EMC ScaleIO

Version 2.x

Performance Troubleshooting Technical Notes

P/N 302-003-503

REV 02

EMC²

Copyright © 2017 Dell Inc. or its subsidiaries All rights reserved.

Published September 2017

Dell believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

THE INFORMATION IN THIS PUBLICATION IS PROVIDED "AS-IS." DELL 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. USE, COPYING, AND DISTRIBUTION OF ANY DELL SOFTWARE DESCRIBED IN THIS PUBLICATION REQUIRES AN APPLICABLE SOFTWARE LICENSE.

Dell, EMC, and other trademarks are trademarks of Dell Inc. or its subsidiaries. Other trademarks may be the property of their respective owners. Published in the USA.

EMC Corporation
Hopkinton, Massachusetts 01748-9103
1-508-435-1000 In North America 1-866-464-7381
www.EMC.com

CONTENTS

Figures **5**

Preface **7**

Chapter 1 **Overview** **9**
Introduction 10

Chapter 2 **Basic Troubleshooting** **11**
Overall system performance 12
Retrieving volume-specific performance information 13
Retrieving host-specific performance information 13
Correlating between the SDC on an ESX server and ScaleIO volumes 14
Correlating between the SDS on a Linux server and ScaleIO volumes 16

Chapter 3 **Advanced Troubleshooting** **17**
Basic verification checks 18
ScaleIO network troubleshooting 18
Troubleshooting SDS connectivity 18
Troubleshooting SDC-to-SDS connectivity 19
Troubleshooting ScaleIO media 20

Chapter 4 **Example** **23**
Reduced IOPS capacity 24
Troubleshooting 24

FIGURES

1 SDS20 24
2 SDS21 25

Preface

As part of an effort to improve its product lines, EMC periodically releases revisions of its software and hardware. Therefore, some functions described in this document might not be supported by all versions of the software or hardware currently in use. The product release notes provide the most up-to-date information on product features.

Contact your EMC technical support professional if a product does not function properly or does not function as described in this document.

Note

This document was accurate at publication time. Go to EMC Online Support (https://support.emc.com) to ensure that you are using the latest version of this document.

Related documentation

The release notes for your version includes the latest information for your product.

The following EMC publication sets provide information about your ScaleIO or ScaleIO Ready Node product:

- ScaleIO software (downloadable as ScaleIO Software <version> Documentation set)
- ScaleIO Ready Node with AMS (downloadable as ScaleIO Ready Node with AMS Documentation set)
- ScaleIO Ready Node no AMS (downloadable as ScaleIO Ready Node no AMS Documentation set)
- VxRack Node 100 Series (downloadable as VxRack Node 100 Series Documentation set)

You can download the release notes, the document sets, and other related documentation from EMC Online Support.

Typographical conventions

EMC uses the following type style conventions in this document:

Bold	Used for names of interface elements, such as names of windows, dialog boxes, buttons, fields, tab names, key names, and menu paths (what the user specifically selects or clicks)
Italic	Used for full titles of publications referenced in text
`Monospace`	Used for: • System code • System output, such as an error message or script • Pathnames, filenames, prompts, and syntax • Commands and options
Monospace italic	Used for variables
`Monospace bold`	Used for user input
[ ]	Square brackets enclose optional values

|	Vertical bar indicates alternate selections - the bar means “or”
{ }	Braces enclose content that the user must specify, such as x or y or z
...	Ellipses indicate nonessential information omitted from the example

Where to get help

EMC support, product, and licensing information can be obtained as follows:

Product information

For documentation, release notes, software updates, or information about EMC products, go to EMC Online Support at https://support.emc.com.

Technical support

Go to EMC Online Support and click Service Center. You will see several options for contacting EMC Technical Support. Note that to open a service request, you must have a valid support agreement. Contact your EMC sales representative for details about obtaining a valid support agreement or with questions about your account.

Your comments

Your suggestions will help us continue to improve the accuracy, organization, and overall quality of the user publications. Send your opinions of this document to techpubcomments@emc.com.

CHAPTER 1

Overview

This chapter provides an introduction to ScaleIO performance troubleshooting. Topics include:

- Introduction 10

Introduction

ScaleIO is a software-only solution that uses existing servers' local disks and LAN to create a virtual SAN that has all the benefits of external storage, but at a fraction of the cost and complexity. ScaleIO utilizes the existing local internal storage and turns it into internal shared block storage. For many workloads, ScaleIO storage is comparable to, or better than, external shared block storage.

This document describes basic and advanced procedures for troubleshooting ScaleIO system performance.

