Pure Storage and Veeam Backup and Replication for Protecting Mission Critical Data

1 + 1 = 3
Pure Storage + Veeam
Better Together

Contents

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 the 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.

The following three core pieces of next generation storage and backup technology will be highlighted within this document and explanations will be provided in 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. Pure Storage 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 archival.
- **Pure Storage FlashBlade:** A fast and dense unstructured data platform from Pure Storage that provides on-premise rapid backup and recovery of archival data. Pure Storage FlashBlade additionally 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 below connectivity diagram illustrates how the various solutions are integrated with one another and illustrates the lack of a single point of failure within the design:

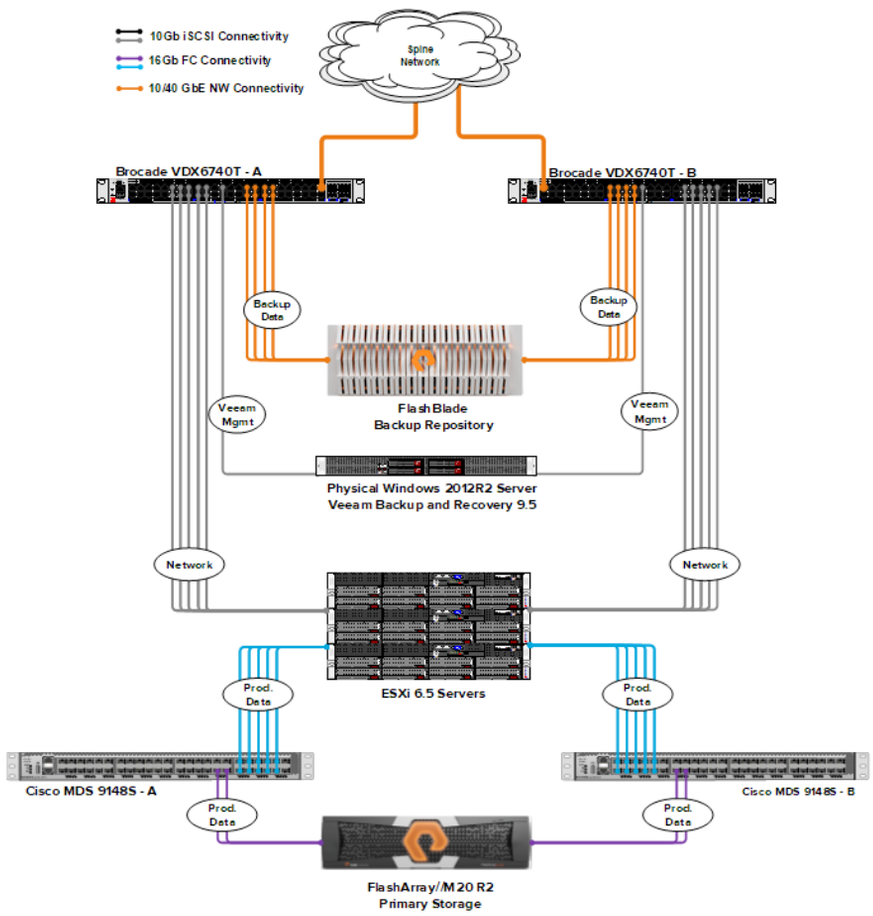


Figure 1: High-Level connectivity diagram of the overall solution

From the above diagram, here are more detailed specifics for each hardware and software elements:

- 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 the Purity Operating Environment version 4.9.3. The array used Fibre-Channel connectivity (though iSCSI would also be fully supported), and will serve as our data source for backup. The vCenter instance, above described 250 desktops, Veeam Proxy servers and all other vSphere-based components were hosted on the FlashArray//M.

- Veeam Backup and 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:

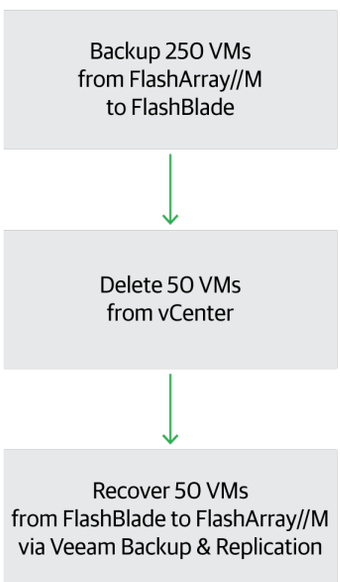


Figure 2: Test process to be executed

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 next diagram:

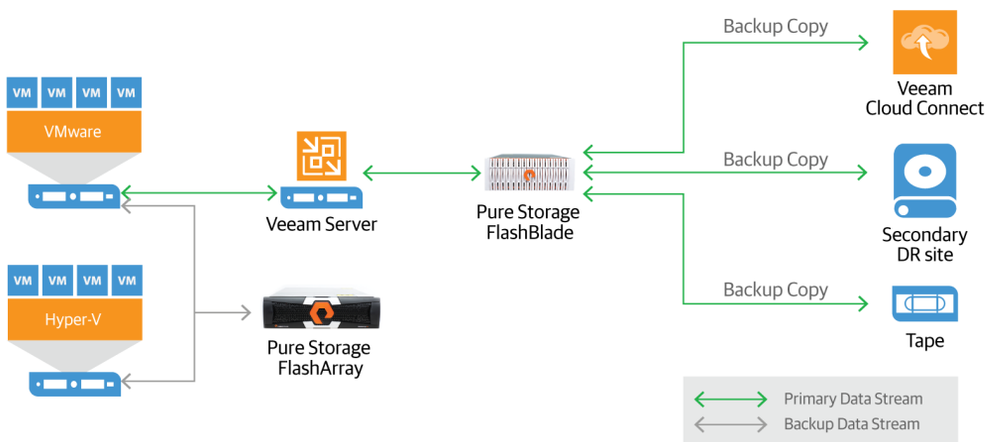


Figure 3: Veeam Backup and Replication dataflow architecture

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 and 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 and 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 and 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 to make is editing the registry of the server running Veeam Backup and Replication to increase the overall throughput and parallelism for 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 and Replication —> MaxSnapshotsPerDatastore**. The updated key and path to it can be seen in the below screenshot:

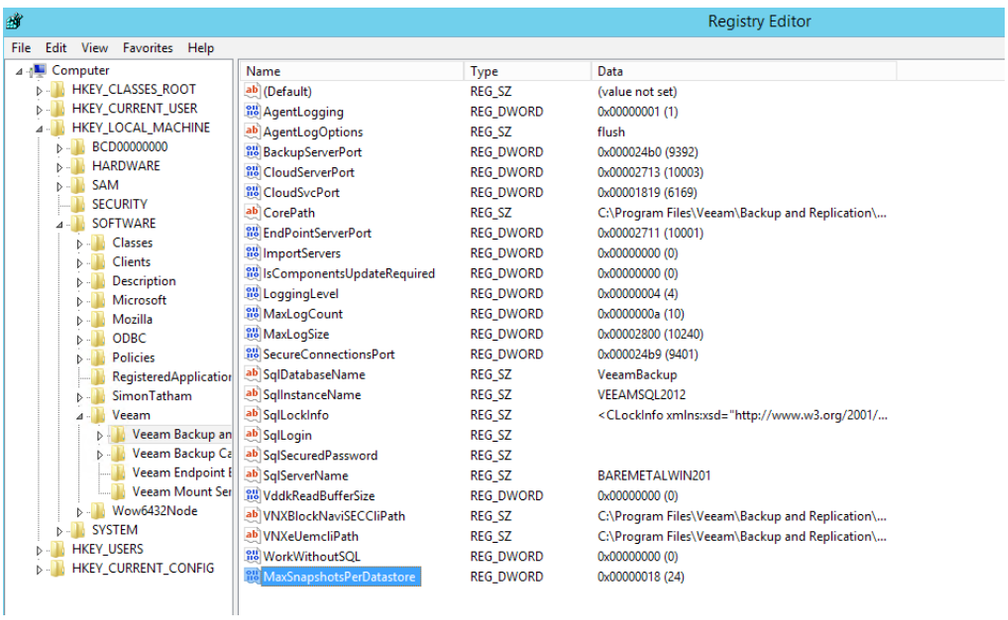


Figure 4: Registry entry modified for increased throughput during testing

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 next diagram illustrates how data is moved for an SMB/CIFS backup repository share:

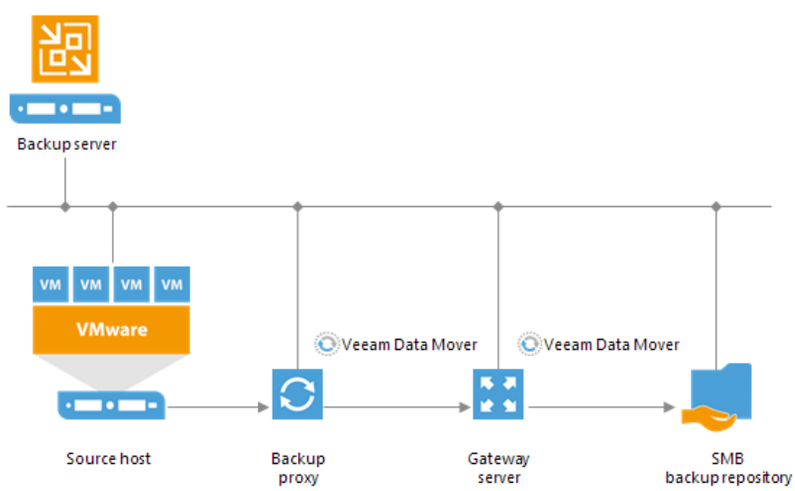


Figure 5: SMB/CIFS data flow diagram for Veeam

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

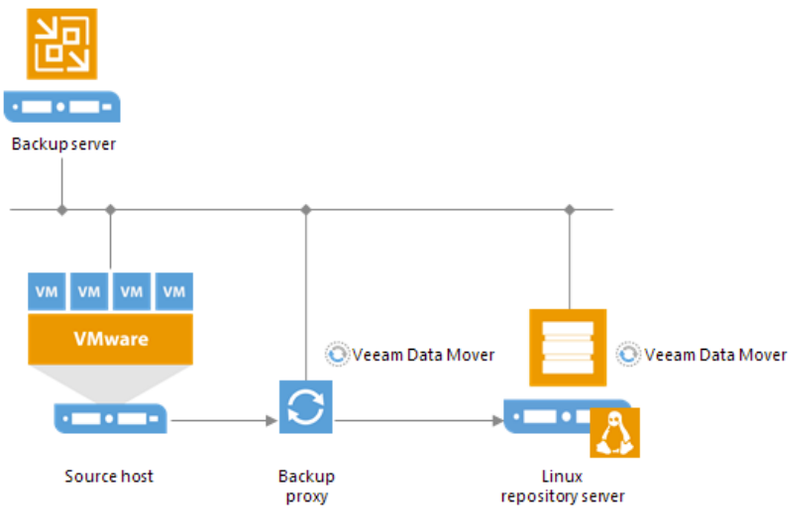


Figure 6: NFS dataflow diagram for Veeam

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 below example will use the SMB protocol, though NFS could also be used.

