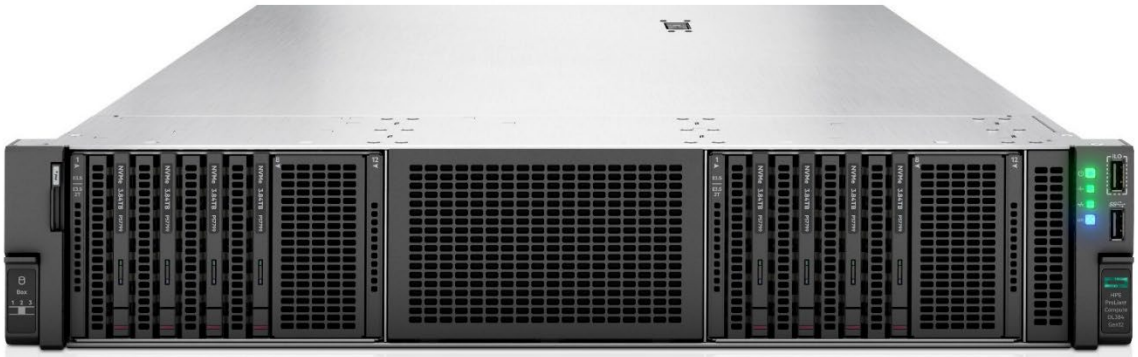
Technical white paper

HPE ProLiant Compute DL384 Gen12 Server

Quick Start Guide



Contents

Abstract

HPE ProLiant Compute DL384 Gen12 is the first server from HPE enabled with NVIDIA GH200 NVL2, optimized for AI inferencing for large language models that require a large memory capacity and/or non-AI workloads like large scale simulation, EDA, weather forecasting, etc.

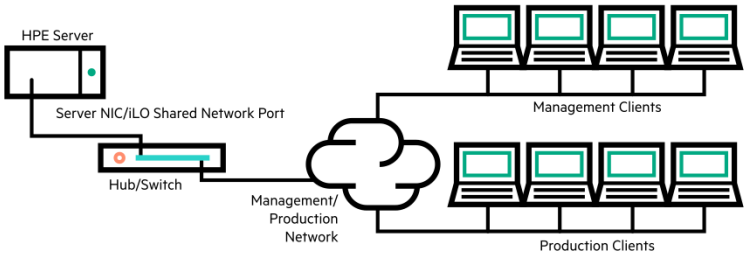
With up to 1.2TB of fast unified memory and up to 10TB/s of bandwidth the NVIDIA GH200 NVL2 is designed to handle large language models fine-tuning and inferencing with Retrieval Augmented Generation (RAG) featuring twice the performance of the previous generation. The HPE ProLiant Compute DL384 Gen12 delivers the best performance per GPU in our HPE ProLiant portfolio, and is ideal for mixed or memory intensive workloads, whether you are doing AI or traditional HPC.

Building on HPE ProLiant as the legendary foundation, the HPE ProLiant Compute DL384 Gen12 delivers a consistent experience across the HPE ProLiant portfolio with HPE iLO management and firmware that delivers robust reliability and security with silicon Root of Trust technology from HPE.

This white paper outlines the ease of deployment utilizing HPE provided tools and utilities such as HPE Integrated Lights-Out (HPE iLO), the HPE Solution Delivery Repository (SDR), and NVIDIA® CUDA® and GPU driver repositories.

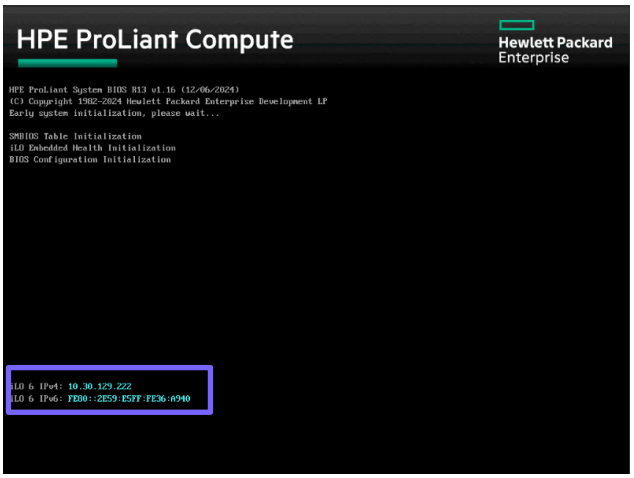
Initial Configuration

The HPE ProLiant DL384 Gen12 features a robust baseboard management controller HPE calls Integrated Lights-Out, or HPE iLO. After the initial racking and cabling of the unit, it is essential to configure HPE iLO for remote management.

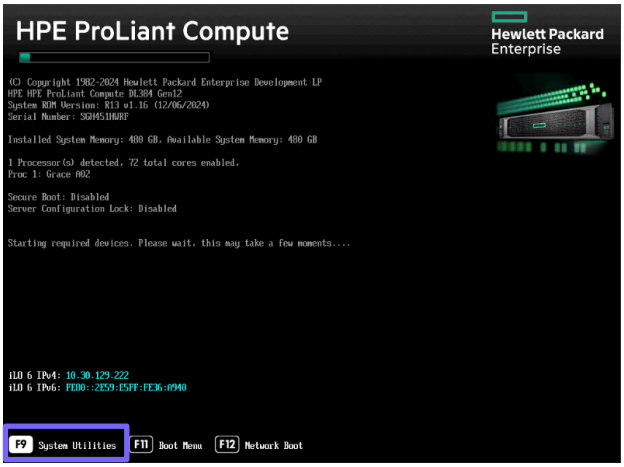


In the default configuration, HPE iLO will acquire an IP address automatically via DHCP. If the DHCP server is accessible to the Administrator, the iLO IP address can be obtained via that mechanism, or the installer can connect a keyboard and monitor to the HPE ProLiant DL384 for first time setup.

With a keyboard and monitor connected, power-on the unit to view the initial Power-On Self-Test (POST). If DHCP is enabled on the management network, the iLO IP address will be displayed in the bottom left corner of the video output.

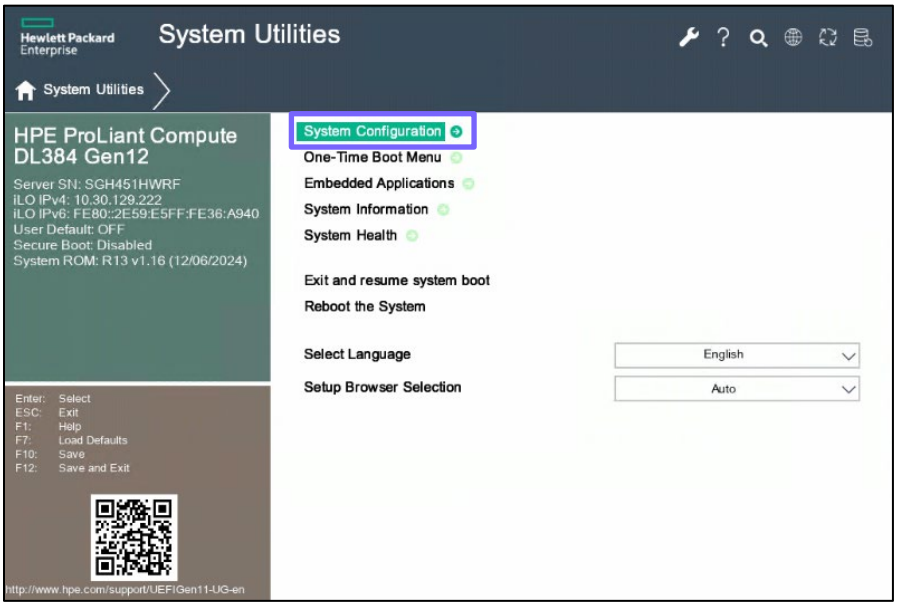


After the early start POST screens, the system will display options to enter System Utilities, a Boot Menu, or Network Boot. If no button is pressed, the system will progress through the default boot order. To enter the System Utilities, press F9.

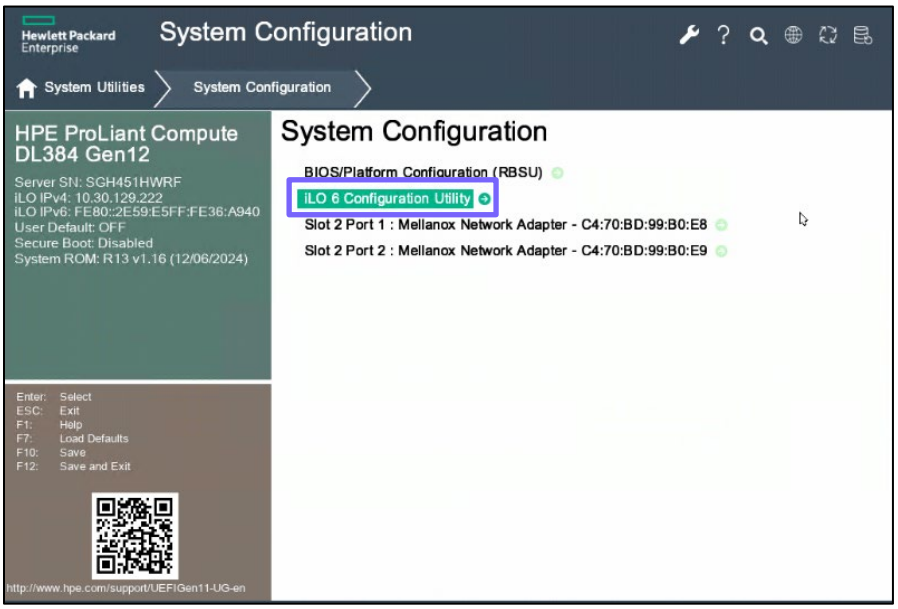


HPE iLO Network Configuration

The System Utilities menu allows administrators to configure both HPE iLO and the system ROM/UEFI. To modify HPE iLO default settings, select System Configuration.

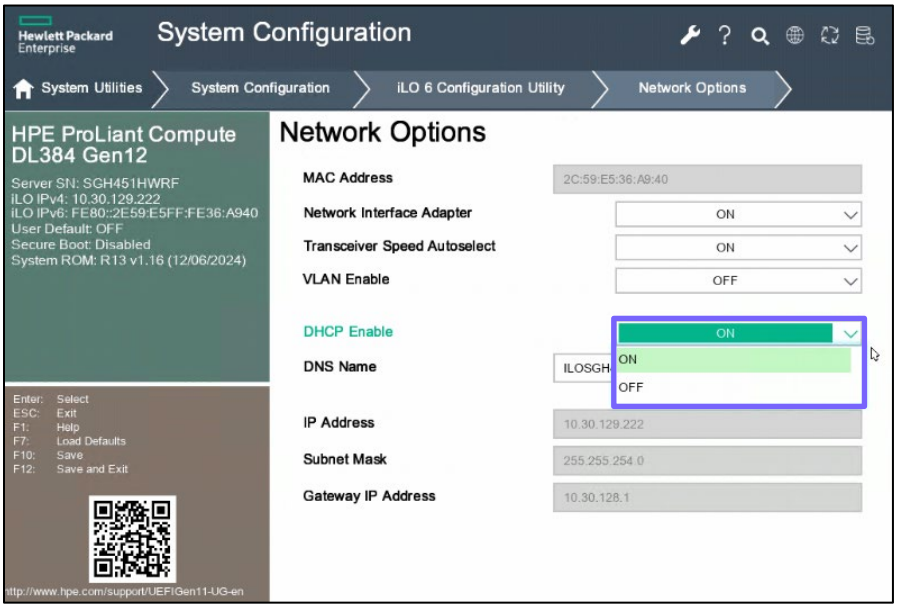


Select iLO 6 Configuration Utility

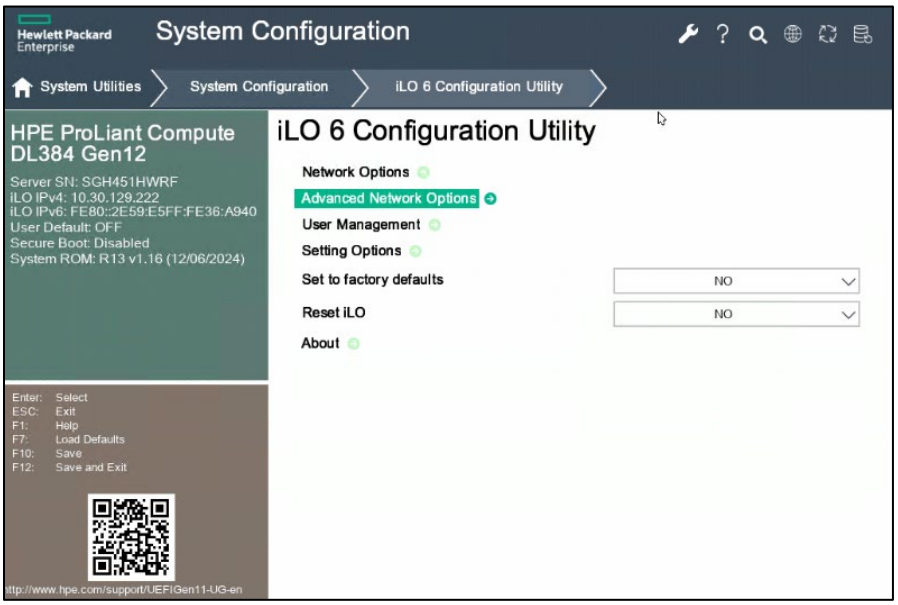


To modify the network configuration, select Network Options.

The Network Options screen allows administrators to disable the default DHCP enabled setting, and assign a static IP Address, Subnet Mask and Gateway IP address.



By returning to the iLO 6 Configuration Utility menu, select the Advanced Network Options menu



This Advanced Network Options menu allows for modification of:

– Static Routes

– DNS Servers

– WINS Servers

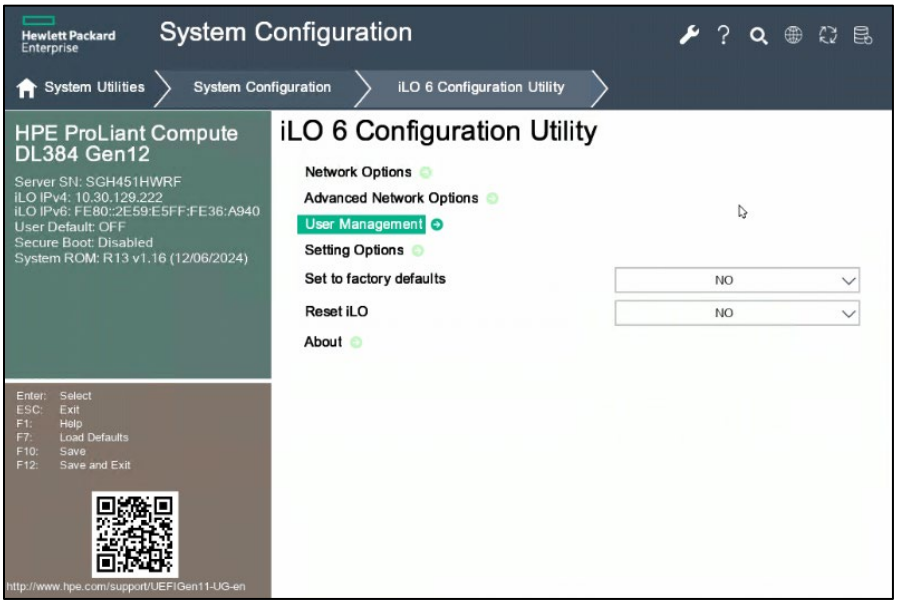
– Fully Qualified Domain Name

These settings and more are also available through the HPE iLO web interface.

