WHITE PAPER

PROTECTING MISSION CRITICAL DATA
WITH BACKUP AND REPLICATION FROM PURE STORAGE AND VEEAM
TABLE OF CONTENTS

INTRODUCTION .......................................................................................................................... 3
ARCHITECTURAL OVERVIEW ................................................................................................. 3
TEST PROCESS .......................................................................................................................... 5
VEEAM BACKUP & REPLICATION 9.5 SETUP AND BEST PRACTICES ........................................ 6
PURE STORAGE FLASHBLADE BACKUP REPOSITORY SETUP .................................................. 8
VEEAM BACKUP & REPLICATION VSphere INTEGRATION SETUP ........................................... 14
VEEAM BACKUP JOB SETUP AND RESULTS .......................................................................... 16
VEEAM VIRTUAL MACHINE RESTORE SETUP AND RESULTS .............................................. 24
CONCLUSION ......................................................................................................................... 28
INTRODUCTION
The purpose of this whitepaper is to showcase the ease of integration, best practices, and expected performance results for data backup and restore using Pure Storage® FlashArray//M, Veeam® Backup & Replication 9.5™, and Pure Storage FlashBlade™ data solutions using realistic, real-world examples. The combination of these innovative, 100% flash-based solutions provides the best performance, highest levels of resilience, easiest implementation, and, perhaps most importantly, extremely low RTO and RPO for your most critical workloads.

In this document, we’ll highlight the following three core pieces of next generation storage and backup technology and provide explanations as to how they seamlessly work together:

- **PURE STORAGE FLASHARRAY//M** Provides high performance primary block storage for the most demanding datacenter workloads, including virtual machines, SQL and Oracle® databases, and more. FlashArray provides built-in snapshots and replication to other Pure Storage arrays, always-on encryption, and 100% non-disruptive operations – including firmware and generational controller upgrades which are traditionally disruptive for legacy storage vendors.

- **VEEAM BACKUP & REPLICATION 9.5** Provides a complete software solution for performing data protection, data replication, and disaster recovery tasks such as VM backup, replication, copying backup files, and carrying out bi-directional disaster recovery procedures for FlashArray//M and FlashBlade. Veeam Backup & Replication also provides integration with a wide range of cloud service providers for off-site data archiving.

- **PURE STORAGE FLASHBLADE** A fast and dense unstructured data platform from Pure Storage that provides on-premise rapid backup and recovery of archival data. Additionally, Pure Storage FlashBlade provides high-performance in a dense-form factor for other unstructured data primary use cases such as machine learning, deep learning, and electronic design automation (EDA), among others. For more information, please follow this link.

ARCHITECTURAL OVERVIEW
The test cases shown here are intended to emphasize the high levels of throughput, workload consolidation, minimized RTO and RPO, and ease of administration that these connected solutions provide. Though we will focus on vSphere virtual machines in our examples, the use cases shown here are easily extensible to other workloads and hypervisors such as SQL databases, VSI, Hyper-V®, and Oracle, just to name a few.
The connectivity diagram below illustrates how the various solutions are integrated with one another and highlights the lack of a single point of failure within the design:

FIGURE 1: High-Level connectivity diagram of the overall solution
From the connectivity diagram, here are more detailed specifics for each hardware and software element:

- Five VMware ESXi™ 6.5 hosts. Each ESXi host features two redundant 10GB network connections and two redundant 16GB Fibre Channel HBA connections for SAN connectivity.
- 250 Microsoft Windows 10™ desktops with 50GB drives (27GB used) were using the above ESXi hosts for compute. The desktops had MS Office 2016™, Adobe Reader®, and numerous iso, pdf, mp4, and many other pre-compressed, commonly used files on the local drive.
- One Pure Storage FlashArray//M20 with 10TB RAW Storage for primary Storage (~50 TB usable assuming a 5:1 data reduction ratio, which is the approximate measured average from the Pure Storage install base) running Purity Operating Environment version 4.9.3. The array used Fibre-Channel connectivity (though iSCSI is also fully supported), and will serve as our data source for backup. The vCenter instance, the 250 desktops described above, the Veeam Proxy servers, and all other vSphere-based components were hosted on the FlashArray//M.
- Veeam Backup & Replication 9.5 installed on a physical Windows® 2012R2 server to provide backup and data orchestration between FlashArray//M and FlashBlade as well as the cloud tier. This component was running on top of a SuperMicro® server.
- One Pure Storage FlashBlade with 7 Blades (half-populated chassis) for the Backup Repository with 8TB per blade running the Elasticity Operating Environment version 2.0.4. Pure Storage FlashBlade supports both NFS and SMB protocols, though this paper will focus on the SMB 2.1 protocol.
- Two paired Brocade® VDX6740T Switches for 1/10/40GB resilient networking.
- Two Cisco® MDS 9148S 16GB Fibre-Channel switches.

