

OPTIMIZED BACKUP AND RECOVERY FOR SAP LANDSCAPES ENABLED BY EMC DATA DOMAIN AND EMC NETWORKER

A Detailed Review

EMC SOLUTIONS GROUP

Abstract

This white paper describes how backup and recovery operations for SAP systems benefit from inline deduplication technology by using the EMC® Data Domain® deduplication storage system with EMC Data Domain Boost (DD Boost) software and EMC NetWorker®. This paper also discusses best practices for administration simplicity and efficiency by using the NetWorker Module for SAP with Oracle and the NetWorker PowerSnap module.

February 2012



Copyright © 2012 EMC Corporation. All Rights Reserved.

EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

The information in this publication is provided “as is.” EMC Corporation makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

For the most up-to-date listing of EMC product names, see EMC Corporation Trademarks on EMC.com.

VMware, ESX, VMware vCenter, and VMware vSphere are registered trademarks or trademarks of VMware, Inc. in the United States and/or other jurisdictions.

Intel and Xeon are trademarks of Intel Corporation in the U.S. and/or other countries.

All other trademarks used herein are the property of their respective owners.

Part Number H8205.1

Table of contents

Executive summary	5
Business case.....	5
Solution overview	5
Key results	6
Introduction	7
Purpose	7
Scope	7
Audience	7
Terminology	7
Technology overview	9
Overview.....	9
EMC Data Domain systems.....	9
EMC Data Domain Boost software	9
EMC NetWorker	9
EMC NetWorker Module for SAP with Oracle	9
EMC NetWorker PowerSnap Module for CLARiiON.....	9
EMC Data Protection Advisor.....	10
VMware vSphere	10
SAP ERP IDES	10
HP LoadRunner	10
Solution architecture.....	11
Solution overview	11
Hardware environment.....	12
Software environment.....	13
Design and configuration	14
Introduction	14
Key configurations	14
SAP storage design	15
DD Boost configuration	16
NetWorker configuration	17
NMSAP configuration	18
PowerSnap specific configuration	21
Data load configuration.....	21

- Testing and validation 22**
 - Test overview 22
 - Test procedure 23
 - Test results 25
 - Results analysis 30

- Conclusion 36**
 - Summary 36
 - Findings 36

- References 37**
 - White papers 37
 - Product documentation 37
 - Other documentation 37

- Additional information 38**
 - Overview 38
 - Backup utility profile 38
 - Backup utility profile for PowerSnap 38
 - PowerSnap configuration file 39

Executive summary

Business case

With the implementation of server virtualization technologies and the multiple copies of SAP systems required for testing purposes, SAP system landscapes are becoming more complex. The number of servers is increasing with virtualization. Every SAP system added to the landscape requires backup. Due to exponential data growth, shrinking budgets, and the increasing need for long-term retention required for governance with auditing processes, the backup demand becomes a challenge. SAP system administrators usually face problems in supporting service-level agreements (SLAs) in the following operation and maintenance phases:

- Completing backups within the designated backup window
- Executing backup tasks without affecting SAP production system performance
- Ensuring that data is recovered within the Recovery Time Objective (RTO) period
- Increasing network bandwidth efficiency during backup
- Reducing administration time spent on tape management, and optimizing the utilization of skilled system administration resources

Solution overview

This solution demonstrates complete next-generation backup and recovery for SAP with inline deduplication technology, enabled by EMC Data Domain duplication storage systems powered by the Intel® Xeon® processors, EMC NetWorker, and EMC Data Protection Advisor.

Data Domain systems and NetWorker are the central components in the solution to optimize the SAP-specific backup and recovery tasks, bringing new levels of simplicity and operational efficiency while providing a reduction in the Total Cost of Ownership (TCO). TCO is a very important factor to consider when supporting SLAs given exponential data growth and reduced IT budgets.

DD Boost significantly increases backup performance, reduces backup windows, and optimizes LAN bandwidth utilization. This is accomplished by distributing parts of the deduplication process to the backup client or to a NetWorker storage node.

NetWorker is used as the backup management tool to control the entire backup process and recovery operations. Through interaction with the NetWorker PowerSnap™ module, NetWorker performs off-host backups, eliminating the performance impact on the SAP production server during backup.

Data Protection Advisor (DPA) provides a comprehensive set of features to analyze data protection operations to ensure that your data is protected and recoverable. Analyzing the backup applications, supporting infrastructure, and target storage, DPA can highlight issues so that they can be addressed before a failure. In addition, DPA dramatically simplifies the evaluation of the metrics of the whole backup landscape.

Key results

This solution demonstrates the following benefits:

- Reduce infrastructure costs using Data Domain systems with deduplication ratios of up to 20:1
- Improve resource utilization using Data Domain systems configured with DD Boost, reducing LAN bandwidth consumption by up to 93 percent
- Reduce by 43 percent the backup windows using Data Domain systems with DD Boost
- Reduce by 16 percent the restore windows using Data Domain systems with DD Boost
- Offload the performance impact on SAP production system during backup with NetWorker PowerSnap Module

Introduction

Purpose

The purpose of this paper is to show that SAP customers can reduce long-term storage consumption while achieving better performance during backup and recovery operations by deploying Data Domain systems with Data Domain Boost software. This paper also explains the cost savings, added flexibility, and additional functionality provided by the validated solution.

Scope

This paper describes a solution that enables SAP customers to improve data protection for mission-critical systems by refining their backup and recovery strategy with inline deduplication technology.

Specifically, this paper describes:

- The deduplication ratio of backup data achieved with Data Domain systems
- The improved performance of backup and recovery operations enabled by DD Boost
- The seamless integration among NetWorker, Data Domain systems, and SAP systems using the NetWorker Module for SAP with Oracle (NMSAP)
- The functionality of clone-based backup enabled by PowerSnap
- The performance monitoring and analysis enabled by DPA

Audience

This white paper is intended for customers, partners, and EMC employees, including IT planners, SAP architects, basis administrators, and any other personnel involved in evaluating, acquiring, managing, operating, or designing an SAP landscape infrastructure.

The knowledge and experience of using BR*Tools for the backup and recovery of SAP with Oracle databases is beneficial.

Terminology

Table 1. Terms used in the white paper

Term	Definition
BACKINT	The external backup binary called by BR*Tools to perform SAP system backup at file system level. It is SAP certified.
Storage node	A host computer (including NetWorker server) that has the physical connection and ownership of the attached storage devices.
AFTD	Advanced file type devices, a NetWorker device type for disk based backup created on local disks or network-attached disks managed by the NetWorker storage node.
Proxy host	Also known as Data Mover or Mount Host, the secondary host, from which PowerSnap implements snapshot-based backup to prevent any performance impact on the backup client.
DEV	SAP development system, usually copied from SAP production system when initially-created
QAS	SAP quality assurance system, which is also referred to as test system, usually copied from SAP production system when initially-created

Term	Definition
PRD	SAP production system.
RTO	Recovery Time Objective is the maximum acceptable length of time that a SAP system can be down after a failure or disaster occurs.
NaviSecCLI	Navisphere Security Command Line Interface. It is used to configure and manage CLARiiON arrays.
BR*Tools	Toolset provided by SAP for Oracle database administration (DBA) tasks.
ESX	VMware hypervisor architectures, built on compatible physical servers, providing the foundation for building and managing a virtualized IT infrastructure.
Virtual Machine	A virtual machine is a tightly isolated software container that can run its own operating systems and applications as if it were a physical computer.
VMDK	Virtual machine disk files, a datastore descriptor.
VMFS	VMware Virtual Machine File System, a cluster file system that enables multiple VMware ESX hosts to concurrently access a single VMDK file.
RDM	Raw device mapping is a mapping file in a VMFS volume that acts as a proxy for a raw physical device. It offers VMFS manageability with the raw device access required by certain applications.
NMSAP	EMC NetWorker Module for SAP with Oracle

Technology overview

Overview

This section briefly describes the following key solution components.

- EMC Data Domain duplication storage system
- EMC Data Domain Boost
- EMC NetWorker
- EMC NetWorker Module for SAP with Oracle
- EMC NetWorker PowerSnap Module for CLARiiON
- EMC Data Protection Advisor
- VMware ESX server
- SAP ERP IDES
- HP LoadRunner

EMC Data Domain systems

EMC Data Domain deduplication storage systems reduce the amount of disk storage by identifying redundant files and data as they are being stored and providing a storage footprint that is 10x-30x smaller, on average, than the original dataset. Data deduplication is performed on incoming data streams and allows only the new and changed segments of data to be identified and stored as unique instances within the Data Domain file system. EMC® Data Domain® deduplication storage systems for backup and recovery use multi-core Intel Xeon processors to deliver faster data deduplication performance and network efficient replication while driving cost savings and efficiency. EMC Data Domain systems leverage the Intel SSE instruction set to increase the amount of parallelism and use system memory more efficiently, which significantly increases overall system performance.

EMC Data Domain Boost software

EMC Data Domain Boost software significantly increases performance by distributing parts of the deduplication process to the backup client or a NetWorker storage node. Without Data Domain Boost, the backup server sends all data—unique or redundant—to a Data Domain system for deduplication processing.

EMC NetWorker

EMC NetWorker backup and recovery software centralizes, automates, and accelerates data backup and recovery across your IT environment. NetWorker delivers centralized backup and recovery operations for complete control of data protection across diverse computing and storage environments.

