PURESTORAGE®

TECHNICAL WHITE PAPER

Rapid Backup and Restore with FlashBlade and Veeam Backup & Replication

High-performance, modern data protection for mission critical environments.

Contents

Introduction

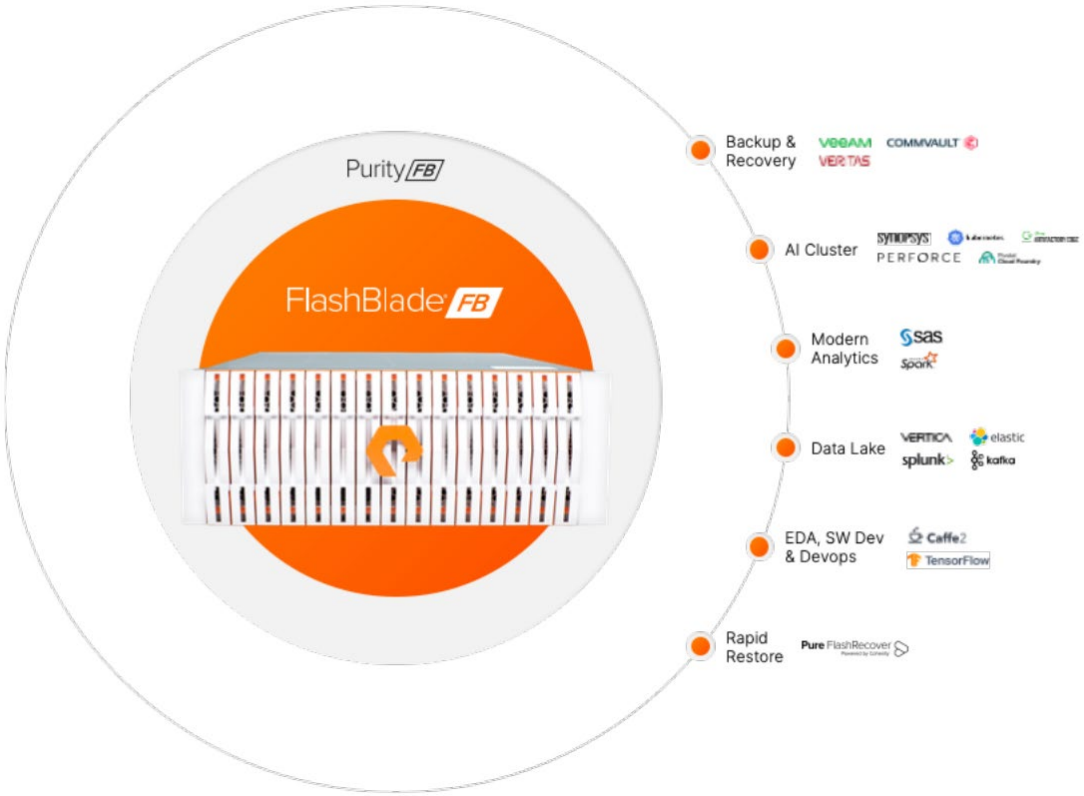
The purpose of this white paper is to share best practices and expected performance results for data backup and restore when using Veeam Backup & Replication version 11 with Pure Storage® FlashBlade®. Pure Storage FlashBlade

FlashBlade is an all-flash, scale-out storage solution that is powered by a distributed file system designed for enormous concurrency across all data types. FlashBlade is a unified fast file and object (UFFO) storage technology that can support thousands of clients while simultaneously hosting multiple file systems and multi-tenant object stores. By simply adding a single blade at a time, up to 150 blades, you can expand FlashBlade to multi-petabyte capacity with linear-scale performance. It is regarded as a data hub because of its inherent scale-out design and ability to drive performance for any type of task. It allows companies to consolidate a variety of workloads on a single platform, from backup to analytics and AI.

Figure 1: FlashBlade Unified Fast File and Object storage platform

Many organizations use FlashBlade to store their data protection backups, taking advantage of fast backup and restore performance while investing in a platform that also consolidates data lakes and other data silos.

Six key innovations underpin a FlashBlade system's ability to scale performance and capacity:



- **High-performance storage**: FlashBlade maximizes the advantages of an all-flash architecture by storing data in storage units instead of crippling, high-latency media, such as traditional spinning disks and conventional solid-state drives. The integration of scalable NVRAM into each storage unit helps scale performance and capacity proportionally when new blades are added to a system.
- **Unified network**: A FlashBlade system consolidates high communication traffic between clients and internal administrative hosts into a single, reliable high-performing network that supports both IPv4 and IPv6 client access over Ethernet links up to 100GB/s.
- **Purity//FB storage operating system**: A symmetrical operating system running on the FlashBlade fabric modules. It minimizes workload balancing problems by distributing all client operation requests among the blades on FlashBlade.
- **Common media architectural design for files and objects**: The single underlying media architecture of FlashBlade natively supports concurrent access to files via a variety of protocols, such as NFSv4.1, NFS over HTTP, and SMB and objects via S3 across the entire FlashBlade configuration.
- **High performance file and object replication**: FlashBlade file and object replication capabilities enable disaster recovery from a secondary site or the public cloud.
- **Simple usability**: FlashBlade alleviates system management headaches as it simplifies storage operations by performing routine administrative tasks autonomously. With a robust operating system, FlashBlade is capable of self-tuning and providing system alerts when components fail.

