
Managing Big Data

It's a question of scale

By Andrew Buss, Freeform Dynamics Ltd, March 2012

Originally published on



Way back in the mists of time, when computers were first developed as servers they used specialised external storage devices – often large tape drives or hard disks – that attached through a proprietary interface. As time went by, the computers shrunk in size but increased in capability. The storage shrunk in size even more – allowing disks to be directly attached within the server.

The development of industry-standard servers built using commodity components with general purpose operating systems in the nineties led to the emergence of standardised interfaces for storage such as SCSI and ATA. These allowed storage to flourish by allowing disks to be used in a huge variety of servers and situations, not to mention PCs and notebooks.

This was great at first as the workgroup server revolution took hold, as the server system - including storage - was a self-contained atomic entity and could be sourced and provisioned easily. But soon this server sprawl led to management challenges, leading to a recentralisation. The server sprawl was reigned in and servers moved back into the data centre to gain back a level of control and order.

With server management increasingly becoming centralised, the model of direct attached storage started to become problematic. Managing the storage of individual machines as requirements changed is operationally taxing and time consuming, as well as very inefficient. Utilisation rates of 30 percent or less are common, with no chance to make use of the capacity if needed elsewhere.

This led to new approaches to storage, where the data resides in and is accessed via a network. Most commonly these are either dedicated Storage Area Networks (SAN) based on Fiber Channel or Ethernet based Network Attached Storage (NAS).

Networked storage resources were complicated to setup, maintain and reconfigure. But once in place they helped to simplify the allocation and management of storage. This worked well for some time, but changes in computing – driven most recently by trends such as data growth, virtualisation and “private cloud” - now mean that storage is becoming more of a headache for many companies. As we can see in Figure 1, storage has become a constraint on the ability of the data centre or server infrastructure to evolve and become more dynamic.

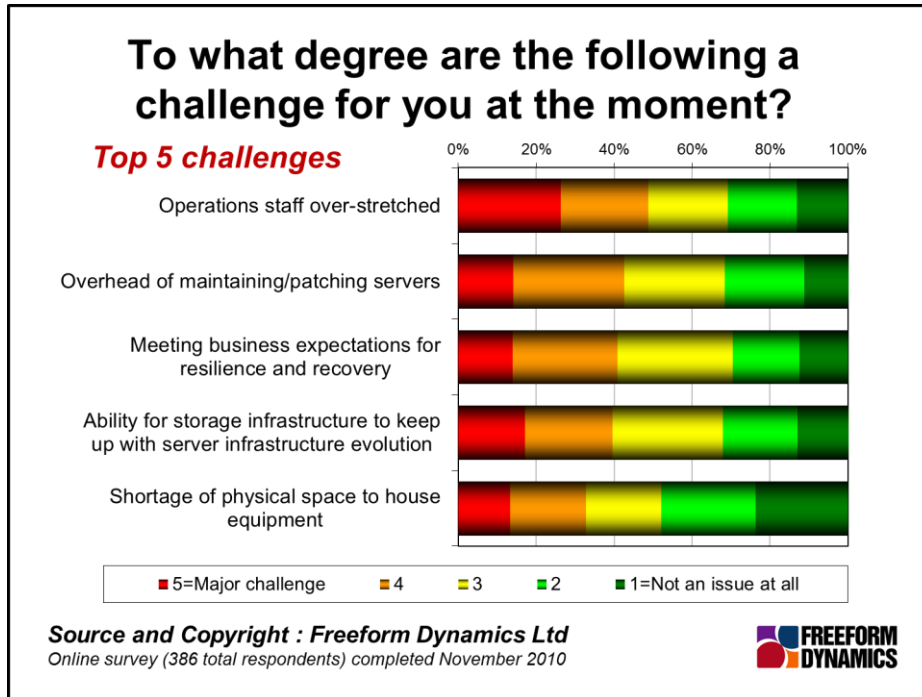


Figure 1

The issue is exacerbated because keeping things running reliably remains challenging, while operations staff are over-stretched and not able to spend enough time or effort on optimising things or doing new stuff. Meanwhile, the storage problem continues to grow daily as data is being created at an ever-accelerating rate, as we can see in Figure 2.

