IBM Systems

Technical white paper

August 2019

iSER host attachment from Linux hosts to IBM SVC and IBM Storwize family controllers

iSER configuration guide for initiator and target

Overview

Challenge
Block storage access over Ethernet using traditional iSCSI protocol is not scalable for high speed Ethernet networks.

Solution
IBM SVC and Storwize family controllers have come up with support for high speed Ethernet ports that are accessible using the iSCSI Extensions over RDMA (iSER) protocol for high performance, low latency access.

The purpose of this paper is to provide detailed, step-by-step instructions on how to discover and establish connections from Linux® hosts to storage from IBM® System Storage® SAN Volume Controller (SVC) using the iSCSI Extensions over RDMA (iSER) protocol over high speed Ethernet links.

This paper is intended for people who are familiar with SVC, Internet Small Computer System Interface (iSCSI) but need a practical example to guide them through the steps of how to export the volumes to Linux hosts using iSER connections over iWARP and RoCE ports on Linux hosts and SVC/ iSER targets in an Ethernet storage area network (SAN).

Getting started

IBM SVC and IBM Storwize® systems were being used as iSCSI targets till 8.1.0.0 release using 1 Gb and 10 Gb Ethernet ports. Release 8.1.1.0 added support for 25 Gb RDMA over Converged Ethernet (RoCE) and Internet Wide-area RDMA Protocol (iWARP) ports which could be used as iSCSI targets. This means that only iSCSI hosts could connect to IBM SVC/Storwize using the 25 Gb ports.

In the 8.2.0 release, IBM introduced support for iSER target on SVC and Storwize platforms using the 25 Gb RoCE and iWARP ports. This feature enables hosts to connect to SVC or Storwize using the iSER protocol enabling high performance and lower latency connectivity from hosts to SVC or Storwize.

Introduction

The iSCSI protocol based on the TCP/IP stack has been the primary protocol used for block data transfers over Ethernet fabrics. Data transfers using traditional iSCSI stacks use processor cycles on host for data transfers and involve multiple copies of data in the I/O path and a high latency. iSER is based on a data mover architecture, which can work on various remote direct memory access (RDMA) technologies such as RoCE, iWARP and InfiniBand. iSER enables direct remote movement of data across application buffers on different hosts and can help to achieve low latency data transfers with a high throughput.

Advantages of iSER

iSER has multiple advantages:

- It frees up the host processor off any data copy work increasing overall system performance.
- It eliminates the need to create multiple copies of data across OS and network boundaries.
- It is vendor and technology independent. Data transfers happen through an API layer called **ib_verbs**, which is available both in user space and kernel space and it is independent of the underlying Ethernet technology. Any Ethernet protocol that supports the ib_verbs API layer and conforms to the RDMA semantics can be used.
- iSER offers latency in the range of 70 micro-seconds for 4KB block transfers to 300 micro-seconds for 256K transfer sizes
- 25 Gbps ports offer 2.5 times the bandwidth of 10Gbps ports with reduced and simplified cabling.
- 25 Gbps ports can be connected to 10 Gbps fabric using SFP+ connectors. The same ports can serve data over iSCSI or iSER protocols from different hosts. Hence, it is possible for legacy 10G iSCSI hosts to establish iSCSI connection to the SVC target through the 25 Gbps ports connected to 10G switches.

Architecture

Software

- IBM SVC software product version 8.2.1.0 or later

Hardware

- SAN Volume Controller model SV1
- Storwize V7000 model 600 or later / IBM FlashSystem® 9100
- RHEL Linux 7.4 with Chelsio T6225-CR dual-port adapter (iWARP)
- RHEL Linux 7.4 + with Mellanox Connect4-LX dual-port adapter (RoCE)

Network

- Cisco or Mellanox 100 Gb switch
- Mellanox or Chelsio 25 Gb adapter

Environment setup

This section describes the lab environment used perform the tests. The set up has two Linux servers. One has a 25 Gb iWARP adapter and the other has a 25 Gb RoCE adapter.

