



white paper



**SECURE DATA PROTECTION WITH DOT HILL'S
BATCH REMOTE REPLICATION**

Replication



TABLE OF CONTENTS

1	Scope	3
2	Prerequisites	3
3	Introduction and Overview.....	3
4	Design Goals	4
5	Data Transport.....	5
6	Batch Replication Example.....	5
7	Summary	9
8	Definitions.....	10

TABLE OF FIGURES

Figure 1	7:00 AM Remote Volume Copy	6
Figure 2	8:00 AM Snapshot Replication.....	7
Figure 3	9:00 AM Snapshot Replication.....	8

1 Scope

Dot Hill's batch remote replication service is part of Dot Hill's R/Evolution Storage Architecture, which encompasses Dot Hill's Data Management Services, or DMS. DMS includes Dot Hill's new snapshot service as well as our new remote replication service. Both software modules are at the core of Dot Hill's data protection strategy. Please refer to Dot Hill's white paper called ***"Snapshots and Data Protection: The Dot Hill Difference"*** for a high-level overview of Dot Hill's snapshot service.

This white paper provides a high-level overview of the Dot Hill's new asynchronous batch remote replication service.

2 Prerequisites

Please note that any time "snapshot" or "snapshot service" is mentioned in this white paper, it refers to block-level sparse snapshot functionality. Asynchronous block data remote replication refers to the asynchronous duplication of data between distinct subsystems. A basic understanding of the functionality provided by block-level snapshots and asynchronous block data replication is assumed. Concepts and terms that are unique to Dot Hill's new batch remote replication service are defined below.

3 Introduction and Overview

Dot Hill's batch remote replication service consists of the asynchronous replication of block-level data from a volume at a local system to a volume at a remote system. This data is batched and duplicated to one or more remote systems by moving the data contained within individual point-in-time snapshots from a local system to one or more remote systems across a transport medium such as TCP/IP (iSCSI) or fibre channel. Batch remote replication will allow a user to designate one or more remote systems as targets that will receive the batch replication data.

The system administrator must create an Array Partition on the remote system(s) that are at least the same size as the array partition of the volume on the local system that is to be replicated, and the volumes are associated with one another in a replication set.¹ It is important to understand that batch replication occurs at the array partition level and therefore the array partitions on the remote systems may have different underlying RAID levels, drive sizes or drive types than the array partition on the local system.

Before the data can be transferred from a volume at a local system to the remote system(s), it must be image stabilized. As mentioned in the definitions above, the data must not change while it is being moved from the local system to the remote system. Dot Hill's snapshot service is used to provide image stabilization of the data. Since a snapshot represents a "point-in-time" state of a volume, it is not possible for the data to change while it is being moved from the local system to one or more remote systems.

Dot Hill's batch remote replication basically consists of scheduling point-in-time snapshots of a volume on the local system in order to image stabilize the data, then transferring the snapshot

¹It is important to understand that remote system volumes must be at least as large as the local system volume, but that it may not be possible to make the volumes exactly the same size due to differences in drives, RAID configuration, etc. Therefore, the remote system volume must be at least the same size as the local system volume being replicated; any excess space on the remote system volume will not be used.

data changes over a TCP/IP (iSCSI) or fibre channel connection to one or more remote systems. Each batch replication transfer consists of the delta data between two successive point-in-time snapshots, or in other words, how the data has changed since the last batch replication. This batch of data is transferred to the remote system(s). For the remainder of this document, snapshot data that has changed since the previous snapshot will be referred to as “snapshot delta data”.

Initially, the local and all remote volumes in a replication set must be synchronized. This may be performed via a remote volume copy or via some other means. Once the volumes have been synchronized, batch replications can be started. For batch replication, the snapshot delta data will be asynchronously transferred from the local system to the remote system and stored on the remote system in the form of snapshot data. Therefore, when a remote system has received all of the delta data for a point-in-time snapshot, the data will be stored on the remote system in the form of a snapshot. This snapshot delta data can be either manually or automatically applied to the volume on the remote system by initiating a roll-forward operation². For example, it is possible for a local system batch replication job to take a snapshot at 7 AM, 8 AM, and 9 AM. The 7:00 AM snapshot data is first transferred to the remote system and is stored on the remote volume array partition. The 8 AM and 9 AM snapshot delta data are also respectively transferred to the remote systems and are stored in the form of snapshots. It is therefore possible to roll-forward the state of the volume on the remote systems to the snapshot state at 8 AM, or 9 AM. After snapshot delta data for an individual point-in-time snapshot has been successfully transferred to all of the remote systems, the snapshot delta data on the local system can be optionally deleted, either manually or automatically, since it is no longer needed.

