

# CRUNCH™ for DPM

Technology and Product Description

© 2013 BridgeSTOR, LLC Poway, CA

# Table of Contents

DPM 2012 Data Deduplication .....	2
Myths and Mysteries About DPM Deduplication .....	2
The Challenges of Deduplicating Microsoft DPM Data .....	3
DPM Deduplication Strategies with CRUNCH .....	1
DPM Storage Requirements <i>without</i> Deduplication .....	1
Estimating Daily Recovery Point Size .....	2
Determining Retention Range Objectives .....	2
DPM Storage Requirements <i>with</i> CRUNCH Deduplication .....	2
Reducing the Number of Recovery Points on Short-Term Disk .....	2
Estimating Deduplicated Recovery Point Capacity Requirements .....	3
Determining Deduplicated Retention Range Objectives .....	3
Using the Deduplicated Long-Term Storage Tier to Maintain a Disk-Based Archive of Recovery Points.....	4
Eliminate the Short-Term Disk Storage Tier; Replace with Long-Term Deduplicated Storage .....	4
A Combination of Techniques .....	5
How BridgeSTOR Deduplicates DPM Data .....	5
Industry-Leading Technological Innovation.....	5
DPM Virtual Tape Interface .....	6
MTF Boundary Alignment .....	6
Block-Level Deduplication for DPM Data .....	6
Thinly Provisioned Storage Containers.....	7
Synthetic Recovery Point Single Instance File Deduplication.....	7
Near-time Data Deduplication.....	8
Linux/Windows “Agentless” Backup.....	8
Storing Deduplicated Data on Physical Tape for Offsite Archive.....	8
Appendix: CRUNCH for DPM ROI Examples .....	10
Base Case .....	10
ROI Model 1: Combining DPM Short-Term Disk with CRUNCH Virtual Tape and Deduplicated Containers.....	11
ROI Model 2: Eliminating DPM Short-Term Disk with CRUNCH Virtual Tape and Deduplicated Containers.....	12

# DPM 2012 Data Deduplication

This white paper describes BridgeSTOR's implementation of data deduplication for System Center 2012 – Data Protection Manager (DPM).

Deduplication is supplied through BridgeSTOR's CRUNCH™ for Microsoft DPM Virtual Deduplication Appliance. The BridgeSTOR Virtual Appliance typically co-resides with the DPM 2012 server on a physical Microsoft Windows Server 2008 R2 Hyper-V system, **adding a new, deduplicated storage tier to DPM.**

BridgeSTOR's CRUNCH for DPM deduplication is optimized to interoperate with the unique storage workloads produced by DPM 2012 when protecting Hyper-V, Exchange, SQL, SharePoint and shared/networked storage (NAS, CIFS) servers. CRUNCH for DPM also supplies an “agentless,” deduplicating backup capability (that operates outside of DPM) to protect Linux file servers.

## Myths and Mysteries About DPM Deduplication

A Google search for “DPM deduplication” tells us a lot about the “disinformation” surrounding the topic. Among the more popular threads are the following:

1. “DPM already has deduplication because it runs on Server 2008 or 2012 that includes native deduplication.”

The short answer is: “No, it does not.” This excellent InfoWorld article tells it like it really is:

“Windows Storage Server 2008 enhances the deduplication capabilities of its predecessor, using SIS-based data deduplication for the Windows File Services, which eliminates identical files on volumes. The duplicates are replaced by pointers that link to files placed in the SIS Common Store. Obviously, for this to work on the backup side, you need to have a SIS-aware backup product. And that's where Microsoft's System Center Data Protection Manager comes into play.

Some people mistakenly believe System Center's Data Protection Manager (DPM) to have deduplication capabilities and feel they have no need for a hardware product to assist with deduplication. That's not true. DPM may use components that are dedupe-like (for example, block-level change tracking), and DPM certainly does an excellent job of using small amounts of storage to fit a large amount of data or a large number of recovery points (giving the impression that traditional SIS or deduplication must be involved). But DPM does not use the traditional compression, SIS, or deduplication features that you will find in a hardware storage platform. The best scenario is to use both DPM and a hardware deduplication product.”

***The fact is that DPM can protect data that has been deduplicated by Windows, but the deduplicated data being protected is returned to its non-deduplicated state when stored by DPM.***

2. “DataDomain products can be used to deduplicate DPM data.”

The myth goes on to describe the DataDomain VTL interface as the key to providing a block-level interface to DPM as short-term disk storage.

The short answer is: “No, it doesn't,” The reason is that DPM requires a native block-level disk storage target for short-term storage. DataDomain's native interface is a NAS (file) based interface and the VTL interface (an extra

cost option) does not present a block-level disk target to DPM.

3. “DPM 2012 will have native DPM deduplication.”

