Dell PowerScale: Solution Design and Considerations for SMB Environments

Abstract

This white paper provides technical details on the design considerations of Dell PowerScale storage and the OneFS operating system with Microsoft Server Message Block (SMB) workloads.

January 2023

Revisions

Date	Description
December 2018	Initial release
April 2019	Updated with new template; added content about SMB performance dataset monitoring introduced on OneFS 8.2.0
December 2019	Updated content about performance dataset monitoring on OneFS 8.2.2
June 2020	PowerScale rebranding
November 2022	Change the dynamic IP and static IP consideration for SMB in SmartConnect
January 2023	Add SmartQoS feature for SMB

Acknowledgments

Table of contents

Executive summary

This white paper provides considerations and common practices to help network and storage architects and administrators plan, configure, monitor, and manage Microsoft Server Message Block (SMB) configuration and design with Dell PowerScale products. This document covers the following topics:

- SMB management design and considerations on the PowerScale cluster
- SMB networking design and considerations on the PowerScale cluster
- SMB security and access control configurations on the PowerScale cluster
- SMB performance tuning and troubleshooting on the PowerScale cluster
- SMB feature introduction, considerations, and performance test results

Audience

The guide is intended for experienced system and storage administrators who are familiar with file services and network storage administration.

The guide assumes the reader has a working knowledge of the following:

- Network-attached storage (NAS) systems
- The SMB storage protocol
- The PowerScale scale-out storage architecture and the PowerScale OneFS operating system
- File-system management concepts including provision and permission
- Integration practices for connecting and establishing authentication relationships with Microsoft Active Directory

For more information about the topics discussed in this paper, see the following publications:

- [PowerScale OneFS Technical Overview](#)
- [PowerScale OneFS Web Administration Guide](#)
- [PowerScale OneFS CLI Administration Guide](#)
- [Current PowerScale Software Releases](#)
- [OneFS Security Configuration Guide](#)

1 SMB design considerations and common practices

1.1 SMB protocol introduction

The SMB protocol is a network file sharing protocol, and as implemented in Microsoft Windows is known as the Microsoft SMB protocol. The set of message packets that defines a particular version of the protocol is called a dialect. The Common Internet File System (CIFS) protocol is a dialect of SMB. Both SMB and CIFS are also available on virtual machines, and several versions of UNIX and Windows operating systems.

The Microsoft SMB protocol is a client-server implementation and consists of a set of data packets, each containing a request sent by the client or a response sent by the server. These packets can be broadly classified as follows:

- Session control packets: Establishes and discontinues a connection to shared server resources.
- File access packets: Accesses and manipulates files and directories on the remote server.
- General message packets: Sends data to print queues, mailslots, and named pipes, and provides data about the status of print queues.

For more detailed information about the SMB protocol, see the Microsoft TechNet article Microsoft SMB Protocol and CIFS Protocol Overview.

Table 1 lists different versions of SMB supported by Windows operating systems. Table 2 lists SMB features supported by PowerScale OneFS versions.

Table 1   SMB version supported by Windows Operation System

Version	Supported Windows operating systems
SMB 1.0 (or SMB1)	- Windows 2000 - Windows XP - Windows Server 2003 - Windows Server 2003 R2
SMB 2.0 (or SMB2)	- Windows Vista (SP1 or later) - Windows Server 2008
SMB 2.1 (or SMB2.1)	- Windows 7 - Windows Server 2008 R2
SMB 3.0 (or SMB3)	- Windows 8 - Windows Server 2012
SMB 3.02 (or SMB3)	- Windows 8.1 - Windows Server 2012R2
SMB 3.1.1 (or SMB3)	- Windows 10 - Windows Server 2016

Table 2   SMB features supported by PowerScale OneFS versions

SMB feature	Supported OneFS versions	Section in this document
SMB multichannel	PowerScale OneFS 7.1.1 and later	SMB multichannel
SMB symbolic links	PowerScale OneFS 7.1.1 and later	SMB symbolic links
SMB server-side copy	PowerScale OneFS 8.0 and later	SMB server-side copy
SMB file filtering	PowerScale OneFS 8.0 and later	SMB file filtering
SMB continuous availability	PowerScale OneFS 8.0 and later	SMB continuous availability and witnessSMB continuous availability and witness
SMB encryption	PowerScale OneFS 8.1.1 and later	SMB encryption

1.2   Networking

In a scale-out NAS environment, the overall network architecture must be configured to maximize the user experience. Many factors contribute to overall network performance. The following sections list some considerations of jumbo frames, link aggregation, and SmartConnect that benefit the user experience on PowerScale systems. For other general design consideration, see the white paper PowerScale Network Design Considerations.

1.2.1   Jumbo frames

Jumbo frames are Ethernet frames where the maximum transmission unit (MTU) is greater than the standard 1500 bytes and could be up to 9000 bytes. The larger MTU size provides greater efficiency as less overhead and fewer acknowledgments are sent across devices, drastically reducing interrupt load on endpoints. However, this is not applicable to all workloads.

To take advantage of the greater efficiencies, jumbo frames must be enabled end-to-end on all hops between endpoints. Otherwise, the MTU could be lowered through Path MTU Discovery (PMTUD) or packets could be fragmented. The fragmentation and reassembly impacts the CPU performance of each hop, which impacts the overall latency.

Here are some general considerations for MTU settings with different scenarios. While the general assumption is that jumbo frames provide performance advantages for all workloads, it is important to measure results in a lab environment simulating a specific workload to ensure performance enhancements.

- Generally speaking, MTU with default value (1500 bytes) often provides adequate performance.
- If a customer uses a 10 GbE network configuration, general guideline is to set 4500 or 9000 bytes for performance advantages and must be enabled end-to end on all hops between endpoints.
- If a customer uses 40 GbE network configuration, general guideline is to set 9000 bytes for performance advantages and must be enabled end-to-end on all hops to ensure the performance.

For detailed information about how to configure MTU, see the white paper PowerScale Network Design Considerations.

1.2.2   Link aggregation

Link aggregation protocol provides methods to combine multiple Ethernet interfaces, forming a single link layer interface, specific to a switch or server. It balances the network traffic leaving the aggregated interfaces.

It is imperative to understand that link aggregation is not a substitute for a higher bandwidth link. Although link aggregation combines multiple interfaces, applying it to multiply bandwidth by the number of interfaces for a single session is incorrect. Link aggregation distributes traffic across links. However, a single session only uses a single physical link to ensure packets are delivered in order without duplication of frames. Thus, the bandwidth for a single client is not increased, but the aggregate bandwidth of all clients increases in an active/active configuration.

Here are some considerations for using link aggregation:

- SMB2 will benefit from link aggregation as a failover mechanism between client network interface cards (NICs).
- No link aggregation configuration is required or desired for SMB3. SMB3 multichannel will automatically detect and use multiple network connections if a proper configuration is identified. For more detail about SMB multichannel, see the SMB Multichannel section.
- Link aggregation is only per PowerScale node, not across PowerScale nodes. Because OneFS is a clustered file system, each node of the cluster is an independent unit with its own operating system. Link aggregation across more than one node is not available or supported.

1.2.3   SmartConnect

SmartConnect enables client connection load balancing and dynamic failover and failback of client connections across storage nodes to provide optimal utilization of the cluster resources. SmartConnect eliminates the need to install client-side drivers, enabling the IT administrator to easily manage large numbers of clients with confidence. And in the event of a system failure, file system stability and availability are maintained. For more detailed information about SmartConnect, see the white paper PowerScale Network Design Considerations.

Here are some considerations for using SmartConnect with SMB:

- There is a server state associated with an SMB connection. An IP failover in a dynamic pool does not replicate that state. Users may counter incorrect behavior and possible file corruption if the application does not handle the server state properly. For failover consideration, it is recommended to use SMB Continuous Availability with PowerScale. For more detailed information about SMB Continuous Availability, see the SMB continuous availability and witness section.
- Static IP is recommended and sufficient for most use cases. SMB also works well with dynamic allocation of IP addresses. In certain workflows, SMB is preferred with dynamic allocation of IP addresses, because IP address consistency is required. Aside from a workflow requirement, it could also be an IT administrative dependence.
- With SmartConnect, if a node that has established client connections goes offline, the behavior is protocol-specific. For SMB workload, the SMB protocols are stateful. When a node that has established client connections goes offline, the SMB connection is broken because the state is lost.
- If a customer has a mixed workload with both SMB and NFS connections to the same PowerScale cluster, it is recommended to have a different SmartConnect zone and separate IP address pool for

each workload. The NFS clients can be put in a dedicated SmartConnect zone that will facilitate failover while the SMB clients are put into another SmartConnect zone that will not participate in failover. This will ensure the SMB clients mount to the "Static node IPs" which do not failover.

1.3 Access zones

Access zones provide a method to logically partition cluster access and allocate resources to self-contained units, thereby providing a shared tenant, or multi-tenant, environment. Access zones support configuration settings for authentication and identity management services on a cluster, so you can configure authentication providers and provision protocol directories such as SMB shares on a zone-by-zone basis. As a general common practice, reserve the System zone for configuration access, and create additional zones for data access. In the following sections, we will focus on the consideration for directory services, access zone separation and Kerberos authentication.

1.3.1 Directory service

An access zone can authenticate users with only one Active Directory domain. Although you can add more than one of the other directory services to a zone, a common practice is to limit each zone to no more than one of each of the directory services. For example, your access zone could contain one Active Directory, one LDAP and one File provider.

1.3.2 Access zones separation

OneFS supports overlapping data between access zones for cases where your workflows require shared data for consolidation consideration; however, this adds complexity to the access zone configuration that might lead to future issues with client access. As a general guideline, overlapping access zones should only occur if data must be shared between zones. If sharing data, it is recommended that the access zones share the same authentication providers. Shared providers ensure that users will have consistent identity information when accessing the same data through different access zones.

In case you cannot configure the same authentication providers for access zones with shared data, Dell Technologies recommends the following common practices:

- Select Active Directory as the authentication provider in each access zone. This causes files to store globally unique SIDs as the on-disk identity, eliminating the chance of users from different zones gaining access to each other's data.
- Set the on-disk identity to **native**, or preferably, to **sid**. When user mappings exist between Active Directory and UNIX users, OneFS stores SIDs as the on-disk identity instead of UIDs. For example, the NFS export in Access Zone A uses LDAP as the authentication provider, meanwhile the SMB share in Access Zone B uses NTLM as the authentication provider. The NFS exports and the SMB share in this example shares the same root data path. In this case, select **native** or **sid** as on-disk identity.

1.3.3 Kerberos authentication

Kerberos is a network authentication provider that works on the basis of tickets to allow communication over a non-secure network to prove their identity to one another in a secure manner. OneFS supports two kinds of Kerberos implementation on PowerScale clusters

- Microsoft Kerberos/Active Directory Kerberos/Microsoft Windows KDC
- MIT Kerberos/MIT KDC

If you configure an Active Directory provider, support for Microsoft Kerberos authentication is provided automatically. If your workflow requires using the SMB protocol, use Microsoft Kerberos.

For using Microsoft KDC/Kerberos with AD users and PowerScale, several considerations and recommendations are listed below:

- It is recommended to authenticate all users with Kerberos because it is a highly secure protocol and the performance is much better than NTLM.
- If you are authenticating users with Kerberos, make sure that both your PowerScale cluster and your clients use Active Directory and the same NTP server as their time source.
- Make sure you do not have duplicated SPNs created for the PowerScale cluster. This can cause an authentication issue.

See Kerberos Authentication and the white paper Integrating OneFS with Kerberos Environment for Protocols for more details about how to configure Kerberos on PowerScale.

1.4 Multi-protocol access

1.4.1 Overview

To provide multi-protocol access, access tokens are generated through the following steps:

1. User identity lookup
2. ID mapping
3. User mapping
4. On-disk identity calculation

The following sections focus on the considerations and common practices of step 2 through 4 and at the end will discuss some options in ACL policy settings.

1.4.2 ID mapping

The ID mapping service's role is designed to map Windows SIDs to UNIX UIDs and GIDs and vice versa in order to associate a user's identity across different directory services.

The followings are some considerations and recommendations for ID mapping:

Use Active Directory with RFC 2307

Use Microsoft Active Directory with RFC 2307 attributes to manage Linux, UNIX, and Windows systems and make sure your domain controllers are running Windows Server 2003R2 or later. Integrating UNIX and Linux systems with Active Directory centralizes identity management and eases interoperability.

If you use Microsoft Active Directory with RFC 2307 attributes to manage Linux, UNIX, and Windows systems, the following fields are required in Active Directory:

- uid
- uidNumber
- gidNumber
- loginShell
- UNIXHomeDirectory

For the detailed configuration steps, see OneFS: How to configure OneFS and Active Directory for RFC2307 compliance.

Do not use overlapping ID ranges

In networks with multiple identity sources, such as LDAP and Active Directory with RFC 2307 attributes, you should ensure that UID and GID ranges do not overlap. It is also important that the range from which OneFS automatically allocates UIDs and GIDs does not overlap with any other ID range. OneFS automatically allocates UIDs and GIDs from the range 1,000,000-2,000,000 which is configurable. If UIDs and GIDs overlap multiple directory services, some users might gain access to other users' directories and files. A typical scenario is when LDAP provides extended AD attributes with a 1000,000+ UID or GID, and this overlap with the one generated on OneFS.

See ID mapping ranges for more information about how to configure ID ranges.

Avoid common UIDs and GIDs

Do not include commonly used UIDs and GIDs in your ID ranges. For example, UIDs and GIDs below 1,000 are reserved for system accounts; do not assign them to users or groups.