Similarly, after the delta data from the oldest snapshot has been applied to the volume at the remote systems, it can be optionally deleted, either manually or automatically, since it is also not needed. These delete operations will minimize the consumption of memory and backing store space on both the local and remote systems. On the other hand, if the batch replication data is kept, the local and/or remote systems can be rolled forward or back to any existing snapshot point.

4 Design Goals

The primary design goal of Dot Hill’s batch remote replication is to utilize existing Dot Hill snapshot code mechanisms, algorithms, data structures and metadata. Since the batch replication data is stored on the remote systems in the form of snapshot data, it is possible to use existing code, algorithms and data structures that perform the following functions:

1. Creation of snapshots
2. Deletion of snapshots
3. Roll-back of snapshots
4. Management of snapshots
5. Failover code that ensures a consistent state in case of RAID controller failure

It is also important to note that since existing snapshot code, algorithms and data structures are used to implement a large portion of the batch replication functionality, the risk of virus infection is significantly reduced as a result of the use of the batch remote replication service.

²Note that the roll forward operation is functionally equivalent to a rollback operation in Dot Hill’s snapshot architecture.

5 Data Transport

Replication data will be transported from a local system to the remote systems using either TCP/IP (iSCSI) or fibre channel. Destination system addresses will be configured on the local system using a standard Dot Hill management interface, including CLI, WBI or SANscape. Retries and other guaranteed data delivery mechanisms will be built in to the protocol to ensure that the data is delivered from the local system to the remote systems without error.

When transferring the snapshot delta data from the local system to the remote system or systems, the local system will maintain a high water mark that indicates the progress of the data transfer to *each* of the remote systems. It is necessary for the local system to maintain this state information for each remote target system because the state of the connection and the quality of the connection to each remote system cannot be guaranteed.

For example, the connection from the local system to remote system #1 may provide excellent bandwidth while the connection to remote system #2 may be intermittent with little or no data transfer for several minutes at a time. In this case, the snapshot delta data transfer to remote system #1 will be “ahead of” the data transfer to remote system #2; hence the need for a water mark that maintains the address of the next data chunk that is to be transferred to each remote system.

6 Batch Replication Example

This section provides an example of batch replication from a local system to a remote system and provides pictorial representations of the snapshot delta data replication process. Batch Replication is configured to snapshot and subsequently replicates the data at 7 AM, 8 AM and 9 AM.

Since the Batch Replication process begins at 7 AM, it is necessary for the local system to take a snapshot at 7 AM in order to image stabilize the local volume. The 7 AM snapshot is taken and a remote volume copy of all the data as it existed on the local system volume at 7 AM is performed. That is, the data for all blocks on the volume from 0 to n-1 as they existed at 7 AM are transferred to the remote system. The remote system will write all of the blocks from the 7 AM image stabilized state to the remote volume. After all of the data blocks have been successfully transferred to the remote volume, the remote volume now has the state of the local volume at 7 AM. Please see Figure 1.

