Backups are essential for short-term data recovery but traditionally have not been designed to serve long-term backup retention needs. However, many corporations still use tape backups for long-term retention needs without considering the benefits of deploying a disk-based retention process. We believe that specialized disk-based retention can provide long-term data retention at considerable cost savings over inefficient tape-based backup retention.

There are several reasons why businesses continue to use tape backup to the extent they do: 1) misconceptions regarding the cost of disk-based long-term backup retention 2) reliance on existing tape libraries typically located in recovery (DR) sites and 3) perceived complexity of using a disk-based backup appliance for both short-term and long-term backup retention.

EMC extends EMC Data Domain systems with DD Extended Retention software to fulfill short-term and long-term backup storage needs by combining multiple tiers in a single appliance. This software addresses many of the challenges organizations are struggling with as they balance backup retention requirements with long-term recoverability.

This Solution Profile will build the case for disk-based long-term backup retention. We'll present the Data Domain system with DD Extended Retention software as an economical and simple solution for long term retention of backups, and discuss its real-world implementations in customer environments.

**The Big Lie: Using Tape Backup for Long-term Retention**

Companies tell themselves the “big lie”: that for long term backup retention tape is the preferred technology and is more cost-effective than disk-based backup. The fact is tape backup in any form is not as cheap as you think and it compounds downstream access problems related to search and discovery. Short-term and long-term backup retention can both perform beautifully on disk-based storage when designed properly. Let’s take a closer look at why tape backup storage is dreadfully inadequate for cost-effective long-term backup retention.

1. **Tape is not as cheap as you think.** Tape cartridges are admittedly cheap to buy, which is attractive to companies retaining large volumes of backup data. But tape automation, transport and storage space are expensive and ongoing. This is particularly true when upgrading and replacing large tape libraries, which are expensive to purchase and take up big chunks of precious data center real estate.

2. **Recoverability and searchability suffer with tape.** Tape's usefulness as a data retention medium only works if the company can prove they have specific data safely stored off-site.
When the DOD or SEC come calling, IT may know they have regulated data committed to tape but finding it is a whole other matter. The same holds true for eDiscovery collection from tape. Collection deadlines are short, yet searching backup tape requires lengthy strolls through backup catalogs – and the fervent hope that the original backup application is available to restore. (Often it is not.)

3. **Keeping tape fresh is a big job.** The only way to keep tape in some semblance of order is to undergo large-scale tape refresh and migration every few years. This is not exactly on the top of IT’s hit list but it is necessary if the company uses backup tape as a de facto archive. Still, many companies do not do it at all and must deal with increasingly unreliable media. However, the ones who do refresh must deal with the time and expense.

Many companies have turned to disk-based backup to increase backup and recovery performance, improve recoverability, and take advantage of deduplication and off-site replication. These are excellent drivers for adopting a disk-based backup storage solution and we support them. However, companies are still left in the lurch when it comes to long-term retention that is measured in years.

Disk-based backup systems add tremendous value to the storage environment by meeting basic challenges. However, traditional disk-based backup systems are not designed for long-term backup retention. For example, some simple VTLs serve as a glorified cache for incoming backup that is immediately backed up to tape. But when it comes to fast recovery and long-term retention, companies need a great deal more. Basic functionality should include deduplication, large and scalable capacity, and segmented but integrated active and retention tiers. DD Extended Retention software enables a two-tiered EMC Data Domain system that fills the bill.

**Data Domain Extended Retention Software**

Data Domain systems deduplicate backup data inline resulting in 10-30x reduction in backup storage required on average. Because Data Domain systems reduce data so effectively, companies can keep data online long-term without sacrificing data center space or spending high amounts on new storage capacity and energy costs.

The Data Domain Operating System includes the Data Domain Data Invulnerability Architecture, built to provide continuous data verification and self-healing, dual disk parity RAID (6) and unique write verification.

Data Domain systems work with all leading backup applications, which lets IT easily integrate them into their existing storage infrastructure. Data Domain systems scale to 65 PB logical capacity to protect deduplicated data. Aggregated throughput across the product family ranges from 1.1 TB an hour to 31 TB an hour.

DD Extended Retention software is purpose-built to enable Data Domain systems to provide long-term retention for multi-vendor backups. Unlike tape or traditional disk systems, the system deduplicates storage for consolidated protection across a broad spectrum of applications. Data Domain systems are qualified with leading enterprise backup software applications and easily integrates into existing enterprise infrastructures. The system supports simultaneous data access methods as a disk-based target through NFS and CIFS file service protocols or as VTL or advanced integration via EMC Data Domain Boost.