Again, not really. Microsoft puts this argument away here:

<http://technet.microsoft.com/en-us/library/jj656644.aspx>

**Quoting from TechNet: “DPM 2012 protects Windows 8 deduplicated volumes.” That is the full extent of Microsoft’s deduplication strategy for DPM.**

## The Challenges of Deduplicating Microsoft DPM Data

Microsoft architected DPM as a data protection product with two storage tiers: a short-term, disk-based tier on which all operational recovery data would be stored and a long-term, tape-based tier intended to be used for removable media destined for offsite storage.

The two-tier approach embraced by Microsoft is similar to disk-to-disk-to-tape (D2D2T) backup strategies that started coming into general use in 2002. A BridgeSTOR predecessor company (Okapi Software led by BridgeSTOR’s CEO John Matze) pioneered disk-to-disk-to-tape backup and coined the term D2D2T.

Deduplication flourished in D2D2T backup strategies and rose to prominence as storage administrators embraced the benefits of backup to disk speed and disk-based data recovery performance. From the outset, storage administrators found that increasing the amount of backup data retained on disk translated into improvements in recovery service levels. Retaining the benefits of disk-speed recovery - while reducing disk capacity requirements through deduplication—created an entirely new market segment.

But DPM is not the same as traditional backup and traditional backup deduplication does not work well with DPM. The reason that DPM deduplication has been elusive rests in the way DPM processes and stores recovery data.

DPM servers create and maintain a *replica*, or copy, of the data that is on protected servers. Replicas are stored in the *storage pool* which is a short-term storage tier as described above. Whether protecting file data or application data, protection begins with the creation of the replica of the data source.

The replica is *synchronized*, or updated, at regular intervals according to the settings that the storage administrator configures. The method that DPM uses to synchronize the replica depends on the type of data being protected. If a replica is identified as being inconsistent, DPM performs a *consistency check*, which is a block-by-block verification of the replica against the data source.

DPM protection agents track changes to protected data and transfer the changes to the DPM server. The protection agent also identifies data on a computer that can be protected and is involved in the recovery process. A protection agent must be installed on each computer to be protected by DPM.

The road block preventing DPM data deduplication at the short-term storage pool level is the consistency check. During a consistency check, relatively little data is transferred and stored but a great deal of data comparison is carried out. This typically translates into thousands of reads and writes—each requiring data reduction and “rehydration” processing to take place.

And although there are variations in the I/O intensity of the consistency check across the various applications

protected by DPM, this fundamental architecture of DPM stands in the way of reasonable deduplication performance.

To add deduplication to DPM while retaining “reasonable” consistency check performance required a rethinking of DPM’s architecture and led BridgeSTOR to the development of a new, long-term, disk-based data retention storage tier through which deduplication could be implemented without disrupting DPM consistency check performance.

## DPM Deduplication Strategies with CRUNCH

BridgeSTOR’s CRUNCH for DPM adds a new, deduplicated storage tier to DPM that offers the corporate data protection administrator new options and great flexibility in the design of their deduplication strategy for DPM. These strategies and options include:

1. Reducing the number of recovery points required on short-term disk (and reducing short-term disk storage requirements) by using the new long-term storage tier to offload recovery points from short-term to long-term, deduplicated storage.
2. Using the new, deduplicated long-term storage tier to maintain a disk-based archive of recovery points.
3. Eliminating the short-term disk storage tier entirely and replace it with long-term deduplicated storage.
4. Apply a combination of the techniques described above to create the data protection strategy for your data center that represents the optimal mix of recovery time, retention/archive and disk cost.

## DPM Storage Requirements *without* Deduplication

DPM can use any of the following for the storage pool:

- Direct Attached Storage (DAS)
- Fiber Channel Storage Area Network (SAN)
- iSCSI storage device or SAN
- Tape for long term data retention

Capacity requirements for the DPM storage pool are variable and depend on the size of the protected data, the rate of data change, the daily recovery point size, the number of protection groups, expected volume data growth rate, and retention range objectives.

Daily recovery point size refers to the total size of changes made to protected data during a single day. It is roughly equivalent to the size of an incremental backup. Retention range refers to the number of days for which you want to store recovery points of protected data on disk.

DPM can store a maximum of 64 recovery points for each volume included in a protection group, and it can create a maximum of 8 scheduled recovery points for each protection group each day.

Microsoft recommends making the storage pool two times the size of the protected data for protection of files. This recommendation is based on an assumed daily recovery point size of approximately 10 percent of the

protected data size and a retention range of 10 days (two weeks, excluding weekends).

## Estimating Daily Recovery Point Size

Microsoft recommends that DPM storage managers make the storage pool two times the size of the protected data (<http://technet.microsoft.com/en-us/library/bb795684.aspx>) while assuming a daily recovery point size of 10 percent of the protected data size. Daily recovery point size is related to data change rate and refers to the total size of all recovery points created during a single day.

DPM also pre-allocates storage capacity (up to 3X the initial allocation) when a new Protection Group is defined.