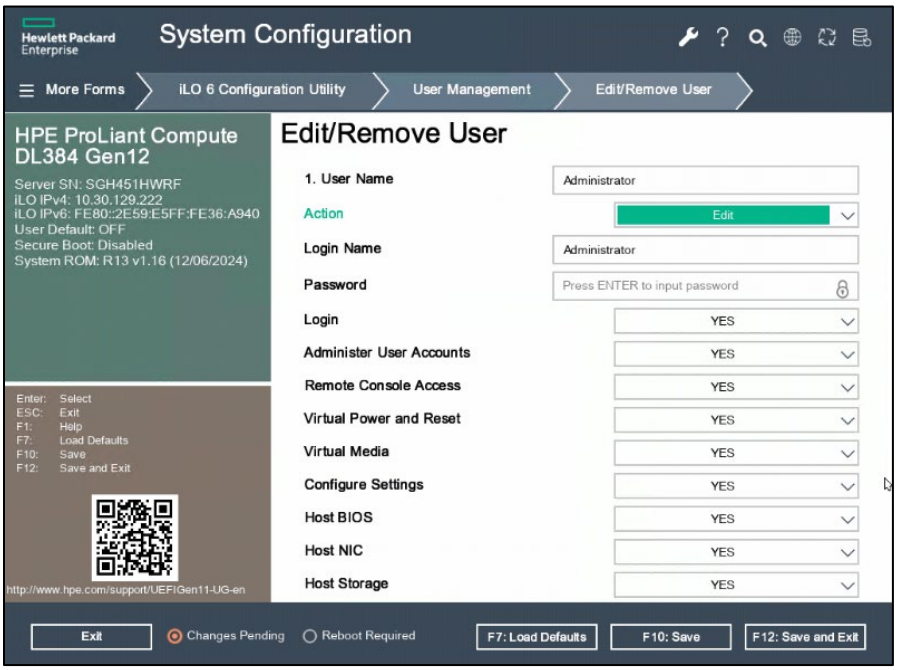
HPE iLO Default Credential Modification and User Creation

Unless specifically requested through HPE Factory Express, every HPE ProLiant Compute server ships with a unique default password for the HPE iLO Administrator account. The pull tag on the front of the server contains the product part number, serial number, and factory applied HPE iLO Administrator account password.

To change the default password in the iLO 6 Configuration Utility, navigate to User Management.



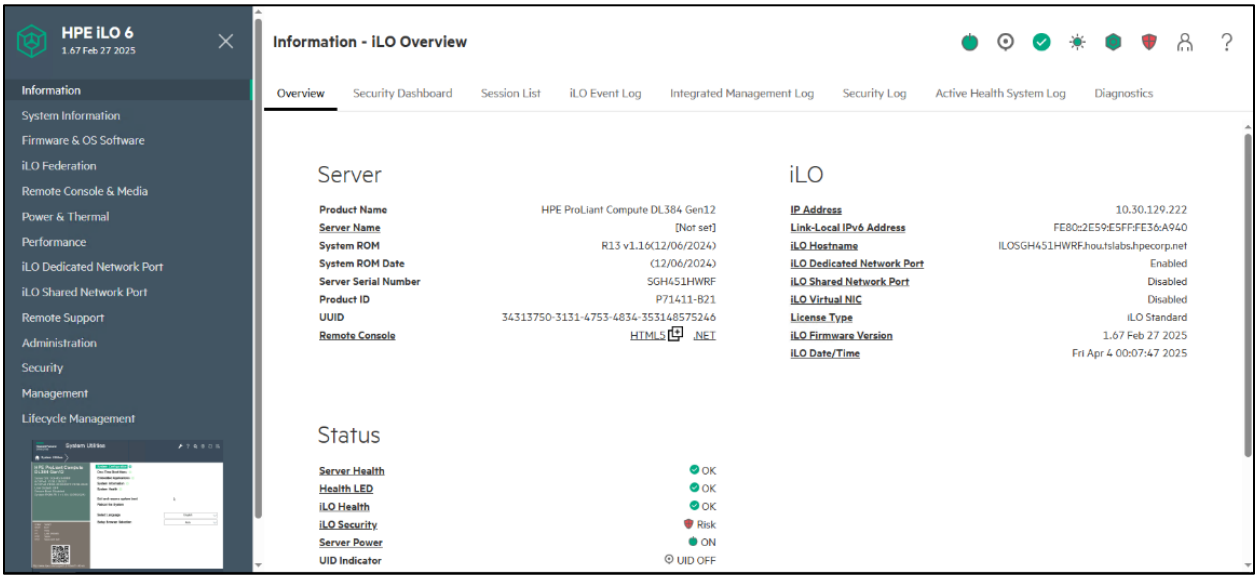
Selecting Edit/Remove User, the Administrator account may be modified.



After the basic steps of acquiring the DHCP IP address or setting a static IP address, and optionally modifying the default Administrator account password, additional configuration and day-to-day operation of the HPE ProLiant DL384 Gen12 platform may be done remotely through a browser and HPE iLO web interface.

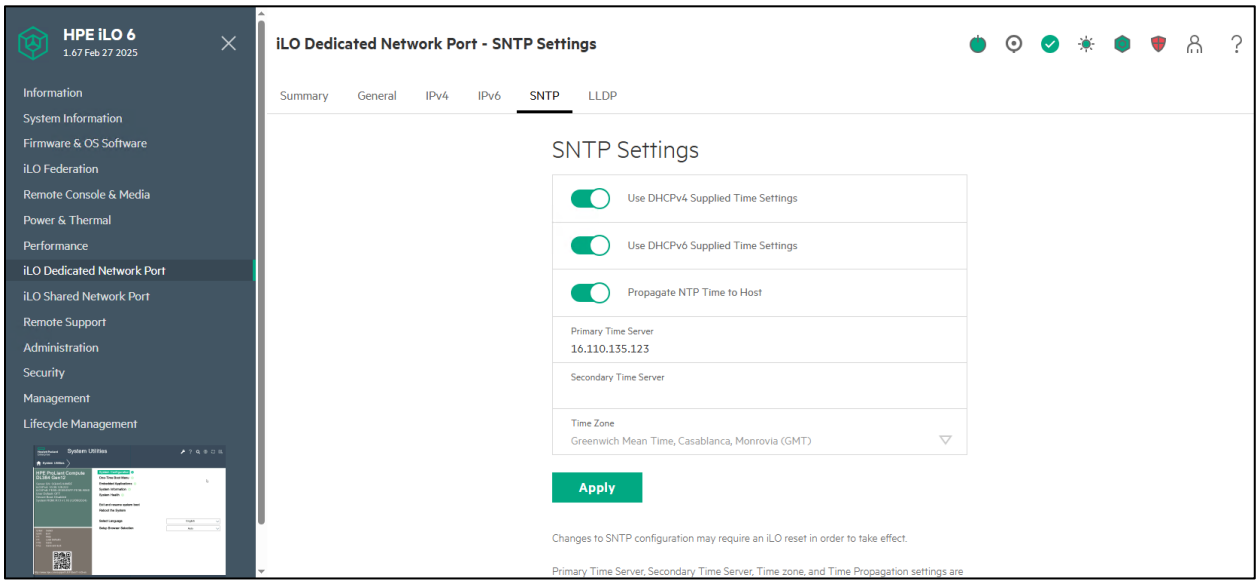
HPE System Inventory

After logging into iLO for the first time, users are greeted with the iLO Overview screen. This displays information about the server such as Serial Number and Product ID. The iLO summary section shows the IP address, Fully-Qualified Domain Name, Licensing Status and the iLO Date and Time.



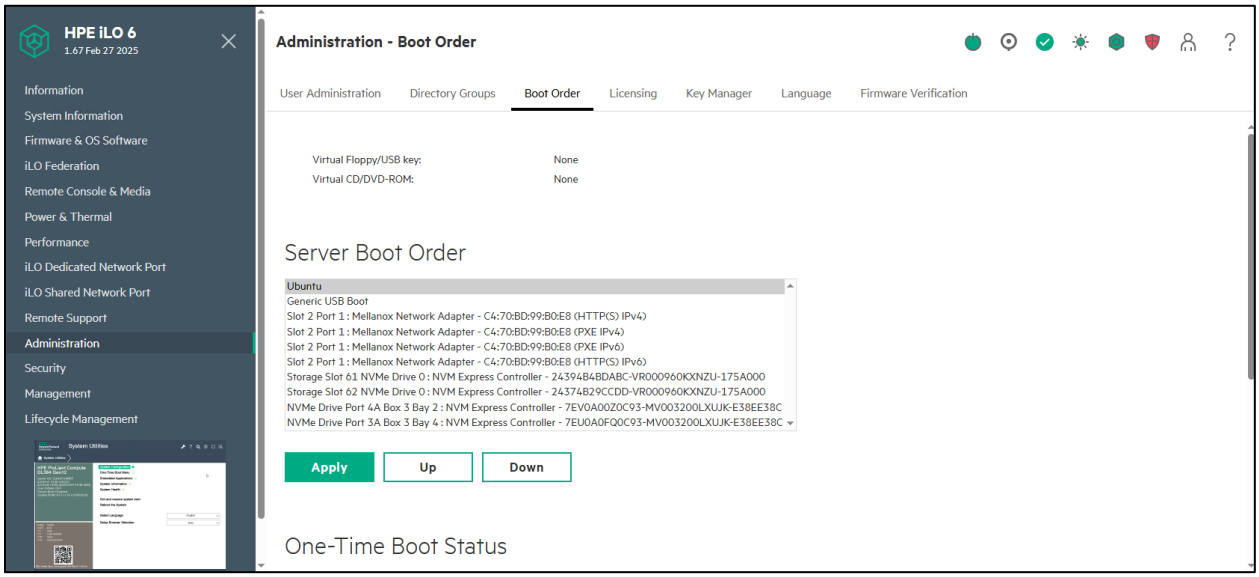
HPE iLO Dedicated Network Port

All network configuration options for the HPE iLO are available through the HPE iLO Dedicated Network Port menu. It is important to ensure proper DNS, and Network Time Protocol settings are applied.

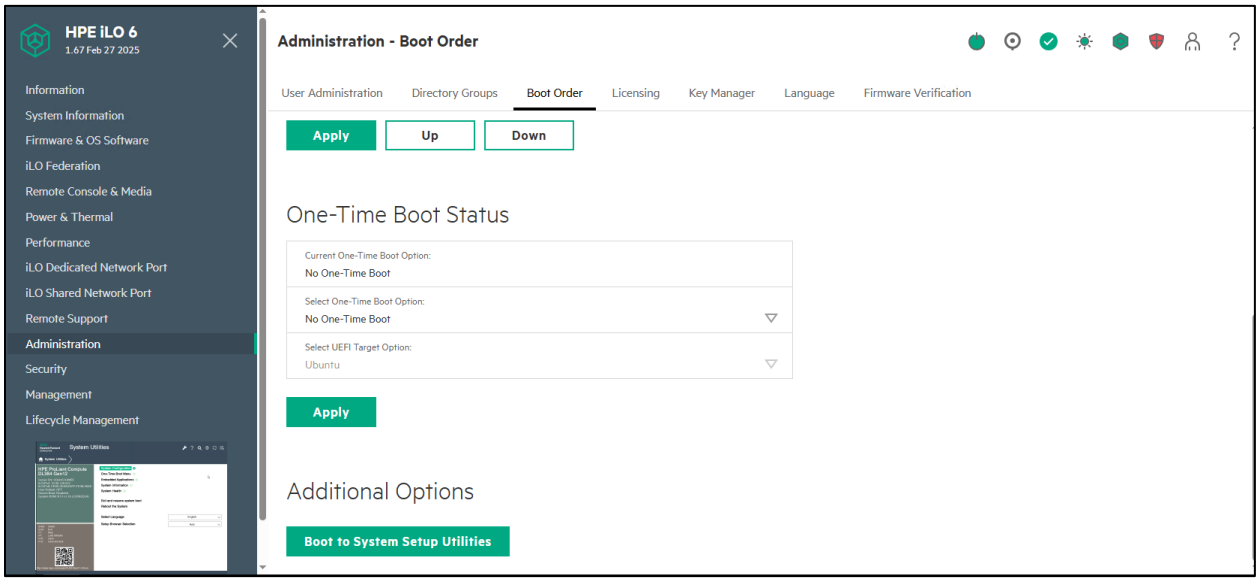


HPE System Boot Order

Through HPE iLO, administrators may choose to change the default or one-time boot order by selecting the Administration tab in the left-hand navigation screen and selecting Boot Order.



In this screen is an option to force the unit to boot to the System Setup Utilities the next time the system is powered on, or rebooted. This is helpful as the administrator does not need to watch the server at POST to press F9 to enter the utility.



