HITACHI
Inspire the Next

Hitachi Dynamic Replicator Software Recovery Solutions for Virtual Server Environments

Data and Application Recovery Considerations When Deploying Virtual Servers

Application Brief

February 2010

Executive Summary

As server virtualization technology penetrates more into the mainstream, hosting more mission critical applications, enterprises will need robust data and application recovery tools that can meet increasingly stringent recovery requirements, while at the same time supporting a common recovery management paradigm across heterogeneous environments. This trend will continue to spread thanks to the increase in utilization ratios for computing resources that enterprises begin to observe as they consolidate their physical servers on virtual servers; improved usage is one of the single biggest advantages to server consolidation.

However, the issue of matching virtual server recovery performance to that which has been demanded of physical machine environments remains. The solutions that provide the most value for server virtualization environments will offer a common recovery management paradigm, centrally managed, that applies to all the hardware and software comprising both the physical and virtual server infrastructure.

This application brief examines the benefits of server consolidation in virtual server environments, its challenges, and the compelling solutions from Hitachi Data Systems, notably Hitachi Dynamic Replicator software, which meet stringent recovery requirements in these environments.

For best results use Acrobat Reader 8.0

Contents

Hitachi Dynamic Replicator Software Recovery Solutions for Virtual Server Environments

Data and Application Recovery Considerations When Deploying Virtual Servers

Application Brief

Introduction

As virtual servers are used more and more to host critical application environments, the tools to manage them must become more robust and, therefore, recovery issues are coming to the fore. A lack of recovery capabilities that are at least equal to those available in physical server environments slow the rate at which virtual servers penetrate the enterprise. There are some significant differences between physical and virtual server environments that have acute impacts on how data protection and disaster recovery solutions can be deployed. Just blindly deploying recovery approaches from physical server environments into virtual ones will have negative impacts on application performance and network bandwidth utilization. Solutions that are designed to take advantage of shared infrastructure and common deployment practices in virtual server environments, however, can address these issues while at the same time offering improved recovery performance. This application brief will discuss the data and application recovery considerations you need to take into account when deploying virtual servers, and then briefly summarize how solutions from Hitachi Data Systems provide compelling local and remote recovery capabilities in these environments.

Today's Virtual Server Environments

Virtual server technology is taking hold in the enterprise because it represents the opportunity to consolidate large numbers of separate, physical servers onto a shared infrastructure that requires less server hardware, supports a shared administrative model that can save significant management costs and provides much more flexibility in scaling and maintaining computing and storage resources. Virtual server deployments that leverage networked storage — either storage area network (SAN) or network attached storage (NAS) — help support that flexibility by making it faster and easier to migrate existing applications and their associated virtual servers to new hardware as well as to provision new servers and storage.

Today's virtual server environments host a number of different operating systems. Heterogeneity is a fact of life, reflected not only in the disparate server and storage hardware, but also more and more in the server virtualization software. In the long run, it is likely that server virtualization technology from VMware, Citrix and Microsoft will co-exist within the same IT shop. As server virtualization technology penetrates more into the mainstream, hosting more mission critical applications, enterprises will need robust data and application recovery tools that can meet increasingly stringent recovery requirements, while at the same time supporting a common recovery management paradigm across heterogeneous environments. And, keep in mind that it will be a rare company that will move entirely to virtual servers in the next five years. Most companies will have some mix of physical and virtual servers, with a common set of recovery problems across both that need to be addressed.

Server Consolidation: The Double Edged Sword

The difficulty of provisioning new resources in purely physical environments has led in the past to the deployment of new applications and new physical servers on a one-to-one basis in many cases. Because of the difficulty of adding resources to physical servers, resources on them were generally overprovisioned by a large margin to accommodate anticipated growth in the application environment. Often only 15 percent to 20 percent of the server's resources were being used at initial deployment. It is a reality that many of these physical servers remained woefully underutilized over time, leading to a huge opportunity for improvement in utilization ratios as virtual servers were deployed. Virtual servers are much easier to provision and reconfigure than physical servers, which allows for them to be sized much more appropriately, meeting real time application requirements and removing this "overallocation" waste. Resources can be easily added as applications grow. As multiple physical servers are consolidated using virtual server technology, utilization ratios for computing resources go way up — both for the virtual servers and the physical servers on which they run. This reality represents one of the single biggest advantages to server consolidation: the ability to better utilize your existing resources.

