

## Is Big Data the death of BI?

On face value, it seems that prior investments in Business Intelligence (BI) products will need to evolve considerably if they are to remain relevant in the Big Data revolution.

Every second of every day, many of the digital products, processes, and information we've come to be reliant on generate vast quantities of data. Smart phones, retail scanners, embedded sensors, social networking, logistics, traffic cameras, and a multitude of other devices capture so much information that the volume required to be processed is often unmanageable with traditional systems. As the need for data warehousing space, processing power, and analytical capabilities grows exponentially due to the low costs required to grow global online businesses, the amount of data required to be processed for competitive business operations has simply outgrown many of the technologies and processes originally designed to manage it.

	TRADITIONAL DATA	BIG DATA
Storage	Gigabytes to Terabytes	Petabytes to Exabytes
Location	Centralised	Distributed
Structure	Structured	Semi-structured to Unstructured
Categorisation	Static Data Model	Flat Schemas
Inter-relationships	Defined and complex	Few and complex

While massive scale data collection makes some companies seem like data hoarders, it serves a purpose; namely to transform the data into information that when organised and analysed helps business to streamline processes, improve productivity, increase profits, and provide other positive corporate benefits. However the speed at which data is now transmitted and received, the quantities being collected, and the variety of sources from which it emanates (the Velocity-Volume-Variety problem) is the source of new challenges. As insights become an increasingly valuable commodity, new architecture, tools, and practices of real value will be required to extract it. Today, as gigabytes have given way to petabytes, corporate business intelligence solutions that were once well equipped to handle the analysis requirements may now find it difficult to deal with the new unstructured scope and chaotic velocity of these data streams. Big Data solutions that are naturally effective at drawing meaning from massive unstructured data collections may be successful at exposing these in a form that enables them to be integrated with other corporate data but may lack at providing the best-of-breed analysis required within a specific market.

Some may question the necessity of gathering such large amounts of information in the first place. Consider successful diamond mining operations for a moment. A single, jewellery-grade gemstone is the by-product of hundreds of hours of painstakingly sifting material-the majority of which turns out to be worthless. Despite the difficulties it presents, every bit must be



Your IT Executive in the Cloud

---

crushed, sifted, and sorted to be sure that nothing of value slips through undetected. Likewise, companies with the ability to execute this process faster and more efficiently than their competitors are accordingly likely to dominate their markets.

These businesses, like successful mine operators, have both the ability to quickly uncover the hidden data “gems” and the capacity to effectively act on the decisions required to set them apart from their competitors.

Big Data grew outward from the need to warehouse enormous quantities of unstructured data, but quickly evolved new methods for processing it as well. Today images and “likes” often carry just as much weight as email addresses and customer satisfaction survey results. Most importantly, the value it holds can be extracted from the midst of even the largest, most unstructured data collections. A business that is ill-equipped to do this, misses important opportunities to capitalise on the value of their unstructured data.

Consider online retail giants such as Amazon. The simplest search involves the instantaneous comparison of your query with that of millions of other shoppers, browsers, ratings, reviews, and other data. Within seconds you receive results tailored to your specific needs, despite the fact that millions of pieces of information from multiple sources may have been processed to do so. Without the slightest bit of retailer intervention, the backend delivers sophisticated analytics that scrutinise data and uncover relationships human users might never have considered. This is where an integrated big data and BI platform that incorporates sophisticated product matching algorithms across large unstructured datasets provides a significantly higher perceived value than a simplistic field-based search and analytics engine.

BI value propositions go well beyond useful the standardised business answers derived from data collections with similar characteristics. Whilst they can easily facilitate the essential operational reporting required for guiding management decision making, they are often underutilised when executives are seeking answers to unstructured queries around irregular or more complex business scenarios. This shortcoming emerges from the tendency of businesses to rely almost entirely on the technology platform rather than establishing a solid intelligence based capability around the BI solution.

Additionally, reporting platforms aren’t always able to offer a clear representation of what is actually impacting business profitability. They form part of a reactive top-down approach that relies on corporate managers to interpret these reports and take meaningful and measurable corrective actions. Many non-managerial individuals are never made aware of the data they need to improve processes and initiate change. A capability must be built around these systems that guarantees useful data is made available to the right people in the business.

The highest business value is exposed through the efficiency of self-serve analytics. When workers at different levels have access to relevant data, the “middle man” becomes unnecessary, and valuable information gets it to the right place faster. For example, production line supervisors might use capabilities to identify productivity problems and devise workable solutions. Marketing departments could identify profitable customers and direct sales efforts in their direction more successfully. Investors might use data to determine ROI, and then direct their financial support to areas where it will have the greatest impact.



Your IT Executive in the Cloud

Mega retailer Walmart is one example of a company which has made good use of Big Data capabilities with the introduction of its search engine, Polaris. To improve online shopping, the platform utilises machine learning capabilities, text analysis, synonym mining, and other semantic data to deliver relevant search results. Polaris is effective because it factors in previously untapped sources of insight to generate accurate, tailor-made product recommendations. Now, nearly anything consumers say, like, or share about a particular product can be used to help others find exactly what they need. As a result, profits from Walmart shoppers who completed purchases online have risen by 10%; an increase equivalent to billions of dollars.