Applying a HPE iLO License

Utilizing the HPE iLO web interface, an HPE iLO Advanced Key may be applied to the server to enable key features including:

– Integrated Remote Console (beyond POST and UEFI menu operation)

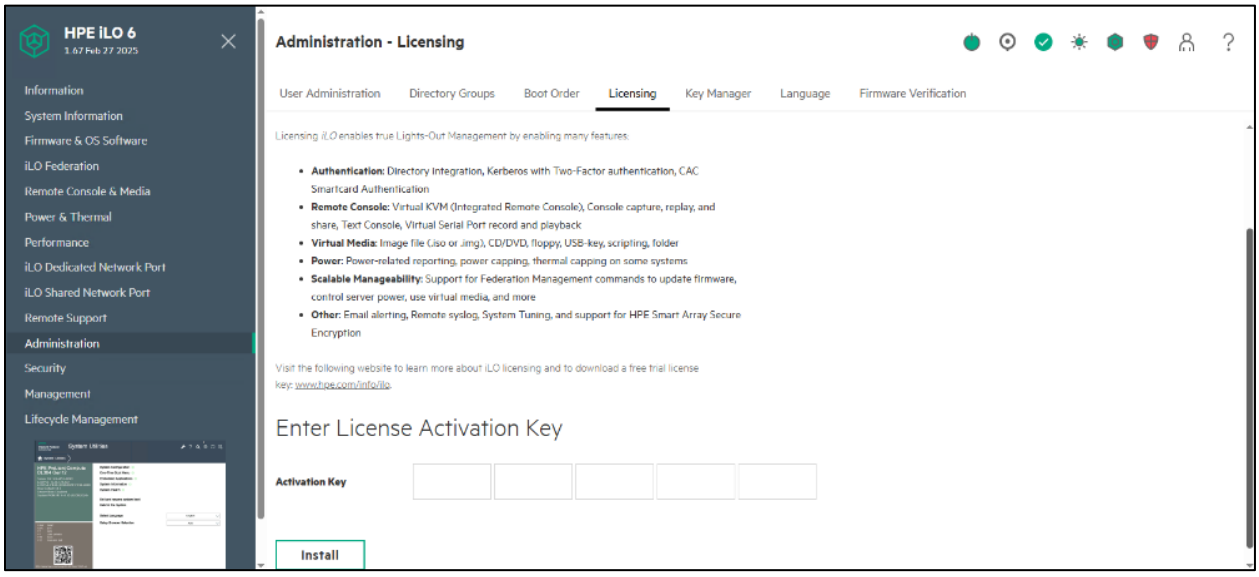
– Virtual Media mount

– Active Directory and LDAP integration

For more features and information see the HPE iLO Licensing Guide at:
support.hpe.com/hpesc/public/docDisplay?docId=sd00002201en_us

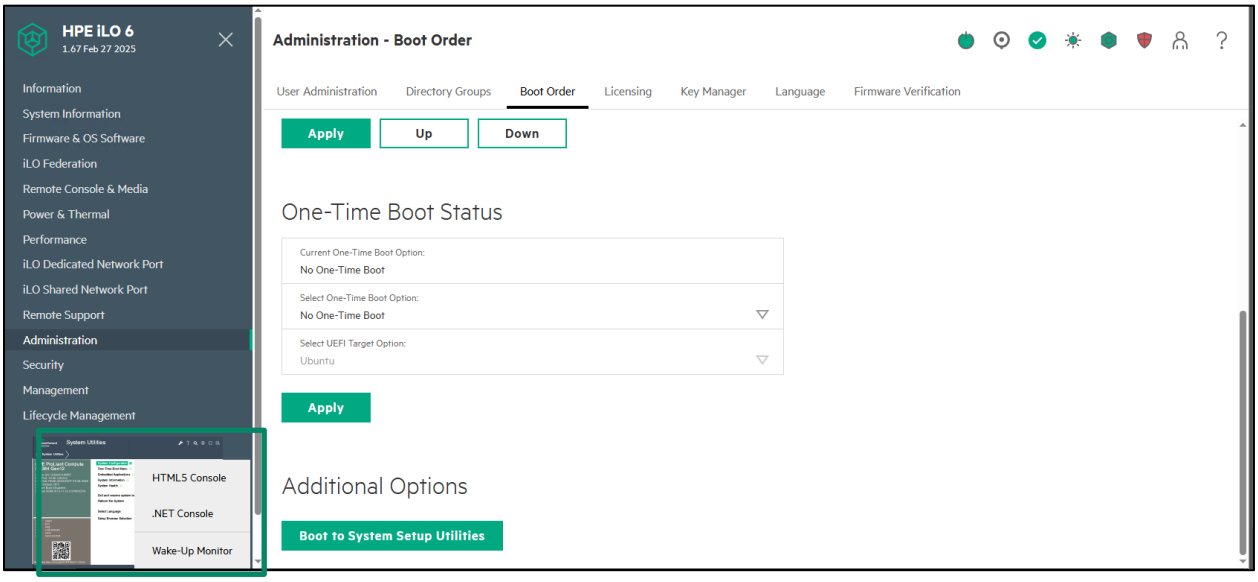
Trial licenses valid for 90-days are available at:

hpe.com/us/en/resources/integrated-systems/ilo-advanced-trial.html

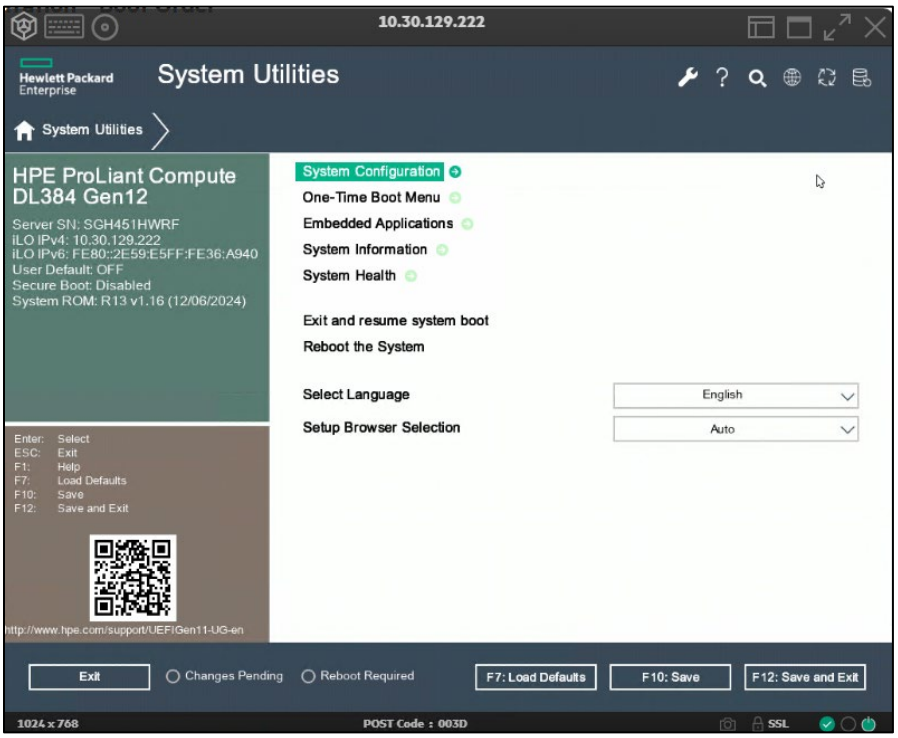


HPE iLO Remote Console

The iLO Remote Console can be launched by clicking the thumbnail in the bottom left corner of the iLO web interface.

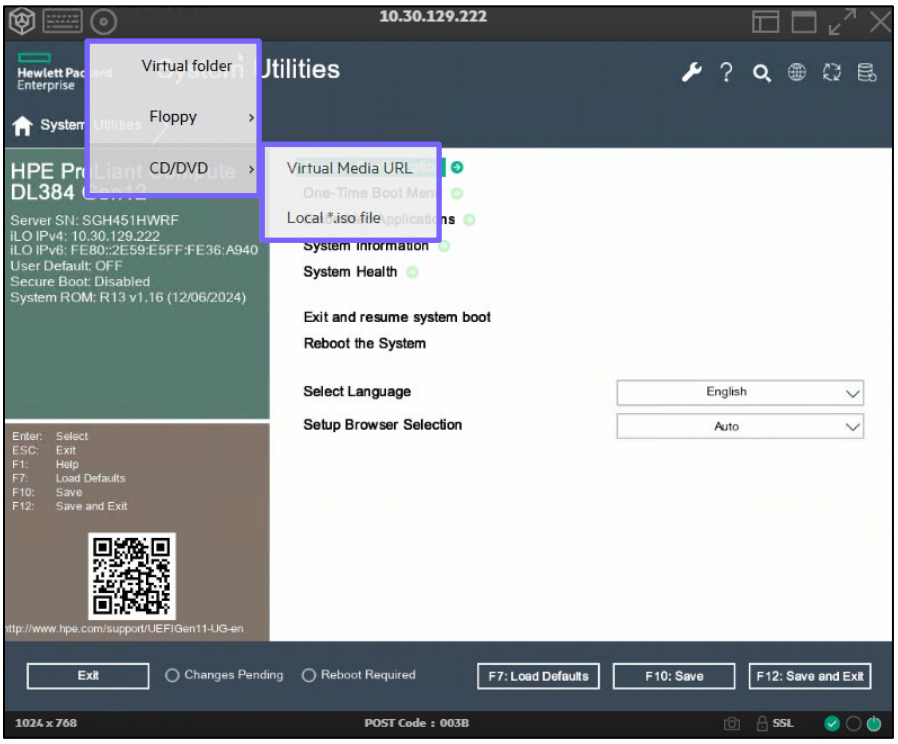


The HTML5 Console option is recommended as it does not require any additional utilities such as Microsoft .NET on the system navigating the iLO web console. This remote console allows for navigation of the system remotely, power on/off buttons and Virtual Media capability.

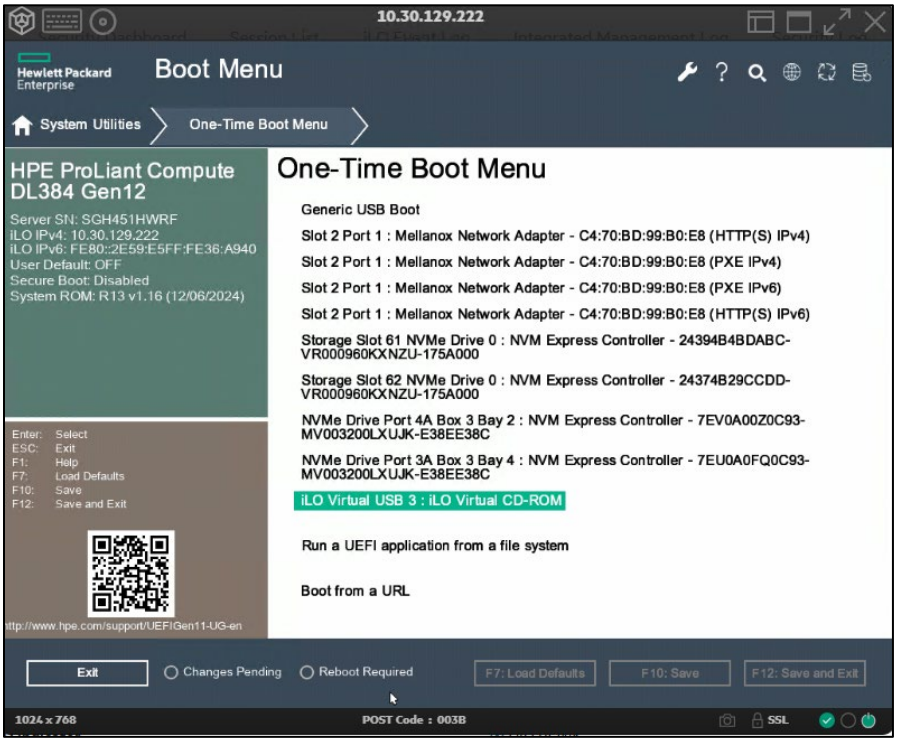


Using HPE iLO Virtual Media

Use the iLO Virtual Media function to mount the ISO, either from a http(s) file share or from a folder located on the local machine accessing the iLO Remote Console.



The One-Time Boot Menu can be helpful forcing the server to boot from the iLO Virtual Media rather than waiting for the system to progress through all PXE enabled NICs and other possible boot sources. Press F11 at POST or navigate to the One-Time Boot Menu through the System Utilities and select the iLO Virtual USB 3: iLO Virtual CD ROM option.



Deploying Canonical Ubuntu Server 22.04 and 24.04

The following section will highlight important steps in deploying Canonical Ubuntu 22.04 and 24.04 on the HPE ProLiant Compute DL384 Gen12 server. It is not intended to provide advice for all installation settings as those vary by use case and organization.

Obtaining the Proper Installation Media Supporting 64k Page Size Kernels

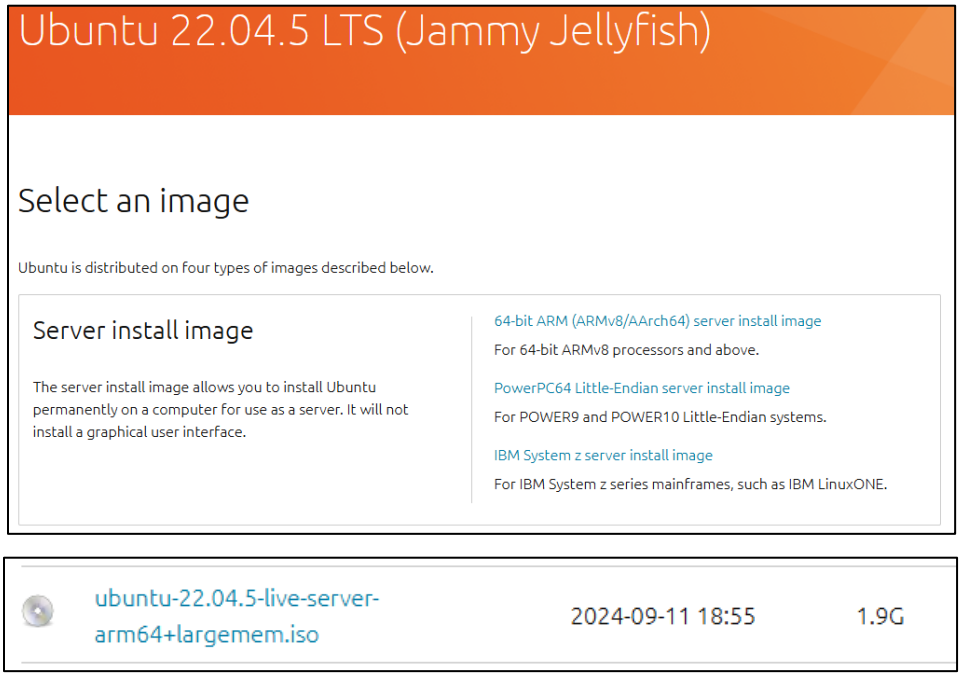
Selecting the proper installation media is necessary to ensure compatibility with the ARM processor contained in the NVIDIA GH200 NVL2 Superchip.

Visit the Canonical download site and click on Alternative architectures to obtain an ARM compatible release.