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

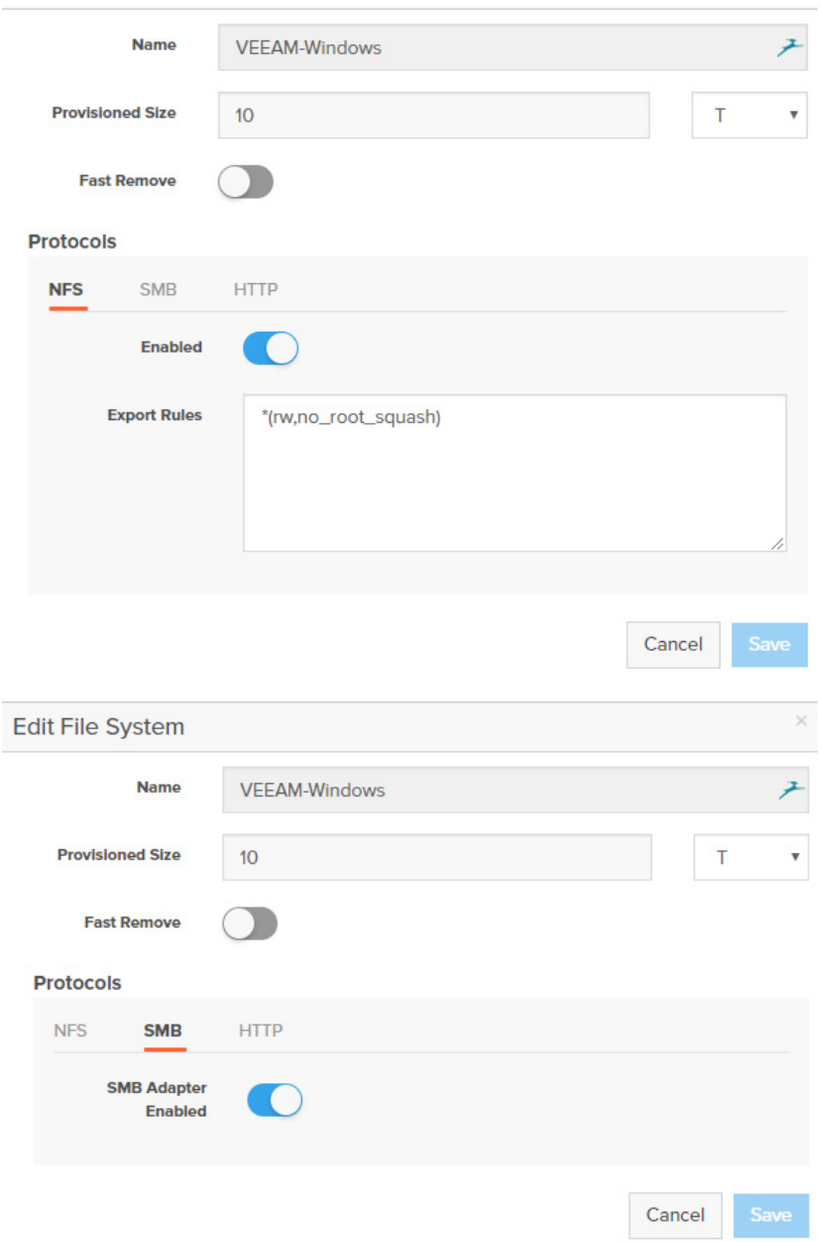


Figure 8: Naming, sizing and enabling the NFS and SMB protocols on a Pure Storage FlashBlade File System

With the backup repository location created on Pure Storage FlashBlade, we next moved over to the Veeam Backup and Replication 9.5 console. Under the **Backup Infrastructure** select the **Backup Repositories** select **Add Repository**. First, provide a name for the repository:

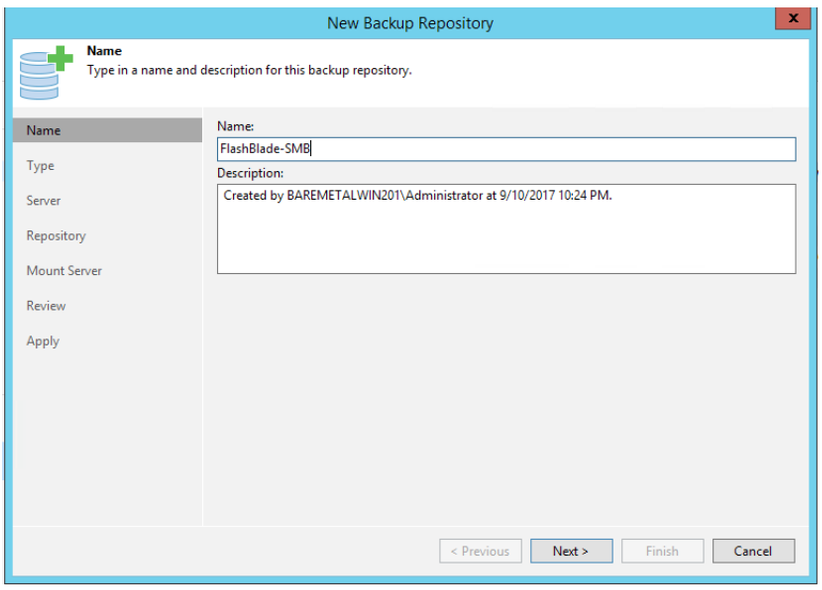


Figure 9: Naming new FlashBlade backup repository in Veeam Backup and Replication

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

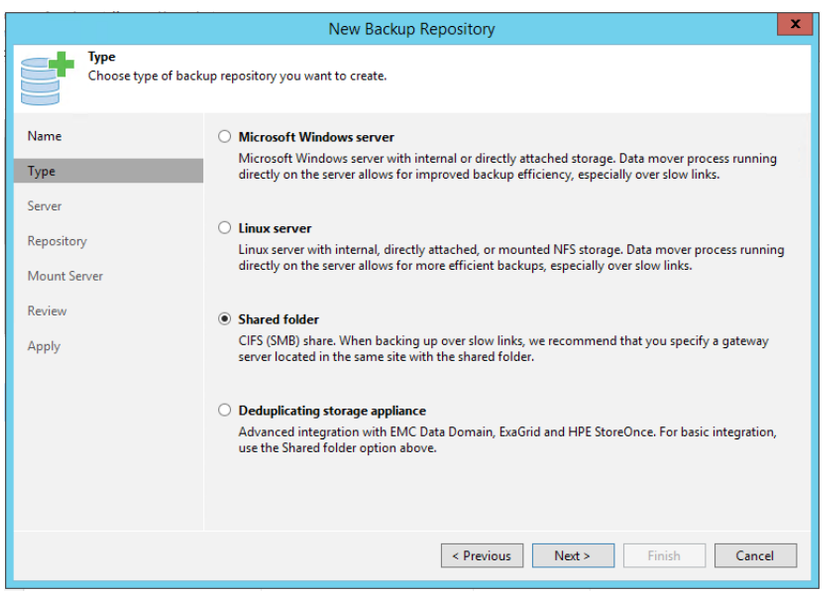


Figure 10: Selecting type of backup repository within Veeam console

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 next screenshot shows.

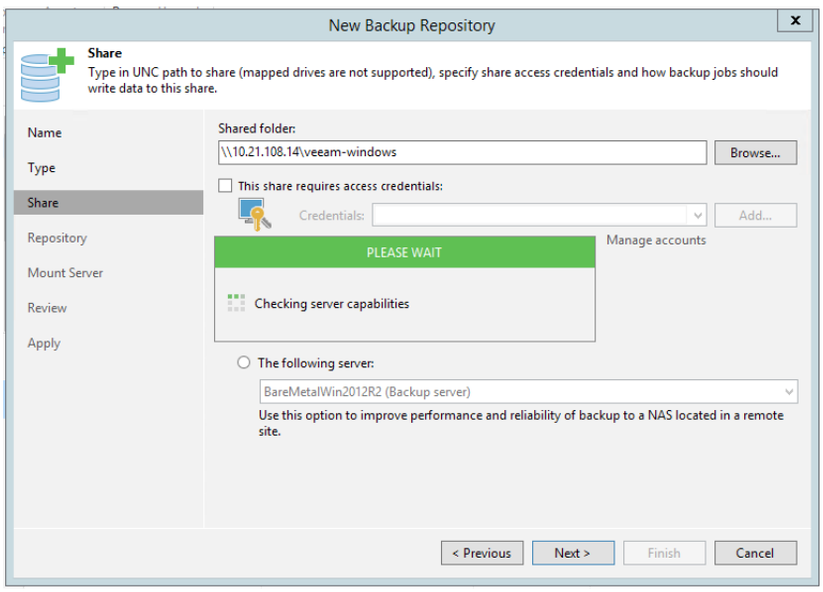


Figure 11: Entering CIFS/SMB data path to FlashBlade within Veeam console

With the path entered correctly the path to the folder and amount of overall and available space are shown. We elected to not 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.

