Image credit: Charis Tsevis

Is big data "bullshit"?

Every few years a new buzz word makes it on to the scene and it can be difficult to separate the hype from the true need. Currently 'big data' holds the title of 'most over-used term' and business leaders that don't have a big data project on the go just aren't up to date with current IT trends. Or are they? Anthony Doesburg *sifts through the big data brouhaha and considers the reality behind the buzz...*
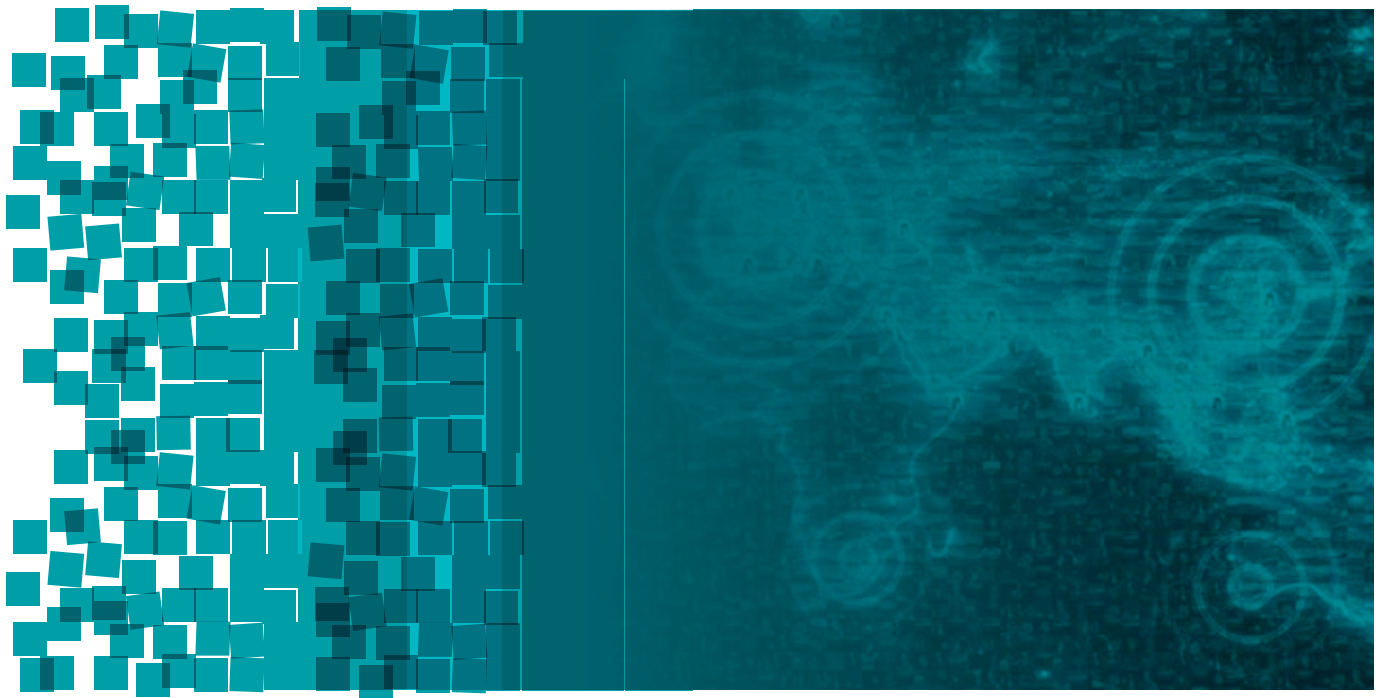
For good or ill, big data is getting bigger. From President Obama's re-election campaign and US retailer Target's identification of pregnant shoppers to whistleblower Edward Snowden disclosing that the National Security Agency has been bugging our communications, examples of big data collection and analytics are everywhere.

IT vendors that until now hadn't been known as purveyors of big data products are joining the bandwagon. In June Hewlett-Packard announced a newly minted big data consulting practice and at Microsoft's July Worldwide Partner Conference in Houston boss Steve Ballmer said big data was one of four focuses of a reorganised company.

The drip-fed revelations by Snowden, an ex-NSA contractor, haven't made him popular with Obama and have angered US allies such as Germany whose citizens have been caught in the NSA net. As noted by a columnist in *The Guardian* newspaper, which was one of the first to report Snowden's disclosures, the leaks also highlight the role of algorithms — or analytical software — for finding gold among the dross.