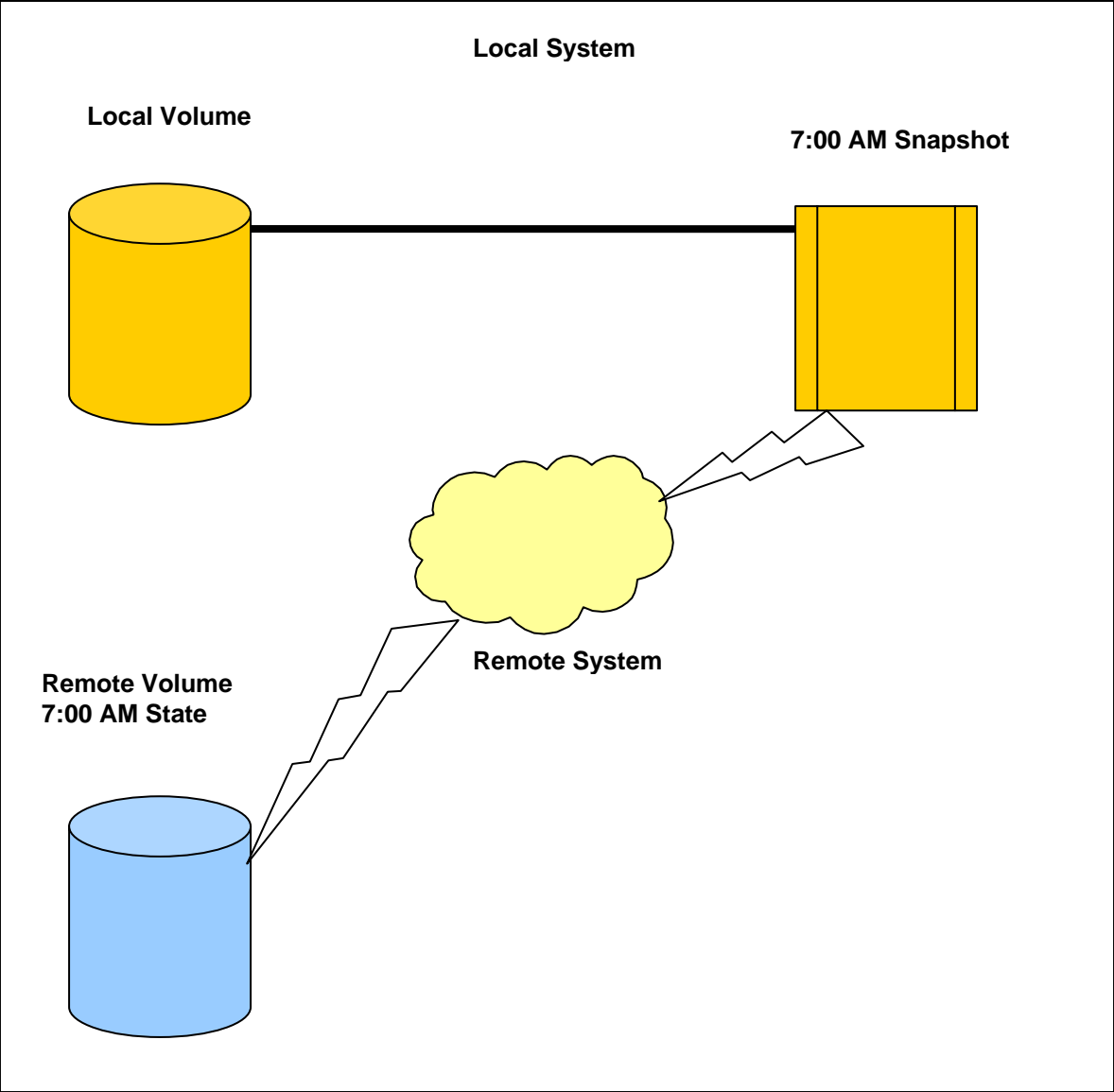


Figure One: 7:00 AM Remote Volume Copy

Batch replication on the local system will take another snapshot of the local volume at 8 AM in order to image stabilize the volume. The new data that has been written to the volume within the time period of 7 AM to 8 AM will now be replicated to the remote system and stored on the remote system as the 8 AM snapshot. Please see Figure 2.