**TEST PROCESS**

Our test process will cover the following three simple steps with instructions, recommended best practices, and performance metrics shown for each phase:

**BACKUP 250 VMs FROM FLASHARRAY//M TO FLASHBLADE**

**DELETE 50 VMs FROM VCENTER**

**RECOVER 50 VMS FROM FLASHBLADE TO FLASHARRAY//M VIA VEEAM BACKUP AND RECOVERY**

FIGURE 2. Test process to be executed
VEEAM BACKUP & REPLICATION 9.5 SETUP AND BEST PRACTICES

Veeam provides data backup and portability between both Pure Storage products, VMware vSphere, and Hyper-V, as well as off-premise cloud solutions. This data orchestration is shown in the following diagram:

![Veeam Backup & Replication dataflow architecture](image)

The major components of Veeam Backup & Replication consist of a physical or virtual management server, physical and/or virtual proxy servers, backup storage repositories, gateway servers or Linux® hosts (for SMB and/or NFS share mounts, respectively), and disk-based or cloud-based backup repositories for cold and/or offsite data storage. The backup proxy and management servers are Windows-based installations that provide increased throughput and resiliency for data movement between hardware appliances. The backup repositories can be Windows or Linux-based network attached storage systems such as Pure Storage FlashBlade.

The Veeam Backup & Replication management server was hosted on a physical Windows server, while the proxy servers were virtualized on the same VLAN as our target desktops to maximize backup throughput and performance to the backup repository. Conversely, Veeam supports a virtualized management server and physical proxy servers as well. However, our rationale for installing Veeam Backup & Replication 9.5 on a physical host is that running Backup and Replication separately from the target vSphere instance would enable faster recovery should the entire target vSphere instance become corrupt or require top-level recovery.

As both Pure Storage FlashArray//M and FlashBlade have hundreds of thousands of IOPS available at low latency, certain values within the Veeam Backup & Replication 9.5 console can be changed to provide higher levels of throughput, enabling backup jobs to be completed much faster than if they were left at default values.

The most important such change is editing the registry of the server running Veeam Backup & Replication to increase overall throughput and parallelism in processing VMs for backup. For our testing, we increased the MaxSnapshotsPerDatastore value from the default dword of 4 to 24.
The registry entry can be found under **HKLM → Software → Veeam → Veeam Backup & Replication → MaxSnapshotsPerDatastore**. The updated key and path to it can be seen in the screenshot below:

![Registry Editor Screenshot](image-url)

**FIGURE 4.** Registry entry modified for increased throughput during testing
PURE STORAGE FLASHBLADE BACKUP REPOSITORY SETUP

Veeam uses multiple physical and/or virtual machines for data movement from the primary storage device to the storage repository. The following diagram illustrates how data is moved for an SMB/CIFS backup repository share:

FIGURE 5. SMB/CIFS data flow diagram for Veeam

NFS is also supported and shown architecturally below, though that protocol is not within the scope of this document.

FIGURE 6. NFS dataflow diagram for Veeam
Furthermore, use **Direct Storage Access** mode when there is end-to-end Fibre Channel connectivity between the ESXi hosts, the physical Windows server running Veeam, and the source FlashArray//M when and where maximum throughput and performance are important requirements. Using Direct storage access has the added benefit of not producing any load on the production network, since backup and replication traffic is relegated to the SAN network only. More detailed information and minimum requirements for use can be found [here](#).