As of the writing of this paper, a fully configured FlashBlade system consists of up to 10 self-contained rack-mounted chassis interconnected by high-speed links to up to four external fabric modules (XFM). At the rear of each chassis, two on-board fabric modules provide high-speed Ethernet interconnects to the blades, other chassis, and clients using TCP/IP. Fabric modules are interconnected, and each contains a control processor and Ethernet switch ASIC. For reliability, each chassis is equipped with redundant power supplies and cooling fans.

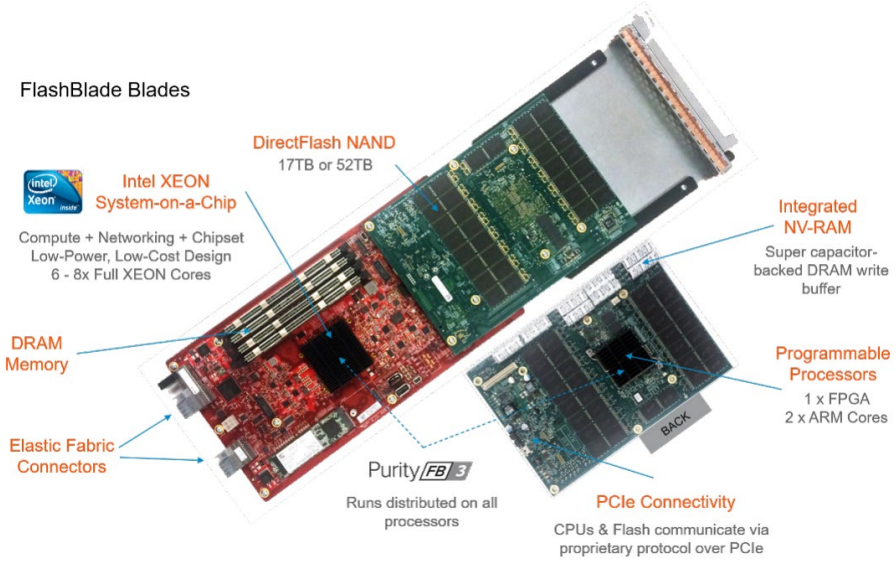


Figure 2: Anatomy of a FlashBlade

FlashBlade is a scale-out storage solution that intersects all three dimensions of fast, big, and simple. FlashBlade delivers unprecedented performance in a small form factor. It is tuned to deliver multi-dimensional performance for any data size, structure, or access. And it delivers significant savings in power, space, and cooling compared to legacy solutions.

Scalability	Performance	Connectivity	Physical
Start with 7 blades and simply add more to scale up to 150	Up to 15GB/s bandwidth with 15 blades in a single chassis	8x 40Gb/s or 32x 10Gb/s Ethernet ports/chassis	4U per chassis
Each blade adds capacity and performance	Up to 24M NFS IOPS in a single cluster with 150 blades	2x FlashBlade External Fabric Modules (XFM) to scale up to 150 blades	1,800 watts per chassis (nominal at full configuration)

Table 1: Pure FlashBlade highlights

Purity FlashBlade OS

Purity/FB, the operating software of FlashBlade,, is the heart of the system, allowing it to grow massively in terms of capacity and speed. Purity/FB has native support for both file and object protocols, allowing for high performance file and object storage on a single platform that can handle mixed workloads at the same time. It contains a variable block metadata engine and a scale-out metadata architecture, and it was built from the ground up for flash. Purity/FB is capable of handling billions of files and objects while offering unrivaled performance for every workload and file size, whether sequential or random access is used. It supports object and file replication on premises, as well as offers cloud integration for disaster recovery. Purity/FB was built from the ground up to provide UFFO storage.

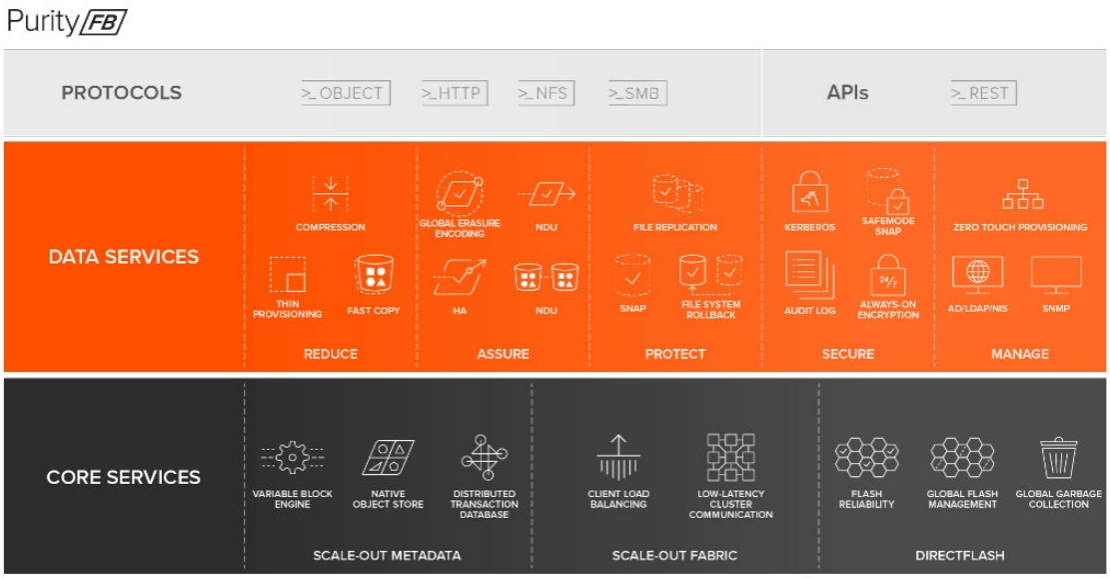


Figure 3: Purity/FB capabilities

Throughout this document, we will use terms related to the solution and its components. To avoid ambiguity, we will list them here:

- **VBR backup server**: The Veeam Backup & Replication server (VBR) performs main management operations; coordinates backup, replication and restore tasks; and controls job scheduling and resource allocation.
- **VBR backup proxy**: This Veeam Backup & Replication server component retrieves data from the source host datastores, processes it, and transfers to the backup repository.
- **VBR backup repository**: The software front-end to the storage target where Veeam Backup & Replication keeps backup files, backup copies, and metadata of replicated VMs.
- **VBR scale-out backup repository (SOBR)**: A repository system with horizontal scaling capabilities for multi-tiered data storage. A scale-out backup repository consists of one or more backup repositories referred to as performance tiers, which require flash-based performance storage, such as Pure FlashBlade. Object storage repositories can be added to provide long-term storage or archive storage. These repositories are called capacity tiers and archive tiers, respectively. A scale-out backup repository aggregates storage devices and systems into a single system and summarizes their capacities.
- **VBR backup extent**: When a backup repository is added to a scale-out backup repository, it is named an extent. Extents are the logical representations of underlying storage tier and named as such, like as performance tier, capacity tier, or archive tier.
- **Pure FlashBlade File System**: A logical container of user defined provisioned capacity. A Pure FlashBlade File System with an enabled protocol like NFS, or SMB, represents an available NAS share to the end user.
- **Pure FlashBlade data VIP**: An IPv4, or IPv6 virtual IP (VIP) address designated for only protocol data, like NFS or SMB. Exporting a FlashBlade file system requires a data VIP. Data VIPs are balanced across all FlashBlade physical interfaces.

Solution Overview

Coupled with Veeam Backup & Replication (VBR), Pure Storage FlashBlade delivers on the promise of cloud-like simplicity and agility with consistent high performance, and predictable low recovery time objectives (RTO) and recovery point objectives (RPO) for backing up the most critical workloads.

This technical white paper is for IT, storage, and backup specialists. It provides guidelines and techniques to optimize backup and restore performance when using the FlashBlade as a Veeam Backup Repository target. Figure 4 below illustrates the solution overview and key components.

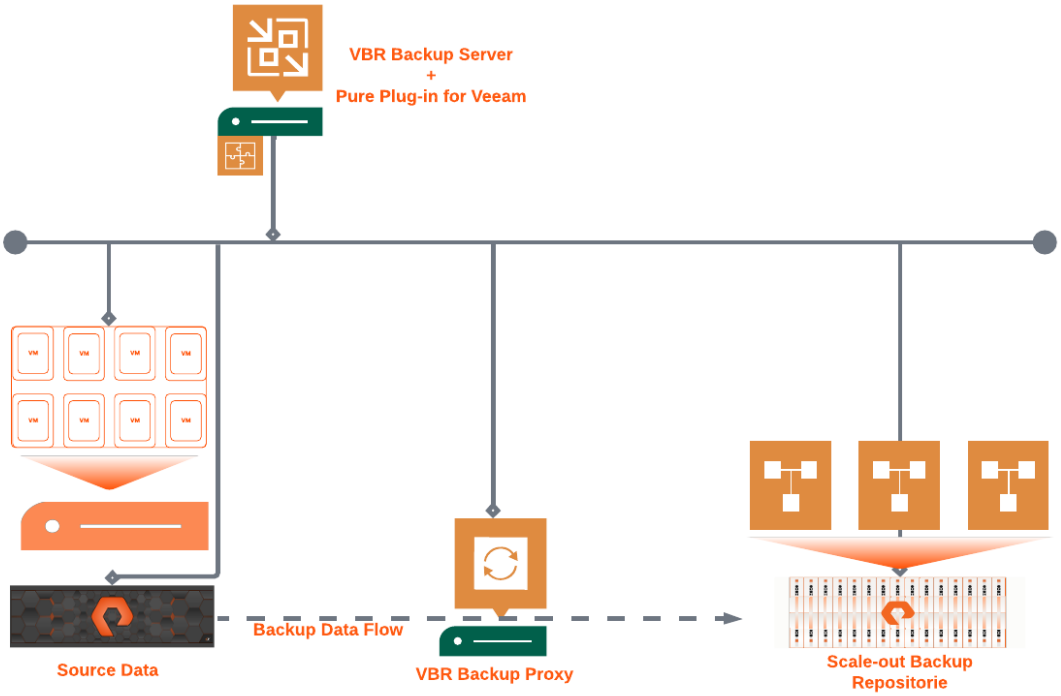


Figure 4: Veeam with Pure FlashBlade solution

Test Environment

This white paper examines Veeam Backup & Replication version 11, and describes best practice configuration when using FlashBlade as a NAS (network attached storage) backup repository. Therefore, the test environment was designed to measure performance differences between protocols (NFS vs. SMB) in conjunction with different scale-out repository configurations and other VBR options.

Compute Details

Veeam Backup & Replication infrastructure environment was built on physical compute to eliminate any workload interference with the VMWare environment which is used as the source for the test backup data. We used two VBR Backup Proxy servers that were also deployed on physical Compute. A single VMWare ESXi Compute was dedicated for this test within the vCenter Cluster environment.

Table 2 details the specification for each component in the test.