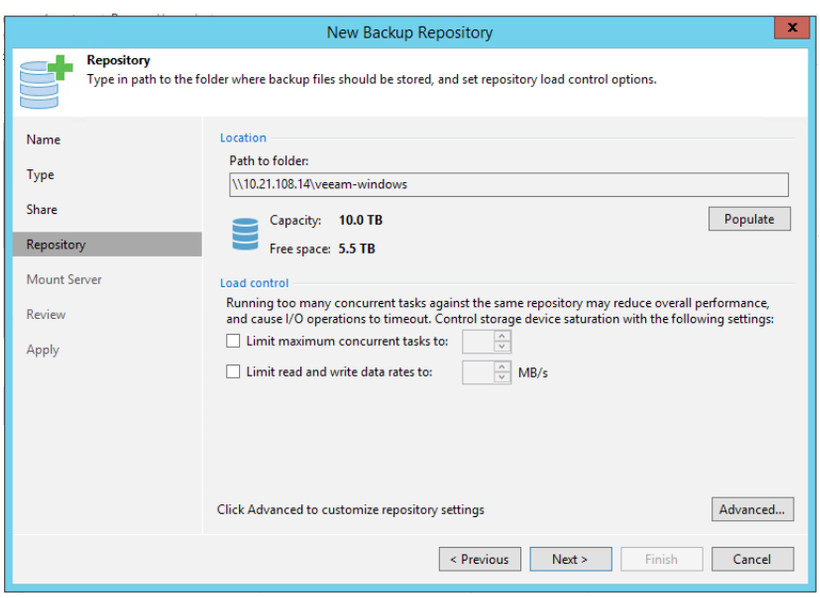


Figure 12: Selecting load control and advanced storage options for FlashBlade backup repository within Veeam console

Next, click on the Advanced… and the check the setting as shown in below figure to maximize overall throughput and take advantage of the performance provided by Pure Storage FlashArray//M and FlashBlade:

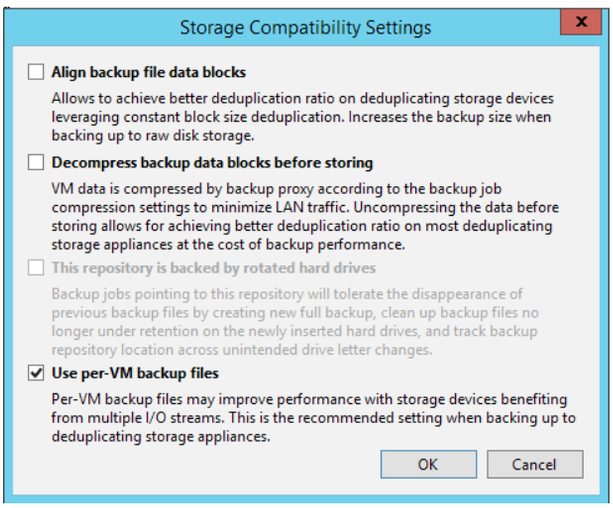


Figure 13: Recommended advanced backup repository storage options in Veeam console

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 and Replication 9.5 installed, however virtualized Windows servers are also suitable for this task and which method to use really depends on network connectivity. That is, we recommend putting the gateway server on as strong of a network connection to the FlashBlade SMB share as possible.

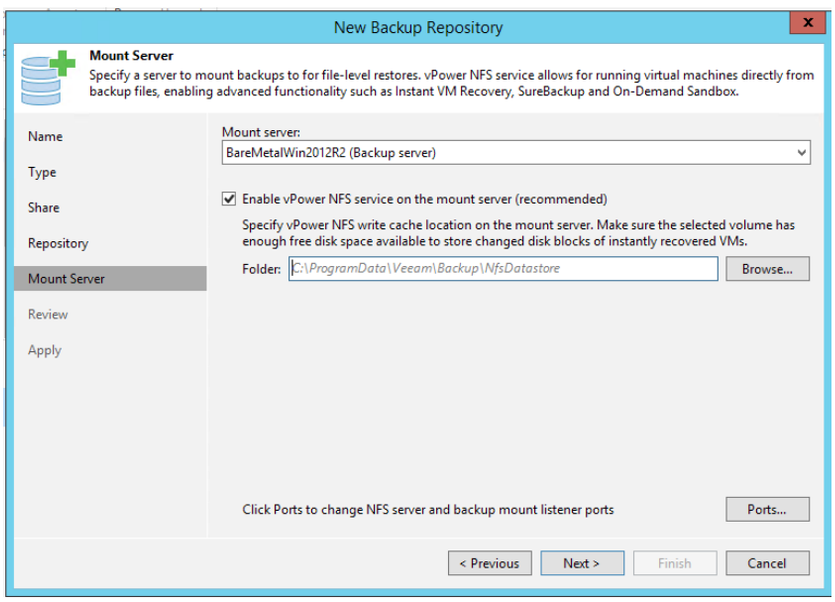


Figure 14: Selecting gateway/mount server for CIFS/SMB backup repository in Veeam console

Finally, review the setting and once confirm click on Apply and Finish to complete adding the FlashBlade as Target repository.

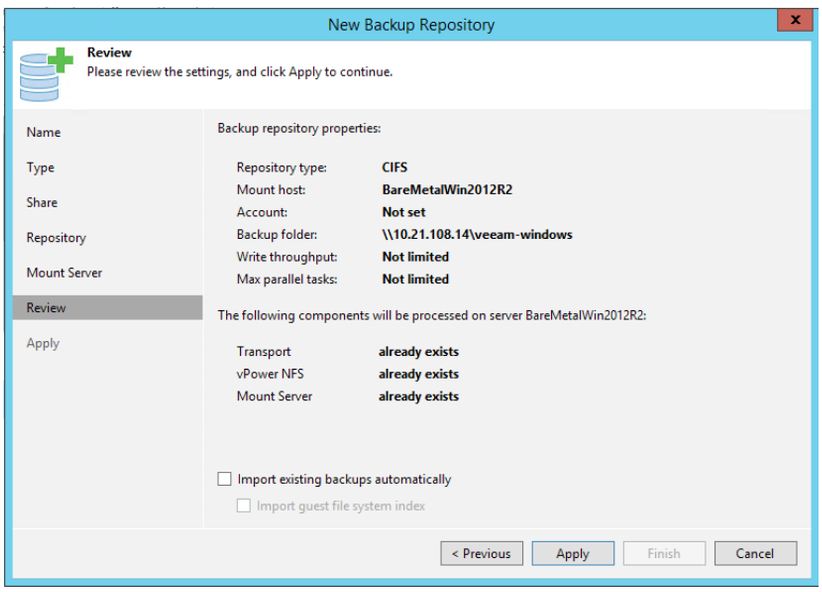


Figure 15: Completing FlashBlade backup repository within Veeam console

Veeam Backup and Replication vSphere Integration Setup

After the backup repository has been setup, we next connected our vCenter instance to the Veaam Backup & Replication console, so that the 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:

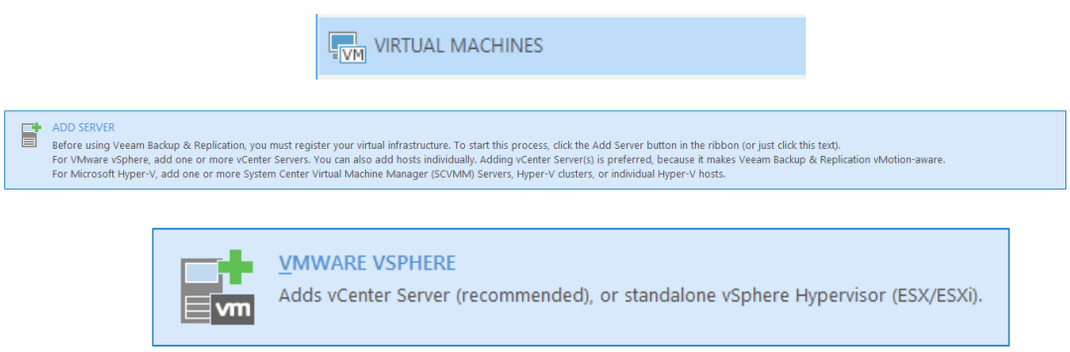


Figure 16: Icons to select in sequence in Veeam console for vCenter integration

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

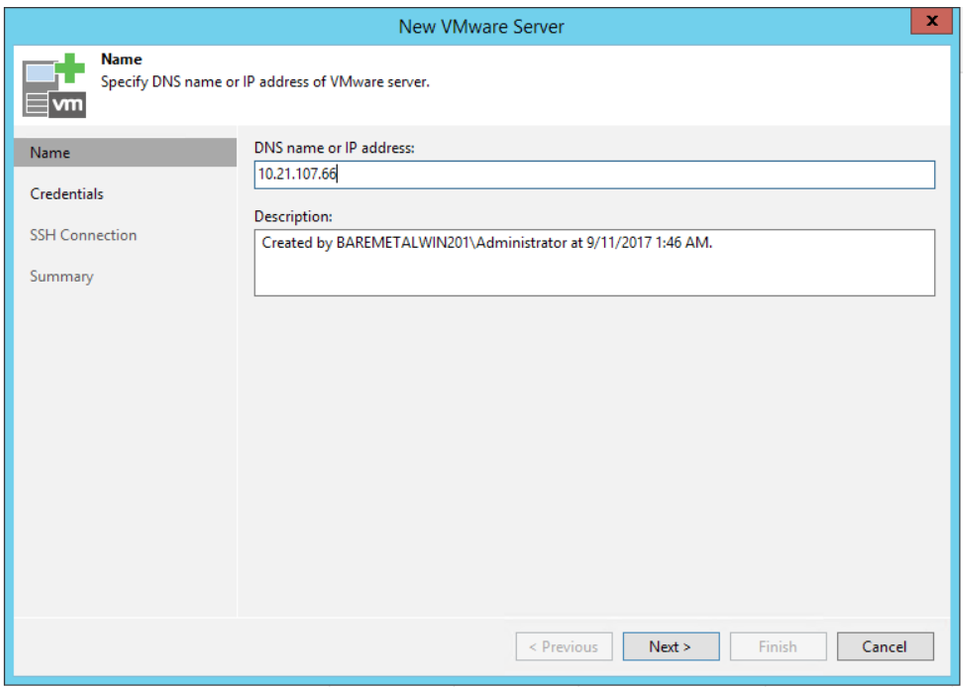


Figure 17: Enter vCenter DNS or IP address

Provide credentials that have administrative rights over your vCenter instance:

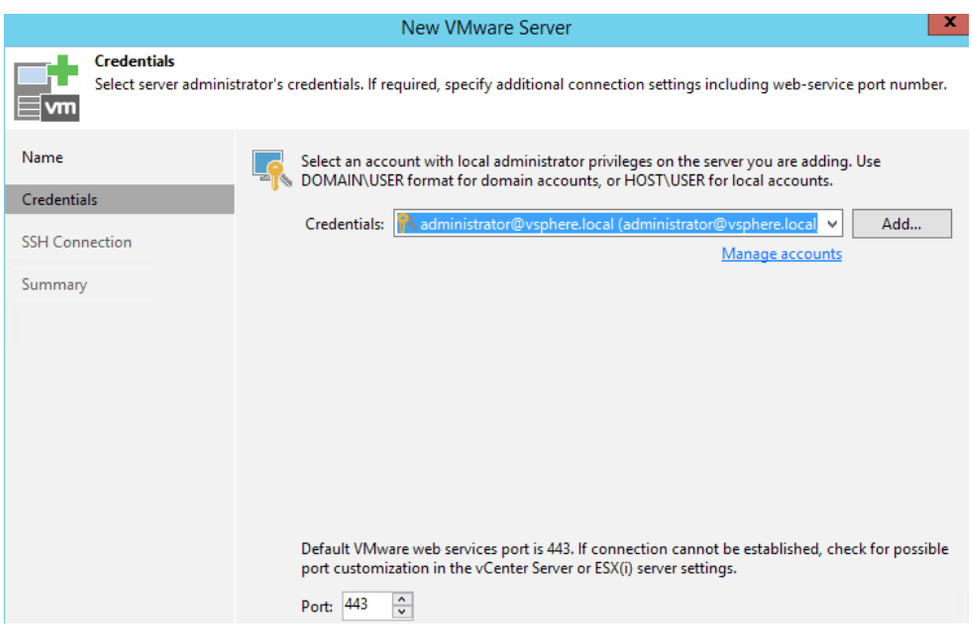


Figure 18: Enter administrative credentials for vCenter from Veeam console

Click **Finish** to add the vCenter instance.

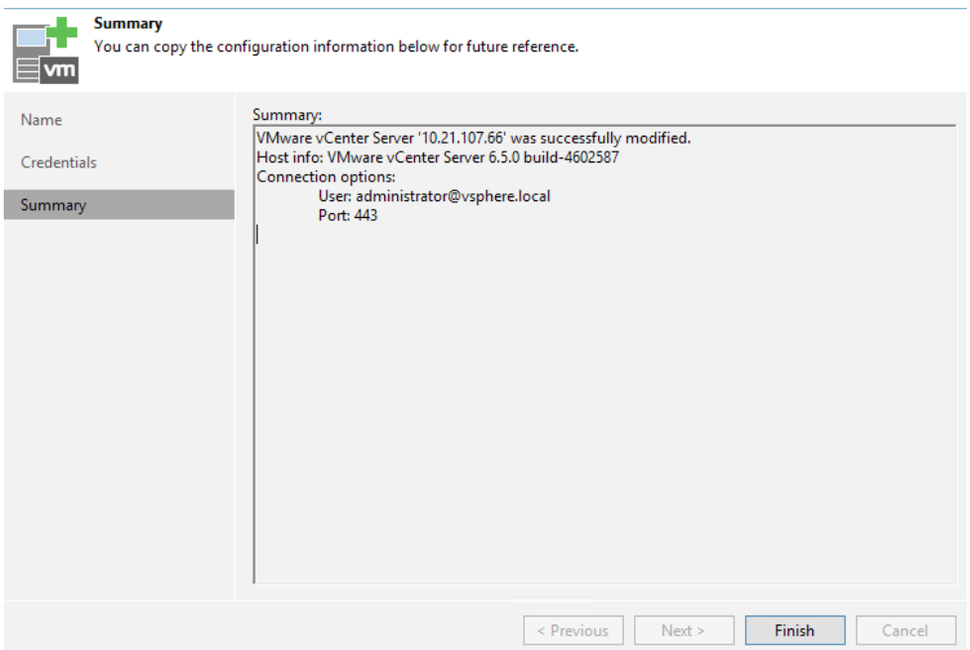


Figure 19: Confirmation of vCenter integration with Veeam

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:

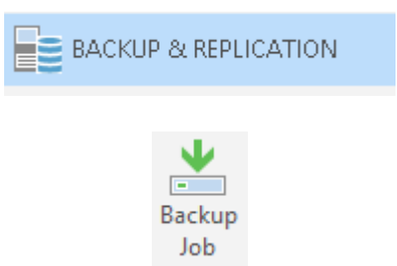


Figure 20: Icons to select within Veeam console for new Backup Job

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

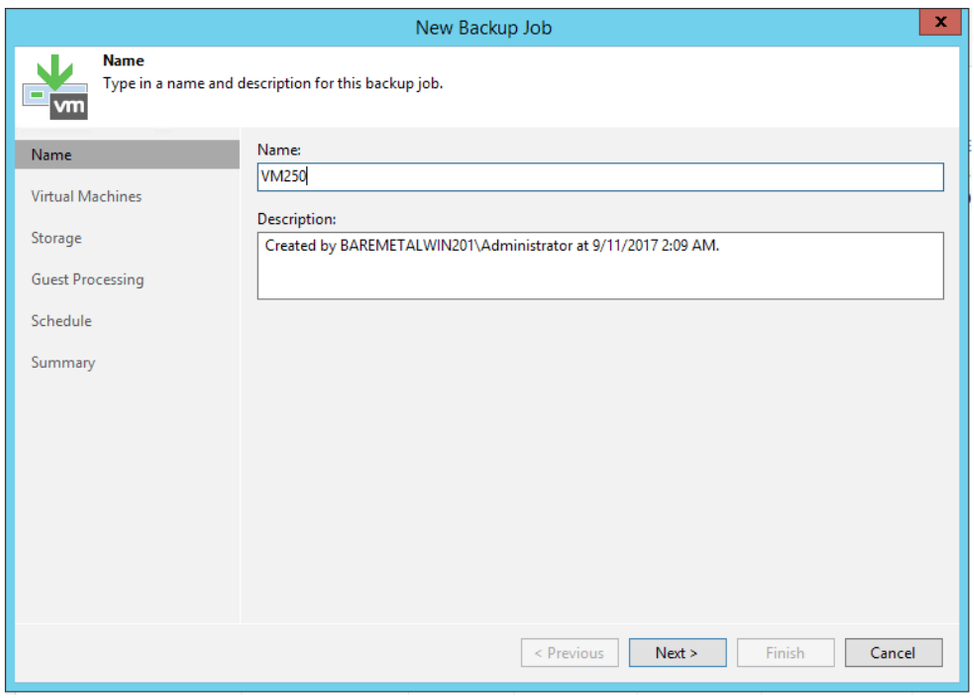


Figure 21: Provide a name for new Veeam Backup Job

As our vCenter instance was integrated to the Veeam console previously, we next selected our target 250 virtual machines for backup.