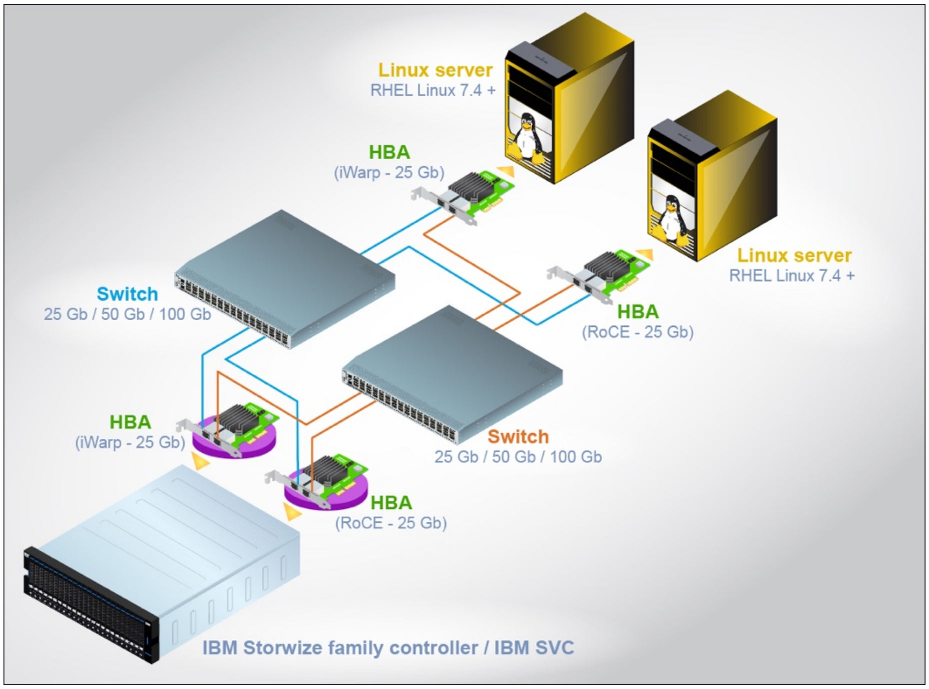


Figure 1: Lab setup with IBM Storwize and Linux hosts using a 25Gb to 100 Gb Ethernet fabric

The iSCSI initiator configuration is same on both SVC and Storwize. In the example used for this white paper, the lab team used SVC as the iSCSI initiator. Same steps are applicable for Storwize also.

Configuring Linux hosts as iSER initiators and IBM Storwize as iSER target

To establish sessions from iSER hosts (initiators) to iSER controllers (targets), first the iSER target has to be configured on the controller. Then, the iSER hosts must to be configured to establish sessions to the controller.

Configuring iSER Storwize target controller

Administrative model for configuring iSER is exactly the same as iSCSI. You can establish iSER host attach connectivity in a few easy steps.

Steps for configuration:

1. Find out the iSCSI qualified name (IQN) of the Linux hosts' servers with RoCE and iWARP ports.

   On host1 and host2 run the following commands:

   ```
   # cat /etc/iscsi/initiatorname.iscsi
   InitiatorName=iqn.1994-05.com.redhat:93b191e0445d

   # cat /etc/iscsi/initiatorname.iscsi
   InitiatorName=iqn.1994-05.com.redhat:93b191e046a1
   ```

   On the Storwize controller, create host objects and add all the SVC initiator's IQN in the host object.

   ```
   svctask mkhost -iscsiname<iqn> -name <host_obj_name>
   ```

   Example:

   ```
   svctask mkhost -iscsiname iqn.1994-05.com.redhat:93b191e0445d -name RHEL_RoCE_Host

   svctask mkhost -iscsiname iqn.1994-05.com.redhat:93b191e046a1 -name RHEL_iWARP_Host
   ```

2. To verify the host object from the Storwize target CLI, run the following command:

   ```
   # lshost
   id name                port_count iogrp_count status  site_id site_name host_cluster_id host_cluster_name
   protocol owner_id owner_name
   0  RHEL_RoCE_Host     1           4            online
   scsi
   1  RHEL_iWARP_Host    1           4            online
   scsi
   ```

3. To enable iSCSI authentication (optional) on the Storwize controller, give a Challenge Handshake Authentication Protocol (CHAP) secret to the host object created in step 1.

