

Solution assurance

PCIe RoCE Express Network Guide



Contents

Introduction..... V

Chapter 1. Install RHEL using RoCE..... 1

 Use VNC to install RHEL..... 4

 Use Kickstart to install RHEL..... 9

Chapter 2. Add the additional RoCE interface..... 13

References..... 17

Notices..... 19

 Trademarks and service marks..... 19

Introduction

This networking guide describes two scenarios using PCIe RoCE network express.

- How to install Red Hat Enterprise Linux (RHEL) in an LPAR (natively) on IBM Z or LinuxONE using the PCIe RoCE Express 2 (RoCE).
- How to add RoCE Express 2 (RoCE) as an additional interface to an existing LPAR (natively).

What is RoCE?

RDMA over Converged Ethernet (RoCE) is a standard network protocol that enables remote direct memory access (RDMA) efficient data transfer over Ethernet.

RoCE is a standard protocol defined in the InfiniBand Trade Association (IBTA) standard. The main advantage of RoCE is it allows direct memory to memory transfer at the application level without involving the CPU. As a result of this, RoCE has low latency and, depending on the workload, can offer a performance advantage. RoCE supports TCP/IP connection and Shared Memory Communications Remote (SMC-R) connection.

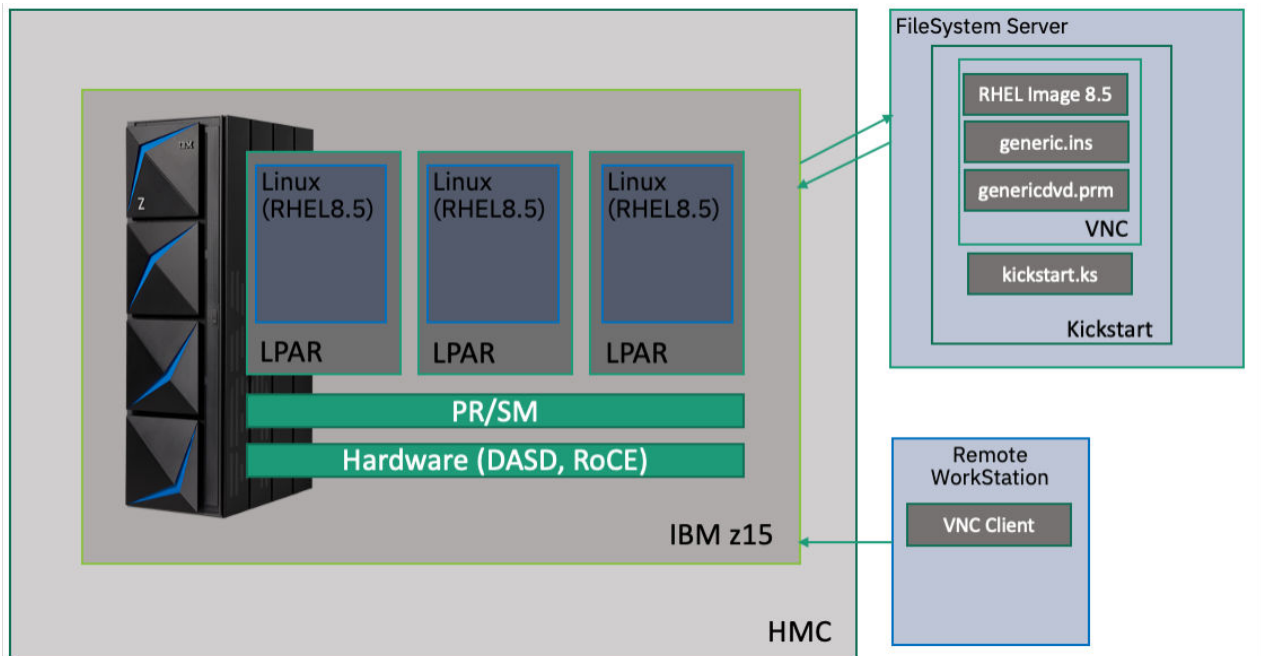
What is PCIe?

Peripheral component interconnect express (PCIe) is an interface standard for connecting high-speed components. In IBM, PCIe slots are used to connect networking devices (RoCe express), storage devices, and so on.

Chapter 1. Install RHEL using RoCE

This topic explains how to install Red Hat Enterprise Linux (RHEL) in an LPAR (natively) on IBM Z or LinuxONE using Virtual Network Computing (VNC) and Kickstart together with the PCIe RoCE Express 2 network card (RoCE).

RHEL installation architecture



Prerequisites

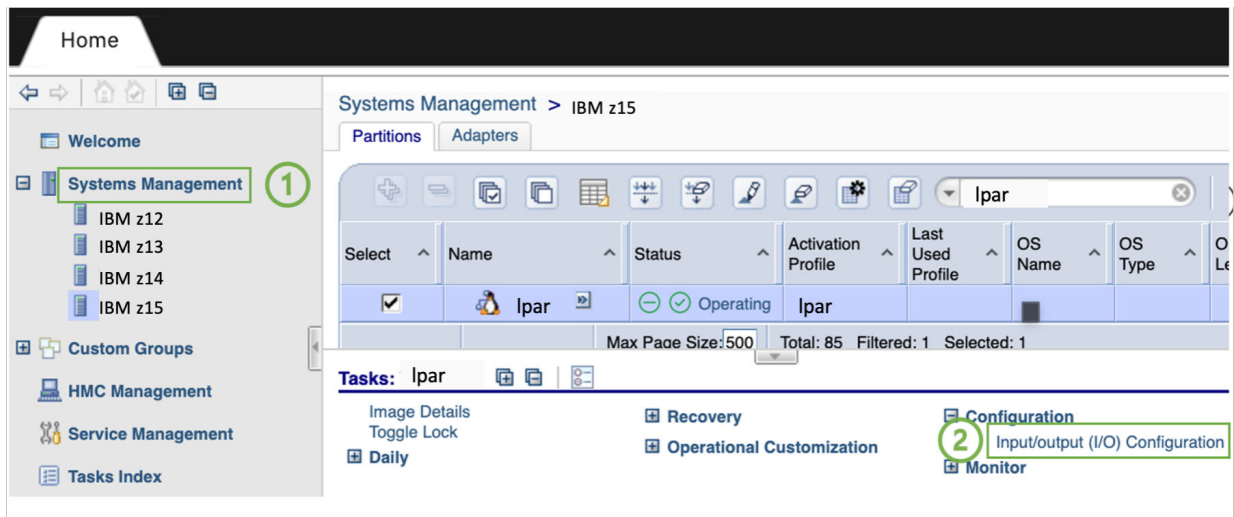
Make sure your installation environment meets the requirements to install RHEL. You need:

- The RHEL DVD ISO image for your preferred version (for the example in this paper, RHEL 8.5 is used)
- For the LPAR (Logical Partition): Make sure the following resources are allocated to your LPAR:
 - Storage device (DASD/SCSI).
 - Memory.
 - CPU (logical IFLs).
 - The RoCE network card. Make sure that RoCE can use HTTP to access the RHEL DVD installation image.
- The IP address for your RHEL machine (check with your network team to get the appropriate IP address)
- For installations that use VNC: VNC software to run the VNC installation from your local system