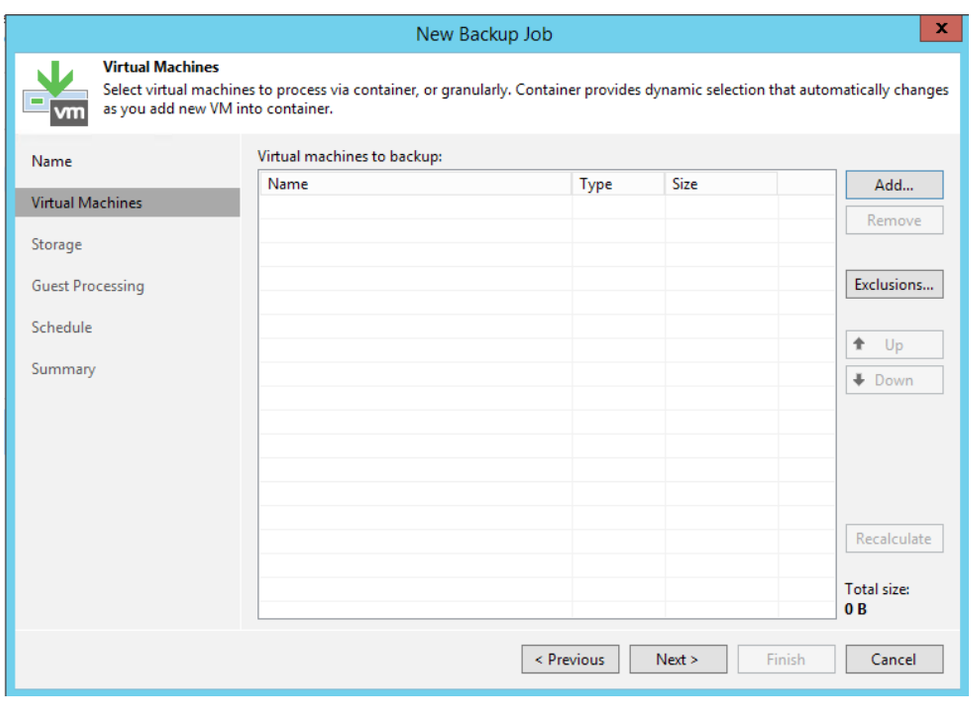


Figure 22: Click 'add' to browse vCenter instance for VMs to add to Backup Job

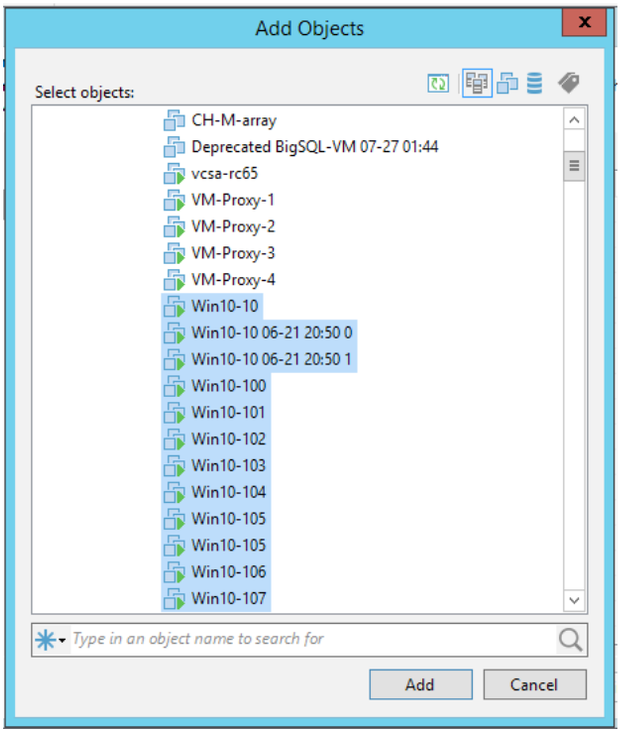


Figure 23: Highlight and select the 250 desktops to add to the Backup Job

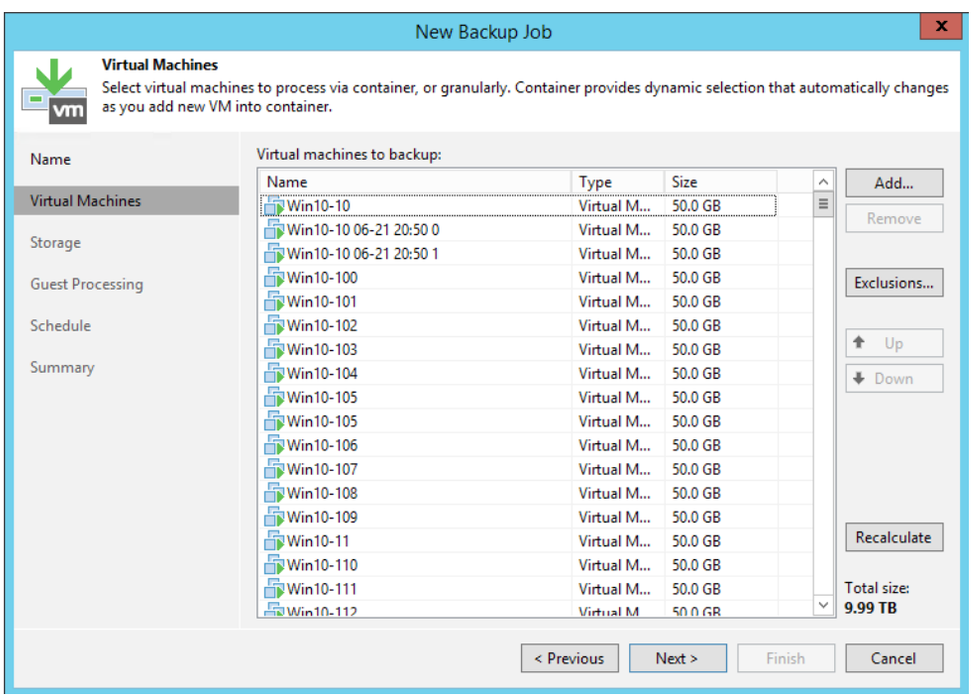


Figure 24: 250 desktops added to Backup Job

With the target VMs selected and added, next we 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.

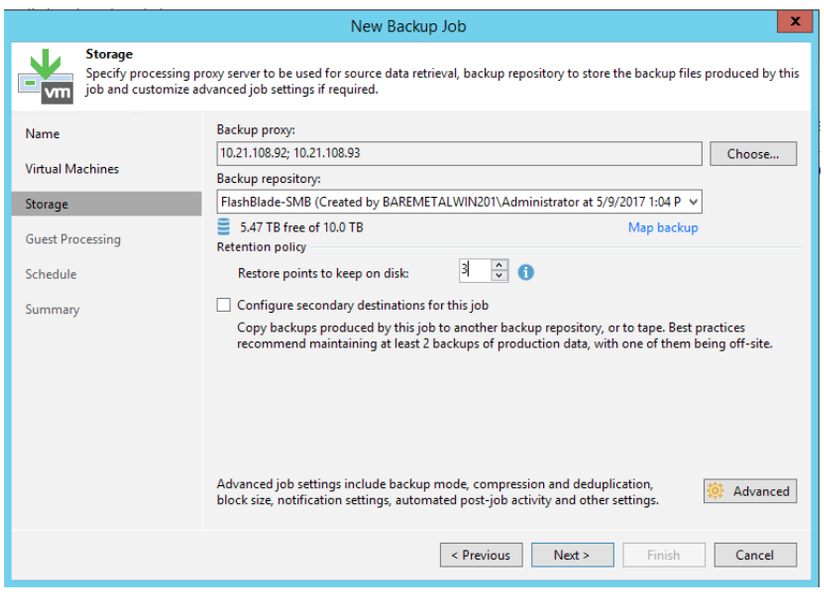


Figure 25: Select the FlashBlade repository, set proxy server affinity, set number of restore points and select any advanced options for Backup Job

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.

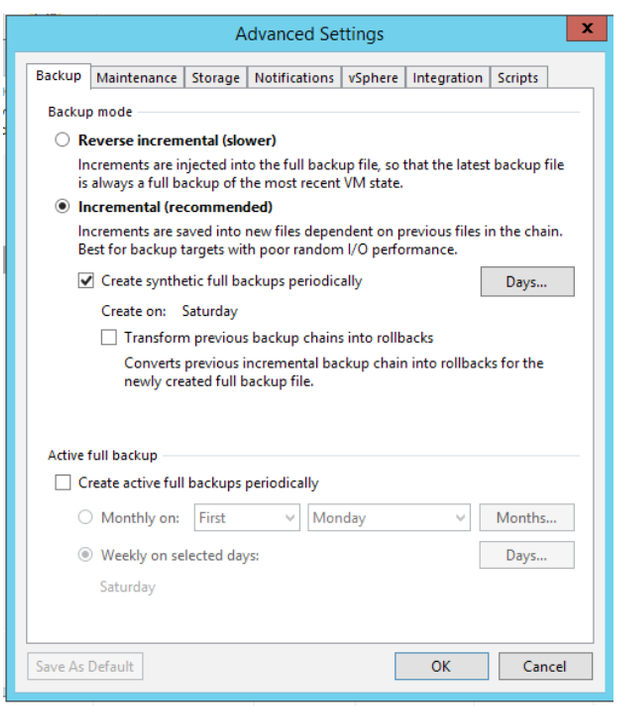


Figure 26: Set Incremental backup mode for best performance

Further selections that we made in our example backup job are shown in the next two screenshots below. 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.

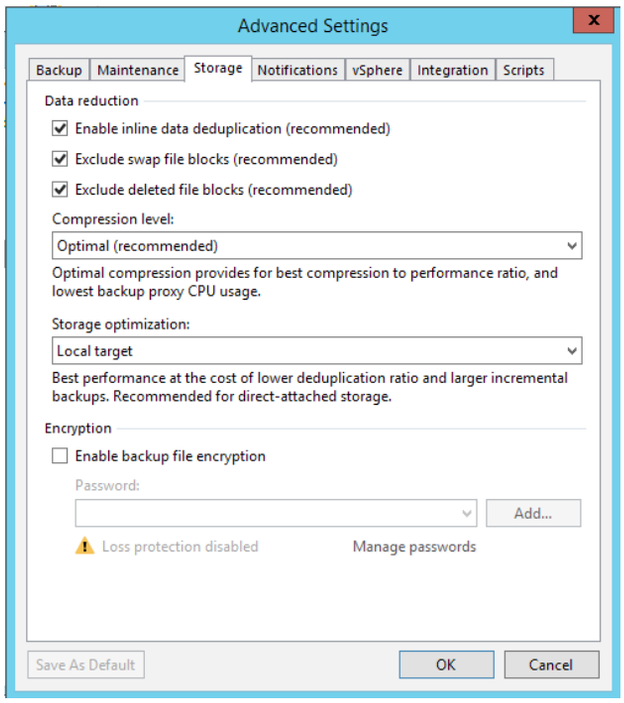


Figure 27: Set compression level to 'optimal' and storage optimization to 'local target' for best overall throughput

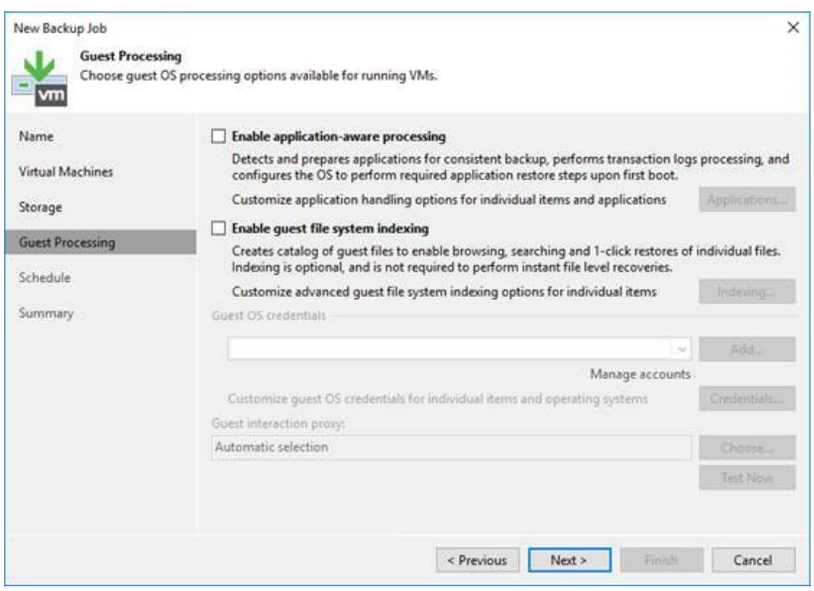


Figure 28: Optional application aware image processing and indexing

With our configuration set, we scheduled our backup to run daily and immediately launched the job upon closing the wizard.