1.4.3   User mapping

The user mapping service combines access tokens from different directory services into a single token. When the names of an account in different directory services match exactly, OneFS automatically combines their access tokens into a single token. On the other hand, you can set rule to modify the user mapping behavior.

You can also test the user mapping rule. For more details, see Test a user-mapping rule.

The followings are some considerations and recommendations for user mapping:

Employ a consistent username strategy

In an environment with two or more identity management systems, the simplest configuration is naming users consistently, so that each UNIX user corresponds to a similarly named Windows user. Before assigning a UID and GID, OneFS searches its other authentication providers, such as LDAP, for other identities with the same name. If OneFS finds a match, the mapping service by default selects the associated UID and group memberships. Naming users consistently also allows user mapping rules with wildcards to match names and map them without explicitly specifying each pair of accounts.

Avoid using UPNs in mapping rules

You cannot use a User Principal Name (UPN) in a user mapping rule. A UPN is an Active Directory domain and username that are combined into an Internet-style name with an @ symbol, such as an email address: jane@example. If you include a UPN in a rule, the mapping service ignores it and may return an error. Instead, specify names in the format DOMAIN\user.com.

Native on-disk identity

The **native** identity option is likely to be the best for a network with UNIX and Windows systems and this is the default setting. In native mode, OneFS favors setting the UID as the on-disk identity because doing so improves NFS performance. OneFS stores only one type of identifier—either a UID and a GID or a SID—on disk at a time. As a common practice, if you change the on-disk identity, you should run the repair permissions job; see the OneFS Administration Guide.

See [PowerScale Multiprotocol Data Access with a Unified Security Model](#) for details of on-disk identity.

1.4.4   ACL policies

A PowerScale cluster includes ACL policies that control how permissions are processed and managed. In the following section, we will explore several options to manage ACL policy manually in the environment.

ACL policies for different environment

For UNIX, Windows, or mixed (Windows + UNIX) environments, optimal permission policy settings are already selected. It is recommended that one of these pre-defined environment templates be used for most workflows. In some cases, one or more of the policy settings might need to be modified. Any modification to the default policy settings will automatically create a **Custom environment** ACL policy.

You can choose to manually configure ACL policies by selecting Custom environment. The following sections discuss the options in detail.

Chmod command on files with existing ACLs

There are six policy options to configure how OneFS processes a chmod command run on a file with an ACL. The option to merge the new permissions with the ACL is the recommended and defaulted approach because it best balances the preservation of security with the expectations of users. However, you can choose to manually configure this option if necessary to support your particular environment.

Table 3 lists the consideration for all the six options you can choose.

Table 3   Options for configuring how OneFS processes chmod

Settings	Consideration
Merge the new permissions with the existing ACL	This is the default and recommended option.
Remove the existing ACL and set UNIX permissions instead	This option can cause information from ACLs, such as the right to write a DACL to a file, to be lost, resulting in a behavior that gives precedence to the last person who changed the mode of a file. As a result, the expectations of other users might go unfulfilled. Moreover, in an environment governed by compliance regulations, you could forfeit the rich information in the ACL, such as access control entries for allowing or denying access to specific groups, resulting in settings that might violate your compliance thresholds.
Remove the existing ACL and create an ACL equivalent to the UNIX permissions	This option can have the same effect as removing the ACL and setting UNIX permissions instead: Important security information that is stored in the original ACL can be lost, potentially leading to security or compliance violations.
Remove the existing ACL and create an ACL equivalent to the UNIX permissions for all the users and groups referenced in the old ACL	This option improves matters over the first two settings because it preserves the access of all the groups and users who were listed in the ACL.
Deny permission to modify the ACL	This option can result in unexpected behavior for users who are owners of the file and expect to be able to change its permissions.

Ignore the operation if file has an existing ACL	This is very similar to the previous setting which results inability to change the permission through chmod. Select this option if you defined an inheritable ACL on a directory and want to use that ACL for permissions.

This setting is located under **ACL Policy Settings** > **General ACL Settings**, as shown in Figure 1.
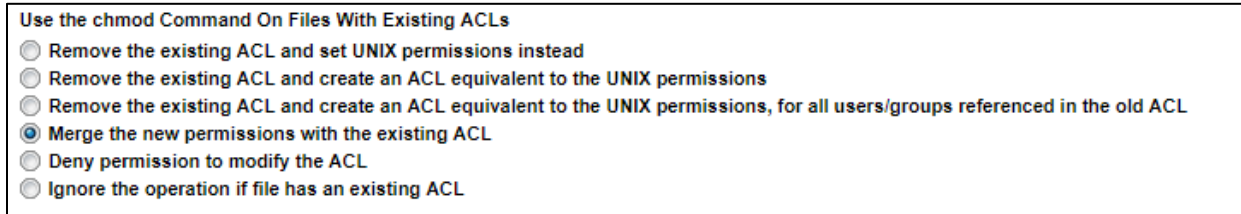
Figure 1   Use the chmod command on files with existing ACLs

Inheritance of ACLs created on directories by chmod

On Windows systems, the ACEs for directories can define detailed inheritance rules. On a UNIX system, the mode bits are not inherited. For a secure mixed environment with SMB and NFS, the recommended and default setting is **Do not make ACLs inheritable**.

You can find this setting under **ACL Policy Settings** > **General ACL Settings**, as shown in Figure 2.
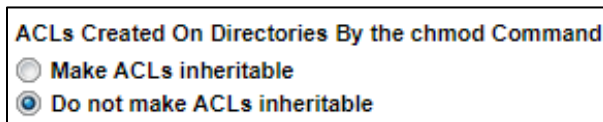
Figure 2   ACLs created on directories by the chmod command

Chown and chgrp commands on files with existing ACLs

For a mixed environment with multiprotocol file sharing, the default option **Modify the owner and/or group and ACL permissions** is the recommended approach because it preserves the ACEs that explicitly allow or deny access to specific users and groups. Otherwise, a user or group who was explicitly denied access to a file or directory might be able to gain access, possibly leading to security or compliance violations.

You can find this setting under **ACL Policy Settings** > **General ACL Settings**, as shown in Figure 3.
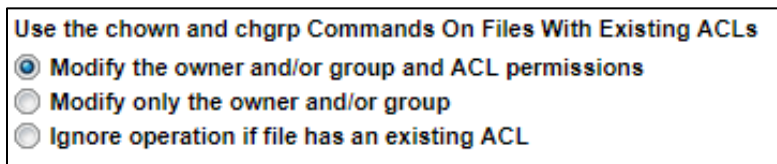
Figure 3   Use the chown and chgrp commands on files with existing ACLs

1.5   SMB share creation

There are several ways to create SMB shares which are listed in the Table 4:

Table 4   Methodologies to create SMB shares

Methodologies	Features	Note

WebUI	• Create and delete SMB share folders • Configure share permission	Ability to enable SMB continuous availability (CA) during creation
PowerScale CLI	• Create and delete SMB share folders • Configure share permission • View and close active SMB sessions • View and close open files	
Microsoft Management Console (MMC)		Inability to create SMB CA enabled share using MMC

MMC provides a simple way to manage SMB share creation. But it requires some configuration at the beginning and more importantly, it does not support enabling SMB CA feature during the creation. For detailed steps for creating an SMB share using MMC, see seeOneFS: How to create new SMB share using MMC. In the following sections, we will explore some of the key options when creating an SMB share.

1.5.1 Overlapping SMB share

OneFS supports overlapping display names for SMB shares if the display name appears only once per access zone. All SMB shares belong to a global list of shares and require unique SMB share names. By default, users see the SMB share name when connecting to the Dell PowerScale cluster; however, you can configure a display name for the SMB share that users see instead.

Display names must be unique within a zone; therefore, if you would like more than one SMB share to display the same name you must add each share to a separate access zone. For example, you can assign the "Home" as the display name for an SMB share in zone A and also assign it to a different share in zone B.

1.5.2 Existing ACL or Windows default ACL

If you have an existing directory structure that you want to add a share to, you most likely do not want to change the ACLs, so you should select the **Do not change existing permissions option** shown as Figure 4.

Figure 4   Directory ACLs

If you are creating a new share for a new directory, you should set the **Apply Windows Default ACLs** option and then once the share is created, go into the Security tab from Windows and assign permissions to users as needed. The selection of **Apply Windows Default ACLs** ends up converting the ACL to:

```
ISI7021-1# ls -led /ifs/tmp
drwxrwxr-x +  2 root  wheel  0 Jul 17 07:46 /ifs/tmp
OWNER: user:root
GROUP: group:wheel
CONTROL:dacl_auto_inherited,dacl_protected
0: group:Administrators allow dir_gen_all,object_inherit,container_inherit
1: creator_owner allow dir_gen_all,object_inherit,container_inherit,inherit_only
2: everyone allow dir_gen_read,dir_gen_execute
3: group:Users allow
dir_gen_read,dir_gen_execute,object_inherit,container_inherit
4: group:Users allow std_synchronize,add_file,add_subdir,container_inherit
```

Note: Do not apply these settings (Apply Windows default ACLs) to the /ifs directory. Doing so may make the cluster inoperable.

1.5.3   Opportunistic lock (oplock) and lease

When a client opens a file using the SMB protocol, the SMB service returns a file ID, or FID, which is used to further reference the specific opening of that file from a specific Windows client. Open mode is the term which the client can specify when it wants to open the file for reading, writing or executing. It is up to the server to validate that the client has sufficient permissions to open the file with the desired open mode. The term share mode specifies the operation type (read, write, and delete) on each file. A share mode can be any combination of read, write and delete. For example, a share mode of 'read' specifies that you are allowing

users to access the file in read-only mode. The server must also check to make sure an open mode does not conflict with any existing share mode before it can return success to the client. Once the open has passed the server's access and share mode checks, the server must do one of the followings:

- Grant the client its requested oplock on the file. This can be an exclusive oplock or a batch oplock. Exclusive oplock grants clients the ability to retain read data, cache data and metadata writes and byte-range lock acquisitions. Batch oplocks are identical to exclusive oplocks, except that they allow clients to cache open/close operations.
- Grant the client a lower-level oplock on the file, called a level II oplock. Level II oplocks also known as shared oplocks grant clients the ability to cache the results of read operations.
- Deny the oplock request.

The SMB oplock is a performance enhancement mechanism whereby the server cooperates with a client and allows the client to aggressively cache data under specific conditions. Oplocks allow a Windows client to cache read-ahead data, writes, opens, closes, and byte-range lock acquisitions. By caching these operations, clients may see a performance gain because the operations can be coalesced.

Starting with SMB2.1, SMB leases were introduced. It shares the same purpose with an oplock, which allows clients to improve performance by reducing network transmission. The newly added types of leases correspond to the new oplock types in SMB2.1. SMB2.1 just gives it a different name to distinguish it from the existing oplock functionality. A lease can be a combination of one or more of the leases types below:

- Read-caching lease: allow caching reads and can be shared by multiple clients.
- Write-caching lease: allow caching writes and is exclusive to only one client.
- Handle-caching lease: allow caching handles and can be shared by multiple clients.

One of the major differences between oplocks and lease is how they deal with multiple file handles (FID) in the same client or application. Oplocks do not allow data caching if there are multiple FIDs for the same file opened by a client or an application, meanwhile, lease allows full data caching on multiple FIDs for the same file opened for a client or application. This enhancement can provide a further performance boost, especially on high latency network.

Both oplocks and leases are supported in OneFS and can help SMB performance in most scenarios and for this reason oplock and leases are enabled by default. However, in some cases, some anti-virus software and old applications do not support this function well. In order to make these applications function well, we recommend disabling oplock and leases for the dedicated SMB share. For details, see OneFS: How to disable opportunistic locking (oplock) on SMB file shares and How to disable oplock leases in OneFS 7.x and later.

For more details of oplocks from Microsoft MSDN see the article Opportunistic Locks.

1.5.4 Access Based Enumeration (ABE)

There are three settings for ABE as listed in Table 5.

Table 5   ABE configurations

Level to enable and disable	ABE configurations	Description
Global	Access Based Share Enumeration	Will only show file shares that the requesters have permission to access

Share	Access Based Enumeration	Will only show the files and directories that the requesters have permission to access
	Access Based Enumeration Root Only	Only the root directory of the share is enabled/disabled for ABE

ABE can restrict the requesters to see only what they have permission to access which is good for security considerations. On the other hand, when ABE is enabled on a top-level directory with thousands of folders and files, PowerScale CPU utilization will be high and could potentially cause performance issues. Based on this point, the recommendation is to enable ABE for root only or turn off ABE for directories that have a large amount of files and subfolders. See OneFS CLI Administration Guide for more details on how to configure Access-based Enumeration.

1.6   SMB administration

This section discusses the considerations and common practices dealing with SMB security, management, and administration using PowerScale OneFS.

1.6.1   /ifs lockdown

By default, the /ifs root directory is configured as an unrestricted SMB share in the system access zone with unlimited access for the Everyone account.

PowerScale cluster administrators must consider whether these configurations are suitable for their deployment and manage the security implications appropriately. It is recommended that you remove this share after initial PowerScale cluster configuration.

1.6.2   Role-Based Access Control

A role is a collection of OneFS privileges that are granted to members of that role as they log in to the cluster. Role-Based Access Control (RBAC) allows delegating specific administration tasks to users.

It is recommended to assign the following privileges for SMB administrators using RBAC:

- SMB Settings (ISI_PRIV_SMB) – Configure SMB server
- WebUI (ISI_PRIV_LOGIN_PAPI) – Log in to Platform API and WebUI

Once the role and the privileges are set up, you can add members to it. Members can be any users from authentication providers such as AD, LDAP or NIS.

