EMC AVAMAR INTEGRATION WITH EMC DATA DOMAIN SYSTEMS
A Detailed Review

ABSTRACT
This white paper highlights integration features implemented in EMC Avamar® with EMC Data Domain® deduplication storage systems to allow high-speed data backup directly to Data Domain systems. The components outlined in this paper will provide IT backup administrators & consultants with a technical foundation to architecting efficient combined data deduplication backup solutions.

February, 2014
# TABLE OF CONTENTS

**EXECUTIVE SUMMARY**  4
   AUDIENCE  4  

**INTRODUCTION**  4  

**EMC AVAMAR WITH EMC DATA DOMAIN INTEGRATION DETAILS**  5
   Architecture Overview & Data Flow  5
   Avamar/Data Domain Systems Centralized Control & Management  6

**CONCLUSION**  10
EXECUTIVE SUMMARY
Implementing EMC Avamar for efficient backups of VMware, desktops & laptops, remote offices, NAS systems, and enterprise applications such as SQL & Oracle databases, Exchange, and SharePoint is very straightforward today.

Beginning with the release of Avamar v6.0 and its ability to direct selected backup workloads to Data Domain systems, the costly need for two separately managed backup environments has disappeared. Now it is commonplace for customers leveraging EMC Avamar, through the use of Avamar interfaces and workflows, to quickly and easily direct enterprise-wide backups to either EMC deduplication platforms while simplifying the backup management infrastructure. Essentially, the integration offers the benefits of Data Domain’s scale and performance, and the simplicity and efficiency of Avamar’s business critical application integration and robust backup and recovery policy management features.

This paper highlights in detail the specific features and architecture surrounding this integration, as well as how the integration of Data Domain systems within the Avamar management framework can enhance the management and optimization of backup operation without additional management overhead.

AUDIENCE
This white paper is intended primarily for backup systems administrators who are responsible for architecting, deploying, and protecting data across the enterprise. Knowledge of basic data deduplication concepts are assumed, and implied throughout this paper. Also, while technical knowledge of current Avamar and Data Domain systems architecture is helpful, it is not required to fully understand the scope of this white paper.

INTRODUCTION
Today, two challenges that face IT backup administrators are:

- Deploying two separate solutions in separate silos to optimize for varying workloads.
- Minimizing administrative costs of managing both solutions separately.

As shown here in Figure 1, a combined backup solution based on EMC Avamar and EMC Data Domain can eliminate the need for multiple backup solutions and costly data silos, which can prohibit capacity savings efficiency. With the features and implementation...
details highlighted in the sections below, one can clearly see how the integration of Avamar with Data Domain systems will address these challenges.

**EMC AVAMAR WITH EMC DATA DOMAIN INTEGRATION DETAILS**

**ARCHITECTURE OVERVIEW & DATA FLOW**

![Diagram of EMC Avamar with EMC Data Domain Systems Integration Architecture](image)

Backup
The diagram shown above in Figure 1 depicts a high level architecture of the combined Avamar and Data Domain solution. At a high level, starting with Avamar v6.0, one can specify whether specific datasets in an Avamar backup policy will target an Avamar Server or EMC Data Domain system.

When specifying an Avamar Data Store as the destination, the Avamar client installed on each host will work as a typical Avamar client performing client side deduplication segment processing and sending those capacity optimized blocks over the WAN/LAN to the Avamar Server including any required metadata. This dataflow does not change from the well-known implementation of previous versions of Avamar.

When specifying a Data Domain system as the backup target of a particular dataset from within the Avamar Administrator interface, the same Avamar client leverages EMC Data Domain Boost™ (DD Boost) software to redirect this data directly to a Data Domain system.

DD Boost extends the optimization capabilities of Data Domain systems by significantly increasing performance by distributing parts of the deduplication process to the client, also simplifying disaster recovery procedures. Enhancements have been made as of the Avamar v6.1 release to further increase performance by leveraging multiple DD Boost streams. This enhancement can introduce an up to 4X increase in backup performance based on performance testing.

DD Boost will transfer the source data, in an efficient transfer method, for processing by Data Domain systems, in place of performing intensive deduplication processing only on the client.