CHAPTER 2

Basic Troubleshooting

This chapter provides basic procedures for troubleshooting ScaleIO performance. Topics include:

- Overall system performance 12
- Retrieving volume-specific performance information 13
- Retrieving host-specific performance information 13
- Correlating between the SDC on an ESX server and ScaleIO volumes 14
- Correlating between the SDS on a Linux server and ScaleIO volumes 16

Overall system performance

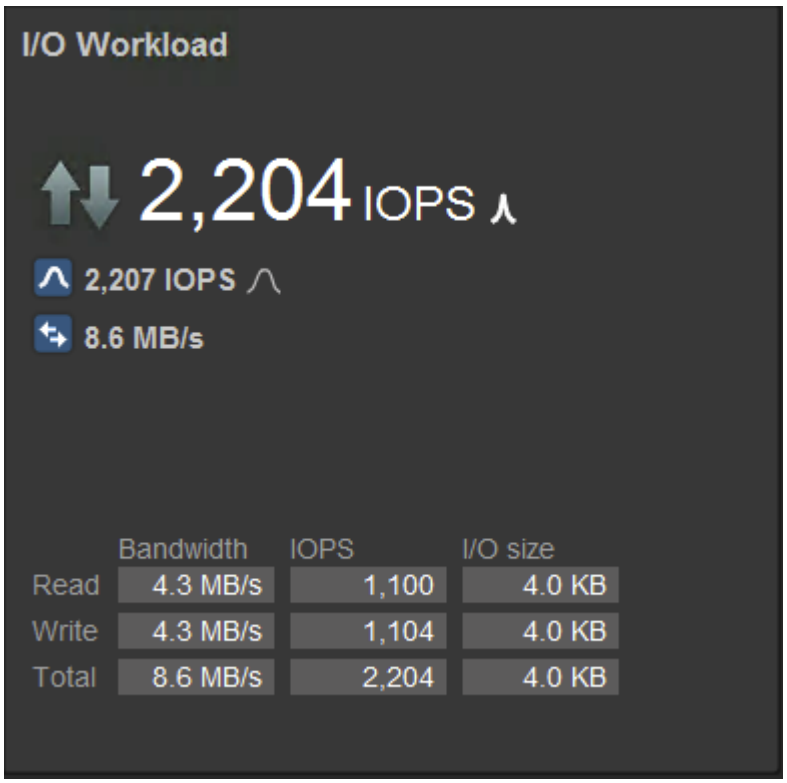
ScaleIO enables you to gather overall system performance numbers using these tools:

- ScaleIO GUI
- SCLI
- REST API

Note

For more information, see the "CLI and REST API command reference" in the *ScaleIO User Guide*.

When using the ScaleIO GUI, overall system numbers are available in the **Dashboard** view > **I/O Workload** tile:



The Dashboard displays the following information:

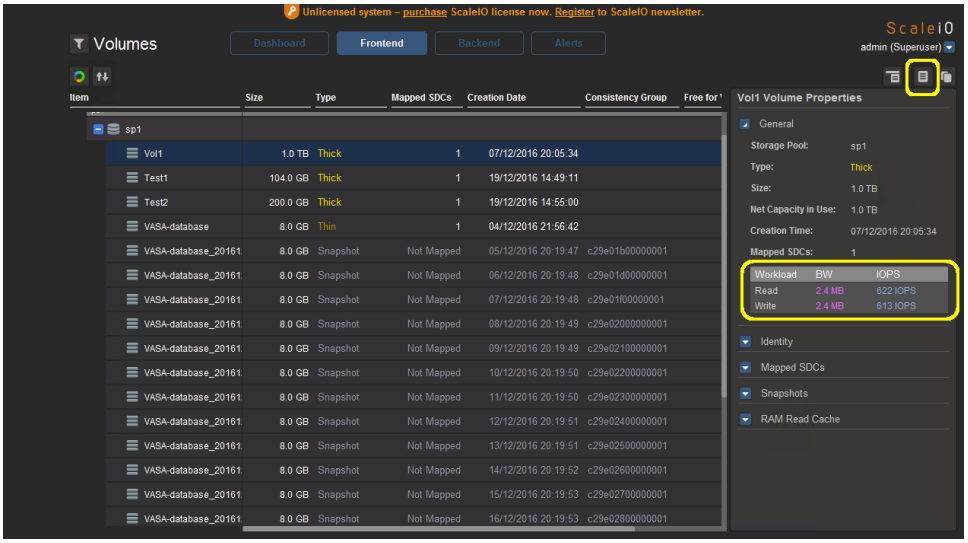
- Overall system IOPS
- Overall system bandwidth
- Read/write statistics
- Average I/O size

Retrieving volume-specific performance information

You can retrieve volume-specific information through the ScaleIO GUI or the SCLI.

