

COLUMNS

Big problems in Big Data

April 1, 2013



By **John Curtis**

Without the right precautions, Big Data can be a lot like sweeping a mess under the rug instead of cleaning it; it will eventually build up, creating more problems.

With data management, I like to think about things in terms of "what happens if " It helps me know what is missing from a solutions perspective, and uncovers interesting problems and lacking lifecycles in trends.

For example: what happens if we give senior citizens two terabyte hard drives; what happens if we give mobile devices the ability to store all of our precious moments (pictures and video) without a backup plan; and **what happens if ad targeting is so good it knows you are pregnant before you do?**

Managing the flood

The problem we have now in Big Data and data warehousing is that we are introducing a lot more of the same stuff people have been working with for years. It seems familiar to them but it is very different in terms of size and speed. Managing a glass of water is very different from 100 million gallons of water but it is still just water, right?

Tools are emerging that are new to most people tasked with managing excessive amounts of data. They are better suited to using the cloud, really distributed resources and virtualization, and can handle the needs of a big data collection system.

The thing we are optimizing for is the number of items being collected. It is no longer just about the data submitted on a form, but about how long the user spent on each field of the form and at which point they abandoned their input.

We now have mobile applications for consumers and for a mobile enterprise that have changed the timing around how data is reported.

In a non-mobile environment, we knew when and where data was being accessed. Through mobile, we have people interacting and creating data, from anywhere and far outside of the planned business hours of the past.

Organizations are currently either logging all this data to look at it later sweeping the mess under the rug or leveraging tools and people to manage it. The camp we have to worry about is the former.

What started as a few cups of water will soon be 100 million gallons of it. Two big problems could exist: the data is stored in a format that will require a lot of heavy lifting for it to be usable, or the data will be missing key pieces or attributes that make it meaningful.

Missing the picture

The purpose of gathering data is to use it to make informed decisions that will keep your business a step ahead of the competition. That is why having access to, and properly managing, data is imperative to a successful campaign.

The worst thing an organization can do with its Big Data campaign is to waste its effectiveness through sloppiness

or mismanagement.

Here is an example: A company wanted to log data from users of its mobile app but it wanted to make registration optional. By not making registration manual and mandatory, the entire exercise is inaccurate.

Instead of learning about the behavior of app users, this company could only make decisions based on the behavior of users that voluntarily chose to register with an app which hardly represents all users. The data that is not captured throws the whole analysis askew and dilutes the effectiveness of any decisions based on those analytics.

The previous example of mismanaged data represents a sloppy mistake that can undermine the whole Big Data campaign.

Without researching what storage amount is necessary for a specific data campaign, or monitoring data storage activity, an organization can run out of drive space and miss data that is critical to proper analysis.

A basic logger storing analytics on one ad can take down an organization's entire portal without foresight.

LEVERAGING BIG DATA comes down to properly managing and organizing the deluge of information received.

Whether companies are sloppy and underestimate the drive space needed to store data or they choose to omit certain information, incomplete data storage results in diluted analysis or weak decisions.

If you are going to do Big Data, then do it right. Investing in data warehousing on the front end is the best way to make informed decisions that propel your company forward.

John Curtis is founder/CEO of [Quotient Solutions](http://QuotientSolutions.com), Austin, TX. Reach him at john.curtis@quotient.net.