The joy of workloads

Clive Longbottom. Service Director

Quocirca Comment

The fallacy of quoted IOPS

Take a look at pretty much any storage system, and the vendor will be quoting input/output operations per second (IOPS) figures. While this does provide a standardised means for a level of comparison, it is, in itself, pretty useless.

Why? Because the standardised test is too specific, and can be gamed to give the highest possible figures. The tests do not use real world workloads – those IOPS may suddenly be badly hit in the real world of your environment.

Let's start with a look at the 'legacy' real world of spinning disks. In a lot of cases, a specific storage system will have been implemented to support a single specific workload. This can then be hyper-tuned to provide the best performance to that single workload. Add another storage workload and things get a little dirty.

Why spinning disks struggle

This is because of the way that spinning disks work. Data is spread across multiple platters in multiple drives. A table of where data is on these platters is maintained. When data needs retrieving, the first action is to go to this table to find where it is. The system then has to wait until the specific area of a specific platter is under a specific head before it can retrieve it. This 'identify/wait/retrieve' cycle then repeats for every block of data that is required. This creates a situation known as 'rotational latency'.

This is with one workload, remember. The 'identify/WAIT/retrieve' is one of the main reasons why flash is so much faster than spinning disks – with flash, that 'WAIT' is instantaneous, as flash is a direct access storage medium.

Let's move on to two workloads on a spinning disk storage system. That first workload has already gone to the table and identified where in the system the first block of data is. It has then gone to the specific drive and platter and is now waiting for the data to appear under the head. The second workload comes along and goes to the table – and finds that the data it needs is on the same disk drive as the first workload is already accessing.

Those darn laws of physics

It is an unfortunate law of physics that a physical object cannot be in two places at exactly the same time (OK – quantum physics may disprove this at a quantum level, but here, we are talking about something much larger). The disk head cannot suddenly stop what it is doing for the first workload to respond to the needs of the second workload – unless priority of service has been applied to over-ride the initial retrieve order.

So, after the 'identify' action, there is suddenly another 'wait' in place, as the system has to wait for the first workload's 'retrieve' action to take place. Now we have an 'identify/wait/wait/retrieve' process in place. To make this worse, as spinning disks become more utilised, they fragment files into smaller pieces – on a clean disk, a full file may be written as contiguous blocks on a single platter. As the disk fills up and more deletions are carried out, the disk has to find spare areas on the disk to write to, so breaking files up into smaller blocks and distributing them around the platters, disks and arrays. Therefore, that 'identify/wait/wait/retrieve' process becomes slower and slower.

Pooling resources

Now extrapolate the problem to a single virtualised pool of storage managing all an organisation's storage workloads. Just how long could that 'identify/wait/…/wait/retrieve' process be? As it is dependent on the speed of moving parts (motors, actuators, spinning platters), only so much can be done to optimise the whole procedure. Therefore, spinning disk will always be sub-optimal when it comes to dealing with mixed workloads.

Being in two places (almost) at the same time

So, over to flash storage. The overall process is still pretty much the same – where the data has to be identified from the table, and then it has to be retrieved. That instantaneous 'wait' is not noticeable in the greater scheme of things. When an extra workload is added, the same issue as there is with spinning disk may still be seen, in that it is not possible for two concurrent retrieve actions to be carried out against the same flash system under the same physical controller. Therefore, the second workload does have to wait until the first workload retrieve has completed – but because the 'retrieve' is so fast, the 'wait' is still hardly noticeable.

Even with multiple workloads being applied against flash, that 'wait' is going to be measured in low microseconds, rather than the milliseconds that spinning disk can introduce into a mixed storage workload environment. Now, throw virtualisation and cloud architectures into the mix and the importance of low latency cannot be over stated.

My IOPS are broken…

The impact on quoted IOPS can be massive. For example, in one real-world case involving a well-known manufacturer's entry-level converged storage array, the buyer was led to believe that they could get around 2,000 IOPS from the system. In real world usage, they were getting closer to 700 IOPS[1]. The vendor explained that real world and hypothetical results were bound to be different – but gaining only 30% of a standardised test result must have left the customer feeling somewhat unhappy. And remember, this was against a single workload.

In recent tests, independent testing group Demartek tested a flash-based system against standardised IOPS workloads and real-world workloads[2]. Demartek's findings were that as expected, the test IOPS were far higher than the real world IOPS: using Iometer standardised workload testing methodologies, performance was found from 100,000+ IOPS for a single workload through to 1,000,000 IOPS for multiple mixed workloads.

Using real-world actual storage workloads, Demartek found that it instead found a reasonably consistent IOPS performance of between 250,000 to 300,000 IOPS irrespective of the number of workloads running.

In these tests, it showed that the user could, at the highest standardised results level, expect around 30% of the quoted IOPS, as was found with spinning disks. However, they could also expect to get two and a half times the performance against the standardised tests in other areas.

Standards, trust and the real world

What this goes to show is twofold – 1) don't trust standardised tests to tell you how a storage system will work in the real world, and 2) if you want consistency in storage performance, then flash is the way to go.

When all is said and done, the aim should be to homogenise an organisation's storage platform. This means that mixed workloads are unavoidable – and that putting these on consolidated storage is the future. The joy of workloads should not be mixed, however – consistency and predictability of high-end performance are what makes the whole experience joyful – and flash is the best way to ensure this.

[1] https://community.emc.com/message/690249
[2] http://www.demartek.com/Reports_Free/Demartek_Violin_Memory_7300_Multiple_Workloads_Evaluation_2015-09.pdf

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