To IT practitioners who've been bombarded by big data hype for years already, none of this will be news. If uncovering a terrorist plot is the nugget sought by the NSA, in the world of business intelligence the pay-off comes in the form of intersecting lines on a chart that might reveal a previously unsuspected customer appetite or the solution to a population or health problem.

The question, said Radio New Zealand broadcaster Kim Hill during a series on big data earlier this year, is whether the phenomenon represents manipulation or empowerment. Target, perhaps spooked by that line of enquiry – it was once confronted by a man whose school-aged daughter was being mailed baby-related bumf by the chain; he had no idea she was expecting – stopped co-operating with the *New York Times* reporter who broke the story about the company's use of data analysis to identify pregnant customers.

Hill pointed out that big data analytics has completely taken over: in 1986, only six percent of the sum total of world data was digital; today, nearly all the written word, music, images and data is in binary form.

She's in good company with that observation: according to Gartner principal research analyst Dan Sommer, it's fast becoming a world of "analytics everywhere". Speaking ahead of a Gartner business intelligence summit in Mumbai in June, Sommer said that by next year analytics will be in the hands of half of potential users, rising to three-quarters by 2020.

"Post 2020 we'll be heading toward 100 percent of potential users and into the realms of the internet of everything."

Cutting through the hype

Yet there are other voices – and research – suggesting big data is just big noise. Although the Obama campaign's use of big data in getting him re-elected last year has been widely trumpeted – campaign workers were apparently able to identify all 69,456,897 Americans who voted for him in 2008 – campaign CTO Harper Reed told a CeBIT audience in Sydney in May that big data was "bullshit".

He wasn't totally dismissive, however. *ITWire* reported Reed qualified his remark by saying data itself is less important than what you do with it.
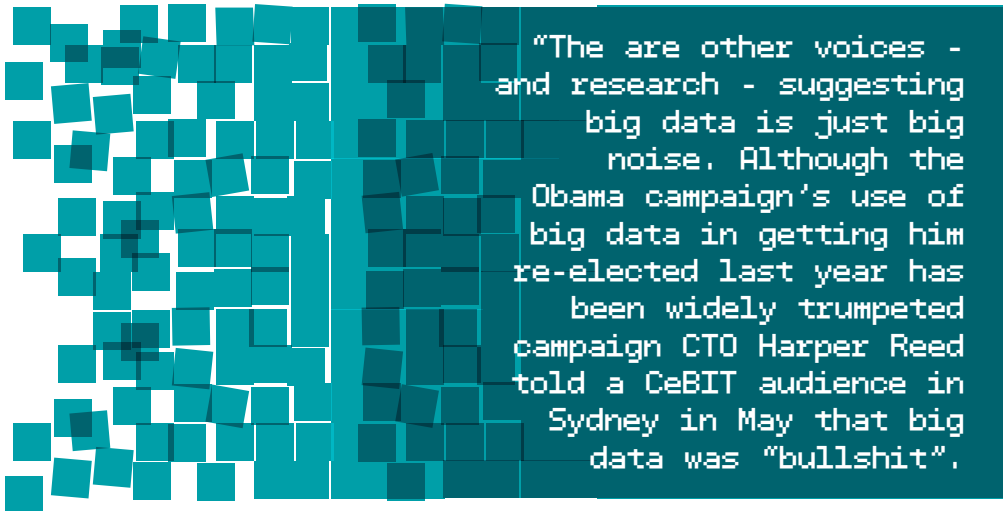
Analyst IDC and customer behaviour researcher Fifth Quadrant have also come out with findings that temper the big data story. An IDC survey this year of 300 Australian organisations found big data analytics is certainly "getting SMB's tongues wagging", (Vern Hue, senior market analyst for Australia and New Zealand, commenting on the survey result).

But his colleague and head of big data research in the two countries, Shayum Rahim, says although Australian enterprises are well on their way in the "big data journey", they are still at a relatively early phase in resolving many of its issues.

In a smaller survey of 64 Australian organisations Fifth Quadrant found customers of less than half the sample were enjoying improved experiences as a result of big data analytics. The author of the Sydney researcher's *'Big Data or Big Hype'* report, Chris Kirby, says only an eighth of the sample rate their analytics activities as "extremely successful", yet three-quarters use big data analytics.

"The greatest success appears to come when organisations adopt an integrated customer analytics strategy that puts quality data in the hands of decision-makers and leadership from executive teams is critical," Kirby says.

