

OpenVMS Technical Journal V6

Case Studies Using EMC Legato NetWorker for OpenVMS Backups



Case Studies Using EMC Legato NetWorker for OpenVMS Backups.....	2
Overview	2
Introduction	2
Case Study 1: Backup of Data at Two Sites	2
Case Study 2: Backup of a Cache Database as Part of a Heterogeneous Medical Information System....	5
Summary	6
For more information	6

Case Studies Using EMC Legato NetWorker for OpenVMS Backups

Siobhán Ellis, Senior Technical Consultant, ENSTOR, Sydney, Australia

Overview

EMC Legato NetWorker is the only product that incorporates OpenVMS into an enterprise-wide data protection solution providing direct backup to tape or disk as well as support for Oracle and Oracle Rdb. This article covers two real solutions that were delivered – one using EMC Symmetrix and CLARiiON Disk Libraries, and the other using host-based OpenVMS volume shadowing, HP storage, and an MSL5000 tape library.

Introduction

Backups are critical for any operating system and OpenVMS is no exception. There have been many solutions on the market, including offerings such as SLS and ABS from HP. Until recently, nearly all of these solutions concentrated on using BACKUP as the data mover. One of the products to break that paradigm is EMC Legato NetWorker.

By using its own data mover, NetWorker can write multiple streams of data in parallel to a single backup device. NetWorker also writes that data in a format that is operating system independent, which means that a tape written on an OpenVMS system can be read by a Windows system. NetWorker can often back up an entire OpenVMS system much faster than the traditional single stream method because the bottleneck often is not the tape device itself, but the process of getting the data off the disk fast enough to keep the tape device spooling. In one such case, NetWorker cut backup times in half without changing the hardware.

NetWorker is also extremely flexible. It provides the ability to share tape devices, tapes, and libraries among systems – even among operating systems – to maximize the use of resources.

This article describes two real backup solutions using EMC Legato NetWorker on OpenVMS.

Case Study 1: Backup of Data at Two Sites

The Scenario

The customer has two datacenters some distance apart. One site is the production site and the other is the disaster recovery site, which also houses development systems. The two sites are connected using TCP/IP and DECnet. There are about 90 OpenVMS servers spread across the two datacenters.

Data is stored on EMC Symmetrix disk arrays. Production data is replicated from Site 1 to Site 2 using EMC SRDF, an array-based duplication, over FCIP (fibre channel over IP) using CISCO MDS. All data, production, and development is in a RAID 1 set; this is locally copied using EMC TimeFinder to create BCVs (business continuation volumes), which is essentially a third copy of the data.

All OpenVMS systems have two fibre channel host bus adapters (HBAs) to provide dual paths to the data, and they are all in clusters. They were currently being backed up using a mixture of manual procedures and TAPESYS.

The Requirements

The customer wanted a solution for their OpenVMS systems only. Their UNIX and Windows systems were already covered by another backup solution from another vendor. Because of stability issues with the other backup product, they were not considering using it for their OpenVMS solution.

The customer needs were:

- A modern backup solution with the following features:

- Extensive reporting
- Administrative graphical user interface (GUI)
- Security
- Reduced backup times
- Ability to keep all backups for two months and occasionally keep a backup for 10 years
- High availability
- No operational impact on the clusters being backed up
- Multiple copies of the data without moving data over the corporate IP data network

The Solution

Figure 1 shows the solution that was designed for this customer's needs. The solutions chosen to address individual needs are discussed in the sections following the figure.