Component	Processor	RAM	Ethernet	Fibre Channel	Operating System
VBR backup server	2 x Intel Xeon E5-26700 @ 2.60 GHz. 64 cores total with HyperThreading enabled	256GB	2 x Mellanox MT27500 family network adapter @ 40 Gbps, in LACP team		Microsoft Windows Server 2022, Build 20348.169.210806
VBR backup proxy	AMD EPYC 7713P 64-Core	64GB	2 x Broadcom NetXtreme @ 25 Gbps, in LACP team	2X Emulex LPe35002-M2-D, 32Gbps. Native MPIO enabled	Microsoft Windows Server 2022, Build 20348.169.210806
VMWare ESXi	AMD EPYC 7713P 64-Core	64GB	2 x Broadcom NetXtreme @ 25 Gbps, in LACP team	2X Emulex LPe35002-M2-D, 32Gbps. Native MPIO enabled	Microsoft Windows Server 2022, Build 20348.169.210806

Table 2: Compute Environment details

VBR Scale-out Backup Repositories Configuration

We tested various configurations of scale-out backup repositories to determine which settings provide the best performance with respect to the protocol in use. The criteria for comparison were:

- Single versus multiple extents on both single and multiple filesystems on the FlashBlade
- Single versus multiple data VIPs (FlashBlade virtual network interfaces)

Table 3 illustrates the different Scale-out Backup Repositories configured for the tests.

Use Case	VIPs	# Extents	No. of FlashBlade NAS Shares
SOBR 01	1	1	1
SOBR 02	4	4	1
SOBR 03	1*	4	4
SOBR 04	4	4	4

*The same VIP was presented four times to simulate multiple extents. The extents were simply different folders on the same NAS share.

Table 3: VBR scale-out backup repositories configurations used for testing.

FlashBlade Configuration

Veeam utilizes FlashBlade as a scale-out backup storage target via NFS and SMB protocols. The FlashBlade system used in testing was configured as shown in Table 4.

Component	Description
Number of blades	15 x 17TB blades
Capacity	240TB raw, (162.46TB usable)
Connectivity	4 x 40Gb/s Ethernet (data) 2 x 1Gb/s Redundant Ethernet (management port)
Operating platform	Purity//FB 3.2.1

Table 4: FlashBlade configuration used in testing

Connectivity

VMWare datastores were presented to ESXi on the Fibre channel SAN and were visible only to VBR backup proxy servers. Pure FlashBlade data network interfaces, and VBR backup proxy servers were isolated on the same VLAN. Management traffic between VBR management server, proxies, VMWare vCenter, FlashArray//M70s, and FlashBlade management interfaces, in addition to Active Directory traffic, were on a separate VLAN.

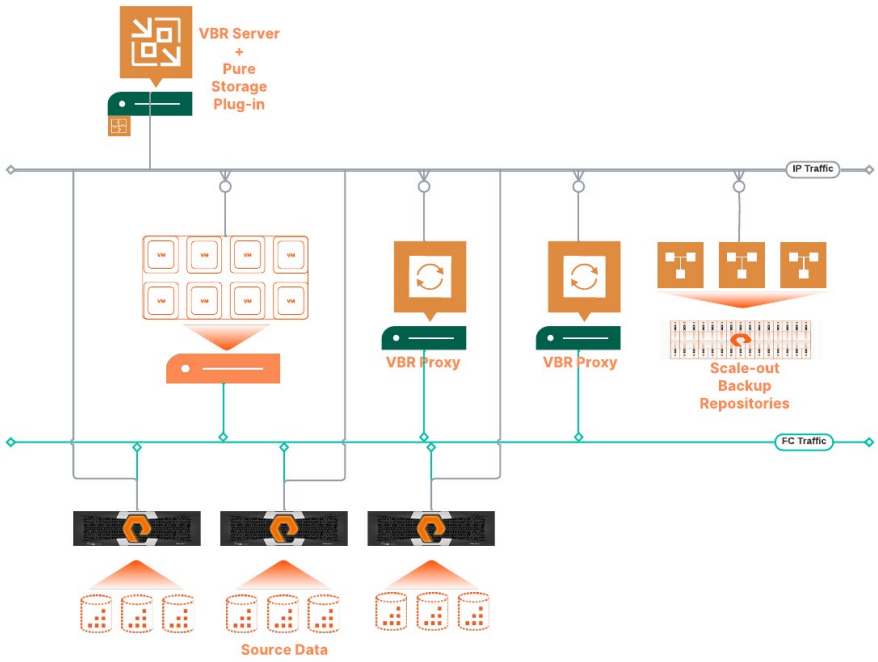


Figure 5: Testing environment connectivity

Test Data

Table 5 details the test and the underlying storage layout.

Component	Description
Test data source	96 virtual machines, with Windows 10
VMDK per VM	100GB
OS space consumed	10GB
Test data consumed space	90GB (10% compressible)
Storage array	3X Pure FlashArray//M70.
Number of datastores	9

Table 5: Backup data source details

Test Procedure and Workflow

The test used the VMs described earlier to perform a series of active full backups against each scale-out repository and measure the average processing rate achieved. Direct SAN with storage integration is the default configuration. Figure 6 illustrates the high-level workflow of each backup job.

Figure 6: High-level backup workflow

During restores, a series of full VM restores we performed to measure average data transfer rate for the entire restore job. Direct SAN mode has been the preferred method of restore due to its speed. Figure 7 describes the high-level restore workflow.

