

Dispelling the Myths of Data Replication

Volume Management and Other Hidden Costs of Array-based Replication

July 19, 2004

TABLE OF CONTENTS

<i>Data Replication</i> _____	3
<i>VERITAS Volume Replicator Replication Overview</i> _____	3
<i>Array-based replication</i> _____	3
<i>Replication Myths Dispelled</i> _____	4
Myth One: I Have Too Many Hosts for VVR _____	4
Myth Two: Host-Based VVR Will Be Expensive _____	4
Myth Three: Host-Based Replication Will Impact My Applications _____	6
<i>Conclusion</i> _____	7

DATA REPLICATION

Information technology professionals today understand that data replication is more effective than traditional tape storage for ensuring business continuity following a disaster. Less widely understood are the differences between VERITAS® host-based replication and array-based replication solutions, and the value each brings to disaster recovery (DR) storage implementations.

VERITAS offers superior manageability, more cost-effective operation, and better application performance than array-based replication products—despite myths to the contrary. These myths are based on the misconception that array-based products bypass the server-level data management aspects of the VERITAS solution. In fact, data storage administrators with array-based solutions routinely perform server-level data management, but without the tools included in the VERITAS host-based solution which greatly simplify that task.

This paper compares the VERITAS and array-based architectures and explains the server-level management requirements of array-based products. Based on that understanding, the myths regarding the manageability, cost, and application performance ramifications of VERITAS replication are explained and dismissed.

VERITAS VOLUME REPLICATOR REPLICATION OVERVIEW

VERITAS offers data replication through the VERITAS Volume Manager™ (VxVM) and VERITAS Volume Replicator™ (VVR). VxVM is an online storage management tool for heterogeneous enterprise computing and storage area network (SAN) environments. Using VxVM, system storage administrators can easily configure, share, add, re-size, and move data volumes between storage systems while business applications run. Administrators manage network-wide data from a single console using the VERITAS command line interface (CLI) or VERITAS enterprise administrator (VEA) GUI.

VVR is an optional component of VxVM that replicates data over a heterogeneous IP network and ensures that the replicated data is accurate and recoverable. VxVM and VVR can replicate data from one server to as many as 32 local or remote locations. One-to-one, one-to-many, and many-to-one data replication configurations are supported, and remote sites can be any distance from the primary server.

ARRAY-BASED REPLICATION

Array-based replication products send I/O directly from the primary storage array to the destination array, using controller-based firmware. Array-based replication is independent of the server, so array-based products are not operating system-dependent. No server-level management is required as long as the complete contents of the array are replicated. However, all the arrays must be identically configured and from the same vendor.

REPLICATION MYTHS DISPELLED

MYTH ONE: I HAVE TOO MANY HOSTS FOR VVR

This myth assumes that the administrative burden of host-based replication products is proportional to the number of application servers. In fact, VVR requires minimal management overhead for each server because LUN and volume mapping between hosts and replicated arrays is fully automated. Administrators control network-wide data replication from a single console, using the java-based VERITAS Enterprise Administrator (VEA) GUI. VEA command sequences can also be incorporated into CLI-based scripts to further simplify management. With these tools, administrators can perform tasks such as resizing a host server volume and mirroring the change across hundreds of other servers with a single click.

These tools dramatically reduce storage management complexity and make VVR far easier to manage than array-based products, which require extensive manual management at the server level. This management is required because array-based products are only host-independent when the entire contents of an array are replicated. Consistently replicating all the data in an array is unnecessary and uses a prohibitively expensive amount of bandwidth. As a result, administrators of array-based replication systems commonly analyze individual servers to determine which volumes need to be replicated and which can be ignored.

Based on the analysis of each server's replication needs, administrators of array-based products manually map data between the source and destination arrays. This time-consuming and error-prone process involves mapping source volumes to the LUNs being replicated on the primary site. These LUNs must then be mapped to target LUNs at the DR site, or the bunker site if the data is being replicated over a long distance. At the DR site, the LUNs must be re-mapped to the host servers used during a disaster. Although LUNs and volumes are dependent on each other there is no direct management relationship that ties them together. Changing volume sizes or moving volumes between arrays requires painstaking rework of the replication configuration, and often requires hosts and operators at both the primary and DR site.