In addition to the backup payload data being transferred optimally to the Data Domain system, relevant metadata generated by the Avamar client software is simultaneously sent to the Avamar Server for storage. This metadata allows the Avamar management system to direct the client in performing restore operations directly from Data Domain systems without first going through the Avamar Server. This functionality, implemented within the Avamar client software, is the same DD Boost API and technology under the covers that are supported on many other Data Domain system implementations today.
Recovery
The process of data recovery, whether from an Avamar Data Store, or from Data Domain systems is completely transparent to the backup administrator. The backup administrator will utilize the same Avamar recovery processes that are native to current Avamar implementations. The Avamar front end will automatically retrieve datasets, which are stored on Data Domain systems. No special retrieval mechanisms or processes are required.

For VMware image backup workloads, Avamar’s integration with Data Domain systems provide for more flexible options in recovering your virtual machines into production. As you may already know, recovering a VM image is rather painless and straightforward leveraging Avamar’s integration with VMware’s vStorage APIs for Data Protection and more specifically Changed-Block Tracking on the restore. This enables customers to restore partially corrupt VMs in minutes instead of the hours typically required.

Introduced in the Avamar 7.0 release, when a VM Image backup is performed to Data Domain systems, the administrator now has the option to leverage Avamar’s Instant VM access feature. With this new capability, the ability to quickly mount and run the VM from the backup image stored on Data Domain systems allows for near-instantaneous recovery time from backup.

Here is how the Avamar and Data Domain systems interact with each other during this process:

1. The backup administrator requests that a particular VM be accessed from backup leveraging Avamar’s VM Instant Access feature.
2. A virtual machine image backup is staged to a temporary location (logical volume) on a Data Domain system.
3. This temporary location is exported as a secure NFS share.
4. The share is mounted as a NFS datastore within a vCenter environment.

At this point the VM is powered on, and is available to service production workloads immediately. Clearly running the VM from backup tier storage is not ideal for long periods of time, however availability is immediate. The final step is to begin the process of moving this VM from the backup tier of storage to Production level VMFS datastores. VMware provides a mechanism for this called storage vMotion. By leveraging this VMware feature we can now copy the VM while it is online from backup storage over to production, restoring workload performance to 100% with the smallest amount of down time.

Replication
Referring back to Figure 1, replication between primary and replica Data Domain systems is also integrated into the Avamar management feature set. This will be controlled in the user interface via the Avamar replication policies applied to each dataset. All typical Avamar replication scenarios will be supported for datasets targeted to Data Domain systems, including:

• Many-to-one, one-to-many, cascading replication
• Extension of data retention times
• Root-to-root

Similar to recovery, the replication process normally configured within the Data Domain systems platform is automated within the Avamar framework and for all practical purposes is transparent to the backup administrator. This replication functionality requires that a remote Avamar server exists, along with remote Data Domain systems.

AVAMAR/DATA DOMAIN SYSTEMS CENTRALIZED CONTROL & MANAGEMENT
Let’s take a look at how some of the processes highlighted in the previous section are controlled from an end user perspective within the Avamar Administrator user interface.

In the example below, we will go through the steps to properly configure Data Domain systems for use with SQL Database Backups. As a starting point, we will begin by logging into the Avamar Administrator.

The first step we will need to perform is provisioning either a newly implemented or already deployed Data Domain system within the Avamar environment so we can then specify it as a backup target for our backup policy datasets.
As you can see in Figure 4, we will click on the “Server” icon in the main menu of the Avamar Administrator to access the Avamar backend system configuration settings.

![Figure 4: Avamar Administrator Main Menu](image)

As shown in figure 5, when accessing the “Server Management” tab within the Avamar Administrator we now have the ability to provision Data Domain systems for use by the Avamar client by simply clicking on the icon in the toolbar to display the Data Domain systems configuration dialog box. To configure Data Domain systems within an Avamar environment there is no need to launch the Data Domain System Manager console UI or CLI interface.

![Figure 5: Specify target Data Domain Systems from within Avamar.](image)
The “Add Data Domain System” configuration dialog box appears as shown below in Figure 6. Starting with the “System” tab, we will specify the system information such as system name and IP address. Additionally, we will need to specify for the Avamar environment the credentials used to access the Data Domain systems and ultimately authenticate the Avamar clients to write backups to the Data Domain backup target. For authentication, we will leverage the DD Boost User account which, unlike the rest of this configuration procedure, is created through the Data Domain System Manager interface. This account is specified during the setup of the Data Domain system via the Data Domain System Manager. For those systems which are already implemented and have the DD Boost license installed, this authentication account should already be configured on the Data Domain systems. For newly implemented Data Domain systems with a DD Boost license, please reference Data Domain systems documentation for the valid procedure to configure this account.