EMC NetWorker Module for SAP with Oracle

The EMC NetWorker Module for SAP with Oracle (NMSAP) is an SAP-certified add-on module for NetWorker and provides a backup and restore interface between the SAP BR*Tool programs and the NetWorker server. Backint, part of NMSAP, provides the interface between BR*Tools and NetWorker.

EMC NetWorker PowerSnap Module for CLARiiON

The NetWorker PowerSnap Module for CLARiiON enhances NetWorker by allowing snapshot-based data protection in conjunction with the EMC CLARiiON storage systems. The backup is virtually instantaneous, regardless of the volume size. The PowerSnap Module software enables frequent backups with little impact on the network or application server.

EMC Data Protection Advisor

EMC Data Protection Advisor (DPA) is an advanced backup reporting and analysis application that gives users complete visibility into everything happening within their backup environment. DPA gathers data from backup servers, backup clients, databases, network switches, storage area network (SAN) switches, and tape libraries. DPA correlates information from each of these areas and provides the ability to perform advanced reporting, troubleshooting, performance management, and capacity planning. DPA provides excellent insight into VMware and Data Domain system environments, allowing the user to identify bottlenecks, load balance, and view the environment from differing perspective in one screen.

VMware vSphere

VMware vSphere is the foundation for a dynamic, self-optimizing IT infrastructure. It is a robust, production-proven virtualization layer that abstracts processor, memory, storage, and networking resources into multiple virtual machines. VMware vSphere allows enterprises to dramatically reduce hardware and operating costs by sharing resources across a virtual environment. The components of VMware vSphere include ESX, ESXi, and vCenter.

SAP ERP IDES

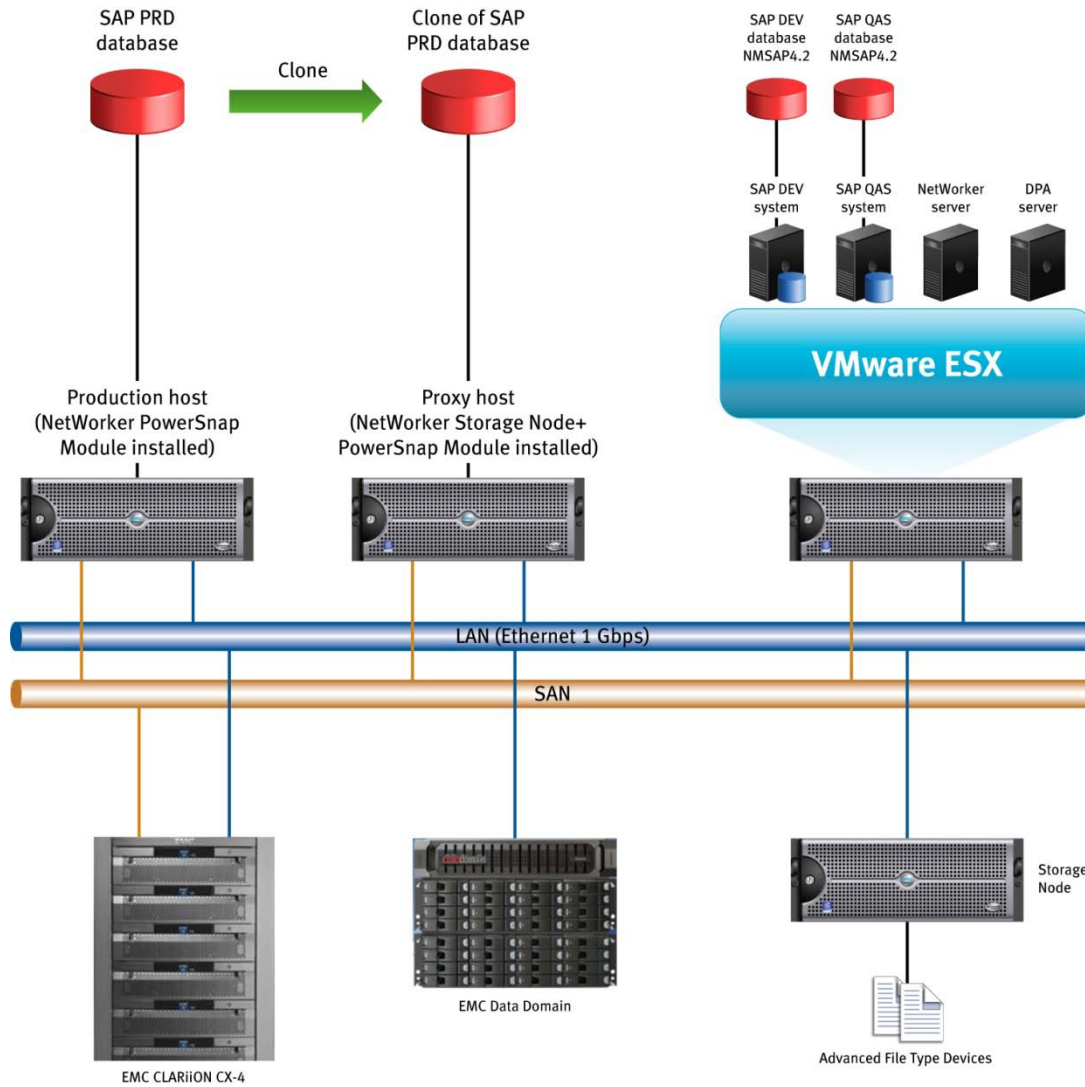
SAP ERP IDES is the "Internet Demonstration and Evaluation System" in the ERP Central Component (ECC) system and represents a model company. It consists of an international group with subsidiaries in several countries. IDES contains application data for various business scenarios that can be run in SAP systems. The business processes in the IDES system are designed to reflect real-life business requirements and have access to many realistic characteristics. IDES uses easy-to-follow business scenarios to show the comprehensive functions of the ECC system. The focal point of IDES, however, is not the functionality itself, but the business processes and their integration.

HP LoadRunner

HP LoadRunner is software that emulates concurrent users to apply production workloads to almost any application platform or environment. HP LoadRunner stresses an application from end to end—applying consistent, measurable, and repeatable loads—and then uses the data to identify scalability issues that can affect real users in production.

Solution architecture

Solution overview Figure 1 shows the high-level architecture of the solution environment.



GEN-001745

Figure 1. Architecture overview

To improve the backup performance, a dedicated network connection between the NetWorker storage nodes and Data Domain systems is recommended. Both 1 Gb and 10 Gb connection can serve for this purpose. A 10 Gb network connection provides ultimate connection speed but requires additional infrastructure costs to implement. In environments where 10 Gb connectivity is not available or affordable, 1 Gb connectivity can also be applied to achieve a private, high-bandwidth data connection, and avoid the latency and complexity of a shared Ethernet connection.

Refer to *EMC NetWorker Data Domain Deduplication Devices 7.6 Service Pack 1 Integration Guide* for more information. In this solution, a shared network topology is applied because only one backup job is running at a time, requiring less network bandwidth.

The SAP Development (DEV) and Quality Assurance (QAS) systems, the NetWorker server and Data Protection Advisor (DPA) are all installed and configured on VMware virtual machines.

To provide best performance and facilitate backup device management, a dedicated NetWorker storage node was set up as the scale-out NetWorker deployment best practice to manage all backups. For the SAP production system, the NetWorker PowerSnap module was installed and configured to mount a clone of the SAP production system on a Proxy host (temporary backup server). This Proxy (or mount) host also had the NetWorker storage node installed on it to offload the backup from the SAP production server, avoiding any performance impact on the SAP production system and maximizing the backup performance with a dedicated host.

The DPA server was used as a central performance management tool to monitor the performance of the backup and recovery operations on the SAP servers and storage nodes. However, for CPU usage, VMware recommends using vSphere client to capture data on the virtualized platform. Refer to *Performance Troubleshooting for VMware vSphere™ 4.1* for more information.

The SAP PRD system and the proxy host reside on physical servers, which are connected to the ESX server through a local area network. EMC CLARiiON is attached to all application servers through the SAN.

Hardware environment

Table 2 details the hardware environment for the use case.

Table 2. Hardware environment

Equipment	Quantity	Configuration
EMC CLARiiON	1	Model: CX4-480 25 x 450 GB 15 K FC disks 10 x 1 TB 7.2 K SATA disks
EMC Data Domain duplication storage system	1	Model: DD670 12 x 1 TB SATA 32 x 2 TB SATA
EMC Data Protection Advisor server	1	VMware virtual machine 4 x vCPU, 6 GB Memory
EMC NetWorker server	1	VMware virtual machine 4 x vCPU, 6 GB Memory
SAP DEV host	1	VMware virtual machine 4 x vCPU, 16 GB Memory
SAP PRD host	1	CPU Type: 4 quad-core 2.93 GHZ Memory: 16 GB

Equipment	Quantity	Configuration
SAP QAS host	1	VMware virtual machine 4 x vCPU, 16 GB Memory
Proxy host (also called backup or mount host, used by PowerSnap)	1	CPU Type: 4 quad-core 2.93 GHZ Memory: 16 GB
Storage Node server	1	CPU Type: 4 quad-core 2.93 GHZ Memory: 16 GB
VMware ESX server	1	CPU Type: 2 quad-core Intel Xeon X5570 2.93 GHZ Nehalem processor family Memory: 96 GB

Note Hyperthreading is enabled on VMware ESX server so that, logically, 16 vCPUs were available.