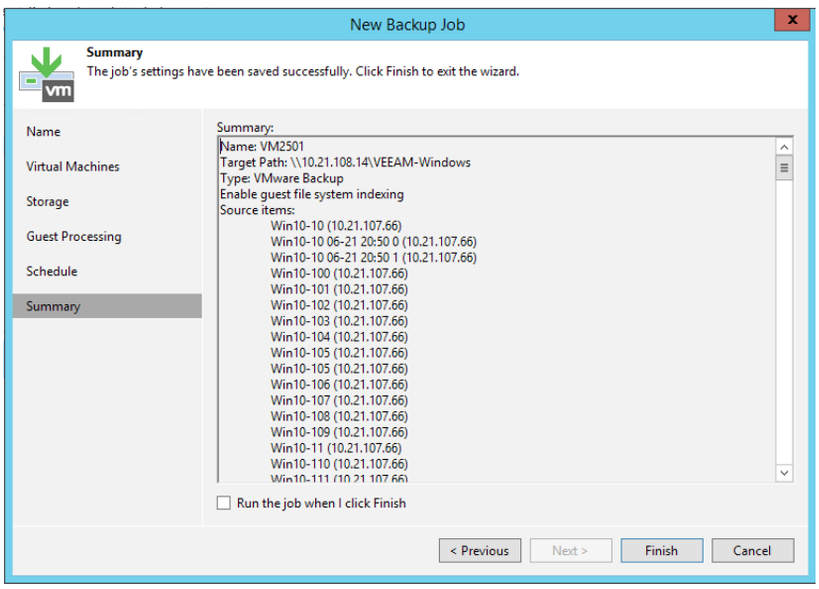


Figure 29: Backup Job summary screen

During this initial protection run for our 250 VMs, we configured three virtual Proxy Servers to backup VM's from FlashArray//M to the FlashBlade data hub to provide load balancing and resiliency. The metrics from the Veeam Backup and Replication Console, Pure Storage FlashArray//M and FlashBlade GUIs, shown below,

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 significant 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.

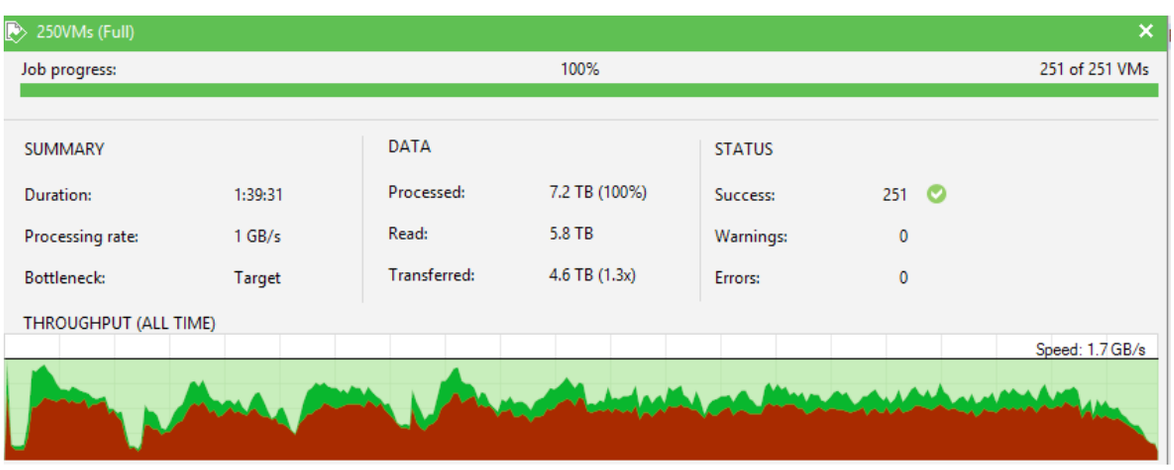


Figure 30: Veeam console showing important metrics during initial 250 VM Backup Job

On Pure Storage FlashBlade we experienced consistent 1GB/s write performance throughout the 250 VM backup creation.

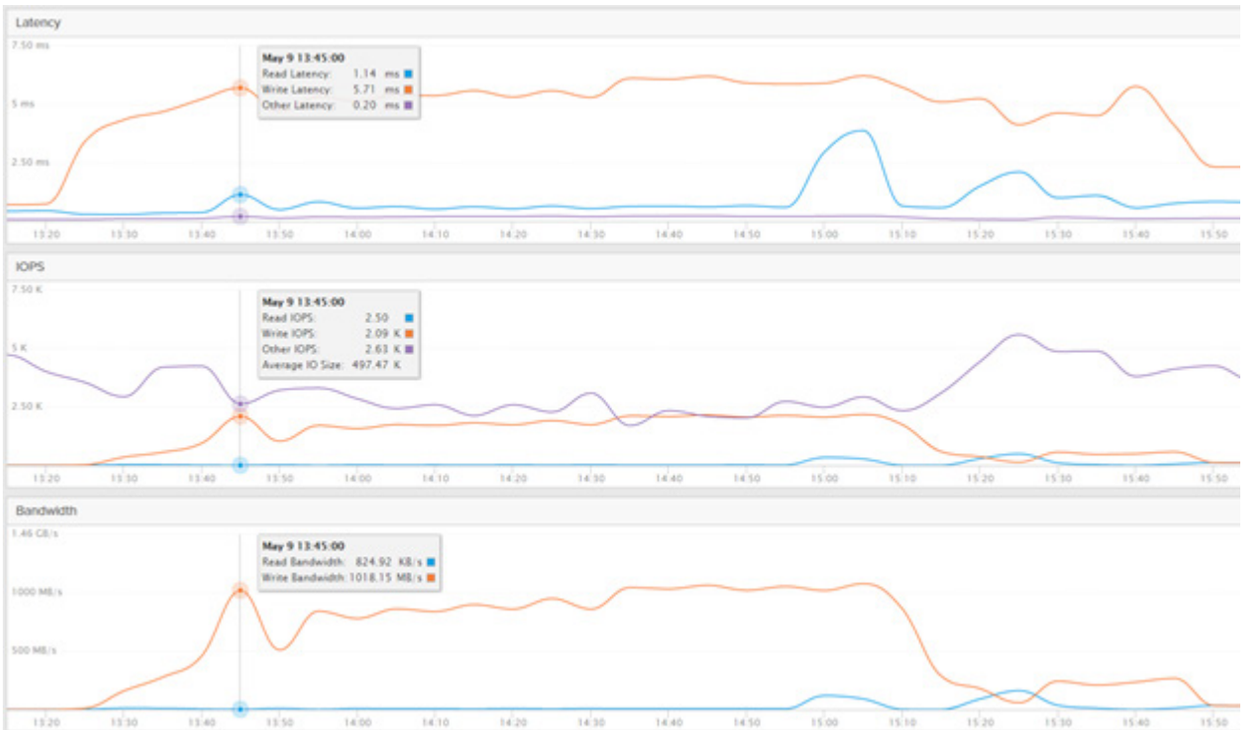


Figure 31: FlashBlade GUI during 250 VM Veeam Backup Job run

The below image shows before and after capacity information from the Pure Storage FlashBlade File System hosting the 250 VM backup data, which agrees with the amount of data shown as transferred in the Veeam console.

VEEAM-Windows	10.00 T	178.87 G	2%	1.0 to 1	177.79 G	NFS, SMB	X
VEEAM-Windows	10.00 T	4.73 T	47%	1.0 to 1	4.72 T	NFS, SMB	X

Figure 32: Pure Storage FlashBlade before (top) and after (bottom) showing space used during 250 VM Backup Job

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 to not adversely impact production workloads.

