Patient Treatments

Accelerated data analysis finds ideal treatments for individual patients

The healthcare organization has spent tens of millions of dollars on collecting data into its enterprise data warehouse but has not yet been able to what it believes is the full potential of that data to improve patient outcomes.

The Business Problem

The databases of a leading healthcare organization hold the records of 13 million patients, and the volumes of records are expanding by the day. Inside that big data are detailed patient information, treatment regimens and outcomes, all waiting to be tapped to help treat current cases — if they can be queried appropriately in time to make a difference.

The difficulties include both the time to set up a relational database system with the data and queries required for the needs of a particular specialty (say cardiology), which often takes four months, and then the time to execute the appropriate queries, which is believed to take half a day, assuming that the problem can be properly solved with RDBMS technology.

The healthcare organization's primary focus is on improving individual patient care, which in the big data world means harnessing the potential of its data. Today, cardiologists apply the standard therapy recommendations, which are for a generic patient, independent of age, clinical history, etc., and they have no automated way of matching a given patient with their data to understand that patient's likely experience. One initial focus is to compare the medical history, including clinical records, lab tests and prescriptions, for a given patient and the patient's family to other patients in their records , with the hope of finding similar patients.

Once these similar patients are identified, a cardiologist can view the outcomes of the various therapies that were tried with them to choose the best option for the current patient. This comparison needs to take place during a typical office visit, with the patient sitting in the cardiologist's office expecting answers — and hope. The cardiologist can wait at most one or two minutes for an answer, and most cardiologists want to tailor the results based on their own experience.

The Technical Challenge

Accomplishing this has proven to be difficult or impossible using an approach based on RDBMSs. It would need to examine the contents of multiple databases housed on multiple servers, perform queries on incomplete data using fuzzy criteria (which is daunting using SQL), and present results in one or two minutes in the form of easy-to-comprehend graphs. As if this weren't difficult enough, the typical RDBMS approach of creating a purpose-built data mart for a given data and query does not support the cardiologists' need to tailor their questions somewhat to match their experience.

The Urika-GD™ Graph Analytics Solution

To discover hidden and unknown relationships in big data with the Urika-GD™ appliance, a relationship warehouse has to be constructed, in this case using event, disease, lab and drug data integrated into a consistent set of relationships, represented as a graph. From that, the Urika-GD system's inference engine augments existing relationships through deduction, such as noting that fluvastatin (a specific statin) is also a member of the statin class, so selecting only patients who have taken statins is more easily executed. As a result, the warehouse contains the most complete set of relationships available. It can be continuously updated either through the addition of new sources of data or through updates to existing sources with additional inference/deduction taking place each time to reveal new relationships, as the data is improved in an ongoing cycle.

Exploiting the intense curation in this relationship warehouse, queries then match the specific relationships needed by a cardiologist. Who are the patients who also had a body-mass index greater than 30 (obese) and family history of high blood pressure before the age of 30, and so started a regimen of statins at that age? Who of them also started an exercise regimen that decreased their body-mass index to less than 25 (normal) and so were able to discontinue statins before the age of 35?

Then the health outcomes of these similar patients can be mined. Of these similar patients, what were the outcomes for those who stayed off statins for five years compared to those who continued low doses of statins? And how did the low-dose results for older, less expensive statins compare with those for newer, more expensive ones? Armed with patient-specific knowledge like this, the cardiologist can make better recommendations that address both patient care and cost effectiveness.

Being able to find and compare similar patients turns the latent value of this healthcare organization's data warehouses into realized value, saving expense, trauma and lives.

About Urika-GD The Urika-GD big data appliance for graph analytics helps enterprises gain key insights by discovering relationships in big data. Its highly scalable, real-time graph analytics warehouse supports ad hoc queries, pattern-based searches, inferencing and deduction. The Urika-GD appliance complements an existing data warehouse or Hadoop® cluster by offloading graph workloads and interoperating within the existing analytics workflow.

About Cray Global supercomputing leader Cray Inc. provides innovative systems and solutions enabling scientists and engineers in industry, academia and government to meet existing and future simulation and analytics challenges. Leveraging more than 40 years of experience in developing and servicing the world's most advanced supercomputers, Cray offers a comprehensive portfolio of supercomputers and big data storage and analytics solutions delivering unrivaled performance, efficiency and scalability. Go to www.cray.com for more information.

THE SUPERCOMPUTER COMPANY