Software environment

Table 3 presents the software environment for the solution.

Table 3. Software environment

Software	Version
Linux Redhat Enterprise	5.2 AS5 U2 64-bit
Oracle Database	10.2.0.4 64-bit
SAP ERP	6.04/NW7.01 IDES 64-bit Unicode
EMC NetWorker	7.6 SP1
NetWorker Module for SAP with Oracle (NMSAP)	4.2
NetWorker PowerSnap Module	2.5 SP1
EMC Data Domain operating system version	4.9.26-226914
VMware vSphere™	4.1-258902
EMC CLARiiON CX Flare	04.29.0.5.014 with Clone enabler
EMC Navisphere® Agent	6.29.6(0.35)
EMC NaviSecCLI	6.29

Design and configuration

Introduction

This section provides the configuration information for this solution including:

- Key configurations or prerequisites for the validated test scenarios
- SAP storage design
- Software components configuration
 - DD Boost configuration
 - NetWorker configuration
 - NMSAP configuration
 - PowerSnap specific configuration
- Data volume generation configuration

The SAP systems presented in this solution follow the SAP traditional recommendation of a three-system landscape shown as DEV, QAS and PRD. This solution compares the performance results between the baseline scenario and the Data Domain scenario with DD Boost. The baseline scenario backs up data to SATA disks on CLARiiON, while the Data Domain scenario backs up data to the Data Domain appliance with DD Boost.

Key configurations

The following points were considered during the tests performed during solution validation:

- Backup data was sent to SATA disks on CLARiiON (presented as AFTD) in the baseline test scenario.
- Backup of SAP was configured to run through BR*Tools with BACKINT.
- DD Boost was enabled throughout the test for backup to the Data Domain system.
- Backup was optimized using multiplexing technology. The backup job was configured to run on two Data Domain devices, each with six data streams, on the SAP production system. SAP non-production systems were configured with two Data Domain devices, each with four data streams.
- Data load was configured for the SAP production system. After the initial backup, the SAP production system was then backed up five times in one percent increments before each full backup. HP LoadRunner was used for data volume generation to simulate data increase in the environment.
- The production system was configured separately to run an archive log backup every two hours. All archive logs were backed up to a dedicated pool associated with AFTD. The archive log shared more unique data segments in its content. By isolating the archive log backup, the Data Domain deduplication ratio interference was eliminated, since the archive logs' random data backup was sent to a separate location.
- The development and testing systems were also involved to validate the feasibility of inline deduplication backup on virtualized platforms. The SAP application and Oracle database were installed on Raw Device Mapping (RDM) volumes. Only one backup of the DEV and QAS systems was performed.

For more information about backup with multiplexing, refer to *EMC NetWorker Release 7.6 Service Pack 1 Administration Guide*.

SAP storage design Table 4 describes the storage design for the SAP DEV and QAS systems built on virtual machines.

Table 4. Storage design pattern for systems built on virtual machines

RAID type	Disk type	LUN size	Purpose	Filesystem
10	VMFS	90 GB	Linux root and swap	/
10	RDM	32 GB	Original redo log	/oracle/<SID>/origlogA /oracle/<SID>/origlogB
10	RDM	32 GB	Mirrored redo log	/oracle/<SID>/mirrlogA /oracle/<SID>/mirrlogB
5	RDM	64 GB	Oracle archive log	/oracle/<SID>/oraarch
5	RDM	128 GB	Database data files	/oracle/<SID>/sapdata1
5	RDM	128 GB	Database data files	/oracle/<SID>/sapdata2
5	RDM	128 GB	Database data files	/oracle/<SID>/sapdata3
5	RDM	128 GB	Database data files	/oracle/<SID>/sapdata4
5	RDM	128 GB	Database data files	/oracle/<SID>/sapdata5
5	RDM	128 GB	Database data files	/oracle/<SID>/sapdata6
10	RDM	32 GB	SAP BIN	/sapmnt/<SID> /usr/sap/<SID> /usr/sap/trans
10	RDM	32 GB	Oracle BIN	/oracle/client /oracle/stage /oracle/<SID>/102_64

Table 5 describes the storage design for the SAP PRD system built on physical machines.

Table 5. Storage design pattern for the system built on physical machines

RAID type	LUN size	Purpose	Filesystem
10	68 GB	Linux root and swap	/
10	32 GB	Original redo log	/oracle/PRD/origlogA /oracle/PRD/origlogB
10	64 GB	Mirrored redo log	/oracle/PRD/mirrlogA /oracle/PRD/mirrlogB
5	32 GB	Oracle archive log	/oracle/PRD/oraarch
5	128 GB	Database data files	/oracle/PRD/sapdata1
5	128 GB	Database data files	/oracle/PRD/sapdata2
5	128 GB	Database data files	/oracle/PRD/sapdata3
5	128 GB	Database data files	/oracle/PRD/sapdata4
5	128 GB	Database data files	/oracle/PRD/sapdata5
5	128 GB	Database data files	/oracle/PRD/sapdata6
10	32 GB	SAP BIN	/sapmnt/PRD /usr/sap/PRD /usr/sap/trans
10	32 GB	Oracle BIN	/oracle/client /oracle/stage /oracle/PRD/102_64

DD Boost configuration

DD Boost increases backup performance and reduces the load on the backup server. DD Boost also reduces the LAN bandwidth requirements by transmitting only unique data segments from the backup server or application client to the Data Domain system. Perform the following steps to configure DD Boost on the Data Domain system:

1. Check the DD Boost license.
2. Create the backup user and set the DD-Boost specific username.
3. Ensure that the Data Domain file systems and NFS services are enabled.
4. Ensure that DD Boost is enabled.
5. Create the storage unit.

For detailed configuration information, see *DD OS 4.9 Administration Guide* and *EMC NetWorker Data Domain Deduplication Devices 7.6 Service Pack 1 Integration Guide*. Refer to the *NMSAP4.2 configuration* section for the detailed setting on NMSAP.

Complete the NetWorker general configuration tasks by following the steps in the *NetWorker 7.6 SP1 Administration Guide*. General tasks include the following:

1. Create the Label Template.
2. Create the Media Pool.
3. Create the Device.
4. Create the Group and Client resources.

The SAP client resource requires the special settings for the **Save set** name and the **Backup command** as shown in Figure 2 and Figure 3, as part of NMSAP configuration.

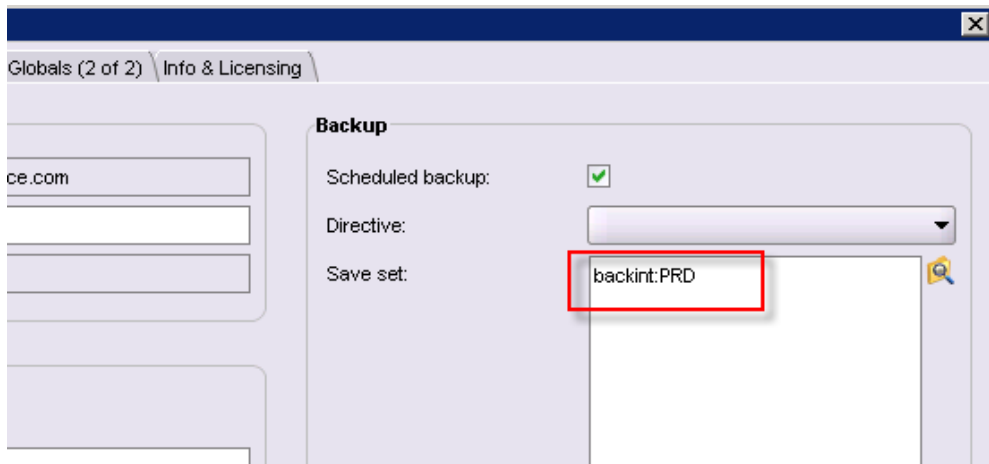


Figure 2. Set the Save set on the SAP client

The **Save set** name formatted as **backint:<sid>** stands for the backup type handled by NMSAP.

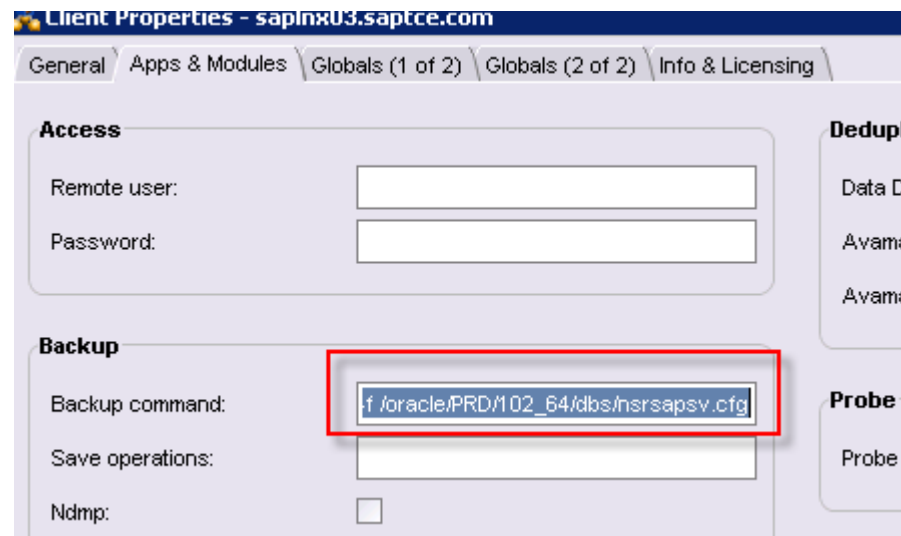


Figure 3. Set the Backup command on the SAP client

The full **Backup command** shown here is:

```
nsrsapsv -f /oracle/PRD/102_64/dbs/nsrsapsv.cfg
```

The command specifies which configuration file NMSAP should call in the event of the backup starting. When you specify the backup command, a manual configuration process is applied to achieve flexible control over the backup of SAP systems. In addition to the manual backup command, you can also use the NetWorker wizard to easily configure backup parameters for SAP systems.

