



 BUSINESS ANALYTICS SERVICES
KNOW MORE. DO MORE.



SOLVING KEY **BUSINESS CHALLENGES** WITH A BIG DATA LAKE



**BIG DATA &
BUSINESS
ANALYTICS**

AUTHOR:

JOHN WILLS

GLOBAL DIRECTOR, CENTER OF EXCELLENCE
HCL BUSINESS ANALYTICS SERVICES

TABLE OF CONTENTS

BUSINESS SPEED AND FLEXIBILITY	3
UNCHAINING INNOVATION	4
COST-EFFECTIVE DATA MANAGEMENT	4
BIG DATA LAKE R&D FACTORY	4
DATA FABRIC FOR THE DIGITAL REVOLUTION	5
HCL'S BIG DATA LAKE SOLUTION	5
HCL EXECUTIVE BIOGRAPHY	6
ABOUT HCL	7

After more than five years of big data evolution, a second generation of more mature technology and architecture options have started emerging. One of the most significant trends is the concept of the big data lake.

The big data lake is a framework that seamlessly integrates core enterprise data assets. It assumes that each enterprise will have an ongoing need for an array of data assets including traditional relational data warehouses, in-memory analytics, columnar/key pair databases, noSQL document databases and of course a distributed file system for the storage of raw data.

Conceptually, the big data lake provides three things.

- A standard staging area for raw data needed by all other data platforms
- A standard transformation environment
- An overarching governance, security and integration mechanisms that link data assets together

It is this third characteristic that distinguishes it as an architecture and framework. The following examples will help illustrate.

- It is what allows a user to be working in a reporting tool consuming data supplied by a data warehouse and then drill into detail from any other data platform such as Hadoop.
- It is the ability for an in-memory analytic system to automatically shift data between memory, near-line relational storage and Hadoop based on usage statistics.
- It is the ability for an analytic model to use data from multiple sources such as Hadoop, a data warehouse and master data to determine the next best recommended product for a customer.

These are interesting technical capabilities but what are the business implications of implementing a big data lake?

BUSINESS SPEED AND FLEXIBILITY

Business access to data has always been controlled and constrained by Information Technology (IT) organizations because of various reasons including the cost of data storage, data skills within the business, security and data quality. However, the dynamics of data usage have changed. Business is now a more sophisticated consumer of data. It also wants to access a broader range of data at a lower cost of storing. The big data lake allows an organization to shift its focus from centralized control to a shared model to respond to the changing dynamics of information management. This means that IT can assist in capturing and consolidating data quickly into the data lake and the business can be given immediate access without waiting weeks and months for it to be cleaned, scrubbed, modeled and loaded. IT can maintain its overall governance and access control but the business accepts responsibility for using data in its raw form as a trade-off for speed of access.

UNCHAINING INNOVATION

In the past data warehouses were used primarily to report and analyze past business performance using structured data based on business transactions. While that is still important, there has been a huge shift to using a broader set of semistructured and unstructured data from internal and external sources to predict behavior and optimize business operations. Instead of the key driver for this innovation being business transactions its focuses on events and the context of the event (location, who, when, comments, environment, etc.). The problem has been a mismatch between this business requirement and the speed at which IT could respond. The result has been 'shadow' IT funding within the business for sourcing, storing and integrating their own data.

The big data lake brings the work of managing the data back out of the shadows and shifts it to IT where it belongs. By doing so, it actually frees up the business to reallocate resources toward higher value analytics for exploration and discovery.

COST-EFFECTIVE DATA MANAGEMENT

One of the great inhibitors of innovation in the past was cost. It was simply not cost-feasible to add the hardware or specialized software required to scale to the data volumes needed by the business to search for experimental ways of using data to gain a competitive advantage. Big data came along and lowered some of the cost burden by providing a commodity priced storage and compute platform but truth be told it was pretty messy. There was really no integration with the enterprise data warehouse (EDW), master data, reference data and other emerging computing capabilities and security was an afterthought. Much of the cost savings was actually given back during expensive custom integration using a small talent pool. The big data lake is changing this with a comprehensive security model and prebuilt integration between data platforms and technologies. This is driving down the cost of custom integration, high priced talent while also taking advantage of commodity priced hardware.

A second aspect of cost that the big data lake helps with is the transformation of data. In the past the movement and transformation of data has required licensing of highly specialized tools (ETL, EAI) and dedicated hardware. The big data lake opens up the possibility to shift this cost to a commodity platform and even extend that to a data-as-a-service approach.