Another important setting we can specify here is “Max used by Avamar”. This setting is used to specify the total number of client backup streams launched simultaneously from Avamar clients managed by this particular Avamar server to Data Domain systems.

This setting is especially important in implementations where Data Domain systems are targeted by backup applications besides Avamar, such as EMC NetWorker. This way we can utilize this setting to assure that the Avamar backup streams do not oversubscribe Data Domain systems if other backup jobs are being targeted to it within the same backup window.

Now that we have specified which Data Domain systems we would like to target for backup, by clicking on the SNMP tab as shown in Figure 4, we will now specify to the Avamar environment more detailed system configuration information involving SNMP to properly enable Avamar to collect, and display data for health monitoring, system alerts and capacity reporting.

As shown in Figure 7, we will specify the “Getter/Setter Port Number” which tells the Avamar server the port on Data Domain systems from which to receive and on which to set SNMP objects. It is usually not needed to change this value from the default value of 161.

Next, we will specify the “SNMP Community String” which specifies to Avamar the community string to utilize which will allow read/write access to Data Domain systems from an SNMP perspective. And finally, we will specify the “Trap Port Number” which specifies which port on the Avamar server that Data Domain systems will utilize to send SNMP traps to the Avamar Server. Again, the default here usually does not need to be changed, refer to your Avamar Server SNMP configuration for the proper settings. Once we have verified all of our settings, we can now click “OK” to complete the provisioning of Data Domain systems for use by Avamar.
As you can see, the process of provisioning Data Domain systems for use with Avamar is done easily and intuitively. Now that we have added the system, we can now return to the “Server Management” tab at any time within the Avamar Administrator UI for quick dashboard view of Data Domain systems statistics such as capacity, DDOS version, and model number.

Our Data Domain system is now properly configured and we can now begin to direct backup policy datasets within Avamar to target this particular Data Domain system appliance. It is also commonplace to repeat this process for multiple Data Domain systems that may be required for a particular Avamar/Data Domain systems joint solution. For the purpose of this paper, we will only configure one Data Domain system.

Next we will take a look at how to specify a dataset policy within Avamar to target backups to it.
In this illustrated environment, I have an already existing policy dataset for backup of a production SQL database. In this case, it is referred to by the dataset called "DDSQLDataset". Please refer to EMC Avamar Administrator guides for creating datasets from scratch.

We will now re-direct the backups of this particular data set from the Avamar Data Store over to the Data Domain systems we just provisioned. We will go ahead and edit the particular dataset we would like to retarget via the Avamar Administrator Policy configuration as shown above in Figure 9. Clicking on the “Options” tab will bring up the configuration settings as shown below in Figure 10.

Newly added to the UI, we are now able to click a checkbox titled “Use Data Domain system” and from the corresponding drop down menu presented to the right of the checkbox, select the Data Domain systems we would like to specifically target for this data set. In this case, I have only provisioned one Data Domain system, so I only have the option of selecting one system, however if more Data Domain systems were provisioned in this environment multiple systems for selection would be shown.

Once we have enabled this checkbox and selected the desired Data Domain systems, we can click “OK” to complete the dataset configuration. In this case we are targeting an SQL dataset, which will be processed via the Avamar Windows SQL Plug-in to target its data to the Data Domain systems.

Upon the next backup policy execution (either manual or scheduled) the SQL dataset will now be sent to Data Domain systems for deduplication and storage.

**CONCLUSION**

With today's infinitely growing backup data sets and the need to reduce backup windows, managing one particular backup implementation can be a serious undertaking on its own. Managing two distinct backup systems to leverage advantages offered by two separate solutions increases the difficulty by a factor of two.

With the features detailed in this paper, EMC Backup and Recovery Systems’ solutions are again changing the landscape and further simplifying the management required in performing enterprise-wide backups with deduplication. Whether your specific environment has leveraged both or only one of the deduplication products mentioned in this paper, the process for augmenting the environment to take advantage of this integrated architecture is easily achieved.

With the integration and control of Data Domain systems utilizing DD Boost within the Avamar client, the IT backup administrator can now eliminate separate backup processes and manage enterprise-wide backup workloads associated with each dataset from a single management interface. Furthermore, backups of VMware, desktops & laptops, remote offices, NAS systems, and enterprise applications are all optimized for faster backups, maximum deduplication and efficient recoveries.