

The perils of not knowing your data

The exponential growth of data is creating a data management headache for large organisations which are increasingly struggling to keep track of not only how much data they have, but also where exactly it is stored.

 By [Johan Scheepers](#) 17 Mar 2020



Johan Scheepers, Country Head at Commvault South Africa

IDC released a study that says that the collective sum of worldwide data will grow by 61% – from 33 zettabytes to 175 zettabytes – by 2025, with as much of the data residing in the cloud as in data centres.

To put this another way, big data growth projections by Domo states that by 2020 every person on the planet should generate 146,880 gigabytes a day, which – taken with population growth – makes it easy to conclude the amount of data we'll create on a daily basis will rise dramatically.

Amid this data explosion, the trend among large corporations is that they are increasingly losing track of what they have and where in terms of data, and there is no doubt that this is becoming a massive issue for companies to manage.



DATA & ANALYTICS

How to be 'data ready' in 2020

Johan Scheepers 4 Feb 2020

The main drivers of exponential data growth include the proliferation of technologies such as the Internet of Things (IoT), which is causing all these huge data feeds and unstructured data changes, while other drivers include multi-cloud adoption, Software-as-a-Service, Machine-to-Machine technologies and tools such as in a virtual environment.

If it isn't already a big issue for some of the large enterprises, it's going to become one soon. The impact of not knowing how much data an organisation has can have severe consequences. From an operational side it obviously has an impact on capacity planning, but the most critical issue that not knowing how much data you have, where it resides and how important it is to your organisation is that it can lead to non-compliance.

Most organisations face this problem, as in some form or other they must either comply with one of the international laws, such as General Data Protection Regulation (GDPR), or local legislation, such as Protection of Personal Information (PoPI) Act, when it eventually comes into effect, or even the Protection of Privacy Act. So, an organisation has data out there that it's not aware of, then non-compliance becomes its biggest challenge.

At the same time, storing data that is harmless to your company, after the required retention period, is another challenge and opens you up to massive data growth that becomes uncontrollable and impacts capacity planning.

Manage your data

Not knowing how much data you have can also significantly raise the cost of trying to manage it, especially in a cloud environment, where storage infrastructure becomes a bottomless pit. Typically, the total cost of ownership and return on investment will increase drastically if you don't keep a handle on your data.

Therefore, you need to start classifying what data is important to the organisation, deciding what the impact of not having this data would be and how long it needs to be kept. A lot of organisations keep everything forever, as opposed to going through a classification process that would give them insight into their data and inform them what they can get rid of.

It is important that enterprises develop good data management practice, which means assigning an appropriate business value to a document at inception and keeping it through its entire lifecycle to deletion. Hence, a good data management strategy will allow you to determine when certain data is no longer needed and you can simply get rid of it.

It is key that organisations work with a vendor who can assist with data management and recovery, by providing insight into their data and help them develop a data management strategy. A vendor can provide technology that will examine your data, index it and give you an understanding of what you have and where it resides.

ABOUT JOHAN SCHEEPERS

Johan Scheepers is Commvault systems engineering director for MESAT
[View my profile and articles...](#)