Inside Analysis article – AllegroGraph: Think of It as Conceptual Computing

Franz Inc. has impressive technology, semantic graph technology that goes by the name of AllegroGraph. I could describe the technical nuts and bolts of this in detail – how it organizes data and indexes it and integrates it – but I'd rather not drag you into the weeds. Just take my word for it that the AllegroGraph technology environment is sophisticated.

If you want to understand what AllegroGraph can do in practice, think of its data layer as storage for every kind of data: structured data from databases or data warehouses, device IoT data and log file data and unstructured data such as text and notes. Now envisage AllegroGraph combined with Hadoop, creating a "semantic data lake" that combines all related definitional data: database schemas, dictionaries of terminology and full-blown ontologies for given areas of business activity. Finally, think of this rich semantic data lake as being coherently integrated and available to a multitude of applications.

The Semantic Data Lake

Such a data lake would be able to support every variety of BI application and every variety of analytics: discovery, predictive analytics, machine learning and so on. In that respect it would be equivalent to a well-organized Hadoop data lake. AllegroGraph is in fact implemented on Hadoop/HDFS and can support all such activity. But it is more than that and capable of much more. This is due primarily to the fact that it stores all data as triples. A triple, if you didn't know, is the smallest unit of data that embodies meaning, which it does in the form of subject-predicate-object (noun-

relationship/verb-noun). All data can be broken down into triples, which is what AllegroGraph does to the data it stores.

However, when AllegroGraph stores data it does more than that. It captures the meaning of the data, modeling the events in the data and annotating the meaning in respect of its unified ontology. (If the word ontology makes your eyes glaze over, then think of this as a unified dictionary of terminology.) It is important to understand precisely what it does. In a sense, the data is being stored according to its meaning in a manner that is analogous to how data is stored in our human memory. AllegroGraph associates data objects one to another meaningfully, creating links between them that can be expressed as graphs.

This can be put to practical use in innovative ways. In this blog post, I'll discuss the big picture.

Semantic Technology Comes of Age

I suspected that semantic technology was approaching maturity when IBM injected heaps of marketing dollars into its Watson technology, began to beat the drum for "cognitive computing" and invented a whole new buzzword. If you're not sure what cognitive computing is, it is (according to a simple definition I found on TechTarget.com) the computer simulation of human thought processes. And maybe that was the impression you got when IBM's Watson beat three past champions at the very "cognitive" game of Jeopardy — it was "out-thinking" them.

In truth, the science of psychology has not yet advanced to the stage where it can rigorously define all human thought processes and describe them definitively, but most aspects of "cognition" are reasonably well understood and can be computerized. However, in my view, IBM's marketing staff has muddied the waters by throwing pretty much everything,

including the kitchen sink, into its description of Watson and its articulation of cognitive computing. It seems to include data analytics of every variety, including machine learning, natural language processing, reasoning (to the limits of AI and including probabilistic algorithms), pattern recognition of every variety, heuristics and semantics.

And naturally, IBM has a collection of software products that provide each of these capabilities in some way. But nevertheless it has not explained cognitive computing in any coherent way. Perhaps it should explain its explanation.

IBM claims that a cognitive system can "understand data, learn from it and reason through it." And therein lies the problem. In the hope of exciting undue enthusiasm, IBM's marketing mavens have decided that cognitive computing equates to understanding. It cannot and does not. Computers can neither understand nor be aware. All they can do is model human cognitive processes.

Conceptual Computing and the Semantic Data Lake

What AllegroGraph provides is, in my view, a breath of fresh air and a more coherent approach. I prefer to think of what it delivers as "conceptual" rather than "cognitive" computing because of the way it computes over concepts rather than just dumb, raw data. By gathering as much definitional data as possible — database schemas and dictionaries of terminology and ontologies — and by rationalizing them and also all associated data on ingest, AllegroGraph creates a "conceptual record" of the data. Each data entity is provided with a context whose meaning is known and can be illustrated simply by displaying a graph of its relationship to other entities.
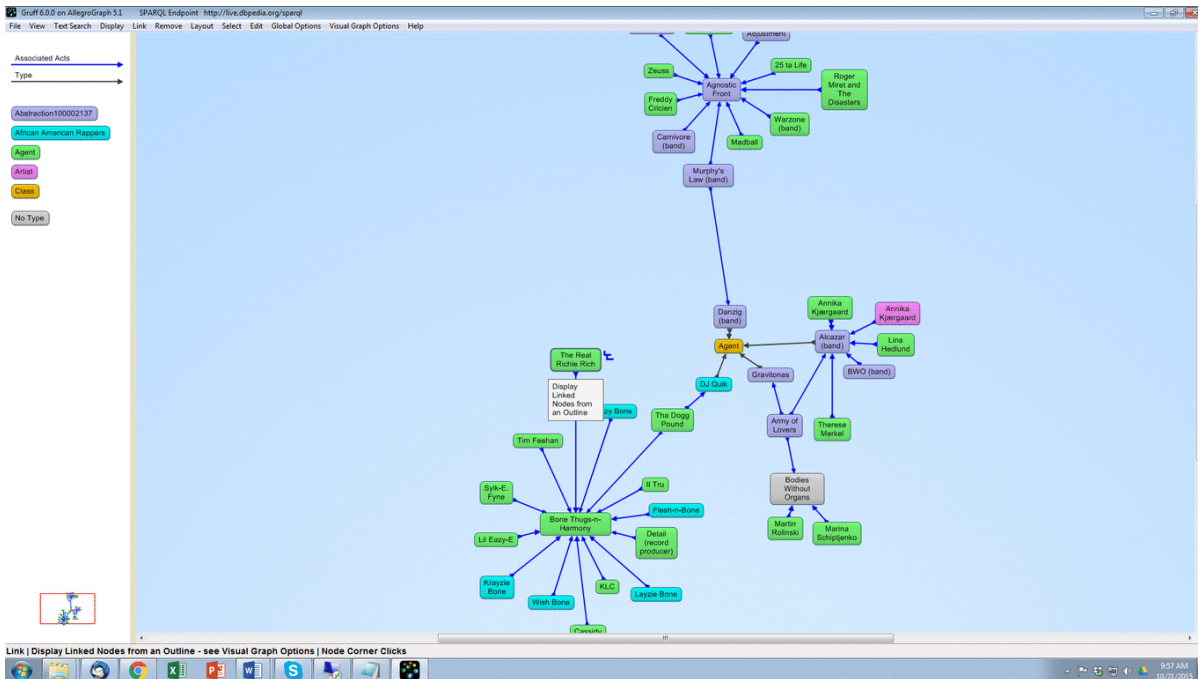
Figure 1. AllegroGraph in Action

AllegroGraph provides a particularly neat way of examining the context of any entity it stores or any instance of it through a contextual browser. The beauty of this lies to some extent in its ability to enhance the user's understanding of the data. The data is arranged in the way the brain probably arranges data by meaningful association. It can thus be traversed in that manner whether one is examining an individual person or a related group of people or a network of people. This is illustrated by Figure 1.

The point is that the Semantic Data Lake provides the conceptual foundation for exploring the data and building applications that process it. And if you want to plug in analytics routines, machine learning routines, pattern matching capability, reasoning software and other semantic applications, then you can. Maybe such a collection of sophisticated applications plugged into the semantic data lake can be thought of as cognitive computing, but the basis of it is the conceptual framework that AllegroGraph provides.