For other general considerations for RBAC, see OneFS Security Configuration Guide.

1.6.3   SMB signing

SMB signing can prevent man-in-the middle attacks within the SMB protocol. However, it will introduce performance degradation which can vary widely depending on the network and storage system implementation. Actual performance degradation can be verified only through tests in the real environment. See The Basics of SMB Signing for more details.

If SMB signing is desired, consider the following two aspects:

- Evaluate the impact before you enable this setting. For details, see OneFS: SMB Security Signatures and OneFS Security Configuration Guide.
- Enable SMB signing for the control path only. This solution requires that clients use SMB signing when accessing all DCERPC (Distributed Computing Environment / Remote Procedure Calls) services on the cluster, but does not require signed connections for the data path. This option requires you to enable four advanced parameters on the cluster. With these parameters enabled, the OneFS server rejects any non-signed IPC request that is initiated by a client. If clients are configured not to sign, they can access files over SMB, but cannot perform certain other functions, such as SMB share enumeration. For the details of how to enable this function, see OneFS Security Configuration Guide.

Enabling or disabling this feature requires two steps, one on the PowerScale side and the other on the client side. See OneFS: SMB Security Signatures for the detailed steps if the client is Windows.

For MAC OS X clients to access PowerScale data using SMB protocol, this setting is controlled with the `dsconfigad –packetsign` command.

1.6.4   Audits

You can audit SMB protocol access on a per-access zone basis and optionally forward the generated events to the Common Event Enabler (CEE) for export to third-party products. See OneFS File System Auditing with Dell PowerScale and Dell Common Event Enabler for a complete list of supported third-party products that can be used in CEE.

Because each audited event consumes system resources, it is recommended that you only configure audit zones for events that are needed by your auditing application.

Before OneFS 8, in the scenario when auditing feature is turned on and many operations for SMB workflow have been audited, SMB performance will be degraded. There may be no high CPU/memory/disk usage symptom, but SMB statistics response time shows slowness. For this case, it is recommended to open a support ticket to Dell Support. In OneFS 8, the values are set correctly by default.

1.6.5   Antivirus

OneFS leverages third-party antivirus software for file scanning. It sends files through Internet Content Adaptation Protocol (ICAP) to a server running third-party antivirus scanning software. These servers are referred to as ICAP servers. ICAP servers scan files for viruses.

There are two high level configurations and the considerations for file scanning as the followings.

On-access scanning

You can configure OneFS to send files to be scanned before they are opened, after they are closed, or both. The general common practice is to enable scan of files on close. It is recommended that you have at least one ICAP server for each node in the cluster. If those ICAP servers are unable to keep up with the workload on the cluster, more ICAP servers will need to be added. It is not uncommon for busy clusters to have two ICAP servers per node or more.

- Enable scan of files on close

  If OneFS is configured to ensure that files are scanned after they are closed, when a user creates or modifies a file on the cluster, OneFS queues the file to be scanned. OneFS then sends the file to an

ICAP server to be scanned when convenient. In this configuration, users can always access files without any delay and this is the most prevalent implementation.

However, it is possible that after a user modifies or creates a file, a second user might access the file before the file is scanned. If a virus was introduced to the file from the first user, the second user can access the infected file.

- Enable scan of files on open

If OneFS ensures that files are scanned before they are opened, when a user attempts to download a file from the cluster, OneFS first sends the file to an ICAP server to be scanned. The file is not sent to the user until the scan is complete. Scanning files before they are opened is more secure than scanning files after they are closed, because users can access only scanned files. However, scanning files before they are opened requires users to wait for files to be scanned, which affects the performance of the PowerScale cluster. For this reason, scan on open is not widely used on large busy PowerScale clusters.

If you configure OneFS to ensure that files are scanned before they are opened, it is recommended that you also configure OneFS to ensure that files are scanned after they are closed. Scanning files as they are both opened and closed will not necessarily improve security, but it will usually improve data availability when compared to scanning files only when they are opened. If a user wants to access a file, the file may have already been scanned after the file was last modified, and will not need to be scanned again if the ICAP server database has not been updated since the last scan. Scanning on both open and close usually comes with a high performance penalty and requires many more ICAP servers to be deployed. For this reason, scanning on both open and close is not widely used.

Policy scanning

Antivirus policies can be run manually at any time or configured to run according to a schedule and they target a specific directory on the cluster. In order to have minimum performance impact, it is recommended to schedule the scanning during off-peak or off-duty hours based on the business requirements.

By leveraging this policy, it is recommended to have at least two ICAP servers per cluster. The number of ICAP servers required really depends on how virus scanning is configured, the amount of data a cluster processes and the processing power of the ICAP servers.

1.6.6 Impact of upgrades

Pre-check and reconfigure unsupported SMB settings

Ensure that SMB settings on the cluster are supported by the version of OneFS to which you are upgrading.

If the SMB settings on the cluster are not supported by the version of OneFS to which you are upgrading, the upgrade might fail. Run the upgrade compatibility check utility to confirm whether your current settings are supported.

The upgrade compatibility check utility is included in the OneFS installation package. Start the upgrade compatibility check utility by running the following command, where <install-image-path> is the file path of the upgrade install image.

```
#isi upgrade cluster assess <install-image-path>
```

If the upgrade compatibility check utility detects unsupported SMB settings, remove or modify the unsupported SMB settings through the command-line interface or web administration interface before you upgrade. If you are upgrading from OneFS 6.0 or OneFS 5.5, remove or modify the settings by editing the `/etc/mcp/override/smbd.xml` file or the `/etc/mcp/override/smbd_shares.xml` file. After you modify your SMB settings, test the workflow.

Backup custom SMB settings

Most settings are preserved during OneFS upgrade. However, some customer settings might not be preserved. Backing up custom settings enables you to reapply any settings that are not preserved during the upgrade process.

The following setting shown in Table 6 is related to SMB and recommended to be backed up by dumping the setting into a text file before upgrading.

Table 6   SMB customized setting

Setting	Description	Recommendation
SMB audit logging	You have custom SMB logging settings configured	If you are upgrading from OneFS 7.0, you must reconfigure SMB audit logging based on the settings you have backed up.

Configuration changes during a rolling upgrade

You can continue to manage data and can modify some cluster configurations during a rolling upgrade, when you absolutely have to. For example, you can modify SMB shares. However, you can only make configuration changes from a node that has not yet been upgraded. If you attempt to make configuration changes from an upgraded node, the changes will not take effect.

Client connections during rolling upgrades

SMB is a stateful protocol which means it maintains a session state for all the open files in the PowerScale node where the client connects. This session state is not shared across the nodes. For a stateful protocol like SMB, it is recommended to use static IP pools. When a node reboots during the rolling upgrades, the static IP on that node is not accessible and SMB1 and SMB2 clients will be forced to disconnect and reconnect to a different node on the PowerScale. Although this is usually instantaneous, it is still a brief disruption. In order to achieve fully non-disruptive operation, SMB3 continuous availability (CA) is required. The SMB CA feature needs to be enabled at share creation time; to enable SMB CA, the following preconditions need to be met:

- SMB3 is supported
- The cluster is running OneFS 8.0 or later
- Clients are running Windows 8 or Windows Server 2012 R2 or later

For more information about SMB CA feature, see the SMB continuous availability and witness section in this paper.

In the case of a rolling upgrades, if your workflow requires the duration of the interruption to be further reduced, you can achieve this by forcing SMB connections to their CA pair before performing the node reboot. For details, see Reducing SMB client impacts during planned node reboots.

1.7 Performance

1.7.1 Overview

In the next section, we will discuss some performance considerations and useful CLI commands to identify performance issues with the SMB protocol and OneFS. For general troubleshooting guide, see the white paper [Dell PowerScale OneFS Cluster Performance Metrics Hints and Tips](#).

1.7.2 PowerScale performance management

Authentication provider

It is recommended to use Kerberos authentication instead of NT LAN Manager (NTLM) for performance reasons. As user load and authentication requests increase, NTLM can degrade performance because NTLM-based authentication inherently requires multiple round trips. For more detail about NTLM and Kerberos authentication, see the Microsoft article [Explained: Windows Authentication in ASP.NET 2.0](#).

For Kerberos authentication, it is important to verify that the service principal names (SPNs) have been set up correctly, otherwise the cluster may fall back to NLTM authentication. See the Kerberos authentication section for more guideline.

Data Access Pattern

PowerScale can be used for different types of workloads. The **Data Access Pattern** setting defines the optimization settings for accessing concurrent, streaming, or random data types. Files and directories use a **concurrent** access pattern by default. To optimize performance, select the pattern dictated by your workflow. Concurrent is a good default setting unless your workflow is very well understood.

Since access patterns can be defined on a per directory basis, it is worthwhile testing and evaluating different access patterns on different workflow data against different tools. Test and validate the access pattern against each data set and the jobs accessing it.

The settings can be configured by from the WebUI by navigating to **File System > Storage Pools > File Pools** and editing the desired file pool policy. Figure 5 shows the configuration of Data Access Pattern through the WebUI.
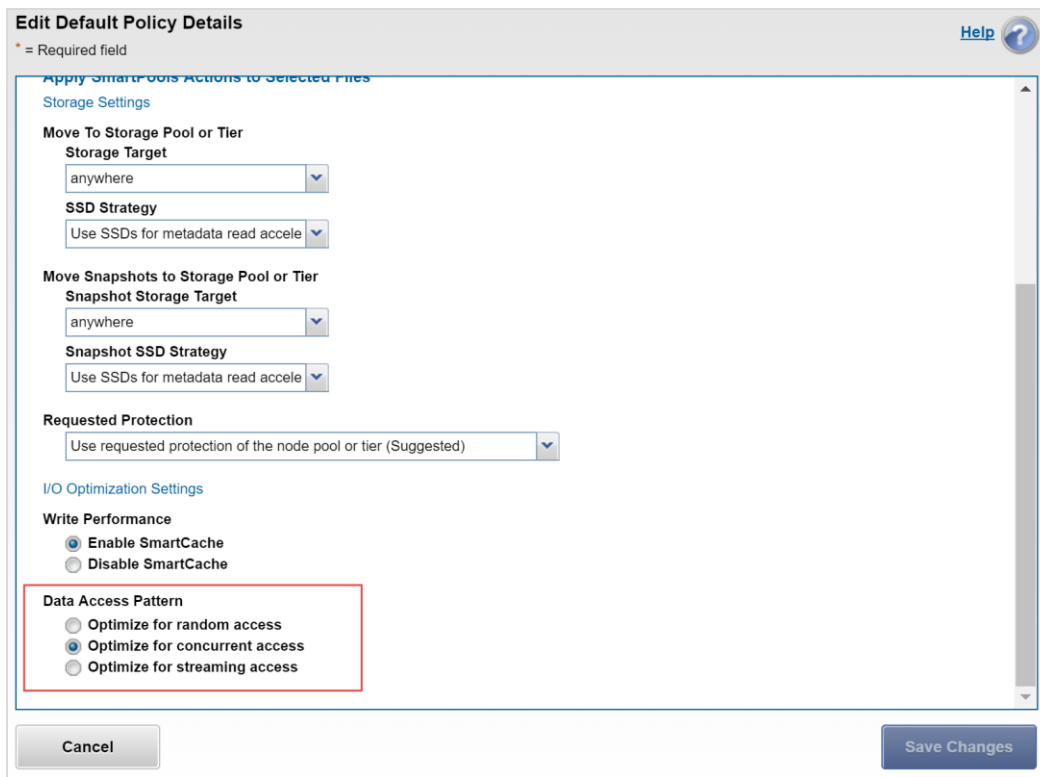
Figure 5   Data Access Pattern configuration in WebUI

Table 7 lists the different types of Data Access Pattern options in OneFS and its general description with different workflow examples.

Table 7   OneFS Data Access Pattern Options

OneFS Data Access Pattern	Description	Examples
Concurrent(default)	Concurrency access is the middle ground with moderate prefetching. Use this for file sets with a mix of both random and sequential access.	General Files and Directories
Streaming	Streaming access works best for sequentially read medium to large files. This access pattern uses aggressive prefetching to improve overall read throughput, and on disk layout spreads the file across a large number of disks to optimize access.	Large SMB Files and Directories A workflow heavy in video editing Increase sequential-read performance on MapReduce jobs
Random	The Random access setting performs little to no read-cache prefetching to avoid wasted disk access. This works best for small files (< 128KB) and large files with random small block accesses.	Typical VMware environment for random IO workflow.

SSDs for performance benefit

Solid-state drives (SSDs) can be used in various ways within OneFS to improve performance. Prior to SmartFlash, or L3 cache, SSDs were used exclusively as file system devices. SmartPools provided the mechanism to use these SSDs primarily as metadata acceleration devices, but also for reducing latency of actual data reads and writes for the appropriate workloads.

Figure 6 is a comparison of L3 cache with the other OneFS SmartPools SSD usage strategies. The principle benefits of L3 cache are around metadata read activity, user data read activity, and assistance with job engine performance.

Assists With	L3	Metadata Read	Metadata Read/Write	GNA	Data on SSD
Metadata read	Yes	Yes	Yes	Yes	No*
Metadata write	No	1 Mirror	All Mirrors	1 Additional Mirror	No*
Data read	Yes	No	No	No	Yes*
Data write	No	No	No	No	Yes*
Job Engine Performance	Yes	Yes	Yes	Yes	No*
Granularity	Node Pool	Manual	Manual	Global	Manual
Ease of Use	High	Medium	Medium	Medium	Lowest

Figure 6   Comparison of L3 cache and other SSD usage strategies.