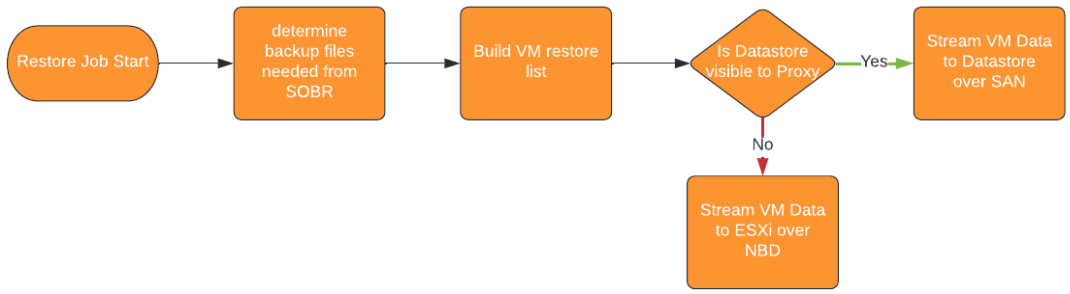


Figure 7: High-level restore workflow

Veeam Backup & Replication 11 Configuration

The Veeam Backup & Replication management and backup proxy servers were hosted on physical Windows 2022. Only the backup proxy servers had access to Fibre channel SAN where VMWare datastores were presented to them. The management server and proxies had access to the management and backup data VLANs.

Pure FlashBlade enables hundreds of thousands of IOPs available to VBR with low latency; various variables in the Veeam Backup & Replication 11 console may be modified to enable greater levels of throughput, allowing backup tasks to be finished significantly quicker than if they were left at default values.

VBR Backup Proxy Settings (VMWare Agent)

A backup proxy is an architecture component that sits between the backup management server and other components of the backup infrastructure; it processes jobs and delivers backup traffic. Below are two of the settings we used.

Transport mode: This determines how the backup proxy will transfer data between the production environment and the backup environment. With the Pure Storage plugin installed and the VMWare data stores visible to the backup proxy, direct SAN access can be performed. In direct SAN transport mode, VBR leverages VMware VADP to transport VM data directly from—and to—FC storage over the SAN, bypassing ESXi hosts and the LAN. Direct SAN access offers the fastest data transfer speed and does not place a significant load on the production network. Direct Storage Access had been selected for our test environment.

Max concurrent tasks: This setting is based on the available CPU cores; we had it set to **64**.

Backup Repository Settings

Veeam Backup & Replication stores backup files in a backup repository. With Pure FlashBlade serving as the storage target, VBR can take advantage of the high throughput, low latency, and parallelism provided by Pure FlashBlade. Below are some of the settings we used.

Limit maximum concurrent tasks: The default value is 4 streams. We unchecked this option to allow for a maximum number of streams to be sent to the FlashBlade. This option is highly dependent on the capability of your infrastructure. To determining the appropriate setting we strongly recommend you conduct a thorough examination of your networking and compute capabilities.

Use per machine backup files: This setting is located under the **Advanced** button. It provides greater parallelism and makes use of the enterprise grade Pure FlashBlade performance capabilities. This option enables a single backup stream per VM, therefore we checked this option.[1]

Backup Job Settings

In Veeam Backup & Replication, back up is a job-driven process. To perform the backup, you need to configure a backup job. When we created the backup jobs, to test the different scale-out backup repositories we left all default storage settings at default values. Therefore we enabled deduplication and optimal compression settings (the default settings).

Performance Results

It's important to denote the performance results mentioned here are infrastructure specific; the testing environment was used to identify performance characteristics rather than maximizing performance of the solution. Modern production environments will likely show better performance results. Please consider using the test results as guidelines, and the settings mentioned on this paper as best practices.

NFS vs. SMB

The purpose of this test was to answer questions concerning the scale-out repository configuration and whether SMB (v2.1) protocol or NFS (v3) protocol is superior for VBR and FlashBlade backup solution. We only focused on the scale-out backup repository type due to its seamless horizontal scaling, which we would recommend for the production environment. Because of this, even when only a single backup repository is configured, it was still added to a scale-out backup repository.

Active-full backups only were used for this test. Figure 8 shows in terms of backup workload NFS is a superior protocol. This is because the VBR NFS client used separate TCP connections per VM whereas the SMB client utilized connection trunking. Connection trunking appeared to be based on the number of FlashBlade data VIPs IP interfaces, and the number of backup proxy servers. SMB performance was noticeably better (SOBR 02), when using multiple data VIPs, while maintaining a single FlashBlade file system share.

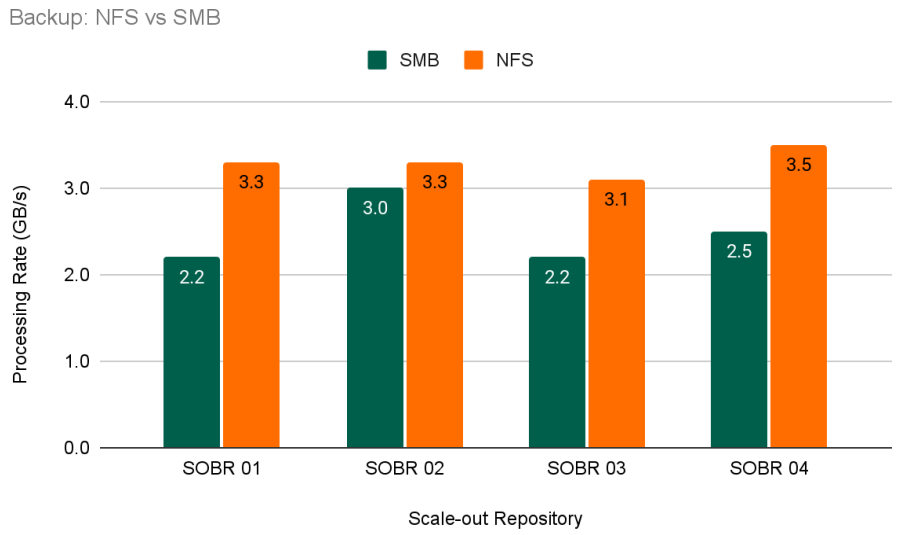


Figure 8: Backup Performance NFS vs. SMB

1 For scale-out backup repository, the default is to have this option checked; therefore it would be unnecessary to have this option checked on a traditional backup repository that is part of a scale-out repository.