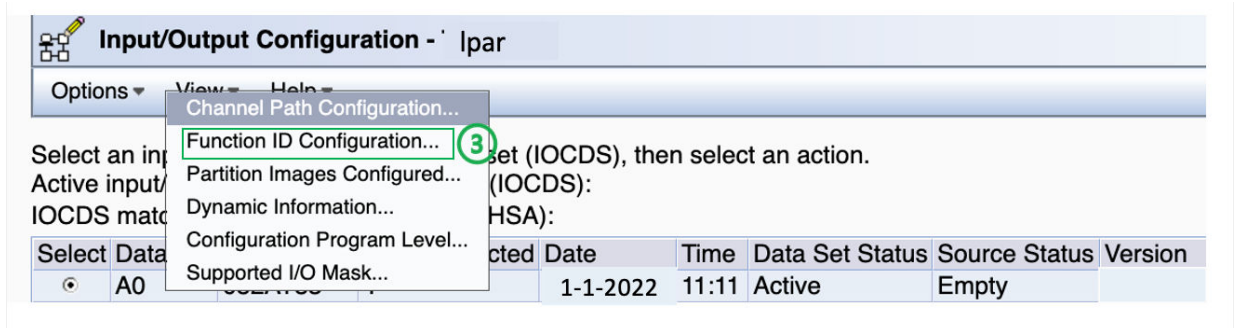
Verify RoCE is attached to your LPAR

Check with your networking team or follow these steps to verify that RoCE is attached to your LPAR:

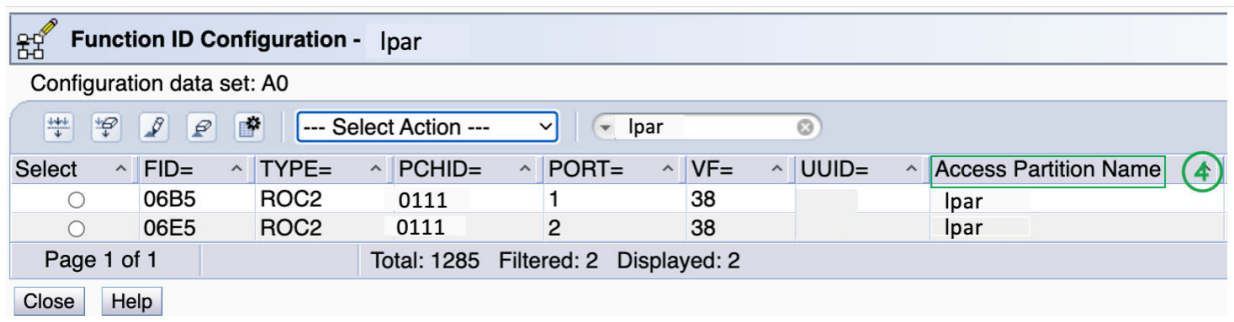
1. Log on to the Hardware Management Console (HMC), then select your LPAR from **System Management**.
2. Select **Configuration, Input/Output (I/O) Configuration**.



3. Select **Function ID Configuration** from the **View** tab.



4. Filter **Function ID Configuration** with the LPAR name. The list of RoCEs attached to the LPAR is displayed.



Create generic.ins and genericdvd.prm files

The files `generic.ins` and `genericdvd.prm` are required for RHEL installations that use VNC and Kickstart.

Create the generic.ins file

`generic.ins` is a mandatory file that is used to load LPAR parameters such as **kernel.img**, **initrd.img**, **generic.prm** and **initrd.addrsz**. Red Hat provides a valid `generic.ins` file for each RHEL DVD ISO for you to use or edit as required. The examples in this paper use the `generic.ins` for both the VNC and Kickstart installations.

The `generic.ins` file provided by RHEL 8.5:

```
images/kernel.img 0x00000000
images/initrd.img 0x02000000
```



```
images/genericdvd.prm 0x00010480
images/initrd.addrsize 0x00010408
```

The `generic.ins` file used for the VNC and Kickstart installations in the example:

```
RHEL8.5/images/kernel.img 0x00000000
RHEL8.5/images/initrd.img 0x02000000
RHEL8.5/genericdvd.prm 0x00010480
RHEL8.5/images/initrd.addrsize 0x00010408
```

Create the `genericdvd.prm` file

The `genericdvd.prm` file contains the list of parameters, such as network source, installation source, storage device, and so on, which are used for the initial RHEL installation setup. The sample installation in this paper uses the parameters listed here:

```
ro ramdisk_size=40000 cio_ignore=all,!condev
inst.repo=<Package_location>
ip=<IP_Address>:<Network_Gateway>:<Network_Netmask>:<LPAR_Host_Name>:<RoCE_Interface_Name>:none

nameserver=<IP_DNS_Server>
rd.dasd=<0.0.0000>
inst.vnc inst.vncpassword=<VNC_Password>
inst.ks=<Kickstart_File_Location>
```

where:

ro

Mounts the root file system (RAM disk in read only mode).

ramdisk_size

Sets the memory size of RAM disk.

cio_ignore=all,!condev

All the I/O devices except the console are to be ignored. This helps to speed up the boot and device detection process with many devices attached to a system.

ip

- `<IP_Address>`: IP address of the LPAR machine.
- `<Network_Gateway>`: Network Gateway IP address.
- `<Network_Netmask>`: Network Netmask length.
- `<LPAR_Host_Name>`: LPAR host name.
- `<RoCE_Interface_Name>`: Interface name of the RoCE device.

nameserver

DNS server IP address.

rd.dasd

The logical address of the DASD storage device attached to the LPAR (`rd.zfcp=` is for SCSI).

inst.repo

Location of the RHEL distribution.

inst.vnc

To run the VNC graphical interface.

inst.vncpassword

VNC login password (length: 6 to 8 characters).

inst.ks

Kickstart file location.

The `genericdvd.prm` file for VNC installation:

```
ro ramdisk_size=40000 cio_ignore=all,!condev
inst.repo=http://example.com/path/rhel8.5
ip=192.168.0.1:192.168.0.1:15:lpars.example.com:enp010:none
```

```
nameserver=192.168.0.1
rd.dasd=0.0.0000
inst.vnc inst.vncpassword=password
```

The genericdvd.prm file for Kickstart installation:

```
ro ramdisk_size=40000 cio_ignore=all,!condev
inst.repo=http://example.com/path/rhel8.5
ip=192.168.0.1::192.168.0.1:15:lpnr.example.com:enp010:none
nameserver=192.168.0.1
rd.dasd=0.0.0000
inst.ks=http://example.com/path/rhel8.5/roce/kickstart.ks
```

Important: Unlike OSA, RoCE does not use the **rd.znet** parameter.

How to find the RoCE interface name

Finding the RoCE interface name is challenging, and it depends upon your distribution. If your distribution follows a "predictable interface name", your RoCE interface name is either **eno<UID_in_dec>** or **ens<FID_in_dec>**.

