Technical White Paper

Dell EMC PowerScale: Common AntiVirus Agent Solution

Abstract

This document discusses general considerations, configurations, performance, and sizing of the Common AntiVirus Agent (CAVA) solution for Dell EMC™ PowerScale™ storage.

September 2021

Revisions

Date	Description
September 2020	Initial release
September 2021	Add DNS delegation when creating the anti-virus pool

Acknowledgments

Author: Vincent Shen

The information in this publication is provided “as is.” Dell Inc. makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Table of contents

Revisions.....2
Acknowledgments.....2
Table of contents.....3
Executive summary.....5
1 Overview.....6
1.1 Architecture.....6
1.1.1 Architecture overview.....6
1.1.2 On-demand scan.....7
1.1.3 Scheduled scan.....8
1.1.4 Manual scan.....9
1.2 Load balance.....9
1.2.1 SMB connection load balancing.....9
1.2.2 CEE connection load balancing.....10
1.3 Supported vendors.....12
1.4 Supportability.....13
2 Deployment and configuration.....14
2.1 Overview.....14
2.2 Create CAVA server in OneFS.....15
2.3 Create an IP Pool.....16
2.4 Create a dedicated access zone: AvVendor.....17
2.5 Create an Active Directory authentication provider for AvVendor.....18
2.6 Update role: AVVendor.....18
3 Performance overview.....20
3.1 Test environment.....20
3.2 Scenario 1: Performance of scan on close.....21
3.2.1 Test methodology.....21
3.2.2 Test results.....22
3.3 Scenario 2: Performance of scan on read.....23
3.3.1 Test methodology.....23
3.3.2 Test results.....24
3.4 Scenario 3: Performance impact on SWBUILD.....25
3.4.1 Test methodology.....25
3.4.2 Test results.....26
4 Sizing.....27

4.1 Overview 27
4.2 General best practices for sizing 27
4.3 Methodology 1 27
4.4 Methodology 2 27
5 General considerations 29
5.1.1 NANON considerations 29
5.1.2 CloudPools considerations 30
5.1.3 SyncIQ consideration 31
5.1.4 Anti-virus vendor considerations 31
A Technical support and resources 32

Executive summary

Many enterprises have strict security policies in place to detect, clean (remove), or quarantine viruses. This is often performed at the individual user level with per-system anti-virus (AV) solutions from third-party security vendors. Many of these same enterprises use large, centralized storage platforms to contain user home directories or group-project repositories. Because these are the same file types that reside on end-user workstations, viruses must not be resident on the storage systems. Since end-user solutions do not work well for centralized storage depots, a different type of solution is required.

Third-party software is often used to scan the storage array through end-user access or manually scheduled policies from a central anti-virus scan server. There are methods to do this process using RPC or with SMB and NFS. However, there are drawbacks to these methods since they use proprietary solutions and non-centralized scanning through NAS protocols.

Common AntiVirus Agent (CAVA) provides an anti-virus solution for Dell EMC™ PowerScale™ storage. It uses an industry-standard Common Internet File System (CIFS) protocol in a Microsoft® Windows Server® environment. CAVA uses third-party anti-virus software to identify and eliminate known viruses before they infect files on the system.

This white paper covers the general considerations, configurations, performance, and sizing of the CAVA solution for PowerScale.

1 Overview

1.1 Architecture

1.1.1 Architecture overview

Figure 1 illustrates a high-level architecture of the CAVA anti-virus solution for PowerScale. Dell EMC Common Event Enabler (CEE) is between the PowerScale cluster and the anti-virus applications. When clients trigger the scanning workflow, Dell EMC PowerScale OneFS™ generates the request to the CEE or CAVA agent through the HTTP protocol. Then, the anti-virus application fetches the scanning files from PowerScale through a hidden SMB share **CHECK$.** These files are scanned by the anti-virus applications, and then CEE sends the response back to the PowerScale cluster.

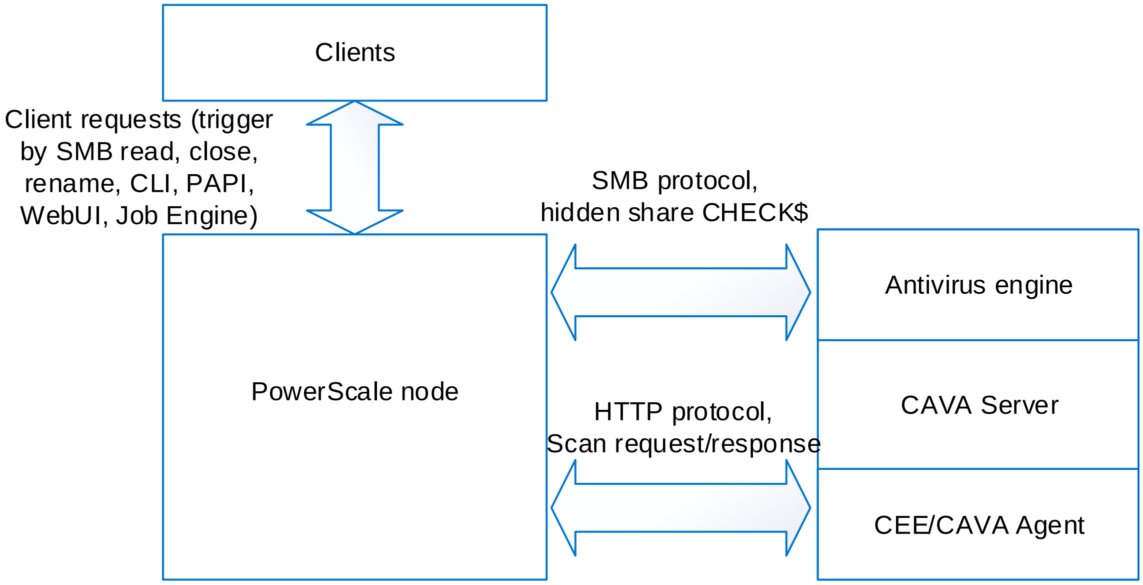

Clients
Client requests (trigger by SMB read, close, rename, CLI, PAPI, WebUI, Job Engine)
SMB protocol, hidden share CHECK$
PowerScale node
Antivirus engine
CAVA Server
HTTP protocol, Scan request/response
CEE/CAVA Agent

Figure 1 The overall architecture of the CAVA solution for PowerScale

The detailed architecture is explained separately by three scanning workflows from the OneFS perspective:

- On-demand scan (also known as on-access scan): This scan is triggered by the proper SMB operation, like a read and close operation, and depends on your scan profile. There are two scan profiles in OneFS:
  - Standard profile: Captures a close and rename operation from an SMB perspective, and triggers the scan operation on the corresponding file.
  - Strict profile: Captures a read, close, and rename operation from an SMB perspective, and triggers the scan operation on the corresponding file.
- Scheduled scan: This scan is triggered by the job engine either manually or by schedule.
- Manual scan: This scan is triggered by the CLI command or the responding PAPI to scan a single file.

Generally, the CAVA solution provides better performance, lower total cost of ownership (TCO), and much less CPU and memory usage from the PowerScale perspective compared with the solution for ICAP.

Note: ICAP is supported, which remains unchanged from the previous version of OneFS.

1.1.2 On-demand scan

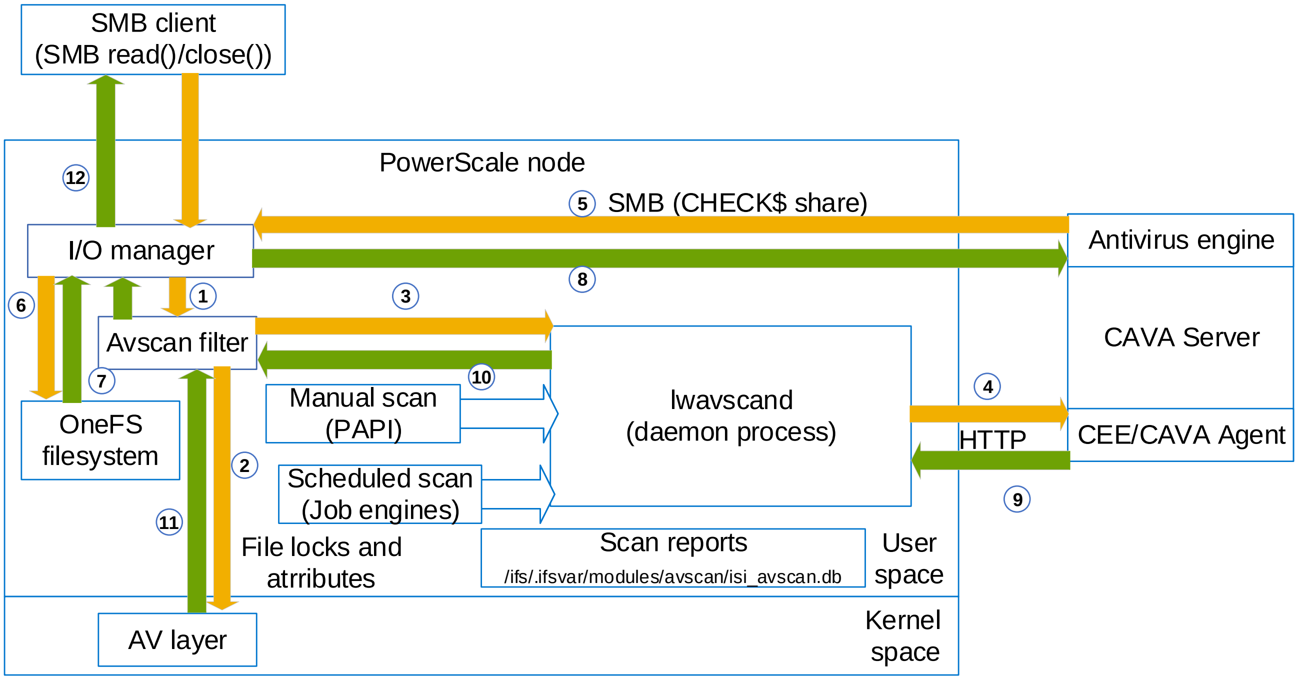


Figure 2 The workflow of the on-demand scan

Figure 2 shows the workflow of an on-demand scan. Depending on which scan profile is selected, the SMB requests with a read or close operation code are captured by the I/O manager in OneFS. It extracts the file path and name to the avscan filter, which can be configured on an access-zone basis. It filters according to the following aspects:

- File extension to include
- File extension to exclude
- File path to exclude

If the file matches all the criteria, an internal process **lwavscand** sends the HTTP scanning request to the CEE/CAVA agent. Simultaneously, OneFS sets the locks on this file. Then, the anti-virus application tries to fetch this file through a hidden SMB share **CHECK$** from the file system in OneFS. CAVA supports only downloading a part of a file for scanning, and this helps the performance. After the corresponding content is downloaded to the CAVA server, it runs the scan with the anti-virus engine, and CEE sends the scan results and response back to the process **lwavscand**. At this stage, some scanning attributes are written to this file and the lock is released. The scanning attributes are listed below:

- Scan time
- Scan Result
- Anti-virus signature timestamp
- Scan current

Then the previous SMB workflow can continue if the file is not infected. Otherwise, the file is denied access. If there are errors during the scan and the scan profile is strict, the setting **Open on fail** determines the next action.

1.1.3 Scheduled scan

Figure 3 shows the workflow of the scheduled scan in OneFS. The scheduled scan is triggered by the job engine. Like other jobs in OneFS, the impact and schedule configuration can be set accordingly. It also implements the filter, which is slightly different from the filter in the on-demand scan. The detailed filtering criteria are listed below:

- File extension to include
- File extension to exclude
- File path to exclude
- File path to include (only in the scheduled scan)

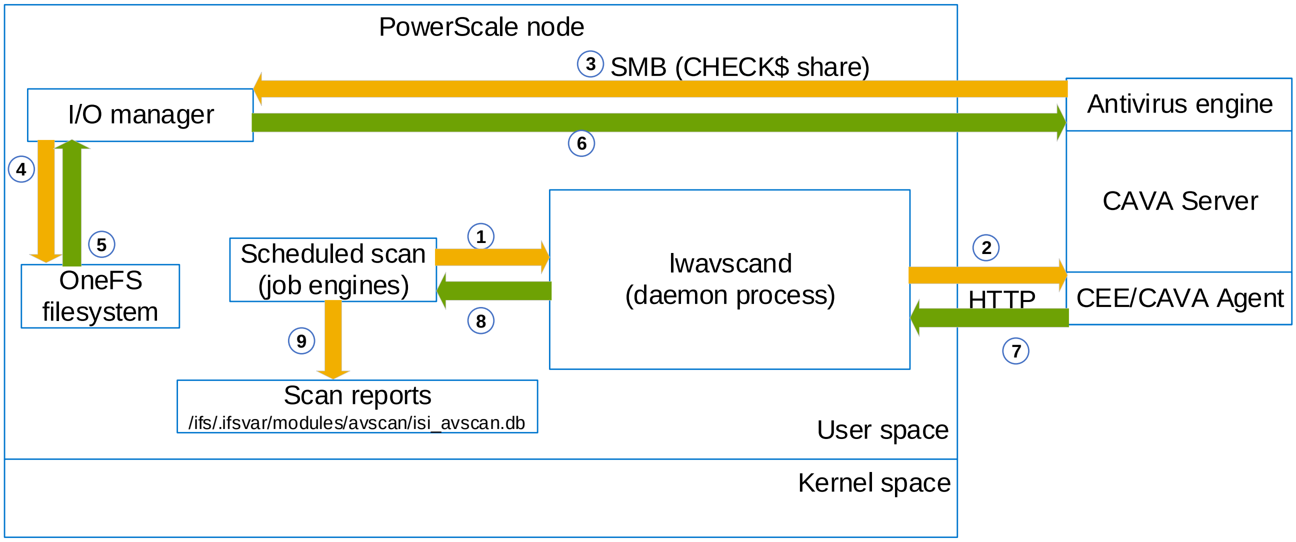


Figure 3 The workflow of the scheduled scan

If the file matches all the criteria, an internal process **lwavscand** sends the HTTP scanning request to the CEE/CAVA agent, and simultaneously, OneFS sets the locks on this file. Then, the anti-virus application attempts to fetch this file through a hidden SMB share **CHECK$** from the file system in OneFS. CAVA supports only downloading a part of s file for scanning, and this helps the performance. After the corresponding content is downloaded to the CAVA server, it runs the scan with the anti-virus engine, and CEE sends the scan results and response back to the process **lwavscand**. All the scanning attributes are recorded under **/ifs/.ifsvar/modules/avscan/isi_avscan.db.**

1.1.4 Manual scan

Figure 4 shows the workflow of a manual scan. The manual scan is triggered by the CLI, PAPI, or web interface. Unlike an on-demand scan or scheduled scan, it has no filters. The scanning files are directly assigned by the parameters. The daemon process **lwavscand** sends the HTTP scanning request to the CEE or CAVA agent, and simultaneously, OneFS sets the locks on this file. Then, the anti-virus application attempts to fetch this file through a hidden SMB share **CHECK$** from the file system in OneFS. CAVA supports only downloading a part of the file for scanning, and this helps the performance. After the corresponding content has been downloaded to the CAVA server, it runs the scan with the anti-virus engine, and CEE sends the scan results and respond to the process **lwavscand**. All the scanning attributes are recorded under **/ifs/.ifsvar/modules/avscan/isi_avscan.db.**

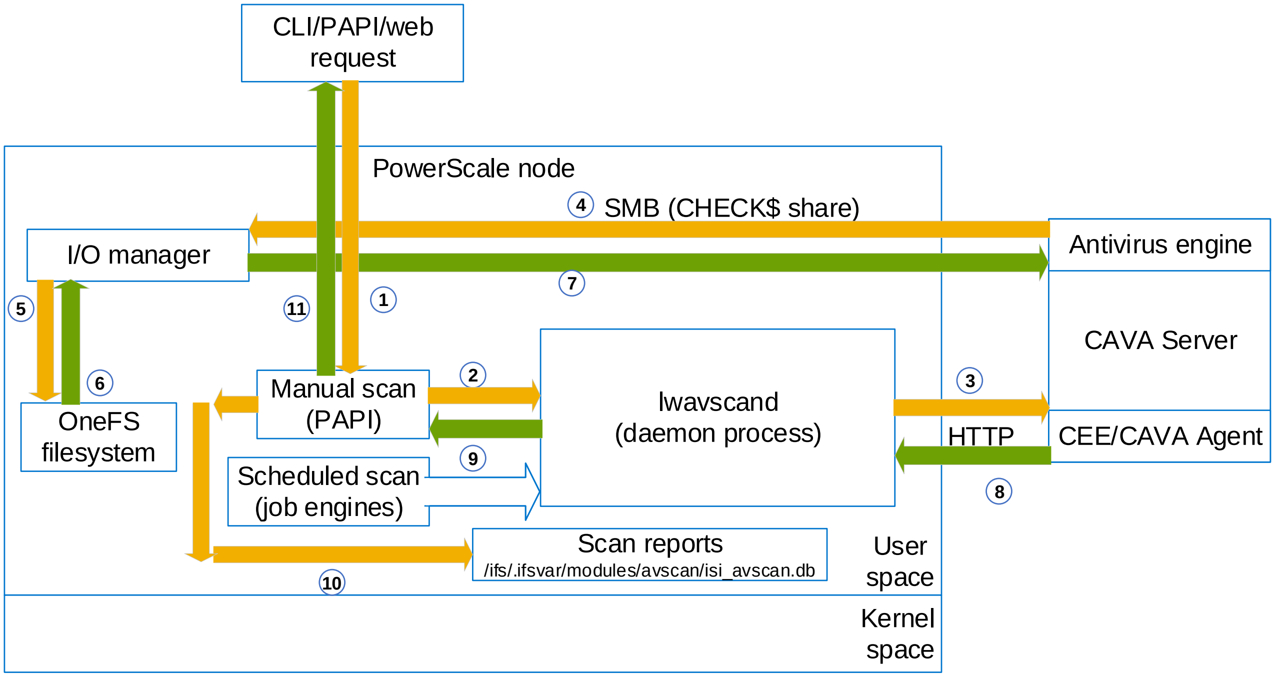


Figure 4 The workflow of manual scan

1.2 Load balance

There are two types of connections between CAVA servers and PowerScale nodes as listed below:

- SMB connection to fetch files or contents for scanning through a hidden share CHECK$.
- CEE connections for scan requests, scan responses, and other functions. These connections are HTTP connections which are shown in the Figure 4.

For each type of connection, OneFS has a different implementation for load balancing. The following sections explain these two methodologies in detail.