To get an estimate of the daily recovery point size for your protected data, you can review an incremental backup for a recent, average day. The size of the daily incremental backup is indicative of the daily recovery point size. For example, if the incremental backup for 100 GB of data includes 10 GB of changed data, your daily recovery point size will probably be approximately 10 GB.

## Determining Retention Range Objectives

Microsoft further recommends that you make the storage pool two times the size of the protected data by assuming a retention range objective of 10 days (two weeks, excluding weekends).

The longer the retention range objective, the fewer recovery points can be created each day. Without deduplication, if your retention range objective is 64 days, you can create just one recovery point each day. If your retention range objective is eight days, you can create eight recovery points each day. With a retention range objective of 10 days, you can create approximately six recovery points each day.

# DPM Storage Requirements *with* CRUNCH Deduplication

## Reducing the Number of Recovery Points on Short-Term Disk

The first CRUNCH implementation strategy is to reduce the number of recovery points required on short-term disk.

DPM can continue to use any of the following for the short-term data retention storage pool:

- Direct Attached Storage (DAS)
- Fiber Channel Storage Area Network (SAN)
- iSCSI storage device or SAN

CRUNCH for DPM adds a new, deduplicated (in this use case) *intermediate-term* data retention tier to DPM. The new deduplication tier works in conjunction with your existing short-term and long-term data retention tiers to offer almost unlimited flexibility in the number of recovery points you can retain while significantly extending the retention period.

Capacity requirements for the DPM short-term retention storage pool continue to be variable and continue to depend on the size of the protected data, the daily recovery point size, expected volume data growth rate, and retention range objectives—but since the bulk of your protection data is now on deduplicated storage, the short-

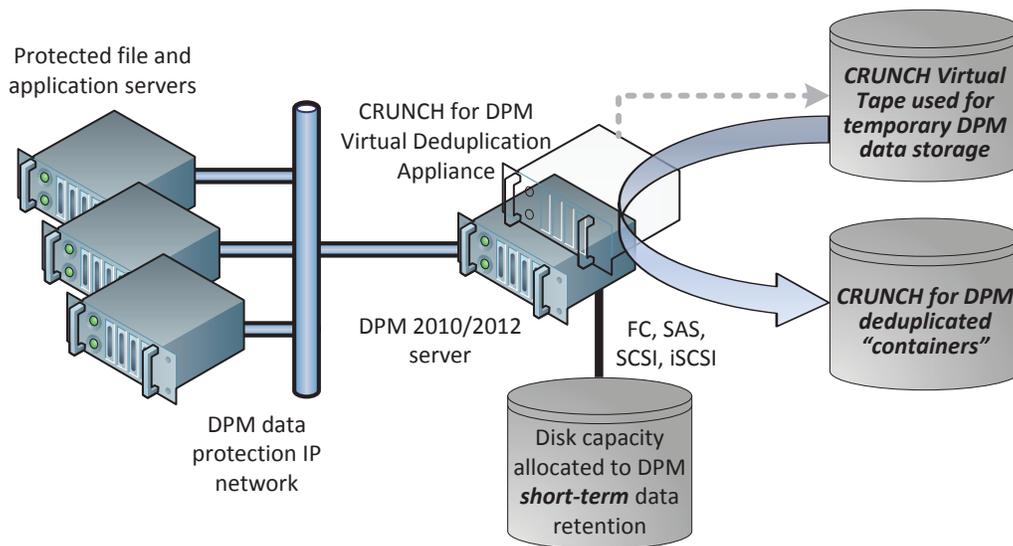
term storage requirements can now be greatly reduced by reducing retention range objectives or moving away from short-term disk entirely for the protection group.

By using the new DPM deduplicated intermediate-term data retention tier, you can reduce the number of recovery points retained on short-term disk by shifting “older” data to deduplicated disk two or three days after creation.

DPM’s maximum of 64 recovery points in the short-term data retention pool is no longer relevant since the bulk of your recovery points have been shifted to the intermediate-term deduplicated storage pool. Retain as many recovery points as you wish, with the most recent one or two on the short-term storage pool for immediate recovery and all the rest of your recovery points on deduplicated storage.

The illustration following describes the operation of BridgeSTOR’s CRUNCH for DPM operating in conjunction with

### BridgeSTOR’s CRUNCH Deduplication for Microsoft Data Protection Manager (DPM 2012)



short-term DPM disk storage.

## Estimating Deduplicated Recovery Point Capacity Requirements

BridgeSTOR recommends that you initially make the short-term data retention storage pool the same size as the protected data plus the daily recovery point size times the number of days in the retention range. This allocation can be adjusted dynamically as utilization and capacity information is collected and analyzed.

## Determining Deduplicated Retention Range Objectives

Our recommendation to make the short-term data retention storage pool the same size as the protected data assumes a retention range objective of 10 days (two weeks, excluding weekends). For the typical enterprise,

requests for recovery of data are concentrated within two to four weeks after data initial creation. A retention range of 10 days provides for recovery of data via normal DPM procedures for up to two weeks after a data loss event.

