THE CLIPPER GROUP
Navigator™
Navigating Information Technology Horizons℠
Published Since 1993 Report #TCG2009015 March 30, 2009

FalconStor's Storage Virtualization and Data Replication Deliver Value to *Both* Business *and* IT

Analyst: Anne MacFarland

Management Summary

In a data center, applications and the business processes they support are of primary importance. One might call them the *A-Team*. Infrastructure services exist to assure they run smoothly. The most basic IT infrastructure services are data replication, and storage virtualization. Data must be replicated for protection. The replication must be sophisticated enough to support a quick restart of the application and access to its data stores should something go wrong with the infrastructure (and there are many levels at which something can go wrong). A remote copy will keep the business whole if the headquarters and its data center loses functionality. Storage is virtualized to let its capacity be used more smoothly and efficiently. The layer of indirection or mapping that storage virtualization provides makes provisioning an application with storage resources less exacting, allows storage be shared safely, and lets additional capacity be added transparently. Storage virtualization facilitates adroit data replication.

Most A-Teams that come to mind, like athletes, have excellent equipment (such as custom, special-purpose footwear) and trainers to help them perform to the best of their ability. Similar qualities in data replication and storage virtualization provide an unobtrusive competence that can preclude outages for backup, performance impairment, or a lack of functionality in the application itself. By contrast, clumsy IT management practices, epitomized by repetitive, manually-invoked full backups, make management a part of the problem. It is as shortsighted as putting shoes that don't fit on your best athlete.

IT infrastructure services have always been important. What is more new is that their quality has to be designed in. Any other approach is risky and subject to human error. With customer and employee self-service, IT blunders are more visible and can affect brand loyalty. As more applications call other applications, or use common data stores, failure has wider ramifications than heretofore. In the data center, the window for intrusive maintenance operations has disappeared. As businesses seek to rein in costs and control complexity, server consolidation has pruned the idle cycles that might be unobtrusively allocated to utility processes. Even "in-the-background" asynchronous operations must be strategically designed.

FalconStor Software, based in Melville, New York, has earned a name for its *IPStor* storage virtualization platform, which is resold by many vendors. FalconStor has used this platform as the basis for its replication and virtualization services. These services are expressed in four products – *VTL* (*Virtual Tape Library* – the backup target), *CDP* (*Continuous Data Protector* - the replication manager, covering a spectrum from continuous to asynchronous), *NSS* (*Network Storage Server* – the storage virtualization gateway now also available as a storage appliance) and *FDS* (*File-interface Deduplication System* – a block-level deduplication engine with a NAS interface). All elements adhere to a philosophy that a good service leaves little evidence of its presence but great satisfaction by its actions. For more details on the FalconStor design imperatives and the products built on them, please read on.

IN THIS ISSUE

FalconStor Utility Design Imperatives

1. Virtualization – the Basis of Transparency

Storage virtualization finds its greatest value not as the abstract mapping layer that allows applications and data to be coupled more loosely than in decades past. Rather, its most pragmatic operational value is in the better utilization of storage and the transparency of adding or retiring storage without a system outage. Its long-term value comes in the way virtualized constructs (such as proxies) and data, together with careful design, can make some basic storage utilities intrude less and scale better. Of course, these qualities must be accompanied by the understanding of and integration with key software elements (applications, hypervisors, and all the various utilities that support them) to provide the right data fully to where it is needed.

Good design in virtualization is often a matter of granularity. Consider the chunks that are used to capture changed data. The bigger the chunks, the more the data that must be replicated and moved – including some that need not be; FalconStor NSS's *MicroScan* supports transmission of *only the changed sections of blocks* – smaller chunks than their competition.

On the deduplication front, FalconStor offers source level (File-interface Deduplication System) as well as target deduplication (using *Single Instance Repository* or *SIR*). The more duplicates (attachments, files, table structures, etc.) that can be identified, the more effective the deduplication is. The ability to dedupe again at a finer, block-level and sub-block level granularity, done after file-level deduplication has taken place will further reduce the data.

The efficiencies given by fine chunking and deduplication are not just a matter of *storage not used*, but of *bandwidth not consumed*, and *time not spent in data transmission* (hence, faster restores) – a matter of space, time, and ultimately money. It is not about doing more or doing less, but about doing better and doing smarter.

2. Componentization: Deftness through Offload

When you are talking data replication, componentization is, by its nature, a limited choice of intimate relationship, for replication is, by its nature, a linear process. To leverage the benefits of the granularity described above, the backup process must know more, yet the topography of today's server-consolidated world gives few omnipotent venues for the replication processes. Backup of virtual machines to virtual storage arrays is an excellent design challenge to consider, for it places stringent requirements on the data protection process.