1.2.1 SMB connection load balancing

OneFS uses a dedicated IP pool and access zone to segregate all CAVA SMB connections from other workloads. Within this IP pool, SmartConnect is enabled to ensure all SMB connections are evenly spread across all the nodes.

1.2.2 CEE connection load balancing

The rules and conditions of how CEE connections are balanced are listed below:

- Maximum connections per CAVA servers = 20
- Number of different CAVA servers a cluster node can connect = 4
- The nth cluster node starts from nth CAVA server

The following example uses an environment with the following:

- 7 PowerScale cluster nodes
- 7 CAVA servers

The calculation is as follows:

$$Maximum\ connections = Maximum\ conections\ per\ CAVA\ server \times number\ of\ CAVA\ servers = 20 \times 7 = 140$$

$$Maximum\ connections\ per\ cluster\ node = \frac{Maximum\ conncetions}{Cluster\ node\ number} = \frac{140}{7} = 20$$

$$Maximum\ connections\ per\ node\ per\ CAVA\ server = \frac{Maximum\ connections\ per\ cluster\ node}{4} = 5$$

The layout for the connections is listed in Table 1.

Table 1 Connection layouts

	First CAVA Server	**Second CAVA Server**	**Third CAVA Server**	**Fourth CAVA Server**	**Fifth CAVA Server**	**Sixth CAVA Server**	**Seventh CAVA Server**
First node	5	5	5	5			
Second node		5	5	5	5		
Third node			5	5	5	5	
Fourth node				5	5	5	5
Fifth node	5				5	5	5
Sixth node	5	5				5	5
Seventh node	5	5	5				5
Total	**20**	**20**	**20**	**20**	**20**	**20**	**20**

A more complicated example has an environment with the following:

- 7 PowerScale cluster nodes
- 5 CAVA servers

The calculation is as follows:

$$Maximum\ connections = Maximum\ conections\ per\ CAVA\ server \times number\ of\ CAVA\ servers = 20 \times 5 = 100$$

$$Maximum\ connections\ per\ cluster\ node = \frac{Maximum\ conncetions}{Cluster\ node\ number} = \frac{100}{7} \approx 14$$

$$Maximum\ connections\ per\ node\ per\ CAVA\ server = \frac{Maximum\ connections\ per\ cluster\ node}{4} = \frac{15}{4} \approx 4$$

The layout for the connections is listed in the Table 2.

Table 2 Connection layouts

	First CAVA Server	**Second CAVA Server**	**Third CAVA Server**	**Fourth CAVA Server**	**Fifth CAVA Server**
First node	4	4	4	2	
Second node		4	4	4	2
Third node	2		4	4	4
Fourth node	4	2		4	4
Fifth node	4	4	2		4
Sixth node	4	4	4	2	
Seventh node		2	2	4	4
Total	**18**	**20**	**20**	**20**	**18**

1.3 Supported vendors

Theoretically, OneFS should support all the anti-virus vendors supported by CEE or CAVA. For the detailed list, see Table 3. CEE regularly updates the table of the supported vendors and their versions. For the most up-to-date list, see the Common Event Enabler Release Notes.

Table 3 Supported vendors

Vendor	Version supported
McAfee® VirusScan	8.8i Patch13
McAfee EndPoint Protection	10.7.0 Update July 2020
Symantec® Protection Engine	8.0
Symantec Endpoint Protection	14.2
Microsoft SCEP	4.10.209.0
Microsoft Defender	4.18.2004
F-Secure ESS	12.12
Kaspersky® Security 10 for Windows Server	10.1.2
TrendMicro® ServerProtect for Storage	6.00 Patch 1
Sophos® Endpoint Security Control	10.8
Computer Associates eTrust	6.0

1.4 Supportability

See Table 4 for the supportability of this feature.

Table 4 Supportability

Supportability	Description
Protocol support	SMB
Legacy ICAP support	Yes
Snapshot scanning support	No
SyncIQ	Transferring anti-virus files attributes using SyncIQ will not be supported.
SmartLock files scanning	The files under SmartLock protection are read-only. OneFS cannot set scanning attributes on them, and in case they are infected, the anti-virus application cannot take proper action against them.

2 Deployment and configuration

2.1 Overview

Figure 5 shows the overall workflow of how CAVA works in OneFS. The anti-virus applications use the SMB protocol to fetch the file or a portion of a file for scanning in a PowerScale cluster. From the anti-virus perspective, a hidden SMB share **CHECK$** is used for this purpose and resides on every anti-virus application server. This share allows access to all files on a PowerScale cluster under **/ifs**. SmartConnect and a dedicated access zone are introduced in this process to ensure that all the connection from the anti-virus application is fully distributed and load-balanced among all the configured nodes in the IP pool. A hidden role **AVVendor** is created by the CAVA anti-virus service to map the **EMC CAVA** service account into OneFS.

Figure 5 also shows the overall steps to configure CAVA in OneFS:

1. Create CAVA servers in OneFS
2. Create IP pool
3. Create a dedicate access zone for CAVA (AvVendor)
4. Create Active Directory authentication provided in the access zone (AvVendor)
5. Update role (AVVendor)

The following sections introduce these steps in detail.

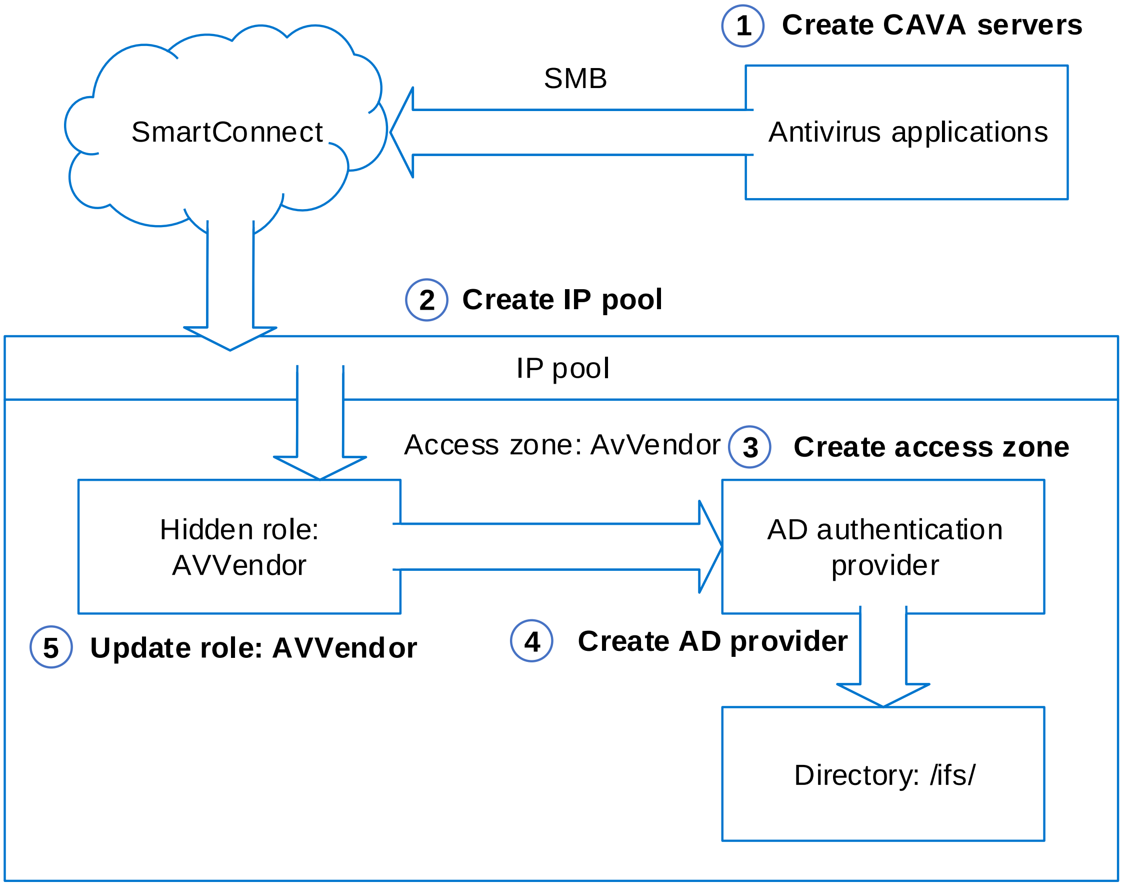


Figure 5 Workflow showing how CAVA works in OneFS

The prerequisites to configure CAVA service in OneFS are listed in Table 5.

Table 5 Prerequisites to configure CAVA service in OneFS

Prerequisites	Description
SMB service	SMB service on OneFS should be enabled to ensure anti-virus applications can fetch data from PowerScale cluster for scanning.
SmartConnect Service IP (SSIP)	SSIP should be configured in the OneFS subnet level. SmartConnect is used by the CAVA service to ensure all the scanning requests are well balanced among all the nodes in the IP pool.
Anti-virus application and Dell EMC Comment Event Enabler (CEE)	See the document from the anti-virus software vendor and Using the Common Event Enabler on Windows Platforms.
Active Directory access	CAVA service in OneFS requires the anti-virus application and PowerScale cluster are in the same domain.
Dedicated IP addresses	All the connections from anti-virus applications are served by a dedicated IP pool in PowerScale. These IP addresses are used to configure the IP ranges in this IP pool. We recommend using exclusive IP addresses which are only available to the anti-virus applications.