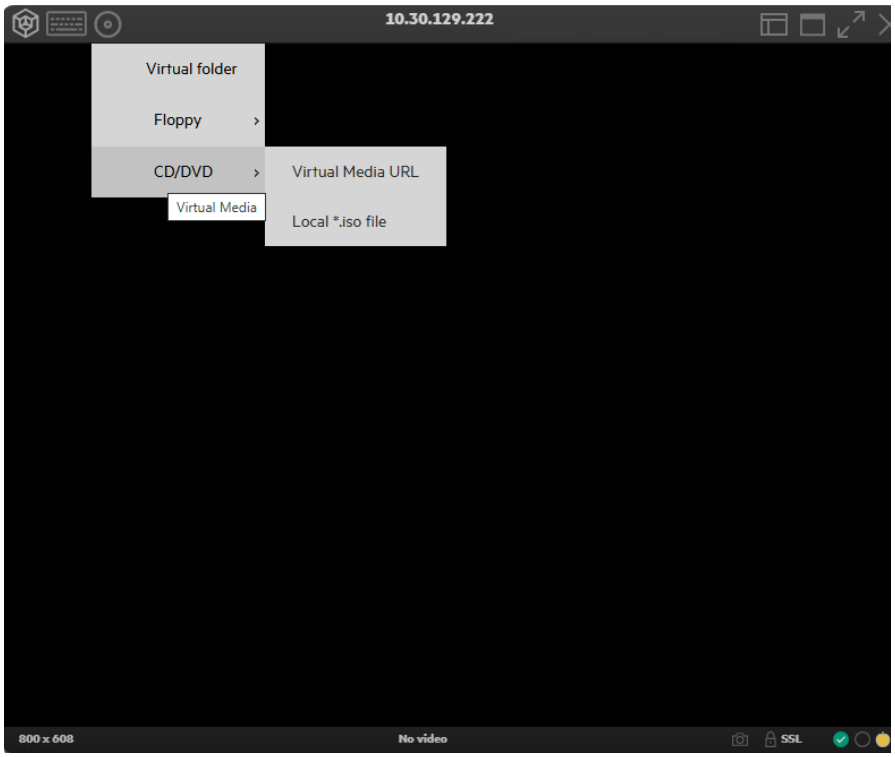
For the most current release visit ubuntu.com/download/server/arm

Older versions such as Ubuntu 22.04 can be downloaded here cdimage.ubuntu.com/releases/

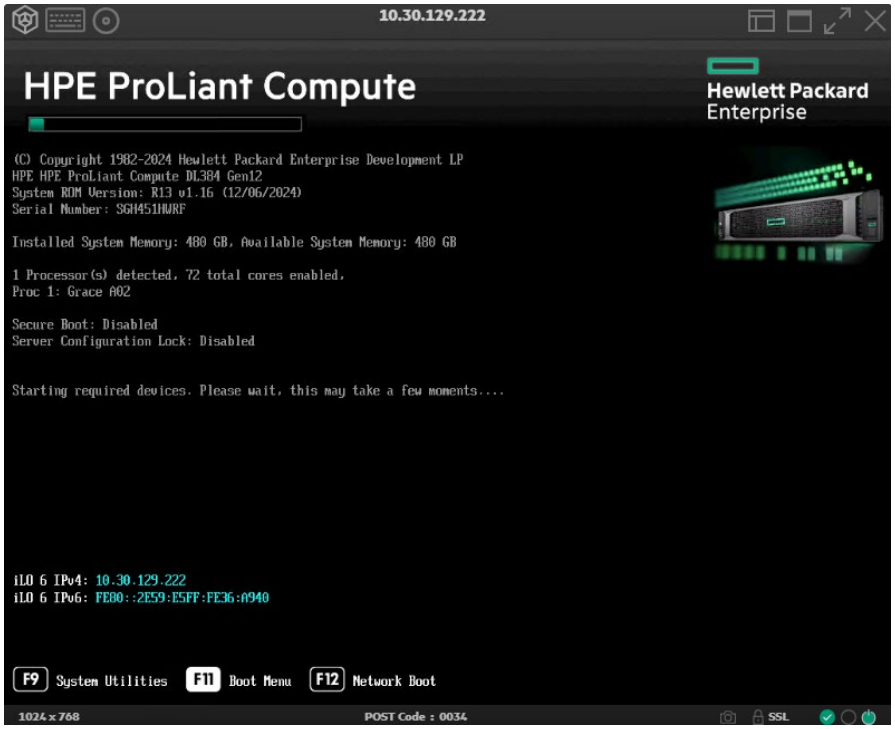
To ensure optimal performance, download installation media specifically built to support 64k page sizes.



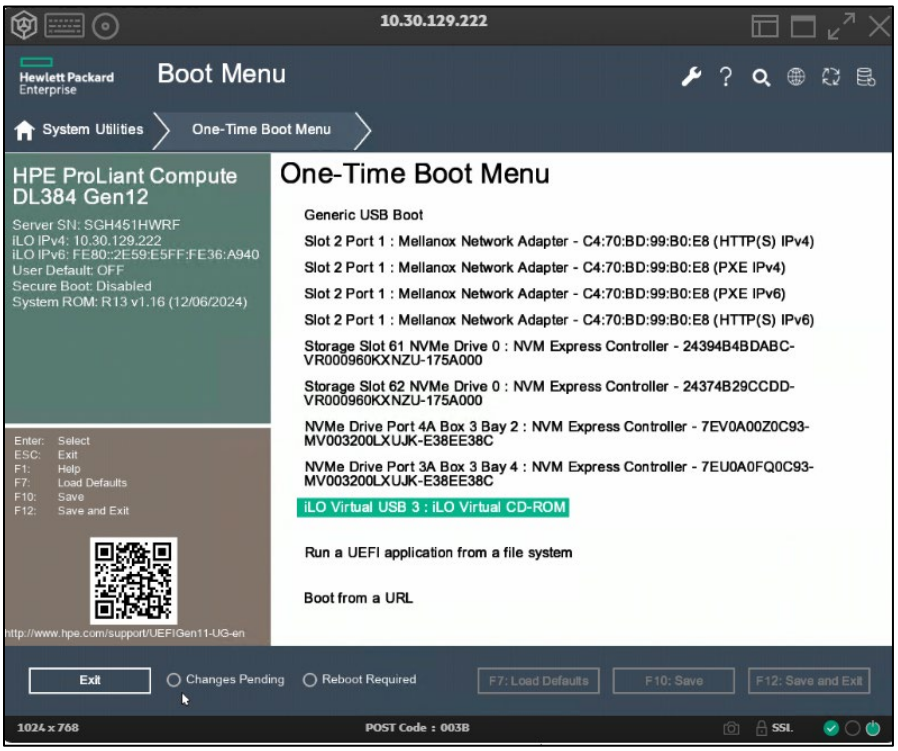
Use the HPE iLO web interface to launch the integrated remote console. Use the iLO Virtual Media function to mount the installation ISO. If the server is not already powered on, use the virtual power button to boot the server.



During POST, users are presented with an option to press F11 to enter a One-Time Boot Menu.



Select the iLO Virtual USB 3: iLO Virtual CD-ROM option to boot the mounted ISO.



With the proper 64k page size compatible installation media mounted, proceed through the installation wizard providing specific network, user, and package selections for the intended use of the system.

Subscribe the Ubuntu Server to the AIP repository

After the initial operating system installation, HPE recommends adding the HPE ProLiant Compute Advanced Integration Pack available from the HPE Software Delivery Repository. This online repository contains drivers, firmware and utilities to assist in managing and lifecycle management of ARM based systems like the HPE ProLiant Compute DL384 Gen12 server.

Additional information on the HPE Software Delivery Repository (SDR) can be found here:

downloads.linux.hpe.com/SDR/getting_started.html

downloads.linux.hpe.com/SDR/project/aip/

Install the HPE Public GPG Key and Update the Local Apt Indexes

The software packages provided by the SDR are cryptographically signed to ensure authenticity. It is necessary to enroll HPE's public keys in Ubuntu's package manager apt. While there are 3 available public keys, the one applied in the following steps are for packages published after 2024.

First download the HPE Public Key, and convert it to GPG format with the following command:

```
curl -fsSL https://downloads.linux.hpe.com/SDR/hpePublicKey2048_key2.pub | sudo gpg --
dearmor -o /usr/share/keyrings/hpePublicKey2048_key2.gpg
```

Run the following command to create a HPE AIP sources file:

```
echo "deb [arch=arm64 signed-by=/usr/share/keyrings/hpePublicKey2048_key2.gpg]
https://downloads.linux.hpe.com/SDR/repo/aip jammy/current non-free" | sudo tee -a
/etc/apt/sources.list.d/hpeaip.list
```

Clean and reload the apt repository by running the following commands:

```
sudo apt clean

sudo apt update
```

To install utilities such as the iLORest Tool run:

```
sudo apt-get install ilorest.
```

```
hpeadmin@dl384:~$ sudo apt-get install ilorest
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
The following additional packages will be installed:
  javascript-common libjs-jquery libjs-sphinxdoc libjs-underscore python3-decorator python3-ilorest
  python3-jsondiff python3-jsonpath-rw python3-pbr python3-ply python3-prompt-toolkit python3-pyaes
  python3-pygments python3-setproctitle python3-tabulate python3-wcwidth
Suggested packages:
  apache2 | lighttpd | httpd python-ply-doc python-pygments-doc ttf-bitstream-vera
The following NEW packages will be installed:
  ilorest javascript-common libjs-jquery libjs-sphinxdoc libjs-underscore python3-decorator python3-ilorest
  python3-jsondiff python3-jsonpath-rw python3-pbr python3-ply python3-prompt-toolkit python3-pyaes
  python3-pygments python3-setproctitle python3-tabulate python3-wcwidth
0 upgraded, 17 newly installed, 0 to remove and 98 not upgraded.
Need to get 2,076 kB of archives.
After this operation, 10.5 MB of additional disk space will be used.
```

To install the Agentless Management Service run:

```
sudo apt-get install amsd
```

```
hpeadmin@dl384:~$ sudo apt-get install amsd
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
The following NEW packages will be installed:
  amsd
0 upgraded, 1 newly installed, 0 to remove and 98 not upgraded.
Need to get 1,057 kB of archives.
After this operation, 5,353 kB of additional disk space will be used.
Get:1 https://downloads.linux.hpe.com/SDR/repo/aip jammy/current/non-free arm64 amsd arm64 3.5.0-1867-ubuntu22 [1,0
57 kB]
Fetched 1,057 kB in 1s (1,821 kB/s)
Selecting previously unselected package amsd.
(Reading database ... 78304 files and directories currently installed.)
Preparing to unpack .../amsd_3.5.0-1867-ubuntu22_arm64.deb ...
Unpacking amsd (3.5.0-1867-ubuntu22) ...
Setting up amsd (3.5.0-1867-ubuntu22) ...
Processing triggers for man-db (2.10.2-1) ...
Scanning processes...
Scanning processor microcode...
Scanning linux images...

Running kernel seems to be up-to-date.

Failed to check for processor microcode upgrades.

No services need to be restarted.

No containers need to be restarted.

No user sessions are running outdated binaries.

No VM guests are running outdated hypervisor (qemu) binaries on this host.
hpeadmin@dl384:~$
```

After installing the Agentless Management Service, information regarding the installed operating system becomes available in the iLO web interface.

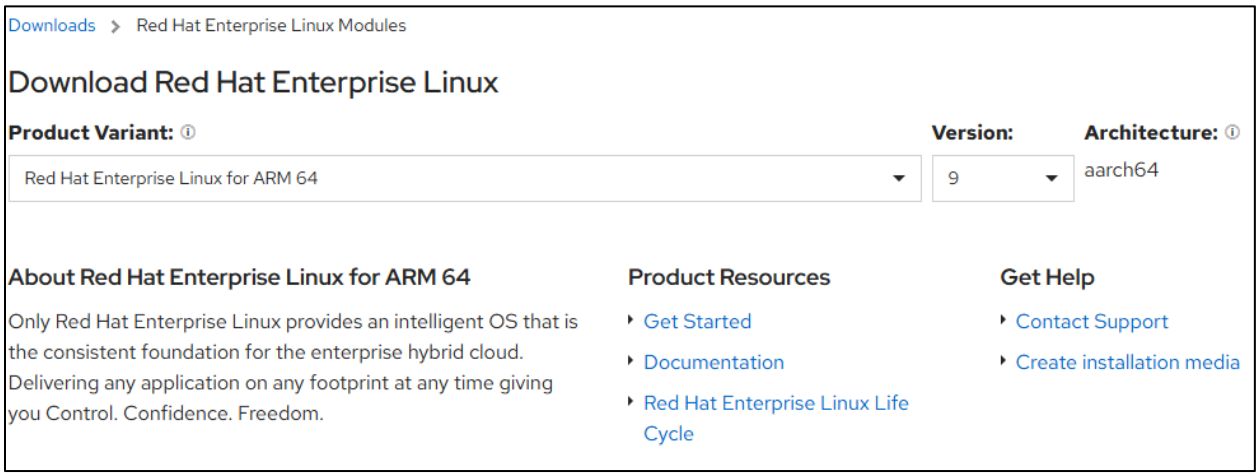
Deploying Red Hat® Enterprise Linux® 9.4

The following section will highlight important steps in deploying Red Hat Enterprise Linux 9.4 and greater on the HPE ProLiant Compute DL384 Gen12 server. It is not intended to provide advice for all installation settings as those vary by use case and organization.

Obtaining the proper OS media

Visit the Red Hat Customer Portal to download the proper media for an ARM 64 Architecture

access.redhat.com/downloads



Install the operating system as you normally would, except for the below recommended steps for 64k Page Size Kernels and the HPE Advanced Integration Pack.

Installation Utilizing 64k Page Size Kernels

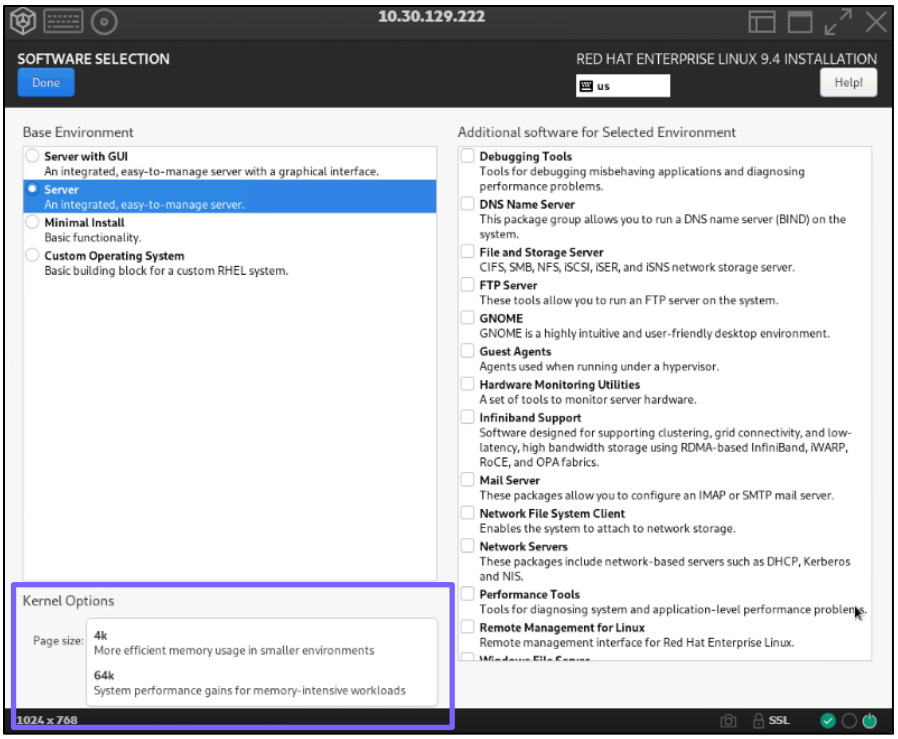
The 64k page size kernel is the recommended option for large datasets on GH200 platforms and memory-intensive workloads as it has significant gains in overall system performance.

You must choose page size on 64-bit ARM architecture systems at the time of a manual installation. When deploying the operating system via Kickstart, the 64k-kernel package must be specified as part of the Kickstart file.

For more information, including enabling 64k-kernel packages please visit:

Installing RHEL Kernel-64k on ARM

Below is an example of manually selecting the 64k-kernel during a manual installation using an iLO Virtual Media mounted ISO.