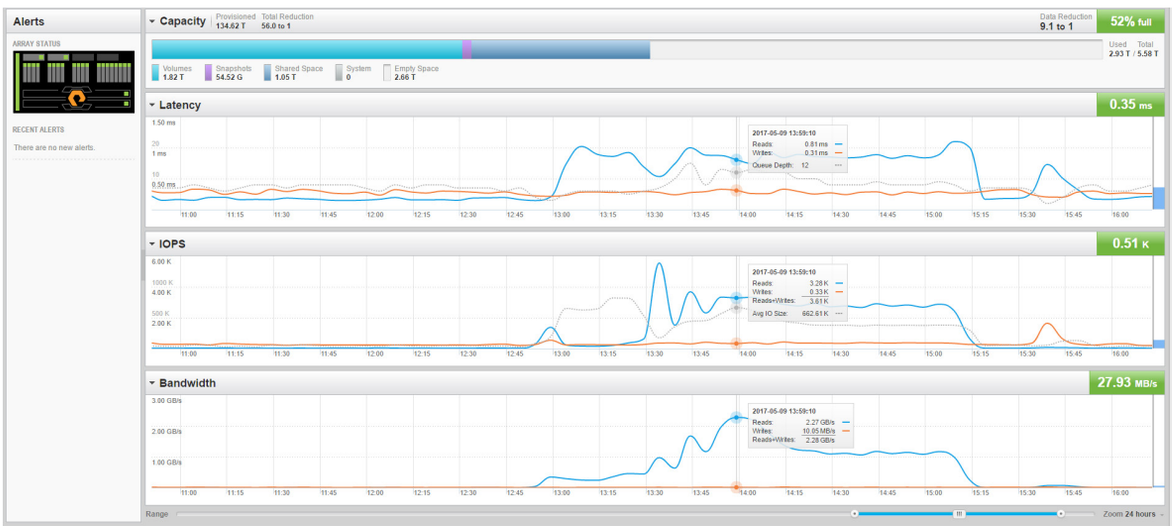


Figure 33: Pure Storage FlashArray//M GUI during 250 VM ingest

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.

We next deleted this backup data from Pure Storage FlashBlade and repeated this experiment with four proxy servers to see what sort of further 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 are our 250 target VMs:

Proxy Server Component	Value
vCPU	10
RAM	16 GB
Hard Disk	500GB
SCSI Controller	VMware Paravirtual
Network Adapters (VMXNET 3)	2
Number of Concurrent Veeam Tasks	24*

**Veeam recommends approximately 2x max concurrent tasks relative to total number of CPUs*

Table 1: Proxy Server configuration used in all testing

We can see from the below table 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 disperse 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.

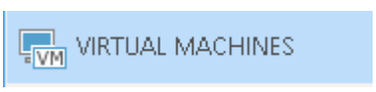
Component	3 Proxy Servers	4 Proxy Servers
Duration	1hr 39 mins	**1hr 29 mins**
Processing Rate	1GB/s	1GB/s
Data Processed	9.7 TB	9.8 TB
Data Read	5.8 TB	5.9 TB
Data Transferred	4.6 TB (1.3x)	4.6 TB (1.3x)
Throughput (ALL TIME)	1.7GB/s	**2.0GB/s**

Table 2: Relative performance of 3 and 4 Proxy Servers during 250 VM Backup Job (initial run)

Veeam Virtual Machine Restore Setup and Results

The next step of the test process was to delete 50 VMs from our 250 VM pool from within vCenter and restore those 50 desktops from the Veeam Backup and Replication as quickly as possible. To start, we first powered off the 50VMs and then deleted them from our vCenter inventory.

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



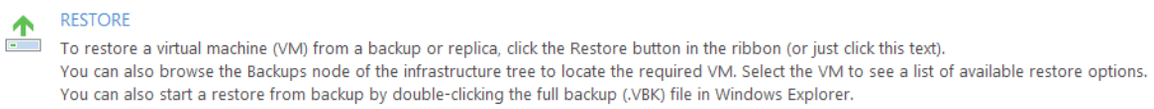


Figure 34: Icons to select in order to start Veeam Restore Wizard

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

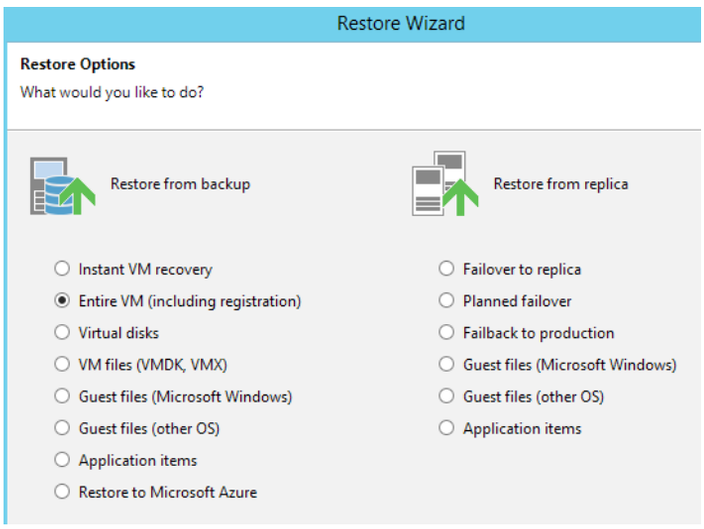
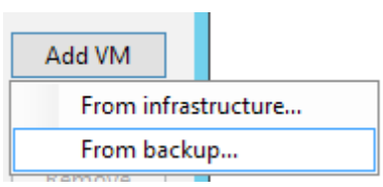


Figure 35: Selecting Entire VM restore option from backup

The next step in the wizard was to add the VMs that we deleted to the recovery job operation from our earlier backup.



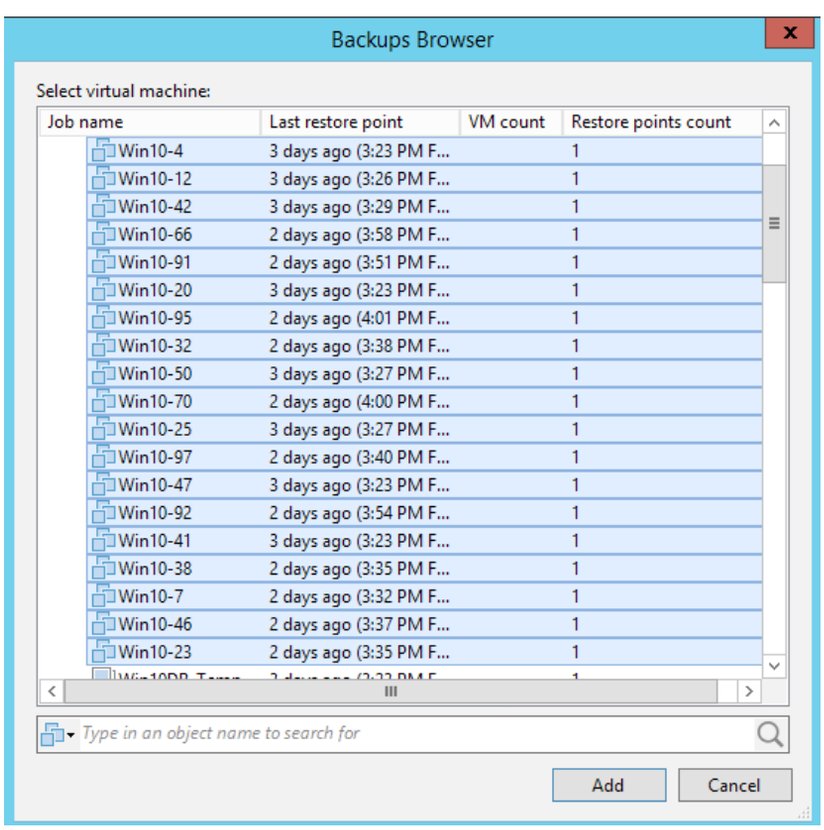


Figure 36: Selecting and adding 50 VMs for restore from Veeam backup on FlashBlade

With the 50 VMs selected, our next selection was to **Restore to the original location** within the vCenter instance.

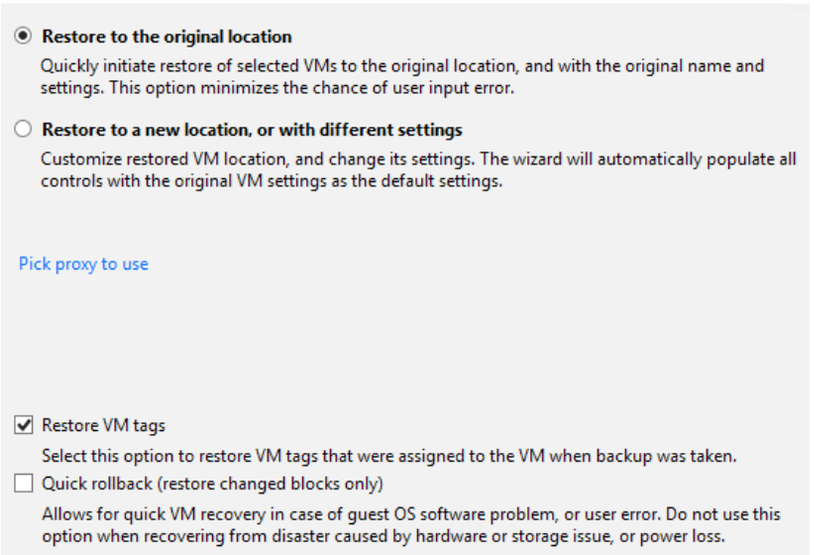


Figure 37: Selecting restore to original location for 50 VMs

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.