You may need to select different SSD strategy based on the workload. For most of workflows with write component, it is recommended to select **Metadata Read/Write** SmartPools SSD option. For repeated random read workloads, the recommendation is to use **L3 cache** and you will observe latency reduced. Figure 7 shows the decision tree of various (non-L3 cache) SmartPools SSD options, and their requirements and dependencies.

For more details about common practices and considerations of different SSD strategy, see article Dell PowerScale OneFS SmartFlash and SmartPool and SSDs.
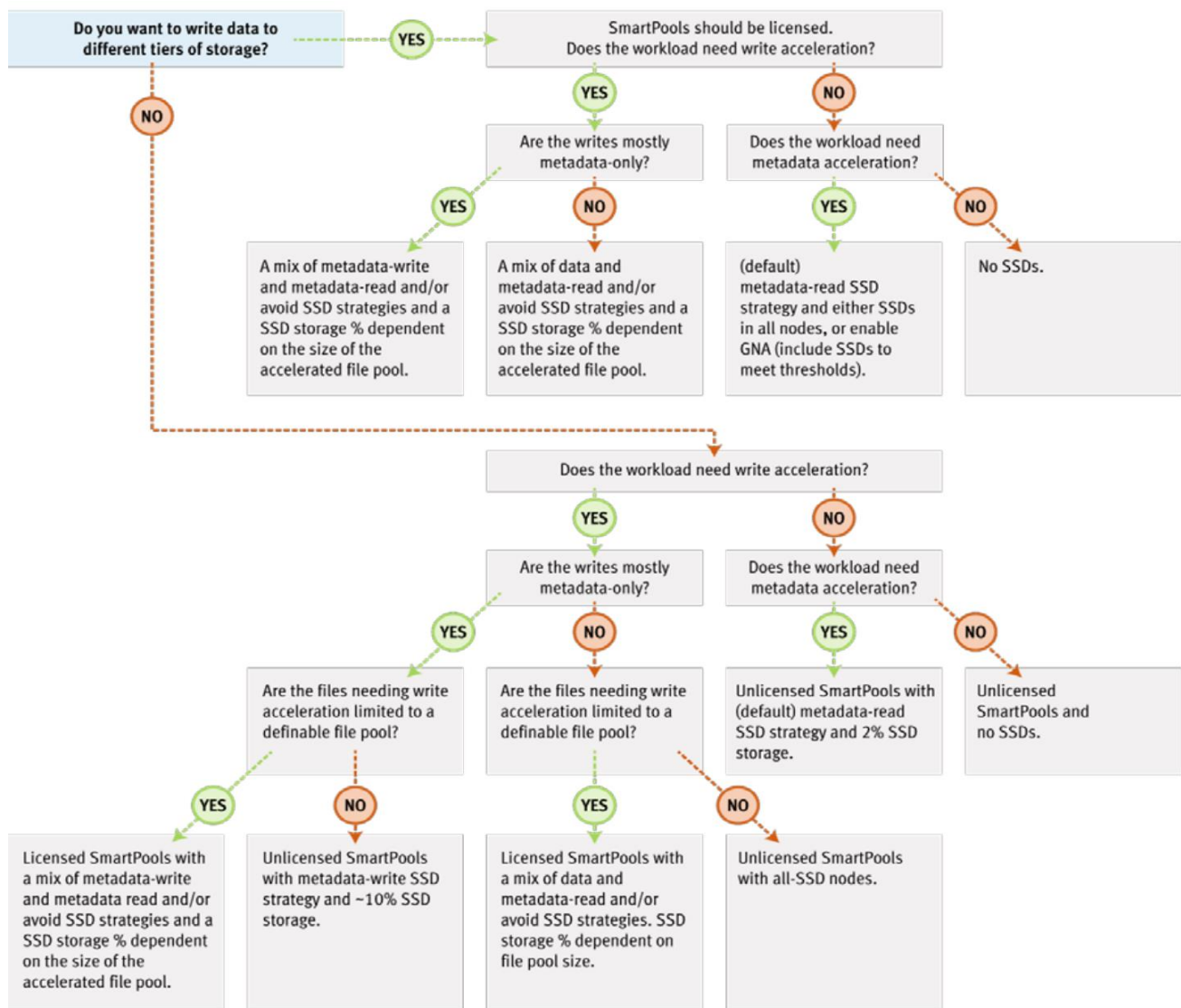
Figure 7   SSD strategy decision tree

1.7.3   isi statistics

isi statistics is a utility to access kernel counters that measure OneFS performance. These counters can give us a better understanding of the latency seen in various parts of the filesystem. It is an excellent tool for users to troubleshoot performance issues including SMB performance. Here are some useful commands for you to troubleshooting SMB performance issues:

Checking SMB clients and connections

You can use the following commands to identify how many clients are connecting to PowerScale nodes with SMB protocols.

```
# isi statistics query current --nodes=all --
stats=node.clientstats.connected.smb,node.clientstats.active.smb1,node.clientsta
ts.active.smb2
```

As Windows client statistics can only show SMB1 and SMB2. SMB3 is considered by Microsoft to be a "dialect" of SMB2, rather than a distinct version. As such, SMB3 client stats are included with the SMB2 statistics. The output displays the current number of SMB sessions per node and how many of those SMB connections are active on each node.

The output in Figure 8 shows there are 36 SMB sessions to node 1 with 8 active SMB2 sessions. The 36 SMB sessions represent clients that were connected to the node, but these clients did not send any requests during the time the command was run. As a result, these sessions are considered idle connections. The 8 active connections represent clients that sent an SMB2 request during the time the counter was collected.

```
x41040g-1# isi statistics query current --nodes=all --stats=node.clientstats.connected.smb,node.clientsta
ts.active.smb1,node.clientstats.active.smb2
   Node  node.clientstats.connected.smb  node.clientstats.active.smb1  node.clientstats.active.smb2
-----------------------------------------------------------------------------------------------------
      1                              36                             0                             8
      2                              33                             0                             6
      3                              33                      I        0                             6
average                              34                             0                             6
-----------------------------------------------------------------------------------------------------
Total: 4
```

Figure 8   Output of isi statistics query command

The number of active SMB2 connections that a node can process depends on the type of node. The more CPUs and RAM that a node has, the more active connections the node can process. The kernel imposes memory constraints on the OneFS protocol daemons, such as the input-output daemon (lwio), and these constraints limit the number of SMB2 connections.

To ensure that a node does not become overloaded with connections, you should monitor the number of SMB2 connections to each node. For example, for OneFS 8.1.1, it is recommended that SMB active connections per node are under 3,000 and the idle connections are under 27,000. Keep in mind that these are maximums that our fastest nodes with lots of memory could conceivably support. Slower nodes and nodes with lower memory configurations will support numbers lower than these. For general SMB connection limits guideline, see the white paper PowerScale OneFS Technical Specification Guide.

You can also use the following syntax to show the SMB2 client statistics displayed in 'top' format where data is continuously overwritten in a single table.

```
# isi statistics client --protocols=smb2 --format=top
```

Checking SMB protocol performance

You can use the following command to break out detailed SMB protocol performance. It can break out by detailed protocol operations with following command:

```
# isi statistics protocol --nodes=all --protocols=smb1,smb2 --interval 5 --
repeat 2 --degraded --sort=Class
```

Figure 9 shows the output of the detailed protocol operations by type, and average latency.

- **Ops** (operations per second) is the rate of operations in the sample time.
- **TimeAvg** is the average amount of latency measured in microseconds (1000 microseconds = 1 millisecond) for the protocol ops to the node.
- **TimeStdDev** measures the standard deviation of ops, the lower the number is, the closer most ops are to the average latency. The larger the number is, the more varied the data set is. A small number for TimeStdDev implies a consistent performance workload on PowerScale.

For more detailed information about how to understand isi statistics protocol output, see the article How to Read and understand the output of Isi Statistics Protocol Command.

```
f800eth-1# isi statistics protocol --nodes=all --protocols=smb1,smb2 --interval 5
 --repeat 2 --degraded --sort=Class
 Ops     In    Out   TimeAvg  TimeStdDev  Node  Proto           Class        Op
-------------------------------------------------------------------------------
3.2k 345.1k 223.6k    107.2       76.6     4    smb2 namespace_write set_info
3.2k 342.3k 221.9k    111.4       77.5     3    smb2 namespace_write set_info
3.2k 340.3k 220.6k    108.7       73.8     2    smb2 namespace_write set_info
3.1k 339.4k 220.0k    109.1       77.2     1    smb2 namespace_write set_info
2.0k   2.1G 169.2k   2826.7     2640.7     3    smb2           write     write
1.9k   2.0G 162.8k   2987.7     2830.0     4    smb2           write     write
1.9k   2.0G 161.2k   2974.0     3046.0     1    smb2           write     write
1.8k   1.9G 154.0k   3199.4     2948.7     2    smb2           write     write
-------------------------------------------------------------------------------
Total: 8
3.1k 339.5k 220.1k    108.7       76.9     1    smb2 namespace_write set_info
3.1k 337.6k 218.8k    106.9       78.1     4    smb2 namespace_write set_info
3.1k 333.9k 216.4k    113.7       79.3     3    smb2 namespace_write set_info
3.0k 328.7k 213.0k    111.4       73.3     2    smb2 namespace_write set_info
2.2k   2.3G 184.2k   2005.3     2361.0     3    smb2           write     write
2.1k   2.2G 180.1k   1992.4     2473.7     4    smb2           write     write
2.0k   2.1G 171.9k   2488.4     2752.4     2    smb2           write     write
2.0k   2.1G 166.2k   2507.3     2975.3     1    smb2           write     write
-------------------------------------------------------------------------------
Total: 8
```

Figure 9   Output of isi statistics protocol command

Table 8 outlines the common expectations about protocol latency time. The output TimeAvg needed converted to milliseconds (ms) if you are comparing to standard expectations in the table.

Table 8   Common expectations about protocol latency time

Namespace metadata		Read	Write
<10ms	Good	Dependent on I/O Size	Dependent on I/O Size
10ms – 20ms	Normal		
>20ms	Bad		
>50ms	Investigate		

You can also use following command to check if the authentication provider is causing an issue. For more detailed information about slow SMB authentication issue, see the article Intermittent slow SMB authentication or share enumeration performance; isi_cbind_d DNS delays.

```
# isi statistics protocol --nodes=all --protocols=lsass_in,lsass_out --totalby
Class --interval 5 --repeat 12 –degraded
```

Checking disks and systems

Once we narrow down the performance issues could be related specific SMB operations, you can also check disks and system performance to identify the root cause by using following commands:

```
# isi statistics drive --nodes=all --interval 5 --repeat 12 --degraded
```

Figure 10 shows the output of disk performance. The most useful statistics are TimeInQ and Queued, along with TimeAvg which is an accurate indication of drive load.

- **TimeInQ** (Time in Queue) indicates how long an operation is queued on a drive. This indicator is key for spindle-bound clusters. A time in queue value of 10 to 50 milliseconds equals Yellow zone, a time in queue value of 50 to 100 milliseconds equals Red.
- **Queued** (Queue depth) indicates how many operations are queued on drives. A queue depth of 5 to 10 is considered heavy queuing.

For more detailed information about how to understand isi statistics drive output, see PowerScale OneFS Cluster Performance Metrics Hints and Tips and How to Understand Isi Statistics Drive Output.

```
nl41080-1# isi statistics drive --nodes=all --interval 5 --repeat 12 --degraded
 Drive  Type  OpsIn  BytesIn  OpsOut  BytesOut  TimeAvg  TimeInQ  Queued  Busy
------------------------------------------------------------------------------
  1:1   SSD    0.0      0.0     0.0      0.0      0.0      0.0     0.0   0.0%
  1:2   SATA  29.2    435.8k    8.0    308.0k    0.0      0.0     0.0   9.0%
  1:3   SATA  24.6    263.8k    6.0    211.4k    0.0      0.0     0.0   7.5%
  1:4   SATA  25.0    386.7k    8.2    206.4k    0.0      0.0     0.0   8.8%
  1:5   SATA  31.4    412.9k    7.2    118.0k    0.0      0.0     0.0   9.8%
  1:6   SATA  25.4    337.5k    6.8    139.3k    0.0      0.0     0.0   8.5%
  1:7   SATA  25.8    290.0k    3.4     63.9k    0.0      0.0     0.0   5.1%
  1:8   SATA  25.6    309.7k    4.8    124.5k    0.0      0.0     0.0   7.1%
  1:9   SATA  19.6    301.5k    7.6    190.1k    0.0      0.0     0.0   6.4%
 1:10   SATA  14.6    317.8k    9.2    201.5k    0.0      0.0     0.0   8.5%
 1:11   SATA  24.0    296.6k    7.8    137.6k    0.0      0.0     0.0   7.9%
```

Figure 10   Output of isi statistics drive command

A good overview of the cluster can be obtained using the isi statistics system command. This will show CPU, core protocols, network, disk, and totals in a single line.

Performance dataset monitoring

For a workload interacting with OneFS through Protocol Operations, System Jobs, or System Services, performance dataset monitoring is introduced to provide improved workload monitoring starting with OneFS 8.2.0. The performance dataset monitoring allows administrators to define metrics in datasets for performance monitoring as needed.