If a Red Hat Subscription Manager was not selected during installation, it is necessary to complete registration with a subscription manager or create a local repository utilizing the installation media.

After the first boot of the newly installed Red Hat Enterprise Linux OS, it is necessary to ensure the 64k kernel is loaded by default.

Upon completion of the reboot, validate the 64k kernel is running by running the below command:

```
getconf PAGESIZE
```

Subscribe the Red Hat Enterprise Linux Server to the AIP repository

After the initial operating system installation, HPE recommends adding the HPE ProLiant Compute Advanced Integration Pack available from the HPE Software Delivery Repository. This online repository contains drivers, firmware and utilities to assist in managing and lifecycle management of ARM based systems like the HPE ProLiant Compute DL384 Gen12 server.

Additional information on the HPE Software Delivery Repository (SDR) can be found here:

downloads.linux.hpe.com/SDR/getting_started.html

downloads.linux.hpe.com/SDR/project/aip/

To add the AIP repository, create a new YUM repository with the following contents:

```
sudo nano /etc/yum.repos.d/aip.repo
```

Contents of aip.repo:

```
[aip]
name=Advanced Integrated Pack
baseurl=https://downloads.linux.hpe.com/SDR/repo/aip/RedHat/9.4/aarch64/current
enabled=1
gpgcheck=1
gpgkey=https://downloads.linux.hpe.com/SDR/hpePublicKey2048_key1.pub,https://downloads.linux.hpe.com/SDR/hpePublicKey2048_key2.pub
```

Run `sudo yum update` to reload the yum repo files and ensure the Advanced Integration Pack is listed as an active repository.

```
[root@DL384 ~]# yum update
Updating Subscription Management repositories.
Unable to read consumer identity

This system is not registered with an entitlement server. You can use "rhc" or "subscription-manager" to register.

Advanced Integrated Pack                                                    2.4 kB/s | 5.9 kB     00:02
Dependencies resolved.
Nothing to do.
Complete!
```

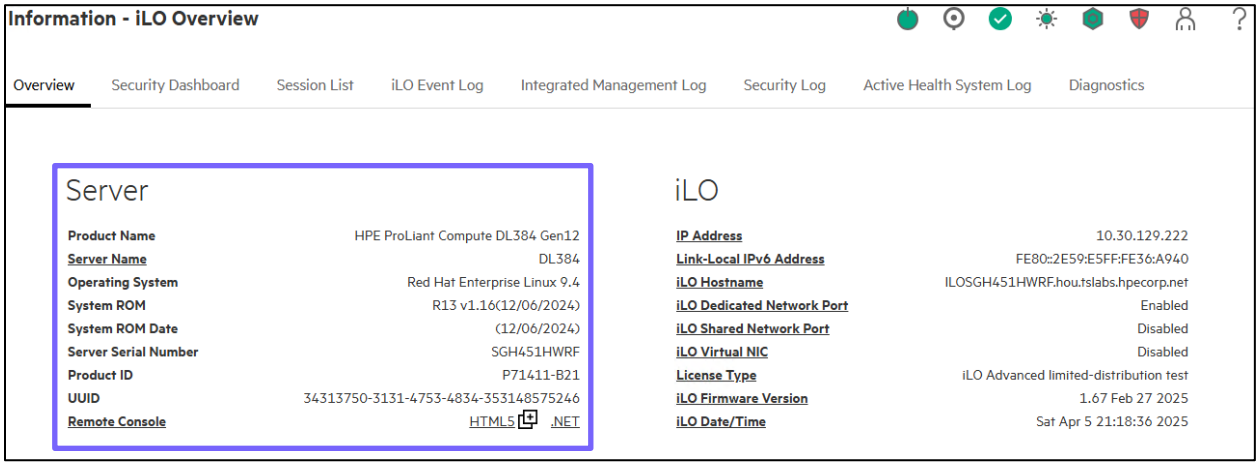
To install utilities such as the iLORest Tool run `sudo yum install ilorest`

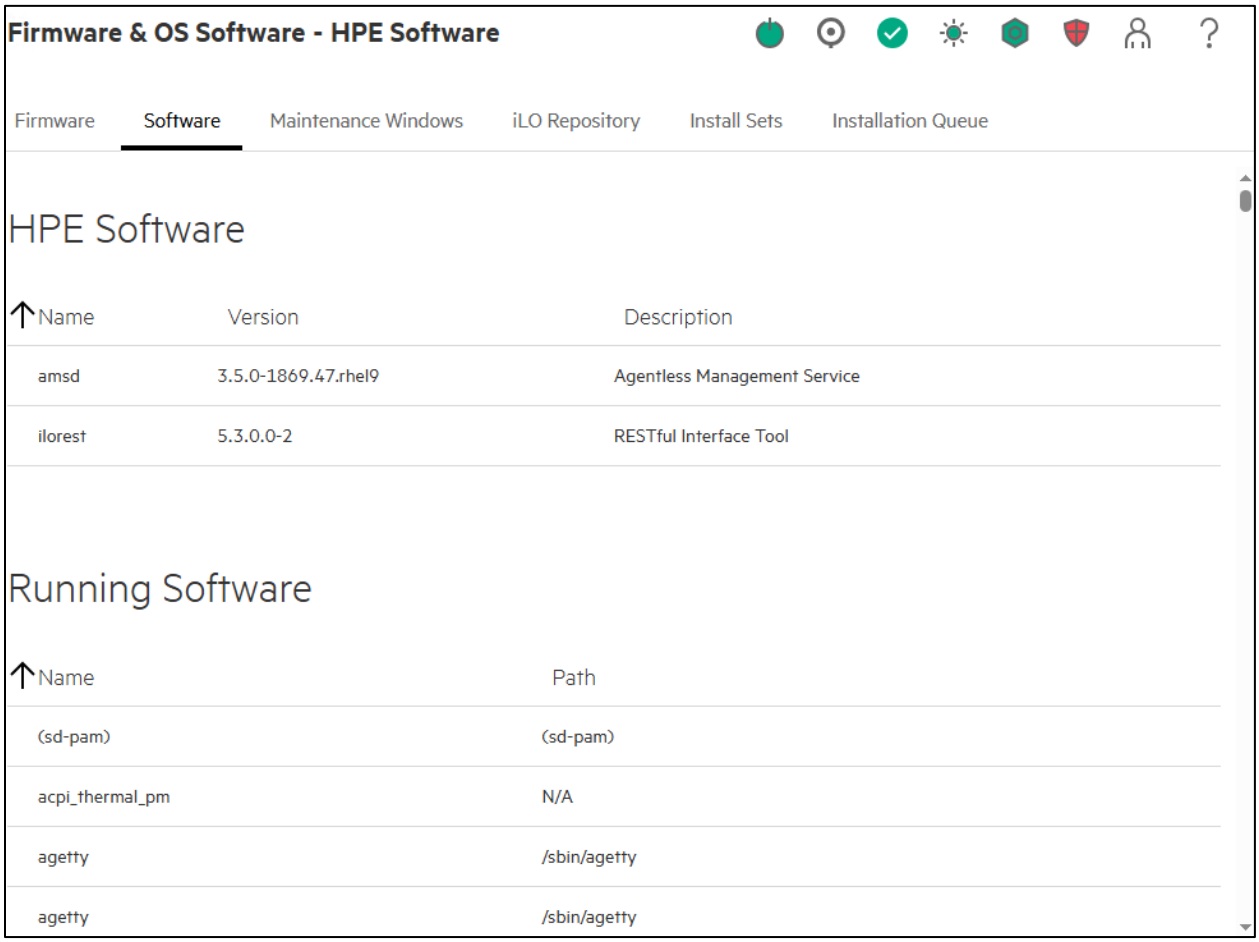
```
[hpeadmin@DL384 ~]$ sudo yum install ilorest
Updating Subscription Management repositories.
Unable to read consumer identity

This system is not registered with an entitlement server. You can use "rhc" or "subscription-manager" to register.

Last metadata expiration check: 0:01:53 ago on Sat 05 Apr 2025 04:15:34 PM CDT.
Dependencies resolved.
================================================================================
 Package                 Architecture            Version                 Repository            Size
================================================================================
Installing:
 ilorest                 aarch64                 5.3.0.0-2               aip                   17 M

Transaction Summary
================================================================================
Install  1 Package
```

To install the Agentless Management Service run `sudo yum install amsd`

```
Last metadata expiration check: 0:00:14 ago on Sat 05 Apr 2025 04:15:34 PM CDT.
Dependencies resolved.
================================================================================
 Package             Architecture            Version                     Repository            Size
================================================================================
Installing:
 amsd                aarch64                 3.5.0-1869.47.rhel9         aip                  1.5 M

Transaction Summary
================================================================================
Install  1 Package
```

After installing the Agentless Management Service, information regarding the installed operating system becomes available in the iLO web interface.



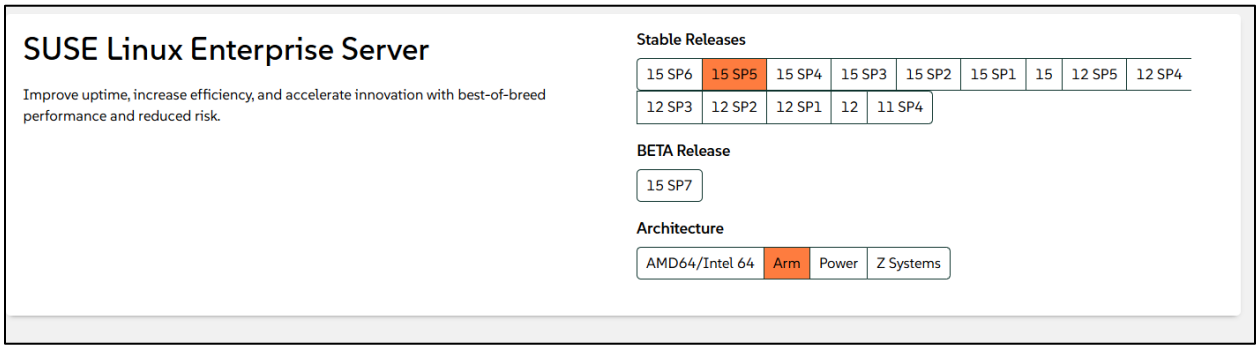


Deploying SUSE Linux Enterprise Server 15 SP5

Obtaining the proper OS media

Selecting the proper installation media is necessary to ensure compatibility with the ARM processor contained in the NVIDIA GH200 NVL2 Superchip.

Visit SUSE's download site to obtain the desired release and ensure the ARM architecture is selected.
suse.com/download/sles/

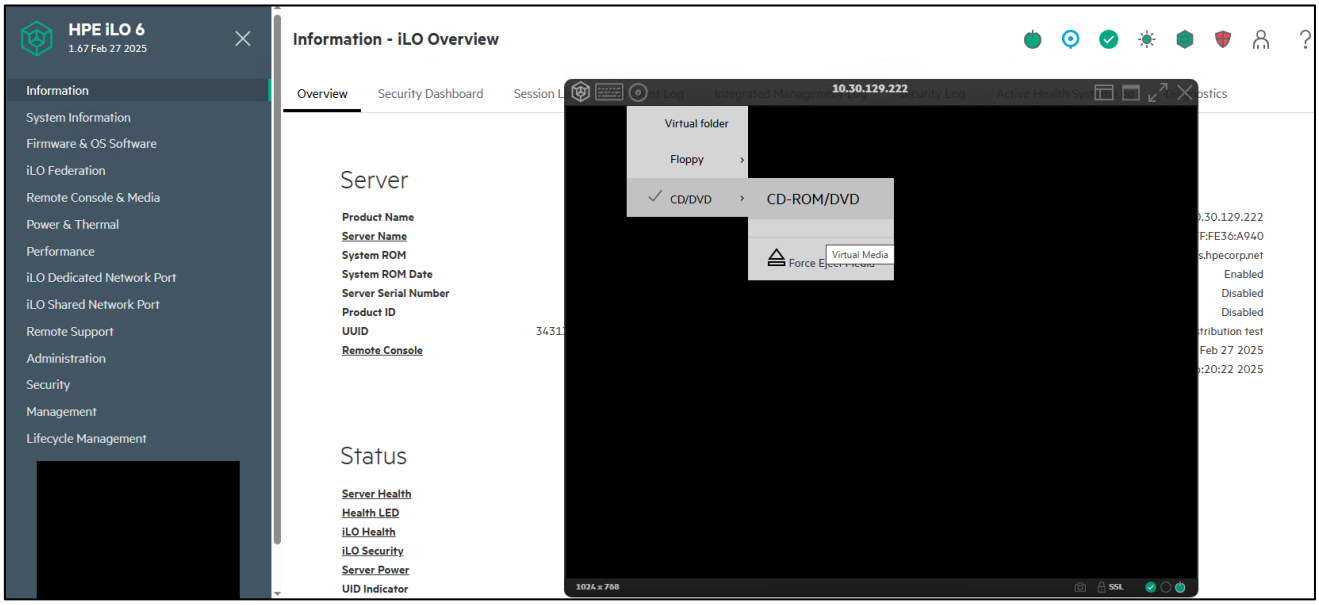


Install the operating system as you normally would, with the exception of the below recommended steps for 64k Page Size Kernels and the HPE Advanced Integration Pack.

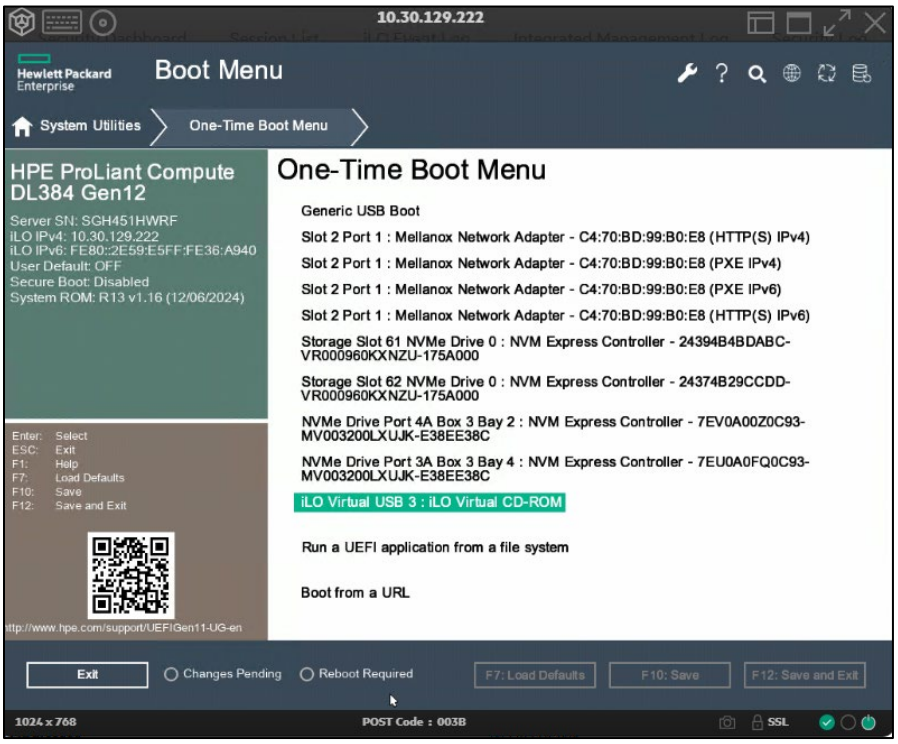
Installation Using Text Console Mode

When booting to a Virtual Media ISO of SUSE Linux 15, it is necessary to perform the installation using a text console screen instead of the graphical user interface.