Retrieving volume-specific information through the ScaleIO GUI

1. On the **Frontend** tab, select **Volumes** view.
2. Select a volume and open the property sheet.
   The volume information is displayed in the **General** section of the property sheet:

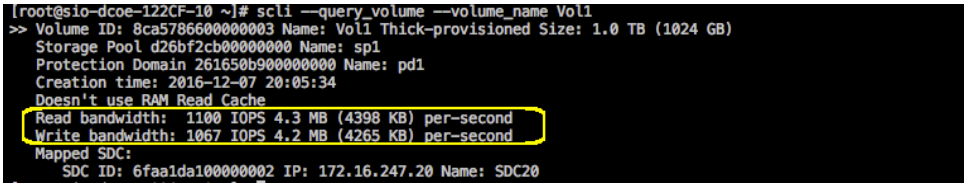


Retrieving volume-specific information through the SCLI

1. Run this command:

```
scli --query volume --volume_name <volume_name>
```

The same information appears as in the property sheet:

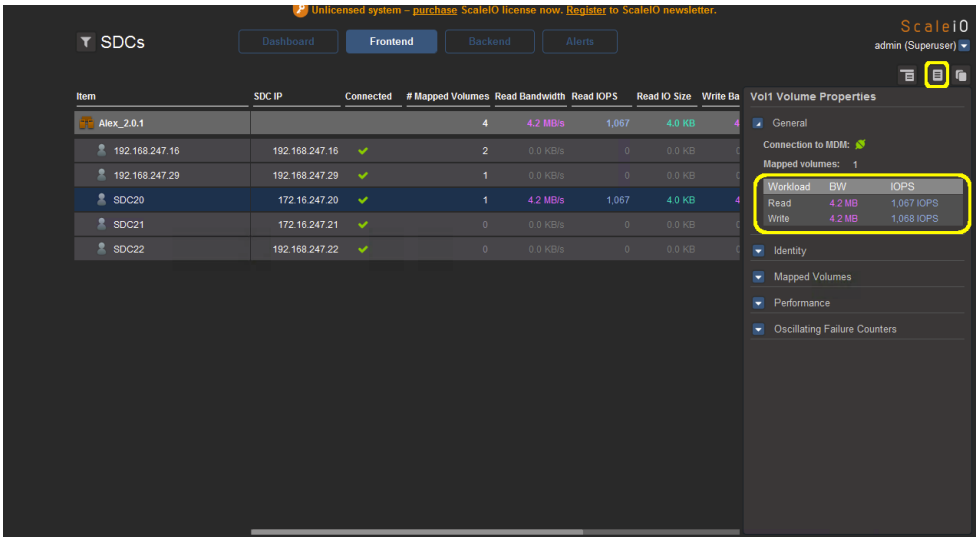


Retrieving host-specific performance information

You can retrieve host-specific information through the ScaleIO GUI or the SCLI.

Retrieving host-specific information through the ScaleIO GUI

1. On the **Frontend** tab, select **SDCs** view.
2. Select a volume and open its property sheet.
   The host information is displayed in the **General** section of the property sheet:

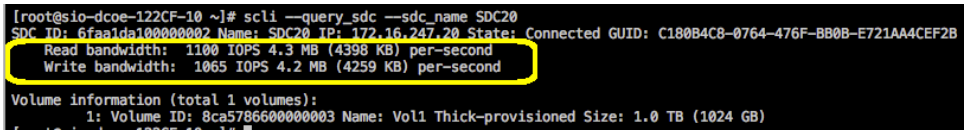


Retrieving host-specific information through the SCLI

1. Run this command:

```
scli --query sdc --sdc_name <sdc_name>
```

The same information appears as in the property sheet:



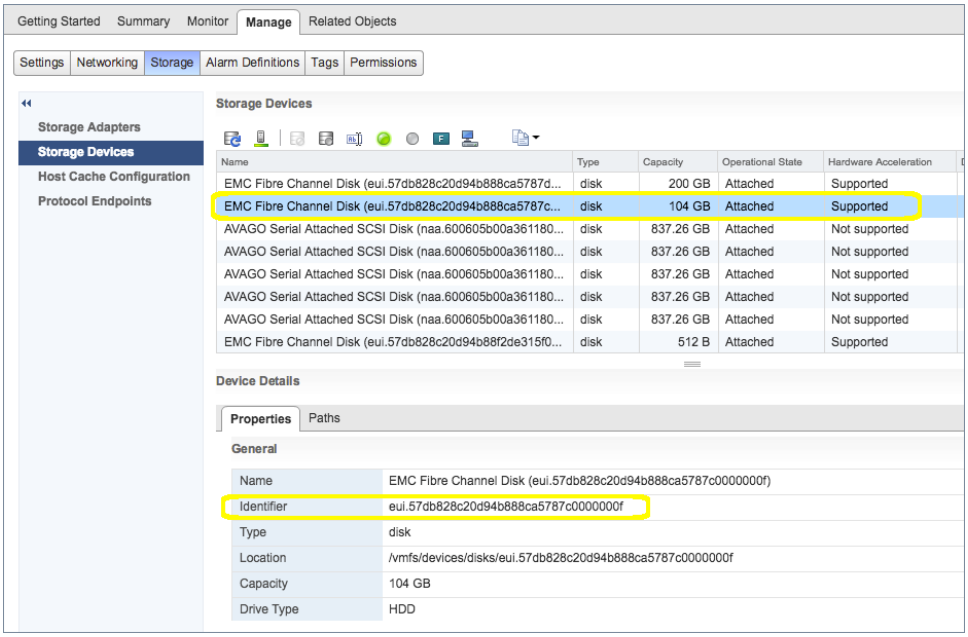
Correlating between the SDC on an ESX server and ScaleIO volumes

You can correlate between the LUN as it appears in the ESX and the logical ScaleIO volume.