To view all supported dataset metric, using the following command. Table 9 shows the details of each metric.

```
# isi performance metrics list
```

Table 9   Supported dataset metrics

Metric	Description
username	The user that the current workload belongs to. For example, a user access data using SMB protocol.
groupname	The user group that the current workload belongs to in OneFS. For any dataset with group metric, only primary groups are reported. However, when the datasets with group metric are pinned or have a filter applied for a group, the supplementary groups will also be scanned and reported. An example is shown in Figure 14.
zone_name	The access zone name that the current workload belongs to.
share_name	The SMB share name that the current workload belongs to.

export_id	The NFS export id that the current workload belongs to. Supported in OneFS 8.2.2 and above.
protocol	Support SMB and NFS protocols in OneFS 8.2.2 and above. Can be set to `smb1`, `smb2`, `nfs3`, or `nfs4`. Only support SMB protocol for OneFS 8.2.0 and OneFS 8.2.1, can be set to `smb1` or `smb2`.
system_name	This is only available for the predefined System dataset. The system name of a given workload: For services started by `isi_mcp`/`lwsm`/`isi_daemon`, this is the service name itself. For SMB protocol, this is named `smb`. For job engine jobs, this is formed with `Job: job_id`. Anything ran using "isi_run -w" is formed with `Run: pid`.
job_type	This is for job engine jobs, it is formed with `job_type[job_phase]`. For example, `AutoBalance[0]` for a AutoBalance job phase 1 execution. For more details about supported job types, see the [OneFS Job Engine](#) white paper.
local_address	Local IP address, CIDR subnet, or IP address range of the client causing the workload.
remote_address	Remote IP address, CIDR subnet or IP address range of the client causing the workload.
path	For SMB protocol only. The path under `/ifs` that the current workload belongs to. It is only reported if they are pinned or have a filter applied. There will be double accounting when using this metric as a file can belong to multiple path. Thus, any double accounting will be aggregated into the `Overaccounted` workload. An example is shown in Figure 15.

When viewing the workload statistics with defined dataset, there are multiple workload types to tell how resources are consumed by the dataset and system. Table 10 lists the details of each workload type.

Table 10   Workload types

Workload Type	Description
Dynamic (shown as "-")	Top-N tracked workloads based on defined dataset. By default, top 1024 workloads will be listed. Can be modified using command `isi performance settings modify`.
Pinned	Make a specific workload always visible regardless of resource usage.
Overaccounted	The sum of all statistics that have been counted twice within a same dataset. This would be used when path metric or groupname metric are used in a dataset. Examples are shown in Figure 14 and Figure 15.
Excluded	The amount of resources consumed by workloads that do not match the filter conditions in a dataset.
Additional	The aggregate of dynamic workloads not in the top-N workloads. The N is 1024 by default.
System	The amount of resources consumed by the kernel.

Unknown	The amount of resources that cannot be categorized as the above workload types.

SMB performance dataset monitoring examples

Use the following command to create a dataset containing metrics (`username`, `protocol`, `share_name`) without filters required. Figure 11 shows the results of performance statistics using the dataset.

```
# isi performance datasets create --name=smb_ds username protocol share_name
```

```
llin-w2ahxyx-1# isi statistics workload list --dataset=smb_ds
   CPU BytesIn BytesOut Ops Reads Writes  L2   L3  ReadLatency WriteLatency OtherLatency    Node UserName Proto ShareName WorkloadType
-----------------------------------------------------------------------------------------------------------------------------------------
  1.3s     0.0     0.0 0.0   0.0   0.0 0.0  0.0       0.0us        0.0us        0.0us cluster       -     -         -        System
471.1ms    0.0     0.0 0.4  32.1  26.8 0.0  0.0       0.0us        0.0us        0.0us cluster       -     -         -      Excluded
179.5ms  68.6M   5.5k 65.5 0.1  11.3k  6.8 0.0  0.0   0.0us       59.1ms        0.0us cluster     bob   smb2  smbshare        -
103.1ms   9.5k  85.3M 81.3 5.4k  0.0  5.0k 0.0  0.0  57.4ms       0.0us        0.0us cluster    root   smb2  smbshare        -
  5.6us    0.0     0.0 0.0   0.0   0.2 0.0  0.0       0.0us        0.0us        0.0us cluster       -     -         -       Unknown
  0.0us    0.0     0.0 0.0   0.0   0.0 0.0  0.0       0.0us        0.0us        0.0us cluster       -     -         -     Additional
  0.0us    0.0     0.0 0.0   0.0   0.0 0.0  0.0       0.0us        0.0us        0.0us cluster       -     -         - Overaccounted
-----------------------------------------------------------------------------------------------------------------------------------------
Total: 7
```

Figure 11   Output using dataset without filters

A dataset also supports adding filters for better visibility. Multiple filters can be applied to the same dataset, and a workload will be included if it matches any one of the filters. Any workload that does not match a filter will be aggregated into an `Excluded` workload. The following commands create another new dataset with applying a filter to view performance metrics only for username "bob". The result appears in Figure 12.

```
# isi performance datasets create --name=smb_ds1 username protocol share_name --
filters=username
# isi performance filters apply --dataset=smb_ds1 --name=bob_filter username:bob
```

```
llin-w2ahxyx-1# isi statistics workload list --dataset=smb_ds1
   CPU BytesIn BytesOut Ops Reads Writes  L2   L3  ReadLatency WriteLatency OtherLatency    Node UserName Proto ShareName WorkloadType
-----------------------------------------------------------------------------------------------------------------------------------------
  1.4s     0.0     0.0 0.0   0.0   0.0 0.0  0.0       0.0us        0.0us        0.0us cluster       -     -         -        System
583.0ms   9.5k  85.3M 81.4 5.7k 26.2 4.8k 0.0  0.0  58.4ms       0.0us        0.0us cluster       -     -         -      Excluded
179.6ms  68.8M   5.5k 65.6 0.1  11.3k  6.8 0.0  0.0   0.0us       59.7ms        0.0us cluster     bob   smb2  smbshare        -
  5.7us    0.0     0.0 0.0   0.0   0.2 0.0  0.0       0.0us        0.0us        0.0us cluster       -     -         -       Unknown
  0.0us    0.0     0.0 0.0   0.0   0.0 0.0  0.0       0.0us        0.0us        0.0us cluster       -     -         -     Additional
  0.0us    0.0     0.0 0.0   0.0   0.0 0.0  0.0       0.0us        0.0us        0.0us cluster       -     -         - Overaccounted
-----------------------------------------------------------------------------------------------------------------------------------------
Total: 6
```

Figure 12   Output using dataset with a filter

To make a specific workload always visible regardless of resource usage, you can pin the workload for a specific dataset using the following command. The result appears in Figure 13.

```
# isi performance workloads pin --dataset=smb_ds1 --name=root_pin username:root
protocol:smb2 share_name:smbshare
```

```
llin-w2ahxyx-1# isi statistics workload list --dataset=smb_ds1
   CPU BytesIn BytesOut Ops Reads Writes  L2   L3  ReadLatency WriteLatency OtherLatency    Node UserName Proto ShareName WorkloadType
-----------------------------------------------------------------------------------------------------------------------------------------
  1.4s     0.0     0.0 0.0   0.0   0.0 0.0  0.0       0.0us        0.0us        0.0us cluster       -     -         -        System
480.7ms  10.0k  89.3M 85.1 2.6k 10.3 8.9k 0.0  0.0  53.6ms       0.0us        0.0us cluster       -     -         -      Excluded
196.5ms  75.2M   6.0k 71.7 0.0  12.5k  7.6 0.0  0.0   0.0us       51.4ms        0.0us cluster     bob   smb2  smbshare        -
  0.0us    0.0     0.0 0.0   0.0   0.0 0.0  0.0       0.0us        0.0us        0.0us cluster       -     -         -     Additional
  0.0us    0.0     0.0 0.0   0.0   0.0 0.0  0.0       0.0us        0.0us        0.0us cluster       -     -         - Overaccounted
  0.0us    0.0     0.0 0.0   0.0   0.0 0.0  0.0       0.0us        0.0us        0.0us cluster    root   smb2  smbshare      Pinned
  0.0us    0.0     0.0 0.0   0.0   0.0 0.0  0.0       0.0us        0.0us        0.0us cluster       -     -         -       Unknown
-----------------------------------------------------------------------------------------------------------------------------------------
Total: 7
```

Figure 13   Output of a pinned workload using dataset with a filter

Shown in Figure 14, a file is written to a cluster with user `bob`, `group02` is the primary group of the user, and `group01` is the supplementary group of the user. We pin the `group01` so the workload is accounted twice, and the total amount is aggregated into the `Overaccounted` workload.

```
llin-w2ahxyx-1# isi statistics workload list --dataset=group_dataset_pinned
    CPU  BytesIn  BytesOut  Ops   Reads  Writes   L2    L3   ReadLatency  WriteLatency  OtherLatency   Node  GroupName  WorkloadType
-----------------------------------------------------------------------------------------------------------------------------------
   1.1s     0.0      0.0    0.0    0.0     0.0    0.0   0.0      0.0us         0.0us         0.0us cluster      -         System
 316.2ms    0.0      0.0    0.0    0.0     0.4   18.8   0.0      0.0us         0.0us         0.0us cluster      -        Excluded
 197.8ms   76.9M     6.2k  73.3    0.0    12.8k   7.3   0.0      0.0us         5.5ms         0.0us cluster      - Overaccounted
 197.8ms   76.9M     6.2k  73.3    0.0    12.8k   7.3   0.0      0.0us         5.5ms         0.0us cluster  group01       Pinned
 197.8ms   76.9M     6.2k  73.3    0.0    12.8k   7.3   0.0      0.0us         5.5ms         0.0us cluster  group02          -
   1.3ms    0.0      0.0    0.0    0.0     0.0    0.0   0.0      0.0us         0.0us         0.0us cluster   wheel           -
   0.0us    0.0      0.0    0.0    0.0     0.0    0.0   0.0      0.0us         0.0us         0.0us cluster      -      Additional
   0.0us    0.0      0.0    0.0    0.0     0.0    0.0   0.0      0.0us         0.0us         0.0us cluster      -       Unknown
-----------------------------------------------------------------------------------------------------------------------------------
Total: 8
```

Figure 14   Groupname metric overaccounted

Shown in Figure 15, a file is written to directory `/ifs/path1/path2`, and the workload is monitored with the path metric at the same time. We find the workload is accounted twice and the total amount is aggregated into the `Overaccounted` workload.

```
llin-w2ahxyx-1# isi statistics workload list --dataset=path_dataset
    CPU  BytesIn  BytesOut  Ops   Reads  Writes   L2    L3   ReadLatency  WriteLatency  OtherLatency   Node              Path  WorkloadType
------------------------------------------------------------------------------------------------------------------------------------------
   1.2s     0.0      0.0    0.0    0.0     0.0    0.0   0.0      0.0us         0.0us         0.0us cluster              -          System
 411.7ms  477.8    523.0    3.2   17.8   586.9  22.1   0.0    891.3us       445.0us         2.0ms cluster              -        Excluded
 120.3ms   47.4M    3.9k   45.7    0.0     7.7k   3.7   0.0      2.3ms        23.2ms       160.9us cluster        /ifs/path1       Pinned
 120.2ms   47.4M    3.9k   45.6    0.0     7.7k   3.7   0.0      3.0ms        23.2ms       187.5us cluster              - Overaccounted
 120.2ms   47.4M    3.9k   45.6    0.0     7.7k   3.7   0.0      3.0ms        23.2ms       187.5us cluster /ifs/path1/path2       Pinned
 156.0us    0.0      0.0    0.0    0.0     0.7    0.5   0.0      0.0us         0.0us         0.0us cluster              -         Unknown
   0.0us    0.0      0.0    0.0    0.0     0.0    0.0   0.0      0.0us         0.0us         0.0us cluster              -      Additional
------------------------------------------------------------------------------------------------------------------------------------------
Total: 7
```

Figure 15   Path metric overaccounted

1.7.4 Client OS performance management

The following sections will discuss the common practices and consideration of client OS performance management and tuning methodologies. Since the environment of clients may vary, the general guideline is the tunings are not identical and the impact must be evaluated before applying them.

Windows parameter management

There are a few SMB parameters can be used to tune SMB client performance. The following section will list the most significant one. For a complete list of the parameters, see Performance Tuning for File Servers.

Bandwidth throttling

Bandwidth throttling can be controlled with the key - DisableBandwidthThrottling (REG_DWORD) under `HKLM\System\CurrentControlSet\Services\LanManWorkstation\Parameters`. This key does not exist by default and has an assumed value of 0. The SMB2 client will try to limit its network throughout on links. In some scenarios, especially like Virtual Desktop Infrastructure (VDI) deployments, the drives mapped to PowerScale are connected by high speed frontend networks, so there is no need to try and limit the throughput of the SMB sessions. This value should be set to 1.

macOS parameter management

Many parameters can be tuned in macOS that affect SMB connections. For details on all the configurations and tuning methodologies, see the white paper PowerScale macOS Performance Optimization which is focused on macOS 10.13 "High Sierra" and 10.14 "Mojave".

2 SMB feature considerations and common practices

New features were introduced in SMB 3.0, and different versions of PowerScale OneFS support the following SMB features shown in Table 11.