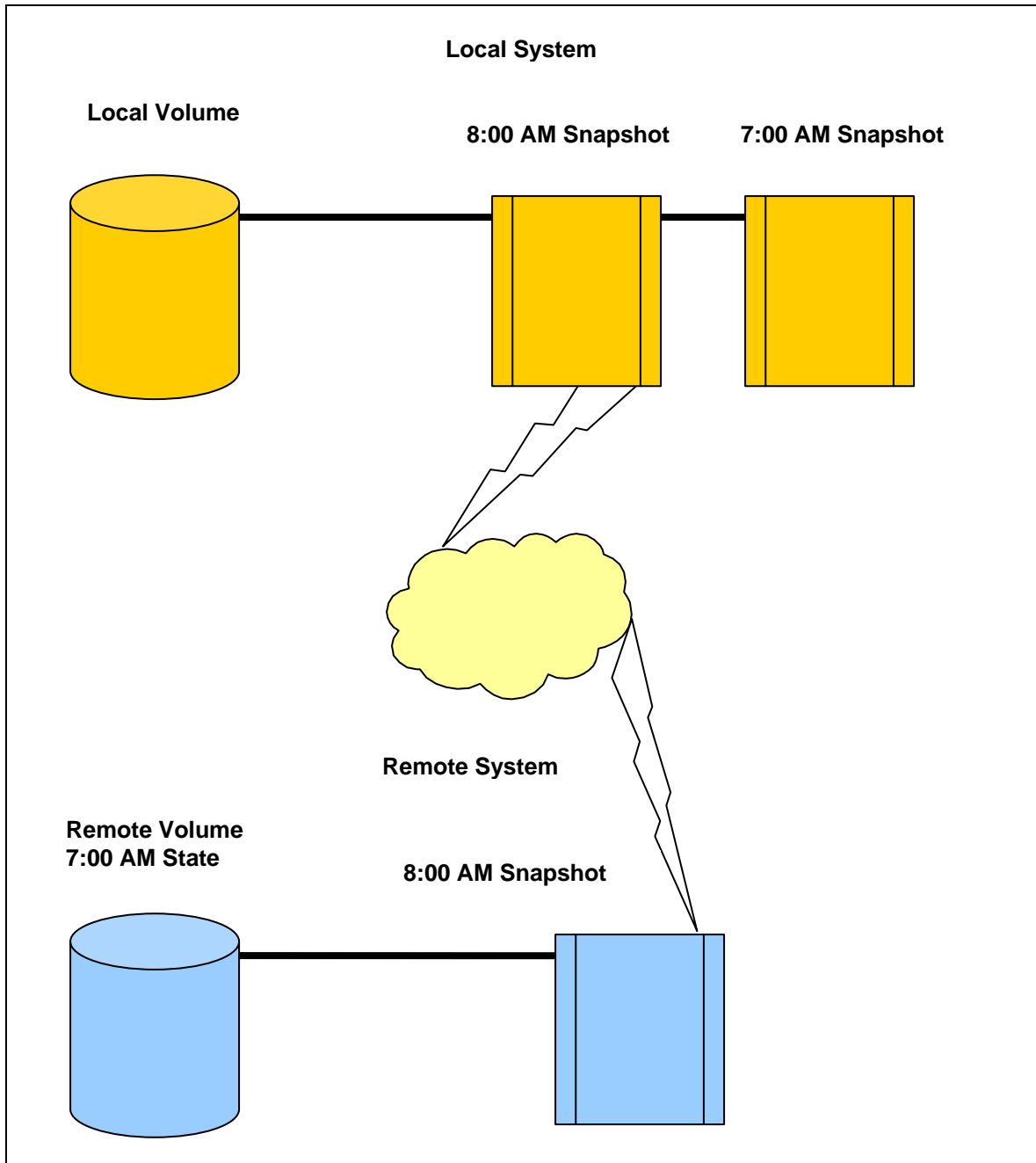


Figure Two: 8:00 AM Snapshot Replication

Batch replication on the local system will take another snapshot of the local volume at 9 AM in order to image stabilize the volume. The new data that has been written to the volume within the time period of 8 AM to 9 AM will now be replicated to the remote system and stored on the remote system as the 9 AM snapshot. Please see Figure 3.