## Using the Deduplicated Long-Term Storage Tier to Maintain a Disk-Based Archive of Recovery Points

With a maximum of 64 recovery points, the longer your retention range objective in DPM, the fewer recovery points you can create each day. For example, if your retention range objective is 64 days, you can create just one recovery point each day. If your retention range objective is eight days, you can create eight recovery points each day. With a retention range objective of 10 days, you can create approximately six recovery points each day.

Many BridgeSTOR DPM customers have described the goal of their deduplication strategy as a way to maintain an archive of data recovery points on disk.

The solution is to use DPM's long-term storage tier coupled with CRUNCH for DPM to maintain multiple (deduplicated) full backups of protected data. Using this approach, you can continue to use up to 64 recovery points on short-term disk, thus having multiple recovery points per day if desired, supplemented with an "unlimited" number of recovery points that are actually full backups on CRUNCH.

CRUNCH for DPM's data deduplication is a capacity-efficient method of storing multiple full backups because of the high level of deduplication that will result.

## Eliminate the Short-Term Disk Storage Tier; Replace with Long-Term Deduplicated Storage

Many BridgeSTOR callers are most interested in greatly reducing – ideally completely eliminating – DPM's voracious storage appetite through data deduplication. Their concern is the cost of storage and reducing that cost is their primary objective and their highest priority in finding a deduplication technology to complement DPM.

BridgeSTOR now offers that alternative.

To set the record straight, eliminating DPM's short-term disk is not without some changes to your DPM strategy. But if cost reduction is your #1 objective, this approach is what you're looking for.

Eliminating short-term disk means that full backups are made to CRUNCH deduplicated virtual tape. You can then either:

Perform full backups to CRUNCH virtual tape daily, or

Send incremental backups to virtual tape (using short term protection to tape). This is a DPM limitation. When this strategy is employed, there will be a full backup job run once a week and incremental backups on other days.

**Note: the frequency needs to be set to daily for DPM to allow incremental backups.**

A benefit of this approach is that the incremental backups performed by DPM will only push the changes to the CRUNCH virtual tape daily, in other words offering a quick way for DPM to write the changed data to tape which allows the data to be subsequently moved offsite.

Using this strategy, when the weekly full backup is run, the differences will be captured by DPM. DPM sends all data to the virtual tape. When CRUNCH runs, it will take advantage of CRUNCH's "synthetic copy" capability

and omit all files that are duplicates. So from the CRUNCH copy perspective, this operation is essentially an incremental backup.

***When using short term to tape to CRUNCH, DPM short term disk is eliminated.***

All protection is performed directly against the protected server.

**Configuration: ST: disk, LT: tape (CRUNCH VT)**

With this setting the only option to tape by DPM is a full. The short term disk will capture the changed data which is seen as an incremental backup by the short-term disk. There is no way to move the data to tape except as a full.

As stated above, once the first full is complete to CRUNCH, subsequent full backups are similar to incremental backups once the Microsoft Tape Format data alignment routines have completed. This is made possible by BridgeSTOR's Single Instance file deduplication technology.

All protection data sent to CRUNCH is done against the DPM server.

You may also perform a full backup weekly and perform daily short-term, incremental backups to CRUNCH virtual tape. This backup strategy is reminiscent of IBM's Tivoli product strategy.

For instance, if a full backup is scheduled weekly and incremental backups are scheduled daily, then the full backup will go to a new CRUNCH virtual tape and hence to a CRUNCH deduplicated container and all subsequent incremental backups for six days will be appended to another new CRUNCH virtual tape and subsequently to a CRUNCH deduplicated container.

Each long-term backup recovery point created will always be on a new CRUNCH virtual tape.

## **A Combination of Techniques**

Coupled with DPM's concept of "protection groups," CRUNCH for DPM gives you the flexibility to employ deduplication where and when it best matches your organization's data protection strategy.

Some protection groups can be protected using DPM's short-term disk without deduplication used anywhere in the process. Other groups can use CRUNCH deduplication in conjunction with DPM short-term disk to achieve the desired blend of recovery point management and long-term disk-based archive. And some protection groups can bypass DPM's short-term disk storage entirely, using the long-term storage capabilities of both DPM and CRUNCH.

These options, never before possible without CRUNCH, are now available to the data protection administrator.

## **How BridgeSTOR Deduplicates DPM Data**

After determining that a new intermediate-term storage tier was the optimal insertion point for DPM deduplication, BridgeSTOR developed a combination of block-based and single-instance, file-based deduplication that optimizes data reduction with minimal operational and performance impact.

### **Industry-Leading Technological Innovation**

Deduplicating data through DPM's long-term data retention interface required the development of a number of

technological elements. The technology breakthroughs incorporated into BridgeSTOR's CRUNCH for DPM include:

- DPM virtual tape interface
- Thinly-provisioned storage containers
- Microsoft Tape Format (MTF) boundary alignment mechanism
- Block-level data deduplication for DPM data
- Single instance file deduplication
- "Near-time" data deduplication

All of these innovations were derived from new or existing BridgeSTOR patents.