Procedure

1. On the vSphere Client **Manage** tab, select **Storage** > **Storage Devices**.
2. Find the ScaleIO volume:

   ScaleIO volumes are exposed to ESX as EMC Fibre Channel Disks.



3. From the **Properties** tab, retrieve the LUN's EUI (Extended Unique Identifier).

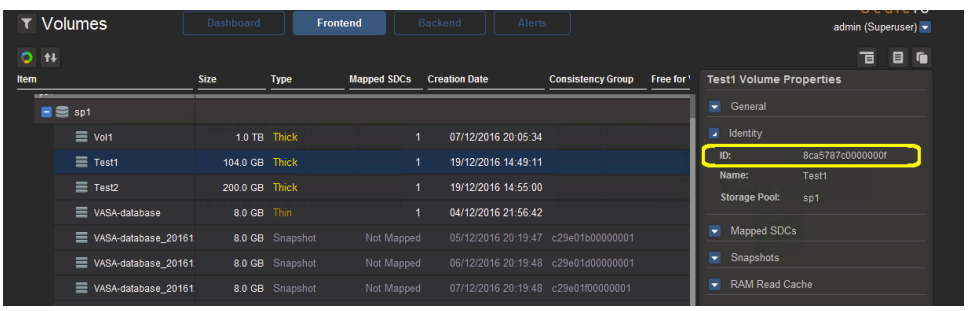
   The ScaleIO volume ID is the last 15 characters of the EUI.

4. Next retrieve the volume ID in ScaleIO:

   - To find the volume ID through the ScaleIO GUI:

     a. On the **Frontend** tab, select **Volumes** view.

     b. Select the volume and open its property sheet.
     The volume information is displayed in the **Identity** section of the property sheet:



   - To find the volume ID through the SCLI, run this command:

```
scli --query all volumes
```

```
Volume ID: 8ca5786600000003 Name: Vol1 Size: 1.0 TB (1024 GB) Mapped to 1 SDC Thick-provisioned
Volume ID: 8ca5787c0000000f Name: Test1 Size: 104.0 GB (106496 MB) Mapped to 1 SDC Thick-provisioned
Volume ID: 8ca5787d00000010 Name: Test2 Size: 200.0 GB (204800 MB) Mapped to 1 SDC Thick-provisioned
Volume ID: 8ca5787e00000011 Name: vvol:mF0r2Mi8SF2U+McdZw7mrg== Size: 8.0 GB (8192 MB) Unmapped Thin-provisioned
Volume ID: 8ca5787f00000012 Name: vvol:OY+U5gi7TkSpP4ueY7y0pg== Size: 8.0 GB (8192 MB) Unmapped Thin-provisioned
```

5. If more than one ScaleIO cluster is mapped to the same ESX, the second part of the EUI can be used to identify the cluster. From the SCLI, run:

```
scli --query cluster
```

The volume ID appears:

```
System Info:
        Product:  EMC ScaleIO Version: R2_0.10000.2072
        ID:       57db828c20d94b88
        Manager ID:       0000000000000000
        Name:     Alex_2.0.1
```

Correlating between the SDS on a Linux server and ScaleIO volumes

You can correlate between the SDS volume ID on a Linux server and a ScaleIO volume.