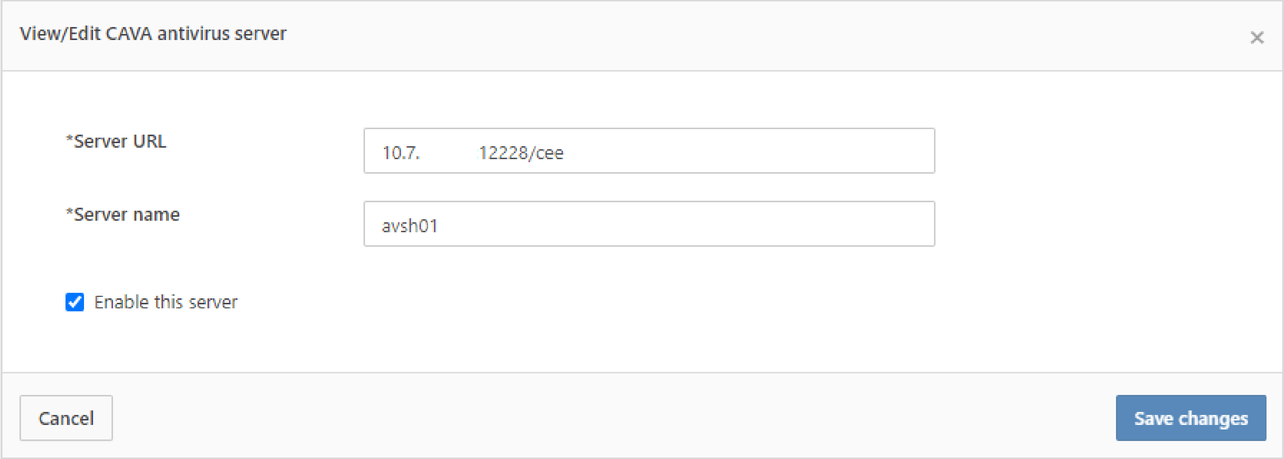
2.2 Create CAVA server in OneFS

Installation of the anti-virus application and CEE is a prerequisite as shown in Table 5. During the configuration of CEE, a domain user is created. This domain user is the service account for Windows service **EMC CAVA** and is used to access the hidden SMB share **CHECK$** to get the files and content for scanning. In this example, the user is **LORG\cavausr.**

After the anti-virus servers are installed and configured, create the corresponding entries in the OneFS so that PowerScale is aware of them. To perform this action, use the following command:

```
# isi antivirus cava servers create --server-name=avsh01 --server-uri=10.7.a.b -
-enabled=1
```

Alternately, use the WebUI as shown in Figure 6.



View/Edit CAVA antivirus server

*Server URL: 10.7.   12228/cee

*Server name: avsh01

☑ Enable this server

Figure 6 Add CAVA server

You can add multiple CAVA servers into OneFS configuration for the proper number of the CAVA servers for a given PowerScale cluster (refer to section 4 about sizing).

2.3 Create an IP Pool

Note: Before performing the following steps, ensure the **Service Enabled** is set to **No** in the CAVA settings.

The purpose of creating an IP pool is to facilitate the connections from anti-virus applications. The dedicated IP pools should only be used by the anti-virus applications. To achieve that result, we recommend that the IP ranges in this IP pool are exclusive and only available to the CAVA servers.

Note: Do not mix the IP range in this dedicated IP pool with others for a regular SMB client connection.

The load balancing for the anti-virus workload is achieved by the SmartConnect zone in this IP pool. Since this is a dedicated IP pool for CAVA servers, all the anti-virus scanning workload should be evenly distributed within the pool. To do this action, use the following CLI command:

```
# isi network pools create groupnet0.subnet0.pool1 --ranges=10.7.u.v-10.7.x.y --
sc-dns-zone "cavacluster1.west.isilon.com" --ifaces=1:ext-1
```

In this example:

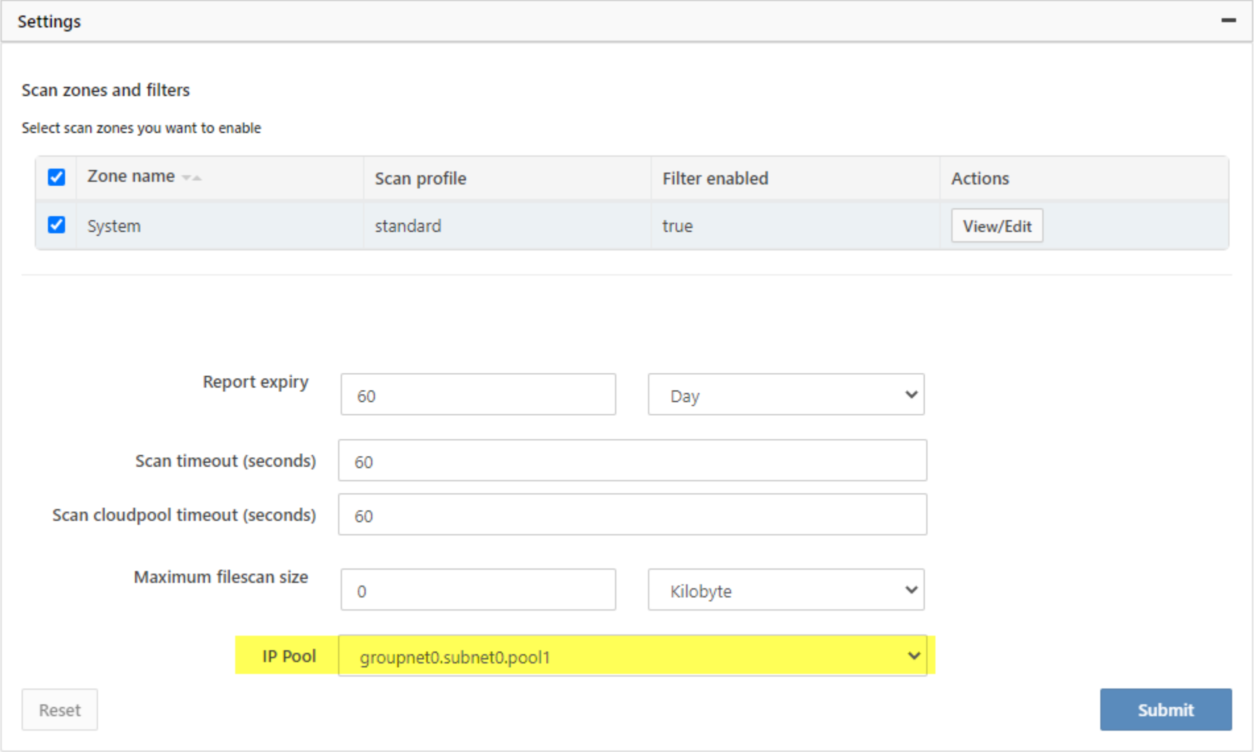
- The IP pool is groupnet0.subnet0.pool1
- The IP range is 10.7.u.v to 10.7.x.y
- The SmartConnect Zone name is cavacluster1.west.isilon.com and make sure the DNS delegion is crated for it
- The network interface is the ext-1 from node 1

After the IP pool is created, associate it with the CAVA configurations by using the following command. After this command, this IP pool is only available to the CAVA servers, and simultaneously, the corresponding access zone of this IP pool are changed to AvVendor.

```
# isi antivirus cava settings modify --ip-pool="groupnet0.subnet0.pool1"
This action will make the IP Pool unavailable to all other users except
antivirus servers. Do you want to continue? (yes/[no]): yes
"
IP Pool groupnet0.subnet0.pool1 added to CAVA antivirus.
Note: The access zone of IP Pool groupnet0.subnet0.pool1 has been changed to
AvVendor.

"
```

This function is also available through the WebUI as shown in Figure 7.



Settings

Scan zones and filters

Select scan zones you want to enable

☑	Zone name	Scan profile	Filter enabled	Actions
☑	System	standard	true	View/Edit

Report expiry: 60 Day

Scan timeout (seconds): 60

Scan cloudpool timeout (seconds): 60

Maximum filescan size: 0 Kilobyte

IP Pool: groupnet0.subnet0.pool1

Figure 7 Associate IP pool with CAVA settings

Note: Be sure to create the DNS delegation for the zone name associated with IP pool created in this step.

2.4 Create a dedicated access zone: AvVendor

A dedicated access zone **AvVendor** is connected to the IP pool that was created in section 2.3. This access zone is automatically created when the CAVA service is enabled in the PowerScale cluster. To enable the CAVA service, use the following command:

```
# isi antivirus cava settings modify --service-enabled=1
```

View the CAVA settings and ensure the **Server Enable** setting is **Yes**.

```
# isi antivirus cava settings view
       Service Enabled: Yes
     Scan Access Zones: System
               IP Pool: groupnet0.subnet0.pool1
         Report Expiry: 8 weeks, 4 days
          Scan Timeout: 1 minute
Cloudpool Scan Timeout: 1 minute
     Maximum Scan Size: 0.00kB
```

At the current stage, AvVendor is created in the access zone list:

```
# isi zone zones list
Name     Path
--------------
System   /ifs
AvVendor /ifs
--------------
Total: 2
```

Note: The ID of the access zone **AvVendor** is **-2**.