In the next few screenshots, we will show how FlashBlade is configured as a new backup repository:

First, a new File System on FlashBlade must be created to be used as the backup repository. The example below uses the SMB protocol, though NFS could also be used.

From the Pure Storage FlashBlade GUI, highlight the **Storage** option and then click the + sign on the right.

![FIGURE 7. FlashBlade File System creation from GUI](#)

For SMB, we recommend enabling both the NFS as well as SMB adapters on the File System for accurate space reporting within the Veeam console.

![FIGURE 8. Naming, sizing, and enabling the NFS and SMB protocols on a Pure Storage FlashBlade File System](#)
With the backup repository location created on Pure Storage FlashBlade, we then moved over to the Veeam Backup & Replication 9.5 console. Under the Backup Infrastructure tab, select Backup Repositories, then select Add Repository. Provide a name for the repository:

![Figure 9: Naming new FlashBlade backup repository in Veeam Backup & Replication]

Next, as we have enabled SMB on our repository, we select the CIFS (SMB) shared folder option:

![Figure 10: Selecting type of backup repository within Veeam console]
Enter a data IP address from Pure Storage FlashBlade and then the File System name (in this format: `<Data IP Address>\SMB Share Name`) within the **Shared Folder** field, as the following screenshot shows.

![Screenshot](image1.png)

**FIGURE 11.** Entering the CIFS/SMB data path to FlashBlade within Veeam console

With the path entered correctly, the path to the folder and the amount of overall and available space is shown. We elected not to throttle any I/O tasks under **Load Control**, though customers concerned about network I/O can certainly do so depending on their own unique environment.

![Screenshot](image2.png)

**FIGURE 12.** Selecting load control and advanced storage options for a FlashBlade backup repository within Veeam console
Next, click on Advanced... and check the setting as shown in the figure below to maximize overall throughput and take advantage of the performance provided by Pure Storage FlashArray//M and FlashBlade:

![Storage Compatibility Settings](image)

**FIGURE 13.** Recommended advanced backup repository storage options in Veeam console

CIFS/SMB shares require a mount (gateway) server for data movement to the FlashBlade backup repository. In our example testing we used the physical host with Veeam Backup & Replication 9.5 installed. However, virtualized Windows servers are also suitable for this task. Which method to use really depends on network connectivity – that is, we recommend putting the gateway server on the strongest possible network connection to the FlashBlade SMB share.

![Mount Server](image)

**FIGURE 14.** Selecting gateway/mount server for the CIFS/SMB backup repository in Veeam console
Finally, review the settings and click on **Apply** and then **Finish** to complete adding FlashBlade as Target repository.

FIGURE 15. Completing the FlashBlade backup repository within Veeam console
After the backup repository has been setup, we then connected our vCenter instance to the Veeam Backup and Replication console, so that inventory can be ingested and protection policies for any single or group of virtual machine(s) from our test group, or even any single file or group of files within a selected desktop, can be scheduled for backup.

First, click on the following series of options in the Veeam console:

**FIGURE 16**. Icons to select in sequence in the Veeam console for vCenter integration

This will spawn the **New VMware Server** wizard, so next we enter the IP address for vCenter (DNS name is also acceptable):

**FIGURE 17**. Enter vCenter DNS or IP address
Provide credentials that have administrative rights over your vCenter instance:

Click **Finish** to add the vCenter instance.
VEEAM BACKUP JOB SETUP AND RESULTS

We moved on to setup our Backup Job for the 250 virtual machines and applied a few specific job settings. The first step is to make the following selections from within the Veeam GUI to begin creating the backup job:

![Backup & Replication](image)

**FIGURE 20.** Icons to select within the Veeam console for a new Backup Job

With the New Backup Job wizard spawned, first give the job a name:

![New Backup Job](image)

**FIGURE 21.** Provide a name for the new Veeam backup job

As our vCenter instance was integrated with the Veeam console previously, we next selected our target 250 virtual machines for backup.
FIGURE 22. Click Add to browse vCenter instances for VMs to add to the backup job.