EBay-owned online ticket marketer StubHub has also realised substantial value through the integration of Big Data solutions with higher analytics. One of the challenges they faced was providing effective fraud alerting and cyber theft protection for customers as well as for the company itself. Some of the most common problems involved consumers who were scammed while using the site's fan-to-fan ticket marketing feature. Not only did users receive bogus tickets; their interaction with the site compromised their banking and other personal information. To address the problem, StubHub elected to apply the R Enterprise component of the Oracle's Advanced Analytics Database option. R is a programming language and software environment for statistical computing and graphics. The use of this R Enterprise component allowed highly skilled system users to run their advanced R language queries across the ticketing database; and in doing so, incidents of fraud reported by site users decreased by as much as 90%. R Enterprise was able to deliver such impressive results because it has the ability to consider patterns and trends a human charged with completing the same task might never consider...or even have time to consider.

Retailers aren't the only ones to benefit from big data. Recently the United States' National Science Foundation, and Japan's Science and Technology Agency have begun collaborating on efforts to improve how the two nations manage disaster response efforts. In such cases, real time information is important because it helps responding agencies accurately assess a situation and formulate an appropriate response. This has become increasingly difficult as data from unrelated sources and of different types have begun to flood in from social media, smart phones, scientific instruments, and more. Researchers from Temple University and the University of Aizu have begun developing smartphone-centric emergency networks with the ability to evolve as disasters do. Timely data capture capabilities and complex algorithms will be invaluable in helping organisations sort through the chaos so they can respond more effectively.

Experts suggest that over the next decade, the need for new Big Data solutions, and those equipped to manage them will increase at least tenfold. For all the many benefits, Big Data also faces some challenges. In addition to data capture, transfer, warehousing, curation, processing and other functions, the difficulty of integrating existing analytical environments with new, Big Data deployments should not be underestimated. In addition, managing such vast quantities of information can make security and privacy more difficult to ensure.

So how do you determine when it is time to make the move to Big Data?



Your IT Executive in the Cloud

---

Your business must be in a position that will benefit from the increased insights derived from massive data sets. As your business grows, there is an increased likelihood that it will take you longer and be less efficient to rapidly draw insights from the data you gather using your current capabilities. When this happens, finding an automated alternative will become essential as it will become simply impossible to transform your data into timely value. Fail to adapt and you will find yourself overwhelmed and overtaken by your forward-thinking competitors.

Though much press surrounds the software, systems and technology pushed by data storage vendors associated with implementing Big Data solutions, we seldom hear of the necessary change agenda required to leverage the full extent of these volumes of data. Without embracing the full extent of governance, change management and structural changes required to harness an information-based organisational capability, BI and Big Data solutions will typically remain underutilised.

It is not a matter whether Big Data and BI solutions can coexist, but rather how best to transition these technologies from simplistic reporting-centric systems into an effective integrated information platform that delivers responsive answers to inform mission critical business questions.

#### References:

A Comprehensive List of Big Data Statistics. Analytics. 1 August 2012. Retrieved from: <http://wikibon.org/blog/big-data-statistics/>

Freifeld, Karen. StubHub Says it was the Victim of Massive Cyber Fraud. Reuters. 23 July 2014. Retrieved from [http://www.huffingtonpost.com/2014/07/23/stubhub-cyber-fraud\\_n\\_5611953.html](http://www.huffingtonpost.com/2014/07/23/stubhub-cyber-fraud_n_5611953.html)

Laskowski, Nicole. Ten Big Data Case Studies in a Nutshell. Tech Target. October 2013. Retrieved from <http://searchcio.techtarget.com/opinion/Ten-big-data-case-studies-in-a-nutshell>

StubHub Taps into Big Data for Insight into Millions of Customers' Ticket-Buying Patterns, Fraud Detection, and Optimized Ticket Prices. Oracle. Retrieved from <http://www.oracle.com/us/corporate/customers/customersearch/stubhub-1-big-data-ss-2244375.html>

R language: [http://en.wikipedia.org/wiki/R\\_%28programming\\_language%29.html](http://en.wikipedia.org/wiki/R_%28programming_language%29.html)

U.S., Japan Bring Big Data and Data Analytics to Disaster Response. Scientific Computing. National Science Foundation. 31 March 2015. Retrieved from <http://www.scientificcomputing.com/news/2015/03/us-japan-bring-big-data-and-data-analytics-disaster-response>



Your IT Executive in the Cloud

---



*David Gee is managing director at G3 Services. He has previously held senior IT leadership positions for Boeing Defence Australia and QSuper; and is a founding member of the CIO Executive Council in Australia.*

*Matthias Fetz is managing director at DataMovers. He has successfully delivered business intelligence solutions for international companies in the finance, mining and retail sectors.*