- If UID uniqueness is enabled (check with your networking team), the RoCE interface name has the format **eno<UID_in_dec>** (<UID_in_dec> UID in decimal notation).

For Example: UID:06b5. A conversion to decimal from hexadecimal is 1717, so the Interface name is **eno1717**.

- If UID uniqueness is disabled (check with your networking team), the RoCE interface name has the format **ens<FID_in_dec>** (<FID_in_dec> FID in decimal notation).

For Example: FID is 06b5. A conversion to decimal from hexadecimal is 1717, so the Interface name is **ens1717**.

Note: You can find the UID and FID of your RoCE in **Function ID Configuration** as described in [“Verify RoCE is attached to your LPAR”](#) on page 1.

Distributions for which the "predictable interface name" scheme does not take effect use these naming formats:

- **enP <UID_or_counter><FID_in_dec>**
- **enP <UID_or_counter>< p0s**
- **enP <UID_or_counter>< p0np0**
- **enP <UID_or_counter>< p0s0np0**
- **enP <UID_or_counter>< p0**

Tip: Multiple dependencies on the hardware and its configuration decide the interface name, and it is difficult to predict. If you are unsure about your network interface name, enter your closest predicted RoCE interface name and search for the correct RoCE interface name by checking the messages in HMC **Operating Systems Messages** during installation, for example:

```
Check for :[...] mlx5_core 0000:00:00.0: enp010: renamed from eth0
```

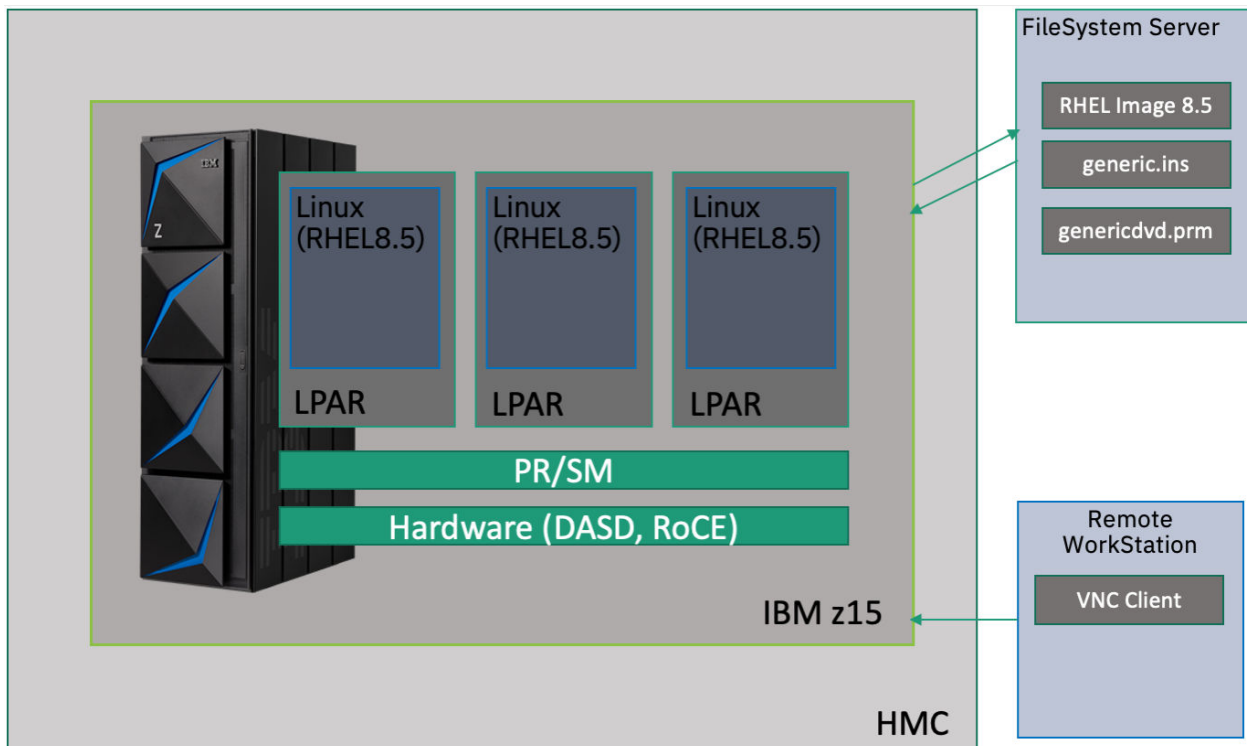
Use VNC to install RHEL

Virtual Network Computing (VNC) allows you to connect to the remote server using a Graphical User Interface (GUI). Use VNC to install Red Hat Enterprise Linux (RHEL) if you are not comfortable using the command line interface.

Prerequisites

Before you install VNC, ensure that your `generic.ins` and `genericdvd.prm` files are created and ready for installation as described in the “[Create generic.ins and genericdvd.prm files](#)” on page 2. In our example, both the `generic.ins` and `genericdvd.prm` files are located in the File System Server, which has access to the LPAR and the Hardware Management Console (HMC).

Architecture



`generic.ins` file:

```
RHEL8.5/images/kernel.img 0x00000000
RHEL8.5/images/initrd.img 0x02000000
RHEL8.5/genericdvd.prm 0x00010480
RHEL8.5/images/initrd.addrsize 0x00010408
```

`genericdvd.prm` file:

```
ro ramdisk_size=40000 cio_ignore=all,!condev
inst.repo=http://example.com/path/rhel8.5/DVD
ip=192.168.0.1::192.168.0.1:15:lpar.example.com:enp010:none
nameserver=192.168.0.1
rd.dasd=0.0.0000
inst.vnc inst.vncpassword=password
```

Use VNC to install RHEL

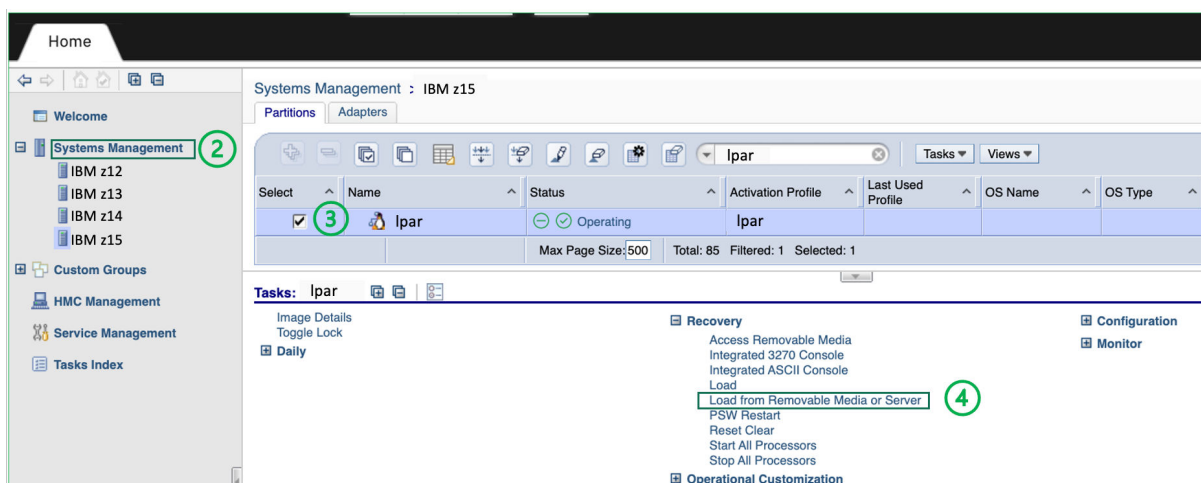
Follow these steps to install RHEL:

