EMC Data Domain Global Deduplication Array

Deduplication storage for large enterprise data centers

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 10-30x, 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 2 EMC Data Domain DD880 controllers. Multi-terabyte 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 14.2 PB of logical storage with a typical enterprise data set and backup policy. Enterprises with over two hundred terabytes of data can use GDA to store and protect 2 months of retention on disk in the same number of floor tiles that would normally provide only a couple days of disk staging. With throughput up to 12.8 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 100 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
EMC Data Domain® deduplication storage systems simplify 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.

Multi-Site Disaster Recovery
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 2 EMC Data Domain DD880 controllers.

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 environment 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.
Multi-site 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 optimized 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 of 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 multi-site 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. 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.

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

Global Deduplication Array seamlessly integrates into backup environments that use Symantec™ NetBackup and Backup Exec using the EMC Data Domain OpenStorage software option. OpenStorage software allows users to enjoy the retention and recovery benefits of inline deduplication as well as managed replication for offsite disaster recovery protection.

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 OpenStorage 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 script-ability along with SNMP monitoring provides additional management flexibility.
Global Deduplication Array Specifications

<table>
<thead>
<tr>
<th>Feature</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capacity, Raw</td>
<td>Up to 384 TB</td>
</tr>
<tr>
<td>Logical Capacity, Standard 1, 3</td>
<td>Up to 5.7 PB</td>
</tr>
<tr>
<td>Logical Capacity, Redundant 2, 3</td>
<td>Up to 14.2 PB</td>
</tr>
<tr>
<td>Maximum Throughput</td>
<td>12.8 TB/hr</td>
</tr>
<tr>
<td>Remote Office Sites</td>
<td>270</td>
</tr>
</tbody>
</table>

1. Mix of typical enterprise backup data (file systems, databases, mail, user files), full backup weekly, incremental backup daily, to system capacity.
2. Mix of typical enterprise backup data (file systems, databases, mail, user files), full backup daily, to system capacity.
3. All capacity values are calculated using Base10 (i.e. 1 TB = 1,000,000,000,000 bytes).

Software

- EMC Data Domain Operating System (DD OS) 4.8 or later
- EMC Data Domain OpenStorage software 2.0 or later
- EMC Data Domain Global Deduplication Array software
- Symantec Veritas NetBackup 6.5 or later
- Symantec Backup Exec 2010 or later

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 OpenStorage, EMC Data Domain Replicator optional software

Management

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

Data Access

- Symantec OpenStorage over 1 GbE or 10 GbE

System Expansion

- Up to 384 TB raw capacity
  - Up to twelve 16 TB expansion shelves
  - Up to twelve 8 TB expansion shelves
  - Support for a mix of 16 TB and 8 TB expansion shelves up to raw capacity of 384 TB

Hardware Platform

- Key components of a Global Deduplication Array:
  - 2 DD880 controllers
  - 2 to 24 ES20 storage shelves, up to 12 ES20 per controller
  - Two dual-port 10 GbE NIC cards (one per controller) for controller interconnect
  - Two 10 GbE NIC cards (one per controller): dual-port copper or 1-port optical for network connectivity

EMC Corporation

Hopkinton, Massachusetts

01748-9103

1-508-435-1000

In North America 1-866-464-7381

www.EMC.com


Take the next step

To learn more about EMC Data Domain deduplication storage, contact your local EMC sales representative or authorized value added reseller, call us at 1-866-WE-DDUPE or visit www.EMC.com.