However, this has a downside that is particularly evident when it comes to data protection. Backup is an operation that puts a heavy load on a server, often demanding more resources for a short period of time than are demanded by an application operating at even peak loads. When multiple physical servers, all of which may have been only 20 percent utilized, are consolidated into the same number of virtual machines running on a single physical server, each virtual machine is generally run at a utilization ratio approaching 100 percent. The headroom that was available before to handle "peak processing" requirements like backup is no longer available, and the result is poor performance — both for backups and for applications — during online backup operations. And, if administrators are performing backups against several virtual machines, all of which reside on a single physical server, at the same time, network bandwidth will be another limiting factor impacting backup performance. Many administrators are surprised to find that, when they just move their existing data protection processes over to a virtual machine environment, they are experiencing significantly poorer performance.

Recovery Solution Requirements in Virtual Server Environments

The realities of virtual server environments — very high resource utilization ratios at the virtual and physical server level and heterogeneity in servers, server virtualization technologies and storage — translate to a specific set of requirements for data and application recovery solutions. In these environments, traditional data protection and disaster recovery solutions do not translate well because they are too resource intensive (on production servers) and often cannot offer the level of heterogeneous support required. And, note that solutions that are designed specifically for one type of server virtualization environment, or only for server virtualization (and not physical) environments, will ultimately end up being not comprehensive enough.

Given that virtual server environments will be hosting more mission critical applications as time goes on, recovery requirements will be stringent. Discrete backup operations (where backups are performed once or more per day as separate tasks) put significant resource load on virtual servers that, given consolidation efforts, can cause much worse performance problems than backups do in physical machine environments. Conventional tape-based infrastructures that are already unable to cope with recovery requirements in physical environments will perform even worse in virtual server environments.

When it comes to disaster recovery, the traditional in house approach has been to replicate from physical servers (or their associated storage) in the local site to physical servers (or their associated storage) at a remote site. This basically requires twice the amount of server and storage hardware, which is one of the reasons that disaster recovery has been such an expensive proposition in the past. Server virtualization

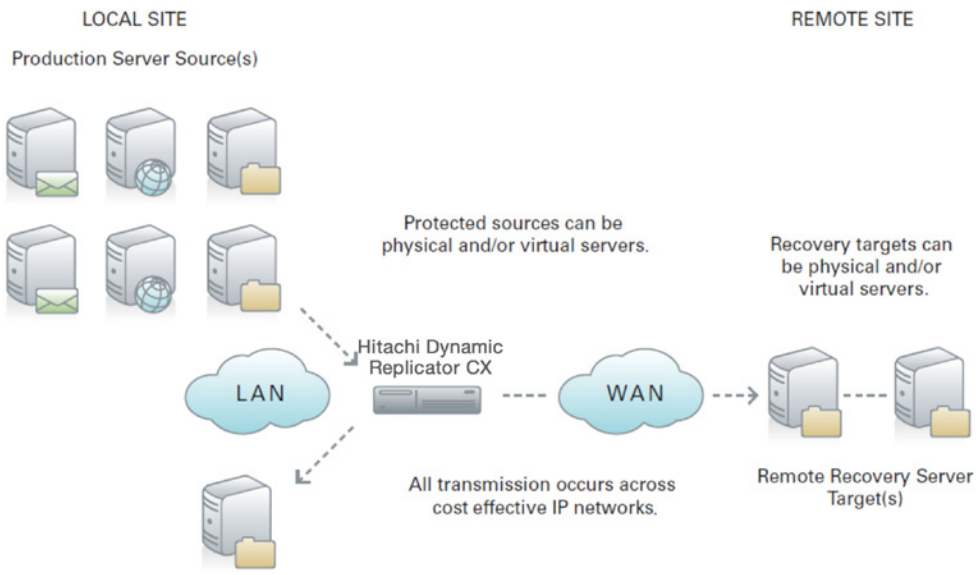
technology can have a significant impact on lowering costs here, however, by supporting replication between physical and virtual servers or just between virtual servers at the local and remote sites.

