Surfing the flash flood of data growth

Clive Longbottom. Service Director

Quocirca Comment

The terror of Exa

Somewhere around the end of the last century, the first terabyte database trumpeted its way onto the planet, with many looking on in awe that so much data could congregate in one place. Then the petabyte database came along in 2008 as Yahoo! tried to keep pace with its data growth. Now, we're looking towards the exabyte database – the German Climate Computing Center is well on its way to a half exabyte one.

This rate of exponential growth in single databases is totally eclipsed, however, by the overall rate of data growth. Looking at all the data that is now being electronically stored on the planet, we see the growth in the relatively chaotic unstructured data completely outstripping its more well organised cousins in the database arena.

Indeed, it is estimated that out of around 15 exabytes of storage shipped in 2013, around two thirds of the data workload on it was unstructured. 2015 is expected to see around 70 exabytes shipped – with nearly three quarters of the data workload being unstructured. By 2020, it is estimated that close to 150 exabytes of storage will be shipped – with 80% of the data workload being unstructured.

Ex(a)panding to zetta

In 2012, there was around 500 exabytes of electronic data stored around the globe. That is expected to grow to 40 zettabytes by 2020 – an exponential growth, with data volumes doubling around every nine months or so.

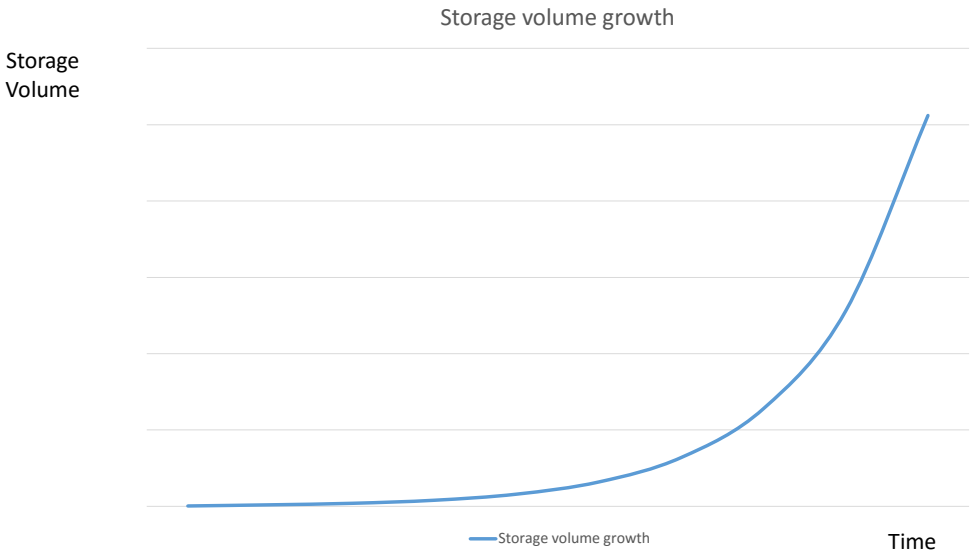

Figure 1

This massive flood of data presents lots of issues to the organisation. Sure – lots of this data will be consumer data of no interest to an organisation, and a lot of it will be dross hiding the gold nuggets of useful information in social media feeds and so on. However, it still needs managing.

How to do this, though? If it was all about the lowest cost of storage, then write it off to tape, put in a storage vault somewhere and forget about it. Actually, not storing it all is cheaper than that – except that it provides no support to the business whatsoever.

Raw costs v. business value

Data storage is not all about the cost of the raw storage – it is more about the optimum cost per workload to provide the desired support to business processes.

Although many still believe that magnetic disk is adequate for some workloads, searching through several terabytes of even structured data will be slow – and this has a direct cost on the business. Move that through to petabytes of dispersed unstructured data and magnetic disks become a major negative impact on the performance of the organisation – many batch data jobs within organisations are already taking hours or even days to complete. This glacial speed of data management is a threat to an organisation's very survival.

Flash-based systems provide that performance that leaves magnetic disk standing. Direct access to data through systems with no moving parts, with lower energy requirements and increasingly dense storage capacities provide much greater capabilities for the all-flash organisation to outpace its competitors in the markets.

