex and changing data types, such as unstructured information, you need a semantic graph database. The prevalence of unstructured data – some estimates indicate that at least 80% of all corporate data is unstructured – makes semantic graphs uniquely suited for MDM. As Dataversity's Jelani Harper clearly stated in his recent article, *Revamping Master Data Management with Graphs*, "The most distinct advantage of augmenting MDM systems with graph databases is reducing the complexity of integrating and querying vastly different source data."

Semantic graphs excel in their ability to correctly connect disparate sources of information. The most robust semantic graph solutions can create these links independently, with no need for predetermined schema or user intervention, making them uniquely well suited for dealing with a real-time data flow. Semantic graphs bring sanity to what is traditionally a labor-intensive and error-prone process.

About the Author: Jans Aasman is Ph.D. psychologist and expert in the Cognitive Science – as well as CEO of Franz.com, an early innovator in Artificial Intelligence and provider of Semantic Graph Databases and Analytics. As both a scientist and CEO, Dr. Aasman continues to break ground in the areas of Artificial Intelligence and Semantic Databases as he works hand-in-hand with organizations such as Montefiore Medical Center, Blue Cross/Blue Shield, Siemens, Merck, Pfizer, Wells Fargo, BAE Systems as well as US and Foreign governments.

Dr. Aasman is a frequent speaker within the Semantic technology industry and has authored multiple research papers, bylines and is one of 15 CEOs interviewed in a new book, "Startup Best Practices".

Dr. Aasman spent a large part of his professional life in telecommunications research, specializing in applied

Artificial Intelligence projects and intelligent user interfaces. He gathered patents in the areas of speech technology, multimodal user interaction, recommendation engines while developing precursor technology for the iPad and Siri from 1995 to 2004. He was also a part-time professor in the Industrial Design department of the Technical University of Delft.