WHITEPAPER

The Modern Virtualized Data Center

Data center resources have traditionally been underutilized while drawing enormous amounts of power and taking up valuable floorspace. Storage virtualization has been a positive evolutionary step in the data center, driving consolidation of these resources to maximize utilization and power savings, as well as to simplify management and maintenance.

There are many entrants into the server virtualization market, including VMWare, Microsoft, Oracle, Citrix, and others; so there is a good chance that you are either running virtualization in a test or development capacity or have taken the leap toward a production virtualized data center. With all of the benefits of virtualization comes a wholesale change in the way that you must plan for implementation, management, backup and recovery.

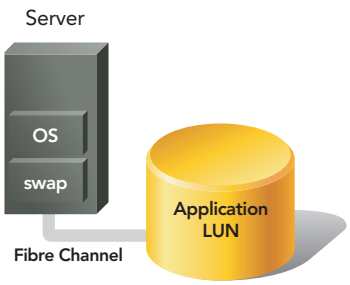
The Zen of Storage Maintenance (in a virtualized datacenter)

Consolidation projects large and small seem to focus on the server and the application. This makes sense, because these projects are often justified by capital and operational expenditure (CAPEX and OPEX) cost savings and in giving appropriate resources to the most business critical applications. Storage has taken a back seat in these projects because most IT managers think of a LUN as a commodity.

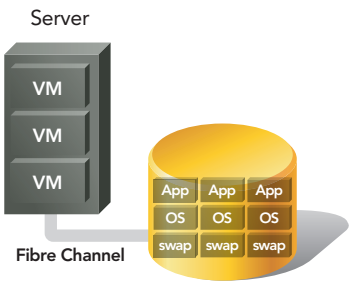
A LUN from a SAN-attached array hasn't changed in decades, and they are all treated equally by the array. Most virtualization software will simply place their files on the LUN or make some kind of raw access meta-data file, so tuning the array for virtualization is pointless. According to the best-practices guides from many virtualization vendors, the value of write cache, read-ahead, and all the value add of the high end array go out the window as the workload from a VM (virtual machine) tend to become more random.

The figures below show the migration from a physical environment to a virtualized environment, depicting the added workload to the SAN, which in turn tends to generate more random IO.

The Good Old Days



The Good New Days



pillar® DATA SYSTEMS

But what if there were a way to provision your storage the way you provision your virtualized servers? You could assign array resources (CPU, cache, and disk) based on the applications that will be using them. This would improve the array's value by greatly increasing its utilization as well as increasing the performance of the applications that run on it. The result would be better server consolidation, resulting in cost savings – and isn't that what this project is all about?

Application-Aware Storage

The Pillar Axiom® is the first storage array that is truly application-aware. It is tuned both automatically and by the administrator for the applications that utilize the array on a file system and LUN level. How does this improve overall system utilization and virtual machine (VM) performance?

With the Pillar Axiom, you gain performance by provisioning your array as you do your VM, assigning resources based on application importance. The Axiom will treat I/O from your provisioned application with a greater priority, giving it more cache and faster spindles than other applications that you have deemed to be less critical to the business. But getting the best value out of the application-aware Pillar Axiom means going against the best practices guides of most virtualization software companies which will suggest that you assign many virtual machines on a single LUN (as individual files, in most cases). Although this seems like a good idea, it actually has an adverse affect both on utilization and on performance due to the issue of capacity vs. spindles.

Let's look at an average scenario: a standard mid-tier array with 300GB of 15K RPM fibre channel drives in a recommended RAID 5 4+1 configuration. This configuration will net about 1.2TB of usable capacity, which seems adequate since most VM software has a 2.0TB limit. But how many VMs will fit on this 1.2TB LUN? Five, ten, twenty – or more? This is what Pillar refers to as a stacked array, where one or multiple physical or virtual LUNs are striped across a RAID group, as shown in the illustration below. You may be working with only five spindles, yet you have up to 20 machines accessing that RAID group. No matter what the individual I/O profile looks like for a single server, the array will see 100% pure random I/O, which is why the write-cache gets overloaded and read-ahead caches become mostly useless.

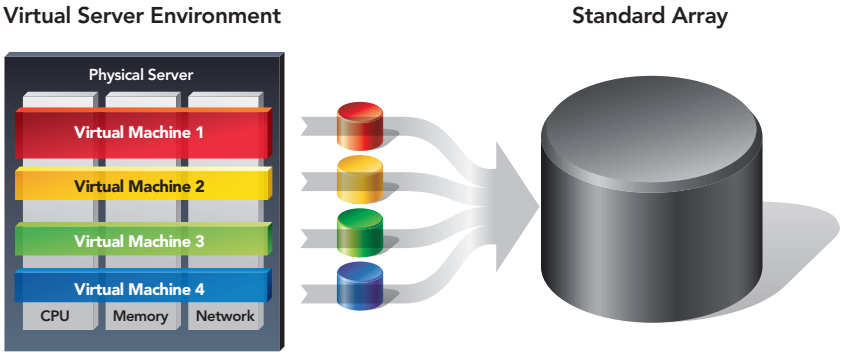
So what is the solution? Adding more spindles won't improve matters; although your IOPs improvies, your LUNs will get larger as you as you add more spindles – essentially throwing away capacity (see diagram below). Although virtual LUNs will help, they won't help enough to mitigate the pure random I/O that your array will see. The solution is application-aware storage.

