

EMC DATA DOMAIN GLOBAL DEDUPLICATION ARRAY

Deduplication storage for large enterprise data centers

ESSENTIALS

Scalable Deduplication Storage

- Fast, inline deduplication with up to 26.3 TB/hour of throughput
- Backup over 175 TB in less than eight hours
- Extended retention providing up to 28.5 PB of logical storage
- 10 to 30x data reduction average

Easy Integration for Large Enterprise Data Centers

- Use of VTL or EMC Data Domain Boost
- Support leading enterprise backup applications
- Minimize footprint and management complexity
- Execute up to 270 concurrent backup jobs
- Protect very large data sets

Multisite Disaster Recovery

- 99 percent bandwidth reduction
- Cost-efficient disaster recovery
- Flexible replication topologies
- Consolidate backup and DR
- Cross-site deduplication and replication from up to 270 remote sites
- Encrypted replication

Ultra-Safe Storage for Reliable Recovery

- Continuous recovery verification, fault detection, and healing
- Dual disk parity RAID 6

Operational Simplicity

- Reduce backup complexity
- Simple system administration

NEXT-GENERATION DATA PROTECTION

EMC® Data Domain® deduplication storage systems have revolutionized disk backup, disaster recovery, and remote office data protection with high-speed, inline deduplication. Backup data can be reduced in size by an average of 10 to 30x, so disk backup storage is now cost-effective for long-term onsite retention, and highly efficient for network-based replication to disaster recovery sites.

EMC Data Domain Global Deduplication Array (GDA) is the industry's highest performance inline deduplication storage system for enterprise backup applications. GDA presents a single deduplication storage pool to the backup application across two EMC Data Domain DD890 controllers. Multiterabyte data sets are dynamically and transparently load balanced across the controllers, simplifying capacity management, performance management, and backup administration.

SCALABLE DEDUPLICATION STORAGE

Global Deduplication Array is a scalable system that offers up to 28.5 PB of logical storage with a typical enterprise data set and backup policy. Enterprises with over 200 TB of data can use GDA to store and protect two months of retention on disk in the same number of floor tiles that would normally provide only a couple of days of disk staging. With throughput up to 26.3 TB/hour, Global Deduplication Array allows more backups to complete sooner while putting less pressure on limited backup windows. The global deduplication file system takes advantage of both controllers and uses this processing power to scale performance. During a single-eight hour backup window, over 175 TB of data can be backed up to GDA, or its performance capabilities can be used to minimize the backup windows substantially. Likewise, GDA uses both controllers to scale restore performance to ensure a rapid recovery when needed.

EASY INTEGRATION FOR LARGE ENTERPRISE DATA CENTERS

Global Deduplication Array seamlessly integrates into backup environments that use EMC NetWorker®, Symantec NetBackup, and Backup Exec using the EMC Data Domain Boost software option. DD Boost software provides users with the retention and recovery benefits of inline deduplication as well as managed replication for offsite disaster recovery protection. Alternatively, GDA can be connected to your backup servers as a virtual tape library (VTL) over Fibre Channel using EMC Data Domain Virtual Tape Library software. GDA is qualified with all leading enterprise backup software and easily integrates into existing infrastructures without change for either data center or distributed office data protection.

Global Deduplication Array addresses requirements for scalability without complexity. It simplifies management by allowing multiple backup servers to simultaneously use a single, combined disk pool. Larger data sets can be accommodated simply by adding more storage to GDA.

Many large enterprise backup infrastructures have to protect thousands of clients and require many concurrent backup jobs to meet the daily backup window. While accommodating very large backup policies, support for up to 270 concurrent backup jobs gives GDA users room to grow their backup environments without having to manage the complexities of hundreds of physical tape devices. For backup environments up to several hundred terabytes, administrators can target all backup policies to a single GDA and leverage a common deduplication storage environment.