## **DPM Virtual Tape Interface**

Virtual tape interfaces present disk storage to the operating system "disguised" as tape. The virtual tape device responds to SCSI commands in a manner that is indistinguishable from physical tape. BridgeSTOR has a rich history with this type of technology and a rich patent portfolio covering this type of interface.

The virtual tape interface not only provides a standard interface to DPM but it also eases integration of deduplication with DPM. Existing DPM data protection strategies can be easily and conveniently enhanced with deduplication by merely adding a "long-term" component to the appropriate DPM protection group policies. Using this methodology, the entire process of protecting data—including the creation of deduplicated backup sets—is managed through the standard DPM command interface.

## **MTF Boundary Alignment**

To understand the requirement for this technology, a short discussion of DPM data formats and an overview of deduplication techniques is necessary.

The DPM replica volume is the most recent copy of the data that has been protected by the DPM server. The format of the replica volume data is the same as the format on the protected server.

The recovery point volume stores the Volume Snapshots of each recovery point. This allows DPM to minimize the amount of disk storage used for protecting changes to data. Changed data is stored at the block level.

Data stored by DPM for long term retention, however, is not stored using the same format as the format on either the protected server or in the DPM replica volume. Instead, DPM long term retention data is stored in the Microsoft Tape Format (MTF).

Since CRUNCH for DPM deduplicates data written using DPM's long-term retention interface, the BridgeSTOR software must deal with data being stored in the Microsoft Tape Format.

It is helpful to view data stored in the Microsoft Tape Format as the equivalent of storing data in a "ZIP" file (although data stored on CRUNCH virtual tape is not compressed as a ZIP file would be). MTF data structures, like Zip files, are archives that store multiple files. Each file is stored separately. A directory is placed within the MTF file. This directory identifies the files that are in the MTF file set and identifies where in the data structure each file is located.

## Block-Level Deduplication for DPM Data

All block-based backup data deduplication systems have the same essential characteristic: they must identify data blocks that are identical. They do this by examining blocks in the input data stream and create an identifier uniquely associated with each block. Unique blocks are stored and all subsequent blocks are compared to the already-stored blocks using the unique identifier. Duplicate blocks are then “deduplicated,” leaving only the unique blocks and a collection of pointers that is used as a “proxy” for the original data.

Block-based deduplication requires that the blocks being examined have the same starting point and length. Block starting points are based on their alignment within the data structure. Once block alignment has been established, the process of examining the data and checking for duplicate blocks continues.

Backup deduplication products, like those from EMC/DataDomain (and others) use many different techniques to establish block alignment. Among the techniques employed is the “sliding window” in which a base location is established and a “window” is created and passed over the data stream seeking out naturally occurring internal block boundaries.

The sliding window approach is a “brute force” approach to achieving data block alignment. In a backup data stream, files are stored in “archives,” in which standard file-system boundaries are discarded and file identities are relegated to the originating backup software. Without an intimate understanding of the backup data format, identifying individual files and aligning data blocks is impossible. For this reason, backup deduplication systems do not (and cannot) use the identification of both duplicate blocks and duplicate files in their deduplication process as CRUNCH for DPM does.

BridgeSTOR’s CRUNCH for DPM uses semantic information and the structure of the content in the MTF-formatted data stream to find the deduplication block boundaries of the data. CRUNCH uses the MTF directory to locate, align and then deduplicate data blocks stored by DPM in the Microsoft Tape Format. In addition, the directory is also used to create BridgeSTOR’s “Synthetic Recovery Point” capability.

## Thinly Provisioned Storage Containers

A frequently heard objection to DPM’s voracious storage appetite revolves around the need to provision the physical capacity that will be required at the outset. BridgeSTOR customers asked for a “thin provisioning” mechanism for DPM capacity that could start out small and be added to dynamically as capacity requirements increased. This requirement became the foundation of BridgeSTOR’s Dynamic Storage Container technology.

CRUNCH stores deduplicated data in “containers”. Containers are dynamic, manageable capacity entities. A single instance of CRUNCH for DPM supports up to 12 containers, with each container capable of holding up to 48 TB of deduplicated data (using a 256KB deduplication block size; the container capacity limit is 3 TB when using a 4KB deduplication block size).

Each container is individually thin provisioned. This means that a container can start out with just a few GB of capacity and grow over time up to its maximum capacity limit.

Additionally, the number of containers used can be dynamically increased from a single unit up to the maximum of 12 containers.

This extreme flexibility enables a CRUNCH for DPM system to grow from just a few GB of storage capacity up to a

maximum of 576 TB (12 containers x 48 TB each, and roughly 1 PB if compressed) as capacity requirements grow in whatever increments the storage administrator desires.