Taking these issues into account, it is clear that virtual server environments require recovery solutions that will have to match the stringent recovery performance requirements demanded in physical machine environments, but they will have to do so in a way that puts much less load on servers to perform data protection operations. It is also clear that any solution must support heterogeneity, not only in server and storage hardware, but also in virtual server technologies and applications. The consolidation strategy that has proved so cost effective on the hardware side should be extended to the recovery side as well. The solutions that provide the most value for server virtualization environments will offer a common recovery management paradigm, centrally managed, that applies to all the hardware and software comprising both the physical and virtual server infrastructure.

Hitachi Dynamic Replicator: Well Matched Recovery Solutions for Virtual Server Environments

Hitachi Dynamic Replicator software offers a single platform that can support transparent backup, disaster recovery from catastrophic failures within minutes and automated application failover/failback to support faster, more reliable restoration of application services. Based around next generation data protection technologies like continuous data protection (CDP), asynchronous replication, recovery automation and wide area network (WAN) optimization, Dynamic Replicator allows administrators to completely dispense with backup as a discrete operation; they can move to a "continuous backup" paradigm that not only enhances existing backup windows and supports very granular recovery capabilities but also provides the industry's most WAN efficient disaster recovery solution. This approach protects data as soon as it is created, a very important consideration for organizations with stringent recovery point objective (RPO) requirements and provides extremely rapid, reliable recovery for organizations with stringent recovery time objective (RTO) requirements. The Hitachi Dynamic Replicator architecture is well suited to meet the recovery requirements of virtual server environments for several reasons (see Figure 1).

Figure 1. Hitachi Dynamic Replicator offers a software-based solution that is well suited to meet both the local and remote recovery needs of virtual server environments.