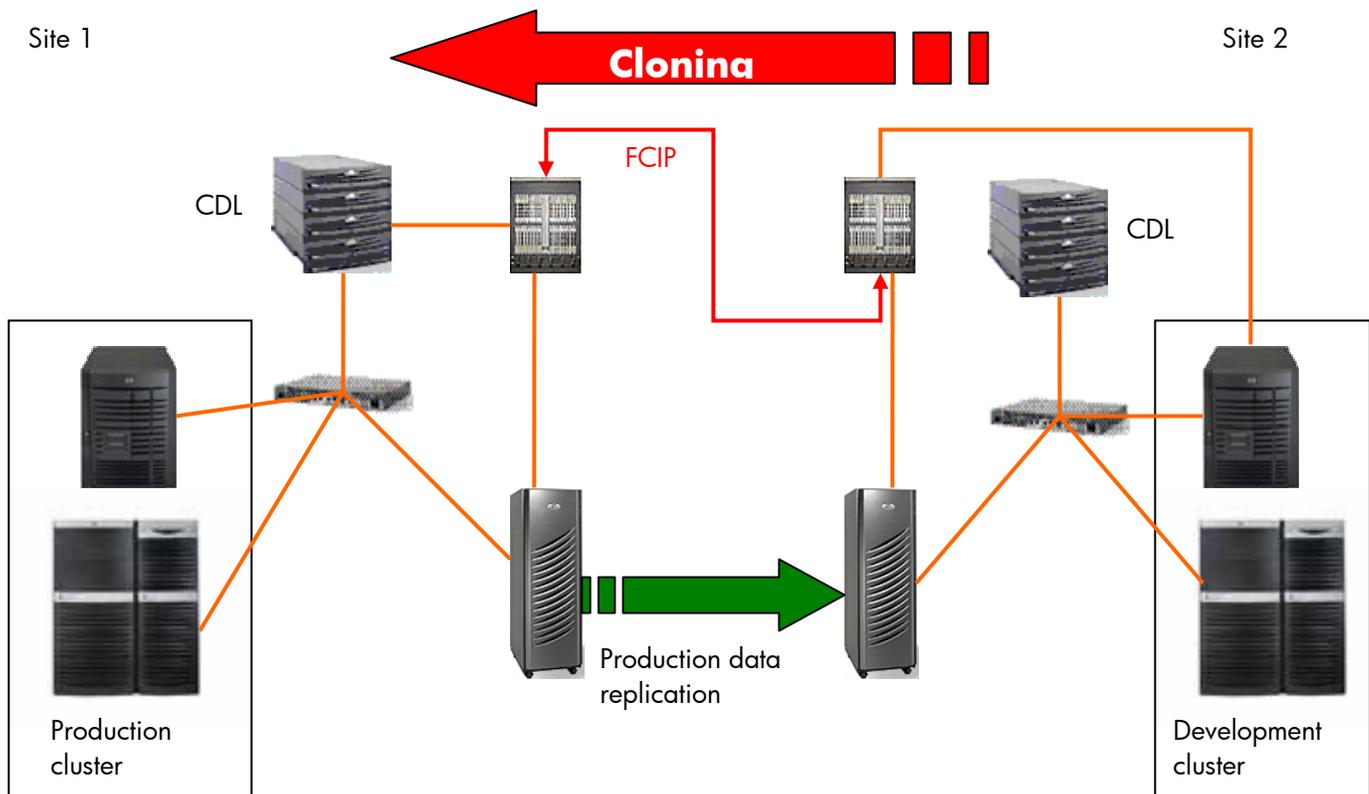


Figure 1 Production Cluster and Development Cluster Connections to SANs and CDLs

A Modern Backup Solution

EMC Legato NetWorker V7 was chosen because of its flexibility to deliver a solution that meets all the basic requirements. In conjunction with NetWorker Management Console (NMC), this solution provides a secure environment where the operations staff can be limited to performing only certain functions using a Java-based, web-start application. NMC also provides sophisticated and extensive reporting, and all actions taken on NetWorker are recorded.

Reduced Backup Times and Storage Capacity

It was determined that a disk-to-disk solution would provide the highest possible throughput for both restores and backups. However, with the amount of data required to be stored, EMC Symmetrix disk was too expensive to use as a backup device. The solution required lower cost disks such as serial ATA disks. EMC's mid-range disk solution is CLARiiON, which, in the spring of 2005, does not support OpenVMS. However, EMC had just launched the CLARiiON Disk Library (CDL), which was a perfect solution.

The CDL is a virtual tape library. It is essentially a collection of fibre-attached, serial ATA disks in a cabinet fronted by a high-availability pair of Linux servers that emulate a tape library using FalconStor VirtualTape Library (VTL) software. HP has just launched its equivalent, the StorageWorks 6000 Virtual Library System.

A CDL was implemented at each site with enough space to keep all data for two months. To ensure that the CDL is always available using fibre channel, two ports were dedicated to local data backup.

Each CDL was set up to emulate an ATL P3000 tape library, which is physically the same as an HP ESL9000, with 8 SDLT 320 drives. Site 1 also had a smaller library set up and was connected to the local MDS, which is described later.

High Availability

The NetWorker server, which holds the media database and client file indexes and controls the execution of backups, was installed on a Sun Solaris server at Site 1. To comply with the customer's standards, high availability was achieved by having a second server available at Site 1 and having the metadata stored on a partition on the Symmetrix; this data was then duplicated to the remote site where a third server was set up and ready to take over from Site 1 if necessary.

As already noted, the CDLs contain a pair of high availability Linux systems. All the disks are configured in RAID 5 sets, giving the CDLs a much better MTBF (mean time between failures) than can be provided by a real tape library.

No Operational Impact on the Clusters Being Backed Up

The customer had recently consolidated their OpenVMS systems, moving many servers onto fewer GS Series servers. This move had freed up a number of smaller AlphaServers, which could now be used as backup appliances in each cluster and as NetWorker storage nodes. An AlphaServer 4000 could push four data streams at once before saturating the CPU, at which point any more streams would degrade overall performance. Each storage node was connected to the local CDL using two fibre channel host-based adapters. At Site 2, one storage node was also connected to the local MDS so it could see the small virtual tape library at Site 1.