Table 11   SMB features supported by PowerScale OneFS versions

SMB feature	Supported OneFS versions	Section in this document
SMB Multichannel	PowerScale OneFS 7.1.1 and later	SMB multichannel
SMB Symbolic Links	PowerScale OneFS 7.1.1 and later	SMB symbolic links
SMB server-side copy	PowerScale OneFS 8.0 and later	SMB server-side copy
SMB File Filtering	PowerScale OneFS 8.0 and later	SMB file filtering
SMB Continuous Availability	PowerScale OneFS 8.0 and later	SMB continuous availability and witnessSMB continuous availability and witness
SMB Encryption	PowerScale OneFS 8.1.1 and later	SMB encryption

2.1 SMB continuous availability and witness

2.1.1 Feature introduction

PowerScale OneFS 8.0 introduced support for SMB Continuous Availability (CA), which can enable users to perform both planned and unplanned disruptive event of PowerScale nodes in a cluster without interrupting server applications storing data on these file shares. It improves the resilience of SMB3-capable client connections to SMB shares during events such as PowerScale node reboots. This feature applies to Microsoft Windows 8, Windows 10, Windows Server® 2012 R2 and Windows Server 2016 clients as part of SMB 3.0 new features.

2.1.2 How SMB continuous availability works

When the SMB client initially connects to the file share, the client determines whether the file share has the continuous availability property set. If it does, this means the file share supports SMB continuous availability. When the SMB client subsequently opens a file on the file share, it requests a persistent file handle. When the PowerScale node receives the request, the PowerScale node will return the file handle along with a unique key (Resume Key). The resume key can resume the handle state after planned or unplanned failover.

Figure 16 shows the workflow between SMB clients and PowerScale nodes when a failure occurs. If a planned move or failure occurs on the PowerScale cluster node to which the SMB client is connected, the SMB client attempts to reconnect to another cluster node with the resume key. Once it successfully reconnects to another node in the PowerScale cluster, the SMB client starts the resume operation using the resume key. When PowerScale receives the resume key, it will recover the handle state to the same state prior to the failure with end-to-end support. For some operations, it can be replayed. For other operations, it

cannot be replayed. From a client perspective, it appears the I/O operations are stalled for a small amount of time.
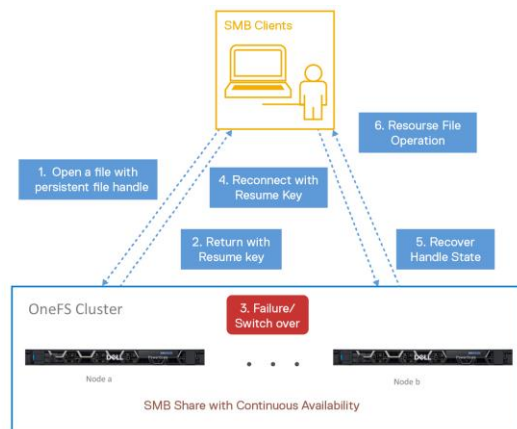

Figure 16   SMB Share with continuous availability

How witness service works with SMB continuous availability

In SMB 3.0, Microsoft introduced a Remote Procedure Call (RPC) based mechanism to inform the clients of any state changes in the SMB servers. This service is called Service Witness Protocol (SWP) which ensures time-critical applications will quickly re-connect to a new node in an PowerScale cluster when there is a failure without waiting for Transmission Control Protocol (TCP) timeouts or SMB timeouts. It will minimize outages and is supported by any PowerScale node in the pool.

Figure 17 shows the workflow between SMB clients and PowerScale nodes with witness service. When the SMB client connects to a file share with CA on an PowerScale cluster, the SMB client will get the witness node list from PowerScale. The SMB client picks up a different cluster node in the same pool and issues a registration request to the witness node for availability events. The witness service then listens to cluster events related to the PowerScale node the SMB client is connected to.

When the node becomes inaccessible, the witness service receives a OneFS Group Management Protocol (GMP) event and notify client failure of the node. The primary role of the OneFS GMP is to help creating and maintaining a group of synchronized nodes. Once receiving the witness notification, clients will immediately failover and reconnect to the new node which significantly speeds up recovery from unplanned failures. The reconnection is reduced from 50-60 seconds (TCP timeouts) to only a few seconds.
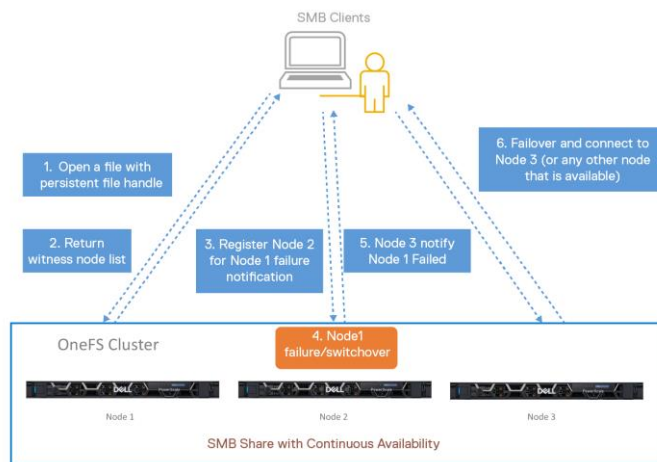
Figure 17   Witness service works with SMB continuous availability

2.1.3   Considerations

These are some key considerations that we recommend during the design and implementation phases:

- SMB continuous availability can only be enabled during share creation. However, users can still update timeout, lockout, and write integrity settings when creating or modifying a share. For more detailed information about SMB continuous availability settings, see the article Enable SMB continuous availability.
- SMB continuous availability is not enabled by default. An existing share needs to be re-created in order to enable CA. You can also use a script with CLI command Isi_smb_ca_share on PowerScale cluster which can help to re-create an existing share to enable SMB CA.
- If a planned reboot of a PowerScale node occurs, the duration of the interruption to the client connection to a CA-enabled SMB share can be further reduced by forcing SMB connections to their CA pair before performing the node reboot. For detail about implementation steps, see the article Reducing SMB client impacts during planned node reboots.
- If a customer uses SMB CA and has write I/O that is mostly sequential in nature, the EC (Endurant Cache) should be turned off on that SMB CA share to ensure performance. Once the EC is turned on, SMB CA on PowerScale will store stable writes on EC across the PowerScale nodes first which could cause EC to be a potential bottleneck. EC can lower the average latency for small and random stable writes workloads. However, EC can become a bottleneck when writes are stable and sequential. If the customer's write I/O is mostly small and random, the EC should remain on. If there is a mix of sequential and random write I/O, additional tests are important to determine the correct setting for EC on that SMB CA share.
- The EC can be turned on or off either globally or on specific directories or files. You can use following command to disable EC per directory. For more detailed settings about EC, see the article OneFS Endurant Cache.

```
# isi set -c coal_only <directory_name>
```

2.2 SMB multichannel

2.2.1 Feature introduction

SMB multichannel establishes multiple network connections to the PowerScale cluster over aggregated network interface cards (NICs), which results in balanced connections across CPU cores, effective consumption of combined bandwidth, and connection fault tolerance. Starting with PowerScale OneFS 7.1.1, OneFS supports the multichannel feature which applies to Microsoft Windows 8 and Windows Server 2012 or later version.

It is important to meet software and NIC configuration requirements to support SMB multichannel on the PowerScale cluster. OneFS can only support SMB3 multichannel when the following software requirements are met:

- Windows Server 2012, 2012R2, 2016, or Windows 8/10 clients
- SMB multichannel must be enabled on both the PowerScale cluster and Windows clients. It is enabled on PowerScale cluster by default.

SMB3 multichannel establishes a single SMB session over multiple network connections only on supported NIC configurations. SMB multichannel requires at least one of the following NIC configurations on the client computer:

- Two or more network interface cards.
- One or more network interface cards that support Receive Side Scaling (RSS).
- One or more network interface cards configured with link aggregation. Link aggregation enables you to combine the bandwidth of multiple NICs on a node into a single logical interface.

SMB multichannel automatically discovers supported hardware configurations on the client that have multiple available network paths. Each node on the PowerScale cluster has at least one RSS-capable network interface card. The client-side NIC configuration shown in Table 12 determines how SMB multichannel establishes simultaneous network connections per SMB session.

Table 12   Client-side NIC Configuration Options

Client-side NIC Configuration	Description
Single RSS-capable NIC	SMB multichannel establishes a maximum of four network connections to PowerScale node over the NIC. The connections are more likely to be spread across multiple CPU cores, which reduces the likelihood of performance bottleneck issues and achieves the maximum speed capability of the NIC.
Multiple NICs	If the NICs are RSS-capable, SMB multichannel establishes a maximum of four network connections to the PowerScale cluster over each NIC. If the NICs on the client are not RSS-capable, SMB Multichannel establishes a single network connection to the PowerScale cluster over each NIC. Both configurations allow SMB Multichannel to leverage the combined bandwidth of multiple NICs and provides connection fault tolerance if a connection or a NIC fails. **Note:** SMB multichannel cannot establish more than eight simultaneous network connections per session. In a multiple NIC configuration, this might limit the number connections allowed per NIC. For example, if the configuration contains three RSS-capable NICs, SMB multichannel might establish three connections over the first NIC, three connections over the second NIC and two connections over the third NIC.

Client-side NIC Configuration	Description
Aggregated NICs	SMB multichannel establishes multiple network connections to the PowerScale cluster over aggregated NICs, which results in balanced connections across CPU cores, effective consumption of combined bandwidth, and connection fault tolerance. **Note:** The aggregated NIC configuration inherently provides NIC fault tolerance that is not dependent upon SMB.

2.2.2   Considerations

These are some key considerations that we recommend during the design and implementation:

- No manual SMB configurations are needed on the Windows machine or on PowerScale to enable this SMB multichannel.
- Do not use LACP on the PowerScale cluster. SMB multichannel will automatically detect the IP addresses of both 10GbE/40GbE interfaces on the client and load balance across each of the two interfaces on the dual-ported NIC.
- SMB multichannel only works between a client and a single PowerScale node. It cannot share the load between PowerScale nodes. With SMB multichannel, Windows client connections to PowerScale node have built-in failover and all throughput is load-balanced between NICs.
- To see SMB multichannel in action, once connected, look at either TCP connections using `netstat`, and/or client statistics using `isi statistics client --numerical` to see at least two connections from clients to each NICs on the PowerScale node.
- The full throughput from the PowerScale cluster for the operation is available.

2.3   SMB server-side copy

2.3.1   Feature introduction

In order to increase system performance, windows clients using SMB2 or SMB3 can use the SMB server-side copy (SSC) feature in PowerScale OneFS.

Windows clients can offload copy/move operations to the PowerScale cluster. In processing such a request, the network round-trip is avoided. Windows clients making use of server-side copy may experience performance improvements for file copy operations, because file data no longer needs to traverse the network. The server-side copy feature reads and writes files only on the PowerScale cluster, avoiding the network round-trip and duplication of file data.

2.3.2   Considerations

These are some key considerations that we recommend during the design and implementation:

- This feature only applies to file copy operations in which the source and destination file handles are open on the same share, and does not work for cross-share operations. When copying data across shares in a OneFS cluster, clients will pull the data from source share to local and write the data to destination share through network.

- It is recommended to upgrade OneFS cluster to a newer version if the cluster is running on the version prior to OneFS 8.0.0.3. For more information, see the article OneFS 8.0.x: Application unable to copy a file between shares.
- This feature is enabled by default on PowerScale cluster, and can only be disabled system-wide across all zones. You can disable the SMB server-side copy using the CLI command line on OneFS by running `isi smb settings global modify --server-side-copy=no`.
- Server-side copy is incompatible with the SMB continuous availability feature. If continuous availability is enabled for a share and the client opens a persistent file handle, server-side copy will be automatically disabled for that file.
- In OneFS versions 8.0.1.X and 8.1.0.X, a combination of having apple extensions enabled and SMB server-side copy being disabled will result in failures for copies in the Finder on macOS X. The workaround is to enable server-side copy on the cluster. Server-side copy is enabled by default. For more detailed information about this issue, see article OneFS 8.0.1.X & 8.1.0.X - If Apple extensions are enabled and SSC is disabled, Mac OS file copies via finder will fail.

2.4   SMB encryption

2.4.1   Feature introduction

OneFS 8.1.1 and above provide SMB encryption to secure access to data over untrusted networks by providing over the wire encryption between the client and PowerScale cluster. It is an on-wire data encryption which prevents an attacker from tampering with any data packet in transit without needing an extra infrastructure.

SMB encryption can be used by any clients which support SMB3 encryption from Windows Server 2012, 2012R2, 2016, Windows Client 8, and Windows 10 and does not require any extra infrastructure management. PowerScale can also be configured to allow accepting or rejecting the old clients that lack the SMB encryption support access.