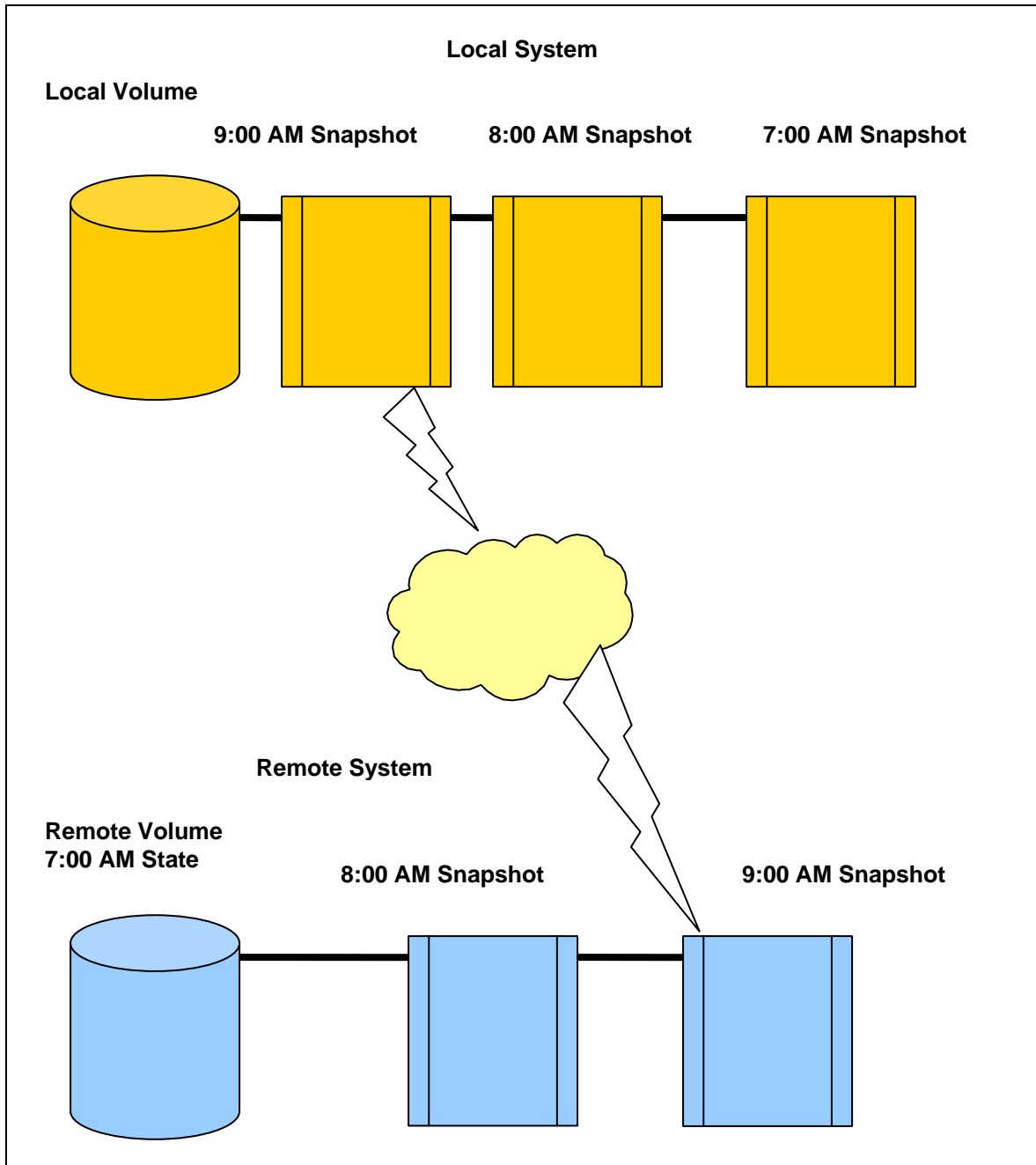


Figure Three: 9:00 AM Snapshot Replication

Since the batch replication data is stored on the remote system as point-in-time snapshots, it is possible for the system administrator to initiate a rollback of the remote volume using the standard snapshot rollback functionality. Therefore, the system administrator could choose to

maintain the remote volume in the 7 AM state, or initiate a rollback operation to roll the volume forward to the 8 AM or the 9 AM state. Please see Figure 3.

It is also possible for the remote system to apply the snapshot delta data to the remote volume immediately upon receipt of the snapshot data. For example, after receiving the snapshot delta data for the 8 AM snapshot, the remote system could immediately apply the snapshot delta data to the remote volume in order to update the state of the remote volume to the 8 AM state. The 8 AM snapshot could then be optionally deleted in order to conserve backing store space.

It is also important to note that the local system can optionally delete a snapshot after transferring all of the snapshot delta data to the remote system. For example, after the local system transfers the data for the 7 AM snapshot to the remote system, the local system could optionally delete the 7 AM snapshot because the data is no longer needed.

7 Summary

Dot Hill's new batch remote replication service provides asynchronous remote replication functionality by building on the existing Dot Hill snapshot service. As mentioned at the beginning, Dot Hill's new remote replication and snapshot services are part of Dot Hill's R/Evolution Storage Architecture, which encompasses Dot Hill's Data Management Services, or DMS. Both software modules are at the core of Dot Hill's data protection strategy.

Dot Hill's new remote replication provides a solid, robust means of replicating data to a remote site, with the added benefit of being able to store multiple point-in-time instances. This provides a means of keeping the data offsite and relatively current while providing additional protections, such as the ability to roll forward and backward between replication snapshot instances and the ability to utilize these snapshots for protection, tape backup or other purposes such as accounting, testing or analysis.