Multiple Copies of the Data Without Moving Data Over the Corporate IP Data Network

Because the production data is replicated at both sites, all that was required was to split the business continuation volumes (BCVs) simultaneously at both sites and back up the data locally. This was done using the preprocessing and postprocessing capabilities of NetWorker, which allow you to run command procedures before and after the backup on a particular client. The steps are as follows:

1. Preprocessing:
 - a. Breaks each disk in the backup saveset out from the BCVs.
 - b. Mounts each disk locally with the correct logical name, so it supersedes the SYSTEM logical.
2. The backup runs and backs up the mounted BCV disk.
3. When the backup is complete, postprocessing does the following:

- a. Dismounts the disks.
- b. Puts the disks back into the BCVs.

Because clients at both sites can split BCVs at the same time, duplicate backups are achieved, but are backed up directly to local CDLs.

The development data was a different matter because it is not replicated from Site 2 to Site 1. Again, the preprocessing and postprocessing capabilities were used to split the BCVs just as with the production data. However, after the backups are finished, NetWorker initiates an automatic clone operation to perform a copy of all the development savesets from the CDL in Site 2 to the small CDL in Site 1.

It took about three weeks to develop and test the preprocessing and postprocessing command procedures.

The Result

All the customer requirements were met, and the solution went into production. With no special tuning, one product system was backed up and cloned in the time it used to take to perform just the backup. This was done with a single clone stream running. It is possible to initiate multiple clone streams and cut cloning time by two thirds.

Case Study 2: Backup of a Cache Database as Part of a Heterogeneous Medical Information System

The Scenario

The customer has a number of healthcare information systems based on the Caché database on OpenVMS, and a number of web, file, and SQL servers running on Microsoft Windows. Storage for OpenVMS is connected using HSG80 controllers, and all disks are shadowed using host-based volume shadowing. Windows systems have direct-attached SCSI disks. OpenVMS was being backed up using scripts based on BACKUP, and Windows was backed up using a workgroup class backup tool.

The Requirements

Customer wanted a solution with the following capabilities:

- Works for both Windows and OpenVMS
- Does not require applications to stop running.
- Runs automatically so support staff can go home at a reasonable hour.

The Solution

A NetWorker server running on Windows with OpenVMS storage nodes filled the bill. Windows web servers, file servers, and SQL servers all are covered by the standard NetWorker and associated modules. Backup is done to an MSL5000 series library, which is shared between Windows and OpenVMS.

The only issue concerned the database because, unlike Oracle or Oracle Rdb, Caché does not have an API with which a backup tool can interface. However, Caché does have a command line that allows you to put the database into a suspend mode, and then later into a resume mode.

The new backup flow is as follows:

1. Start the Backup.
2. Use NetWorker preprocessing to do the following:
 - a. Suspend the Caché database.
 - b. Drop disks from the shadow sets.
 - c. Mount the disks in the group with their correct logical names, thus overriding the system logicals.
 - d. Resume the database.
3. Back up the disks on the system.
4. Use NetWorker postprocessing to minimerge the disks back into the shadow sets.
5. Automatically clone the data to ensure there are two copies, one for offsite and one for onsite.

The command procedures took about one week to write and test.

The Result

Backups across Windows and OpenVMS are running reliably and consistently. The applications do not shut down and support staff no longer needs to worry whether the data is being backed up.

Summary

OpenVMS is an excellent operating system for many reasons, some of which are its robustness and reliability. Consequently, system managers are often reluctant to use a backup utility that is not native to OpenVMS. While NetWorker performs no magic tricks, it does use the RMS APIs for OpenVMS, thus guaranteeing correct backup and restoration of data because it is using basic and legal OpenVMS system calls. Indeed, in tuning NetWorker to work best with OpenVMS, the NetWorker for OpenVMS engineering team worked closely with the RMS engineering team.

In these two scenarios, one using EMC storage and the other using HP storage and OpenVMS volume shadowing, NetWorker on OpenVMS was easily customized to meet the particular requirements of each customer to deliver a reliable data protection solution. These solutions were chosen as examples because of their complexity and the author's intimate involvement with the service delivery. There are many standard installations of NetWorker on OpenVMS where no customizations are performed at all. There are also other variants, particularly with the Caché database, where HP's EVA storage is used and snapshots are performed rather than splitting shadow sets.

With the addition of NetWorker support for Oracle and Oracle Rdb databases, the possibilities of using NetWorker on OpenVMS are boundless with minimal customization. Using NetWorker on OpenVMS, you can help to bring OpenVMS operations into the enterprise, reducing the total cost of ownership by sharing backup resources with other operating systems and reducing training needs because NetWorker works across many operating systems.

NetWorker is the most mature and best heterogeneous data protection tool available. Now that these capabilities are available on OpenVMS, the best operating system, the combination of NetWorker and OpenVMS makes an unbeatable backup solution.

For more information

See the ENSTOR web site at www.enstor.com.au.

Find out more about the EMC Legato NetWorker on the [NetWorker product page](#).

Find out more about NetWorker on OpenVMS in the [OpenVMS Journal V4](#) and [OpenVMS Journal V1](#).

© 2005 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice. The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

