



May 2012

Document M41

RESEARCH NOTE

INFORMATICA: BRIDGING THE GAP BETWEEN TRADITIONAL DATA AND BIG DATA

THE BOTTOM LINE

Companies seeking to maximize the value of their analytics investments must learn to integrate Big Data into their current data management strategies. Informatica 9.5 focuses on this challenge by supporting the accelerated growth of information, the unique challenges of social and machine data, and the management difficulties of processing Big Data through desired business logic. Informatica has taken important steps towards the evolution of Big Data from a complex tool accessible only to data analysts and developers into a broadly available information source for line-of-business employees.

INFORMATICA 9.5 IN CONTEXT

On May 15, Informatica announced the launch of Informatica 9.5. This version is intended to provide multiple big data capabilities that will allow enterprises to take advantage of multiple Big Data attributes. Although Informatica is typically best known for its data integration and data governance capabilities, the 9.5 release adds new support for Hadoop, natural language processing, social networking data which Informatica supports with its Social Master Data Management (MDM) offering, and drag-and-drop data mapping to increase the usability of Big Data and traditional data.

There are three aspects of Big Data that Informatica is most interested in supporting with this release, each of which line up favorably with the key drivers Nucleus has identified for Big Data success.

BIG TRANSACTIONAL DATA

Inbound transactional data has quickly increased as the automation of data and the expansion of data sources has led to the increased growth of data storage and volumes. As data has grown throughout every department, companies have started to realize the opportunities of these sustained data collection efforts through the effective aggregation and integration of these large transactional data sources. A major manufacturer obtained an ROI of 942 percent by using Big Data capabilities to examine purchasing and cost-related data in all of its vendors' databases, leading to vendor consolidation and reduced cost of goods sold (Nucleus Research *m20 - The Big Returns from Big Data*, April 2012).

To support the increasing fragmentation of data and the proliferation of data sources, Informatica 9.5's cloud integration services allow data from any source to be included within any cloud based application, leading to the union of cloud applications with real-time data updates. Data is only useful when it is provided to the correct stakeholders in the correct formats at the right time. Until data is properly delivered in context, data collection is a trivial exercise. By extending cloud data integration capabilities directly into cloud applications, Informatica provides a vital step in making data ubiquitous and relevant to the enterprise.

INTERACTION-BASED DATA

Enterprise data has traditionally been limited to transactions that served as direct inputs for enterprise applications. Over time, the challenges of supply chain, manufacturing, and field services required specialized tracking of machines and devices that fell outside of the traditional enterprise data model, but each of these exceptions were handled as a department-specific solution that could not be broadly scaled throughout the enterprise. With the spread of subscription-based billing, social networking, machine-to-machine interactions, and other automated machine data sources, every department and every organization now has this potential problem of tracking and analyzing these interactions.

The ability to effectively parse social, transactional, and machine data is necessary to take advantage of the external sources associated with Big Data. Informatica has integrated social data with traditional data sources to create a Social MDM solution that provides additional perspective on customers and audiences. In addition, Informatica 9.5 supports natural language processing to quickly translate text-based social information from semi-structured and unstructured data inputs to structured data that meet the human language requirements of a business environment.

THE PROCESSING OF BIG DATA

The challenges of processing Big Data are not just based on volume. As the business logic associated with processing becomes more complicated, traditional data syntax becomes more unwieldy and less efficient than newer frameworks, such as Hadoop, to support these needs. In addition, as multiple groups within an organization seek to access the same information simultaneously, a Big Data approach is more likely to lead to line-of-business success in accessing data in a timely fashion.

All of these Big Data problems constitute challenges for companies that understand the value of Big Data but may not be sure how to fully support a Big Data deployment. Nucleus finds that Big Data projects have led to an ROI of 241 percent, which justifies these efforts (Nucleus Research *m20 - The Big Returns from Big Data*, April 2012). However, the success of these efforts is dependent on using Big Data solutions that do not simply focus on the scalability of data volumes, but also consider the challenges of data

sources and context as well as the challenges associated with complex business logic and simultaneous data access.

Although Hadoop and other data processing platforms have gained acceptance over the past couple of years as core tools to translate massive volumes of Big Data into human-friendly results, these tools tend to be technically challenging to integrate with existing enterprise data deployments. To ease this challenge, Informatica 9.5 has created an integrated development environment designed to support business processes based on a visual layout, which will be beta released in July. This allows Big Data processing and transformations to be translated into business friendly formats more easily, which is a necessary step in making Big Data ubiquitous. Nucleus found that the largest incremental gains in analytics deployments occurred when these technologies were deployed across a majority of an organization to move from the Tactical stage to the Strategic stage of the Analytic Enterprise (Nucleus Research *m17 – The Four Stages of the Analytic Enterprise*, April 2012).

SUMMARY AND RECOMMENDATION

Big Data represents the convergence of growing relational databases, increasing numbers of relevant data sources, greater amounts of social and machine-based data, the cloud as a scalable storage tool and computing resource, and the business demand for additional context associated with strategic and tactical decisions. This convergence has made Big Data difficult to fully manage and integrate, since this combination is not a single trend but the combination of multiple key trends in business analytics.

An integrated Big Data processing environment must include a strategy to identify all relevant data sources associated with an analytics project, transform all incoming data into relevant business formats, process and analyze the data through business logic, and then provide the data on an as-needed basis to all relevant employees. Informatica has taken all of these trends into account in its Informatica 9.5 release to connect traditional data integration environments with the emerging world of Big Data.

Nucleus recommends that businesses seeking to maximize the benefits of Big Data should pursue solutions that reflect the multi-faceted nature of Big Data. By doing so, these Analytic Enterprises can effectively integrate Big Data into their current business analytics and business intelligence projects.