About Dot Hill

Dot Hill is the market leader in providing flexible storage offerings and responsive service and support to OEMs and system integrators, from engagement through end of life. Founded in 1984, Dot Hill has more than two decades of expertise in developing high-quality, competitively priced storage products. Focused on delivering global 24x7x 365 technical support, the company has more than 100,000 systems in use worldwide as well as numerous OEM and indirect partners. With its patented technology and award-winning SANnet II® and RIO Xtreme™ families of storage and its Dot Hill Storage Services, Dot Hill makes storage easy. Headquartered in Carlsbad, Calif., Dot Hill has offices in China, Germany, Israel, Japan, Netherlands, United Kingdom and the United States. More information is available at <http://www.dothill.com>.

8 Definitions

Array: One or more disk drives across which raw host data is spread for purposes of redundancy and performance. A RAID set.

Array Partition: Is a logical disk, often referred to, simply, as a “partition.” An array partition provides access to a virtual disk drive which can read or write fixed blocks through logical block addressing, or LBA, which is a technique that allows a computer to address a hard disk that is larger than a certain capacity. An array partition can have an assigned valid LUN number, in which case it is host addressable.

Asynchronous Batch Replication: See “Batch Replication”.

Asynchronous Replication: Replication (duplication) of data from one source (local system) to one or more targets (i.e., remote systems). This replication is done in an asynchronous manner. That is, I/Os may be complete on the local system and response sent back to the host prior to the replication being completed (or in some cases not even started) on the remote system.

Backing Store: A backing store is the location where all snapshot data is stored – whether preserved for a snapshot or written directly to a snapshot. It occupies an array partition but does not have a LUN number assigned to it. This makes it invisible to the host operating system. The backing store will never be mapped to a LUN.

Batch Replication: Asynchronous remote replication of block level data from a volume at a local system to a volume at one or more remote systems by moving the data contained within individual point-in-time snapshots from a local system to one or more remote systems across a transport medium such as TCP/IP (iSCSI) or fibre channel.

Container: Is a structure that is used to reference an array or an array partition. Context is based on the function.

Copy-On-Write, or COW: The process by which snapshot data is preserved to the backing store. A copy-on-write operation is typically kicked off when a new write to the master volume occurs, requiring that the existing data on the master volume be preserved for the snapshot(s) which need the existing data.

COW: see Copy-On-Write above.

Image Stabilization: The act of freezing the state of block-level data at a point-in-time such that the block-level data remains constant and cannot change. Point-in-time snapshots are used to “image stabilize” block-level data for use by a backup utility or a data replication process. In other words, the data must not change while it is being moved from the local system to the remote system.

LUN: A logical unit number. More precisely, it refers to a LUN identifier, which is an integer value which identifies a set of host address mappings, which default to LUN identifier N = LUN N. Each LUN maps 1:1 to an array partition describing the logical disk.

Master Volume: Is a physical volume which contains LUN information. It is equivalent to a traditional array partition in current RAID controller terms but has been enabled for snapshots.

Original Volume: Alias for Master Volume.

Preserved Data: Data that is preserved for a snapshot or set of snapshots. That is, it is data that is saved as part of a copy-on-write, or COW, operation.

Promote Snapshot: Alias for rollback.

RAID Set: A group of devices that form a single RAID entity, such as a RAID 5 array. Equivalent to an array.

Remote Volume Copy: A volume copy performed over some transport such as TCP/IP (iSCSI) or fibre channel.

Rollback: If a critical error occurs on a master volume for any reason, the administrator may elect to replace the master volume information with the master volume data as it existed at an earlier point in time. The process of replacing the master volume data with older data is called rollback, since this rolls the data on the master volume back to an earlier point in time.

Snapshot: A snapshot is a point in time capture of data on a master volume (or another snapshot – not supported in initial implementation). The snapshot itself will define an array partition but is a virtual volume. That is, the snapshot volume itself does not exist. Instead all data associated with a snapshot volume resides either on the master volume or on the backing store associated with the snapshot.

SN: Serial Number. Each array partition, including each snapshot, will have a unique serial number assigned to it.

Volume Copy: Copies all data that existed on a volume at a specified point in time (i.e., where a snapshot is taken) and copies it to another volume, thus creating a fully independent copy of the snapshot.

Write Data: On a snapshot, the snapshot write data is data that is written directly to a snapshot. This is different from preserved data associated with a snapshot.