ORIENTATION GUIDE

Storage in the virtual era

Revisiting the requirements on data storage

Freeform Dynamics, 2010

In a nutshell:

The growth in use of server virtualisation technology is having a knock-on effect on other infrastructure capabilities including storage. As it becomes a core technology (of this we have no doubt), virtualisation can both help and hinder our efforts as we aim to store and manage information for our organisations. In this paper we remind ourselves of the requirements on core storage and considering the impacts, both positive and negative, of server virtualisation.

Key points:

- Server virtualisation has some clear benefits including cost efficiency and IT flexibility. However these are still early days for the technology.
- From a storage perspective, there's plenty good about server virtualisation. However storage infrastructures were created historically based on the 'silo' principle – this cannot be broken without consequences.
- As well as delivering storage consistently to meet expectations, requirements on storage include availability and performance; accessibility and security; adaptability and management.
- Storage virtualisation may be a key tool in dealing with the bottlenecks that server virtualisation can cause. However it can't be done by itself, a view of the relationship between the physical and virtual environments needs to be maintained.
- A number of new storage capabilities are now available for businesses of all sizes. We consider how these might address the needs of disaster recovery, remote site access and making storage an operational cost, in the virtualised environment.

As businesses grow ever more reliant on the exploitation of digitised content, the issue of how we handle data has never been more important. Virtualisation is just one element making the task of information management tougher than ever before, as it brings a number of benefits but needs to be approached in the right way. If you are trying to help your business work out its priorities, scope out your existing IT capabilities or liaise between the two, this paper is for you.

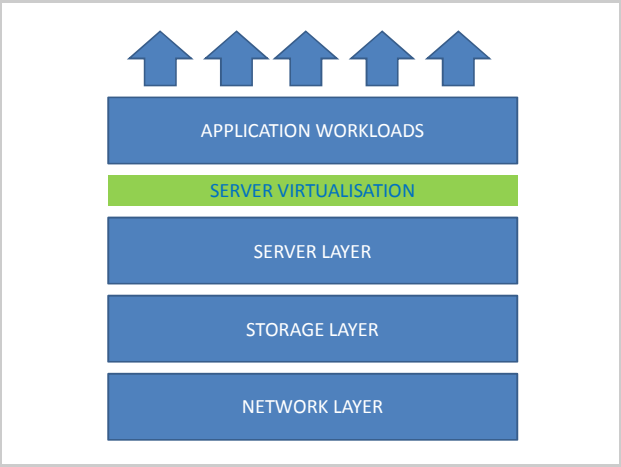
The insights upon which this paper is based are taken from a number of research projects designed, executed and interpreted independently by Freeform Dynamics. This paper has been authored independently by Freeform Dynamics and is sponsored by HP.

The impact of server virtualisation

Storage, together with processing and networking, is one of the central pillars of information technology infrastructure. More than anything else, storage is considered as something that should just work -- like a dog burying a bone, and then expecting it still to be there when he comes back, storage is considered by all but storage managers, as something straightforward to the point of even being quite boring.

However, while disk capacities may not have kept up with processor speeds, there have nonetheless been some fantastic innovations in how data is stored and managed. Such developments have not just been in the drives, though today we are seeing solid-state disk, capacities of over 2TB, perpendicular recording and so on. Also however, what we know as 'storage' encompasses many advances in how information flows are managed from server to disc.

As well as ploughing their own furrow when it comes to innovation, storage technologies have to keep up with new developments in servers and networking, and the application architectures that all three infrastructure pillars need to support. Today for example, we are seeing x86 server virtualisation offering new opportunities to balance and manage application workloads across physical resources. While the principle of virtualisation may not be new, outside of the mainframe world the practical existence of an abstraction layer between physical resources and their logical application workloads certainly is (Figure 1).



Server virtualisation provides an abstraction layer between the physical server infrastructure and the workloads it supports.