Since array-based products work only at the array level, no tools are provided to ease this management burden. As a result, VVR is far easier to manage than array-based systems, regardless of the number of servers connected.

MYTH TWO: HOST-BASED VVR WILL BE EXPENSIVE

This myth focuses on the cost of the software licenses required for each server when using a host-based replication product. In fact, the excessive storage costs required to support array-based products far exceed the server license cost of host-based products. Array-based products require eight copies of each data set that is replicated (see sidebar). VVR needs only two complete copies plus a small amount of additional capacity for a replication log. Even in large IT environments, the investment in multiple VERITAS server licenses is more cost-effective than the extensive storage costs of array-based products.

Array-based products also require additional hardware investments because identical disk controllers and external equipment to convert disk channel protocols for transmission over WANs and LANs are needed at the primary and secondary sites. This fact locks customers into a single-vendor solution and equally high-end products at the primary and secondary sites. In contrast, VxVM and VVR customers can use low-cost storage at secondary sites because the products support heterogeneous replication.

When all expenditures are compared, VVR is significantly more cost-effective than array-based solutions.

Storage Comparison for VVR and Array-based Replication Products

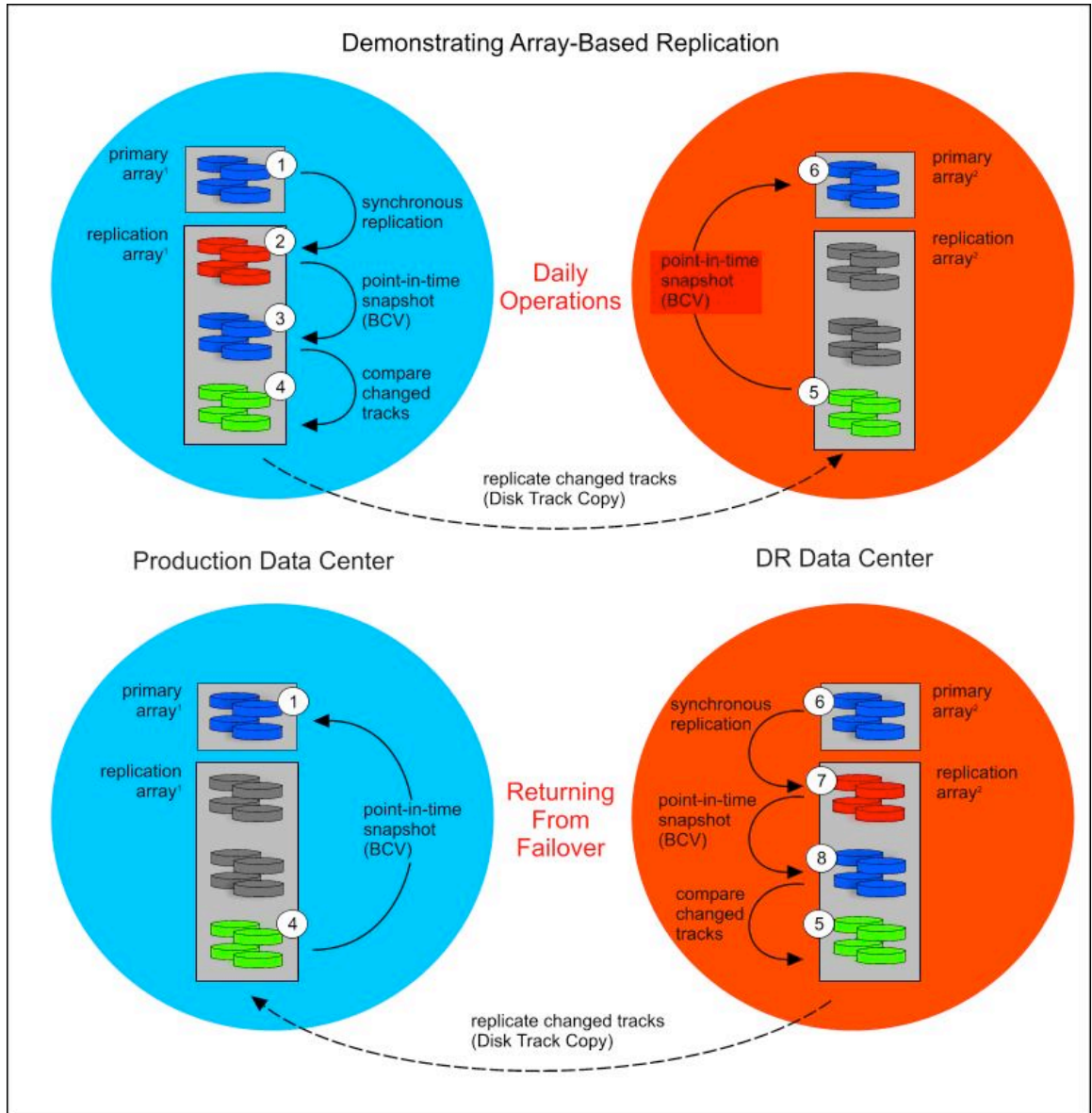


Figure 1: Demonstrating Array-based Replication