## **Synthetic Recovery Point Single Instance File Deduplication**

CRUNCH for DPM brings a new level of operational ease and convenience to data recovery. The data stored on application servers protected by DPM is recoverable through an intuitive and easy to use Windows Explorer-based interface.

CRUNCH uses the data supplied by DPM to build a “synthetic” or a “synthesized” full backup view. BridgeSTOR’s synthetic recovery combines a full backup with subsequent changed data backups to create a backup set that is identical to what would have been created had the last backup been a full backup and makes this view available as a Windows share that can be used for recovery.

The system administrator performing a recovery operation sees the equivalent of a full backup from each day from which to select recovery data. Files that are unchanged from day to day are not backed up again, but are represented by pointers in the recovery directory that point to the actual file instance.

The Synthetic Recovery Points in a CRUNCH container are “expired” after a user-defined number of days, up to as long as 99 years.

## **Near-time Data Deduplication**

In a CRUNCH for DPM system, data is first written to virtual tape at whatever level of throughput is available. DPM data is then deduplicated (and optionally compressed) as data flows from the virtual tape to the deduplicated data containers for storage. Once complete, the virtual tape is “expired” and returned to DPM for the next operation.

This strategy enables CRUNCH to deduplicate DPM data in an offline process that has no performance impact on DPM operations.

## **Linux/Windows “Agentless” Backup**

The deduplication “engine” in the CRUNCH for DPM product is also capable of deduplicating a network share—whether created by Windows, Linux, Samba or other application/OS. CRUNCH does this on a user-defined schedule by reading the contents of a share and simultaneously deduplicating the data and writing it to a CRUNCH container. As with DPM data, Synthetic Recovery Points can be used to further optimize the use of storage space.

CRUNCH provides this capability without the need to install software “agents” on the servers or workstations being protected.

When using CRUNCH as a backup deduplication technology, it is the user’s responsibility to quiesce any I/O activity to/from the file(s) being backed up. It is a “best practice” to perform a snapshot of the data to be backed up and backup the snapshot.

# Storing Deduplicated Data on Physical Tape for Offsite Archive

CRUNCH deduplicated DPM data is stored in “containers” which are Linux shared volumes. To transfer that data to tape for long term, offsite data retention (in deduplicated form), a user has several options:

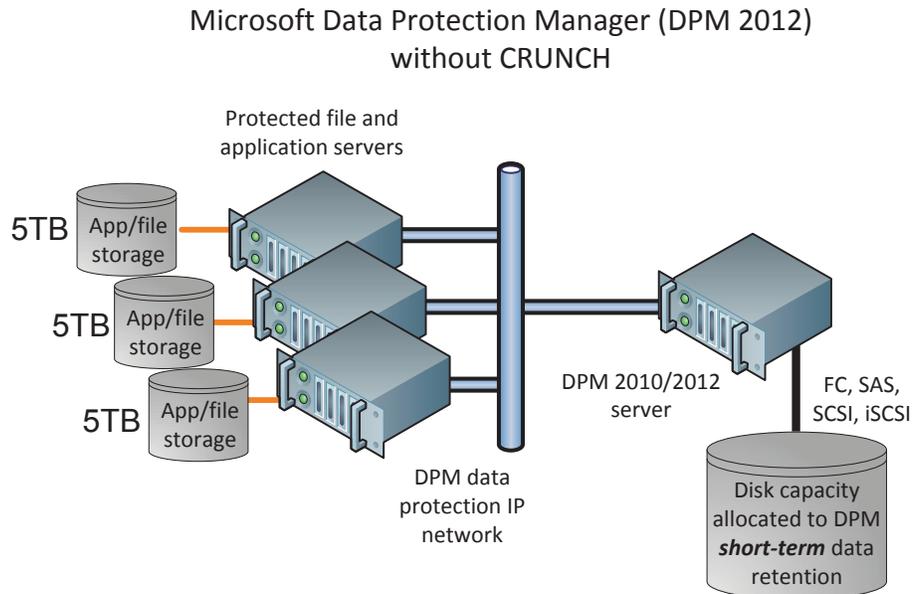
1. Copy the CRUNCH data from the Linux shared volume to a Windows share. Then use DPM to backup the CRUNCH data to tape. The data will then be stored in deduplicated form on the physical tape along with all the necessary metadata needed for rehydrating the data for later restore. (This step can be automated by adding a “post copy” command to the CRUNCH job.)
2. Use the Linux tar application, running in a physical Linux server with physical tape attached, to tar the CRUNCH data to physical tape. The data will then be stored in deduplicated form on the physical tape along with all necessary metadata. (The Linux system will SMB (CIFS) mount the data in the CRUNCH VM.)
3. Use any backup product that supports Linux clients to backup the CRUNCH data (stored in a standard Windows SMB share) to physical tape. The data will then be stored in deduplicated form on the physical tape.

Method #1 can be performed through DPM scheduling. Method #2 and #3 are not scheduled through DPM.

