



Big data" barely needs an introduction. Big data, we're told, is the answer to everything; it's the ultimate solution to the engineer's dilemma that "what cannot be measured cannot be improved". Networked sensors and reporting tools in everything, everywhere, gives us unprecedented abilities to monitor the world in realtime. And the data that produces is big.

But even while many South African firms are still coming to terms with the concept of big data, the global conversation has already moved on. There's an understanding that the future isn't about big data – it's about what comes next.

Data by itself isn't a solution; too much data can be as paralysing and uninformative as too little, the informational hay that buries the needle of useful knowledge. It's not about the data; it's about how you make sense of it.

SO WHAT EXACTLY IS BIG DATA?

Not all data is big, says Jeff Fletcher of Internet Solutions' Research and Innovation team. "Microsoft SQL has been the business standard

DID YOU KNOW?

The Michael J Fox foundation is investigating new ways of diagnosing and treating Parkinson's Disease, using vast amounts of data captured from sensors worn by sufferers.

for databases. The top end SQL server can store roughly five petabytes, which is really a big amount of data in the grander scheme of things. However, it is impractical for anyone to have that amount of data in a SQL server because of the costs involved in the licensing and the size of the servers which would be required.

"While it is still big in terms of what most people consider to be big data, it is tiny compared to what Google or the Large Hadron Collider at CERN would have. What most businesses are going to encounter will be less than that, and they want analysing capability at a reasonable price."

A general rule of thumb is that if it fits on anything from an Excel spreadsheet to a 4TB

hard drive, it's not considered a "big data". At the 4TB or 5TB breakpoint, it is necessary to start thinking about how to process it. The data will need to be broken into parallel parts in order to process it within any reasonable time.

The rise of big data brings with it technology challenges that companies are still addressing. Running database queries over "big" databases on a single server would take an insurmountable amount of time. So data has to be split into chunks that can be processed in parallel. This in turn brings other challenges; with the location-agnostic nature of cloud computing, data may have to be moved long distances, fast. Metrofibre, a fibre-optic networking company in Gauteng, recently began offering a 100Gbps, low-latency >

APACHE'S HADOOP IS POSSIBLY THE MOST IMPORTANT NAME IN BIG DATA RIGHT NOW. THE SOFTWARE FRAMEWORK IS DESIGNED FOR DISTRIBUTING COMPUTING TASKS OVER LARGE NUMBERS OF MACHINES TO SPEED UP DATA PROCESSING IN A RELIABLE WAY.



DATA ANALYSIS



THE EU IS SO CONCERNED ABOUT THE COMPETITIVE ADVANTAGES OF COMPANIES WITH THE CAPACITY TO ANALYSE BIG DATA THAT IT HAS BEGUN INVESTIGATING WHETHER OR NOT IT SHOULD BE TREATED AS AN ANTI-TRUST ISSUE AT SOME POINT IN THE FUTURE.

product across its network specifically to enable these kinds of big data transactions.

BIG DATA OR NOT BIG DATA?

Given the amount of investment required to properly enter the world of big data, it's surprising that so many firms seem unprepared about what to do with it.

In 2013, the Harvard Business Review was one of the first to pick up on this. "The biggest reason that investments in big data fail to pay off," it said, "is that most companies don't do a good job with the information they already have. They don't know how to manage it, analyse it in ways that enhance their understanding, and then make changes in response to new insights."

Peter Sondergaard, a senior vice-president at Gartner, is of the same opinion: "[Big data] will not be transformative, as anybody can store it. Good analysis of big data will be worth a bit more, but data is inherently dumb."

Firms which are gathering big data must determine how they will use it to make business decisions, and the process of making data usable in order to base decisions on it has been coined "decision engineering". Companies are faced with finding the correct analytic talent to implement this decision engineering.

LOOKING TO THE FUTURE

For that, the field of predictive analytics is coming to the fore. Internet Solutions' Fletcher says this is exactly what his company is

BIG DATA WILL NOT BE TRANSFORMATIVE, BECAUSE DATA IS INHERENTLY DUMB...

looking at now. As one of the major providers of internet services, it has access to incredible amounts of information, and it wants to learn how to exploit it.

"The aggregation of our network capacity utilisation data allows us to better determine when we need to upgrade something," says Fletcher. "When we need to upgrade a transit or peering link, the longer we wait to upgrade, the better it is from a cost perspective. But the later we wait to upgrade, the worse it could be for our customers."

"Now we can more accurately predict the rate of change of utilisation and upgrade link

WHERE IS THE TALENT?

Specific skills and competencies are required to enable a business to find data-driven solutions, and firms such as Gauteng's Tarsus Technologies have hired in-house data scientists to begin learning how they can impact their own decision-making and that of their customers.

Data scientists are specialists in statistical modelling, natural language processing, computer science and machine learning skills, and use these skills to garner meaning from large amounts of data. They develop advanced predictive algorithms and self-learning models that extract actionable information and highly accurate predictions from these big data sets. As Internet Solutions' Fletcher says: "To make a decision, I need information, not data."

The problem is that data scientists are still a rare breed, and good ones are expensive. Analysts at McKinsey are predicting demand for data scientists will be 60% higher than supply by 2018. So if you're looking for a career change or trying to encourage the kids into a career that's challenging and well paid, you know where to start.

capacity at exactly the right time. It's cost saving that can add to the bottom line and maintains a consistent customer experience."

These are, however, fields that might still see some failure before they see great success. Just because more companies have access to toolkits only universities and Google could afford doesn't mean that big data is easy.

Even so, says Wesley Govender, a director at Deloitte South Africa, firms that aren't starting to think harder about how they implement big data strategies are asking for trouble.

"Companies need to get smarter with how they harness the immense benefits of these changes, or be left behind by those that do," Govender says. **ICT**

DON'T COLLECT IT ALL

MANY FIRMS RECOMMEND THAT YOU ARE MORE CAREFUL ABOUT WHAT YOU COLLECT. "IT'S BEST TO HAVE MORE EFFICIENT USE OF DATA THAN MORE DATA," SAYS DELOITTE SOUTH AFRICA'S WESLEY GOVENDER. IT'S WORTH REMEMBERING DATING SITE ASHLEY MADISON WAS ATTACKED PRIMARILY FOR HORDING USER DATA.