   ```
   chhost -chapsecret<chap secret><host object id / name>
   ```

   Example:

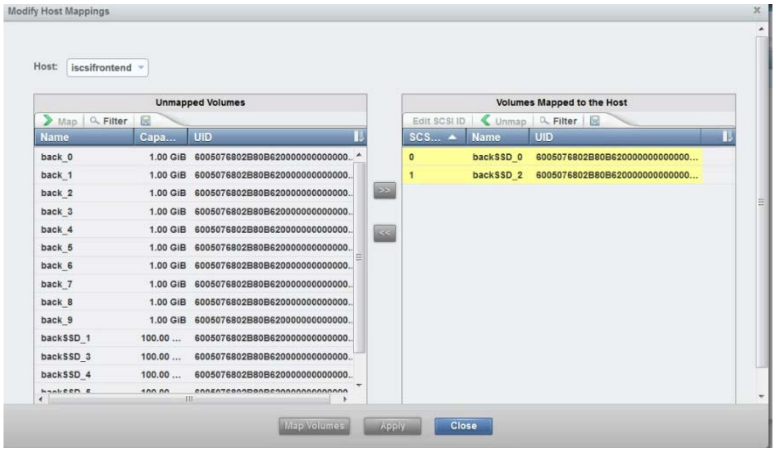
   ```
   svctask chhost –chapsecret hostchap RHEL_RoCE_Host
   ```

4. Map the volumes to the hosts.

```
svctask mkvdiskhostmap -host <host object id/name><vdisk id/name>
```

Example:

```
svctask mkvdiskhostmap -host RHEL_RoCE_Host backSSD_0
svctask mkvdiskhostmap -host RHEL_iWARP_Host backSSD_1
```



Modify Host Mappings

Host: iscsifrontend

Unmapped Volumes

Name	Capa...	UID
back_0	1.00 GiB	6005076802B80B620000000000000...
back_1	1.00 GiB	6005076802B80B620000000000000...
back_2	1.00 GiB	6005076802B80B620000000000000...
back_3	1.00 GiB	6005076802B80B620000000000000...
back_4	1.00 GiB	6005076802B80B620000000000000...
back_5	1.00 GiB	6005076802B80B620000000000000...
back_6	1.00 GiB	6005076802B80B620000000000000...
back_7	1.00 GiB	6005076802B80B620000000000000...
back_8	1.00 GiB	6005076802B80B620000000000000...
back_9	1.00 GiB	6005076802B80B620000000000000...
backSSD_1	100.00 ...	6005076802B80B620000000000000...
backSSD_3	100.00 ...	6005076802B80B620000000000000...
backSSD_4	100.00 ...	6005076802B80B620000000000000...

Volumes Mapped to the Host

SCS...	Name	UID
0	backSSD_0	6005076802B80B620000000000000...
1	backSSD_2	6005076802B80B620000000000000...

Figure 4: Mapping the VDisk to the newly created host object

5. Configure the target IPs on the Storwize controller using the `–host` flag to allow login from the iSER initiator, if not done already.

```
svctask cfgportip -node <node id> -ip<ipv4 addr> -gw<ipv4 gwaddr> -mask <mask>   -host yes
<port id>

svctask cfgportip -node <node id> -ip_6 <ipv6 addr>  -prefix_6 <prefix> -gw_6 <ipv6
gwaddr>   -host_6 -yes <port id>
```

Example:

```
svctask cfgportip -node 1 -ip 172.168.33.100 -gw 172.168.33.1 -mask 255.255.255.0 -host
yes 6

svctask cfgportip -node 1 -ip 172.168.34.100 -gw 172.168.34.1 -mask 255.255.255.0 -host
yes 7

svctask cfgportip -node 2 -ip 172.168.33.200 -gw 172.168.33.1 -mask 255.255.255.0 -host
yes 6

svctask cfgportip -node 2 -ip 172.168.34.200 -gw 172.168.34.1 -mask 255.255.255.0 -host
yes 7
```

If the IPs are already configured, change the `host` flag to `yes` using the following CLI:

```
svctask cfgportip -node <node id> -host yes <port id>
svctask cfgportip -node <node id> -host_6 -yes <port id>
```

Example:

```
svctask cfgportip -node 1 -host yes 6
svctask cfgportip -node 1 -host yes 7
svctask cfgportip -node 2 -host yes 6
svctask cfgportip -node 2 -host yes 7
```