# Appendix: CRUNCH for DPM ROI Examples

## Base Case

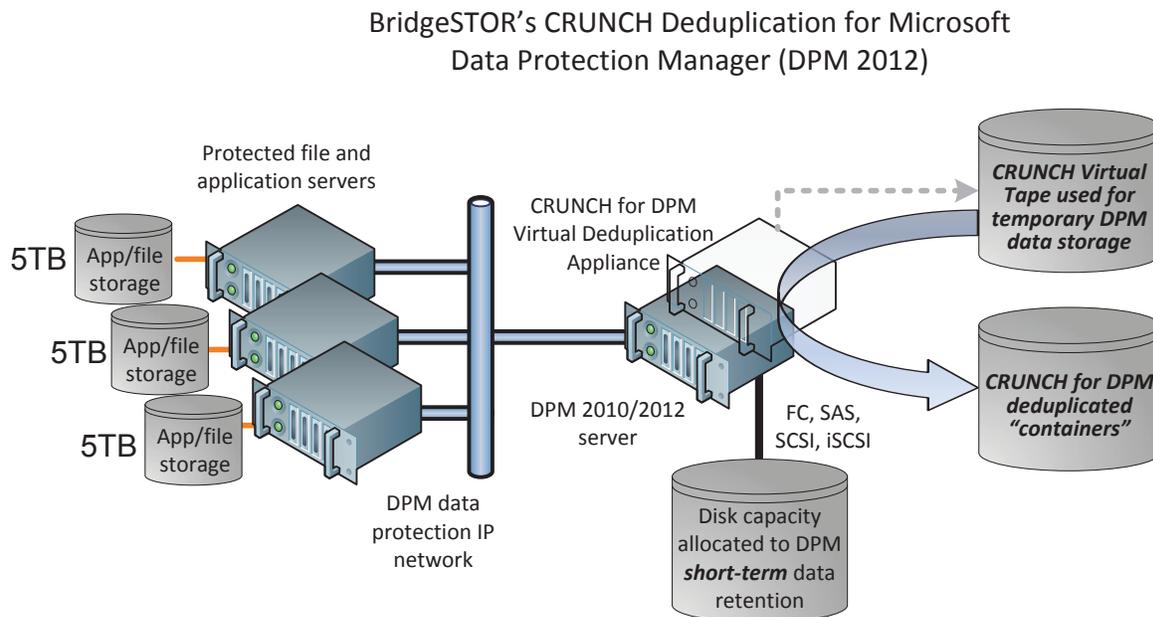
The deployment examples detailed in this Appendix use the following configuration (without CRUNCH for DPM) as the base case. CRUNCH for DPM-enabled configurations are contrasted with this configuration.



- Start with 3 application/file servers each with 5TB of data
- DPM performs the initial replica, copying  $3 \times 5\text{TB} = 15\text{TB}$  to DPM's short term disk storage
- Data changes at the rate of 10% per day
- After day 0, there is 1.5TB of new/changed data stored on DPM's short term disk storage
- After day 1, there is another 1.5TB of new/changed data stored on DPM's short term disk storage (total changed data, 3TB. Total data  $15\text{TB} + 3\text{TB} = 18\text{TB}$ )
- After day 2, there is another 1.5TB of new/changed data stored on DPM's short term disk storage (total changed data, 4.5TB. Total data  $15\text{TB} + 4.5\text{TB} = 19.5\text{TB}$ )
- This continues
- After one week (5 days), DPM short term disk storage =  $15 + 1.5 + 1.5 + 1.5 + 1.5 + 1.5 = 22.5\text{TB}$

After 64 days (because DPM only does up to 64 recovery points), DPM short term disk storage =  $15 + 64 \times 1.5 = 111\text{TB}$