1. Log on to the HMC portal with sufficient user privileges.

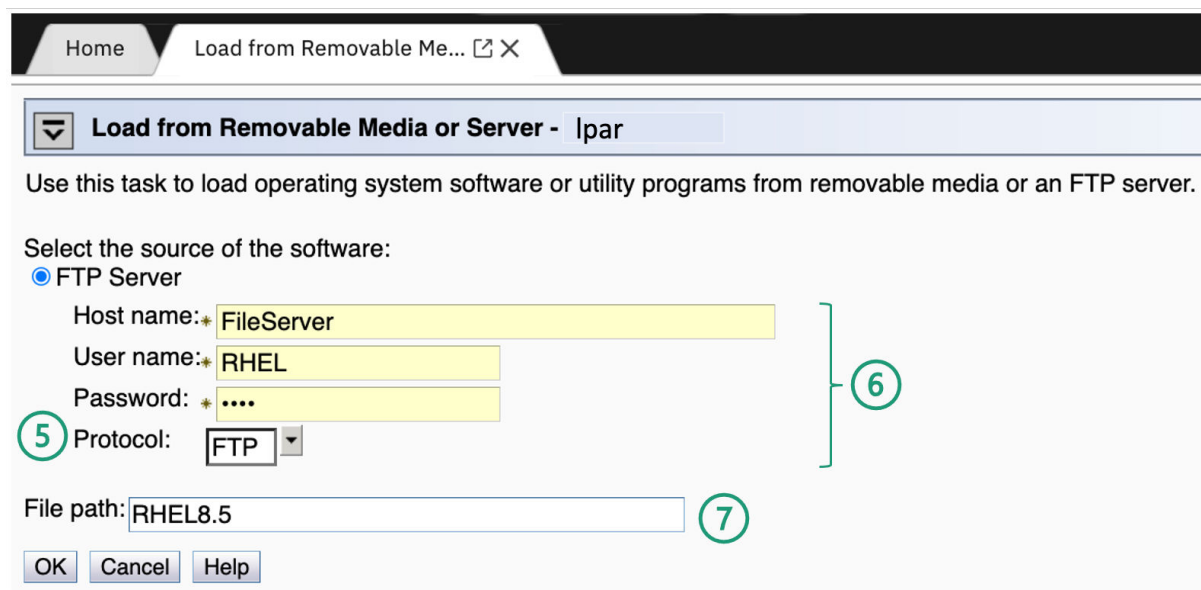
Login to the Hardware Management Console ①



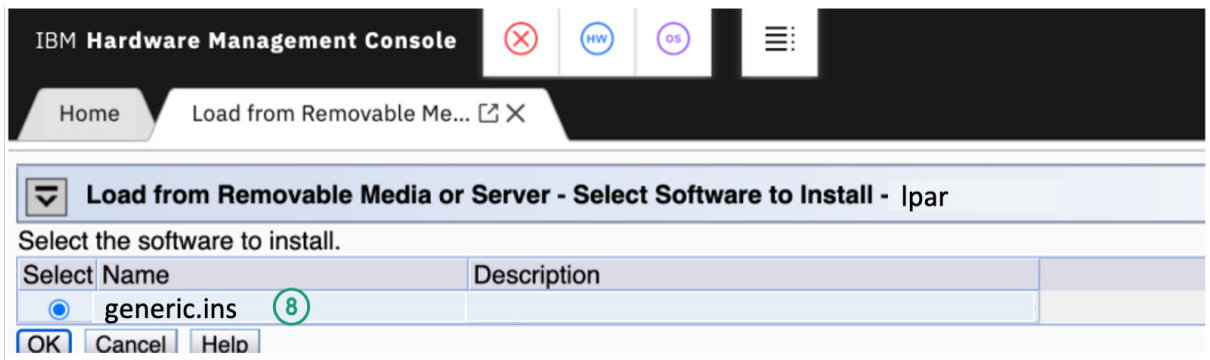
2. Expand **Systems Management**, then select your machine.
3. Check the box next to your selected LPAR.
4. Expand **Recovery**, select **Load from Removable Media or Server**.



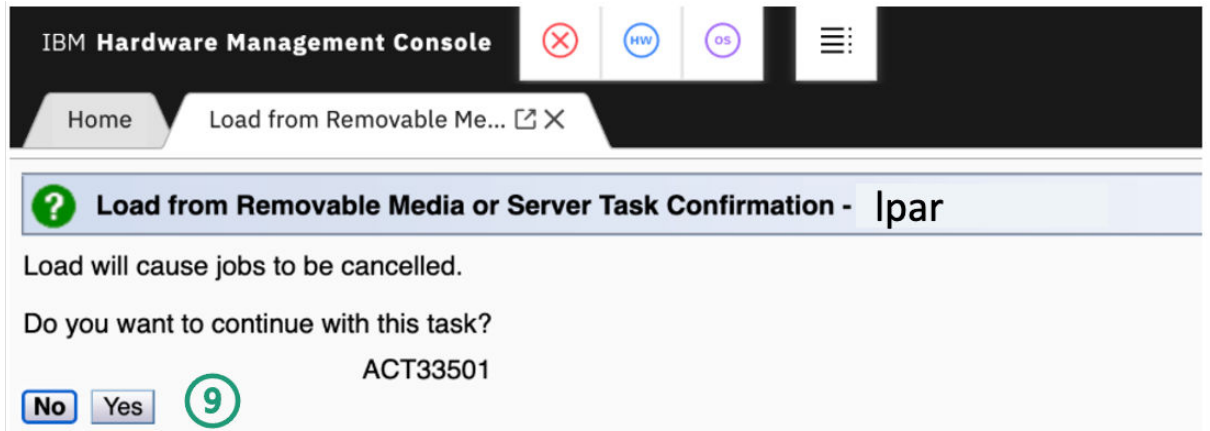
5. In **Protocol**, select FTP. Note that FTPS and SFTP are also supported.
6. Provide the Host name, User name and Password for the server from where the .ins file is loaded.
7. In File path, enter the .ins path, select **OK**.