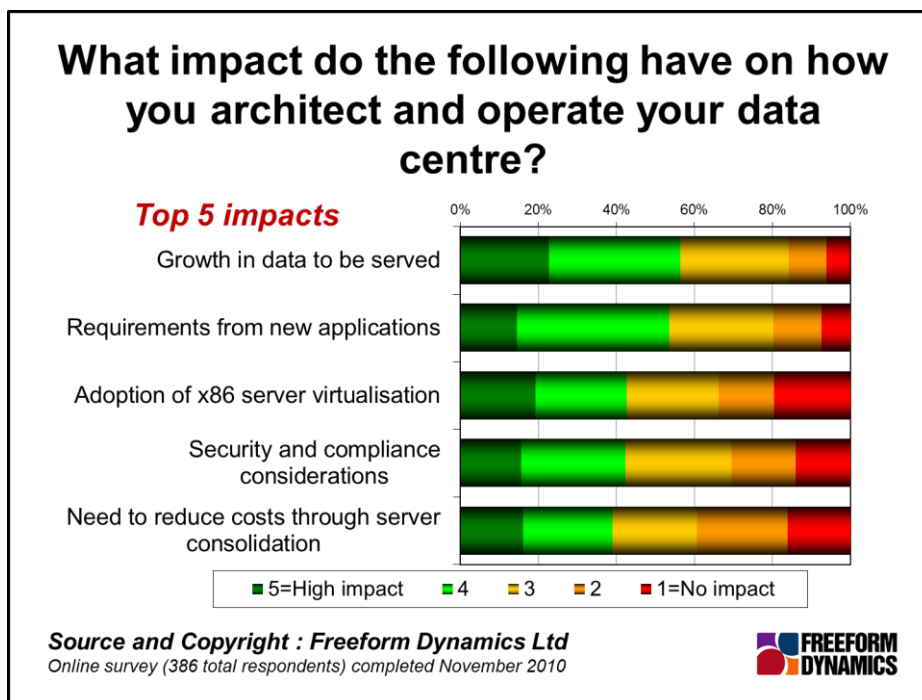


Figure 2

This growth in storage is typically due to a number of factors, the primary drivers being a continuing decrease in the cost per bit to store data coupled with the increasing density of storage devices, increasing computerisation of business processes and the emergence of new types of data such as video, voice or machine data such as log files and interactions. And let's not underestimate the

impact of users just creating more of their own files and expecting them to be available whenever and forever.

The outcome of all this is that many organisations, large and small, are starting to be overwhelmed by storage – by the amount they have to store, by getting it to the systems that matter, and by the huge variety of the types of data being held. While many may not think of it as such, this is very much a “Big Data” problem, whatever the scale.

Many of you may think that Big Data is yet another hyped-up fad that will pass with time or impact only the largest or most demanding organisations with massive amounts of data. However, the reality is that it is a question of scale and the ability to deal with it in terms of investment capability, existing infrastructure and skills.

Companies of all sizes are feeling the effects of the data explosion - a mid-size organisation may struggle to deal with a few tens of terabytes of data, while this may be thought of as relatively simple by a global company regularly dealing with petabytes or even exabytes themselves.

Merely investing in bigger and faster storage arrays, multi-gigabit networks and lots of flash cache may alleviate the symptoms – but like putting lipstick on a pig – the underlying issues remain. Dealing with the problem of Big Data requires a change in mind-set from trying to manage storage systems to trying to manage information.

This brings up the big challenge of Big Data, which is the lack of an information management strategy. Often the mass of data that businesses hoard is not stored because it is necessary due to legal or compliance requirements, nor is it stored because it is indexed for later retrieval and analysis, generating valuable insight for the business.

In many cases the data is stored en masse because it can be done and it is easier to do this in the short term than tackling the root cause of the problem. As the old saying goes – garbage in leads to garbage out. There is little point trying to optimise the IT infrastructure for Big Data, let alone other initiatives, without tackling the underlying issues of knowing what data is being stored in the first place, and for what purpose.

We know - from survey after survey - that getting senior management to invest in management tools is quite a big ask. They don't always get the link between better management and increased service quality and business satisfaction with IT, but are happy to spend money on shiny new applications and services. Yet investment in management is exactly what is needed to provide the foundations for a capable information infrastructure to support the organisation and enable it to react and adapt in a fast moving business environment.

About Freeform Dynamics



Freeform Dynamics is a research and analysis firm. We track and report on the business impact of developments in the IT and communications sectors.

As part of this, we use an innovative research methodology to gather feedback directly from those involved in IT strategy, planning, procurement and implementation. Our output is therefore grounded in real-world practicality for use by mainstream business and IT professionals.

For further information or to subscribe to the Freeform Dynamics free research service, please visit www.freeformdynamics.com or contact us via info@freeformdynamics.com.

Terms of Use

This document is Copyright 2012 Freeform Dynamics Ltd. It may be freely duplicated and distributed in its entirety on an individual one to one basis, either electronically or in hard copy form. It may not, however, be disassembled or modified in any way as part of the duplication process. Hosting of the document for download on the Web and/or mass distribution of the document by any means is prohibited unless express permission is obtained from Freeform Dynamics Ltd.

This document is provided for your general information and use only. Neither Freeform Dynamics Ltd nor any third parties provide any warranty or guarantee as to the suitability of the information provided within it for any particular purpose.