Table 6 depicts the number of TCP connections that were utilized during the backup, along with the individual scale-out repository configuration

Component	SOBR 01	SOBR 02	SOBR 03	SOBR 04	Component
FlashBlade data VIPs	1	4	1	4	FlashBlade data VIPs
FlashBlade shares (file systems)	1	1	4	4	FlashBlade shares (file systems)
VBR proxies	2	2	2	2	VBR proxies
SMB connection	2	8	2	8	SMB connection
NFS connection	96	96	96	96	NFS connection

Table 6: NFS vs. SMB backup streams

Similarly, during restore tests, NFS again exhibited superior performance. It was evident, however, that the four FlashBlade file system shares boosted the performance of SMB (SOBER 04), to a degree similar to that of NFS. Figure 9 shows the restore performance using the same scale-out repositories as those used for backup testing.

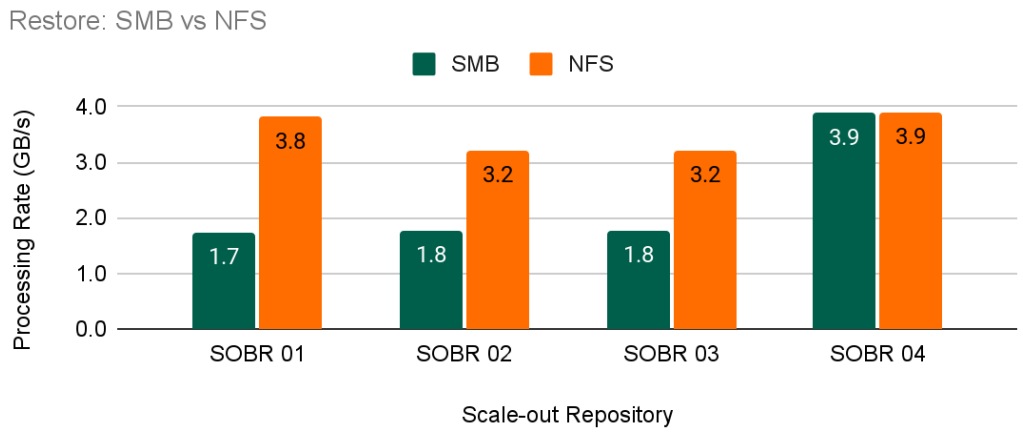


Figure 9: Restore Performance NFS vs. SMB

Single vs. Multiple Extents

Setting up a Veeam Backup & Replication scale out backup repository entails adding extents. In Veeam, extents are simply represented by folders, typically pointing to additional capacity with some performance characteristics tied to them.

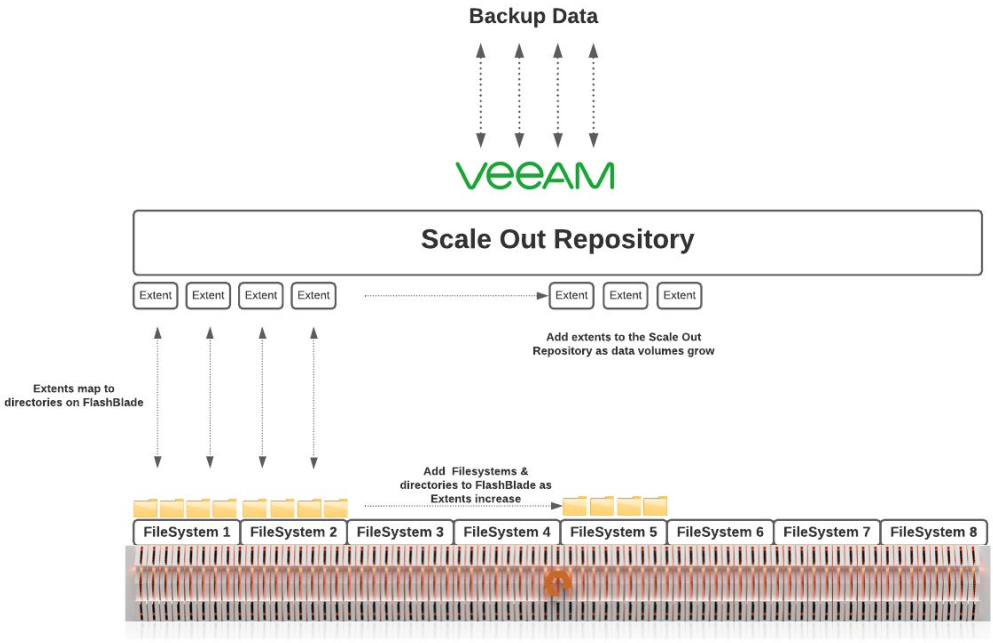


Figure 10: Single vs. multiple extents

The goal of this test was to determine if using multiple FlashBlade file system shares would increase backup throughput and would be superior to creating multiple folders (extents) on the same file system share. As with other tests we used our two VBR proxy Servers. Table 3 details the SOBR configurations used during these tests. Figure 11 shows the difference between SOBR 01, and SOBR 02, both reflecting single extent , and multiple extents respectively on a single file system share. There was no difference in performance between using a single extent versus using multiple extents when the underlying configuration is a single file system share.