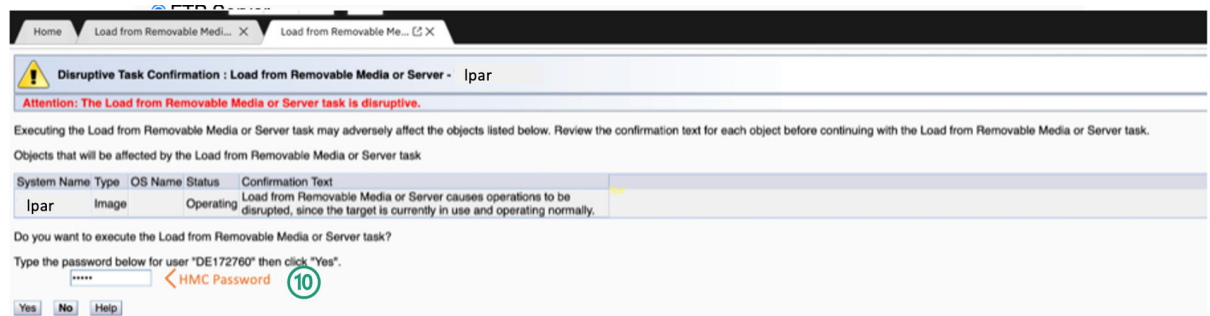
8. A new screen is displayed with your generic .ins file. Select **OK**.



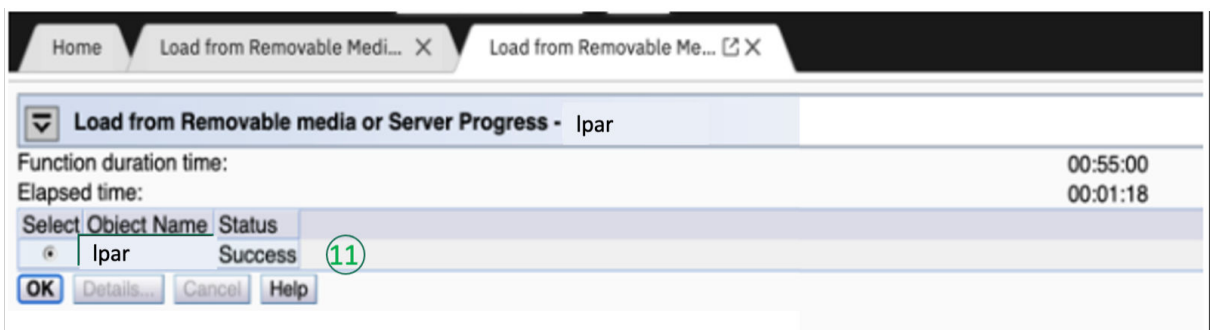
9. Select **Yes**.



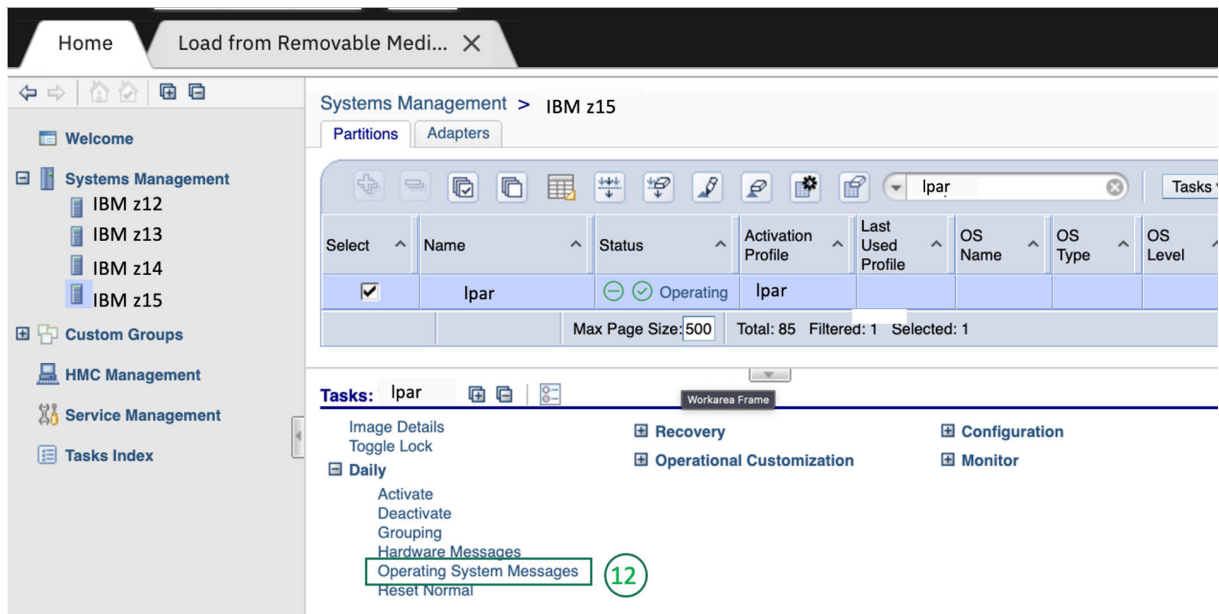
10. Enter the HMC password for confirmation, select **Yes**.



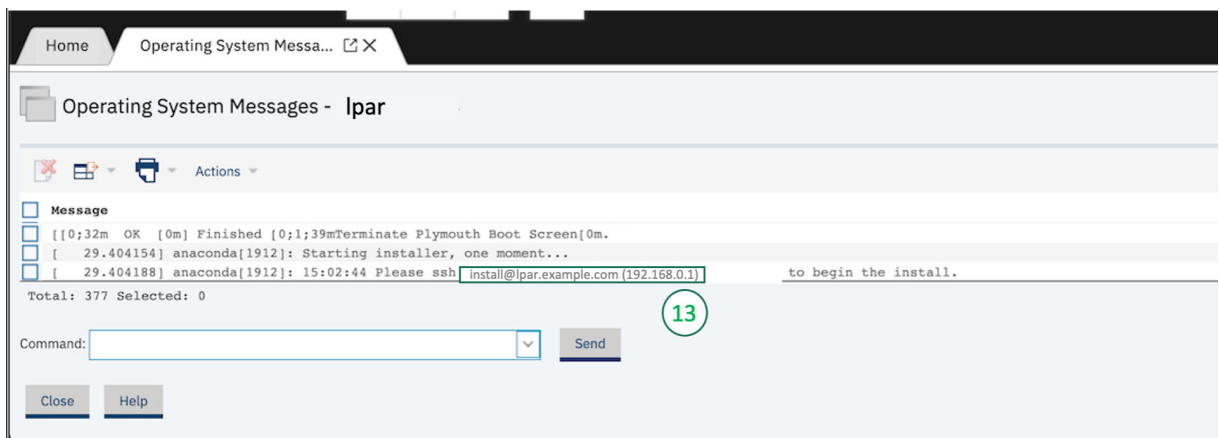
11. The **Load from Removable media or Server Progress** screen displays the time taken to load RHEL image. After the RHEL image has loaded, select **OK** to exit the screen.