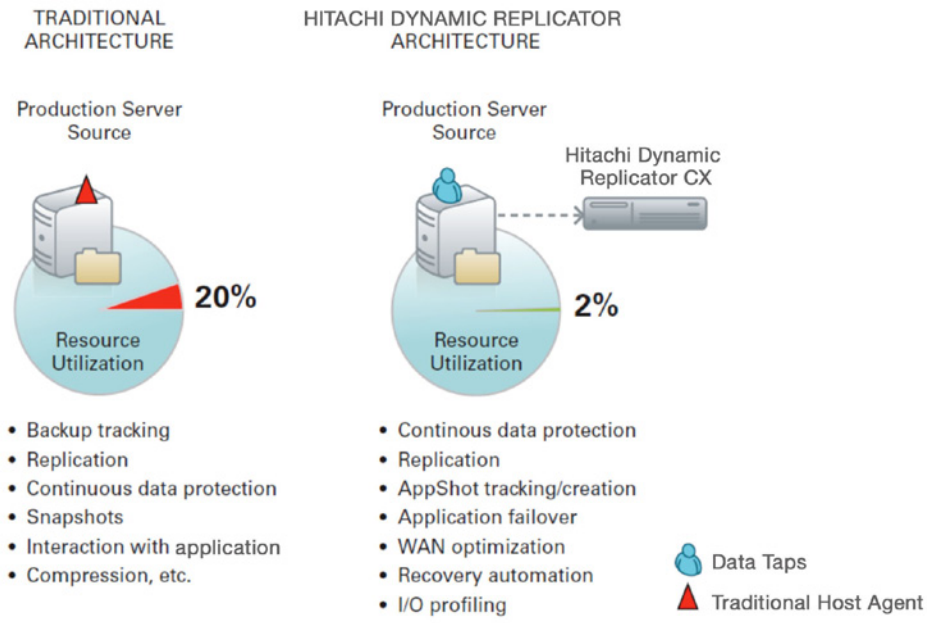


Low Overhead

Traditional backup and/or disaster recovery approaches often place software on the server(s) that will be protected, and that software can take up to 15 percent to 20 percent of the resources of that server during operation. When significant server headroom was already available, this was not much of a problem, but it is a significant issue in virtual server environments.

Hitachi Dynamic Replicator uses a unique hybrid recovery architecture that effectively offloads all the recovery processing to a network-based, dedicated server called the CX. Any host-based components that were being used in the past for backups (like backup agents) or disaster recovery (like replication agents) are removed and replaced with a single, very lightweight filter driver that Dynamic Replicator calls a “data tap.” The data tap is a simple, kernel level component that performs one task: it splits off writes in real time as they occur, sending a copy of the original write across the local area network (LAN) to the CX. Because of its simplicity, the data tap imposes at most one percent to two percent overhead on a virtual server — much less than the 15 percent to 20 percent that can be imposed by application agents (see Figure 2).

Figure 2. The inherent weakness of host-based approaches is that richer functionality has a bigger hit on production application performance. Hitachi Dynamic Replicator software’s model allows functionality to be added without impacting production performance in any way.



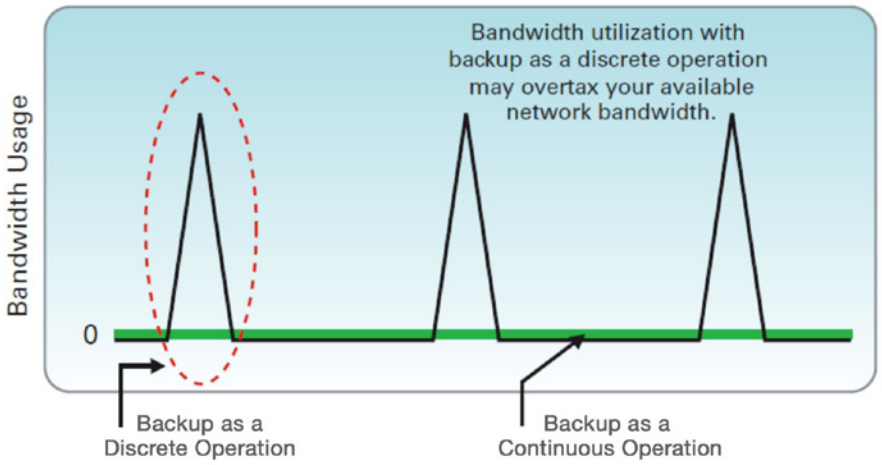
Despite its minimal impact on protected servers, this approach offers very rich functionality. The CX is effectively a dedicated appliance whose resources are used to drive this functionality — adding more functionality does not impact production server resources. Dynamic Replicator software’s rich feature set includes five foundation technologies: CDP, asynchronous replication, application failover/failback, WAN optimization and recovery automation.

CDP

Hitachi Dynamic Replicator software’s CDP functionality effectively moves organizations away from the concept of backup as a discrete operation that impacts protected servers and results in significant spikes in network bandwidth requirements by moving to a continuous approach, which captures data in real time as it is

created. Dynamic Replicator captures only the data changes, a factor that minimizes the amount of data that has to be sent across the network, and the transmission of those changes is spread across 24 hours each day. Compare that to the legacy approach, which bundles up all the data that has to be sent across the network (either the full backup or, in the case of incrementals or differentials, just the changes) and attempts to send it all at once. The traditional approach puts a huge load on the network during backup, whereas the Dynamic Replicator approach results in very low impacts on protected servers and network bandwidth that in most application environments are negligible (see Figure 3). The use of CDP results in low overhead on CPU, memory and network bandwidth resources — all features that are critical for recovery solutions running in virtual server environments.