Procedure

1. On the Linux server, run:

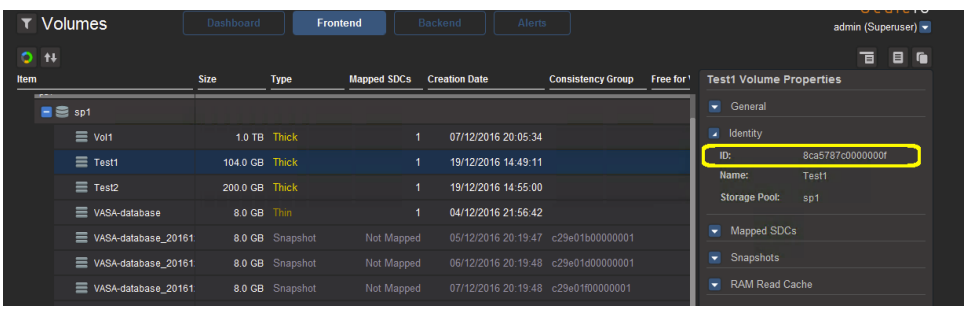
```
ls -l /dev/disk/by-uuid/
```

2. Find the UUID (Universally Unique Identifier) of the ScaleIO volume:

```
[root@sio-dcoe-122CF-10 ~]# ls -l /dev/disk/by-id/
total 0
lrwxrwxrwx 1 root root 12 2016-12-31 14:26 emc-vol-57db828c20d94b88-8ca5786600000003 -> ../../scinia
lrwxrwxrwx 1 root root  9 2016-12-04 21:42 scsi-3600605b00a363c301d1414f913784b70 -> ../../sda
lrwxrwxrwx 1 root root 10 2016-12-04 21:36 scsi-3600605b00a363c301d1414f913784b70-part1 -> ../../sda1
lrwxrwxrwx 1 root root 10 2016-12-04 21:36 scsi-3600605b00a363c301d1414f913784b70-part2 -> ../../sda2
lrwxrwxrwx 1 root root 10 2016-12-04 21:36 scsi-3600605b00a363c301d1414f913784b70-part3 -> ../../sda3
```

The ScaleIO volume ID is the last 15 characters of the UUID.

3. Next retrieve the volume ID in ScaleIO:
   - To find the volume ID through the ScaleIO GUI:
     a. On the **Frontend** tab, select **Volumes** view.
     b. Highlight the desired volume and open the property sheet.
     The volume information is displayed in the **Identity** section of the property sheet:



   - To find the volume ID through the SCLI, run this command:

```
scli --query all volumes
```

```
Volume ID: 8ca5786600000003 Name: Vol1 Size: 1.0 TB (1024 GB) Mapped to 1 SDC Thick-provisioned
Volume ID: 8ca5787c0000000f Name: Test1 Size: 104.0 GB (106496 MB) Mapped to 1 SDC Thick-provisioned
Volume ID: 8ca5787d00000010 Name: Test2 Size: 200.0 GB (204800 MB) Mapped to 1 SDC Thick-provisioned
Volume ID: 8ca5787e00000011 Name: vvol:mFOr2Mi8SF2U+McdZw7mrg== Size: 8.0 GB (8192 MB) Unmapped Thin-provisioned
Volume ID: 8ca5787f00000012 Name: vvol:OY+U5gi7TkSpP4ueY7y0pg== Size: 8.0 GB (8192 MB) Unmapped Thin-provisioned
```

CHAPTER 3

Advanced Troubleshooting

This chapter provides advanced procedures for troubleshooting ScaleIO performance. Topics include:

- Basic verification checks 18
- ScaleIO network troubleshooting 18

Basic verification checks

The following are basic verification checks to perform when troubleshooting ScaleIO performance:

Procedure

1. Check that the ScaleIO system configuration conforms to ScaleIO best practices.

   For details, see the *ScaleIO Performance Fine-Tuning Technical Notes*.

2. Check that no rebuild or rebalance is in progress.

   Rebuild and rebalance consume network bandwidth and generate disk I/Os. If rebuild or rebalance is affecting application I/O, limit the impact on system performance one of the following ways:

   - Set network throttling in order to limit the rebuild or rebalance network load to a desired value for a given Protection Domain.
   - Set I/O priority for a given Storage Pool in order to limit rebuild/rebalance I/O, or set a dynamic bandwidth policy.

   For information see the *ScaleIO User Guide*.

3. In the ScaleIO GUI **Dashboard** and **Alerts** tabs, check whether there are any failed components.

   When performing I/O, ScaleIO utilizes all of the SDS devices. Increasing or reducing the number of devices will have a direct effect on ScaleIO performance.

4. In the ScaleIO GUI **Backend** tab, check whether any SDS or Fault Set components are in maintenance mode.

   A component that is in maintenance mode does not serve I/Os, causing a reduction in total ScaleIO performance numbers.

ScaleIO network troubleshooting

ScaleIO supports a number of methods of troubleshooting the network for issues that may affect system performance:

- Troubleshooting SDS connectivity on page 18
- Troubleshooting SDC-to-SDS connectivity on page 19
- Troubleshooting ScaleIO media on page 20

Troubleshooting SDS connectivity

SDS connectivity problems will affect ScaleIO performance. Use the following checks to troubleshoot connectivity between the SDSs:

Procedure

1. Check the status of each network interface used by the SDS, using the tools appropriate to the operating system.

2. Check the packets lost counter.

Note

In some operating systems, the packets lost counter resets only within system reboot. It is therefore recommended that you check whether the number of packets lost increases over time.

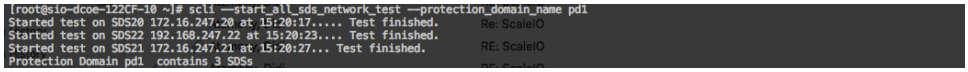
3. From the SCLI, run the ScaleIO internal network test in order to verify the network speed between all the SDS nodes in the Protection Domain:

Note

For accurate results, you must execute this test while there are no application, internal rebuild or rebalance I/Os in the system.