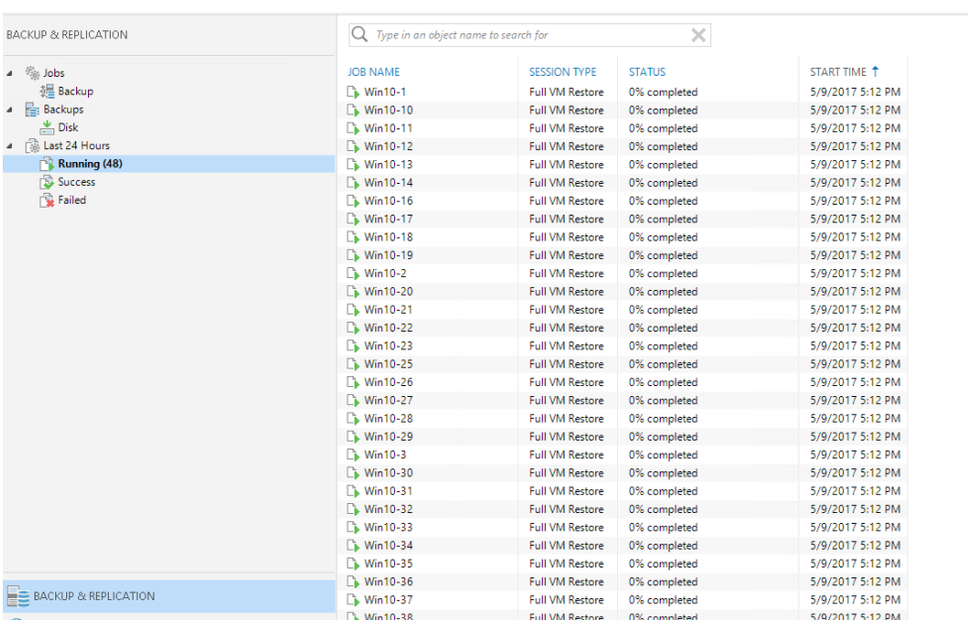


Figure 38: 50 VM restore job running

Users can optionally check progress on a single VM by double-clicking on any single instance:

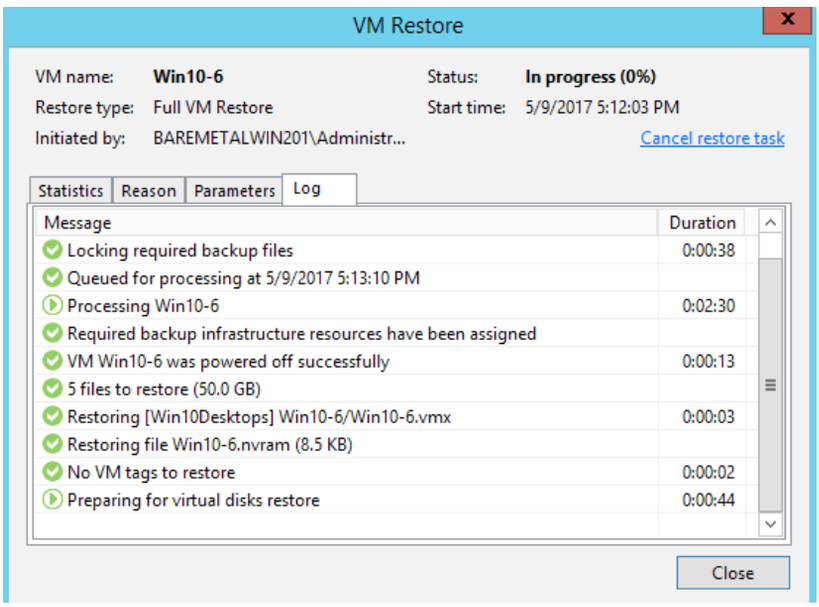
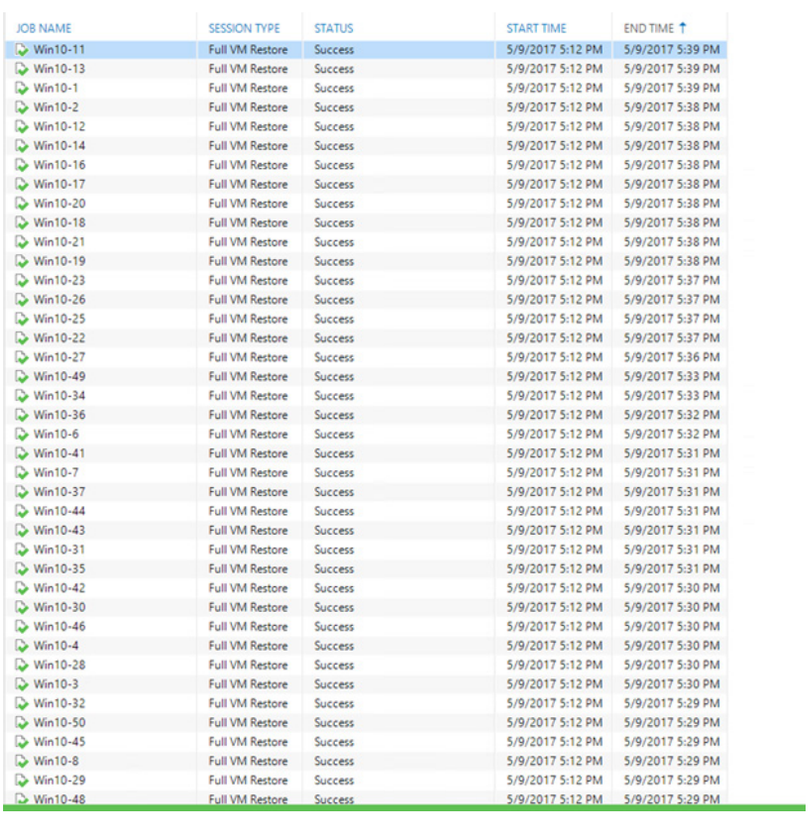


Figure 39: Individual VM restore progress example

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 this level of throughput using legacy storage would require many rack units (if not several racks) of legacy spinning disk to provide comparable bandwidth to this solution. With Pure Storage FlashBlade, we accomplished this in just 4U with only half of the chassis populated.



JOB NAME	SESSION TYPE	STATUS	START TIME	END TIME ↑
Win10-11	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:39 PM
Win10-13	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:39 PM
Win10-1	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:39 PM
Win10-2	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:38 PM
Win10-12	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:38 PM
Win10-14	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:38 PM
Win10-16	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:38 PM
Win10-17	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:38 PM
Win10-20	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:38 PM
Win10-18	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:38 PM
Win10-21	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:38 PM
Win10-19	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:38 PM
Win10-23	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:37 PM
Win10-26	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:37 PM
Win10-25	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:37 PM
Win10-22	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:37 PM
Win10-27	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:36 PM
Win10-49	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:33 PM
Win10-34	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:33 PM
Win10-36	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:32 PM
Win10-6	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:32 PM
Win10-41	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:31 PM
Win10-7	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:31 PM
Win10-37	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:31 PM
Win10-44	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:31 PM
Win10-43	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:31 PM
Win10-31	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:31 PM
Win10-35	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:31 PM
Win10-42	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:30 PM
Win10-30	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:30 PM
Win10-46	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:30 PM
Win10-4	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:30 PM
Win10-28	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:30 PM
Win10-3	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:30 PM
Win10-32	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:29 PM
Win10-50	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:29 PM
Win10-45	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:29 PM
Win10-8	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:29 PM
Win10-29	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:29 PM
Win10-48	Full VM Restore	Success	5/9/2017 5:12 PM	5/9/2017 5:29 PM

Figure 40: Completed 50 VM restore operation confirming very fast recovery

The Pure Storage FlashBlade GUI confirms the speed at which our 50 virtual machines were read from the Veeam archive hosted on it and back to the Pure Storage FlashArray//M so that they could begin servicing workload with minimal recovery time.

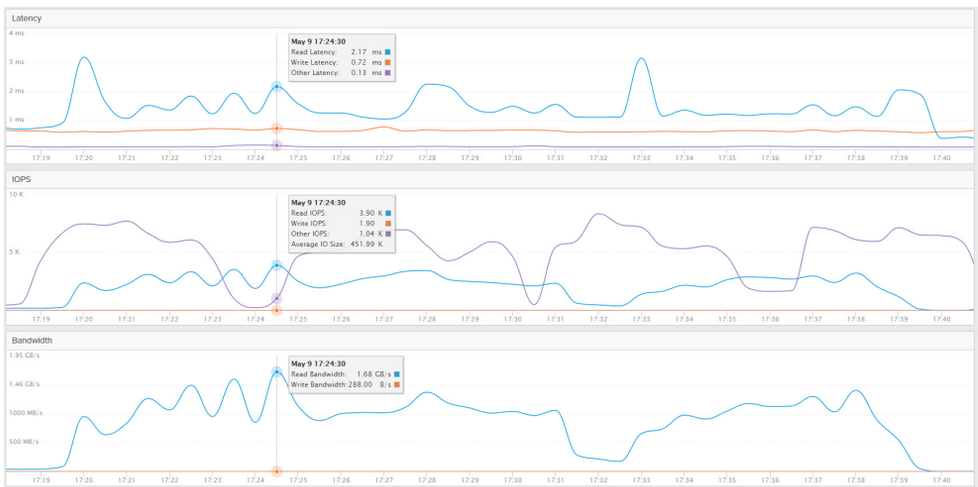


Figure 41: Pure Storage FlashBlade GUI showing high read throughput during 50 VM restore job

Conclusions

Through this simple demonstration, we have proven 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 of this example from the interfaces shown in this guide across the entire business.

As data capacity requirements grow in both primary and secondary use cases, Pure Storage FlashBlade, FlashArray//M and Veeam Backup & Replication can non-disruptively scale and be upgraded completely transparently to your customers. The benefits of running all-flash for both primary and backup data workloads provide 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 focused on just keeping their 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 data is secure and simple to restore when necessary.

About Veeam Software

Veeam® recognizes the new challenges companies across the globe face in enabling the Always-On Business™, a business that must operate 24.7.365. To address this, Veeam has pioneered a new market of Availability for the Always-On Enterprise™ by helping organizations meet recovery time and point objectives (RTPO™) of < 15 minutes for all applications and data, through a fundamentally new kind of solution that delivers high-speed recovery, data loss avoidance, verified protection, leveraged data and complete visibility. Veeam Availability Suite™, which includes Veeam Backup & Replication™, leverages virtualization, storage, and cloud technologies that enable the modern data center to help organizations save time, mitigate risks, and dramatically reduce capital and operational costs.

Founded in 2006, Veeam currently has 49,000 ProPartners and more than 255,000 customers worldwide. Veeam's global headquarters are located in Baar, Switzerland, and the company has offices throughout the world. To learn more, visit http://www.veeam.com.


AVAILABILITY for the Always-On Enterprise™
VEEAM
Veeam makes the Fortune 500 Available. 24.7.365
To enable its **Digital Transformation,** 70% of the Fortune 500 rely on Veeam to ensure Availability of all data and applications. **24.7.365**