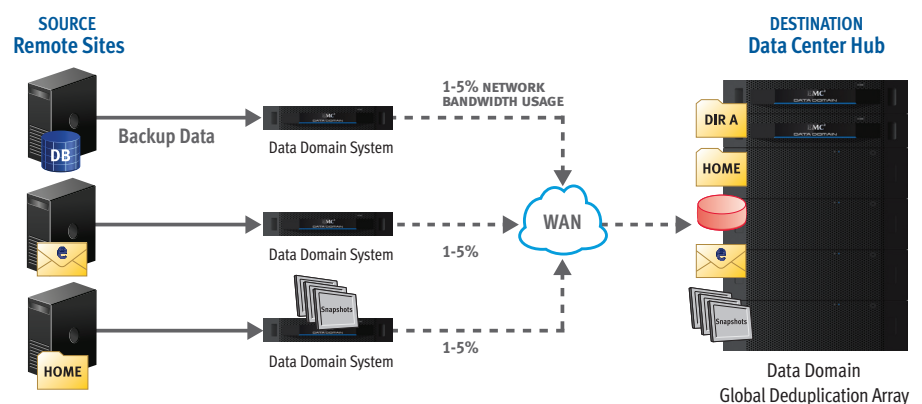
The innovative global deduplication technology inherent in GDA minimizes the need to reconfigure complex backup policies or load balance policies for performance or capacity management. Consequently, very large data sets can be easily protected with administrative simplicity while maximizing the overall deduplication efficiency and minimizing the physical storage footprint.

MULTISITE DISASTER RECOVERY

Taking advantage of EMC Data Domain Replicator software, Global Deduplication Array can be included in flexible replication topologies that can meet a large variety of disaster recovery (DR) configurations for multiple remote sites and large data centers. Using DD Boost managed file replication, the bandwidth required for replication to or from GDA is reduced by up to 99 percent, which dramatically reduces the time needed to create duplicate copies of backups for consolidation or disaster recovery purposes. This reduction of data transferred over the wide area network makes network-based replication fast, efficient, and reliable.

For organizations with large data centers, one or more GDA systems can be deployed as the target for local backup policies. The system can also be the primary DR target for a different geographically distributed data center. Additional multisite protection is possible by simply mirroring the entire backup contents to another GDA located at a remote DR site.

GDA can also be used to consolidate backups from up to 270 remote sites, when using DD Boost-managed file replication. Cross-site deduplication takes place across all of the remote sites and local backups, minimizing bandwidth utilization even further, since only the first instance of data is transferred across any of the WAN segments. If confidentiality is required, deduplicated and compressed data can be encrypted in-flight when being replicated between Data Domain systems, independent of the replication topology used.



Multisite Disaster Recovery

When using DD Boost, Global Deduplication Array can provide a replication target to consolidate backups from up to 270 remote sites. Cross-site deduplication takes place across all of the remote sites and local backups, minimizing bandwidth utilization, since only the first instance of data is transferred across any of the WAN segments.

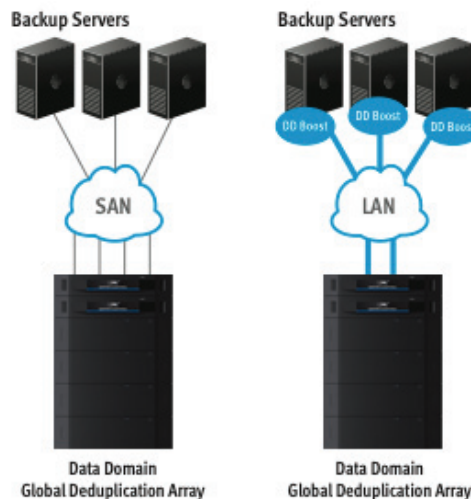
ULTRA-SAFE STORAGE FOR RELIABLE RECOVERY

The EMC Data Domain Data Invulnerability Architecture provides the industry's best defense against data integrity issues. Continuous recovery verification, fault detection, and self-healing protects data during the initial backup and throughout the data lifecycle. GDA is configured with dual disk parity RAID 6, so two disks can fail simultaneously and the system will remain healthy. Fans and power supplies are redundant and easy to replace for added system resilience.

OPERATIONAL SIMPLICITY

GDA is simple to manage. It can be centrally managed from the EMC Data Domain Enterprise Manager graphical user interface (GUI). Administrators can see a consolidated view across both controllers, and manage the file system, replication, and DD Boost interface centrally, simplifying the deployment. It also provides administrators with a consolidated view of capacity and performance utilization and proactively reports by email on complete system status via autosupport.