Figure 3. The difference in network overhead between the traditional and the Hitachi Dynamic Replicator approaches is significant, with Dynamic Replicator requiring far less network bandwidth.



Asynchronous Replication

Replication only sends “changed” data across networks, so it requires a lot less bandwidth than “backups,” which send an entire backup data set across the network for each backup. And, because Hitachi Dynamic Replicator combines it with CDP, the solution supports an infinite number of potential recovery points, unlike replication by itself, which would only support the latest recovery point; with just replication, if that latest point is corrupt, then there are no other recovery options. Even compared to incremental or differential backups, which do send less data across networks during backups, the Hitachi Data Systems approach supports a much faster, less risky recovery paradigm. That’s because not only can this approach access an infinite number of recovery points, but it can also access them immediately (without having to “build” the recovery set from multiple tapes) and directly from disk.

Application Failover/Failback

Traditional approaches to automated application recovery for physical servers use shared disk architectures and separate availability clustering software products. Virtual server environments offer some optimizations here, leveraging networked storage and administrative tools designed to migrate servers and/or storage to separate physical hardware, but these tools are not integrated with data recovery tools and tend to be specific to a particular virtual server technology. Dynamic Replicator offers a single integrated recovery solution that

covers both data and applications, and it can be used to set up partially or fully automated recovery procedures (depending on administrative preferences) for both that apply to local as well as remote recoveries. And because Dynamic Replicator uses a “shared nothing” disk architecture (instead of the shared disk architecture used by availability clustering software) combined with CDP, it provides recovery options even in cases of data corruption (something traditional failover cannot address). These automated recovery procedures can be set up to run across heterogeneous server and storage environments, all the key enterprise applications (including Microsoft® Exchange, SQL Server® and SharePoint® as well as Oracle, MySQL, Blackberry Server, any file systems, etc.), and all three server virtualization environments (VMware, Citrix, Microsoft).

WAN Optimization

WAN bandwidth tends to be more limited than LAN bandwidth, so for the WAN hop Hitachi Dynamic Replicator uses a number of different WAN optimization technologies to keep the amount of data sent across the WAN to an absolute minimum. Dynamic Replicator leverages transmission control protocol (TCP) optimization, various types of compression and bandwidth shaping (quality of service policies established by the administrator) to tailor the flow of data across the WAN, ensuring not only that minimal data is sent across the network to support recoverability but also that the solution will play well with other applications using the WAN.

Recovery Automation

Recovery automation is used to shorten recovery times, in particular for failed application services, and to reduce the risk associated with recovery operations. Conventional data protection approaches focus on the data, leaving application recovery up to administrators. Manually rebuilding servers in the event of a disaster can be very time consuming and fraught with risk, particularly if they are multiserver applications like SharePoint or SAP, whereas the use of automated recovery procedures will allow those same application services to be brought back up quickly, with less risk and with minimal dependence on the skill set of the administrator. Recovery automation, particularly when applied in conjunction with virtual server technology, can also be used to support more frequent testing of disaster recovery plans. Traditional disaster recovery testing is disruptive and manually intensive, putting a huge impact on production operations. Using Dynamic Replicator’s AppShot and recovery automation technologies with virtual server targets at the disaster recovery site, disaster recovery plans can be tested without impacting production environments and with minimal staff and costs. AppShots can be selected at the remote site and mounted on targeted recovery servers, applications can be restarted and application services can be verified — all without impacting production applications at the local site at all. This supports the type of “low impact” that will help make more frequent disaster recovery testing viable, and more frequent disaster recovery testing leads to higher confidence and more predictable recovery from various disaster recovery scenarios. In addition, it uses a common set of processes, centrally managed, that can be used across any and all server virtualization environments.