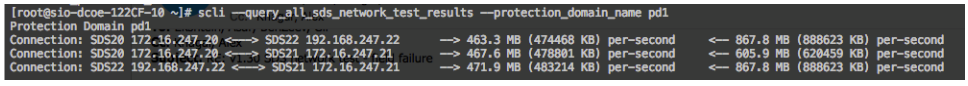
a. Initiate the test:

```
scli --start_all_sds_network_test --protection_domain_name
<protection_domain_name>
```



b. Review the results of the test:

```
scli --query_all_sds_network_test_results --
protection_domain_name <protection_domain_name>
```



4. In the SCLI, check the network latency numbers for the last 5 seconds of each SDS:

Note

You must execute this command while there are application I/Os in the system.

```
scli --query_network_latency_meters
```

Troubleshooting SDC-to-SDS connectivity

In cases where a network bottleneck is suspected, run an educated workload on the ScaleIO volume in order to check the quality of the network between the SDS and SDC.

This test should be run only in cases where the SDS devices are fast enough to accommodate the workload. Otherwise, the SDS devices may be the cause of the bottleneck, thereby creating misleading test results.

Procedure

1. Install an I/O generator on the host to which the ScaleIO volume is attached. (For example, FIO.)

2. Run a workload with a large I/O size (64K).

Results

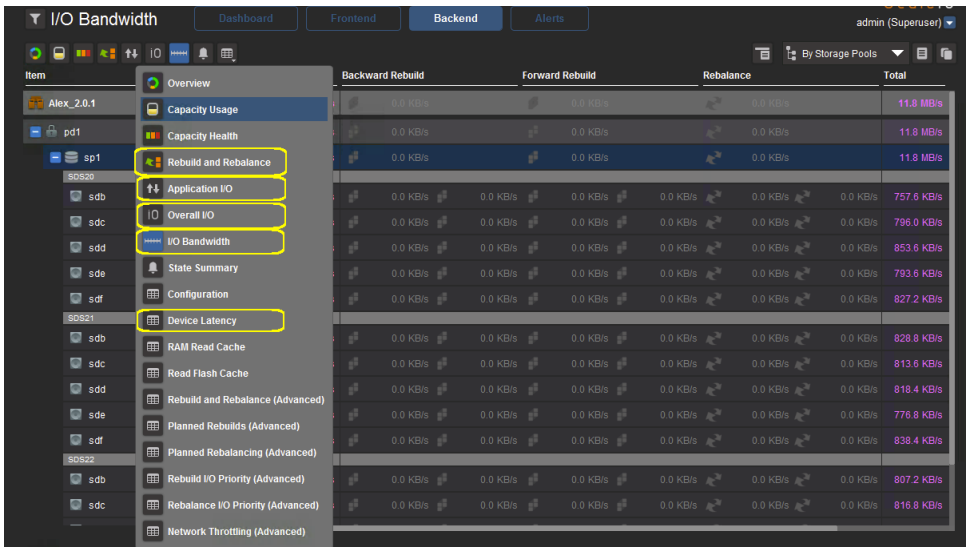
For each 10Gb of network interface, the SDC should generate approximately 1GB of network bandwidth.

Troubleshooting ScaleIO media

ScaleIO supports a number of methods of

Procedure

1. In the ScaleIO GUI **Backend** tab, monitor each device in the system for concurrent:
   - Application I/O
   - Overall I/O
   - I/O bandwidth
   - Device latency



2. From the SCLI, perform an internal ScaleIO device test:

Note

For accurate results, you must execute this test while the system has no application or internal rebuild or rebalance I/Os.

---

a. Start the device test:

```
scli --start_device_test --sds_name <name>
```

b. Review the results:

```
scli --query_device_test --sds_name <name>
```

Output similar to the following should appear:

```
[root@sio-dcoe-122CF-10 ~]# scli --query_device_test --sds_name SDS20
Queried results from 5 devices
fff98eed00020000: /dev/sdb
   Results from Tue Nov  8 18:43:40 2016
     5040 random reads, 8K each, total of 39.4 MB (40320 KB) in 20 seconds (20131 milliseconds). Bandwidth: 2.0 MB (2002 KB) per-second,  IOPS: 250
   Results from Sun Nov 20 15:34:47 2016
     16384 random reads, 8K each, total of 128.0 MB (131072 KB) in 46 seconds (46791 milliseconds). Bandwidth: 2.7 MB (2801 KB) per-second,  IOPS: 350
fff98eef00020001: /dev/sdc
   Results from Tue Nov  8 18:43:40 2016
     5045 random reads, 8K each, total of 39.4 MB (40360 KB) in 20 seconds (20161 milliseconds). Bandwidth: 2.0 MB (2001 KB) per-second,  IOPS: 250
   Results from Sun Nov 20 15:34:47 2016
     16384 random reads, 8K each, total of 128.0 MB (131072 KB) in 46 seconds (46811 milliseconds). Bandwidth: 2.7 MB (2800 KB) per-second,  IOPS: 350
```