The two-tiered system uses two logical tiers to enable short-term and long-term retention of backup data. Incoming backup data lands on the active tier, which is built for high performance backup and short-term recovery. The retention tier scales easily using storage shelves to add multiple retention units. The system achieves up to 31 TB per hour of throughput and capacity scales up to 65 PB logical.
ACTIVE AND RETENTION TIERS

The active tier of a Data Domain system with DD Extended Retention software is meant for short-term retention for operational recovery purposes.

User-set policies move deduplicated data from the active tier to the high capacity retention tier: a massive secondary tier that uses the same Data Domain controller, management and namespace. The user defines data movement between the tiers within Data Domain Enterprise Manager. Users may easily optimize the system to suit both short and long-term retention needs.

The two-tiered architecture enables administrators to deploy a short-term active tier for fast ingest of backup data, and a retention tier for cost-effective long-term backup retention. This gives customers the best of both worlds: the active tier deduplicates data inline for exceptionally efficient short-term retention, and the retention tier provides highly efficient and cost effective online storage for backup data.

Data movement policies are set within the Data Domain system enabling transparent movement from the active tier to the retention tier. Moving backup data to the retention tier provides faster retrieval than is possible from tape and enables companies to keep backup data online for long periods of time at a fraction of the rack space and energy demands of uncompressed backup or tape data.

By using the system for long-term backup retention, users achieve much better levels of storage optimization gaining immediate access to online backups and speeding up business processes that depend on long-term stored data.

Data Domain system with DD Extended Retention software
Data Domain Extended Retention Software Key Features

<table>
<thead>
<tr>
<th>Key Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy Integration</td>
<td>EMC retains Data Domain’s backup vendor-neutral architecture. Data Domain systems support leading backup applications via VTL, CIFS, NFS or Data Domain Boost.</td>
</tr>
<tr>
<td>Efficient replication and simple management</td>
<td>Supports network-efficient replication from all Data Domain systems, enabling customers to replace tape for DR. With Data Domain Boost, applications maintain total control over data retention as local and remote backups can be managed from the backup console.</td>
</tr>
<tr>
<td>Highly scalable capacity for long-term backup retention</td>
<td>Deduplicated logical capacity scales up to 65 PB behind the single controller.</td>
</tr>
<tr>
<td>High backup throughput</td>
<td>The high performance Data Domain controller ensures that the data movement process does not impact backup speed, which is especially important in critical data environments. Data Domain systems achieve fast throughput speeds of 31 TB per hour.</td>
</tr>
<tr>
<td>Centralized administration and policies</td>
<td>Centralized management console uses wizards and dashboards to ease device configuration and system monitoring. A simple policy engine automates movement between the tiers.</td>
</tr>
</tbody>
</table>

DD Extended Retention in the Real World

We spoke with IT professionals from companies that are using DD Extended Retention. We were interested in their long-term backup and recovery environment and in the business issues they intended to solve.

CUSTOMER #1: INTERACTIVE ENTERTAINMENT INDUSTRY

This leading global interactive entertainment software company develops, publishes and distributes interactive software worldwide for video game systems, personal computers, wireless devices and the Internet.

This company's primary backup and recovery objective is to minimize and eventually eliminate their dependency on tape. They have found that tape is cumbersome to manage, requires manual intervention in the data center, and has ongoing and increasing costs associated with offsite storage and retrieval. By deploying a Data Domain system with DD Extended Retention software, the company aims to consolidate and retain data at all of its studios and then replicate it to a remote data center.

The company has dozens of offices and studios that each back up hundreds of terabytes a week, and that must observe strict retention policies. In this demanding environment, a tape-only solution
makes it difficult to support long-term retention and make the most efficient use of backup windows. For this company, the Data Domain system’s ability to support their 3-year retention policy and integrate with their existing EMC NetWorker backup solution via DD Boost was particularly important. Features such as inline deduplication allow the company to quickly, easily and efficiently backup their data. Subsequently the retention tier allows them to move the requisite backup sets into the retention tier for long-term retention and can also be replicated for offsite storage.

The company’s data is increasing rapidly, which generates critical requirements around deployment, flexibility, scalability and integration with their existing NetWorker environment. DD Extended Retention’s ability to quickly deploy with their existing infrastructure with policy based data movement and web and command line interfaces makes a compelling case for a full blown deployment for this company at all of its studios. Furthermore, a modular design makes the solution scalable and allows the company to incrementally increase system capacity.