FIGURE 23. Highlight and select the 250 desktops to add to the backup job.
With the target VMs selected and added, we then selected our backup repository on FlashBlade. In this step, certain proxy server(s) can be selected specifically, or Veeam can automatically select one or more proxy VMs based upon network connectivity and availability relative to other active backup jobs.
Incremental backup jobs were used for our testing, though reverse incremental can certainly be used if the use case or customer preference calls for it.

![Incremental backup mode configuration](image1)

**FIGURE 26.** Set Incremental backup mode for best performance

Further selections that we made in our example backup job are shown in the following two screenshots. Compression level was set to **Optimal**, as we found this provided the best blend of archive size and throughput performance. Storage optimization was set to **Local Target**, as we had strong network connections throughout the design. Finally, **Guest File Indexing** was **enabled** so that individual files could be restored from within each guest OS.

![Compression and storage optimization settings](image2)

**FIGURE 27.** Set compression level to **Optimal** and storage optimization to **Local Target** for best overall throughput
With our configuration set, we scheduled our backup to run daily and immediately launched the job upon closing the wizard.

During this initial protection run for our 250 VMs, we configured three virtual Proxy Servers to backup VMs from FlashArray//M to the FlashBlade data hub to provide load balancing and resiliency. The metrics from the Veeam Backup & Replication Console, FlashArray//M, and FlashBlade GUls are shown in the following screenshots.
We can see that our entire dataset was ingested and backed up in approximately one hour and forty mintutes, which highlights the huge performance gains provided in running an all-flash datacenter, as legacy backup repository solutions require significantly more time to complete. Worth noting is that subsequent backup runs will generally take a fraction of the time and space of the initial run, as the only data moved will be changes to the original dataset. The Veeam console below confirms fast VM data backup as well as backup data space efficiencies from setting the compression level to Optimal.

On FlashBlade we experienced consistent 1GB/s write performance throughout the 250 VM backup creation.
The image below shows before and after capacity information from the Pure Storage FlashBlade File System hosting the 250 VMs' backup data, which agrees with the amount of data shown as transferred in the Veeam console.

![FIGURE 32. Pure Storage FlashBlade before (top) and after (bottom) showing space used during 250 VM Backup Job](image)

The FlashArray//M hosting the 250 desktops was easily able to provide the read throughput necessary to ingest the 250 VMs during the backup run while still maintaining very low latency so as not to adversely impact production workloads.

![FIGURE 33. Pure Storage FlashArray//M GUI during 250 VM ingest](image)

Worth noting again from this experiment is that the initial VM ingest will always be the most ‘expensive’ in terms of bandwidth, IOPS, and time. That is, subsequent backup job runs will only update the changes made to the target VMs, rather than running a full VM backup, as was done in this first example.

Next, we deleted the backup data from Pure Storage FlashBlade and repeated this experiment with four proxy servers to see what sort of additional efficiencies could be gained, if any.
All proxy VMs had the following hardware characteristics and were hosted on the same Pure Storage FlashArray//M and vCenter instances as our 250 target VMs:

<table>
<thead>
<tr>
<th>PROXY SERVER COMPONENT</th>
<th>VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>VCPU</td>
<td>10</td>
</tr>
<tr>
<td>RAM</td>
<td>16 GB</td>
</tr>
<tr>
<td>HARD DISK</td>
<td>500GB</td>
</tr>
<tr>
<td>SCSI CONTROLLER</td>
<td>VMWARE PARAVIRTUAL</td>
</tr>
<tr>
<td>NETWORK ADAPTERS (VMXNET 3)</td>
<td>2</td>
</tr>
<tr>
<td>NUMBER OF CONCURRENT VEEAM TASKS</td>
<td>10*</td>
</tr>
</tbody>
</table>

*Veeam recommends approximately 1 concurrent task per vCPU

TABLE 1. Proxy Server configuration used in all testing

We can see from the table below that we found no significant difference in adding an additional proxy server other than the job completion time decreasing by approximately ten minutes and the overall throughput increasing slightly.