IDC found more than 80 percent of organisations have either deployed or have plans to launch big data analytics in the next 12 months, which Rahim says illustrates that Australian organisations' efforts range in maturity »

> "The are other voices – and research – suggesting big data is just big noise. Although the Obama campaign's use of big data in getting him re-elected last year has been widely trumpeted campaign CTO Harper Reed told a CeBIT audience in Sydney in May that big data was "bullshit".

from ad-hoc and experimental discovery to advanced analytical capability that drives decision-making.

"Technology vendors and service providers will really need to think about where in this maturity model they will engage customers. Do they go with the low-hanging fruit and target customers with budgeted projects, or do they expand their pipeline and work with those customers who have just begun their business analytics journey?"

What's more, he says, organisations are being very strategic in the use of data analytics to deliver specific line-of-business outcomes. "We spoke to senior IT executives who acknowledged a shift in focus from IT priorities to LOB requirements when it came to big data initiatives."

He sees a warning there for CIOs: they will need to be more closely linked to lines of business and their specific goals to remain relevant "in rapidly changing organisational landscapes".

Entering the mainstream

If there are obstacles to the adoption of business intelligence and analytics, ethical questions such as Hill hints at seemingly aren't one of them. Gartner's Sommer said ease of use, performance and relevance are the sticking points, but disruptive technologies such as Facebook and web browsers are helping overcome them.

And usability is being pushed to new levels by what Gartner calls interactive visualisation or data discovery, the tools which enable data mash-ups. This segment is growing three times faster than traditional business intelligence front-ends, with Gartner expecting the market to be worth US$1 billion by the end of next year, dominated by vendors such as QlikTech, Spotfire and Tableau.

"However, MicroStrategy, IBM, Microsoft and SAS have all launched rivalling products in the past year, propelling that entire segment into a newer, much more competitive phase," Sommer

said. "What this means is that data discovery has arrived as a mainstream architecture."

Cynics might say the subject is being over-hyped but one thing is irrefutable: data is piling up at an enormous rate.

Gems amongst the dross

Big data is classified as structured or unstructured, the former representing about 20 percent of the total. The unstructured 80 percent, which encompasses everything from machine and sensor data to contact centre logs, is growing three times faster than structured data. Lurking amongst the vast amounts of unstructured data is what is known as dark data – data organisations might not even know they have.

Every enterprise action from a tweet to a customer paying an invoice or making a purchase makes the pile bigger. Much of it goes unexamined, yet it can have real value. A customer transaction that might not immediately result in a purchase, for instance, could be an opportunity to create awareness of a product they might ultimately buy.

Or, as cosmetics company Elizabeth Arden discovered, there could be significant savings hidden in the heap waiting to be revealed by big data analytics' all-seeing eye. Using QlikView data visualisation software, the New Zealand subsidiary of the US cosmetics maker found that a couple of its retailers were chewing through product testers at a much greater rate than others.

Visibility of business data went "from night to day" when Elizabeth Arden began using QlikView in place of manually entering figures from a green-screen ERP system into a spreadsheet, says finance chief Tony Goddard.

"It was hopeless," says Goddard of the company's pre-QlikView business intelligence efforts. "If we noticed a particular sales trend it was very slow trying to go back to the original data to try to interpret it."

QlikView, in contrast, makes it easy for users to analyse sales results for particular products and customers over any period.

"We can see how much we sold to the customer at what discount, the cost of goods and the gross margin. We can get into profitability by customer, by brand and by SKU."

Goddard says the marketing team love it because it gives them an answer when he presses them on where their budget goes.

"They're able to say, 'we got this result from that spending'," he explains.

The sheer volume of data, however, is holding organisations back from making sense of it. By 2015, says HP, organisations with more than 1000 employees in industries such as banking, manufacturing and communications will have an average of 14.6 petabytes (million gigabytes) of data.

Coping with your data

The pioneer users of business intelligence tools were financial services organisations and Fifth Quadrant' study suggests they are also the leaders in big data analytics – more than a quarter of respondents to its survey were banks, insurance companies or finance firms. But HP says interest in analytics is spiking among business-to-consumer and particularly telecommunications companies and airlines.

Aside from drowning in large volumes of data, other words starting with 'v' – velocity and variety – neatly sum up additional challenges faced by big data newbies. When it is considered that data comes from sources as diverse as the web, social media, partners, customers and the factory floor, arriving at a rate of a terabyte an hour from a single industrial machine, the potential difficulties can be glimpsed.