## ROI Model 1: Combining DPM Short-Term Disk with CRUNCH Virtual Tape and Deduplicated Containers

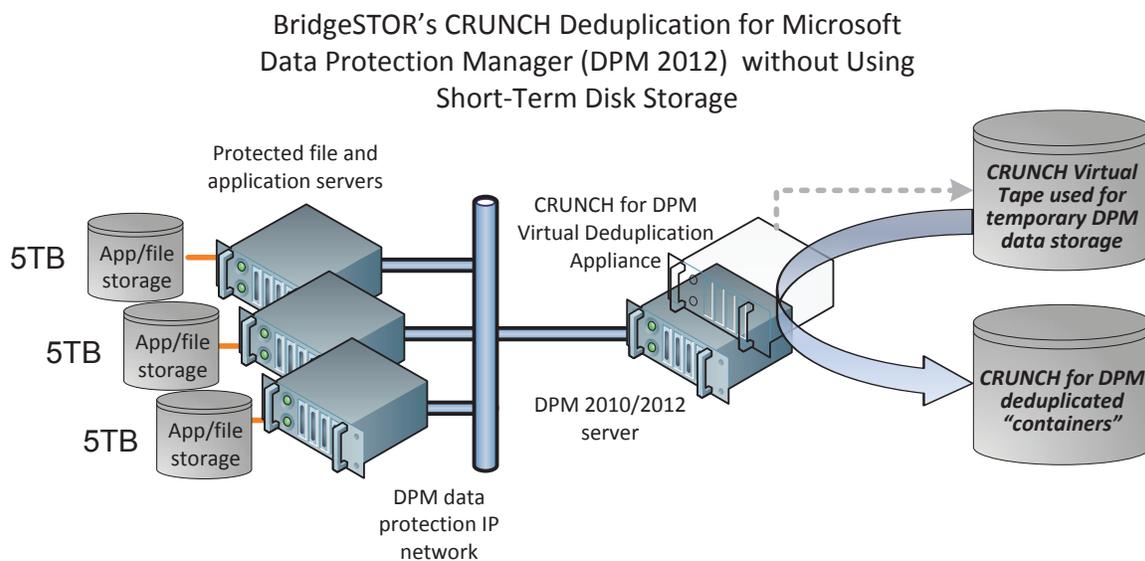


- Start with 3 application/file servers each with 5TB of “data”
- DPM performs the initial replica, copying 3 x 5TB = 15TB to DPM’s short term disk storage
- Data changes at the rate of 10% per day
- After day 0, there is 1.5TB of new/changed data stored on DPM’s short term disk storage
- After day 1, there is another 1.5TB of new/changed data stored on DPM’s short term disk storage (total changed data, 3TB. Total data 15TB + 3TB = 18TB)
- After day 2, there is another 1.5TB of new/changed data stored on DPM’s short term disk storage (total changed data, 4.5TB. Total data 15TB + 4.5TB = 19.5TB)
- This continues
- After one week (5 days), DPM short term disk storage = 15 + 1.5 + 1.5 + 1.5 + 1.5 + 1.5 = 22.5TB
- Also after one week, the 22.5TB described above are written to CRUNCH VT requiring an ADDITIONAL 22.5TB of capacity (even though this capacity is temporary and will be reused)
- The 22.5TB is CRUNCHEd to containers. Data reduction is assumed to be 65%. CRUNCH containers will then require  $22.5 * (1 - 0.65) = \sim 8\text{TB}$
- CRUNCH VT capacity requirements can be further reduced by configuring more protection groups, each having smaller capacity requirements. In the example, creating three protection groups (one for each application and file server), coupled with CRUNCH scheduling, could reduce the physical capacity

requirements of CRUNCH VT storage by as much as 2/3.

- All subsequent recovery points are CRUNCHED. Assume a 65% reduction in size.
- After 64 days, DPM short term disk storage = 22.5TB
- Virtual tape capacity required = 22.5TB
- CRUNCH container storage =  $8 + 64 * 1.5 * (1 - 0.65) = \sim 42$  TB
- Total of all storage = 42 (CRUNCH) + 22.5 (short term) + 22.5 (virtual tape) = 87TB
- Compared to 111TB without CRUNCH or about a 21.6% reduction overall after reaching 64 recovery points

## ROI Model 2: Eliminating DPM Short-Term Disk with CRUNCH Virtual Tape and Deduplicated Containers



- Start with 3 application/file servers each with 5TB of “data”
- Data changes at the rate of 1% per day
- There is NO short term DPM disk storage
- Data is backed up through DPM directly to CRUNCH-deduplicated disk storage daily and can be maintained as long as 99 years for long-term protection
- File and folder data recovery are performed through Windows Explorer
- Disk capacity used for CRUNCH virtual tape (temporary storage awaiting deduplication) can be as little as 5TB (backing up in 5TB increments) to 15TB (backing up all servers in a single job)
- After day 0, there is 15TB of deduplicated data residing on CRUNCH deduplicated container storage. CRUNCH typically reduces an initial backup by up to 65% meaning that CRUNCH may require as little as 5.25TB (up to

about 9.75TB depending on your data's "reducibility") to store the protection data for all the servers

- After day 1, CRUNCH containers will grow by the amount of unique data change ( $1\% \times 15\text{TB} = 0.15\text{TB}$ )
- After day 2, there is another 0.15TB of new/changed data stored on CRUNCH deduplicated container storage
- This continues
- After one week (5 days) of daily DPM backups, CRUNCH deduplicated container storage will increase by the amount of the deduplicated storage ( $5 \text{ days} \times 0.15\text{TB}/\text{day} = 750\text{GB}$  before deduplication)
- After 64 days: Virtual tape capacity remains at 5 to 15TB based on your data protection strategy
- CRUNCH container storage will be dependent upon the reducibility of your data, but may be as little as 5.25TB (up to about 9.75TB depending on your data's "reducibility") plus  $64 \times 0.15\text{TB}$  (64 daily changes (9.6TB)) or roughly 15TB to 20TB

Compared to 111TB without CRUNCH, this strategy reduces DPM storage requirements by as much as 80% to 90% overall after reaching 64 recovery points.