Interestingly, Dynamic Replicator also uniquely offers an I/O profiling tool. Administrators can use this tool to determine exact bandwidth requirements to meet defined RPO/RTOs prior to deployment. After deployment (and during normal operation) it can be used to track and instantaneously monitor RPO capabilities in real time (based on available network bandwidth, I/O write rates and tracking, where data is as it is captured on the protected server, sent to the CX and then from there to a recovery target) for trend analysis and capacity planning purposes.

Flexibility

The use of CDP completely changes the traditional model for protecting data, but it also supports much more granular recovery capabilities that effectively make data recoverable the instant it is created (rather than the instant it is backed up). As writes are captured, they are written into a time stamped log managed by the CX (although physically residing on a target server). This allows the administrator to retroactively select any previous point in time to use as a recovery point. That time might be the most recent point in time, it might be

a point that is tagged as marking an application consistent recovery point (what Hitachi Dynamic Replicator calls an “AppShot”), or it might be the latest point in time right before data corruption occurred about 20 minutes ago. What is interesting about this technology is its flexible ability to meet any RPO requirements, from the most lax to the most stringent, using the same data capture approach across all applications (regardless of whether they are hosted on a physical or a virtual server). Because Dynamic Replicator offers a common model across heterogeneous environments — server, virtual server and storage it supports much lower administrative overhead than approaches that are locked in to specific physical servers and operating systems, virtual server environments, storage, storage architectures or applications.

Leveraging Dynamic Replicator software’s built in CDP technology, this Hitachi Data Systems recovery solution can instantly generate one or more physical or virtual, disk-based copies of any retroactively selected application state at any time, without impacting production in any way. An AppShot, once created, can be used for any number of administrative operations, depending on the server on which it is mounted. Throughout the course of a day, an administrator may create an AppShot reflecting last quarter’s financial close for reporting purposes, then an AppShot of what the Oracle or SQL database looked like at 7 a.m. this morning to load into the data warehouse, then a copy of what the Exchange server looked like at 11 a.m., right before a patch was applied; therefore, if the patch causes problems, it is very easy to immediately revert back to a recent, working state for Exchange. The flexibility of this approach allows administrative operations to be time shifted to achieve optimal productivity, and it is clearly not available in recovery products that do not leverage CDP (and not even in some that do, depending upon how they are architected). Note that all of these operations can be performed while still fully preserving Dynamic Replicator software’s flexible recovery capabilities.

Dynamic Replicator’s approach is perfectly suited to providing recovery capabilities for application environments that have different RPO/RTO requirements as well. The data capture method is the same regardless of what those requirements are, making deployment across different environments uniform. Simply eliminating backup jobs and moving to a much less resource intensive “continuous” approach, can provide better protection at lower cost and with less overhead.

When recovering virtual servers, Dynamic Replicator can support not only the restoration of data files but also the “bare metal restore” capabilities, which can be so helpful in disaster recovery scenarios where full system restores must be performed.

Dynamic Replicator’s heterogeneous support also provides significant flexibility that leads directly to cost savings. It preserves existing investments in hardware and software, while providing maximum freedom of choice in new server, storage and application purchases going forward. It supports using different types of storage and/or storage architectures [SAN, NAS, direct attached storage (DAS)] simultaneously in different locations. This type of flexibility translates directly into more cost effective deployments.

Note also that Dynamic Replicator appliances, which are dedicated x86 servers running Dynamic Replicator’s CX software, can be deployed on either physical or virtual servers. This provides additional deployment flexibility, which can be used to streamline management and lower deployment costs.