Figure 1

From our research we know there are three reasons why organisations are using server virtualisation:

- First and foremost, organisations need to reduce costs while maintaining service levels – server virtualisation is seen as a cost effective option in terms of requiring less physical hardware, easier provisioning and reduced licensing overhead.
- Second, there is a drive to improve the responsiveness and flexibility of IT. Virtualisation simplifies operational management – e.g. a virtual machine can be moved between physical servers such that the underlying hardware can be upgraded.
- Third, many organisations are looking to improve their power consumption. Many physical servers are under-utilised: roughly 5-10 virtual machines can be migrated onto a single physical server, and the rest switched off with immediate benefit.

Server virtualisation has its benefits, to be sure, but these are still early days for the technology. Many organisations are still only piloting virtualisation, or they use it specifically for discrete, non-mission-critical workloads such as test, research and development environments. Unsurprisingly, less than a quarter of respondents to a Freeform Dynamics study earlier this year said they had anything approaching 'expert' level experience of virtualisation.

Reviewing core requirements on storage

Innovations such as server virtualisation do impose new constraints, but these are not the only forces on the storage layer. Despite the economic pressures that need to be taken into account as much as ever today, information growth remains a primary concern for many organisations – indeed, it was the number one criterion reported in a recent study concerning the deployment of new server infrastructure, so one can only imagine where it sits with respect to storage infrastructure!

Before we go on then, let's revisit the fundamentals of what we need of our storage. First and foremost, we absolutely need to know that when something is stored, it can be retrieved consistently and reliably. Building on this, we can consider the following criteria:

Availability and performance. Two perspectives may be taken: user and workload. From a user perspective, storage availability and performance come down to service level criteria, which can be measured in seconds – though remember a user's 'storage experience' may also be based on the time taken for information to cross the network. Meanwhile however, applications may be even less tolerant than users than users when it comes to performance. Latency thresholds (based on Input/Output Operations per second, or IOPs) need to be stuck to, otherwise the storage layer can become a bottleneck to applications. And finally, information may need to be kept available and performance maintained at an acceptable level in the case of a major failure, be it caused through a systems crash, fire or flood.

Accessibility and security. Information must be accessible, this much goes without saying. Accessibility boils down to standards in terms of data formats, file systems supported and indeed the standards used to access storage devices, either directly or over a network (such as SATA, Fibre Channel, iSCSI, and so on). It also imposes requirements on how information is shared between workloads. As a counter to accessibility, security criteria need to ensure that only the applications and users who are authorised to access certain information, can access it.

Adaptability and manageability. The storage layer needs to be able to adapt to changing conditions and requirements: the faster it can do this, the less costly it will be. In concrete terms for example, adding capacity to a storage 'pool' might require devices to be turned off or reconfigured – the time delay is an overhead in both operational and business terms. Also, while it would be good (in principle) that the storage layer responds automatically to new needs or thresholds being reached, in practical terms this will require human intervention so some level of management reporting and configuration capability will always be required.

Storage infrastructures have evolved through the years to support the needs of specific workloads. Two main paths exist, depending on the type of information to be stored:

- File-based information (unstructured content – documents, media files and so on) has tended to be stored either on storage directly attached to servers or using a storage device attached to the local network, known as Network-Attached Storage (NAS).
- Block-based information (generally structured content – that is, databases, data warehouses etc) has also moved from the server to be stored in specialised storage environments known as Storage Area Networks (SANs).

Server virtualisation impacts each of these approaches. Let's take a look.

What impact does server virtualisation have on storage?

To understand how server virtualisation impacts the storage layer, we need first to consider how larger IT environments have evolved. Apart from once-a decade consolidation activities, in general IT is added to each time a new application is deployed – with a new set of servers and a new set of storage devices, in a so-called 'silo'.