Administrators can also manage the GDA through a command line interface (CLI) over SSH. SNMP monitoring allows administrators to easily integrate the GDA with existing heterogeneous SNMP monitoring tools. Simple scriptability along with SNMP monitoring provides additional management flexibility.



GDA can be set up as the largest Data Domain deduplication storage system using VTL for backup infrastructure relying on Fibre Channel networks or using DD Boost when leveraging IP networks.

SPECIFICATIONS

| Global Deduplication Array Specifications | |
|--|---------------|
| Logical Capacity, Standard ^{1,2} | Up to 11.4 PB |
| Logical Capacity, Redundant ^{1,3} | Up to 28.5 PB |
| Maximum Throughput (VTL) ⁴ | 10.7 TB/hr |
| Maximum Throughput (DD Boost) ⁵ | 26.3 TB/hr |
| Remote Office Sites ⁶ | 270 |

1. All capacity values are calculated using Base 10 (i.e., 1 TB = 1,000,000,000,000 bytes).
2. Mix of typical enterprise backup data (filesystems, databases, e-mail, developer files), full backup weekly, incremental backup daily, to system capacity.
3. Mix of typical enterprise data (filesystems, databases, e-mail, developer files), full backup daily, to system capacity.
4. Maximum throughput achieved using VTL interface and 8 Gb/s Fibre Channel.
5. Maximum throughput achieved using DD Boost and 10 Gb Ethernet.
6. Maximum remote office site fan-in ratio requires DD Boost.



SOFTWARE

EMC Data Domain Operating System (DD OS) 5.1 or later

EMC Data Domain Boost

EMC Data Domain Virtual Tape Library (VTL)

EMC Data Domain Global Deduplication software

Software Features

Global deduplication, Data Invulnerability Architecture including end-to-end verification (ongoing) and integrated dual disk parity RAID 6, snapshots, telnet, FTP, SSH, email alerts, scheduled capacity reclamation EMC Data Domain Boost, EMC Data Domain Virtual Tape Library (for open systems), and EMC Data Domain Replicator optional software

Management

EMC Data Domain Enterprise Manager, SNMP, and command line management interface

Data Access

DD Boost (for use with Symantec OpenStorage and EMC NetWorker) over 1 GbE or 10 GbE, or tape library emulation (VTL) over Fibre Channel

SYSTEM EXPANSION

- Up to 24 expansion shelves with 2 TB drives; up to 12 shelves with 2 TB drives per controller
- Up to 32 expansion shelves with 1 TB drives; up to 16 shelves with 1 TB drives per controller
- Support for a mix of expansion shelves with 2 TB drives or 1 TB drives up to maximum external storage capacity
- Support for a mix of ES30 and ES20 shelves up to maximum external storage capacity

HARDWARE PLATFORM

Key components of a Global Deduplication Array:

- 2 DD890 controllers
- 2 to 24 shelves with 2 TB drives, up to 12 per controller, up to maximum external storage capacity
- 2 to 32 shelves with 1 TB drives, up to 16 per controller, up to maximum external storage capacity
- Two dual-port 10 GbE NIC cards (one per controller) for controller interconnect
- Up to four dual-port 10 GbE optical or copper NIC cards (two per controller) or four dual port 8 Gbps VTL HBA (two per controller) for host connectivity

CONTACT US

To learn more about how EMC products, services, and solutions can help solve your business and IT challenges, contact your local representative or authorized reseller—or visit us at www.EMC.com.

EMC², EMC, Avamar, Data Domain, Global Compression, NetWorker, and the EMC logo are registered trademarks or trademarks of EMC Corporation in the United States and other countries. All other trademarks used herein are the property of their respective owners. © Copyright 2011 EMC Corporation. All rights reserved. Published in the USA. Data Sheet 10/11 H7027.4

EMC Corporation
Hopkinton, Massachusetts 01748-9103
1-508-435-1000
In North America 1-866-464-7381
www.EMC.com

EMC Backup Recovery Systems
Santa Clara, California
95054
1-408-980-4800
In North America 1-866-933-3873

EMC²
®