12. Select the **Home** tab, expand **Daily**, select **Operating System Messages**.



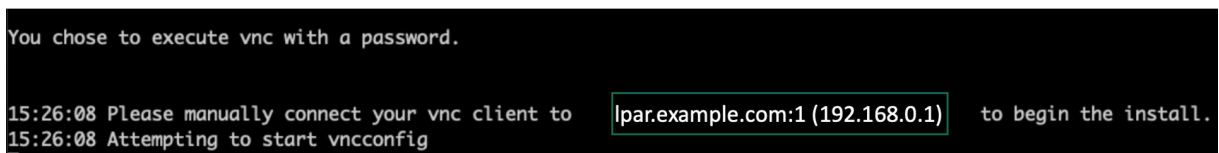
13. Kernel parameters included in the `genericdvd.prm` file are loaded automatically. After the parameters have loaded, the installation process to connect to the new system with IP or hostname is available



14. Open your **Remote System** and type the **ssh** command as displayed in **Operating System Messages**, for example:

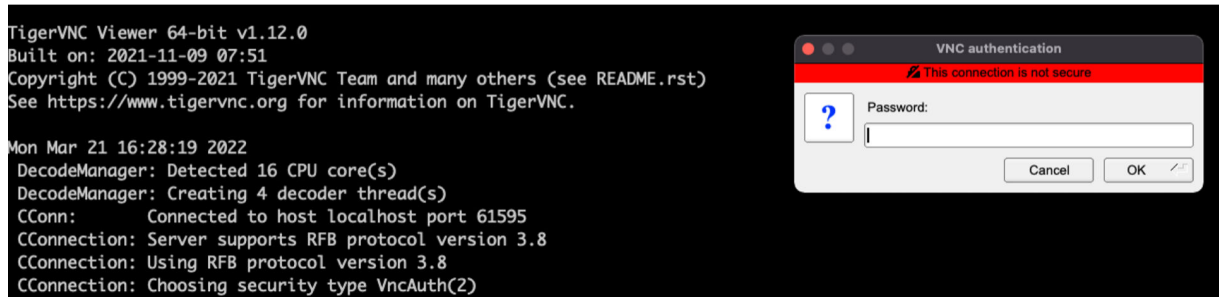
```
ssh install@192.168.0.1
```

15. A new window is displayed with VNC client details **lpar.example.com:1(192.168.0.1)**.



16. Open the VNC terminal to proceed with the installation. For our example, **VNCTiger** is used.

```
TigerVNC Viewer 1.12.0.app % vncviewer 192.168.0.1:1
```



17. The **Welcome to Red Hat Enterprise Linux** installation screen is displayed after a successful VNC connection. Select your language to proceed with the installation.

18. Enter the required data and select **Begin Installation**.

Note: If you use DASD disks, the installer shows the DASD disk as 0B free. This is because the disk is not formatted. Select the disk you want to use, then select **Done**. This starts the formatting process.

19. Select reboot after the installation is complete.

Now your new system is ready to use.

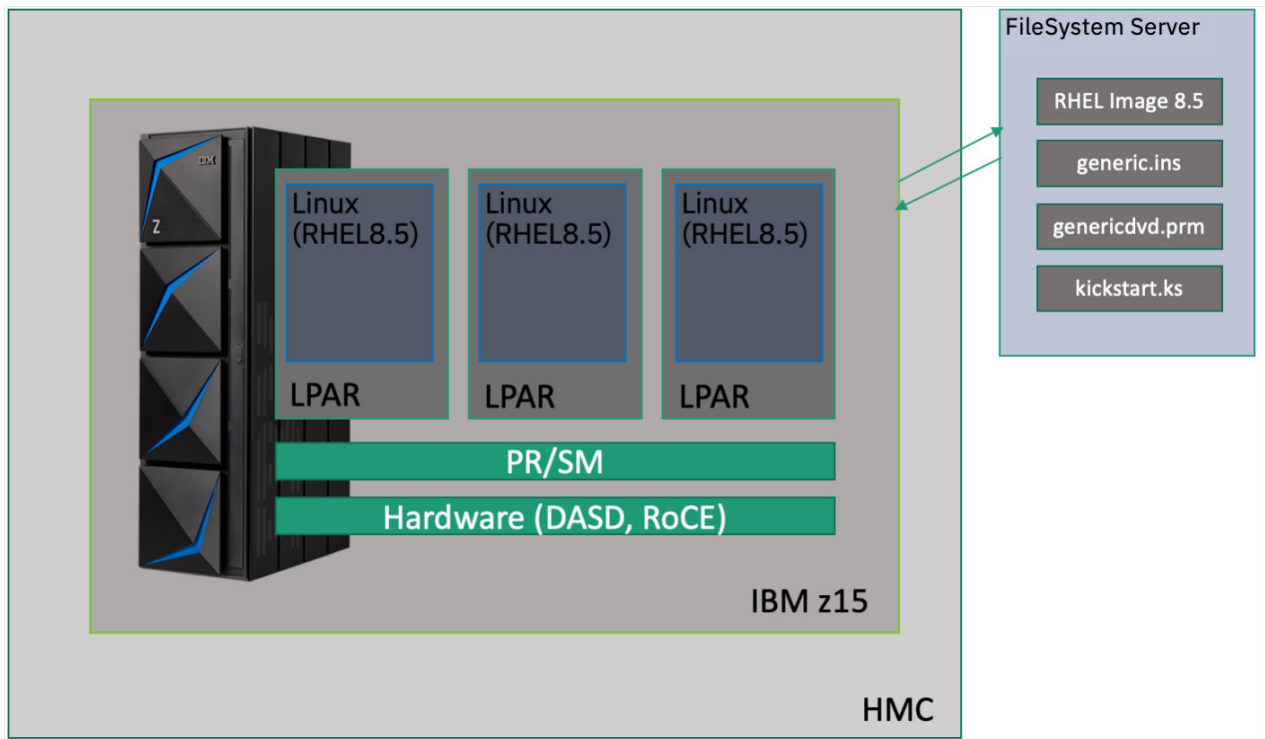
Use Kickstart to install RHEL

A Kickstart file is a simple text file that contains configuration information for a Red Hat Enterprise Linux (RHEL) installation. The system reads this configuration information at boot time and then runs the installation process automatically.

Prerequisites

Before you proceed with the Kickstart installation, ensure that your `generic.ins` and `genericdvd.prm` files are created and ready for installation as described in the “Create `generic.ins` and `genericdvd.prm` files” on page 2. In our example, both the `generic.ins` and `genericdvd.prm` files are located in the FileSystem Server, which has access to the LPAR and the Hardware Management Console (HMC).

Architecture



generic.ins file:

```
RHEL8.5/images/kernel.img 0x00000000
RHEL8.5/images/initrd.img 0x02000000
RHEL8.5/genericdvd.prm 0x00010480
RHEL8.5/images/initrd.addrsz 0x00010408
```

genericdvd.rpm file :

```
ro ramdisk_size=40000 cio_ignore=all,!condev
inst.repo=http://example.com/path/rhel8.5/DVD
ip=192.168.0.1::192.168.0.1:15:lpas.example.com:enp010:none
nameserver=192.168.0.1
rd.dasd=0.0.0000
inst.ks=http://example.com/roce/kickstart.ks
```

Important: Add **inst.ks=** to the genericdvd.prm file to provide your Kickstart file location. In this paper the Kickstart file is located in FileSystem Server.

