

Assessing the big data challenge

The business of data analysis and collection

Consider whether your company needs new methods to conquer mountains of data

By **Jacob Stoller**

IT TRENDS ARE OFTEN PRESENTED

by the industry as threats, and what is being dubbed “big data” is a classic example. The sheer volume and complexity of data that organizations are collecting, the argument goes, is overwhelming existing IT systems, and requires the adoption of a new generation of tools. Those who fail to adopt such solutions risk falling behind and putting themselves at a competitive disadvantage.

Nobody would dispute that data stores are growing dramatically. Companies everywhere are seeing increases of 20 per cent annually. IT market intelligence firm IDC reported that in the second quarter of 2011, storage companies shipped 31 per cent more disk capacity than they had shipped in the same period in 2010.

While such growth may seem alarming, it shouldn't be cause for panic, says Barry Cousins, senior research analyst for Info-Tech Research in London, Ont. “People have run out of capacity for years,” he says. “This is not new.”

Cousins also points out that traditional database tools — such as SQL, Oracle and DB2 — are “rapidly expanding the scope of what they can deal with” and that improvements will help most organizations keep up with their data growth.

Complexity

Big data, however, isn't just about size. Organizations need to look at the

growth drivers and the kinds of data they're accumulating.

Two factors in particular should be watched carefully. Because data is being collected from more sources (notably smartphones and other wireless devices), systems are seeing more diversity of formats and data types. Furthermore, the collection of this data is automated, making growth difficult to control.

In addition, organizations are seeing a growing proportion of unstructured data, which can be loosely described as data that doesn't fit neatly into a pre-defined database. For example, when a customer fills in a name and an address on an online form, the information subsequently collected is structured data. If the input is an email or a written paragraph from a Facebook page, the information is unstructured data.

Unstructured data often can't be analyzed with traditional database tools, yet there's an increasing interest in it as decision makers look beyond traditional sources to better understand their business environments.

“Unstructured data is of more value than it used to be because there are tools that are emerging that help to organize and structure it,” says Ian Anderson, senior manager of Information Management Shared Services at TD Bank.

Tougher questions

Janet Pierce, vice-president of Professional Programs for CMA Ontario, points out that the real driver



Ask three essential questions to determine if big data is the answer.

behind the growth of data is the questions analysts are expected to ask and answer.

Questions about “grey area” issues such as pricing are particularly demanding. When is the optimum time for a retail chain to lower its prices? What is the appropriate response to fuel price fluctuations for a shipping company? “It was never easy to answer some of these questions,” says Pierce. “The answer was always ‘it depends.’”

Often, these questions are driven by competitive practices within a particular industry, compelling organizations to collect and analyze more information just to stay in the running. “The banking industry is always trying to understand its customers better by having a deeper knowledge of their preferences and behaviours,” says Anderson.

The real question here is whether the answers that analysts seek from data are worth the cost of a big data solution or, for that matter, any IT solution. “It’s the age-old question: how do you build a business case for an IT investment?” says Pierce. The focus, she suggests, should be on identifying the areas where decision makers are handicapped by a lack of information.

Establish what’s needed

A variety of technologies are being positioned as big data solutions, but the most common solution is software that allows multiple processing and storage devices to work simultaneously to execute large data-crunching chores.

Getting a system like this one working can require substantial upgrades to an organization’s IT infrastructure. Cloud solutions may make this transition easier, but this technology is not for the faint-hearted. Asking three essential questions will help you determine if big data is the answer for your organization.

1. Do you have a big data problem or an IT infrastructure weakness?

Many of the problems that seem to point to a big data solution are just weaknesses in an existing IT infrastructure.

“Make sure you’ve got a big data problem before you start spending time there,” says Cousins. “The strategy is to look inward and get clear on the actual limitations posed by your own data in all areas: storage, processing, moving, analyzing, reporting, backup-restore. I think the majority will find that they’ve just got a classic data problem.”

2. Are you missing critical information that a big data solution will help you capture?

On the analytics side, people should be equally clear about the most critical questions they’re asking, and whether an inability to answer them could spell trouble down the road. The starting point, Pierce suggests, should be industry risk factors such as fraud in the banking industry, oil prices in transportation or price optimization in retail.

3. If you need a big data solution, what are your analytical requirements?

“Anybody who has got an enterprise management system has already done the prioritizing of what the risks are,” says Pierce. “They now need to tie in their analytics requirements with what they’ve determined are the key risks in their organization.”

As with so many IT phenomena, the primary concerns about big data are issues that organizations have been dealing with for some time. Those who have defined their information requirements clearly and are getting access to the data they need from IT have nothing to fear from big data. But those who have fallen behind will face a steeper climb if and when they decide it’s time to catch up. ■

Jacob Stoller is a Toronto-based writer and researcher.