2.5 Create an Active Directory authentication provider for AvVendor

All the anti-virus application servers and PowerScale cluster should be in the same domain. Use the following CLI command to join the PowerScale cluster into the domain. In this example, the domain name is **lorg.west.isilon.com.**

```
# isi auth ads create lorg.west.isilon.com --user administrator
```

Then, add the authentication provider to the access zone AvVendor:

```
# isi zone zones modify AvVendor --add-auth-providers=lsa-activedirectory-
provider:lorg.west.isilon.com
```

2.6 Update role: AVVendor

The CHECK$ share is a hidden share that allows access to all files on the cluster. It is used exclusively by the anti-virus software running on a Microsoft Windows® server. Since the CHECK$ share allows access to all files on the cluster, any user accessing the share must have a unique privilege. The hidden **ISI_PRIV_AV_VENDOR** privilege in the AVVendor role is added to give the user account that is used by the anti-virus software access to the CHECK$ share.

Use the following CLI command to assign the user **LORG\cavausr** to the role **AVVendor** in the access zone **AvVendor**:

```
# isi auth roles modify AVVendor --zone=AvVendor --add-user lorg\\cavausr
```

At the current stage, the configuration for the CAVA service on the PowerScale is finished. Ensure the **System Status** is **RUNNING** by using the following command:

```
vshen-p21evix-1# isi antivirus cava status
        System Status: RUNNING
        Fault Message: -
          CEE Version: 8.7.7.0
          DTD Version: 2.3.0
            AV Vendor: Symantec
```

3 Performance overview

In general, CAVA has better performance than ICAP. In the following sections, a detailed report is explained to introduce and compare the performance between CAVA and ICAP in a PowerScale cluster.

The performance results may vary depending on the PowerScale model, total number of nodes, anti-virus application vendors, size of the CAVA or ICAP servers, workload, file size distribution, and other factors. The following example only gives an overall idea of what the CAVA performance looks like and how it compares with ICAP.

The following sections cover the following:

- Test environment
- Scenario 1: Performance of scan on close
- Scenario 2: Performance of scan on read
- Scenario 3: Performance impact to SWBUILD

3.1 Test environment

Table 6 shows the components in the test environment for scenario 1 and scenario 2:

Table 6 Test environment for the scenario I and II

Components	Number of the components	Details
PowerScale cluster	4-node cluster	• H500
CAVA servers	3 virtual machines	• 8 vCPUs, 16 GB memory • Symantec Endpoint Protection 14.2
Clients	3 physical servers	To generate SMB workload: • For scenario 1, these servers generate create workloads. • For scenario 2, these servers generate read workloads.

Table 7 shows the components in the test environment for scenario 3:

Table 7 Test Environment for the scenario III

Components	Number of the components	Details
PowerScale cluster	4-node cluster	• H600
CAVA servers	4 hosts	• In this scenario, we use a custom script to simulate what anti-virus applications could do to make the test cases more generic.
Clients	4 servers	• To generate SMB workload

3.2 Scenario 1: Performance of scan on close

3.2.1 Test methodology

In this scenario, the client servers generate different sizes of files which are 10 KB, 100 KB, 1 MB, 10 MB, 100 MB, and 200 MB. From PowerScale perspective, a standard scan profile is applied to ensure the scan occurs when the file is closed. Regarding the detailed CAVA configuration on PowerScale, see the following profile:

```
vshen-p21evix-1# isi antivirus cava filters view System
                 Zone: System
              Enabled: Yes
         Open-on-fail: Yes
      File Extensions: *
File Extension Action: include
 Scan If No Extension: No
        Exclude Paths: -
         Scan-profile: standard
         Scan-on-read: No
        Scan-on-close: Yes
       Scan-on-rename: Yes
 Scan Cloudpool Files: No
```

3.2.2 Test results

Figure 8 shows how the number of files scanned per second changes along with the increase of the average file size. With the increased size of the files, the overall trend of the files scanned per second is decreased.

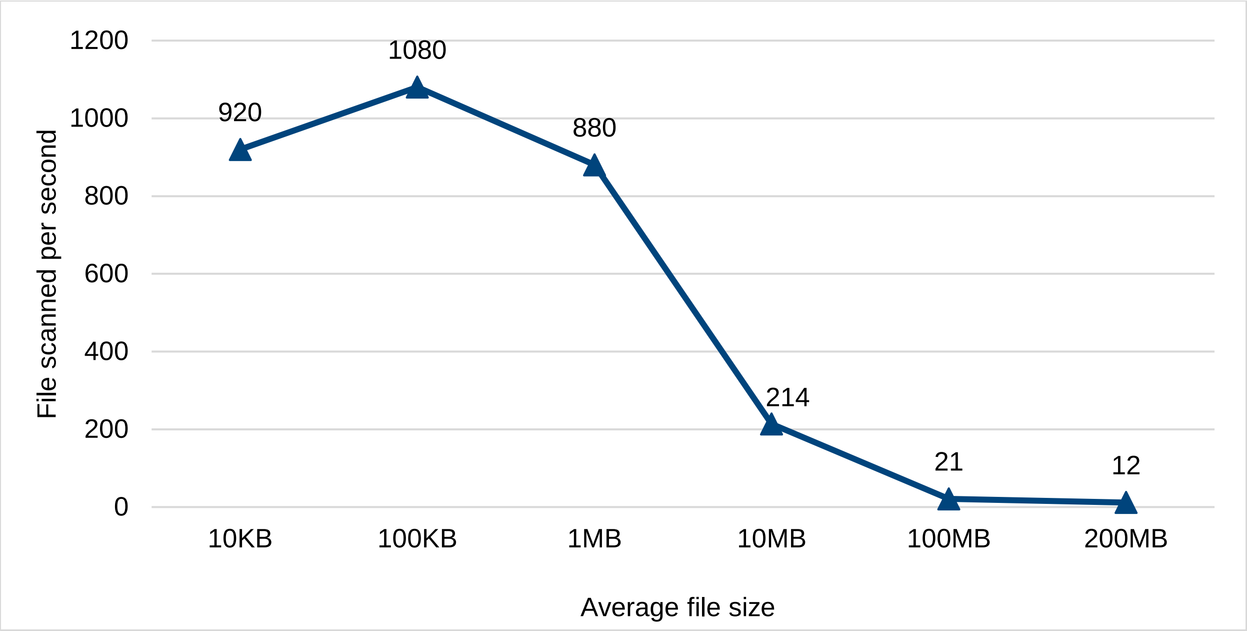


Figure 8 File scanned per second (scan-on-close)

Figure 9 shows the relationship between average file size and the scanning throughput. Theoretically, the results equal the files scanned per second times the average file size in each category. With the increase of file size, the scanning throughput increases. In this case, after it reaches 2 GB/sec, it goes flat.

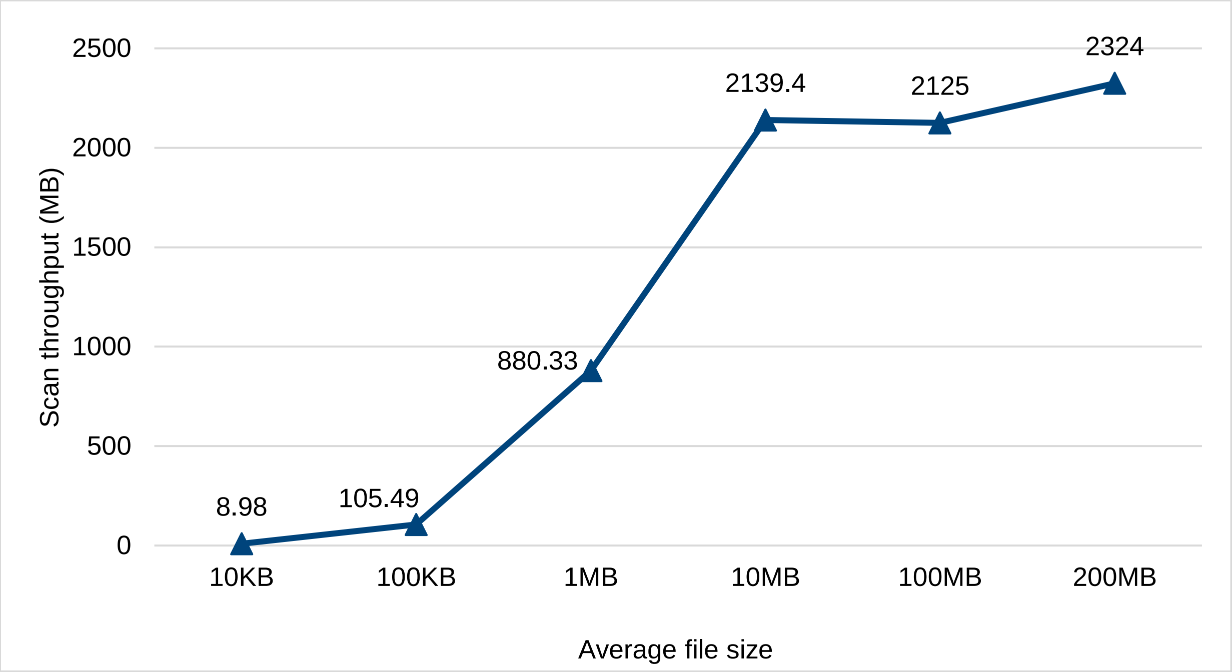


Figure 9 Scan throughput (scan-on-close)

The average CPU utilization on the PowerScale cluster is stable under 8% no matter the overall scan throughput or various average file sizes. For more details, see Figure 10.