For more information about the backup command profile, refer to *EMC NetWorker Module for SAP with Oracle 4.2 Administration Guide*. For backup with Data Domain, the DD device must be created and bound to the relevant pool associated with Net Worker group resources.

NMSAP configuration

The solution is validated with the following NMSAP configuration specified for SAP systems. Refer to the non-production system configuration in Table 6, and production configuration in Table 7.

Table 6. DEV or QAS– initDEV.sap or initQAS.sap

Profile characteristic	Value
Database size	497 GB
Used space in the database	415 GB
Installed agent	NMSAP
BR*Tools - Backup device type	util_file_online
BR*Tools - Backup type	online_cons
BR*Tools - Backup mode	all_data

Note Database size is the total size of all Oracle data files, excluding the temporary tablespace, while used space in the database refers to the space occupied by actual data segments in the Oracle database. The data volume increase is calculated by comparing the used space in the database with the occupied space.

Table 7. PRD – initPRD.sap

Profile characteristic	Value
Database size (initial)	502 GB
Used space (initial)	437 GB
Installed agent	NMSAP, PowerSnap (valid in PowerSnap enabled backup)
BR*Tools - Backup device type	util_file_online
BR*Tools - Backup type	online_cons
BR*Tools - Backup mode	all_data

The BR*Tools parameters are explained below.

The profile value is described as follows:

- **util_file_online**
Use external backup programs for file-by-file backup. Use this parameter value for an online backup, if it is supported by the external backup program. This dynamically sets and ends the backup status of the tablespaces to be saved and greatly reduces the volume of the offline redo log files during the backup.
- **online_cons**
The backup is performed when the database is open. Both the database files and the offline redo log files generated during the backup are copied to the same volume. As a result, you have a logically consistent dataset available.
- **all_data**
Back up the files of all tablespaces except pure index tablespaces or empty tablespaces.

The **util_file_online** must work along with the backup utility profile, as demonstrated in the Additional information section.

For more information about BR*Tools configuration, refer to the *SAP online help*.

DD Boost specific configuration

The parameters in the backup utility profile are specific to the DD Boost functionality as shown in Table 8.

Table 8. PRD – initPRDpsdd.utl

Profile characteristic	Value
NSR_DIRECT_ACCESS	no
NSR_DEVICE_INTERFACE	DATA_DOMAIN

Note The parameter NSR_DIRECT_ACCESS determines the location where duplicate data is computed and the data transmission path.

- The **Yes** value means duplicate data computing occurs on the SAP server and backup data is transmitted directly to the Data Domain appliance.
- The **No** value means duplicate data computing occurs on the NetWorker storage node and backup data is transmitted to the Data Domain appliance through the NetWorker storage node.

The direct access mode is activated on the application client only. In the backup scenario for SAP production systems with PowerSnap enabled, the parameter value must be set to **No**. This forces parts of the deduplication process to take place on the NetWorker storage node installed on the proxy host. In the meantime, the production data volume is first mounted on the proxy host. After the data is read and deduplicated, it flows from the proxy host to Data Domain appliance.

For non-production systems, direct access is activated. This causes data to be deduplicated on the application client directly without storage node engagement. The deduplicated data is then directly transmitted from the SAP server to Data Domain system.

For more information about the backup utility profile, refer to *EMC NetWorker Module for SAP with Oracle 4.2 Administration Guide*.

The clone-based backup for the SAP production system was validated using the following configuration parameters.

Table 9 lists the configuration characteristics of the proxy host using PowerSnap.

Table 9. PowerSnap configuration on Proxy host

Profile characteristic	Value
Installed agent	PowerSnap
Synchronized volumes	Database data files
Snapshot type	Clone
PowerSnap backup type	Immediate live backup (serverless snapshot backup)

Note The database control file, the BR*Tools profile, and log files are always backed up directly from the production host.

Immediate live backup refers to a process where clone volumes are split and immediately backed up by the PowerSnap module to the NetWorker backup device. The clone volumes are automatically resynchronized with the source volumes after backup.

For more information about the backup and restore definition enabled by NMSAP with PowerSnap, refer to the *EMC NetWorker Module for SAP with Oracle 4.2 Administration Guide*.

HP LoadRunner provides a way to generate data in a customer-simulated environment. The following workload scenarios are used in this case:

- Procure To Pay (P2P): this process goes through a series of steps to create a purchase requisition for the material, a purchase order with reference to the purchase requisition, changes to conditions, a goods receipt, and an invoice for the purchase order.
- Order to Cash transaction (OTC): this process covers a sell-from-stock scenario, which includes the creation of a customer order with six line items and the corresponding delivery with subsequent goods movement and invoicing.

Tasks defined to run in LoadRunner have the following attributes:

- 50 users for P2P
- 100 users for OTC
- Duration: 2 hours and 20 minutes

The target is to generate one percent of new data in the total used space of the database.

Testing and validation

Test overview

SAP backup is an important yet challenging task for system administrators. The objective of this solution is to:

- Validate the SAP backup optimization by using a Data Domain system and NetWorker along with the additional software options in terms of backup storage consumption, backup window, network utilization, and restore performance.

A typical backup strategy recommended by SAP may include:

- Backup cycle of 28 days
- Full online backup every working day in the cycle
- Full offline backup at least once in the cycle
- Keeping the last verified full offline backup of each cycle safe for long-term usage

To fulfill these requirements, SAP customers need to continue to grow the backup storage or the tape volume to ensure that the backup space is large enough to maintain the constantly growing data. It is also difficult to manage the growing volume of backup media using a tape library.

In traditional tape-based backup environments, various applications share the same network, which causes network bottlenecks during busy hours and consequently affects the backup performance. The network bottlenecks and impact on backup performance are not acceptable for 7x24 systems such as SAP. In addition, SAP customers can have other pain points when running backup jobs outlined as below:

- Triple backup space requirement for SAP DEV, QAS, and PRD
- CPU and disk contention on the SAP server caused by backup job
- Difficulties in monitoring and reporting different backup tasks

To address all of these problems, EMC offers the Data Domain system, NetWorker, and DPA backup solution for SAP customers. The purpose of this solution is to:

- Validate the benefits of deduplication for SAP backup, particularly with regard to completing backup across different SAP systems (DEV, QAS, and PRD)
- Validate the improved SAP backup performance in terms of backup duration and network utilization by using DD Boost with NMSAP
- Validate the functionality of clone-based backup as it relates to eliminating the performance impact on the backup client
- Validate the features of backup monitoring and reporting tools with EMC Data Protection Advisor
- Validate the application integration that NMSAP provides to the SAP administrators

The following key performance indicators are recorded and compared between the baseline test and the Data Domain system test for validation purposes.

- Deduplication ratio (logical backup size/real backup size)

- Time required to complete backup and recovery jobs
- LAN utilization during backup
- CPU utilization on the host or hosts. For the production scenarios, this includes the CPU utilization of the PRD host and the proxy host.

Two types of tests were performed as part of solution validation.

- Baseline test
- Data Domain system test

The general steps in each scenario include:

- Initial backup for PRD, DEV, and QAS
- Five subsequent backups of PRD with a one percent change between backups
- Restore data for PRD, DEV, and QAS

Notes

- In the baseline test, advanced file type devices (AFTD) were created on SATA disks presented by CLARiiON, acting as the backup destination.
- All backup operations involve an online full backup for the Oracle database, on which the SAP system runs. A daily full backup of the SAP system is required to protect data from accidental loss, according to typical data protection policies.
- In both baseline and Data Domain system test scenarios, the SAP PRD system included an initial backup and subsequent backups with delta load. HP LoadRunner was used to load data in the SAP PRD system five times between backups, to simulate changes created by business activities that took place in the SAP PRD database.

Test procedure

The test procedures below were performed to validate the results for the baseline test and the Data Domain system tests respectively.

Baseline test

The testing team performed the following steps to complete the baseline test of the SAP production system:

1. Performed the initial backup for PRD.
2. Performed one-time backup for DEV and QAS.
3. Ran LoadRunner to load new data on PRD, approximately one percent of the database capacity.
4. Performed backup again for PRD.
5. Repeated steps 3 and 4 four times to generate five sets of data in total along with the individual SAP PRD database backup set.
6. Removed the PRD database and restored from the latest PRD backup.
7. Removed the DEV and QAS databases and restored from their backups.

Data Domain system test

The testing team performed the following steps to complete the backup and recovery in the Data Domain scenario:

1. Performed the initial backup for PRD with PowerSnap.
2. Performed one-time backup for DEV and QAS.
3. Retrieved the PRD database by restoring from the relevant baseline backup.
4. Performed another backup for PRD with PowerSnap.
5. Repeated steps 3 and 4 until all databases backed up in baseline test were restored and backed up to the Data Domain system.
6. Removed the PRD database and restored from the latest backup
7. Removed the DEV and QAS databases and restored from their backups.