CUSTOMER #2: FINANCIAL SERVICES COMPANY

This large investment bank has offices all over the world that were driving needs around backup and archiving. Tape-based backup proved insufficient to meet the needs of multiple data centers. Weekly full backups scheduled over weekends took longer and longer to complete, and restore speeds failed service level agreements (SLAs) for recovery objectives. They also found that storing regulated data on long-term tape met basic compliance regulations, but was inadequate for litigation eDiscovery searches.

By purchasing a Data Domain system with DD Extended Retention software this customer tackled these challenges. The active tier enables the company to economically replace its tape backup infrastructure with an extremely fast and highly reliable deduplicated backup solution, while the retention tier enabled the company to efficiently retain backup data onsite long term instead of shipping tape to an off-site vault. In addition, Data Domain replication features allow the company to replicate data bi-directionally and many-to-one for vastly improved disaster recovery between data centers. This is a huge improvement over backing up to tape, and enabled the bank to backup and restore well within their SLAs.

The company also purchased EMC SourceOne for Exchange software to archive Exchange data and store it on the same Data Domain system, enabling them to consolidate backup and archive into a single system. The company gained a tremendous sense of control over their data with Data Domain systems. The system offers massive backup data reduction, improves DR, and increases accessibility of backup and archive data.

**Key Benefits**

- **Supports evolving backup and long-term retention needs.** We strongly support the use of disk-based backup appliances for short-term and long-term storage. But the reality is that while many enterprises have adopted some form of short-term disk-based backup storage, they rarely optimize long-term backup storage needs. Data Domain systems with DD Extended Retention now enable efficient, cost effective long-term backup retention.

- **Cost-effective.** Replacing tape with a Data Domain system lets IT save money and lower risk by replacing tape with highly efficient disk-based backup. Keeping a massive tier of storage behind a single controller yields up to 65 PB of logical capacity. Scalability enables IT to grow system deployments and minimize tape for immediate backup and for long-term retention. Centralized management also lowers costs by decreasing administrative overhead.
• Highly efficient tiered architecture. Data Domain systems support leading enterprise backup software applications, which enables IT to leverage their existing backup investments. With DD Extended Retention, the active and retention tiers enable simple, policy-driven movement based on data retention requirements, and can accept incoming data from up to 270 remote sites to support long-term retention for a distributed enterprise.

• Long-term data retention. A Data Domain system with DD Extended Retention software eliminates the need for tape for long-term retention. The retention tier is highly scalable and manageable, and Data Invulnerability Architecture provides continuous recovery verification, fault detection and self-healing to protect data throughout its lifetime.

• Granular recovery. The retention tier can consist of multiple retention units, which provide fault isolation and granular recovery over the long term. Therefore, a disaster only affects an individual unit and rather than a system-wide impact, only the affected unit need be restored. Similarly, replication is performed on a retention unit basis and if a replication is interrupted, IT only needs to re-replicate the affected unit to complete synchronization.

Taneja Group Opinion

Operational backup and recovery typically involves data retention periods of weeks or months, while long-term data retention is measured in years. Today’s most common method of long-term data retention is simply to keep backup tapes longer. This disjointed, manual and expensive process carries distinct risks and costs for users.

EMC has done the smart thing by creating the DD Extended Retention software option, enabling a high capacity retention tier on a powerful deduplication storage platform for long-term retention of backup data. We recommend corporations using Data Domain for long-term backup retention also investigate archiving software for functionality that backups cannot provide. When they do deploy archive software, the same Data Domain system will provide safe archive storage as well as high performance backup protection.

EMC’s solution supports the present and future of corporate data retention. The backup market should take a very serious look at DD Extended Retention for powerful and cost-effective short-term and long-term backup retention.

NOTICE: The information and product recommendations made by Taneja Group are based upon public information and sources and may also include personal opinions both of Taneja Group and others, all of which we believe to be accurate and reliable. However, as market conditions change and not within our control, the information and recommendations are made without warranty of any kind. All product names used and mentioned herein are the trademarks of their respective owners. Taneja Group, Inc. assumes no responsibility or liability for any damages whatsoever (including incidental, consequential or otherwise), caused by your use of, or reliance upon, the information and recommendations presented herein, nor for any inadvertent errors that may appear in this document.