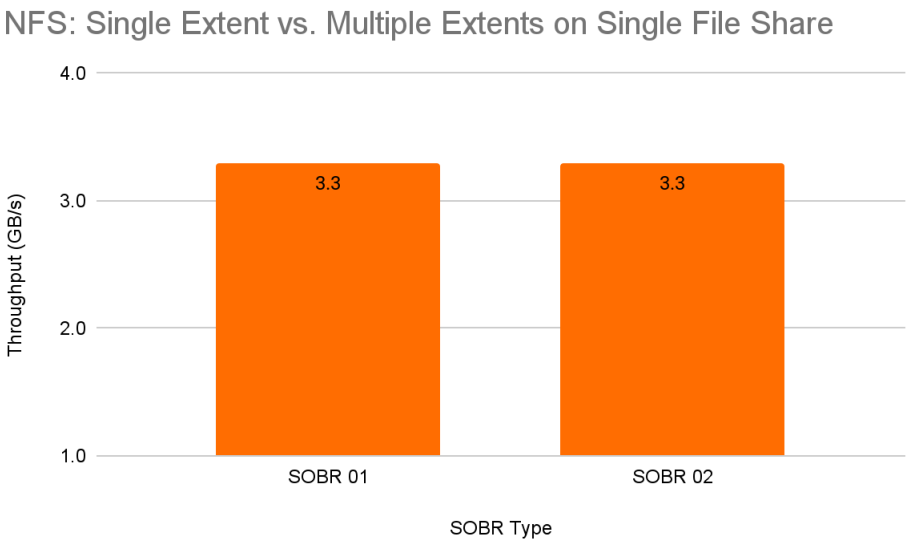


Figure 11: Scale-out repository with single, vs. multiple extents over single file system

During the next test we've examined SOBR 01, to SOBR 04 which had multiple extents (each extent on its own FlashBlade file system share), and we were able to drive a little more throughput. Figure 11, shows the testing results. We've noticed backup processing rate improvements, however it's not significant enough to warrant the complexity introduced with multiple extents.

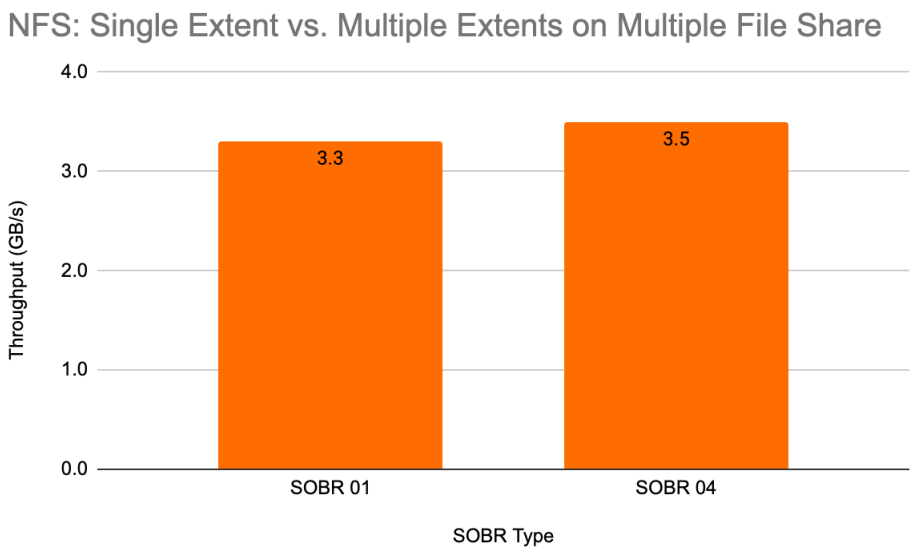


Figure 12: Scale-out repository with single, vs. Multiple Extents over multiple file systems

Synthetic Full Performance

An active full backup retrieves all the VM data from the source datastores but is resource intensive and consumes a large amount of network bandwidth. Synthetic full backups create identical backup files as active full backups. A synthetic full backup creates a VBK file which contains the data of the entire virtual machine. In contrast to active full backups, Veeam Backup & Replication does not retrieve VM data from the source datastore. Instead, it combines the incremental backups and full backups in the backup repository to form a new full backup.

We compared the performance impact of creating a synthetic full backup on the FlashBlade versus creating an active full backup based on the protocol used. As illustrated in Figure 13, the FlashBlade has similar performance utilization characteristics to an active full backup workload