Notes:

- Steps 3-5 were intended to get the exact same copy of the database loaded with LoadRunner in the baseline scenario. This ensured consistency of the data load process and results throughout the test scenarios. This process was also used to minimize the elapsed time required to generate the 1 percent changes in the database using HP LoadRunner.
- Backup of DEV and QAS was performed right after the initial backup of PRD. This was to observe the effect of cross-client deduplication by aligning SAP non-production systems with the SAP production system in the initial backup stage.

Test results

Initial backup results

All performance charts and data presented below were captured by DPA.

Figure 4 and Figure 5 indicate the CPU utilization and network interface utilization of the SAP production system during the initial backup in baseline test.

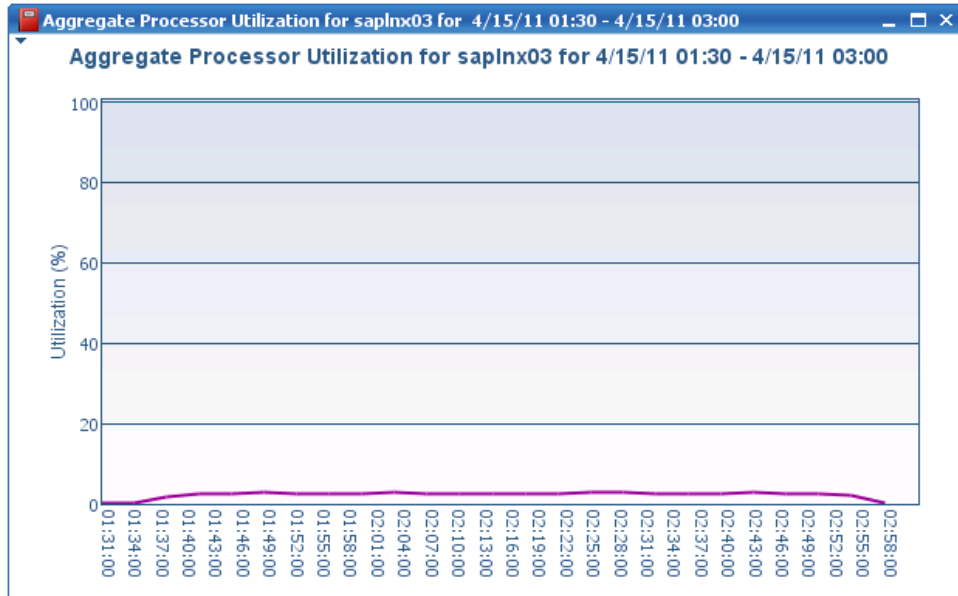


Figure 4. Baseline test: Low CPU utilization of SAP PRD system in initial backup

The CPU utilization of PRD stays at low level while the network utilization keeps at very high level during backup in the baseline test.

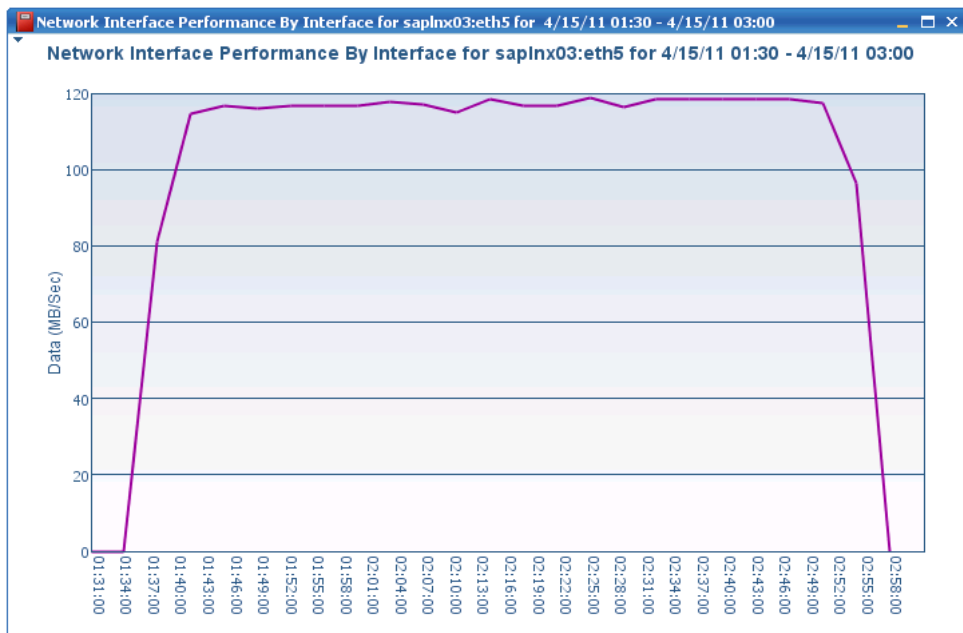


Figure 5. Baseline test: High network utilization of SAP PRD system in initial backup

Figure 6 and Figure 7 indicate the CPU utilization of the SAP production system and the proxy host during the initial backup in Data Domain system test.

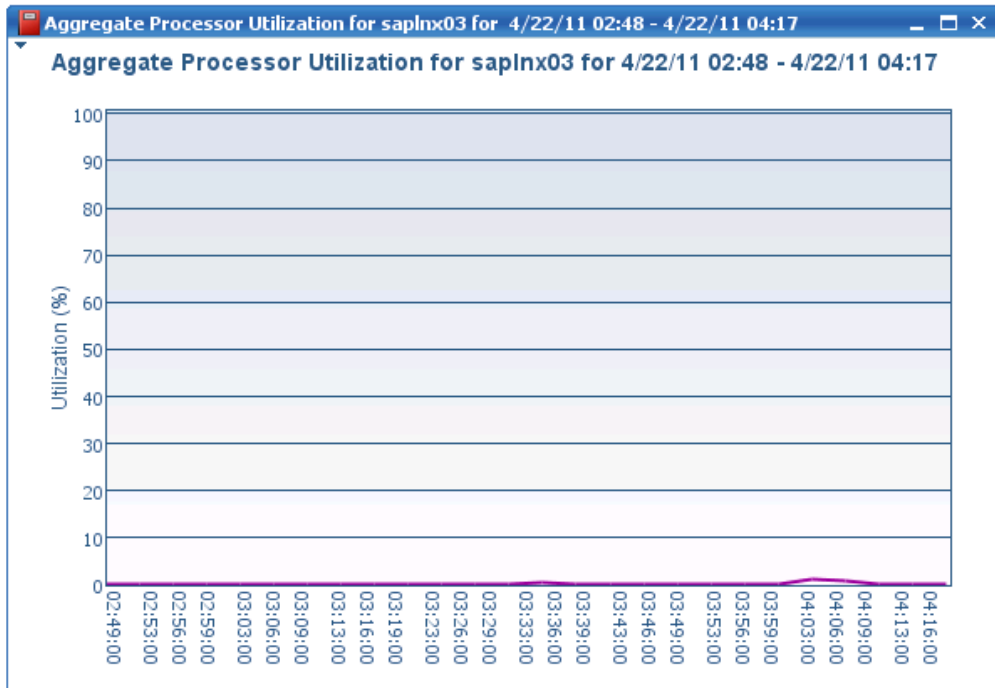


Figure 6. Data Domain test: CPU utilization of SAP production system during the initial backup

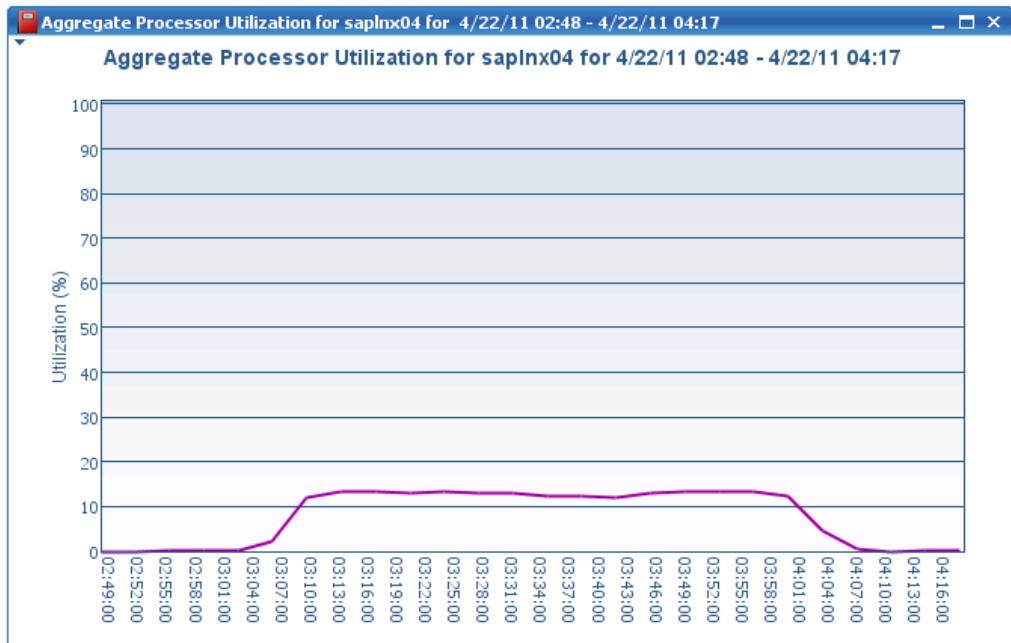


Figure 7. Data Domain system test: CPU utilization of the proxy host of SAP production system during the initial backup

From the above two figures, there is almost no CPU load on SAP production system while more than ten percent CPU load takes place on the proxy host. By backing up data from the proxy host, where the replica of SAP data volumes is mounted, CPU impact caused by the backup is eliminated from the SAP production server.