In virtualized, consolidated server environments, traditional host-based software is not aware enough of what is going on in the virtual machines, just as the European nations and their colonies back four centuries ago. Better awareness is needed to evoke snapshots of all that is necessary at a simultaneity that will allow quick restore with no data loss.[1] As it has developed to support more guests and more sophisticated resource sharing, the *ESX Server* hypervisor from VMware (as an example) does not have the spare cycles to support additional functionality without performance impairment. As virtual machines get more sophisticated and do more, this situation will only grow worse. Therefore, putting the replication software on the ESX Server gives visibility into the cost of intrusion. How to protect burgeoning data subject to burgeoning usage becomes a grand challenge.

Let us assume a VMware environment – others are similar, though most are less complex[2]. VMware's ESX Server must be instructed to quiesce the engine and the data (VMDK file) of a guest application for a snapshot to be taken. All the data pointers have to be captured simultaneously for the environment to be fully and easily recoverable. However, to replicate the data from within the ESX Server will consume processor cycles, impairing ESX Server functionality, and, perhaps, causing a flurry of dysfunctionality on the guests.

FalconStor's *HyperTrac Backup Accelerator for VMware Consolidated Backup* (VCB) software[3] provides a differently designed solution. It works with VMware *Virtual Center* and ESX Server instances to trigger a snapshot,[4] quiescing both the application and its

[1] This is not a new problem, but the granularities at which systems now function make the server no longer a device but a solar system, if not a universe.

[2] IBM's z/VM is part of a different management scenario entirely.

[3] VCB is usually implemented on a storage gateway or array, together with an instance of VMware ESX Server.

[4] This works with any application running on *Windows* or *Linux* File Systems

file system (VMDK) briefly. FalconStor's HyperTrac for VCB software (not the production VM ESX Server software) then moves the sets of pointers and the data to a *VCB Proxy* for backup. This takes the process out of the way of VMware operations and leaves them unimpaired.

The proxy is, essentially, a shadow of the production environment – but one that consumes few resources, since there are no production reads and writes to it. As changes are posted in ESX logs, they are fed to the VCB Proxy, which keeps the shadow up to date. When a recovery is needed, it is done from the VCB proxy instead of using a somewhat-stale snapshot and the resyncing with log files. The VCB proxy also becomes a source for testing and auditing, as well as for back up and recovery. With this scenario, there are no backup agents on the VMware production virtual machines or on VMware ESX server. The solution is less intrusive to business operations. Recovery is more straightforward. It can be either from cache or from disks controlled by FalconStor software, obviating the need to be restored to a local host's internal storage. This is where the breadth and depth of FalconStor's products comes into play.

FalconStor Products

FalconStor leverages the following design points (components) throughout its set of solutions.

FalconStor Network Storage Server (NSS)

FalconStor's *Network Storage Server* has been managing storage virtualization since the turn of the Century. The Version 6.0 *Virtual Appliance for VMware Infrastructure* provides access to shared, virtualized storage, and the integration with VMware's resilience products such as *VMotion.* It is certified by VMware as a *Storage Virtual Device*.

This appliance leverages the granularity and the proxy approach to replication described above, in addition to the benefits of traditional Thin Provisioning,[5] to provide LAN-free access to data for processing and granular recovery. NSS is also certified with *Windows Server 2008 Failover Clustering* and *Hyper-V* virtual server environments. It supports Fibre Channel, iSCSI and Infiniband protocols.

In other countries, FalconStor has been selling its software both on a gateway (user- or distributor-supplied) appliance and as part of an array. Now, FalconStor is offering the *NSS HC series* arrays in the United States. It features dual active-active controllers, up to 336 drives, and up to 32 GB of cache. Long a champion of iSCSI, FalconStor's arrays support a mix of Fibre Channel (2 and 4 GB) and iSCSI protocols in the same box. *The NSS HC Series* supports a mix of SAS and SATA drives but in the same shelf, though mapped to different volumes. This gives small data centers, as well as branch and home offices great flexibility to meet their needs with well-designed data replication and storage virtualization.

FalconStor Continuous Data Protector (CDP)

FalconStor's *Continuous Data Protector* can support frequency of protection (replication) ranging from synchronous mirroring to asynchronous transmission of changes. It now supports Microsoft's *Windows 2008* and *Hyper-V*, and *Linux* platforms.

FalconStor CDP captures writes at a fine level of granularity. This affects how much storage is used. The software is aware of both the application and the file system and, by use of the shadow proxy strategy described above, supports rapid recovery even in environments with many instances of an application on many virtual machines. This is more than just competing in the hype-battle on compression-ratios.[6]

FalconStor HyperTrac Backup Accelerator for VMware Consolidated Backup (VCB)

HyperTrac for VCB focuses on the challenge of virtual environments. Its solution, described earlier in the design section of this paper, epitomizes the scalable design that underlies FalconStor software. HyperTrac for VCB is a cornerstone for the management of data protection in highly virtualized environments.