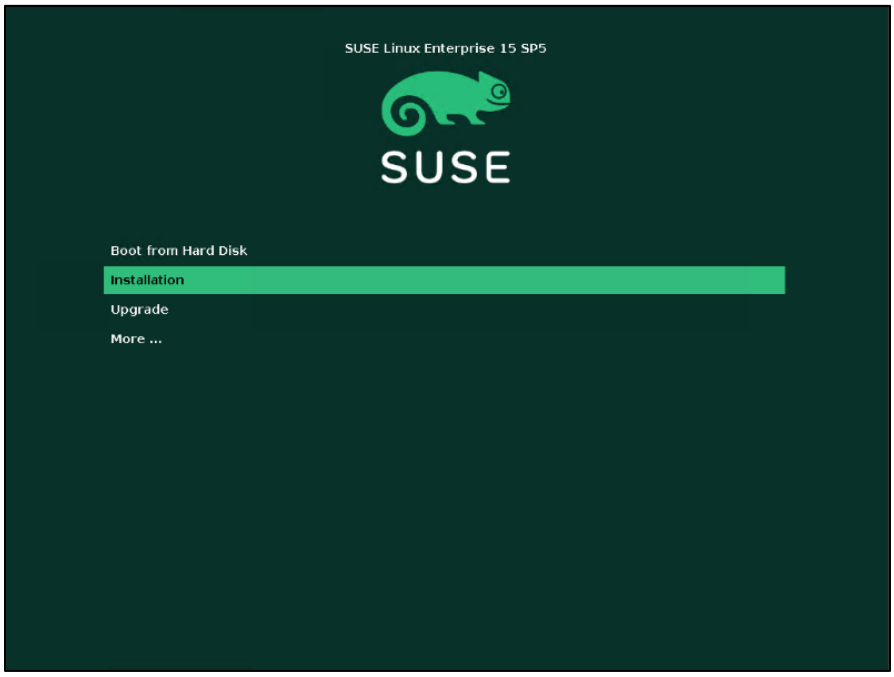
Launch the HPE iLO Remote Console from the iLO web interface as described in a previous section. Using the iLO Virtual Media capability, mount the SLES 15 SP5 image.



During POST press F11 to invoke the One-Time Boot Menu

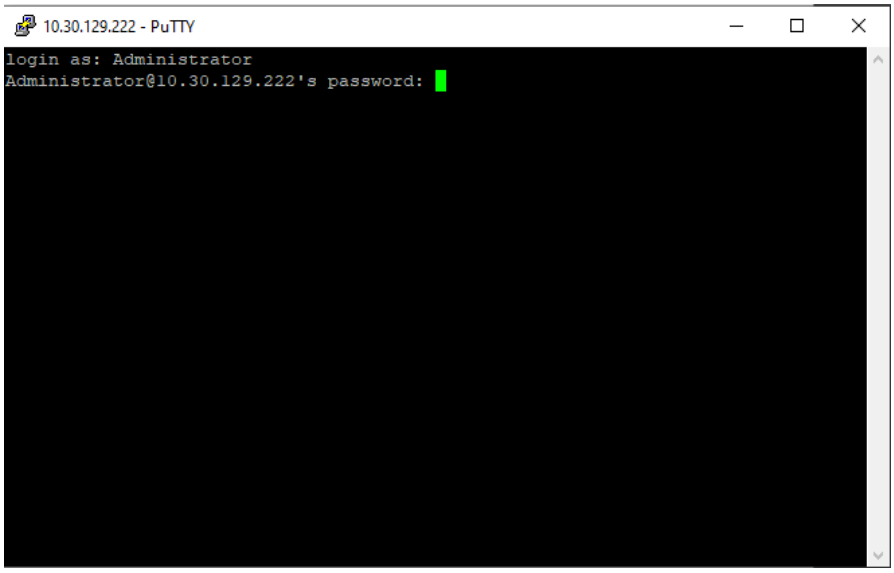


The server will boot to the ISO boot menu. The default selection of Installation will start after 20 seconds of inactivity, otherwise press ENTER to continue booting to Installation.

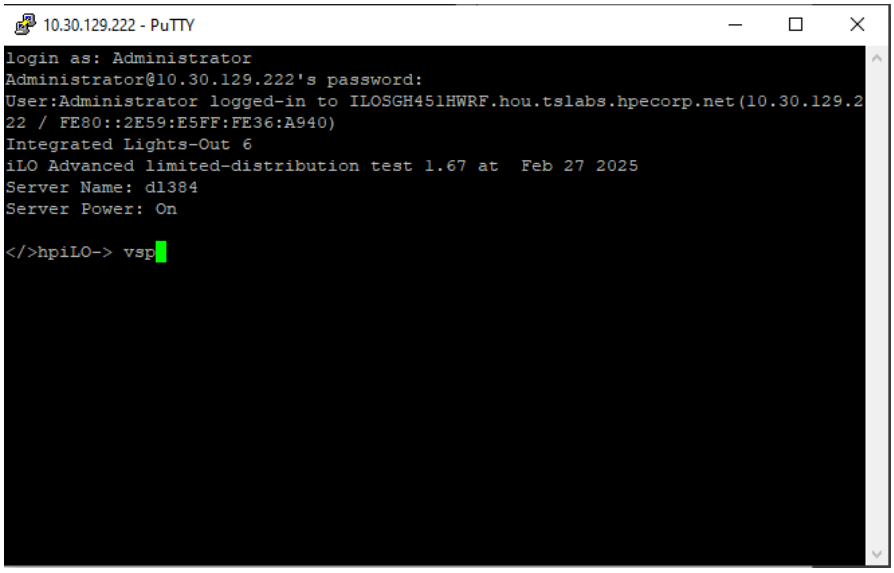


To continue the installation, it is necessary to SSH to the iLO IP address or Fully-Qualified-Domain Name utilizing a tool such as Putty.

Login with the iLO Administrator credentials or another equivalent user if created during the initial setup.

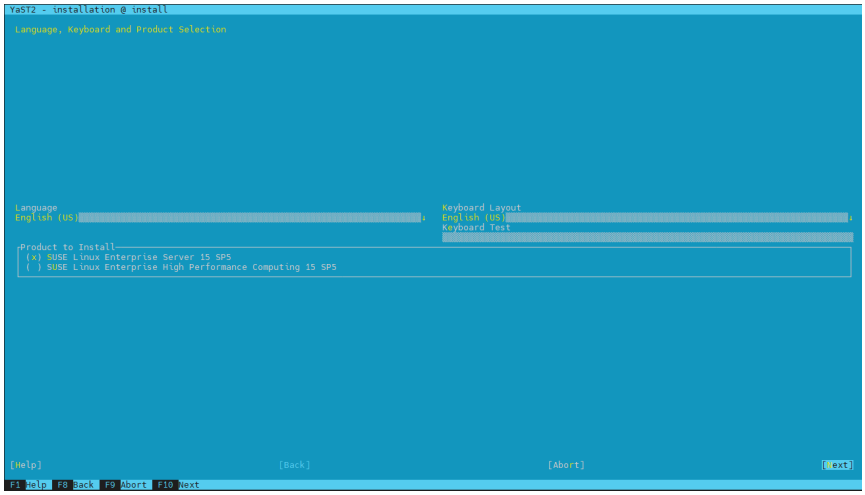


Launch the Virtual Serial Port by entering the command `vsp`



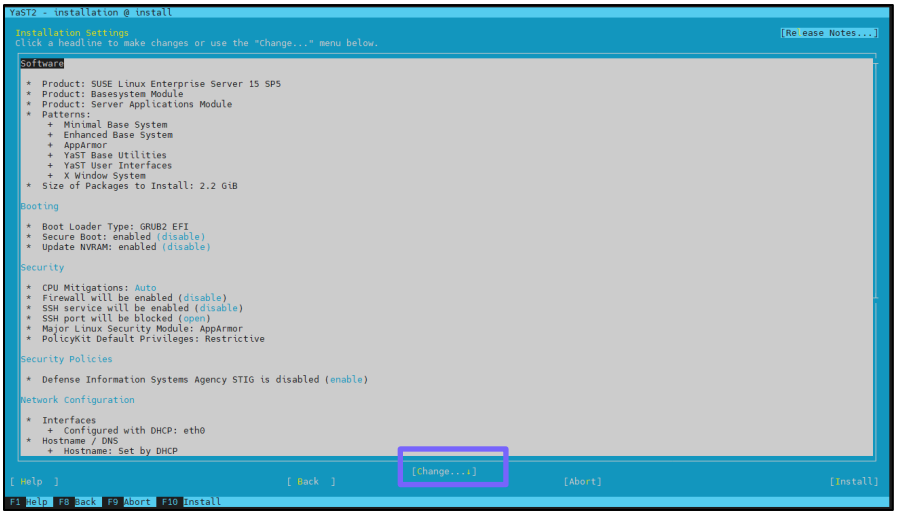
Installation Utilizing 64k Page Size Kernels

The SLES YaST2 text-based installation wizard will start as the server boots to the iLO Virtual Media mounted ISO.

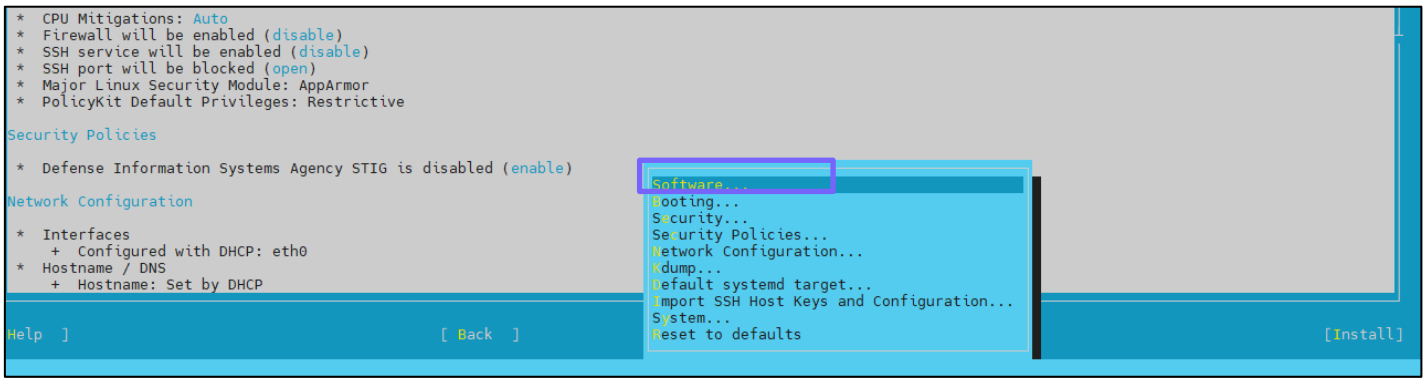


Using the keyboard cursors, function keys, tab, and ALT keys, navigate the installation wizard to select the proper language, keyboard layout, and SUSE Linux Enterprise Server 15 SP5 to start the installation.

On the Installation Settings screen at the end of the installation wizard, use the Tab key to select Software and press enter.



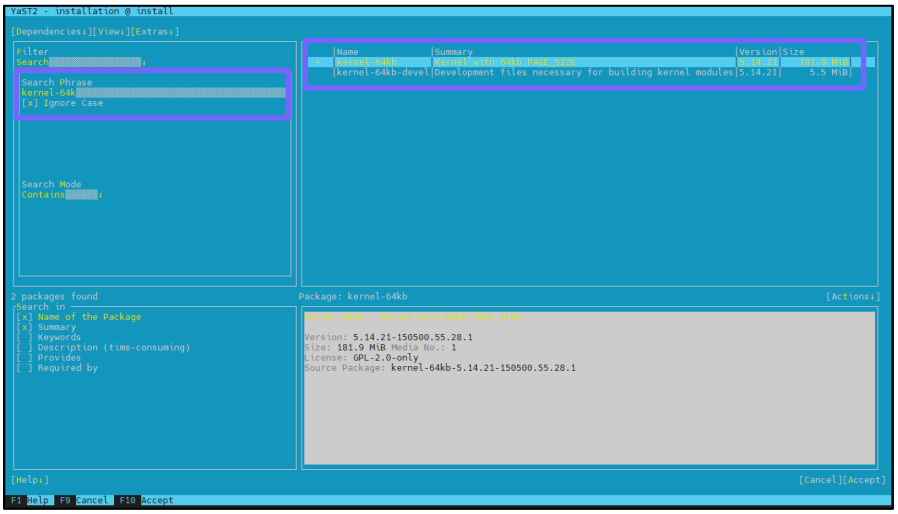
Select the Software submenu and press Enter.



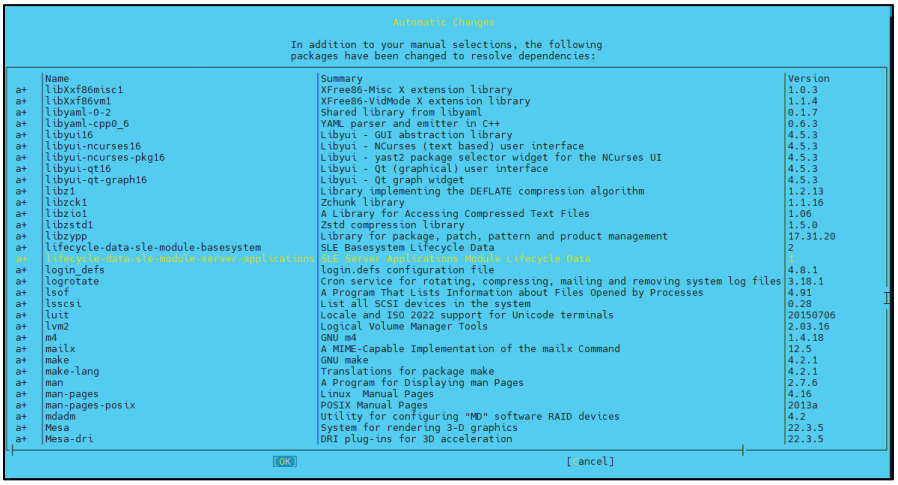
Use the Search function to find the following packages.