Figure 8 indicates the network interface utilization captured from the proxy host of SAP production system during the same period.

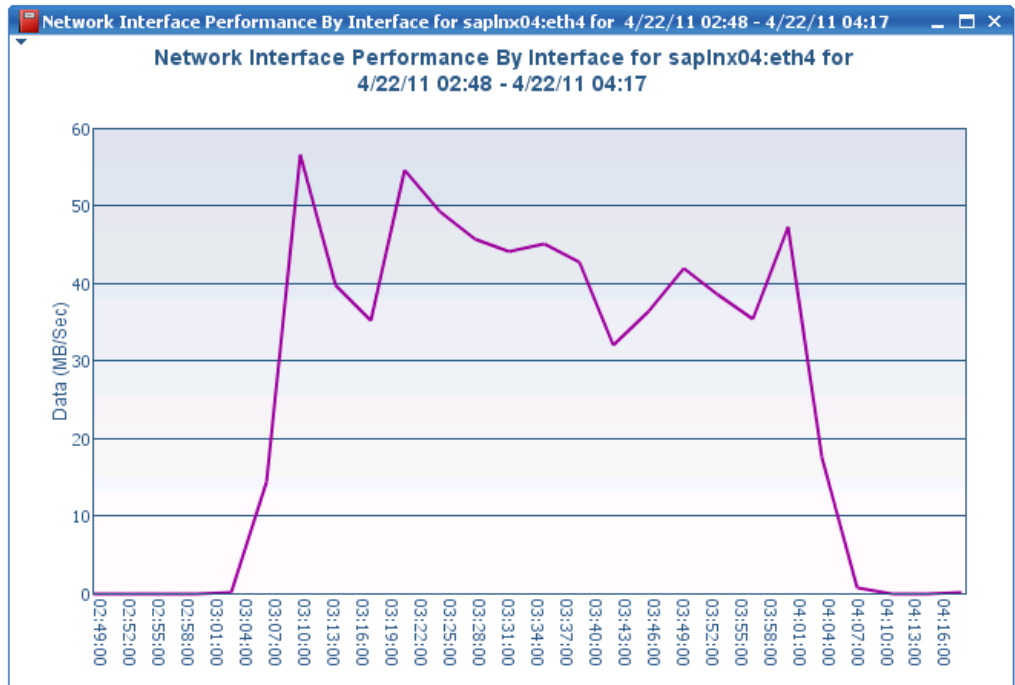


Figure 8. Data Domain system test: Network interface utilization of the proxy host of SAP production system during the initial backup

As shown in Figure 8, the network utilization of the proxy host is also reduced by backing up to a Data Domain system with DD Boost.

More performance metrics are indicated from Table 10 to Table 13.

Table 10. Production performance statistics of initial backup in baseline test

SAP System ID (SID)	CPU utilization (%)	Network throughput (MB/s)	Database size (GB)	Used space in DB (GB)	Backup size (GB)	Duration (min)
PRD	2.86	113.55	502	437	502	79.07

Table 11. Non-production performance statistics of initial backup in baseline test

SAP System ID (SID)	CPU utilization (%)	Network throughput (MB/s)	Database size (GB)	Used space in DB (GB)	Backup size (GB)	Duration (min)
---------------------	---------------------	---------------------------	--------------------	-----------------------	------------------	----------------

DEV	13.74	109.68	497	415	497	79
QAS	13.24	108.65	497	415	497	79.5

Notes

- The database size and used space in the database are interpreted in the note in Table 6.
- The backup size is the total size of the content physically written to the backup device.

Table 12. Production performance statistics of initial backup in Data Domain test

SAP System ID (SID)	CPU utilization (%)		Network throughput (MB/s)	Database size (GB)	Used space in DB (GB)	Data Domain Backup size (GB)	Duration (min)
	SAP host	Proxy host					
PRD	0.14	11	35.47	502	437	133.4	68.6

Table 13. Non-production performance statistics of initial backup in Data Domain test

SAP System ID (SID)	CPU utilization (%)	Network throughput (MB/s)	Database size (GB)	Used space in DB (GB)	Data Domain Backup size (GB)	Duration (min)
DEV	21.3	11.7	497	415	29	45.2
QAS	21.25	1.56	497	415	3.1	41.22

Notes

- DD Backup size means the data physically stored on the Data Domain system.
- The CPU utilization for SAP non-production system is captured by vSphere client. VMware recommends using vSphere client as the CPU performance monitoring tool since it is more accurate than in-guest tools, such as *top* and *mpstat*. For more information, refer to *Performance Troubleshooting for VMware vSphere™ 4.1 white paper*.

As shown from Table 10 to Table 13, the backup size of all SAP systems combined in the baseline is around 1.5 TB while the real space used in the Data Domain system to accommodate that data is only 167 GB, which reflects a 9:1 deduplication ratio. More importantly, by backing up PRD first, much less space is required for the Data Domain system to back up data from DEV and QAS, indicating global deduplication support by Data Domain system.

Other facts noted in the above tables include the optimized network utilization and reduced backup window. The network utilization decreases significantly from 108-113 MB/s to 1.6-35 MB/s, almost 85 percent less bandwidth requirement. The backup time was reduced by 35 percent from 80 minutes to 52 minutes on average.

With the deployment of the NetWorker PowerSnap module, the additional CPU utilization generated by DD Boost is also offloaded from the SAP Production system, and still the backup window is reduced compared to the baseline.

Subsequent backup result – PRD only

Table 14 and Table 15 compare the results of subsequent backups for SAP production system between the baseline test and the Data Domain test.

Table 14. Performance statistics of subsequent backups in baseline test

Backup sequence	CPU utilization (%)	Network throughput (MB/s)	DB size (GB)	Used space in DB (GB)	Backup size (GB)	Duration (min)
1	2.76	109.65	502	442	502	79.22
2	2.86	112.19	502	445	502	78.75
3	2.45	110.74	502	449	502	79.48
4	2.86	109.98	506	455	506	80.1
5	2.95	108.10	512	460	512	82.05

Note The backup size was identical for the first four backups and started to increase after that. It is because there was still free space available in the database at the earlier stages of the data load. With data loading into the database over time, some table spaces became full and needed to be expanded, which explains the change of database as shown in Table 14.

Table 15. Performance statistics of subsequent backups in Data Domain test

Backup Sequence	CPU utilization (%)		Network throughput (MB/s)	DB size (GB)	Data volume increase (GB)	Data Domain Backup size (GB)	Duration (min)
	SAP host	Proxy host					
1	0.05	15.47	2.2	502	5	5	45.5
2	0.1	15.22	1.97	502	4	3.2	45.83
3	0.12	14.24	2.47	502	4	3.9	47.92
4	0.08	15.11	2.05	506	6	4.9	46.85
5	0.11	15.2	2.28	512	5	0.9	46.43

In the baseline test, the backup size perfectly matches the database size, indicating that NetWorker just stores the backup data to AFTD as it is. This approach presents a big challenge to backup space management, requiring additional storage resources, increasing maintenance costs, and limiting the retention period with the same backup storage.

In the Data Domain system test, backup data physically written to the Data Domain system (as shown in the Data Domain Backup size column) is much smaller than the logical size of backup content (as shown in the DB size column). Only unique data is transferred across the network during backup using Data Domain with DD Boost, which is a great improvement in terms of LAN bandwidth requirements, reducing the storage requirements to preserve the backed up information, increasing dramatically the length of the retention period, raising the governance and probability of a faster recovery when compared to traditional methods.

The sustained low network utilization and optimal backup window is achieved in the course of subsequent backups. The network utilization decreases significantly from 110 MB/s to 2.2 MB/s, almost 99 percent less. The backup time is reduced by 43 percent from 80 minutes to 46 minutes on average.

Restore result

The restore time captured in Table 16 is used to compare the restore time between the baseline test and the Data Domain system scenario.

Table 16. Restore result of baseline test and Data Domain test

SID	Restore size (GB)	Duration when restoring from AFTD (min)	Duration when restoring from DD(min)
PRD	512	113.5	81.67
DEV	497	92	85.5
QAS	497	89	80.5

Comparing the results between the baseline and Data Domain scenario, restore time is reduced by 16 percent from 98 minutes to 82 minutes, proving that Data Domain optimizes not only for the backup task but also for the restore operation.

Results analysis

The following sections consolidate the test results and provide in-depth analysis across the following test scenarios:

- Backup storage footprint
- Backup performance
- Restore performance
- Performance impact during backup

Backup storage footprint

Figure 9 shows the graphical comparison of the combined storage requirements of a one-time backup of all three SAP systems, PRD, QAS, and DEV during the baseline tests and the Data Domain tests.