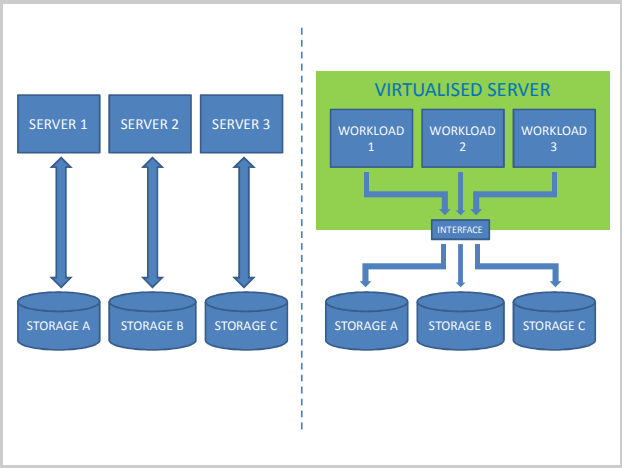
While this approach causes problems of its own (physical space for a start, as well as the manageability overhead of trying to deal with multiple, disparate hardware and software platforms), it nonetheless has one major benefit: that each application's storage requirements have been considered more or less individually. The paths between application processing and storage, and

the bandwidth required, can be calculated relatively straightforwardly for individual applications. And while bottlenecks have always existed, the siloed nature makes them relatively easy to spot.

From the storage perspective there is an additional downside, however. As each storage system was acquired fully specified up front to cope with expected loads, a lot of spare capacity exists today that has never held valuable data, but which is unavailable for anything else to use. Storage is today one of the highest costs on the hardware and software acquisition budget, and even higher on operational maintenance.

Enter server virtualisation – which, as we have said, offers a great way (in principle) to consolidate multiple workloads onto a single physical server. The benefits of server consolidation via virtualisation are generally well appreciated. In principle again, these can also result in more efficient use of storage, by breaking the silos of storage that were bought to serve the needs of particular applications.

What's not to like? The downside of server virtualisation (the clue is in the name) is that it doesn't currently pay sufficient attention to the physical environment beyond the server. It is all very well in principle to shift multiple workloads onto a single server. But rather than having their own routes to the physical storage layer, information flows now take place using the same physical interface (Figure 2).



By moving from the siloed model, server virtualisation may nonetheless be a cause of physical bottlenecks.

Figure 2

From talking to early adopters of server virtualisation, we know this can easily cause a bottleneck if the workloads are data intensive and/or they require access to concurrent physical resources. As mentioned earlier, at least silos made bottlenecks easier to spot. Not so with virtualisation: a number of IT managers have told us of difficulties understanding what physical resources a virtual machine is actually using. Management tools are not yet sufficiently mature to provide such a view, and IT staff do not yet have sufficient experience of the new environments to make any such mapping straightforward.

All is not lost however – and we need to remember these are still early days for virtualisation in the distributed systems environment. Management tools vendors are stepping up to the plate and working with virtualisation vendors to ensure the physical and virtual worlds are both taken into account – and we are starting to see the fruits of their labours. Meanwhile, while traditional storage infrastructure (NAS, SAN and the like) evolved without any forethought of virtualisation, new capabilities from a number of vendors are starting to 'bake in' recognition of server virtualisation.

An important element in all of this is storage virtualisation, which we consider below.

Bringing storage virtualisation into the mix

The mechanisms to manage and deliver disk capacity to applications and users – that is, all the innovations we consider under the banner of 'storage' – have not been standing still. Indeed, storage virtualisation has been around as a mainstream technology for longer than server virtualisation.

This is not the place to go into technical detail about storage virtualisation. But let's recall the principle – once again, it is about implementing an abstraction layer between physical resources (in this case, disks and storage networking) and logical resources which can be served up to applications. As we have discussed however, the abstraction layer needs to have a pretty good grasp of what's going on in the physical world, such that bottlenecks can be avoided, or at least spotted and dealt with.

Storage virtualisation certainly has its place – not least because it makes the overall pool of storage easier to manage, and enables the allocation of storage as and when it is needed (so-called thin provisioning) – this has knock-on benefits in terms of reduced power utilisation. It also offers more opportunity to support disaster recovery scenarios, as data can be replicated and moved within the overall storage pool with minimum disruption to users and applications.