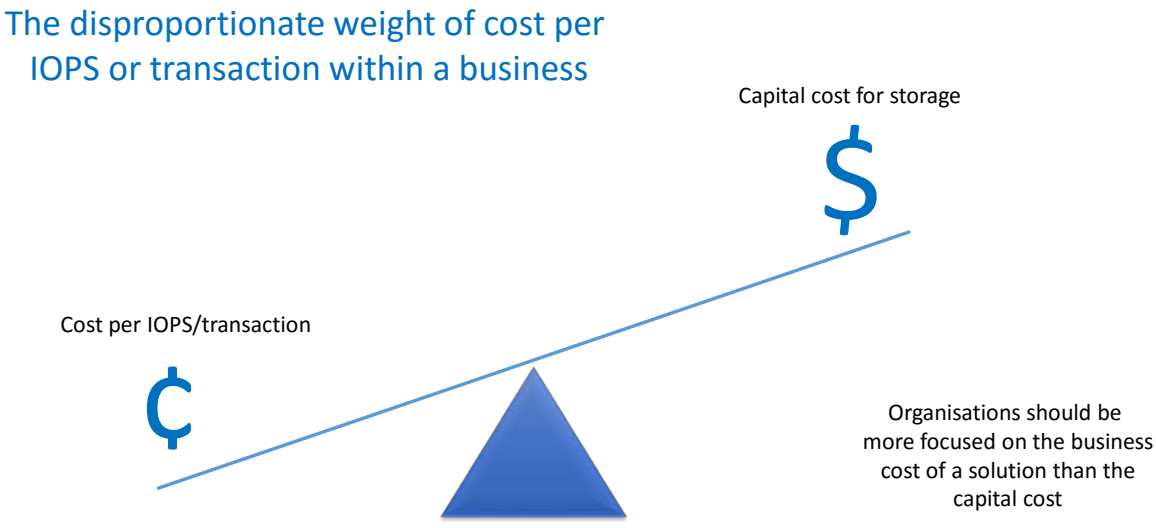

Figure 2

And it doesn't mean spending the earth. As detailed in previous pieces here and here, the EFFECTIVE storage volume of flash is far greater than magnetic disk, making the real cost per byte of flash storage at least as good as magnetic – and improving all the time.

When we then take this and turn it into cost per effective IOPS or transaction, that massive boost in performance that flash provides means that magnetic disk should no longer be a consideration.

It's all in the IOPS

For example, if we take a high-performance, IOPS-focused disk array, it is likely that the cost per raw GB of storage will be higher than a standard magnetic disk array, due to the use of higher performance disk drives, of volatile memory to act as a caching layer and intelligent software aimed at trying to ensure that the 'hot' data (that which is being accessed more often) is cached effectively. This is no simple feat of going down to the local computer store and looking at a sub-$100 4TB disk drive and using simple multipliers.

Indeed, even with a massively engineered IOPS-focused magnetic disk-based storage array, you are still likely to be looking at cost per IOPS.

Now – let's look at a similar flash-based array. With the lowest-cost enterprise-class MLC-based flash substrate, the performance is still going to be many times that of magnetic disk. And, by many times, we are talking about orders of magnitude. So, even if we are looking at comparing RAW capacity, the $ per IOPS or transaction would still be lower for flash than for magnetic; when we look at EFFECTIVE costs, then that $s per IOPS comes down to cents. And that cost differentiation should carry far more weight with the business than the up-front capital cost of any system.

Adding lustre to Hadoop clusters

Even with unstructured data, the capability to run Hadoop clusters much faster, so enabling the continuous filtering of data to reduce volumes, as well as being able to run NoSQL databases at blistering speed means that an organisation that moves to an all-flash storage architecture will be able to uncover insights from mixed mass data far more rapidly – and so make better decisions.

So – Quocirca's advice is to start your move over to a flash-based storage architecture as soon as possible. Look at those workloads that will benefit most from being run on a massively performant platform, and gets those over as soon as possible. Prioritise other workloads, replacing magnetic with flash when it makes sense. Not through timed replacement based on expected useful lifespan, on sweating the magnetic disk assets for as long as possible, nor through replacing only when the nominal book value has slipped below a specific point.

Replace when it makes business sense to do so. This is harder to calculate than when looking at the measures outlined above – but makes more sense. As soon as the storage system fails to meet the needs of the organisation, it needs replacing. For every minute that any part of an IT platform is underperforming against the business' needs, the business is losing money.

This is all that counts as far as the business is really concerned – if you as an IT person can show that the benefit they are getting from a system outweighs the cost, then there should be no reason not to go for it.

About Violin Memory

Violin Memory transforms the speed of business with continuous data protection through high performance, always available, low cost management of critical business information and applications.

Violin's All-Flash optimized solutions accelerate breakthrough CAPEX and OPEX savings for building the next generation data centre. Violin's Flash Fabric Architecture (FFA) speeds data delivery with chip-to-chassis performance optimization that achieves lower consistent latency and cost per transaction for Cloud, Enterprise and Virtualized mission-critical applications. Violin's All-Flash Arrays empowered by our enterprise data management software solutions enhance agility and mobility while revolutionizing data centre economics.

Founded in 2005, Violin Memory is headquartered in Santa Clara, California.

Further details are available at http://www.violin-memory.com

Violin MEMORY

About Quocirca

Quocirca is a primary research and analysis company specialising in the business impact of information technology and communications (ITC). With world-wide, native language reach, Quocirca provides in-depth insights into the views of buyers and influencers in large, mid-sized and small organisations. Its analyst team is made up of real-world practitioners with first-hand experience of ITC delivery who continuously research and track the industry and its real usage in the markets.

Through researching perceptions, Quocirca uncovers the real hurdles to technology adoption – the personal and political aspects of an organisation's environment and the pressures of the need for demonstrable business value in any implementation. This capability to uncover and report back on the end-user perceptions in the market enables Quocirca to advise on the realities of technology adoption, not the promises.

Quocirca research is always pragmatic, business orientated and conducted in the context of the bigger picture. ITC has the ability to transform businesses and the processes that drive them, but often fails to do so. Quocirca's mission is to help organisations improve their success rate in process enablement through better levels of understanding and the adoption of the correct technologies at the correct time.

Quocirca has a pro-active primary research programme, regularly surveying users, purchasers and resellers of ITC products and services on emerging, evolving and maturing technologies. Over time, Quocirca has built a picture of long term investment trends, providing invaluable information for the whole of the ITC community.

Quocirca works with global and local providers of ITC products and services to help them deliver on the promise that ITC holds for business. Quocirca's clients include Oracle, IBM, CA, O2, T-Mobile, HP, Xerox, Ricoh and Symantec, along with other large and medium sized vendors, service providers and more specialist firms.

Full access to all of Quocirca's public output (reports, articles, presentations, blogs and videos) can be made at http://www.quocirca.com