The first step in helping customers grapple with their mounting data stores is to understand what problem the customer is trying to solve or the objective he is trying to achieve.

Glen Rabie, head of Melbourne-based data visualisation tool company Yellowfin, says the ideal is that customers will have worked out how to store and manage their data before contemplating how to slice and dice it. But the reality is many come at the problem the other way around.

"Often people buy backwards. They'll look at what we do and say 'that's fantastic — I want those lovely shiny charts', but they don't really think about the structure their data is in. They're not necessarily ready yet to put a presentation layer in front of it."

Rabie says Yellowfin will offer advice about the plethora of possibilities, but the choices aren't simple, either for large or small to medium-sized operations.

"When thinking about SME solutions, the

»

> "A fundamental question for an organisation undertaking a big data or business intelligence project is 'what are its objectives?'. It can't expect a return on the investment, for instance, if it is not prepared to act based on what the analytics brings to light, if necessary changing the way it operates."
>
> Glen Rabie, head of Yellowfin

market is actually quite complex. There are a lot of different products and solutions so it can be difficult for people to work out what's appropriate in terms of infrastructure."

Standards in the big data world are ill-defined, in contrast to the traditional database market in which SQL and relational databases won the day, and it can be hard to validate the claims made for particular products, he says.

"Every product will tell you it does all things, as opposed to understanding what your needs are and easily being able to find the right match."

From a visualisation perspective, the key is that data access not be a bottleneck.

"If you have to wait 30 minutes for a query to run then no tool on the planet is going to be appropriate because you're going to get very frustrated every time you want to change a chart," Rabie says.

"So we help our customers to think about what is the most appropriate place to put their data, should it be SQL Server, or an analytical database like Vectorwise or Netezza, or should it be a Hadoop or MongoDB solution. If they've already solved those issues, that's fantastic."

Studies in success

"An even more fundamental question for an organisation undertaking a big data or business intelligence project is 'what are its objectives?'. It can't expect a return on the investment, for instance, if it is not prepared to act based on what the analytics brings to light, if necessary changing the way it operates," says Rabie.

At Elizabeth Arden, that was easy. When it found some of its 130 retailers were using unjustifiably large quantities of testers, which represent a significant cost to the company, it asked the sales reps looking after those accounts to find out what was going on. "Once they knew we had this visibility, suddenly their demand on our testers went down and we're saving money," says Goddard.

At Yellowfin customer Macquarie University in Sydney, data visualisation led to big, and financially beneficial, changes to grant applications and its supply chain, Rabie says.

Savings aside, an intimate understanding of their data can be the only advantage available to organisations in highly competitive markets. If it helps them keep customers close, that's much cheaper than acquiring new ones.

How else than by poring over sales data, for instance, might Australia's largest department store group Myer have found that customers at certain rural department stores who bought beds also bought chainsaws?

Mark Fazackerley, Australia and New Zealand head of MicroStrategy, says when the unlikely association between the two product lines was discovered, Myer rearranged its store layout accordingly. The theory was that holiday-home owners were trimming trees and replacing beds on the same cycle.

"When you consider the millions of point of sale transactions Myer handles, to sift through and find that correlation is quite an achievement," Fazackerley says.

Similarly, New Zealand electrical goods wholesaler JA Russell used MicroStrategy to uncover slow-moving stock in its warehouse. "It was able to adjust its forward ordering and pricing to try to alleviate that."

The trend among business intelligence users is towards "democratising" data collection and analysis, says Fazackerley, away from "guys in lab coats" who generated reports only they understood to giving staff the ability to find "information that matters in a time-frame that allows them to act on it and generate results".

The trouble is democracy, as Obama might say, can be messy: putting BI in the hands of the masses could be a potential IT management nightmare. However, that's not how it worked out at real estate agent Ray White.

Brisbane Tableau consultant and former Ray White analyst Nathan Krisanski says the firm opened up its data to 1000 franchise owners and 10,000 sales people without them even knowing what tool they were using.

"Tableau gave us the ability to push data to our people in a format that made sense to them and required little interaction to obtain basic information. But it also provided end users with transparency and deep-dive analysis if they wanted to understand more."

And it was done in a controlled environment, Krisanski says.

First things first, though, says Yellowfin's Rabie, who warns that big data advocates shouldn't go off half-cocked.

"They need to know why they're doing it and if they find out interesting things be in a position to do something about it.

"If they don't have their ducks in a row, they're just wasting money."