The Storwize system with the IPs configured will look like:

```
4:1:node1:172.168.33.101:255.255.255.0:172.168.33.1::::ec:0d:9a:9c:e5:fb:Full:configured:2
5Gb/s:no:active:yes:0::0:::::4:1:yes::1:RoCE:yes
4:1:node1:::::::ec:0d:9a:9c:e5:fb:Full:configured:25Gb/s:yes:active::0::0:::::4:1:::0:RoCE
:yes
5:1:node1:172.168.34.101:255.255.255.0:172.168.34.1::::ec:0d:9a:9c:e5:fa:Full:configured:2
5Gb/s:no:active:yes:0::0:::::4:2:yes::1:RoCE:yes
5:1:node1:::::::ec:0d:9a:9c:e5:fa:Full:configured:25Gb/s:yes:active::0::0:::::4:2:::0:RoCE
:yes
6:1:node1:172.168.33.100:255.255.255.0:172.168.33.1::::00:07:43:42:2a:38:Full:configured:2
5Gb/s:no:active:yes:0::0:::::6:1:yes::3:iWARP:no
6:1:node1:::::::00:07:43:42:2a:38:Full:configured:25Gb/s:yes:active::0::0:::::6:1:::0:iWAR
P:no
7:1:node1:172.168.34.100:255.255.255.0:172.168.34.1::::00:07:43:42:2a:30:Full:configured:2
5Gb/s:no:active:yes:0::0:::::6:2:yes::3:iWARP:no
7:1:node1:::::::00:07:43:42:2a:30:Full:configured:25Gb/s:yes:active::0::0:::::6:2:::0:iWAR
P:no

4:2:node2:172.168.33.201:255.255.255.0:172.168.33.1::::ec:0d:9a:9c:e6:63:Full:configured:2
5Gb/s:no:active:yes:0::0:::::4:1:yes::1:RoCE:yes
4:2:node2:::::::ec:0d:9a:9c:e6:63:Full:configured:25Gb/s:yes:active::0::0:::::4:1:::0:RoCE
:yes
5:2:node2:172.168.34.201:255.255.255.0:172.168.34.1::::ec:0d:9a:9c:e6:62:Full:configured:2
5Gb/s:no:active:yes:0::0:::::4:2:yes::1:RoCE:yes
5:2:node2:::::::ec:0d:9a:9c:e6:62:Full:configured:25Gb/s:yes:active::0::0:::::4:2:::0:RoCE
:yes
6:2:node2:172.168.33.200:255.255.255.0:172.168.33.1::::00:07:43:42:28:58:Full:configured:2
5Gb/s:no:active:yes:0::0:::::6:1:yes::3:iWARP:yes
6:2:node2:::::::00:07:43:42:28:58:Full:configured:25Gb/s:yes:active::0::0:::::6:1:::0:iWAR
P:yes
7:2:node2:172.168.34.200:255.255.255.0:172.168.34.1::::00:07:43:42:28:50:Full:configured:2
5Gb/s:no:active:yes:0::0:::::6:2:yes::3:iWARP:no
7:2:node2:::::::00:07:43:42:28:50:Full:configured:25Gb/s:yes:active::0::0:::::6:2:::0:iWAR
P:no
```

iSER initiator configuration on Linux hosts

The following sections provide the steps for configuring iSER on the Linux host.

Installing the Linux software iSER initiator

To install the Linux software iSER initiator for iSER host attachment, set the IQN, target discovery, and authentication. Also, enable multipathing for the Linux hosts.

1. Configure the iSER-initiator daemon by editing the /etc/iscsi file.

   ```
   [root@R64SRV6 ~]# cd /etc/iscsi/

   [root@R64SRV6 iscsi]# ls

   initiatorname.iscsi  iscsid.conf

   [root@R64SRV6 iscsi]#
   ```

2. Manage the software by using the `chkconfig` command and the **service** command. The service on Red Hat Enterprise Linux (RHEL) systems is named *iscsid*. The service on SUSE Linux Enterprise Server (SLES) systems is named *open-iscsi*. Consider the examples for RHEL and SLES systems.

   For RHEL systems:

   ```
   [root@R64SRV6 iscsi]# chkconfig --list | grep "iscsid"
   iscsid          0:off   1:off   2:off   3:off   4:off   5:off   6:off

   [root@R64SRV6 iscsi]# chkconfig iscsid on

   [root@R64SRV6 iscsi]# chkconfig --list | grep "iscsid"
   iscsid          0:off   1:off   2:on    3:on    4:on    5:on    6:off
   [root@R64SRV6 iscsi]#
   ```

   For SLES systems:

   YaST and YaST2 provide an appropriate method to properly configure iSER services and initiator. At a command line for iSER services in SLES, enter the following commands:

   ```
   #/etc/init.d/open-iscsi start

   #chkconfig open-iscsi on
   ```

3. Use the default settings for the iSER initiator. These settings are described in /etc/iscsi/iscsid.conf.

Setting the IQN for the iSER initiator

The iSCSI initiator IQN that you configure on the host side must never be the same as the iSER target IQN on the system node.

