Solution Brief

**Diagnosis: Positive – How a health organization's
vastly accelerated data analysis finds ideal treatments
for individual patients**

The Business Problem

There are 13 million patients in the records of a leading healthcare organization's databases, 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 healthcare organization has spent tens of millions of dollars on collecting data into its enterprise data warehouse but has not yet been able to reap the full believed potential of that data to improve patient outcomes. 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, once set up, the time to execute the appropriate queries, which are 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 a given patient's medical history, including clinical records, lab tests, and prescriptions for the patient and the patient's family, to the patients in their records to find similar patients.

Once these similar patients are identified, a cardiologist can view the outcomes of the various therapies that have been tried on those similar other patients to choose the best option for the 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 1-2 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 1-2 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™ Solution

To discover hidden and unknown relationships in big data with the Urika 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 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 intentional 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 taking into account 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

YarcData's Urika is a big data appliance for graph analytics. Urika helps enterprises gain business insight by discovering relationships in big data. It's highly-scalable, real-time graph analytics warehouse supports ad hoc queries, pattern-based searches, inferencing and deduction. Urika complements an existing data warehouse or Hadoop cluster by offloading graph workloads and interoperating within the existing analytics workflow. Subscription pricing or on-premise deployment of the appliance eases Urika adoption into existing IT environments.

About YarcData

YarcData, a Cray company, delivers business-focused real-time graph analytics for enterprises to gain insight by discovering unknown relationships in big data. Adopters include the Institute of Systems Biology, the Mayo Clinic, Noblis, Sandia National Labs, as well as multiple deployments in the US government. YarcData is based in the San Francisco bay area and more information is at www.yarcdata.com.