Figure 1 shows how array-based products use multiple interim copies of replicated data, resulting in high storage costs. In daily operations, four copies of the data are required at the primary array. The first copy (copy #2) is created in a replication array. This copy backs up the primary array, and allows the primary to continue functioning without the performance degradation caused by the replication process. A snapshot of copy #2 creates a consistent point-in-time data image (copy #3). Firmware compares this snapshot with the previous snapshot, creates a copy of the changed data (copy #4), and sends this data to an identical storage array in the remote data center (copy #5).

Once all of the changed data blocks have been received at the remote data center, copy #5 is identical to copy #3. However, as soon as the process of receiving new changed blocks begins again, copy #5 is no longer identical to copy #3 and is therefore corrupt and cannot be used in a disaster recovery scenario. To avoid this situation, a snapshot of copy #5 is taken (copy #6) as soon as changed blocks have been applied, and this copy is used for disaster recovery if an outage occurs.

Two additional copies of the data are required for an actual recovery scenario. In a disaster, operations shift to processing against copy #6, which replicates to copy #5. Another copy (copy #7) is required at the remote data center to be the target of the snapshot previously performed at the local data center to produce copy #3.

Again, to avoid sending all data across the network, another copy (copy #8) is required to compare with copy #7 in order to determine which data blocks have changed. The changed data blocks are then sent across the network to copy #3, which then performs a snapshot to copy #1. This process requires a total of eight copies of the original data.