3. From the SCLI, query the latency of the device for the past 5 seconds:

```
scli ---query_device_latency_meters --sds_name <name>
```

Output similar to the following should appear:

```
[root@sio-dcoe-122CF-10 ~]# scli --query_device_latency_meters --sds_name SDS20
SDS SDS20 returned information on 5 devices
Device: /dev/sdf
  Average read size: 0 Bytes
  Average read latency: 0 (microseconds)
  Average write size: 0 Bytes
  Average write latency: 0 (microseconds)
```

CHAPTER 4

Example

This chapter provides an example of a performance problem in a ScaleIO system and how to troubleshoot it. Topics include:

- Reduced IOPS capacity 24

Reduced IOPS capacity

Description

A system that used to serve more than 2000 IOPS can now only serve up to 1000 IOPS.

Troubleshooting

In this example, the system has three SDS nodes. Each node server contains five HDDSs and two 10Gb NICs. Because this is a converged environment, the same NICs are being used for both SDSs and SDCs.

Procedure

1. From the SCLI, run an SDS network test:

```
scli --start_all_sds_network_test --protection_domain_name pd1
```

The test results in the following output:

```
[root@sio-dcoe-122CF-10 ~]# scli --query_all_sds_network_test_results --protection_domain_name pd1
Protection Domain pd1
Connection: SDS22 192.168.247.22 <--> SDS20 172.16.247.20      --> 839.3 MB (859488 KB) per-second     <-- 1.1 GB (1150 MB) per-second
Connection: SDS20 172.16.247.20 <--> SDS21 172.16.247.21       --> 898.2 MB (919803 KB) per-second     <-- 853.3 MB (873813 KB) per-second
Connection: SDS22 192.168.247.22 <--> SDS21 172.16.247.21      --> 906.2 MB (927943 KB) per-second     <-- 1.1 GB (1113 MB) per-second
```

The results of the test are not perfect and instead of ~2GB (20Gb) of bandwidth, only half of the expected value is seen. However, the bandwidth still does not explain the performance issue because the system includes HDDs. Therefore, only 1GB of bandwidth should be enough. Additional tests are required.

2. Run a device test on each host:

```
scli --start_device_test --sds_name SDS20
```

```
scli --start_device_test --sds_name SDS21
```

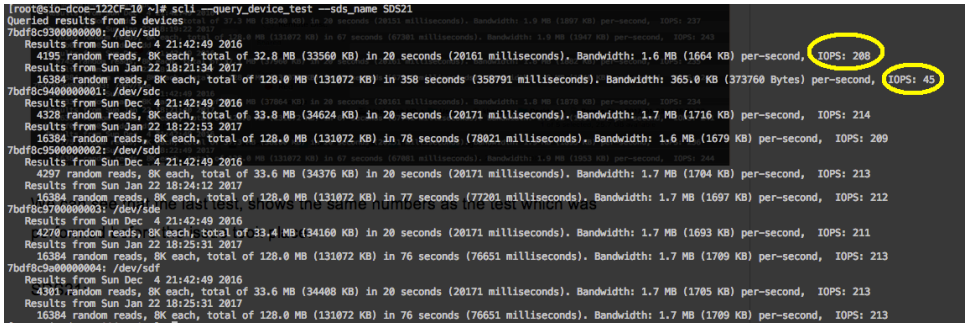
The tests result in the following output:

Figure 1 SDS20

```
[root@sio-dcoe-122CF-10 ~]# scli --query_device_test --sds_name SDS20
Queried results from 5 devices
7bdf8c9600010000: /dev/sdb
   Results from Sun Dec  4 21:42:49 2016
     4818 random reads, 8K each, total of 37.6 MB (38544 KB) in 20 seconds (20151 milliseconds). Bandwidth: 1.9 MB (1912 KB) per-second,  IOPS: 239
   Results from Sun Jan 22 18:18:13 2017
     16384 random reads, 8K each, total of 128.0 MB (131072 KB) in 72 seconds (72741 milliseconds). Bandwidth: 1.8 MB (1801 KB) per-second,  IOPS: 225
7bdf8c9800010001: /dev/sdc
   Results from Sun Dec  4 21:42:49 2016
     4780 random reads, 8K each, total of 37.3 MB (38240 KB) in 20 seconds (20151 milliseconds). Bandwidth: 1.9 MB (1897 KB) per-second,  IOPS: 237
   Results from Sun Jan 22 18:19:22 2017
     16384 random reads, 8K each, total of 128.0 MB (131072 KB) in 67 seconds (67301 milliseconds). Bandwidth: 1.9 MB (1947 KB) per-second,  IOPS: 243
7bdf8c9b00010002: /dev/sdd
   Results from Sun Dec  4 21:42:49 2016
     4745 random reads, 8K each, total of 37.1 MB (37960 KB) in 20 seconds (20161 milliseconds). Bandwidth: 1.8 MB (1882 KB) per-second,  IOPS: 235
   Results from Sun Jan 22 18:20:31 2017
     16384 random reads, 8K each, total of 128.0 MB (131072 KB) in 67 seconds (67541 milliseconds). Bandwidth: 1.9 MB (1940 KB) per-second,  IOPS: 242
7bdf8c9d00010003: /dev/sde
   Results from Sun Dec  4 21:42:49 2016
     4733 random reads, 8K each, total of 37.0 MB (37864 KB) in 20 seconds (20161 milliseconds). Bandwidth: 1.8 MB (1878 KB) per-second,  IOPS: 234
   Results from Sun Jan 22 18:21:40 2017
     16384 random reads, 8K each, total of 128.0 MB (131072 KB) in 67 seconds (67741 milliseconds). Bandwidth: 1.9 MB (1934 KB) per-second,  IOPS: 241
7bdf8c9f00010004: /dev/sdf
   Results from Sun Dec  4 21:42:49 2016
     4775 random reads, 8K each, total of 37.3 MB (38200 KB) in 20 seconds (20151 milliseconds). Bandwidth: 1.9 MB (1895 KB) per-second,  IOPS: 236
   Results from Sun Jan 22 18:22:49 2017
     16384 random reads, 8K each, total of 128.0 MB (131072 KB) in 67 seconds (67081 milliseconds). Bandwidth: 1.9 MB (1953 KB) per-second,  IOPS: 244
```