- kernel-64k
- kernel-default

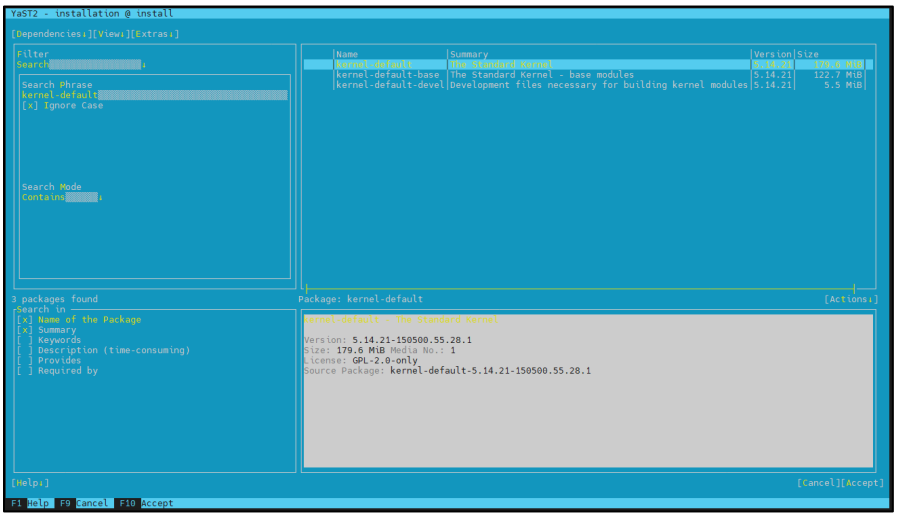
For both of the above packages, use the spacebar or enter key to insure a “+” is next to the package needing installation. The following image adds kernel-64k to the list of packages to install.



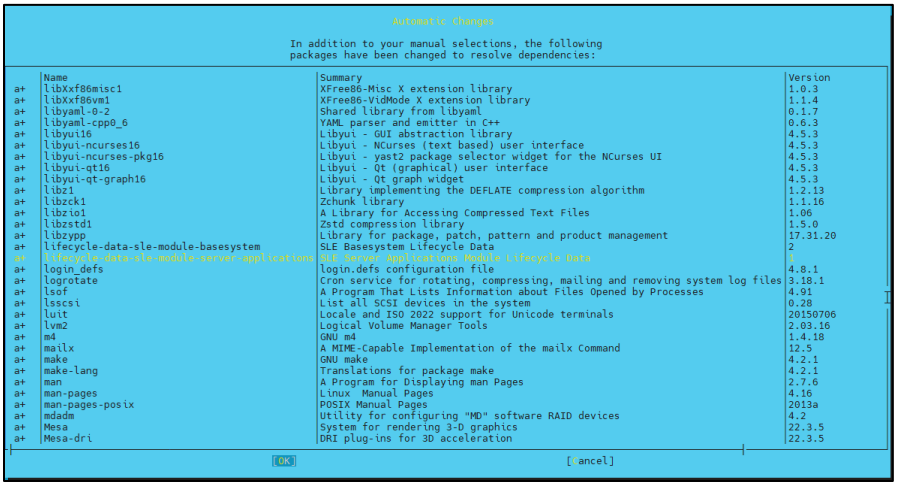
Selecting Accept or pressing Alt+a displays a screen of dependencies that will be installed as a result of adding the kernel-64k package. Select OK.



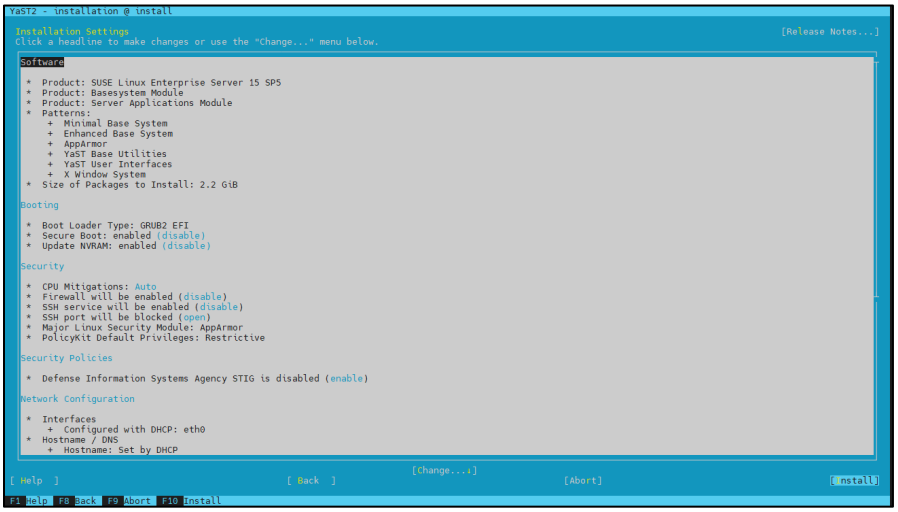
Repeat the same search action for kernel-default. Press enter to deselect the package kernel-default.



Navigate to Accept or press alt+a to display the automatic changes to resolve dependencies screen and click OK to continue.



With those two changes applied, navigate to Install or press alt+i to continue.



Confirm the installation to start installing the OS to local disk.

After the installation completes, the server will automatically reboot. Login to the OS via the iLO Remote Console or Virtual Serial Port to obtain the IP address of the OS to establish an SSH connection for the remaining steps of installing the HPE utilities and NVIDIA CUDA and GPU drivers.

Note:

Installation of the CUDA Toolkit and NVIDIA GPU Driver requires additional OS packages. Ensure your system is using an appropriate subscription manager or local repositories for installation of the component dependencies.

Subscribe the SLES Server to the AIP repository

After the initial operating system installation, HPE recommends adding the HPE ProLiant Compute Advanced Integration Pack available from the HPE Software Delivery Repository. This online repository contains drivers, firmware and utilities to assist in managing and lifecycle management of ARM based systems like the HPE ProLiant Compute DL384 Gen12 server.

Additional information on the HPE Software Delivery Repository (SDR) can be found here:

downloads.linux.hpe.com/SDR/getting_started.html

downloads.linux.hpe.com/SDR/project/aip/

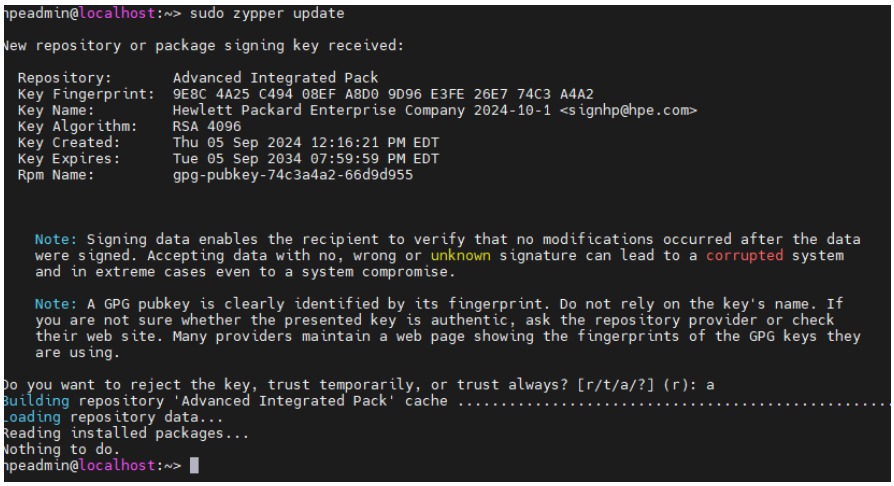
To add the AIP repository, create a new ZIPPER repository with the following contents:

```
sudo vim /etc/zypp/repos.d/aip.repo
```

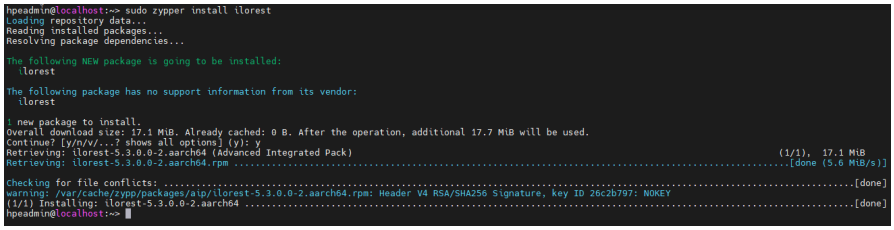
Contents of aip.repo:

```
[aip]
name=Advanced Integrated Pack
baseurl=https://downloads.linux.hpe.com/SDR/repo/aip/SuSE/SLES15-SP5/aarch64/current
enabled=1
gpgcheck=1
gpgkey=https://downloads.linux.hpe.com/SDR/hpePublicKey2048_key1.pub,https://downloads.
linux.hpe.com/SDR/hpePublicKey2048_key2.pub
```

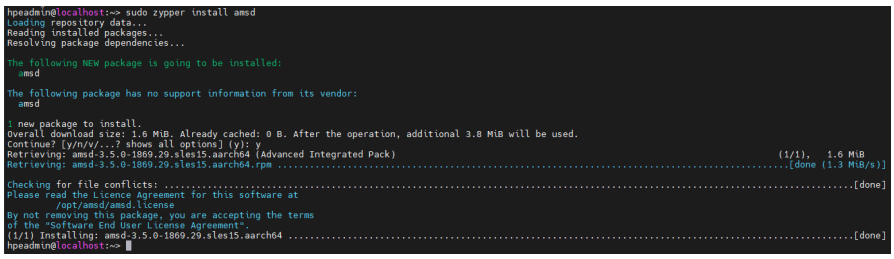
Run `sudo zypper update` to reload the repo files and ensure the Advanced Integration Pack is listed as an active repository.



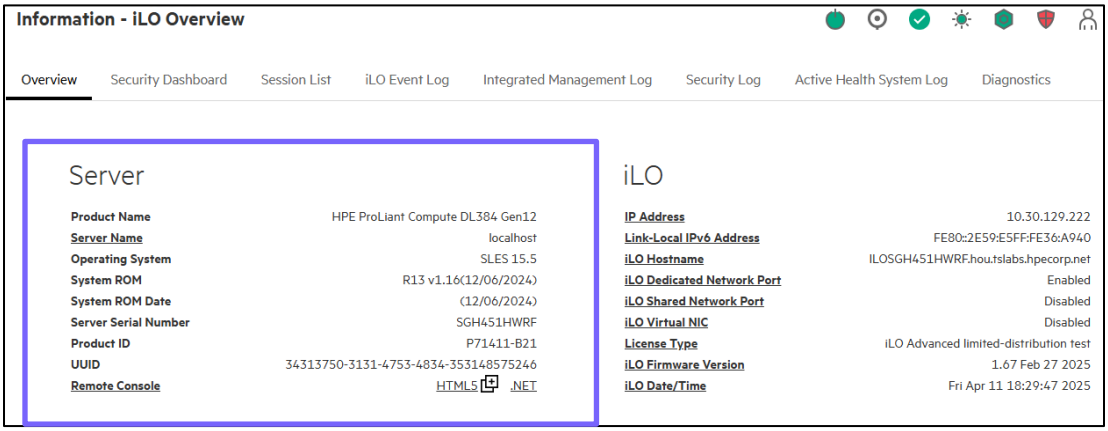
To install utilities such as the iLORest Tool run `sudo zypper install ilorest`

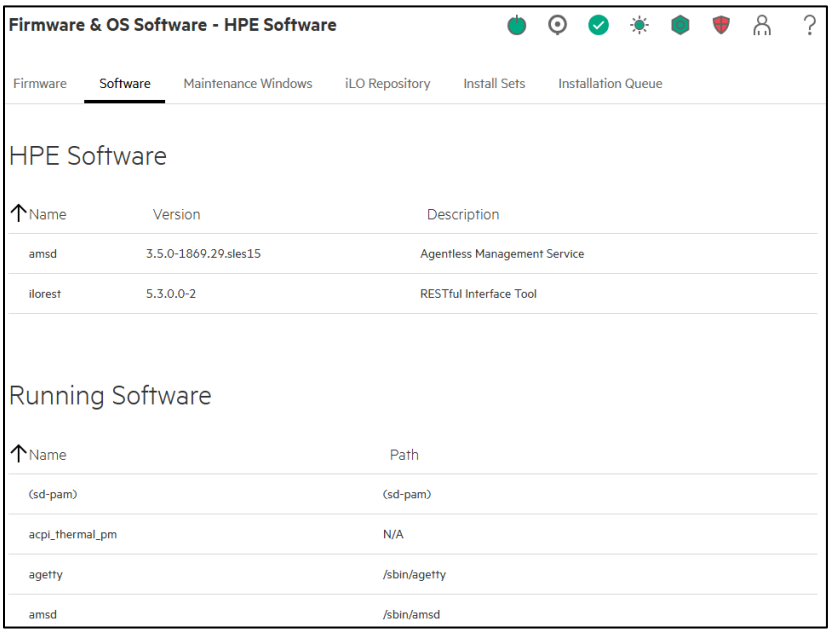


To install the Agentless Management Service run `sudo zypper install amsd`



After installing the Agentless Management Service, information regarding the installed operating system becomes available in the iLO web interface.





Installing NVIDIA GPU Driver and the CUDA Toolkit

Note:

Installation of the CUDA Toolkit and NVIDIA GPU Driver requires additional OS packages. Ensure your system is using an appropriate subscription manager or local repositories for installation of the component dependencies.