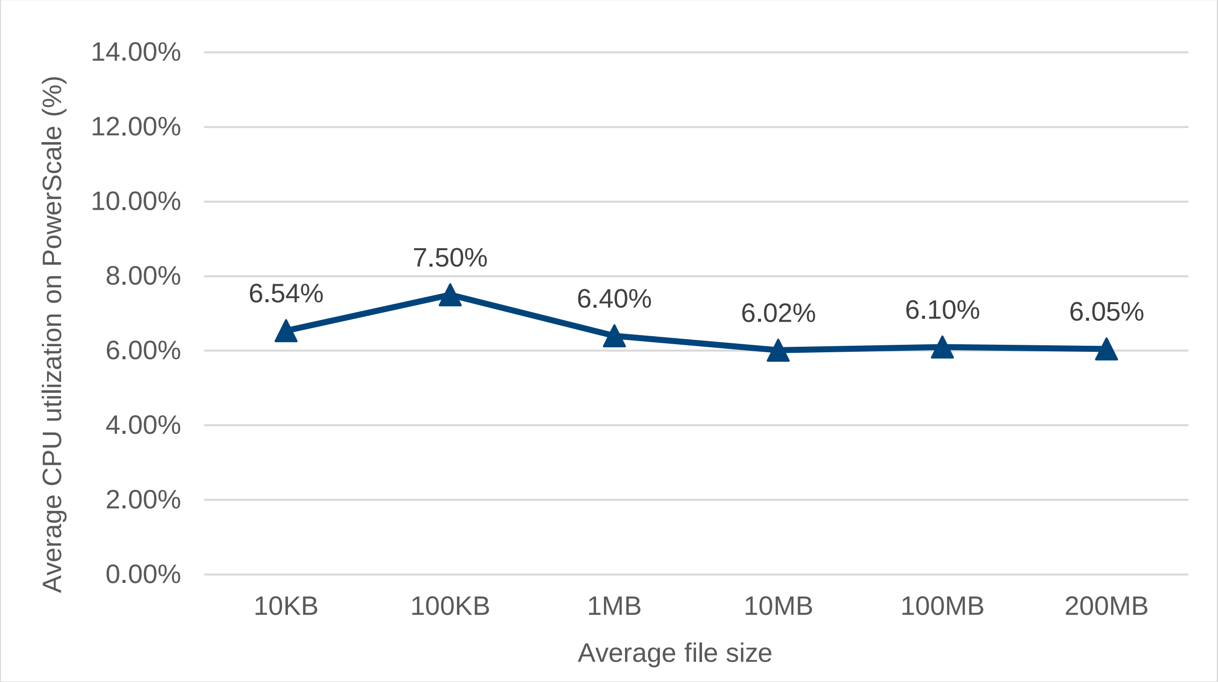


Figure 10 Average CPU utilization on PowerScale (scan-on-close)

3.3 Scenario 2: Performance of scan on read

3.3.1 Test methodology

In this scenario, the client servers generate read requests against the files which have been generated in scenario 1. The CAVA strict profile is applied to ensure **scan-on-read** is enabled. Regarding the detailed CAVA configuration on PowerScale, see the following profile:

```
vshen-p21evix-1# isi antivirus cava filters view System
                 Zone: System
              Enabled: Yes
         Open-on-fail: Yes
      File Extensions: *
File Extension Action: include
 Scan If No Extension: No
        Exclude Paths: -
         Scan-profile: strict
         Scan-on-read: Yes
        Scan-on-close: Yes
       Scan-on-rename: Yes
 Scan Cloudpool Files: No
```

3.3.2 Test results

Figure 11 illustrates how the number of files scanned per second changes along with the increase of the average file size. With the increased size of the files, the overall trend of the files scanned per second is decreased.

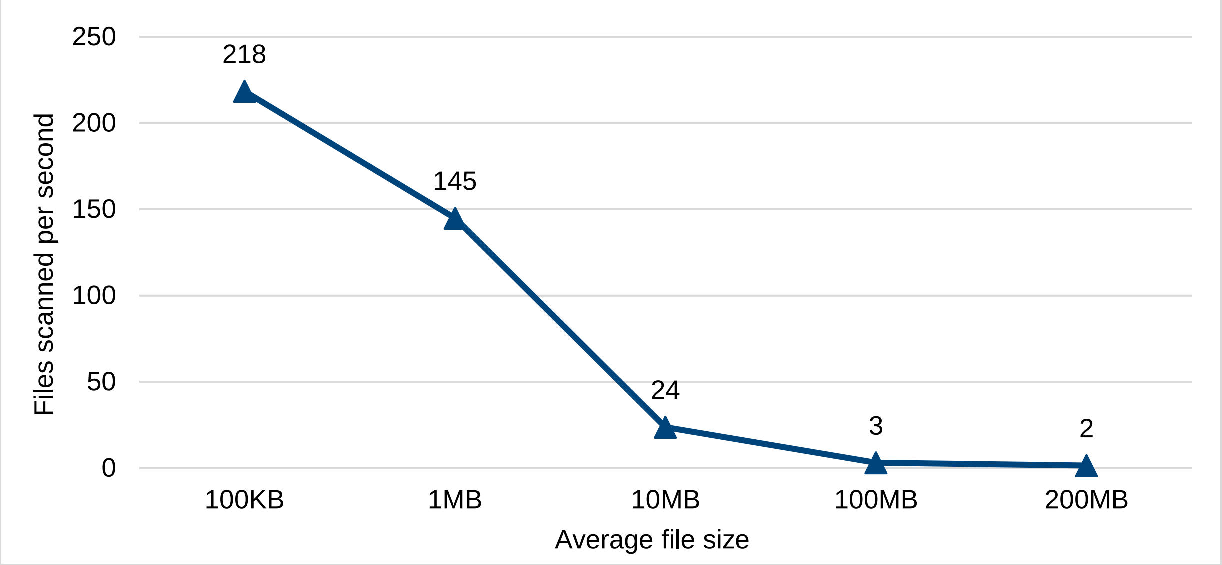


Figure 11 Files scanned per second (scan-on-read)

Figure 12 shows the relationship between the average file size and scanning throughput. Theoretically, the results equal the files scanned per second times the average file size in each category. With the increase of file size, the scanning throughput increases. In this case, after it reaches 300 MB/sec, it goes flat. Compared with the results in scenario 1, it is obvious that the scan efficiency is much higher than the outcome in this scenario. Scan-on-read can provide better protection again viruses, however, the scanning efficiency is lower than scan-on-close. We recommend carefully selecting the scan profile to match your business requirements and performance level.

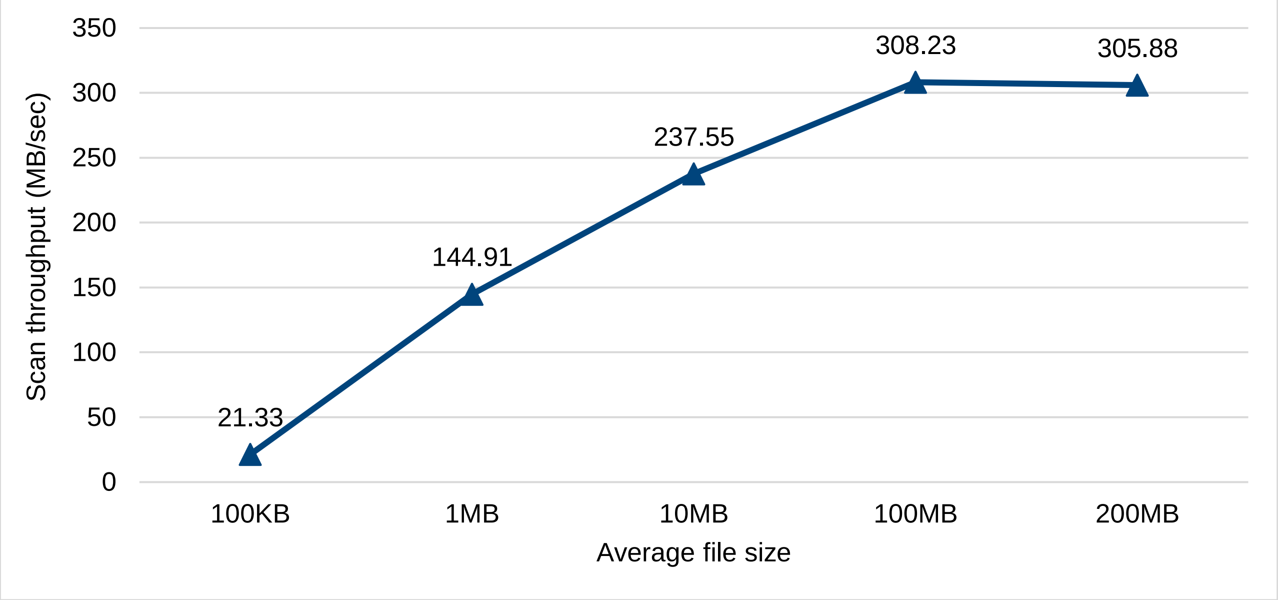


Figure 12 Scan throughput (scan-on-read)

The average CPU utilization on PowerScale cluster is stable around 6% no matter the overall scan throughput or various average file size. For more details, see Figure 13.