Tip: Always format storage, such as DASD, before you use it avoid failure during the installation.

There are multiple ways to create a Kickstart file. For our example, the Kickstart file is copied from the /root/anaconda-ks.cfg location of a manually installed RHEL 8.5 and edited as required.

Kickstart.ks file:

```
# Use text install
text --non-interactive

# Use network installation
url --url="http://example.com/path/rhel8.5/DVD/"

# Keyboard layouts
keyboard --xlayouts='us'

# System language
lang en_US.UTF-8

# System timezone
timezone Europe/Berlin
```



```

# Network information
network --hostname=lpar.example.com

# Run the Setup Agent on first boot
firstboot --disable

# Do not configure the X Window System
skipx

# Generated using Blivet version 3.4.0
ignoredisk --only-use=dasda
# Clear the Master Boot Record
zerombr
# Partition clearing information
clearpart --initlabel --cdl --drives=dasda

# Create all partitions automatically
autopart --type=plain --nohome --noboot

# Manually create /boot and / partitions
#part /boot --fstype=ext4 --size=512 --ondisk=dasda
#part / --fstype=ext4 --ondisk=dasda --size 1024 --grow

# Root password
rootpw --plaintext redhat

# Create a non-root user with the password "redhat"
user --name=nonroot --password=nonroot --gecos="Non Root User"

# Disable the firewall
firewall --disable

%packages
@^minimal-environment
@core
kexec-tools
python
%end

%addon com_redhat_kdump --enable --reserve-mb='auto'
%end

%post --log=/root/ks-post.log
#####
# SSH configuration
#####

# allow root login
sed -i -e '/PermitRootLogin/ c\PermitRootLogin yes' /etc/ssh/sshd_config

#####
# Repositories
#####

# Repo Appstream
cat <<-EOF > /etc/yum.repos.d/bistro-appstream.repo
[bistro-appstream]
name=Appstream Bistro
baseurl=http://example.com/path/rhel8.5/DVD/AppStream
enabled=1
gpgcheck=0
skip_if_unavailable=False
EOF
%end

```

Use Kickstart to install RHEL

1. Follow the steps from 1 to 12 as described in [Install RHEL using VNC](#).
2. After the installation has completed, the **Operating System Message** stops at <lpar> login:, as shown here:

Red Hat Enterprise Linux 8.5
Kernel

lpar login:

Total: 359 Selected: 0

Command:

Send

Close

Help

Total: 359 Selected: 0

Close Help

Now your new system is ready to use.

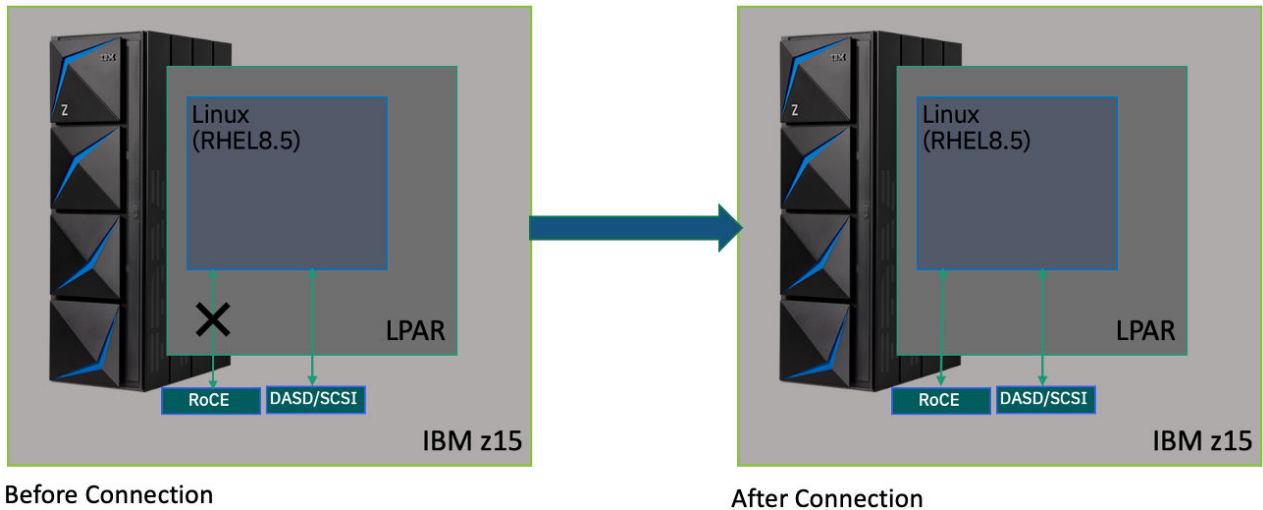
For more information about Kickstart, see the [Kickstart Command reference from RHEL](#).

Chapter 2. Add the additional RoCE interface

In this exercise, you learn how to add PCIe RoCE Express 2 network card (RoCE) in an LPAR guest with Red Hat Enterprise Linux 8.5 (RHEL 8.5). Let's assume a new RoCE is attached to the existing LPAR, and it needs to be configured from scratch by using one of these network managers:

- nmcli
- nmtui

Architecture



Prerequisites

The example in this paper uses this environment:

- IBM z15
- LPAR with RHEL 8.5
- PCIe RoCE Express 2 network card (RoCE) (25GB)

Note:

Use the Hardware Management Console (HMC) to verify that RoCE is attached to the LPAR as described in [“Verify RoCE is attached to your LPAR” on page 1](#).

Important: The **smc-tools** command lists attached RoCE details, powering ON/OFF, checking the connection, and so on. Install **smc-tools** before you configure RoCE. It can be installed from the RHEL repository, for example, `yum install smc-tools`.

```
eg:
# smc_rnics
-----
FID   Power  PCI_ID      PCHID  Type          PPrt  PNET_ID      Net-Dev
-----
6e5   1       0001:00:00.0 0122   RoCE_Express2 0      NET25        ens1765
```

Activate PCIe

As part of the first step, power on the PCI slot to bring RoCE online.

1. Use the `smc_rnic -a` command to list both the online and offline RoCEs attached to your LPAR. In the example, there are two RoCEs:
 - one RoCE is powered on (6e5)
 - one RoCE is powered off (6b5).

Eg:

```
# smc_rnics -a
```

FID	Power	PCI_ID	PCHID	Type	PPrt	PNET_ID	Net-Dev
6b5	0						
6e5	1	0001:00:00.0	0122	RoCE_Express2	1	NET20	ens1765

2. To power on the RoCE, use the `smc_rnics -e` command.

Eg:

```
# smc_rnics -e 6b5
```

3. Use the `smc_rnics` command to verify that the RoCE is powered on.

Eg:

```
# smc_rnics
```

FID	Power	PCI_ID	PCHID	Type	PPrt	PNET_ID	Net-Dev
6b5	1	0000:00:00.0	0111	RoCE_Express2	0	NET20	ens1717
6e5	1	0001:00:00.0	0111	RoCE_Express2	1	NET20	ens1765

Note: For more information about how RoCE got its interface name, check [How to find the RoCE interface name](#).