In practical terms however, it is not enough to consider either server virtualisation, or storage virtualisation, in isolation from each other. What both server and storage virtualisation give us, is a number of options, tools in the tool chest as it were, that make it possible to revisit how things are done. This is particularly true with storage virtualisation, as there is more than one way to implement the technology. With this in mind, let's look at some practical examples.

Getting down and dirty with virtualisation

Rather than seeing server and storage virtualisation as the answer to life, the universe and everything, let us consider a number of real-world scenarios and how virtualisation gives some new options to deal with them. The first couple consider the use of modern storage capabilities to support server virtualisation; in the third, we consider storage virtualisation and its role in disaster recovery scenarios.

1. Delivering IT services to remote branches

Problem: Many organisations have offices distributed over large distances. Providing IT services to these locations can be challenging, especially as it has become unlikely for skilled IT staff to be located on site. Delivering services to such remote locations can be difficult, as well as managing storage and data protection challenges that result.

Solution: one approach now being utilised is for organisations to deploy virtual servers on remote sites, as supporting such systems from a distance can be far more straightforward than with physical systems. Where VMware ESX virtual servers are used, it can be possible to employ software to use the device's physical storage as a virtual SAN appliance. This approach enables remote management of the storage and can deliver many of the cost and availability benefits associated with SAN systems in the remote office without having to physically install a SAN on site.

2. Making storage associated with virtual servers an operational cost

Problem: Even capital-rich companies are looking to how they can move towards a model where they only pay for what they use. This is particularly applicable in the virtual server environment: once workloads are consolidated, attention turns to the cost of storage. As has been discussed the physical acquisition of storage that remains unused over years is expensive and wasteful. Being able to make incremental changes to available storage is an essential element of paying only for storage when it is needed.

Solution: Many applications need storage to be 'associated' with them from day one, even though they do not fill it with data immediately. The capability to 'thin provision' storage', that is, fool the application into thinking it has a large pool of data at its disposal whilst physically giving it only what it uses, offers tremendous benefits. Traditionally this has been challenging but thin-provision capabilities available in some virtual storage offerings can reduce storage requirements by a

significant factor. This is a capability that will become widely used, but only after IT budgets evolve the ability to fund incremental storage acquisitions rather than having to fund storage 'up front'.

3. Information availability in the smaller business.

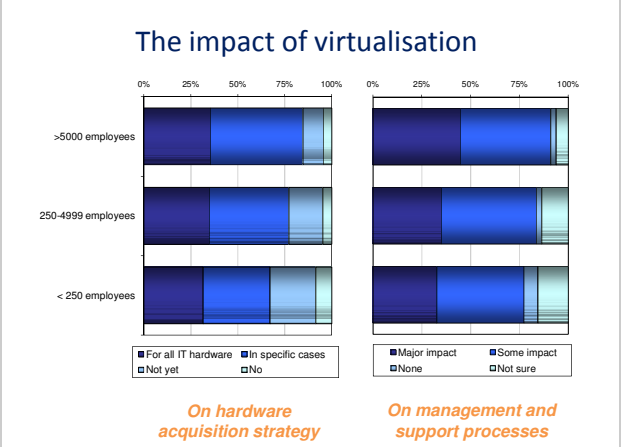
Problem: Rare as they have become in modern storage systems, physical errors do infrequently occur on the spinning disks used to hold data. This is an issue faced by larger and smaller organisations alike.

Solution: The traditional, almost default, step taken to enhance data availability has been to use the 'RAID' capabilities of storage systems where data is written, or 'striped', across all available storage disks in the cluster. This approach lets you do things like replicate different types of data as appropriate whilst providing better data availability in the face of hardware problems. Some newer storage virtualisation and management solutions available to non-enterprise customers include capabilities such as 'synchronous write' across the network – which enables information to be replicated to another server room, or indeed off-site.