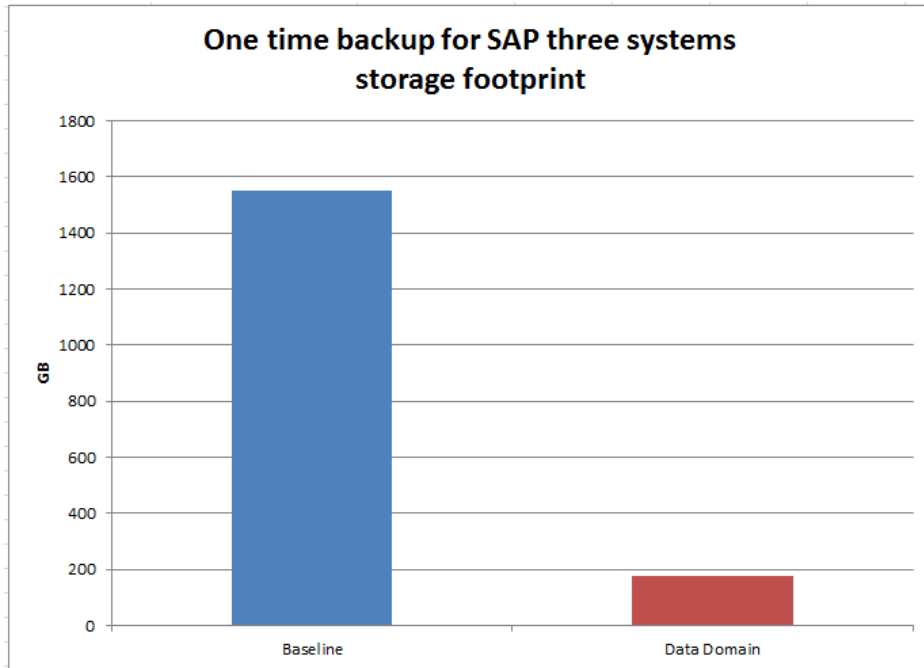


Figure 9. One-time backup for SAP three systems storage footprint

As shown in the chart, the footprint requirement of the three systems triples the storage requirement over a single system at around 1.5 TB. Backing up to the Data Domain system saved 1.3 TB of backup storage space only requiring 167 GB, due to deduplication and compression to protect the same 1.5 TB data set, providing a 9:1 deduplication ratio.

The result obtained in our lab makes clear the potential benefit of the inline deduplication provided by the Data Domain systems regarding storage requirements.

Figure 10 shows the comparison between the storage requirements over time between the CLARiiON (baseline) and the Data Domain system.

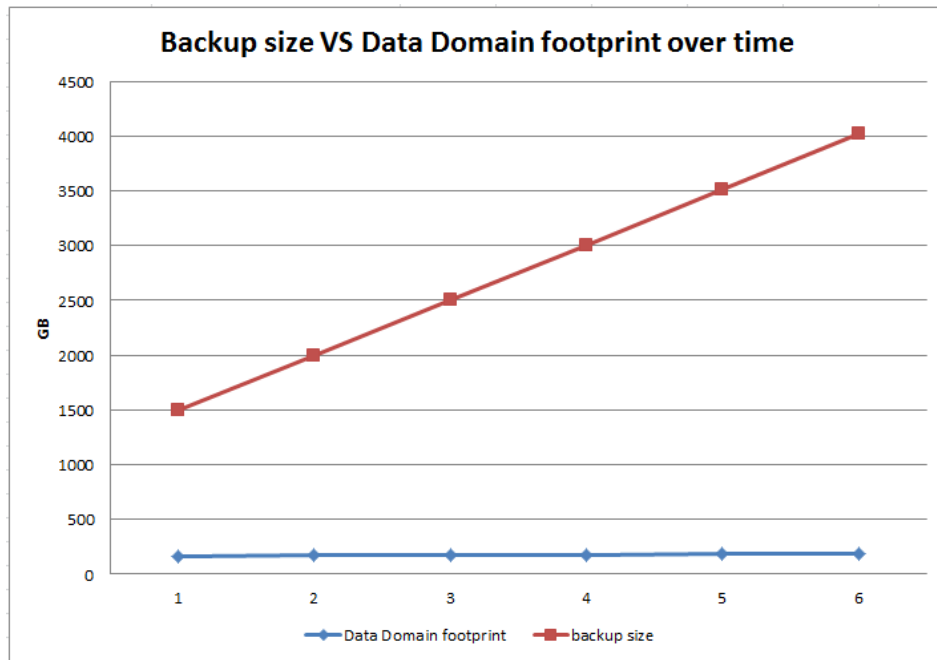


Figure 10. Backup size versus the Data Domain system footprint over time

The backup storage requirements over time (5 days) increased exponentially using the traditional (baseline AFTD) backup to the CLARiiON, while the storage requirements on the Data Domain system were almost flat, enabling the customers to save more data with less space. Over time, this improves the deduplication ratio ... from 9X to 21X on the validated scenario.

Note Results may vary, since the deduplication ratios, are dependent on the frequency and volume of changes and on the similarity of the data between the SAP systems backed up to the same Data Domain system.

Backup performance

Figure 11 shows a graphical comparison between the baseline tests and the Data Domain tests regarding the Network bandwidth utilization.

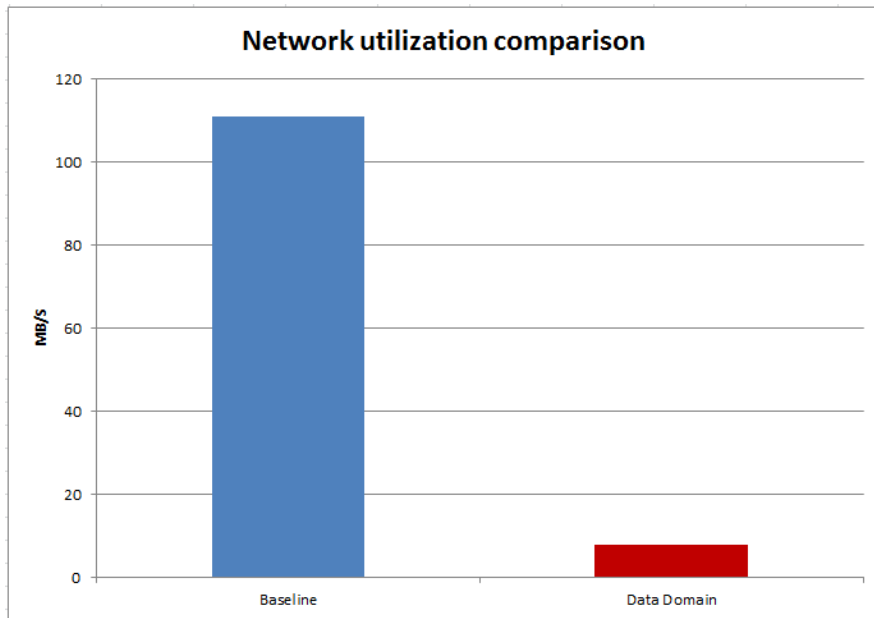


Figure 11. Network utilization comparison

The consumption of the network bandwidth was reduced by 93 percent with the introduction of DD Boost, comparing the baseline and Data Domain tests performed on the SAP Production system.

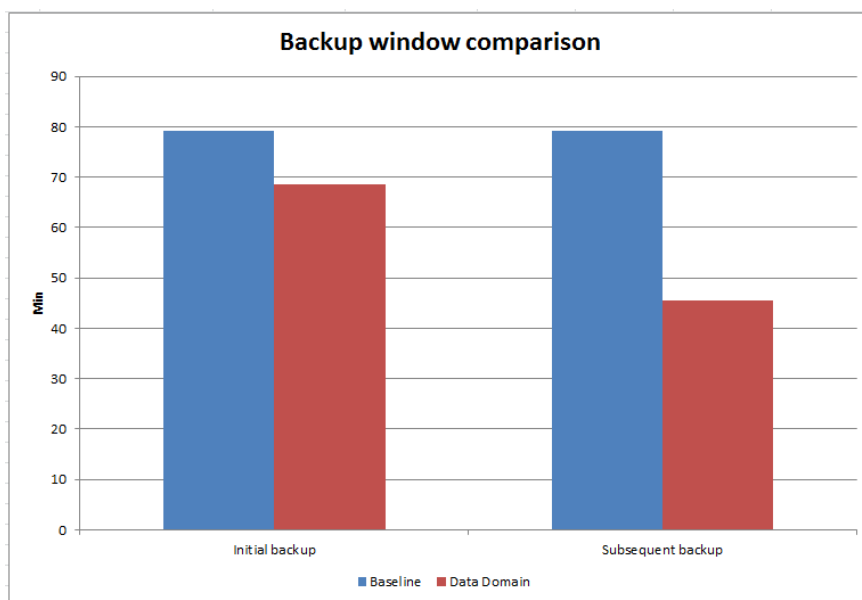


Figure 12. Backup window comparison

Figure 12 indicates a reduction in the backup window between the initial backup and subsequent backup for the SAP production system. In the initial backup scenario, backup efficiency improves by 14 percent. In subsequent backup scenarios, the backup process is accelerated by 43 percent, 1.75X faster. **Restore performance**

Restore performance is important for SAP customers because the mission-critical system needs minimum recovery time in case of failure. The Data Domain system improves the restore performance by allowing restore operations 1.38X faster than the one operated from AFTD.