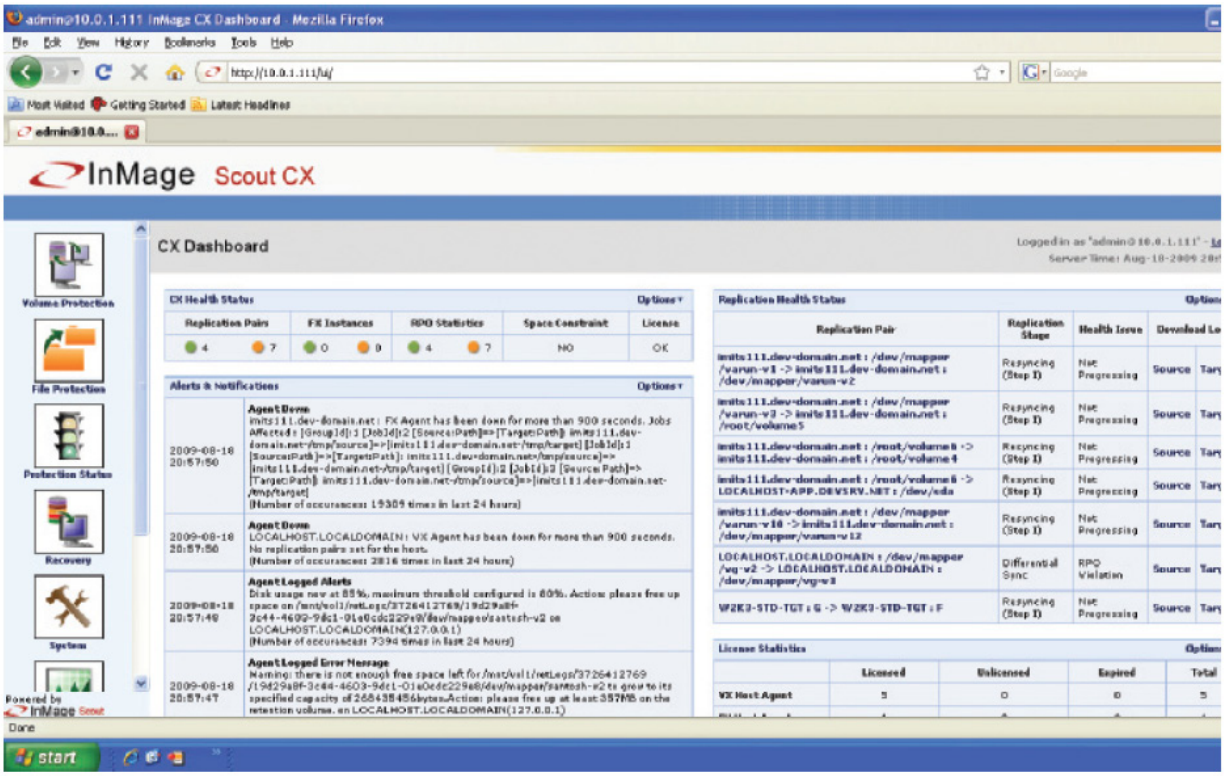
Centralized Management

A single CX supports one or more protected servers, depending on I/O write rates and the size of the server that hosts the CX software. The CX software can be installed on smaller x86 servers to cost effectively support smaller configurations, or larger servers to support higher performance. The CX can be replaced (for either maintenance or upgrade purposes) without impacting the availability of production applications in any way.

Because the data streams of multiple protected servers are flowing through a single CX, it becomes a focal point for tracking the I/O rates of various servers. The I/O profiling capability mentioned earlier allows monitoring of the I/O load from each protected server to determine how those I/O loads change over the course of the day or over time. As configurations scale, additional CXs can be added, and those can all be

managed centrally from a single, Web browser–based management graphical user interface (GUI). This allows centralization of all recovery management operations, for all infrastructure (heterogeneous servers, storage and different server virtualization technologies and applications) with a single integrated recovery solution that covers both local and remote scenarios.

Figure 4. Hitachi Dynamic Replicator supports an easy to use GUI, as well as a more powerful command line interface.