Another capability becoming available is 'SAN stretching', whereby half of a logical SAN may be located physically in another location, possibly in a service supplied by an outside agency for smaller organisations that do not have multiple computer rooms. SAN stretching can offer combined high availability and disaster recovery capabilities at a cost effective price and offers higher levels of service than may be available using traditional offline tape for these purposes.

Adopting a joined up approach

For better or worse, server virtualisation appears to be inevitable in many organisations. But it cannot be considered in a vacuum: and we know from research that it will have an impact on both hardware acquisition strategy, and on management processes (figure 3).



Server virtualisation is having a profound impact on both hardware acquisition strategy, and management and support processes.

Figure 3

What does this mean from a storage perspective? Traditionally storage has been an "afterthought", acquired as part of a bigger system. But it is becoming clear that, for virtualised server environments to run efficiently, it is necessary for them to consider the interface with storage in a more joined up manner. The question to ask is, "What about the storage?" but not from a siloed, application-specific perspective, but from the point of view of the environment as a whole.

Meeting the needs of the virtualised server environment may well require some investment in storage infrastructure – and we know from out research feedback that this can be difficult to justify. The good news is that a number of capabilities now exist that were previously only available to enterprise customers, but now are an option for companies with smaller IT budgets. However it is important to think about your needs first in business, rather than purely technical terms.

You will no doubt already have a good grasp of where the challenges lie in your existing environment, but what may be lacking is the big-picture view – knowing what workloads are seen as a priority from the business perspective, and expectations in terms of service levels. It's by building such a picture that you can understand what needs to be prioritised, as well as giving the starting point for a business case that takes into account not just the up-front capital costs but also the impacts on the existing environment, including the inevitable re-training and ongoing management overheads.

From a technical perspective as well, there's a caveat: given the state of flux of both server and storage virtualisation, it is not just a case of assuming it will all 'just work'. Some virtual machine documentation will casually mention that in order to take advantage of a particular capability you will need a particular networked storage solution type. Again, we'd advise against jumping in with both feet before understanding exactly what you are trying to achieve – since the investment in the wrong storage virtualisation solution, or failure to configure the selected option correctly, could result in operational costs that could dwarf the capital costs.

To summarise, then, virtualisation of both servers and storage do offer benefits in terms of flexibility and efficiency, but only if they are implemented in the right way. Here are some considerations that might help you on your way to adoption:

- Make sure you have a clear idea of what storage virtualisation means to you and the benefits you expect from it.
- Conversely, if you are buying in hardware and software for virtualisation, consider what other business benefits the products could bring to strengthen your business case.
- A significant amount of the benefit from virtualisation will come in operating and running costs so be sure to factor all of these (such as saved rack space and power) into your business case
- Be ready to change IT (and business) procedures to take full advantage of virtualisation
- Prototype and experiment with your target hardware and proposed solutions to get a feel for the real likely level of benefit in your own environment
- To get the full benefits, virtualisation needs to be implemented in a structured, strategic way rather than as a set of point solutions for specific projects
- Make sure that the management tools deliver what you need to keep the platforms operational with minimum administrator time
- IT budgets may need to become flexible if "thin provisioning" or just in time storage acquisition is to be used.
- And don't forget to test your ability to protect and recover data from a virtualised storage system!

As a final thought, good management processes are perhaps the most important element of all. The fact that virtualisation, by its very nature, helps to hide infrastructure complexity is an advantage on which many benefits are built but without good management approaches and solutions the hiding of complexity can also become a factor in its downfall. It is likely that ongoing management capabilities, your ability to undertake them and modify your operational procedures to take advantage of virtualisation will ultimately be the key to success or failure.

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.

About HP

HP, the world's largest technology company, simplifies the technology experience for consumers and businesses with a portfolio that spans printing, personal computing, software, services and IT infrastructure. More information about HP (NYSE: HPQ) is available at http://www.hp.com/.

HP document reference number: 4AA1-5918EEW