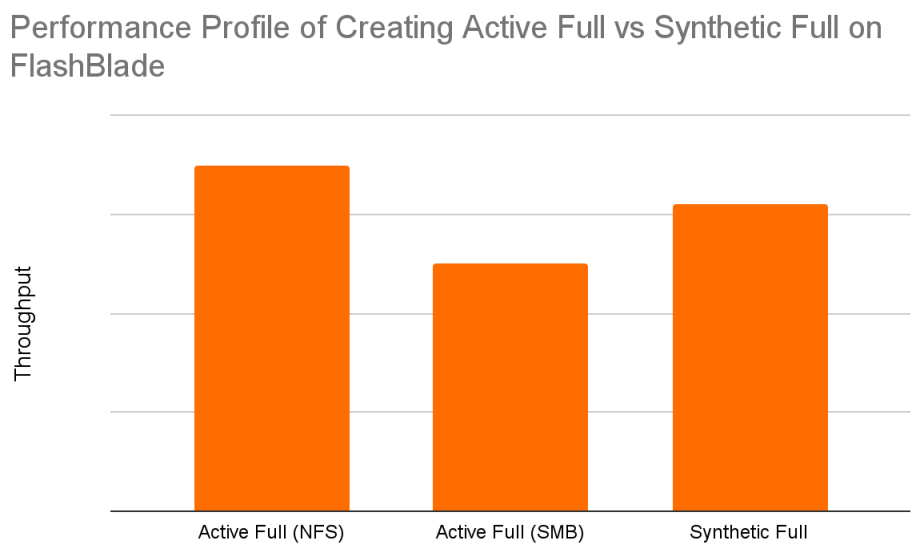


Figure 13: Active vs. synthetic backup workload profile

Given the observed workload profile on the FlashBlade, we recommend considering synthetic full backup as a viable alternative to active backup only when the backup schedule is known and controlled in such a way as not to adversely affect other backup jobs, or tasks.

Compression Settings

Veeam Backup & Replication complements Pure FlashBlade in terms of capacity savings. Deduplication and compression are provided as part of Veeam Backup & Replication by default. Pure FlashBlade provides global compression as a standard feature, significantly enhancing the capacity-saving potential across multiple backup repositories.

Veeam Backup & Replication provides a repository feature called “Decompress backup file data blocks before storing” when enabled (the default is disabled). Veeam Backup & Replication will decompress the backup files before writing to target storage. Our virtual environment (10% compressible) did not benefit when enabling this option.

Our recommendation is to leave Veeam Backup & Replication data reduction options at default values. Enabling “Decompress file data blocks before storing” will (potentially) increase CPU utilization on VBR proxies, but may also increase capacity saving on the FlashBlade. Careful consideration should be given to understand the additional benefits at the cost of more CPU cycles.

Ransomware Protection with SafeMode

Pure Storage SafeMode™ provides built-in protection of your data in the event of a ransomware attack by frequently protecting Veeam Backup & Replication Backup repositories with read-only snapshots from which you can recover your data. SafeMode helps secure backup data since snapshots can’t be modified, deleted, or encrypted even if administrative credentials have been compromised.

Veeam Backup & Replication stores backups on disk using a self-contained file-based approach. Metadata needed for recovery is included with the backup files. In the event a recovery is required, all metadata needed for a successful recovery is available within the snapshot. There are a couple of ways to recover Veeam Backup & Replication backups with SafeMode:

- **Method #1**: Direct "restore" of the snapshot, which will revert the content of the NAS share back to the time of when the snapshot was taken. This method will overwrite the contents of the NAS share storing the Veeam Backup Repository files.
- **Method #2**: Create a repository using the SafeMode snapshot itself. In this case, Veeam Backup & Replication will import the backup metadata into its database by scanning the backups contained in the snapshots. This method preserves the original repository content for forensic analysis. Keep in mind that Veeam Backup & Replication won't allow you to add an NFS read-only snapshot. The work around is use the SMB protocol to mount the NFS snapshot. For example, if the source repository of the snapshot is NFS, then you will need to use SMB protocol when creating the recovery repository from its snapshot. You should create a new scale-out repository for new backups.

Conclusion

Veeam Backup & Replication and Pure Storage FlashBlade deliver the entire suite of data performance, protection, and data mobility. As data capacity requirements grow, FlashBlade and Veeam Backup & Replication can be non-disruptively scaled and upgraded transparently without impact to end user operations.

FlashBlade with Purity/FB dynamically tuning itself to the workload, unlocks the main benefits of Veeam Backup & Replication Scale-Out Backup Repository: flexible capacity management, practically unlimited cloud-based storage capacity, and predictable performance.

Protocol tests showed that NFS consistently performed better than SMB, regardless of the VBR scale-out backup repository configuration. The scale-out backup repository configuration tests revealed a single extent from a single FlashBlade file system share is marginally less performant than that of using multiple extents from multiple FlashBlade file system shares, however the difference was not substantial. As the environment grows, we recommend adopting a simpler configuration that uses a single extent from a single FlashBlade filesystem share. As the capacity requirements grow, simply expand the FlashBlade file system share.

Synthetic full backups do drive significant workload on the FlashBlade. For environments that can't endure regular active full backups, then synthetic full backups are probably your only option. Our recommendation is to adopt a single approach to full backups; if synthetic full backups are the least impactful, they should be the standard for creating full backups—and of course should be scheduled during low business activities.

Veeam Backup & Replication comes with deduplication, and optimal compression turned on by default. Pure FlashBlade provides global compression, and auto tuning based on the workload. Our recommendation is to leave default values intact. Test showed that decompressing backup files before writing to the backup repository option drives backup proxies to higher CPU utilization, with little benefits to overall data reduction.

Finally, we recommend enabling SafeMode, as it is the simplest and most effective method to protect Veeam Backup & Replication repositories from accidental deletion, or intentional ransomware attack.

About the Author

Tamer Swidan is a senior solution architect with Pure Storage. He is responsible for defining Pure Storage solutions and reference architectures for protecting and recovering primary workloads such as Oracle, SQL, and VMware. Tamer has 19 years' experience working in and with data protection hardware and software solutions, serving as an end user, a subject matter expert consultant, and data protection solution architect. Tamer joined Pure Storage in October 2020.

Pure Storage, Inc.
650 Castro Street, #400
Mountain View, CA 94041