FalconStor Virtual Tape Library (VTL)

FalconStor VTL is a legacy name for contemporary enterprise backup. It was designed

[5] Thin Provisioning is the practice of virtually provisioning applications with the capacity of storage they are programmed to expect (usually very generous), while physically provisioning it only as needed.

[6] FalconStor feels it does well there, too – but the proof is in the pudding and generalizations of what to expect are not actionable.

originally as a point of aggregation to better feed and utilize physical tape libraries – a function that persists for industries like media and entertainment, high-performance computing, and others with a need for huge low-cost archives of massive data.[7] For those with merely huge amounts of data, the VTL has become a popular back-up target, as it is optimized for throughput.

FalconStor's granularity and pre-synchronized restorable image are of obvious value in this role. Accessorized with this provisioning and clustered block-level deduplication,[8] the VTL can support a single instance repository. Such a repository is the Holy Grail of storage cost avoidance.

FalconStor VTL understands the formats used by classic enterprise backup products. It supports Copan's *MAID* (*Massive Array of Idle Disks*) architecture. It can be acquired as software only, a turnkey appliance, or integrated into a Virtual Appliance for VMWare Virtual Infrastructure.

Conclusion

Utility processes, such as storage virtualization and data replication, become killer apps in a very negative sense if they are not deft and unobtrusive. They limit how much work a system can process. They create bottlenecks that can be moved – but not removed. We need infrastructure utilities to be like the butler who facilitates a life of seamless ease for the important folks (like the A–Team, mentioned earlier).

Enterprises are finding new ways to leverage their data to optimize both operations and strategies. More gregarious routes to market and new detailed documentation of business, such as click-stream and sensor data, enrich and clutter the corporate data landscape. If ever there was a need for deft butlers to keep things civilized and productive, it is now.

The more you ask of your infrastructure, the more that quality counts. This is true of workforces, and it is also true of the software utilities that support the data that supports your business. It is even truer of the utilities that support the vast scale of cloud environments, where the expectations are higher.

The software you use to ply your business must be secure. It must be bug-free. Now, with the ballooning of business data, it must also be scalable. This is a matter of design, not just of maintenance. Look for utilities that are well designed. Take a close look at FalconStor's solutions.

[7] Environmental reform will drive repositories of this type.

[8] That this block-level deduplication is offered as a post-process is another example of the non-disruptive approach of the shadow-proxy discussed earlier.

About The Clipper Group, Inc.

The Clipper Group, Inc., is an independent consulting firm specializing in acquisition decisions and strategic advice regarding complex, enterprise-class information technologies. Our team of industry professionals averages more than 25 years of real-world experience. A team of staff consultants augments our capabilities, with significant experience across a broad spectrum of applications and environments.

- ***The Clipper Group can be reached at 781-235-0085 and found on the web at www.clipper.com.***

About the Author

Anne MacFarland **is Director of Data Strategies and Information Solutions for The Clipper Group.** Ms. MacFarland specializes in strategic business solutions offered by enterprise systems, software, and storage vendors, in trends in enterprise systems and networks, and in explaining these trends and the underlying technologies in simple business terms. She joined The Clipper Group after a long career in library systems, business archives, consulting, research, and freelance writing. Ms. MacFarland earned a Bachelor of Arts degree from Cornell University, where she was a College Scholar, and a Masters of Library Science from Southern Connecticut State University.

- ***Reach Anne MacFarland via e-mail at Anne.MacFarland@clipper.com or at 781-235-0085 Ext. 128. (Please dial "128" when you hear the automated attendant.)***

Regarding Trademarks and Service Marks

The Clipper Group Navigator, **The Clipper Group Explorer**, **The Clipper Group Observer**, **The Clipper Group** *Captain's Log*, **The Clipper Group Voyager**, **Clipper Notes**, and *"clipper.com"* are trademarks of The Clipper Group, Inc., and the clipper ship drawings, *"Navigating Information Technology Horizons"*, and *"teraproductivity"* are service marks of The Clipper Group, Inc. The Clipper Group, Inc., reserves all rights regarding its trademarks and service marks. All other trademarks, etc., belong to their respective owners.

Disclosure

Officers and/or employees of The Clipper Group may own as individuals, directly or indirectly, shares in one or more companies discussed in this bulletin. Company policy prohibits any officer or employee from holding more than one percent of the outstanding shares of any company covered by The Clipper Group. The Clipper Group, Inc., has no such equity holdings.

Regarding the Information in this Issue

The Clipper Group believes the information included in this report to be accurate. Data has been received from a variety of sources, which we believe to be reliable, including manufacturers, distributors, or users of the products discussed herein. The Clipper Group, Inc., cannot be held responsible for any consequential damages resulting from the application of information or opinions contained in this report.