Figure 18 shows the different client connection behavior after enabling SMB encryption on PowerScale cluster. In this configuration, Windows 7 client connection is rejected because it lacks the SMB encryption support. Windows 10 client data access will be encrypted on the wire to protect data against snooping.
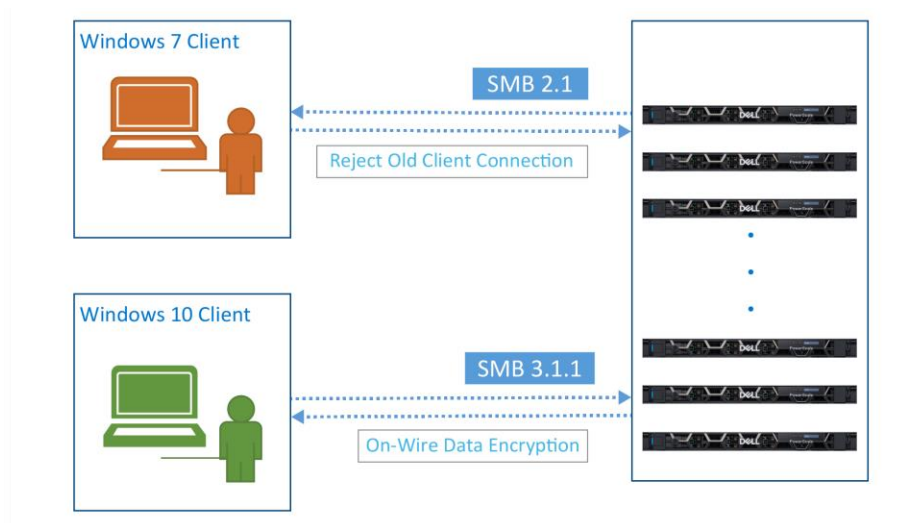

Figure 18   SMB3 encryption on a PowerScale cluster

The encryption algorithm of Windows Server 2012 R2 (or Windows 8) and Windows 2016 (or Windows 10) are different. There are currently three SMB3 dialects which are all supported by PowerScale OneFS 8.1.1 and above:

- SMB 3.0 with AES-128-CCM encryption (Windows 8 or Windows Server 2012)
- SMB 3.0.2 with AES-128-CCM encryption (Windows 8.1 or Windows Server 2012R2)
- SMB 3.1.1 with added AES-128-GCM encryption (Windows 10/Windows Server 2016)

SMB encryption has been enhanced in SMB 3.1.1. The AES-128-GCM mode offers a significant performance gain comparing to SMB 3.0.x. On the PowerScale side, the encryption and decryption happen in the kernel level with Intel CPU extensions for hardware acceleration to gain a performance benefit for next generation PowerScale clusters. It can be easily managed at the global, access zone and individual share level on Dell PowerScale:

- For global level, on-wire data between clients and PowerScale clusters will be encrypted after authentication.
- For access zone level, on-wire data between clients in the access zone will be encrypted after authentication.
- For share level, on-wire data between clients and share will be encrypted once clients can have access to the share.

2.4.2   Considerations

These are some key considerations that we recommend during the design and implementation:

- SMB encryption zone settings retain their autonomy. Changes in global settings do not override existing access zone or share settings. If there are no explicit zone-level settings, then the PowerScale node will check the global settings.
- Only the global level encryption can be configured with the WebUI. Figure 14 shows the global level SMB encryption settings in the WebUI. To configure encryption settings by access zone or by share, run CLI command `isi smb settings <option>` to configure. For more detailed information about CLI commands, see [OneFS CLI Administration Guide](#).
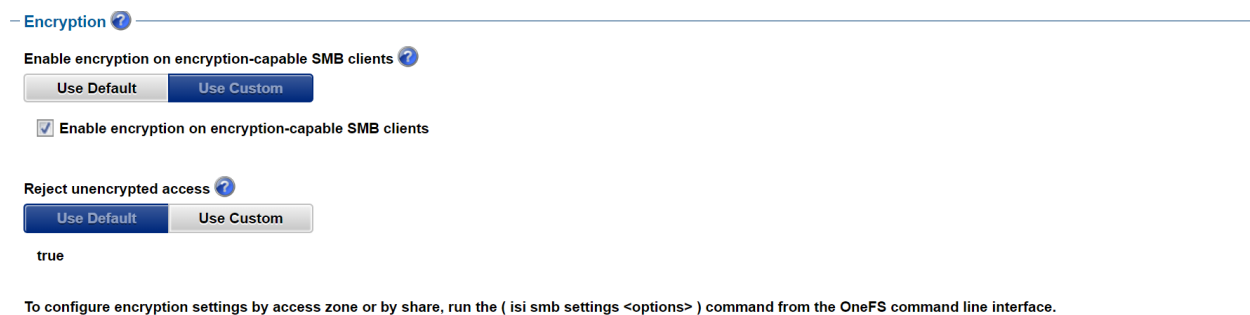

Figure 19   Global level SMB encryption settings in the WebUI

- Making SMB encryption configuration changes requires either restarting or refreshing (less intrusive) the SMB server. You can refresh SMB server service by running the command `/usr/likewise/bin/lwsm refresh srv` on the PowerScale cluster.
- It is recommended to make sure **RejectUnencryptedAccess** is set correctly, and refresh the PowerScale node by changing the configuration. You can use following CLI command to configure the option:

```
# isi smb settings global modify --revert-reject-unencrypted-access
```

- Once you enable encryption on the PowerScale cluster, client will always operate with encryption.

2.5 SMB symbolic links

2.5.1 Feature introduction

PowerScale OneFS 7.1.1 and later enable SMB2 clients to access symbolic links in a seamless manner. Many administrators deploy symbolic links to virtually reorder file system hierarchies, especially when crucial files or directories are scattered around an environment.

In an SMB share, a symbolic link (also known as a symlink or a soft link) is a type of file that contains a path to a target file or directory. Symbolic links are transparent to applications running on SMB clients, and they function as typical files and directories. Support for relative and absolute links is enabled by the SMB client. The specific configuration depends on the client type and version.

OneFS exposes symbolic links through the SMB2 protocol, enabling SMB2 clients to resolve the links instead of relying on OneFS to resolve the links on behalf of the clients. To transverse a relative or absolute link, the SMB client must be authenticated to the SMB shares that the link can be followed through. However, if the SMB client does not have permission to access the share, access to the target is denied and Windows will not prompt the user for credentials.
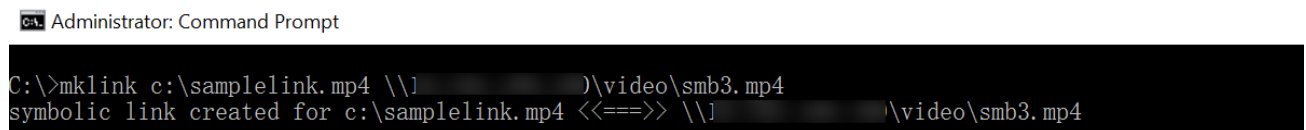
Windows and SMB make a distinction between links to files and links to directories. Once a link is set to be of a certain type, this flag cannot change. Directory links are fully supported in OneFS 8.0 and beyond.

How Windows creates a symbolic link with OneFS

Windows symlinks are created using the SMB protocol by writing a regular file in the initial Create request, then converting it into a symlink with a Set Reparse Point request.

Whenever a PowerScale OneFS node completing an SMB2 request encounters a symbolic link, it returns a Symbolic Link Error Response, which includes the target path (stored in the inode), the remaining section of the original file path, and whether the target path is relative or absolute. From here the client is expected to formulate another request to PowerScale OneFS node, using the details from this response. If there are three symlinks in the path, this process will be repeated three times as each link in the path is reached.

Figure 20 shows an example to create a symbolic file link on Windows client that points to a file on a PowerScale cluster. For detailed steps about how to implement symbolic link with PowerScale, see the Dell article How to create a directory symbolic link in Windows that points to an PowerScale cluster.

Figure 20   Create a symbolic file link that points to a file on a PowerScale cluster

2.5.2 Considerations

These are some key considerations that we recommend during the design and implementation:

- To create a symbolic link from Windows client, it is recommended to **Run as administrator** on the Command Prompt with `mklink` command. Otherwise, it may return with error message "You do not have sufficient privilege to perform this operation". By default, members of the Administrators group have rights to create symbolic links shown in Figure 21. This user right should only be given to trusted users. Symbolic links can expose security vulnerabilities in applications that are not designed to handle them. For more information, see the Microsoft article Create symbolic links.
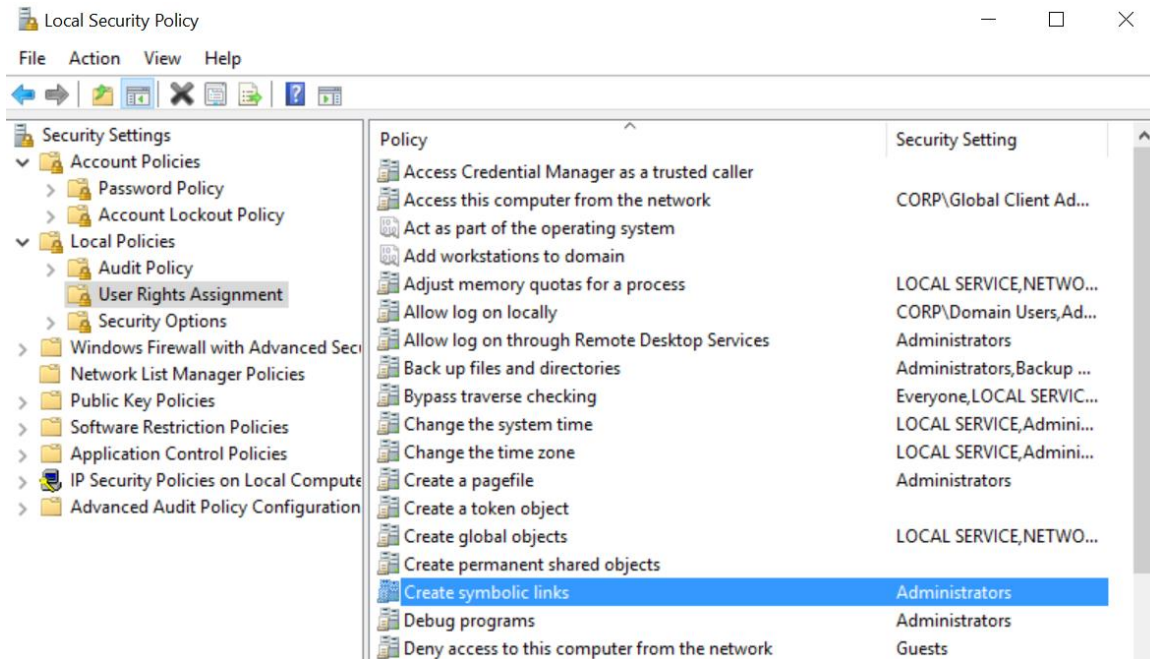
Figure 21   Create symbolic link security policy setting

- If users want to manage symbolic links through mapped network drives with SMB shares, it is recommended to use net use in administrator console with the `mklink` command.
- When deleting a symbolic link, the target file or directory still exists. However, when deleting a target file or directory, a symbolic link continues to exist and still points to the old target, thus becoming a broken link.
- When you create a symbolic link, it is designated as a file link or directory link. Once the link is set, the designation cannot be changed. Directory links are fully supported in OneFS 8.0 and beyond.
- To delete symbolic links, use the `del filename` command in Windows client.
- You can use the following command on Windows client to show all the symbolic links on an SMB share mapped to drive `O`.

```
> dir /AL /S O:\
```

- PowerScale OneFS provides the ability to disable all SMB symbolic link features through a registry key. This key is called `SMB2Symlinks` and can be accessed using the OnefsConfigSMB2Symlinks function. To enable or disable symlinks, run the `lwregshell` utility from the OneFS CLI as shown in Figure 17:

```
# cd /usr/likewise/bin
# ./lwregshell
> cd HKEY_THIS_MACHINE\Services\lwio\Parameters\Drivers\onefs
> ls                              /* to check the current value */
```

```
            > set_value "SMB2Symlinks" 0        /* 0 to disable, 1 to enable */
```

```
[HKEY_THIS_MACHINE\Services\lwio\Parameters\Drivers\onefs]
  "ACECountGuess"                   REG_DWORD        0x00000014 (20)
  "AuditGlobalSacl"                 REG_BINARY       01,00,10,80,14,00,00,00,24,00
,00,00,34,00,00,00,00,00,00,00,01,02,00,00,00,00,00,05,20,00,00,00,20,02,00,00,0
1,02,00,00,00,00,00,05,20,00,00,00,20,02,00,00,02,00,08,00,00,00,00,00
  "ChangeNotifyInterval"            REG_DWORD        0x00000064 (100)
  "ConfigVersion"                   REG_SZ           "0"
  "DotSnapAccessibleChild"          REG_DWORD        0x00000001 (1)
  "DotSnapAccessibleRoot"           REG_DWORD        0x00000001 (1)
  "DotSnapVisibleChild"             REG_DWORD        0x00000000 (0)
  "DotSnapVisibleRoot"              REG_DWORD        0x00000001 (1)
  "FakeGlobalMultiString"           REG_MULTI_SZ[0] "fake1"
                                    REG_MULTI_SZ[1] "fake2"
                                    REG_MULTI_SZ[2] "fake3"
  "FileAttributeEncryptedIgnored"   REG_DWORD        0x00000000 (0)
  "GuestUser"                       REG_SZ           "nobody"
  "HandleLeaseHardThreshold"        REG_DWORD        0x0000c350 (50000)
  "HandleLeaseSoftThreshold"        REG_DWORD        0x0000afc8 (45000)
  "IgnoreEAs"                       REG_DWORD        0x00000000 (0)
  "LegacyIrpDispatch"               REG_DWORD        0x00000000 (0)
  "OnefsCpuIoMultiplier"            REG_DWORD        0x00000002 (2)
  "OnefsCpuMultiplier"              REG_DWORD        0x00000004 (4)
  "OnefsMaxIoWorkers"               REG_DWORD        0x00000010 (16)
  "OnefsMaxWorkers"                 REG_DWORD        0x00000080 (128)
  "OnefsNumIoWorkers"               REG_DWORD        0x00000000 (0)
  "OnefsNumWorkers"                 REG_DWORD        0x00000000 (0)
  "SMB2Symlinks"                    REG_DWORD        0x00000001 (1)
```

Figure 22   Check the current value of SMB2Symlinks on the PowerScale cluster

2.6   SMB file filtering

2.6.1   Feature introduction

OneFS 8.0 introduces new file filtering capabilities. SMB file filtering allows PowerScale administrators to control what type of files can be written by these protocols to a cluster, based on include or exclude filter lists. OneFS file filtering can be used across SMB clients to allow or disallow writes to an export, share, or access zone.