We recommend a minimum of two proxy servers for resiliency and load-balancing. However, additional proxy servers should be considered if target VMs are in geographically dispersed locations and/or if minimizing backup duration as much as possible is important, or for much larger numbers of virtual machines. As always, we recommend experimenting in your own unique environment in order to determine the optimal configuration for your use cases and data.

<table>
<thead>
<tr>
<th>COMPONENT</th>
<th>3 PROXY SERVERS</th>
<th>4 PROXY SERVERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>DURATION</td>
<td>1HR 39 MINS</td>
<td>1HR 29 MINS</td>
</tr>
<tr>
<td>PROCESSING RATE</td>
<td>1GB/S</td>
<td>1GB/S</td>
</tr>
<tr>
<td>DATA PROCESSED</td>
<td>9.7 TB</td>
<td>9.8 TB</td>
</tr>
<tr>
<td>DATA READ</td>
<td>5.8 TB</td>
<td>5.9 TB</td>
</tr>
<tr>
<td>DATA TRANSFERRED</td>
<td>4.6 TB (1.3X)</td>
<td>4.6 TB (1.3X)</td>
</tr>
<tr>
<td>THROUGHPUT (ALL TIME)</td>
<td>1.7GB/S</td>
<td>2.0GB/S</td>
</tr>
</tbody>
</table>

TABLE 2. Relative performance of 3 and 4 Proxy Servers during 250 VM backup job (initial run)
The next step of the test process was to delete 50 VMs from our 250 VM pool within vCenter and restore those 50 desktops from Veeam Backup & Replication as quickly as possible. To start, we first powered off the 50 VMs and then deleted them from our vCenter inventory.

From there, we returned to the Veeam Backup & Replication 9.5 console and made the following icon selections to kick off the recovery operation:

Once the Restore Wizard was launched, we selected the **Entire VM** restore option **Restore from backup**:

The next step in the wizard was to add the VMs that we deleted to the recovery job operation from our earlier backup.
With the 50 VMs selected, our next selection was to **Restore to the original location** within the vCenter instance.

With the setting changes made, we kicked off the recovery operation. The recovery job can be viewed in the running jobs section of the Veeam console.
Users can optionally check progress on a single VM by double-clicking on any single instance:

Upon recovery job completion, we found that we could restore the 50 VMs in approximately 33 minutes as the next screenshot highlights. This extremely fast recovery operation again emphasizes the huge advantages that an all-flash datacenter running Veeam's optimized software stack provides. Achieving throughput comparable to this solution using legacy storage would require many rack units (if not several racks) of legacy spinning disk.

With Pure Storage FlashBlade, we achieved this speed in just 4U, with only half of the chassis populated.
The Pure Storage FlashBlade GUI confirms the speed at which our 50 virtual machines were read from the Veeam archive hosted on it and transferred back to the Pure Storage FlashArray//M so that they could begin servicing workloads with minimal recovery time.
CONCLUSION

Through this simple demonstration, we have shown that the combination of Pure Storage FlashArray//M, Veeam Backup & Replication, and Pure Storage FlashBlade delivers the entire suite of data performance, protection, and data mobility. Additional workloads can easily be mixed and managed alongside this example from the interfaces shown in this guide – across an entire business.

As data capacity requirements grow in both primary and secondary use cases, FlashBlade, FlashArray//M, and Veeam Backup & Replication can be non-disruptively scaled and upgraded – the process is completely transparent to your customers. Running all-flash for both primary and backup data workloads provides transformative performance for primary workloads, extremely fast recovery and mobility, and extreme density, which minimizes datacenter real estate as well as power and cooling costs. Veeam is a natural partner in this space as they provide a software-based solution that makes data orchestration across these on-premises solutions and cloud-tier providers effortless to manage from a single pane of glass.

Finally, systems administrators can move away from tedious and repetitive tasks that are just focused on keeping infrastructure online, and instead move up the stack to concentrate on tasks that will improve the company as whole, with the comfort and knowledge that their data is secure and simple to restore when necessary.