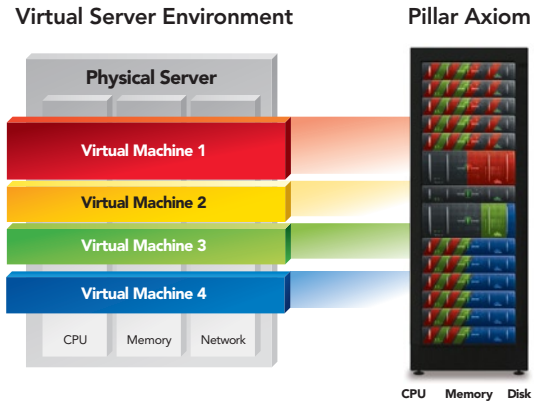
The application-aware Pillar Axiom has a rich feature set that mitigates each of these issues which, in turn, improves VM performance, increases array utilization, and provides for much greater VM consolidation.

Let's return to the scenario above with the LUN using a RAID 5 4+1 configuration. Although the Pillar Axiom supports RAID 5, each LUN, no matter how large or small will get striped across at least four RAID-5 disk groups, which increases the spindle count to at least 24 disks. Now that we have more spindles to work with, we need to differentiate the performance of the individual LUNs and tie them to their respective VMs running individual applications.

The administrator assigns resources on the server by virtue of CPU, memory, and network, however on the Axiom the administrator will either use a pre-existing profile for their application or create a custom profile based on CPU, cache, and disk layout (fibre channel or SATA). This will map them to the individual VM as either a single VMDK (virtual machine disk format) or RDM (raw device mapping) (using VMware terminology as an example). No stacking VMDKs on a single LUN using this approach, Pillar will take care of this aspect to keep the random I/O random, and the sequential I/O sequential to take advantage of the write-cache and read-ahead array functions.

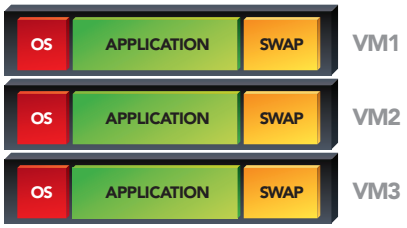
The result is an array that shows the same resource preferences as your servers. This improves performance for the applications to which you assigned greater server resources, while not punishing the workload from other applications. This inherently drives greater array utilization and Pillar has found that it also increases the amount of VMs that can be placed on a physical server by reducing or eliminating the fibre channel I/O bottleneck in high-end consolidation projects.





Provisioning Virtualized LUNs

When an array treats all I/O on a first-come, first-serve basis, provisioning VMs onto LUNs is quite simple – as shown here:



Although this would seem to be a logical way to deploy VMs, it does nothing to inhibit a virtually 100% random I/O workload which is the toughest workload that an array can endure. The more random the workload, the more spindles that you will be forced to buy, and the lower the overall system utilization will become, thus increasing the overall costs of your virtualization deployment (think $/GB).

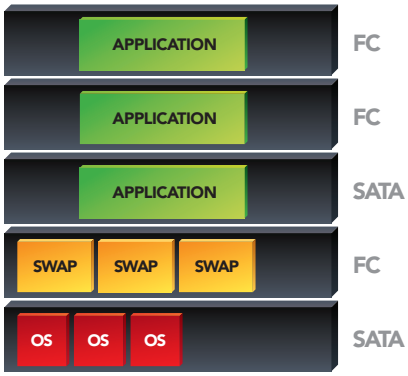
So, what can be done to mitigate this random I/O? To do so, one must characterize the workload that is added when migrating from physical to virtual infrastructure. As mentioned previously, before virtualization, the OS and SWAP space were local to the server hard disk, and the application was the only user of the SAN. For virtualized infrastructure, the OS, SWAP and application must all reside on the SAN. So let's characterize the requirements of each:

Operating system – generally not highly utilized files. Most files would have sporadic access with relatively low I/O requirements after boot.

Application – this should be the easiest to characterize as workloads vary by use and are well-known, be they database, e-mail, content management, etc.

SWAP – this one usually throws admins for a loop. SWAP is part of the OS, but can be moved from the OS partition to another LUN. And SHOULD be moved from the OS partition to another LUN. This is because SWAP is essentially memory operations to disk. This is 100% random, high performance, low latency I/O. Whereas you can deploy the OS on a SATA disk, deploying SWAP on SATA is tantamount to placing a time-bomb within your server. Although a server should never use SWAP, when it does, it will usurp control of the array and any applications sharing that array have a tendency to fail due to I/O timeout.

Armed with this knowledge, a better way to deploy your VMs is depicted here:



Based on the performance and business requirements of the application, they can be placed within a LUN on FC or SATA disks. The OS VMs can all be placed on a single LUN and may even be de-duped using writeable snapshots (Pillar refers to this as a "writeable clone"). Since the OS is only used on boot, and then sporadically, there is no reason to use high performance disk. Finally, SWAP should all be placed on the fastest disk your array will support, but can be placed within a single LUN (as most SWAP files are small, between 6–24 GB in size). By placing all SWAP files within a single FC LUN, this gives them the random IOPs they require yet quarantines them to a single set of disks, where they can do no damage to the overall performance of the array.

But there is only one caveat to the above: it can only be done on an array with quality of service. As the Axiom does not treat all I/O the same, it can differentiate services to each of the above LUNs to ensure that no one LUN or Application seizes control of the system. This mitigates contention, creates more of a sequential I/O workload as well as dramatically improves system capacity utilization.

Summary: Keeping the Data in the Data Center

Particularly in a virtualized data center, consolidating DAS, NAS, SAN, and other disparate storage resources onto a networked storage pool can go a long way toward improving efficiencies. Consolidating storage on the heterogeneous, application-aware Pillar Axiom not only holds the pooled storage on an easy-to-access platform, but it allows for all management and maintenance to be performed through a single interface, greatly reducing management effort and cost.

With the highest utilization rate in the industry – up to 80%, more than double the industry average, and with the lowest power and space consumption per GB of storage, the Pillar Axiom is fast becoming the storage system of choice for the virtualized data center.