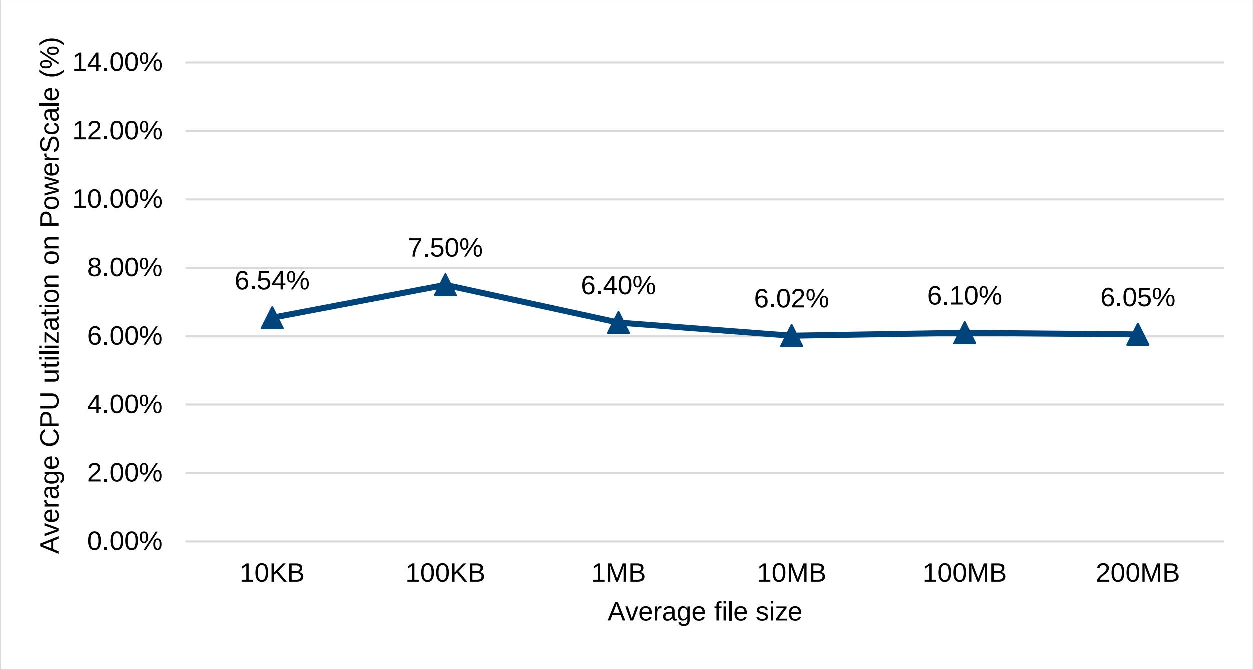


Figure 13 Average CPU utilization on PowerScale (scan-on-read)

3.4 Scenario 3: Performance impact on SWBUILD

3.4.1 Test methodology

In this scenario, we use SPEC SFS 2014 to generate and simulate the real-world workloads with **SWBUILD**. We included the following subscenarios to better understand the performance impact from the CAVA workload:

- Baseline: There is no CAVA workload, and the only workload running on the cluster is SWBUILD from SPEC SFS 2014.
- Path excluded: The path for SWBUILD workload has been excluded from the CAVA configuration.
- Standard profile: Scan-on-close, scan-on-rename.
- Strict profile: Scan-on-close, scan-on-rename, scan-on-open.

Note: For comparison, all four subscenarios use the same requested operation rate that is configured in SPEC SPS 2014.

3.4.2 Test results

Figure 14 shows the results of the achieved operations per second in the four subscenarios defined in the previous sections. The standard scanning profile, or if we exclude the workload path from the CAVA configuration, has almost no performance impact compared with the baseline result. The standard profile implements scan-on-close, which means any file newly written or modified is submitted for virus scanning in the background after it is closed by the client. Scan-on-close is not synchronous in the host I/O path, and it slows down the incoming workload. In this case, the incoming workload proceeds unabated.

In the case of the strict profile, the performance drops by 34%. This is because the strict profile adds scan-on-read; any file not previously scanned, when opened for read by a client, is submitted for scanning at that time and it is synchronously in the host I/O path.

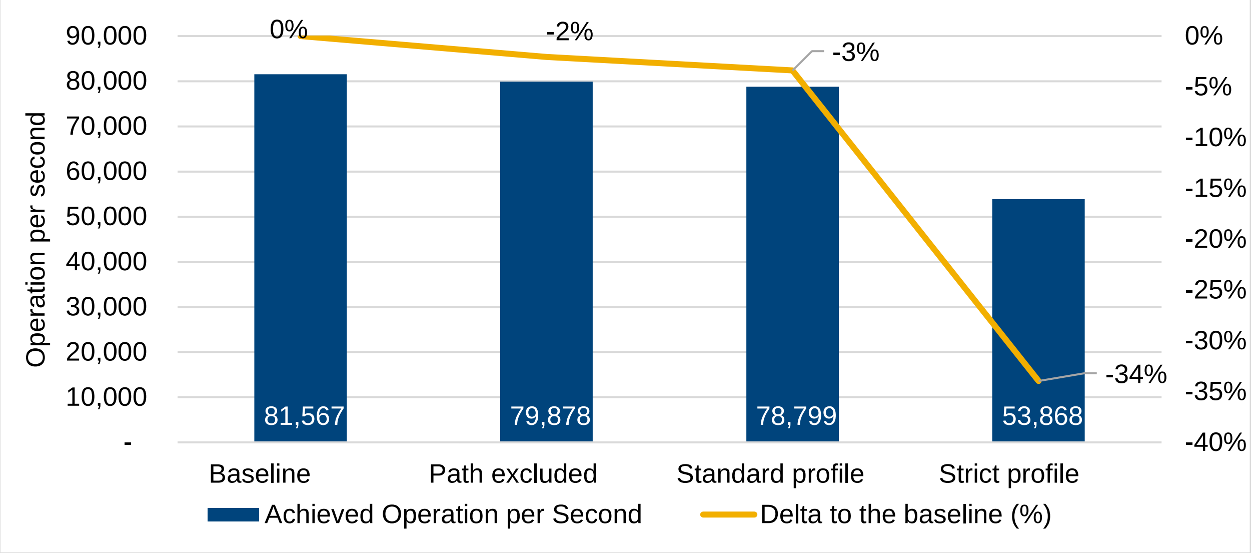


Figure 14 Achieved operation per second

Figure 15 shows the average latency in the four subscenarios defined in the previous sections. The standard scanning profile, or if we exclude the workload path from the CAVA configuration, has almost the same latency compared with the baseline. In the strict profile, the latency increases by 52%.

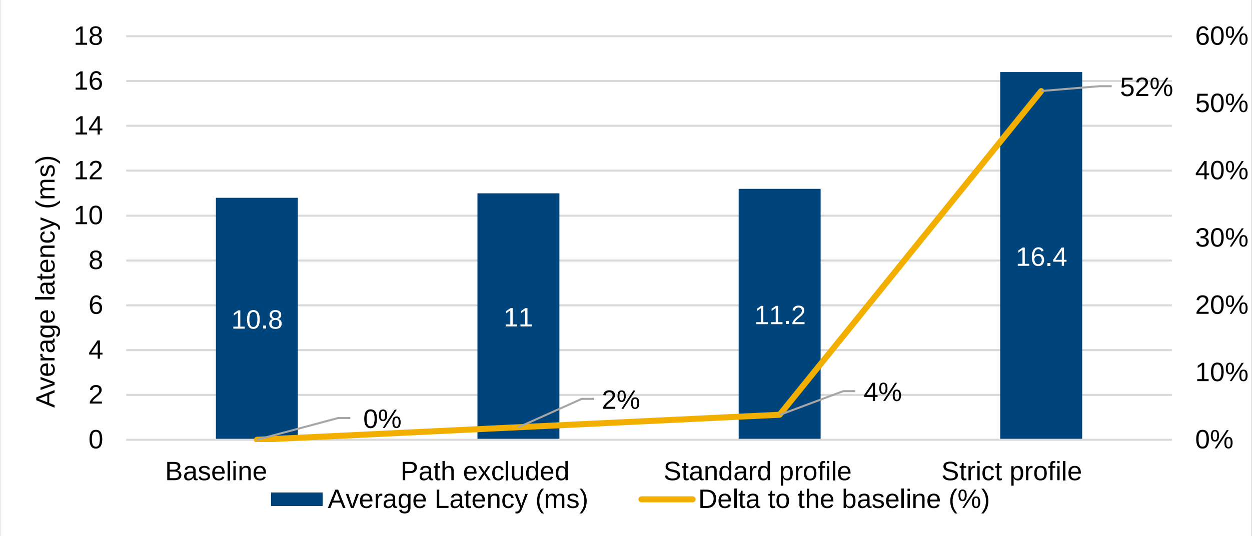


Figure 15 Average latency

4 Sizing

4.1 Overview

In this section, two sizing methodologies are introduced for sizing the number of CAVA servers with a given PowerScale cluster. Both methodologies can only come out with an essential number of the CAVA servers. For more granular sizing which is based on various workloads, see the CEE CAVA calculator described in Using the Common Event Enabler on Windows Platforms.

4.2 General best practices for sizing

No matter which methodology to use, we recommend using the following general best practices for sizing:

- Use Windows machines with at least 16 GB memory and two-core processors.
- Use a minimum of two CEE or CAVA machines for redundancy.

4.3 Methodology 1

Methodology 1 only requires having the cluster node number as the known condition. The number of CAVA servers follows the following algorithm.

$$\text{The number of CAVA servers} = \frac{\textit{The number of nodes in a PowerScale cluster}}{4}$$

For example, in a 22 node PowerScale cluster, The number of CAVA servers is:

$$\text{The number of CAVA servers} = \frac{22}{4} = 5.5 \approx 6$$

In this case, we need six CAVA servers.

Methodology 1 only provides an essential calculation that is based on the total number of nodes as the known condition. This methodology suits both the on-demand scan and scheduled scan. For some scenarios where customers do not have details for the I/O pattern of their workflow, they can use this methodology for sizing.

