Big data becomes big intelligence

By Bryan Balfe   25 Jul 2014

Big data has become a reality; however it is not the same reality for every company or every user.

The explosion of data is creating different problems and opportunities. For example, the medical provider required to store scanned images for each patient's lifetime faces a very different challenge to the FMCG brand now offered an unprecedented depth of customer purchasing-behaviour data. The end-user despairing over the time taken to locate a file or email has a different set of challenges to the legal team struggling with new, big data-inspired compliance demands.

According to Gartner, a survey of 720 companies asked about their plans to invest in big data gathering and analysis revealed that almost two-thirds are funding projects or plan to this year, with media/communications and banking firms leading the way. The research firm insisted that 2013 was the year of experimentation and early deployment for big data. Adoption was still at the early stages with less than 8% of all respondents indicating their organisation has deployed big data solutions. Twenty percent were piloting and experimenting, 18% developing a strategy, 19% were knowledge gathering, while the remainder had no plans or didn't know.

This is, therefore, a critical phase in the big data evolution. While storage costs have come down in recent years, organisations cannot possibly take a "store everything" approach to big data and hope to realise the full long-term benefit. The issue is not only what data to retain and where, but how to extract value from that data - not just now but in the future as big data technologies, including analytics, become increasingly sophisticated.

Significant management challenges

In addition to the huge expansion in data volumes, organisations also now have access to new content types. While this depth of data offers exciting opportunities to gain commercial value, it also creates significant management challenges. How should the business protect, organise and access this diverse yet critical information that increasingly includes not only emails and documents, but also rich media files and huge repositories of transaction-level data?

At the heart of a successful big data strategy is the ability to manage the diverse retention and access requirements associated with both different data sources and end-user groups. While today a large portion of the data in a typical enterprise does not get regularly accessed for a year or more, this is definitely set to increase as big data strategies evolve. Many organisations are gleefully embarking upon a "collect everything" policy on the basis that storage is cheap and the data will have long-term value.

Certainly, inexpensive cloud-based storage is enabling big data strategies, but the reality is that while it is feasible to store

all the data in the cloud, even with fast connections retrieving that 5Tb of data from the cloud back into the organisation would take an unfeasibly long time. Furthermore, cloud costs are increasing, especially as organisations add more data; and even cheaper outsourced tape back-up options still incur escalating power and IT management costs.

In addition, the impact of unused data sitting on primary storage extends far beyond higher back-up costs; time-consuming end-user access leads to operational inefficiency and raises the risk of non-compliance.

Take a far more intelligent approach

Organisations cannot take a short-term approach to managing the volumes of big data and hope to realise long-term benefits. There is a clear need to take a far more intelligent approach to how, where and what data is stored. Is it really practical to take a back-up of an entire file server simply because some of the documents need to be retained for several years to meet compliance requirements? Or is there a better way that extracts the relevant information and stores that in a cheaper location, such as the cloud?

To retain information and avoid a cataclysmic explosion in data volumes, organisations need to take a far more strategic approach to data archive and back-up. What information must be kept on expensive local storage and what can be sent to the cloud or another location? And what policies will be put in place to take data ownership away from end-user control? By taking a strategic approach to archiving data, based on the property of each data object, organisations avoid the problems caused by end-users applying their own "retain everything" policies.

By deleting the local data source and moving it to a virtual data repository, an organisation avoids duplication and inconsistency whilst still ensuring information can be retrieved in a timely and simple fashion. Policy-driven rules for data retention can be based on criteria such as file name, type user, keyword, tagging or Exchange classifications, while tiering can be applied based on content rules to any target, including tape or cloud.

This intelligent retention model needs to be backed up by effective data retrieval. The key to this process is context indexing, which enables end-users to apply simple key word search to access any data. Organisations have the option to context index either live data or secondary data, in back-up or archive. In both cases, rather than context index the entire data resource, by applying the right filters and policies organisations can also prioritise the most valuable and frequently accessed data sources. Context indexing critical corporate data in this way ensures the business always has the option to access and retrieve information rapidly.

Combining intelligent storage policies with content indexing reduces data volumes, enables organisations to use the most appropriate storage media for each data object and facilitates rapid access to business critical information.

Growing pressure on IT to deliver

Demands from individuals to explore and exploit big data will put growing pressure on IT to deliver more than additional storage resources. What happens when it takes the CEO over 15 minutes to find and access an essential document? Or when the legal team cannot retrieve vital information to prove compliance? Or when the brand manager cannot exploit expensive retailer data and analytics investment to understand customer behaviour?

The key to transforming big data into big intelligence is content and context. By managing big data retention and storage based on content and its inherent value to the business, organisations will be well placed to harness this data not only to address immediate problems, but also to improve strategic insight. From predicting demand for new products and services to transforming the speed with which every end-user can retrieve corporate documents, it is those organisations that consider retention strategies from day one that will be best placed to realise the big data vision.

ABOUT BRYAN BALFE

Channel Manager at CommVault in South Africa

View my profile and articles...

For more, visit: https://www.bizcommunity.com

View my profile and articles...