Edit the /etc/iscsi/initiatorname.iscsi file to specify the IQN. The following example shows a sample name:

```
InitiatorName= iqn.1994-05.com.redhat:93b191e0445d
```

After you change the initiator name, you must start the iSCSI-initiator service again.

Configuring more than one Linux host Ethernet port for iSER

You can use one of the following two methods:

- Configure different subnet IP addresses for the host ports and the system ports. Each target IP of a system node is accessible only from one host IP of the corresponding subnet. As a result, pairs of host and system ports are in separate subnets.
- In the /etc/iscsi/ifaces subdirectory, create a separate *iface* configuration file for each network object that you want to bind a session to. Each *iface* config file must have a unique name that is less than or equal to 64 characters in length.

The following *iface* example shows a simple case with two network devices (eth0 and eth1):

Procedure

1. Edit the /etc/iscsi/ifaces/eth0 file to specify the iface.net_ifacename:

```
 iface.net_ifacename = eth0
```

2. Similarly, edit the /etc/iscsi/ifaces/eth1 file:

```
iface.net_ifacename = eth0
```

iSER configuration required on the Linux host

The SVC or the Storwize system supports iSCSI and iSER sessions on the same target Ethernet port from different iSER hosts. To allow the same:

- The standard iSER target port for iSER login over RoCE ports is 3260.
- The standard iSER target port for iSER login over iWARP ports is 860.

Perform the following steps on the Linux host before starting the actual iSER login process:

1. Stop the port mapper service if it is running on the Linux host.

```
 # service iwpmd stop
```

2. Turn on the Address Resolution Protocol (ARP) filtering on the Linux host.

To avoid ARP flux on Linux, use the following setting on the host:

```
echo 1 > /proc/sys/net/ipv4/conf/all/arp_filter
```

Or, to make this behavior persistent, add the following entry in /etc/sysctl.conf:

```
net.ipv4.conf.all.arp_filter=1
```

The overall iSER session creation is divided into two steps:

1. iSER target discovery for Linux host
2. iSER login from the Linux host to the discovered targets

iSER target discovery for Linux hosts

This section explains the two ways to perform iSER target discovery.

Discover iSER target portals by using the send targets method

In this method, you can either perform RoCE discovery from a Linux host or iWARP discovery from a Linux host.

RoCE discovery from a Linux host

Run the following discovery command for a two-port 25 Gbps Ethernet (RoCE) adapter:

```
# iscsiadm --mode discovery --type sendtargets --portal x.x.x.x:3260 -I iser
```

For example:

```
# iscsiadm -m discovery -t sendtargets -p 10.20.90.21:3260 -I iser

10.20.90.21:3260,1 iqn.1986-03.com.ibm:2145.testcluster.node2
10.20.100.21:3260,1 iqn.1986-03.com.ibm:2145.testcluster.node2
```

As you can see from the output, 10.20.90.21 and 10.20.100.21 are the available RoCE targets IPs.

iWARP discovery from Linux host

Run the following discovery command for a two-port 25 Gbps Ethernet (iWARP) adapter:

```
#  iscsiadm --mode discovery --type sendtargets --portal x.x.x.x:860 -I iser
```

For example:

```
#iscsiadm -m discovery -t sendtargets -p 10.20.70.11:860 -I iser

10.20.70.11:860,1 iqn.1986-03.com.ibm:2145.testcluster.node1
10.20.80.11:860,1 iqn.1986-03.com.ibm:2145.testcluster.node1
```

This output from a successful discovery lists the target endpoints. Each endpoint is qualified by an IP address, a port number, a target portal group tag, and a target IQN.

Discover iSER targets on your Linux host by using iSNS

To discover iSER targets by using iSNS, complete the following steps:

1. Modify the iSNS server settings in the /etc/iscsi/iscsid.conf file by entering the IP address and the port number that is used by your iSNS server:

   ```
   isns.address = <iSNS server IP address>
   isns.port = <iSNS server port>
   ```

2. Make the iSNS configuration changes effective by restarting the iSER initiator service.

3. Run the iSNS discovery command to generate a list of all iSER target portals that are registered with the iSNS server:

   ```
   # iscsiadm --mode discovery --type isns
   ```

iSER login from the Linux host to the discovered targets

Initiate login from the iSER initiator to your selected system targets by using the following appropriate command examples for RHEL.