Always visit NVIDIA's web site to determine the latest version of the CUDA toolkit, NVIDIA driver, download documentation and release notes. Select the version that is appropriate for your installation

Links to the latest and archived versions of NVIDIA CUDA and associated GPU drivers can be found here:

developer.nvidia.com/cuda-downloads

Canonical Ubuntu Instructions

The below commands are necessary to install the following:

- Repository for NVIDIA CUDA and GPU drivers
- NVIDIA CUDA toolkit 12.8
- Open Kernel NVIDIA GPU drivers
- Set NVIDIA services to start on every boot

```
sudo wget
https://developer.download.nvidia.com/compute/cuda/repos/ubuntu2204/sbsa/cuda-
keyring_1.1-1_all.deb

sudo dpkg -i cuda-keyring_1.1-1_all.deb

sudo apt-get update

sudo apt-get install cuda-toolkit-12-8

sudo apt-get install nvidia-open

sudo mkdir /lib/systemd/system/nvidia-persistenced.service.d

sudo dd status=none of=/lib/systemd/system/nvidia-persistenced.service.d/override.conf
<< EOF
[Service]
ExecStart=
ExecStart=/usr/bin/nvidia-persistenced --persistence-mode --verbose
[Install]
WantedBy=multi-user.target
EOF

sudo reboot now
```

Red Hat Enterprise Linux Instructions

The below commands are necessary to install the following:

- Red Hat kernel development packages for 64k page files
- Kernel-headers
- Matching core and development packages for the installed kernel
- Extra Packages for Enterprise Linux (EPEL)
- Repository for NVIDIA CUDA and GPU drivers
- CUDA version 12.8
- Open Kernel NVIDIA GPU drivers

```
sudo dnf install kernel-64k-devel-$(uname -r| sed 's/+64k//g')
sudo dnf install kernel-headers-$(uname -r| sed 's/+64k//g')
sudo dnf install kernel-devel-matched-$(uname -r| sed 's/+64k//g')
sudo dnf install https://dl.fedoraproject.org/pub/epel/epel-release-latest-9.noarch.rpm
sudo dnf config-manager --add-repo
https://developer.download.nvidia.com/compute/cuda/repos/rhel9/sbsa/cuda-rhel9.repo
sudo dnf clean expire-cache
sudo dnf install cuda-toolkit-12-8
sudo dnf module install nvidia-driver:open-dkms
sudo systemctl enable nvidia-persistenced
sudo reboot now
```

SUSE Linux Enterprise Server Instructions

The below commands are necessary to install the following:

- Repository for NVIDIA CUDA and GPU drivers
- CUDA version 12.8
- Open Kernel NVIDIA GPU drivers

Note: It may be necessary to remove a driver installed during the OS installation.

```
sudo zypper remove nvidia-driver-G06-kmp-64kb-570.124.06-1.aarch64
sudo reboot
```

```
sudo zypper install kernel-64kb-devel
sudo usermod -a -G video <username>
sudo zypper addrepo
https://developer.download.nvidia.com/compute/cuda/repos/sles15/sbsa/cuda-sles15.repo
sudo zypper refresh
sudo zypper install cuda-toolkit-12-8
sudo zypper install nvidia-open-64k
sudo mkdir /usr/lib/systemd/system/nvidia-persistenced.service.d/
```

Use a text editor such as vi, vim, nano to create a 70-nvidia-persistenced.conf file.

```
sudo sudo vim /usr/lib/systemd/system/nvidia-persistenced.service.d/70-nvidia-
persistenced.conf
```

Insert the below text at the top of file in the editor of your choice:

```
[Service]
Type=
Type=forking
```

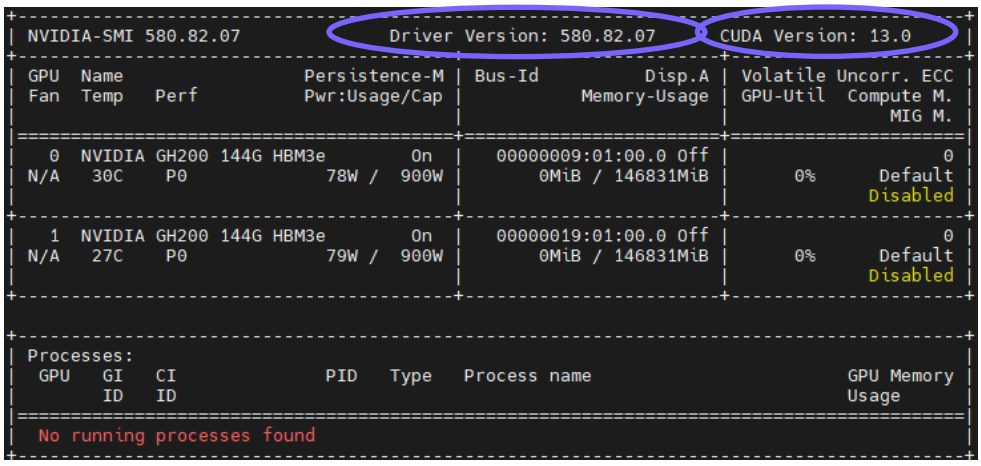
Run the following commands to enable nvidia persistence and reboot.

```
sudo systemctl enable nvidia-persistenced

sudo reboot now
```

Verification of a Successful Installation

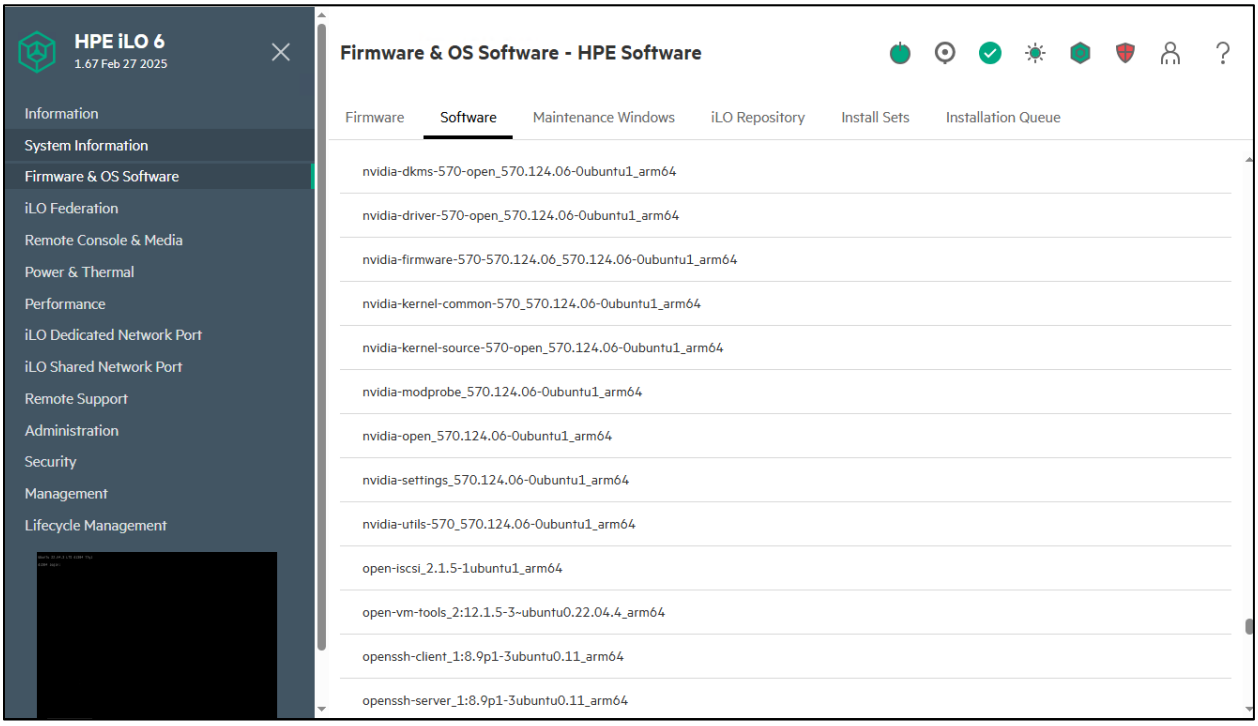
After installation of NVIDIA CUDA and GPU drivers, and the required reboot, run the following command to display information about the NVIDIA GH200 NVL2 devices installed in the HPE ProLiant Compute DL384 Gen12 server.

```
nvidia-smi
```



Validating the Installation Through iLO

In HPE iLO, the firmware and software versions installed on the platform are available through the Firmware and OS Software menu item on the left-hand navigation pane if the Agentless Management Service is installed.

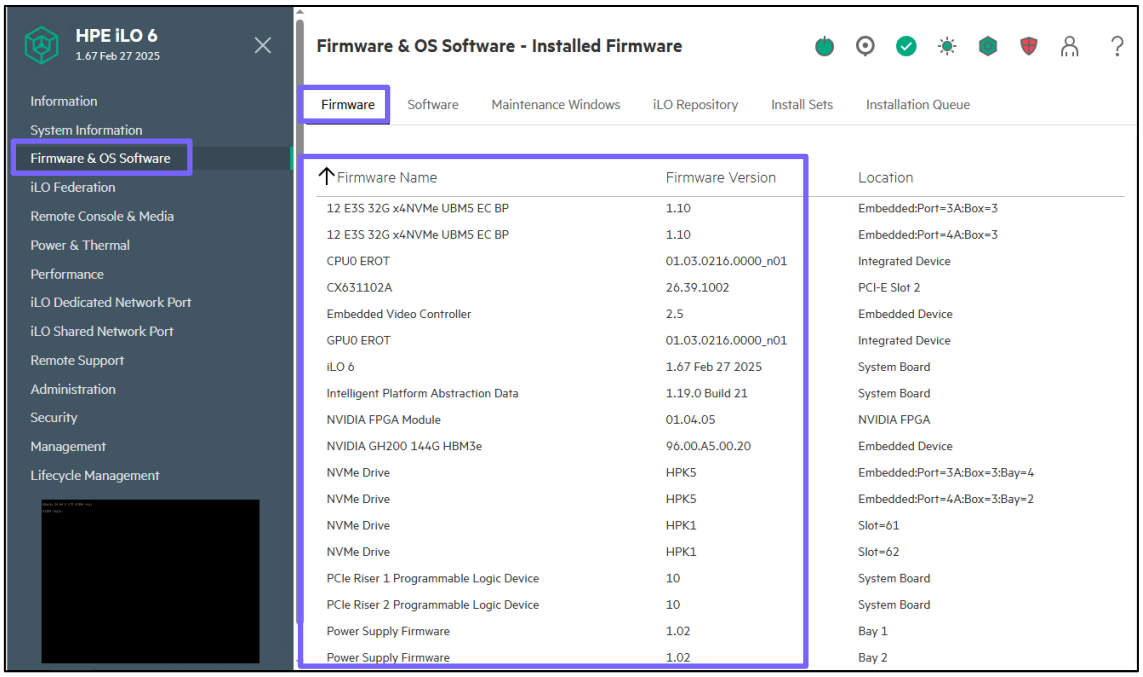


Updating Firmware and Drivers Through Software Delivery Repository

HPE provides tools through the HPE Software Delivery Repository to assist in locating, downloading and applying firmware packages through the operating system. Each OS deployment section of this paper included steps for adding the HPE ProLiant Compute Advanced Integration Pack. Refer to the appropriate OS distribution section if the HPE repos were not added at the time of installation.

Locating installed packages on the HPE ProLiant Compute DL384 server:

The HPE iLO webpage contains information on all firmware components for the server and installed options.



Utilities Required to Locate and Apply Firmware Updates

HPE provides a simple command-line browser that interacts with the HPE Firmware Pack for ProLiant online repository. This python3 script provides the following two commands:

- Use the `fwlist` command executes on the local server to return the host server firmware information. The response from `fwlist` contains a target ID that can be used as a key in `fwget` commands to filter searches.
- The `fwget` command can search, locate, and download firmware packages including rpm and fwpkg files. The downloaded firmware packages may then be flashed with the `fwget` and companion tool `ilorest`.

Download `fwget`

- Direct download - downloads.linux.hpe.com/SDR/project/fwpp/fwget.py
- RHEL8, RHEL9, SLES15 - downloads.linux.hpe.com/SDR/project/fwpp/fwget-1.0.4-1.noarch.rpm
- Ubuntu - downloads.linux.hpe.com/SDR/project/fwpp/fwget-1.0.4-1_all.deb

Examples:

Installing `fwget` in Ubuntu:

```
sudo wget https://downloads.linux.hpe.com/SDR/project/fwpp/fwget-1.0.4-1_all.deb
sudo dpkg -i fwget-1.0.4-1_all.deb
```

After installing `fwget`, it is necessary to edit the `.fwget.conf` file. The file is generated after executing a `fwget` command the first time.

```
sudo fwlist spaced_display
```

Edit the `~/.fwget` file to update the SDR URL, and provide an iLO username, password and IP address or FQDN.

```
{
"_comment": "For Gen9 and earlier, generate token at:
http://downloads.linux.hpe.com/SDR/project/fwpp/fwget.html",
"_comment": "Replace fwpp-gen11 with fwpp-gen10/fwpp-gen9/fwpp-gen8 as appropriate.",
  "ilo_username": "username_here",
  "ilo_password": "password_here",
  "ilo_address": "IP_Address",
  "token": "na",
  "sdr_url" : "http://downloads.linux.hpe.com/SDR/repo/aip/pool/non-free"
}
```

Re-run the command below to obtain the firmware installed on the system.

```
sudo fwlist spaced_display
```

Additional information on the HPE Software Delivery Repository (SDR) can be found here:

downloads.linux.hpe.com/SDR/project/aip/

downloads.linux.hpe.com/SDR/project/fwpp/fwget.html

Running Basic System Health Checks

STREAM is an industry standard microbenchmark for measuring sustained memory bandwidth. The STREAM benchmark is specifically designed to work with datasets much larger than the available cache on a system, and therefore provide results that are more indicative of the performance of very large, vector style applications.

NVIDIA provides guidance on basic Grace and Grace Hopper tests to validate the system is configured correctly and check the performance and memory bandwidth using STREAM.

Download and configure the STREAM benchmark by running the following commands.

```
wget https://www.cs.virginia.edu/stream/FTP/Code/stream.c
gcc -Ofast -mcpu=neoverse-v2 -fopenmp \
  -DSTREAM_ARRAY_SIZE=120000000 -DNTIMES=200 \
  -o stream_openmp.exe stream.c
```

Execute the test by setting the number of OpenMP threads (`OMP_NUM_THREADS`) based on the number of Grace CPU cores. As a refresher, the NVIDIA GH200 Superchips have 72 cores per GH2000, while the GH200NVL2 features 144 CPU cores.

Run the following command on a DL384 Gen12 with a single GH200 Superchip with 72 cores.

```
OMP_NUM_THREADS=72 OMP_PROC_BIND=spread ./stream_openmp.exe
```

The following command is suitable for DL384 Gen12 systems with dual NVIDIA GH200 NVL2 Superchip with 144 cores.

```
OMP_NUM_THREADS=144 OMP_PROC_BIND=spread numactl -m0,1 ./stream_openmp.exe
```

Another example command below would be used on systems with dual GH200 Superchips with 72 cores.

```
OMP_NUM_THREADS=72 OMP_PROC_BIND=spread numactl -m0,1 ./stream_openmp.exe
```

The microbenchmark produces output like the following.

```
-------------------------------------------------------------
Function  Best Rate MB/s Avg time   Min time   Max time
Copy:     733697.5   0.003554   0.002617   0.010737
Scale:     734835.6   0.003589   0.002613   0.013019
Add:      729664.5   0.005199   0.003947   0.016373
Triad:     717656.6   0.005295   0.004013   0.016040
-------------------------------------------------------------
Solution Validates: avg error less than 1.000000e-13 on all three arrays
```

For more information visit the NVIDIA Grace Performance Tuning Guide:

docs.nvidia.com/grace-perf-tuning-guide/system.html

Conclusion

This white paper has demonstrated the HPE and NVIDIA best practices for installing Canonical Ubuntu, Red Hat Enterprise Linux and SUSE Linux Server on a HPE ProLiant Compute DL384 Gen12 server. In addition to the base operating system, the paper highlighted critical steps such as installing 64k Kernel options and installing the NVIDIA CUDA libraries and appropriate drivers for the NVIDIA GH200 superchip.

Together HPE and NVIDIA have delivered a purpose-built platform to provide an AI Inferencing engine for unmatched performance and manageability.

Resources

Visit the below resources for more information.

HPE ProLiant Compute DL384 Gen12

HPE ProLiant Compute DL384 Gen12 Server User Guide

NVIDIA Ubuntu Linux Installation Guide

NVIDIA Red Hat Linux Installation Guide

NVIDIA SUSE Enterprise Linux Installation Guide

NVIDIA Grace Performance Tuning Guide

a00147767enw

HEWLETT PACKARD ENTERPRISE

hpe.com