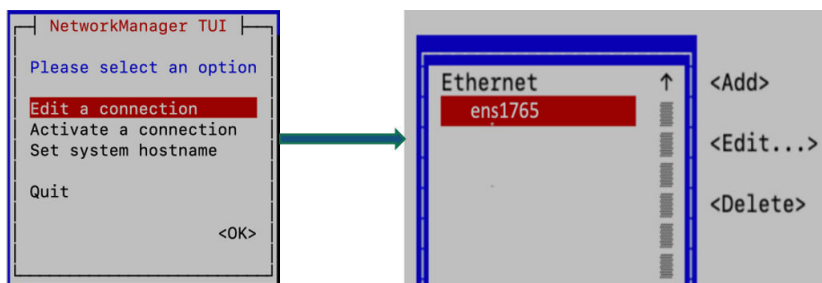
Use nmtui to set up RoCE

To establish a permanent IP for RoCE, you can use `nmtui`.

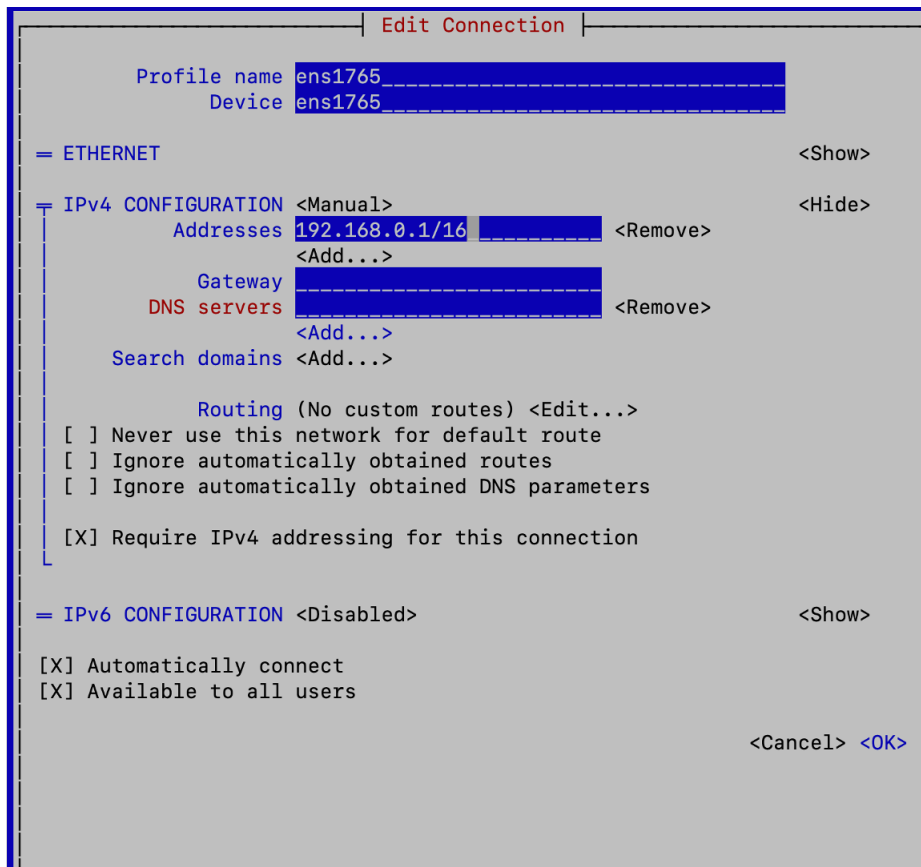
Make sure you have the IPv4 addresses for the configuration and then follow these steps:

1. `# nmtui`

Select **Edit a connection**, then select the RoCE **ens1765**.



2. Expand **IPv4 CONFIGURATION** and update the address with the IP. Select **OK**.



Note: Provide the Gateway, DNS, or SearchDomain as required.

Use nmcli to set up RoCE

To establish a permanent IP for RoCE, you can use **nmcli**. Make sure you have the IPv4 addresses for the configuration and then follow these steps.

1. Enter the command to list the RoCE device, for example:

```
# nmcli c s
NAME      UUID                                  TYPE      DEVICE
ens1765   11111111-1111-1111-1111-111111111111 ethernet  ens1765
```

2. Select the UUID/Name for the RoCE for which you are trying to set up the connection, then provide the IP address for the RoCE.

```
nmcli c e 11111111-1111-1111-1111-111111111111
nmcli> set ipv4.addresses 192.168.0.1/16
nmcli> save persistent
nmcli> quit
nmcli c down ens1765
nmcli c up ens1765
```

Verify the nmtui and nmcli setup

After the RoCE setup is complete, use **nmcli** to verify the setup:

```
# nmcli
ens1765: connected to ens1765
"Mellanox MT27710"
ethernet (mlx5_core), 82:28:9B:1B:28:2C, hw, mtu 1500
inet4 192.168.0.1/16
route4 192.168.0.1/16
```

```
inet6 fe80::d8c5:3a67:1abb:dcca/64  
route6 fe80::/64
```

References

- Red Hat documentation:
 - [Installing RHEL in an LPAR](#)
 - [RHEL Kickstart Reference](#)
- RoCE Express:
 - [Networking with RoCE Express on Linux on IBM Z \(mainframe\)](#)
 - [PCIe Express 2 support](#)

Notices

References in this publication to IBM products, programs, or services do not imply that intends to make these available in all countries in which IBM operates. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Subject to IBM's valid intellectual property or other legally protectable rights, any functionally equivalent product, program, or service may be used instead of the IBM product, program, or service. The evaluation and verification of operation in conjunction with other products, except those expressly designated by IBM, are the responsibility of the user. IBM may have patents or pending patent applications covering subject matter in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
USA

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently IBM created programs and other programs (including this one) and (ii) the mutual use of the information, which has been exchanged, should contact:

IBM Deutschland Research & Development GmbH
Department 3282
Schönaicher Strasse 220
D-71032 Böblingen
Federal Republic of Germany
Attention: Information Request

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

Any pointers in this publication to websites are provided for convenience only and do not in any manner serve as an endorsement of these websites. Use of these materials is at your own risk.

Trademarks and service marks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at "Copyright and trademark information" at www.ibm.com/legal/copytrade.shtml.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

The registered trademark Linux® is used pursuant to a sublicense from the Linux Foundation, the exclusive licensee of Linus Torvalds, owner of the mark on a worldwide basis.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Red Hat®, JBoss®, OpenShift®, Fedora®, Hibernate®, Ansible®, CloudForms®, RHCA®, RHCE®, RHCSA®, Ceph®, and Gluster® are trademarks or registered trademarks of Red Hat, Inc. or its subsidiaries in the United States and other countries.

UNIX is a registered trademark of The Open Group in the United States and other countries.
Other company, product, and service names may be trademarks or service marks of others.