Cost Savings

One of the primary drivers of server consolidation is cost savings, and Hitachi Dynamic Replicator contributes strongly to that in virtual server environments. Most agents or other software components that are being used to gain access to application data on production servers can effectively be replaced with this simple, light-weight filter driver, since application consistent images (AppShots) of any application data set can be created at any time and mounted on any network attached server for any reason, whatsoever. In environments with 10 or more servers, the savings add up quickly due to license and maintenance fee savings. For example, agents tend to be more complex products that often have to be patched or otherwise upgraded when new hardware or software releases come out, and these maintenance operations all cost money and impact the availability of production applications. Dynamic Replicator software's data tap rarely needs to be updated for new hardware or software releases due to its simplicity, promoting lower costs and overall higher availability.

When deploying disaster recovery solutions, using virtual servers as recovery servers at remote locations significantly lowers hardware infrastructure costs. This may not support the same levels of performance that could be achieved if dedicated physical servers were used to mirror production configurations at primary sites, but it can certainly support the restoration of application services at remote sites until primary sites can be brought back online. And the savings will be significant, since multiple physical servers at a primary site can be "mirrored" with multiple virtual servers running on a much smaller number of physical servers at the remote site.

Summary

The realities of server virtualization deployments demand recovery solutions that can meet a different set of requirements than those traditionally posed by physical server environments. Solutions must be comprehensive in that they support both data and application recovery either locally or remotely. As more mission critical application environments migrate to virtual servers, recovery solutions will need to be able to meet more stringent recovery requirements, and they will be evaluated across metrics like impact of data protection operations on production environments, ability to meet RPO/RTO requirements and recovery reliability. The use of disk-based recovery technologies offers significant advantages in any setting that has stringent recovery requirements, not only because of its advantages over tape for RTO and recovery reliability performance, but also because it gives access to other disk-based technologies. These technologies, such as CDP, asynchronous replication and WAN optimization, can eliminate backup impacts on production environments, meet the most stringent RPO requirements and obviously minimize or in some instances entirely eliminate tape handling.

Given its use of foundation technologies like CDP, asynchronous replication, application failover/failback, WAN optimization and recovery automation, Hitachi Dynamic Replicator software fully meets these requirements. Dynamic Replicator uses a unique hybrid recovery architecture that offloads all data protection processing from production servers, deftly handling the very high utilization ratios of virtual servers that cause significant problems for conventional data protection or disaster recovery approaches. Its data capture and data movement methods minimize the amount of data that has to be sent across LANs or WANs to enable recovery at either local or remote locations, resulting in network bandwidth requirements significantly below standard backup and/or replication product requirements.

Dynamic Replicator software's flexibility, not only in supporting heterogeneous environments but also in its ability to partially or fully automate recovery operations on an application-by-application basis, contributes to a comprehensive recovery solution that can be used across physical, virtual and mixed physical/virtual environments, as well as provides a common recovery management paradigm across all three major server virtualization environments (VMware, Citrix, Microsoft). Its uniform deployment methodology, regardless of the application or recovery requirements, simplifies the installation and ongoing management of the solution, particularly when it is compared to other solutions, which are specific to one type of server virtualization environment, a particular server or storage device, or a particular application. And, when all the savings — in backup and/or replication agents, minimization or elimination of the effects of conventional data protection technology (backup windows, data loss on recovery, lengthy recovery times,
and poor recovery reliability), lowered administrative overhead, ability to support existing hardware and software, and lowered bandwidth requirements — are taken into account, it provides a very cost effective solution as well.

Hitachi Data Systems Corporation

Corporate Headquarters 750 Central Expressway, Santa Clara, California 95050-2627 USA
Contact Information: + 1 408 970 1000 www.hds.com / info@hds.com

Asia Pacific and Americas 750 Central Expressway, Santa Clara, California 95050-2627 USA
Contact Information: + 1 408 970 1000 www.hds.com / info@hds.com

Europe Headquarters Sefton Park, Stoke Poges, Buckinghamshire SL2 4HD United Kingdom
Contact Information: + 44 (0) 1753 618000 www.hds.com / info.emea@hds.com