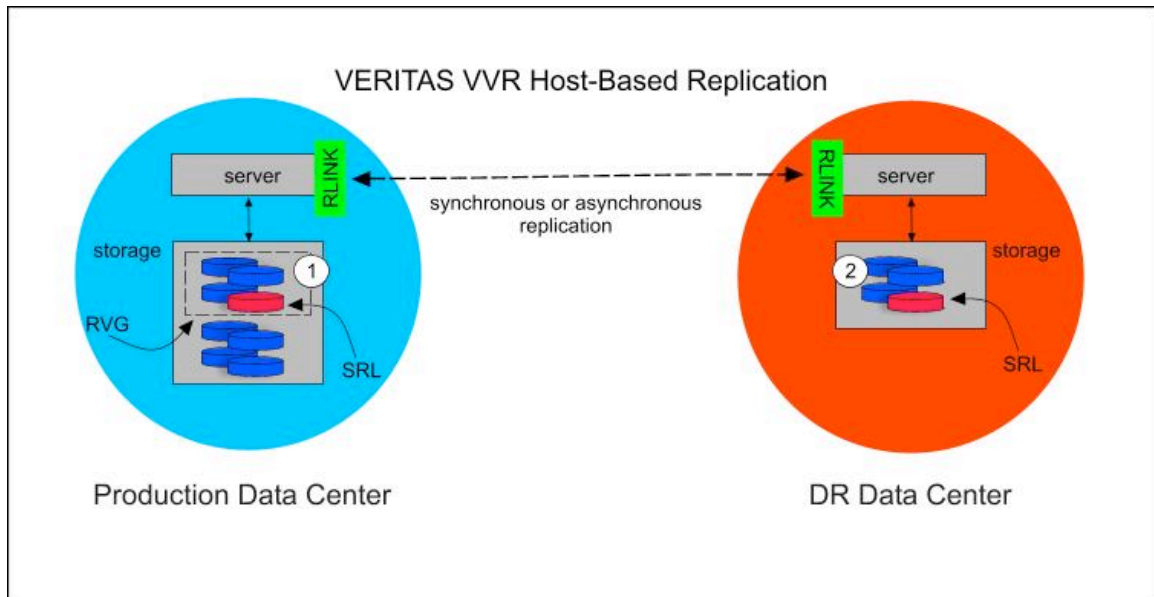


Figure 2: Demonstrating Array-based Replication

By comparison, VVR needs only two complete copies plus a small amount of additional data for their VVR SRL. The local array data (copy #1) is replicated, and changed blocks are simultaneously written to the SRL. The replication writes directly to copy #2 at the remote data center. There is no need for an additional copy to safeguard against an incomplete replication-type corruption because the SRL always contains the information to bring copy #2 back to full data integrity. When comparing these two-plus copies to the eight copies required for array-based products, it is clear that array-based products require almost four times the storage of VVR, which adds significant expense to the implementation.

MYTH THREE: HOST-BASED REPLICATION WILL IMPACT MY APPLICATIONS

This myth assumes that a host-based replication process will steal server resources (CPU cycles, memory, I/O bandwidth, I/O slots) from the resident applications, potentially degrading performance. In fact, testing by Waltham, MA-based GiantLoop Testing and Certification (GTAC) Lab found no support for this claim.

GTAC provides unbiased evaluations of data movement technologies, products, and solutions. In database performance tests using VxVM and VVR with synchronous replication, the database performance was roughly equal to the baseline performance without replication. The same test conducted over a distance of 130 km showed database performance degradation of less than three percent compared to the baseline without replication. In fact, GTAC testing found far greater application performance degradation in an array-based product, with degradation of over 13 percent when compared with the baseline.

GTAC evaluated the performance characteristics of VERITAS with an industry standard OLTP (on-line transaction processing) benchmark from Quest Software called Benchmark Factory®¹. This tool can generate both online and deferred warehouse type transactions, non-uniform in nature, applied to an Oracle database.

BenchMark Factory simulated an enterprise database environment that emulated typical enterprise-class data center networking configurations. These configurations included several components such as high-end fault tolerant servers, enterprise-class storage arrays, director-class Fibre Channel SAN switches, DWDM and SONET transport devices (Nortel OPTera™, Metro 5200, Cisco Systems ONS 15540, and Cisco Systems ONS 15454), and several hundred kilometers of Corning SMF-28™ optical fiber. For the primary site host was a Sun Microsystems Sun Fire 4810 server with 8 CPUs and 8 GB of memory. The secondary site consisted of a Sun E450 server with 4 processors and 4 GB of memory.

GiantLoop evaluated VERITAS volume replication technologies on their own merits and in comparison with an array-based replication product. All products were tested in multi-site, remote data replication configurations with approximately one terabyte of data storage per site. The performance, stability, and reliability of both products were tested and evaluated over various network configurations, traffic loads, and distances.

CONCLUSION

VxVM and VVR are easier to manage, more cost-effective, and support higher application performance than array-based data replication products. Myths to the contrary are based on faulty reasoning and scenarios that do not reflect realistic enterprise environments.

VERITAS addresses all of today's requirements for cost-effective data replication in a heterogeneous environment. As business enterprises pay increasing attention to disaster recovery implementations and the cost, ease of management, and application performance impact of replication products, VxVM and VVR will be the solution of choice.

¹ GiantLoop Testing and Certification (GTAC) Lab. [Benchmark Test Results: VERITAS and EMC Replication Technologies](#), Prepared for VERITAS. March 2003.

VERITAS ARCHITECT NETWORK

VERITAS Software Corporation

Corporate Headquarters
350 Ellis Street
Mountain View, CA 94043
650-527-8000 or 866-837-4827

For additional information about VERITAS Software, its products, VERITAS Architect Network, or the location of an office near you, please call our corporate headquarters or visit our Web site at www.veritas.com.