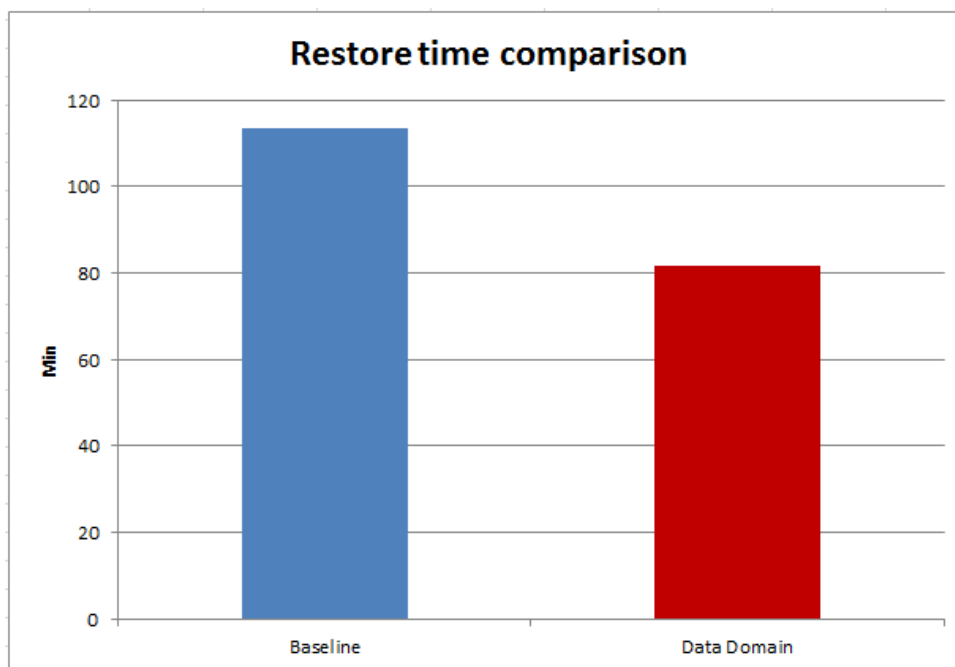


Figure 13. Restore time comparison

Performance impact during backup

Networker PowerSnap module was used to clone the SAP PRD system to the proxy host. The CPU and network bandwidth utilization on the SAP production system was offloaded to the Proxy host, eliminating any backup load from the SAP Production system, while safeguarding its performance and stability. The dedicated backup resources in the proxy host maximized the backup efficiency minimizing the overall backup duration.

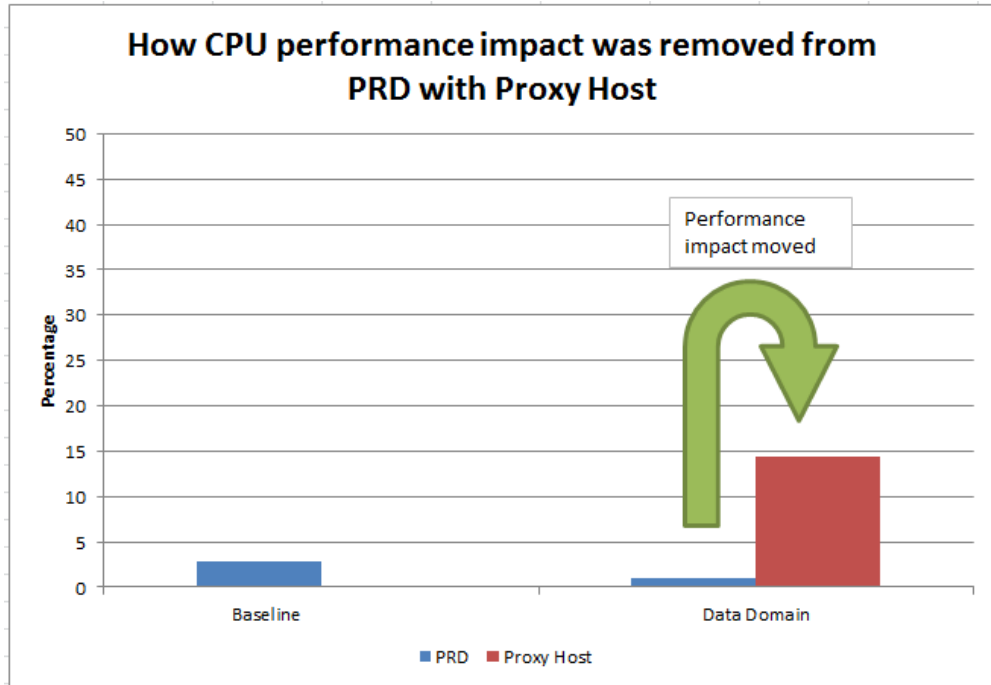


Figure 14. Comparison of CPU utilization of PRD and Proxy host between baseline and the Data Domain system tests

Conclusion

Summary

A comprehensive backup solution is required to meet today's demanding SLAs of data protection for SAP landscapes. This solution demonstrates the best-in-class backup solution for SAP systems with inline deduplication technology provided by the Data Domain system powered by Intel Xeon, NetWorker, and DPA.

SAP customers can better utilize their SAP backup hardware and optimize their backup and recovery strategy while achieving significant cost savings with Data Domain systems. With lower TCO, it is easy for SAP customers to maintain the infrastructure to better protect their SAP environments.

Data Domain systems allow SAP customers to significantly reduce backup storage costs while backing up multiple SAP systems in an SAP recommended backup cycle. Inline deduplication also enables longer onsite retention of backup data to fulfill the requirement of auditing or revisiting old data for governance. Enhanced by DD Boost, backup performance is improved significantly, making the backup process more efficient compared to a traditional tape-based backup process.

Findings

Combining Data Domain systems with NetWorker brings SAP backup and recovery to a new level by:

- Eliminating duplicate data while backing up multiple SAP systems multiple times, achieving deduplication ratios up to 20:1
- Reducing the backup window by 43 percent with DD Boost
- Reducing LAN utilization by up to 93 percent during backup with DD Boost
- Refining the SAP backup strategy with reduced restore time
- Removing performance impact on the SAP server during backup

The Data Domain system and NetWorker offerings with Data Protection Advisor increase SAP data protection levels and enhance backup and restore management tasks. The solution removes the risk involved in SAP backup and recovery operations and increases the flexibility of SAP administration.

References

White papers

For additional information, see the white papers listed below:

- *EMC Backup and Recovery for SAP Enabled by EMC NetWorker Module for SAP with EMC Avamar – A Detailed Review*
- *EMC Backup and Recovery for Microsoft Applications Deduplication Enabled by EMC CLARiiON and Data Domain – A Detailed Review*

Product documentation

For additional information, see the product documents listed below:

- *EMC NetWorker Release 7.6 Service Pack 1 Installation Guide*
- *EMC NetWorker Release 7.6 Service Pack 1 Administration Guide*
- *EMC NetWorker Module for SAP with Oracle Release 4.2 Installation Guide*
- *EMC NetWorker Module for SAP with Oracle Release 4.2 Administration Guide*
- *EMC NetWorker PowerSnap Module for CLARiiON Version 2.5 Service Pack 1 Installation and Administration Guide*
- *EMC NetWorker Data Domain Deduplication Devices 7.6 Service Pack 1 Integration Guide*
- *VMware ESX 4.1 and vCenter 4.1 Installation Guide*
- *Performance Troubleshooting for VMware vSphere™ 4.1*
- *EMC DD OS 4.9 Administration Guide*
- *EMC Data Protection Advisor 5.8 Administration Guide*
- *Master Guide for SAP ERP 6.0*

Other documentation

For additional information, see the documents listed below:

- *SAP ADM505 – Database Administration Oracle*

Additional information

Overview

This section provides the sample configuration files used in this white paper.

Backup utility profile

The backup utility profiles are listed in Table 17.

Table 17. PRD – initPRD.utl

Profile characteristic	Value
NSR_NWPATH	/usr/sbin/
parallelism	12
savesets	12
expiration	3 Months
pool	467PRD
server	networker
client	saplnx03.saptce.com
ssNameFormat	new
NSR_SAVESET_RETENTION	3 Months
NSR_SAVESET_BROWSE	3 Months

Backup utility profile for PowerSnap

The backup utility profile for PowerSnap is listed in Table 18.

Table 18. PRD – initPRDps.utl

Profile characteristic	Value
ps_backup_mode	yes
ps_archive_mode	no
ps_restore_mode	yes
ps_inquire_mode	yes
ps_exclude_backup_bi_run_nums	2
ps_exclude_archive_bi_run_nums	1;2
ps_exclude_backup_paths	/oracle/PRD/102_64/dbs/*;/oracle/PRD/sapreorg/*;/oracle/PRD/sapbackup/*
ps_opaque_pfilename	/oracle/PRD/102_64/dbs/nsrsapps.cfg

The **ps_opaque_pfilename** parameter defines the path where the PowerSnap configuration file is located.

The PowerSnap configuration profiles are listed in Table 19.

Table 19. PRD – nsrsapps.cfg

Profile characteristic	Value
NSR_CLIENT	saplnx03.saptce.com
NSR_DATA_MOVER	saplnx04.saptce.com
CLAR_SNAP_POOL	/nsr/res/clariion.res
NSR_SNAP_TYPE	emcclar
FRAME_IP	<CLARiON's IP address>
NSR_PS_SAVE_PARALLELISM	12
EMCCLAR_SNAP_SUBTYPE	mirror