4.4 Methodology 2

Methodology 2 is much more granular than the methodology 1 since it accounts for the various workloads. This methodology only applies to the on-demand scan due to the sizing and because expected performance can vary based on different job performance impact levels.

The assumption for this methodology is as follows:

$$\text{Typical antivirus scan roundtrip time (RTT) for each CAVA connection} = 40\text{ms}$$

The known condition for this methodology is as follows:

$$\text{Define the total number of the requests per second which will trigger the on demand scan as R}$$

$$\text{Maximum connections per CAVA servers} = 20$$

The algorithm is as follows:

$$\text{Total scans can be done per second for each CAVA connection} = \frac{1000}{40} = 25$$

$$\text{The number of CAVA connections required } = \frac{R}{25}$$

$$\text{Minimum number of CAVA servers required} = \frac{The\ number\ of\ CAVA\ connections\ required}{Maximum\ connections\ per\ CAVA\ server} = \frac{R}{25 * 20} = \frac{R}{500}$$

In the following example, the scan profile is **standard**, the SMB close requests are **1,000 operations per second**, and the SMB rename requests are **200 operation per second**. The number of CAVA servers required is as follows:

$$Minimum\ number\ of\ CAVA\ servers\ required = \frac{1000 + 200}{500} = \frac{1200}{500} = 2.4 \approx 3$$

5 General considerations

5.1.1 NANON considerations

Not all nodes on network (NANON) is supported in CAVA configurations. More details are provided as follows:

- Scheduled scan: This works by default. In OneFS 9.1.0.0, an improved job engine has been implemented to detect a list of nodes not having connection to CAVA servers, and then inform job engine not to distribute scan jobs to these nodes. This means only the PowerScale nodes with the front-end connection to the CAVA servers will be triggered to run scheduled scan.
- On-demand scan or manual scan: As long as the PowerScale node which triggers the on-demand scan or manual scan has the front-end connection to the CAVA servers, it works. Otherwise, a CELOG event - SW_AVSCAN_CAVA_SERVER_OFFLINE is raised for notification.

The following is an example in which two IP pools are segregated.

Table 8 IP pools configuration

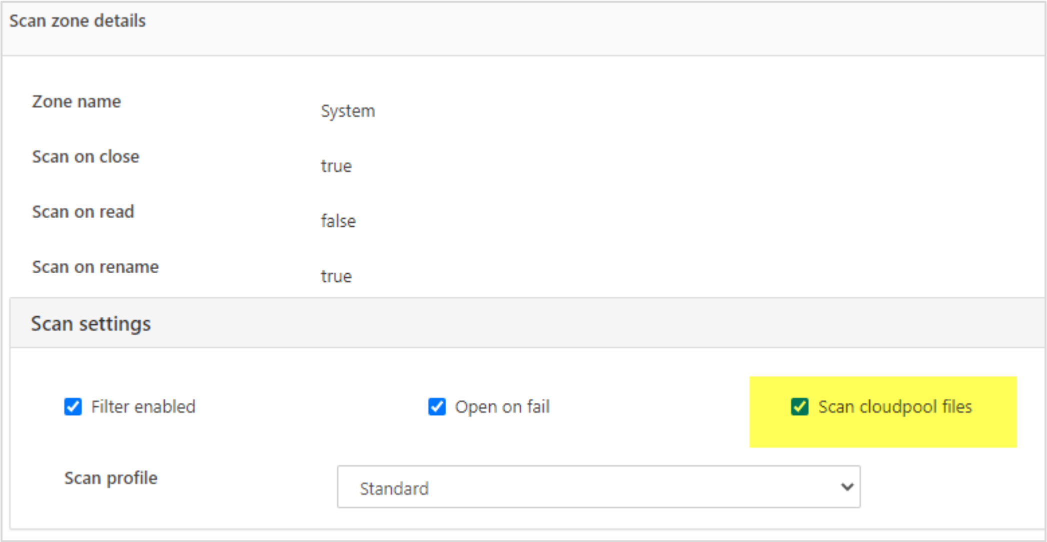
IP pools	Nodes connected
Pool for business workload	• Node 1 • Node 2 • Node 3 • Node 4
Pool for CAVA workload	• Node 1 • Node 2 • Node 3

In this example, the following behavior is expected:

- If the on-demand scan or manual scan is triggered from Node 1 to Node 3, it works.
- If the on-demand scan or manual scan is triggered from Node 4, it fails with CELOG event SW_AVSCAN_CAVA_SERVER_OFFLINE.
- The scheduled scan works properly.

5.1.2 CloudPools considerations

The CAVA configuration in OneFS supports scanning files in CloudPools. By default, scanning of CloudPools files is disabled to prevent the unexpected cost of file callback. To enable this setting, go to the scan zones settings in the CAVA configuration tab. Figure 16 shows the details.



Scan zone details

Zone name: System
Scan on close: true
Scan on read: false
Scan on rename: true

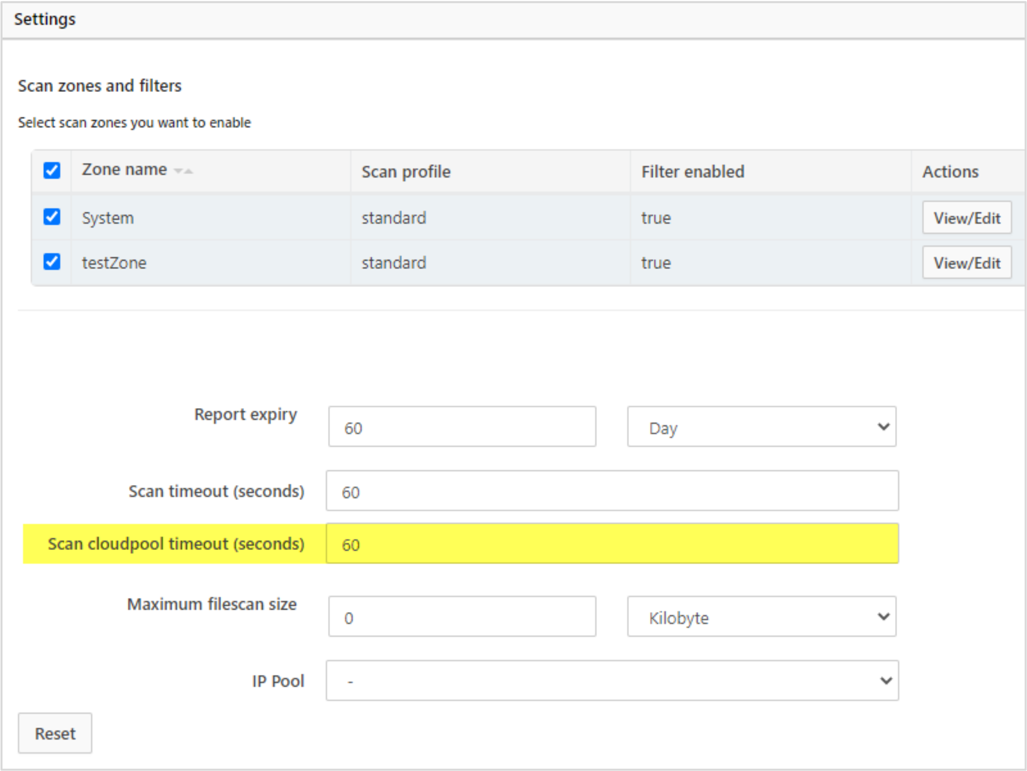
Scan settings

☑ Filter enabled ☑ Open on fail ☑ Scan cloudpool files

Scan profile: Standard

Figure 16 Enable scanning CloudPools files

OneFS has a separate, configurable scan timeout for CloudPools files, since scanning a CloudPools stub file may take more time than a regular on cluster file. To configure the timeout value for scanning CloudPools files, go to the settings section under CAVA configuration tab. See Figure 17 for details.



Settings

Scan zones and filters

Select scan zones you want to enable

☑	Zone name	Scan profile	Filter enabled	Actions
☑	System	standard	true	View/Edit
☑	testZone	standard	true	View/Edit

Report expiry: 60 Day
Scan timeout (seconds): 60
Scan cloudpool timeout (seconds): 60
Maximum filescan size: 0 Kilobyte
IP Pool: -

Figure 17 Configure timeout value for scan CloudPools files

When the stub file is read, in the kernel level, CloudPools fetches the file content and stores the data in the BCM cache. Since the anti-virus application is the first SMB client to read from the file, it triggers the file-block fetches.

For the scheduled scan, OneFS bypasses the BCM cache because it would continuously overflow the BCM cache, making it worthless. For the on-demand scan or manual scan, the BCM cache would not overflow as often, and the subsequent SMB client read would result in a cache hit.

5.1.3 SyncIQ consideration

Transferring anti-virus files attributes using SyncIQ is not supported. Files without those attributes are scanned again.

5.1.4 Anti-virus vendor considerations

Theoretically, OneFS should support all anti-virus vendors supported by CEE or CAVA, but some anti-virus applications may behave differently than others.

The details of the infected files are maintained the anti-virus-application level. We recommend checking log files once the infection is detected. For an infected file, OneFS denies all access to it, but the detailed action is taken by the anti-virus application. Consult your anti-virus vendors or documentation to understand how they deal with the infected files.

A Technical support and resources

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

Storage and data protection technical white papers and videos provide expertise that helps to ensure customer success with Dell EMC storage and data protection products.