BIG DATA LAKE R&D FACTORY

It would be a mistake to think of the big data lake as an end-point used to produce reports and analysis much like the EDW. Instead it's an R&D factory aimed at producing a product (analytic techniques and models) that can be deployed and embedded as part of a business process. The use of this process then initiates the next cycle by feeding new data back into the factory. This concept of continuous process improvement, experimentation and innovation is not new but having the means and capability to accomplish it is.

It may also help to consider and contrast roles of the EDW and big data lake. The EDW provides a self-service environment primarily for reporting to look back at what happened. The big data lake provides a self-service environment for the development of analytics used to proactively effect and optimize business operations.

While this cycle of improvement might be welcomed by a number of businesses it's not a process that is well-defined and understood.. To solve this problem the goal should be to define and implement a big data lake governance process that defines when, how and who promotes experimental work into a more traditional development and deployment process.

DATA FABRIC FOR THE DIGITAL REVOLUTION

Nimble companies with visionary leaders think first about the digital experience their customers have with their company and are looking for ways to enhance this experience. To enable this they are looking for a new data fabric that supports the democratization of data by opening it up to business experts who can figure out how to put it to use. Consider the difference between video rental and on-demand streaming, shrink wrap software against online app purchases, online shopping, online banking, and more. Leading companies in all industries are simply leapfrogging those stuck in old processes revolving around bricks and mortar instead of digital first. The big data lake is of course the core of this new metaphorical data fabric. It is the next generation of data architecture and a framework for a cost-effective and flexible data environment that can be used to reinvent business and compete in the digital economy.

HCL'S BIG DATA LAKE SOLUTION

HCL's Business Analytics Services organization provides an end-to-end big data framework and operating model that can be used for exploration of all sources including social media, internet of things (IoT), machine-to-machine (M2M), and business events and transactions.

It is used by chief information officers, chief data officers and chief marketing officers for rapid access to a growing array of data sources that allows the business to quickly derive insight and monetize big data.

It includes a comprehensive big data operating model that includes three phases – consolidation, liberation and process for closing the analytic loop – enabled by an agile analytics development method and governance process.

For more information, please contact HCL at ets.bis@hcl.com.

HCL EXECUTIVE BIOGRAPHY



JOHN WILLS

Global Director, Center of Excellence
HCL Business Analytics Services

HCL

John Wills leads HCL's global Center of Excellence Business Analytic Services with responsibilities for solution development, labs, intellectual property and knowledge management.

Wills has 24 years of business intelligence experience in a number of diverse leadership roles including responsibility for consulting, sales, marketing and product development.

ABOUT HCL

About HCL Technologies

HCL Technologies is a leading global IT services company working with clients in the areas that impact and redefine the core of their businesses. Since its emergence on the global landscape, and after its IPO in 1999, HCL has focused on 'transformational outsourcing', underlined by innovation and value creation, offering an integrated portfolio of services including software-led IT solutions, remote infrastructure management, engineering and R&D services and business services. HCL leverages its extensive global offshore infrastructure and network of offices in 31 countries to provide holistic, multi-service delivery in key industry verticals including Financial Services, Manufacturing, Consumer Services, Public Services and Healthcare & Life Sciences. HCL takes pride in its philosophy of 'Employees First, Customers Second' which empowers its 91,691 transformers to create real value for customers. HCL Technologies, along with its subsidiaries, had consolidated revenues of US\$ 5.4 billion, for the Financial Year ended as on 30th June 2014. For more information, please visit www.hcltech.com

About HCL Enterprise

HCL is a \$6.5 billion leading global technology and IT enterprise comprising two companies listed in India – HCL Technologies and HCL Infosystems. Founded in 1976, HCL is one of India's original IT garage start-ups. A pioneer of modern computing, HCL is a global transformational enterprise today. Its range of offerings includes product engineering, custom & package applications, BPO, IT infrastructure services, IT hardware, systems integration, and distribution of information and communications technology (ICT) products across a wide range of focused industry verticals. The HCL team consists of over 95,000 professionals of diverse nationalities, who operate from 31 countries including over 505 points of presence in India. HCL has partnerships with several leading global 1000 firms, including leading IT and technology firms. For more information, please visit www.hcl.com



Hello there! I am an Ideapreneur. I believe that sustainable business outcomes are driven by relationships nurtured through values like trust, transparency and flexibility. I respect the contract, but believe in going beyond through collaboration, applied innovation and new generation partnership models that put your interest above everything else. Right now 95,000 Ideapreneurs are in a Relationship Beyond the Contract™ with 500 customers in 31 countries. **How can I help you?**

Relationship[™]
BEYOND THE CONTRACT

HCL