This feature allows certain types of file extensions to be blocked, for files which might cause security problems, productivity disruptions, throughput issues, or storage clutter. Configuration can be either be a blocklist, which blocks explicit file extensions, or an allowlist, which explicitly allows writes of only certain file types. For example, to prevent users from saving MP3 files to their OneFS based home directory, a blocklist exclude rule for `*.mp3` files can be configured.

2.6.2   Considerations

These are some key considerations that we recommend during the design and implementation:

- If you choose to deny file writes, you can specify file types by extension that are not allowed to be written. OneFS permits all other file types to be written to the share, shown in Figure 23.

Figure 23   Deny file extension writes setting of SMB share on the PowerScale cluster

- If you choose to allow file writes, you can specify file types by extension that are allowed to be written. OneFS denies all other file types to be written to the share, shown in Figure 24.

Figure 24   Allow file extension writes setting of SMB share on the PowerScale cluster

3　SmartQoS

3.1　SmartQoS Overview

SmartQoS provides the ability for administrators to apply an operations ceiling or limit to specific workloads to prioritize desired business workloads:

1. Applicable to NFSv3, NFSv4, NFSoRDMA, SMB, and S3
2. Enables IT infrastructure teams to achieve performance SLAs
3. Enables throttling of rogue or low priority workloads and therefore gives priority to other business critical workloads
4. Helps minimize DU events due to overloaded clusters.

Note: SmartQoS is a post commit feature, which means it will only work after the upgrade has been committed successfully.

3.2　Configuration ops limitation

SmartQoS is implemented based on the partitioned performance in OneFS. Concepts such as dataset, workload, and so on are still applied to SmartQoS. Use the following steps to configure an ops limitation using SmartQoS:

- Enable SmartQoS in OneFS
- Create a SmartQoS dataset
- Pin a workload from the dataset
- Set an ops limitation for a pinned workload

3.2.1　Enable SmartQoS in OneFS

The SmartQoS feature is by default enabled. Use the following CLI command to view the current settings:

```
 # isi performance settings view
                 Top N Collections: 1024
        Time In Queue Threshold (ms): 10.0
 Target read latency in microseconds: 12000.0
Target write latency in microseconds: 12000.0
           Protocol Ops Limit Enabled: Yes
```

To disable the SmartQoS feature, use the following command:

```
# isi performance settings modify --protocol-ops-limit-enabled=false
protocol_ops_limit_enabled: True -> False
```

3.2.2　Create a SmartQoS dataset

A dataset is used to categorize workloads by various metrics, including:

- export_id
- local_address
- protocol

- share_name
- zone_name
- groupname
- path
- remote_address
- username

In this example, we use protocol and path to create a test dataset.

1. Use the following command to create the dataset:

```
# isi performance datasets create --name ds1 protocol path
Created new performance dataset 'ds1' with ID number 1.
Note: Resource usage tracking by 'path' metric is only supported by SMB and NFS.
```

You can use the WebUI for the same purpose：

a. Select **Smart QOS** under **Cluster management** in the WebUI:
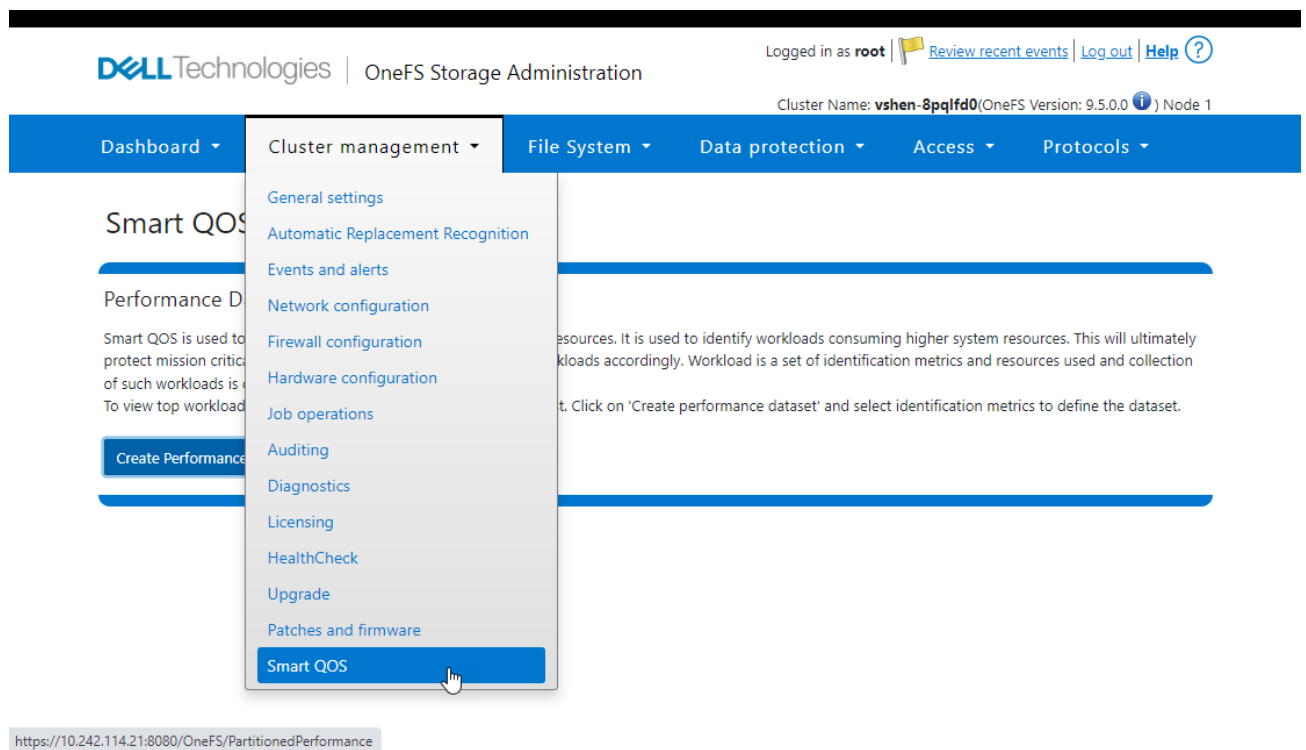
Figure 25   Smart QOS in the dropdown menu

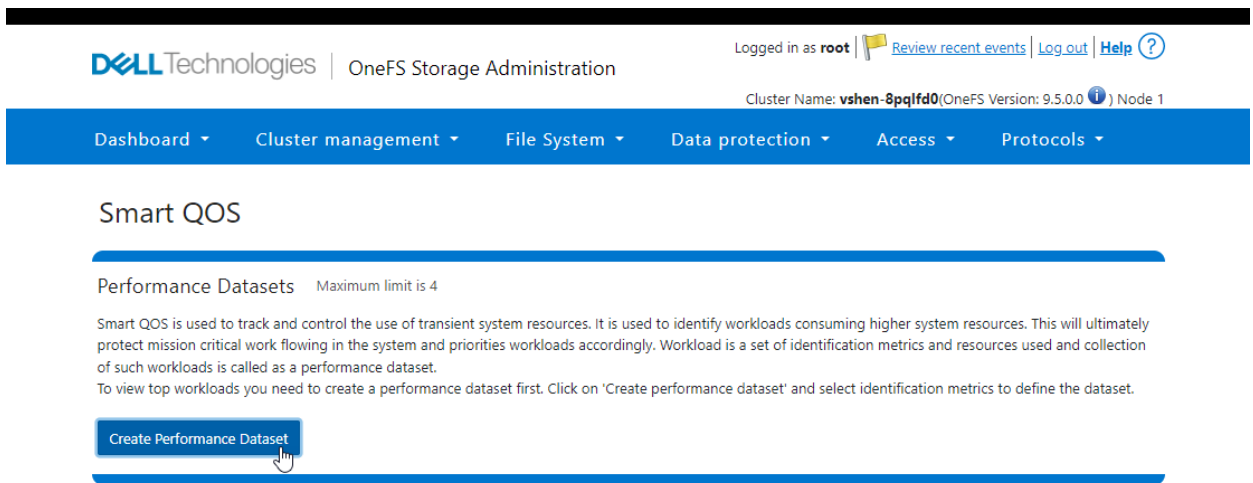b. Click the **Create Performance Dataset** button:

Figure 26   Click Create Performance Dataset

      c.   Specify your **dataset name**, select the **Path** and **Protocol** checkboxes, then click **Save**:
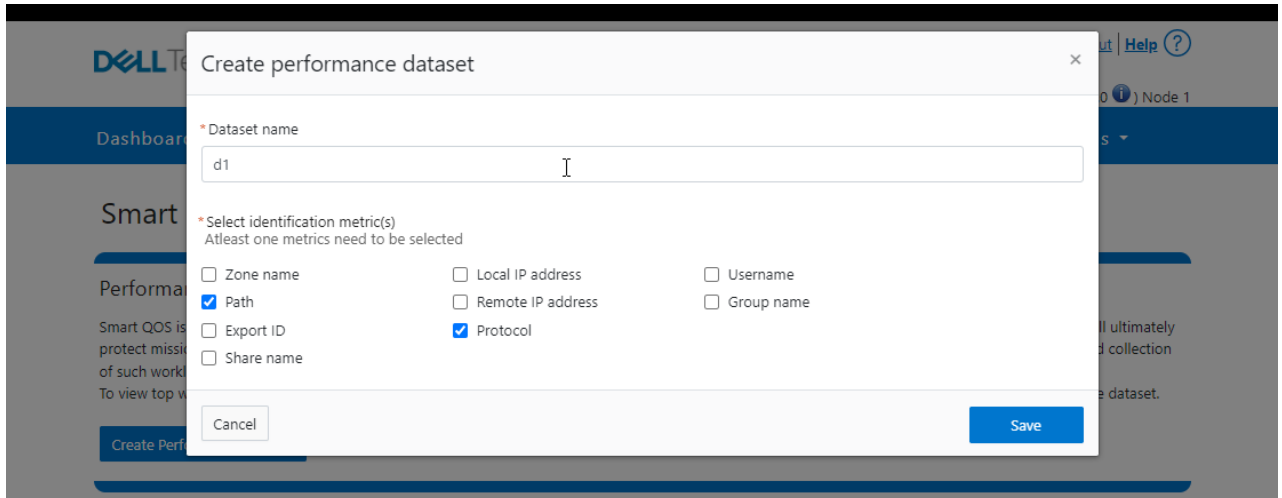

Figure 27   Create a performance dataset

    2.   Use the following command to list all the datasets

```
# isi performance datasets list
```

3.2.3   Pin a workload from the dataset

After you have created the dataset, you can pin a workload from there by specifying the metric values. In this example, we use the dataset created in the section Create a SmartQoS dataset and set the metric values to the following to pin the workload we need:

- **Protocol** is `smb2`
- **Path** is `/ifs/data/client_export`

Use the following command:

```
# isi performance workloads pin ds1 protocol:smb2 path:/ifs/data/client_export
```

```
Pinned performance dataset workload with ID number 100.
```

To pin the workload in the WebUI:
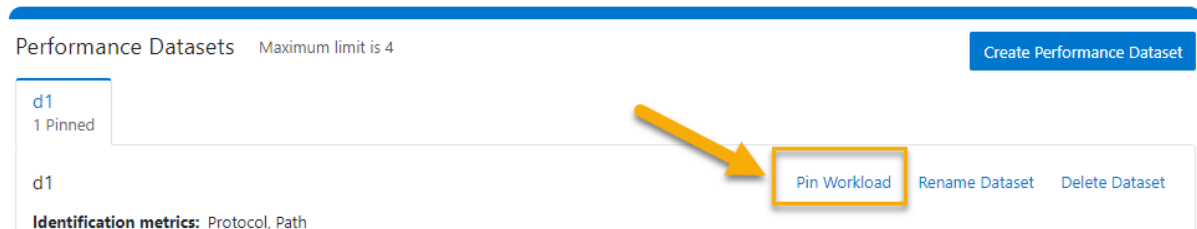
1. Click **Pin Workload**.

Figure 28   Pin workload

2. Select the protocol from the drop down list, specify the path, then click the **Pin workload** button.

To list all the pinned workloads from a specified dataset, use the following command:

```
# isi performance workloads list ds1
```

3.2.4   Set an ops limit for a pinned workload

For a pinned workload in a dataset, you can use the following command to set the limit for the protocol ops:

```
# isi performance workloads modify <dataset> <workload ID> --limits
protocol_ops:<value>
```

In this example, the name of the dataset is ds1 and the workload ID is 100. We also set the ops limit to 10:

```
# isi performance workloads modify ds1 100 --limits protocol_ops:10
protocol_ops: 18446744073709551615 -> 10
```

A Technical support and resources

Dell.com/support is focused on meeting customer needs with proven services and support.

Storage technical documents and videos provide expertise that helps to ensure customer success on Dell Storage platforms.

A.1 Related resources

Dell Technologies documentation

The following documentation provides additional and relevant information. Access to these documents depends on your login credentials. If you do not have access to a document, contact your Dell representative.

PowerScale Network Design Considerations

PowerScale Multiprotocol Data Access with a Unified Security Model

OneFS CLI Administration Guide

OneFS Security Configuration Guide

PowerScale OneFS Cluster Performance Metrics Hints and Tips

Dell PowerScale OneFS SmartFlash

Dell PowerScale OneFS: macOS Network Storage User Experience and Performance Optimizations

Microsoft documentation

The following documentation on the Microsoft website provides additional helpful information:

Microsoft SMB Protocol and CIFS Protocol Overview

Performance Tuning for File Servers

The Basics of SMB Signing

Opportunistic Locks