RoCE login from the Linux host

To log in to a two-port 25 Gbps Ethernet (RoCE) adapter on node1, enter the following command:

```
iscsiadm --mode node --targetname target_IQN --portal x.x.x.x:3260 –login
```

For example :

```
iscsiadm --m node --targetname iqn.1986-03.com.ibm:2145.testcluster.node2 --portal 10.20.90.21:3260
–login
```

iWARP login from the Linux host

To log in to a two-port 25 Gbps Ethernet (iWARP) adapter on node2, enter the following command:

```
iscsiadm --m node --targetname iqn.1986-03.com.ibm:2145.testcluster.node1 --
portal 10.20.70.12:860 --login
```

When the login command completes, the host can access the volumes that are mapped to the host from the target after the SCSI devices are rescanned.

Best practices for iSER host-attach connectivity

It is recommended to adhere to the following best practices for iSER host-attach connectivity:

- Same ports can be used for iSCSI and iSER host-attach concurrently.
- A single host can establish either iSCSI or iSER host-attach session to an SVC or a Storwize target.
  - iSCSI sessions should be logged out before attempting iSER host-attach. Similarly, iSER sessions should be logged out before attempting iSCSI sessions.
  - iSER host-attach sessions will be rejected if an iSCSI host-attach session from the same host is present. Similarly, iSCSI host-attach sessions will be rejected if an iSER host-attach session from the same host is present.
- A maximum of four iSER host-attach sessions from an iSER initiator to each target node can be established.
- A one-one connectivity between an initiator host port to the target port is recommended.
- More than one session from one or more initiator host ports to the same target port will be rejected.
- A single initiator host port can establish sessions to multiple target ports (not recommended).
- Single iSER host can use either IPv4 or IPv6 to establish iSER host-attach sessions
  - IPv6 session requests will be rejected if hosts have existing IPv4 host-attach sessions. Similarly, IPv4 session requests will be rejected if hosts have existing IPv6 host-attach sessions.
  - This restriction does not apply for iSCSI sessions. iSCSI hosts can still have concurrent IPv4 and IPv6 host-attach sessions to the same target.
- RoCE or iWARP ports are iSER un-interoperable
  - A Microsoft® Windows® iWARP initiator can establish iSER host-attach session to iWARP target ports only. Because the target supports both RoCE and iWARP ports, it is necessary to initiate target discovery and login from the Windows initiator host ports to the SVC or Storwize target ports only.
  - Attempting to established iSER host-attach connections from Windows iWARP ports to SVC/ Storwize RoCE ports will fail.

Features and limits

Some other features and limits of iSER host attach sessions from Linus host to IBM SVC/ IBM Storwize iSER targets are described in the following sections.

Authentication

Only uni-direction CHAP is supported for iSER host-attach sessions.

Session limits

Different number of total iSER host-attach sessions is supported on different SVC/Storwize platforms. Refer to the *Limits and Restrictions* page for product and release to understand the exact limits. Refer to V8.2.1.x Configuration Limits and Restrictions for IBM System Storage SAN Volume Controller. A different link is available for each major release of IBM SVC and IBM Storwize family controller.

Reservation support

iSER supports SCSI reservations. The specific feature provides the following support:

- SCSI2 and SCSI3 reservation
- SCSI3 Persistent reservation over multiple iSER host-attach sessions

Resources

The following websites provide useful references to supplement the information contained in this paper:

- IBM Storwize V7000 support page
  ibm.com/storage/support/storwize/v7000
- IBM Storwize family home page

  ibm.com/it-infrastructure/storage/storwize
- IBM Storwize V7000 supported hardware list
  ibm.com/support/docview.wss?uid=ssg1S1003703
- IBM Storwize V7000 Knowledge Center
  ibm.com/support/knowledgecenter/ST3FR7
- IBM SVC Knowledge Center
  ibm.com/support/knowledgecenter/STPVGU
- IBM Redbooks
  ibm.com/redbooks/redbooks/pdfs/sg247432.pdf

About the authors

Rakesh Kumar Chinhara is a software engineer in the SVC team at IBM Systems Labs India. He has around10 years of experience in IT. He holds a bachelor’s degree in Electrical Engineering from the Biju Patnaik University of Technology. You can reach Rakesh at: rchinhar@in.ibm.com.

Anuj Chandra is a Software Architect in the SVC team at IBM Systems Labs India. He has more than 19 years of experience in IT. He holds a bachelor’s degree in Computer Science from the University of Pune and has studied Business Management from IIM Calcutta. You can reach Anuj at anuj.chandra@in.ibm.com.

IBM Systems
3039 Cornwallis Road
RTP, NC 27709

Produced in the United States of America