The test on SDS20 shows the same numbers as the test that was performed before the issue took place.

Figure 2 SDS21

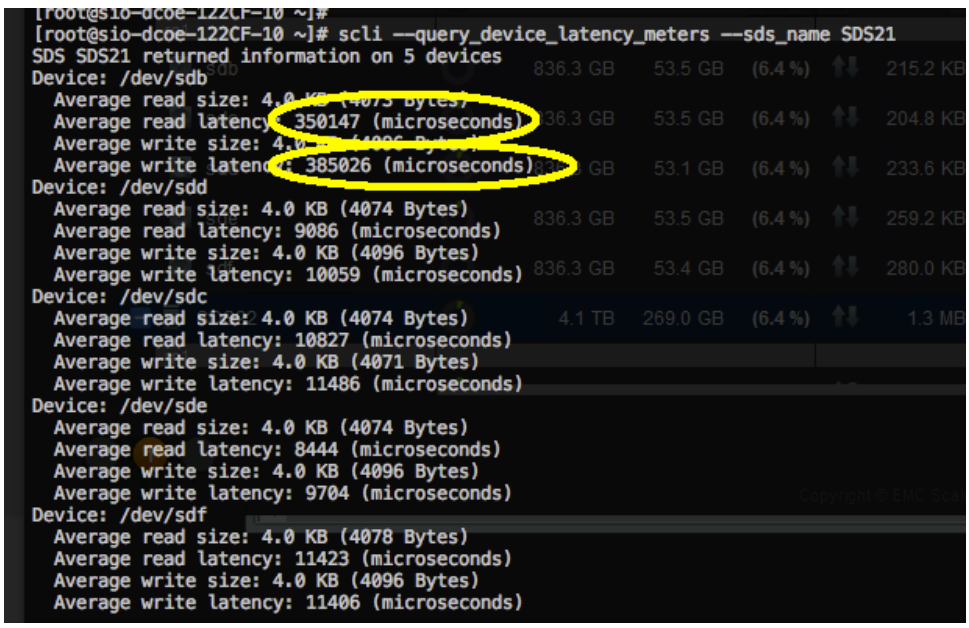


The test on SDS21 shows that the device named sdb, which was previously serving 208 IOPS, now serves only 45. Further tests are required to determine whether sdb is causing the drop in IOPS.

3. Check the device latency on SDS21 by running I/Os on the ScaleIO volumes:

```
scli ---query_device_latency_meters --sds_name SDS21
```

The test results in the following output:



These results clearly show that the latency of the device named "sdb" is significantly higher than that of any other device.

4. Disable sdb to verify whether the system performs better without it:

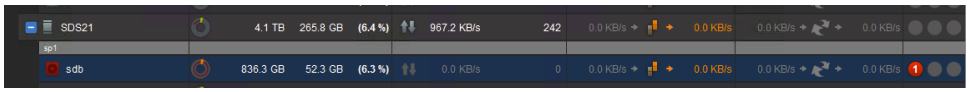
```
Echo offline >  /sys/block/sdb/device/state
```

Note

In this example, because the SDS is running on a Linux system, a Linux command is used to disable it.

---

5. In the ScaleIO GUI **Backend** tab, verify that the sdb device is disabled:



6. In the ScaleIO GUI **Dashboard** tab, check the **I/O Workload** tile:

The Dashboard shows that although there is a rebuild running in the system, the overall performance now is nearly twice as high: 1791 IOPS, instead of the 1000 IOPS previously detected.

Results

The tests reveal that sdb suffers from performance issues.