

VMware in the Kuria Cluster

Introduction to Virtualization Software

Virtualisation software creates an abstraction layer over physical hardware, enabling it to be segmented into multiple virtual machines (VMs). Each VM functions independently, running its own operating system and applications. Virtualisation allows better resource utilisation and flexibility in managing workloads.

Components of Virtualization Software

- **Virtual Machine (VM):** A software-based representation of a physical computer. Each VM operates independently, often with its own guest operating system (OS).
- **Guest OS:** The operating system running within a VM, independent of the host OS.
- **Hypervisor (Virtual Machine Monitor):** A small software layer that enables multiple OS instances to run on a single physical machine. It ensures that VMs can share the same physical resources without interfering with each other.

Types of Hypervisors

There are two types of hypervisors, each with a different architecture:

- **Type-1 Hypervisor (Bare Metal):** Runs directly on the host's hardware without needing a host operating system. Examples include VMware ESXi and Microsoft Hyper-V. It is commonly used in data centres and for large-scale deployments.

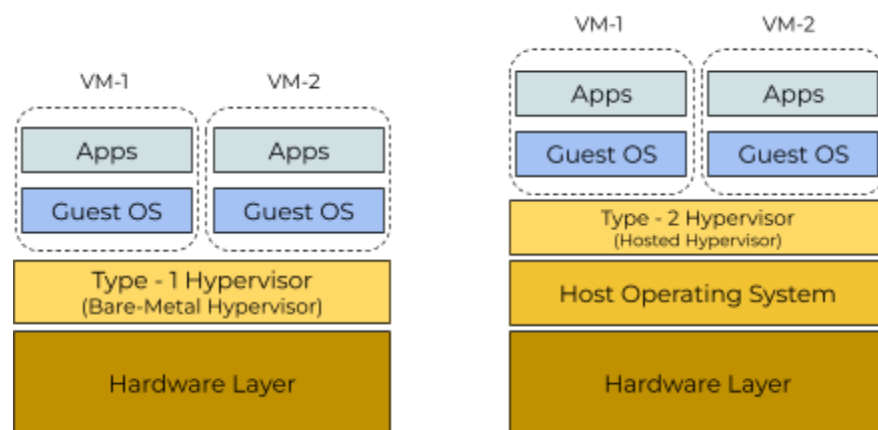


Fig: Types of Virtualization Architecture

- **Type-2 Hypervisor:** Runs on a host OS and manages virtual machines within it. Examples include VMware Workstation, Oracle VM VirtualBox, and Parallels Desktop. These are typically used for personal or desktop-level virtualisation.

Virtualisation in the Kuria Cluster

The Kuria cluster employs VMware ESXi, a Type-1 hypervisor, to virtualise the master node. VMware ESXi offers a high-performance, secure environment to manage VMs and allocate resources effectively. The virtual machines are managed through a web interface called VMware ESXi Host Client, which can be accessed by the cluster administrator using confidential credentials.

Key Features of VMware ESXi Host Client

- **Virtual Machine Management:** Create, configure, and manage VMs.
- **Resource Allocation:** Allocate CPU, memory, and storage to VMs.
- **Monitoring and Reporting:** Track the performance and health of VMs and physical hardware.
- **Snapshot Management:** Take snapshots for backup and disaster recovery.
- **Network Configuration:** Manage virtual networks and network interfaces.
- **Security Management:** Implement security policies such as virtual firewalls and access controls.
- **Storage Management:** Manage data stores and VM storage resources.

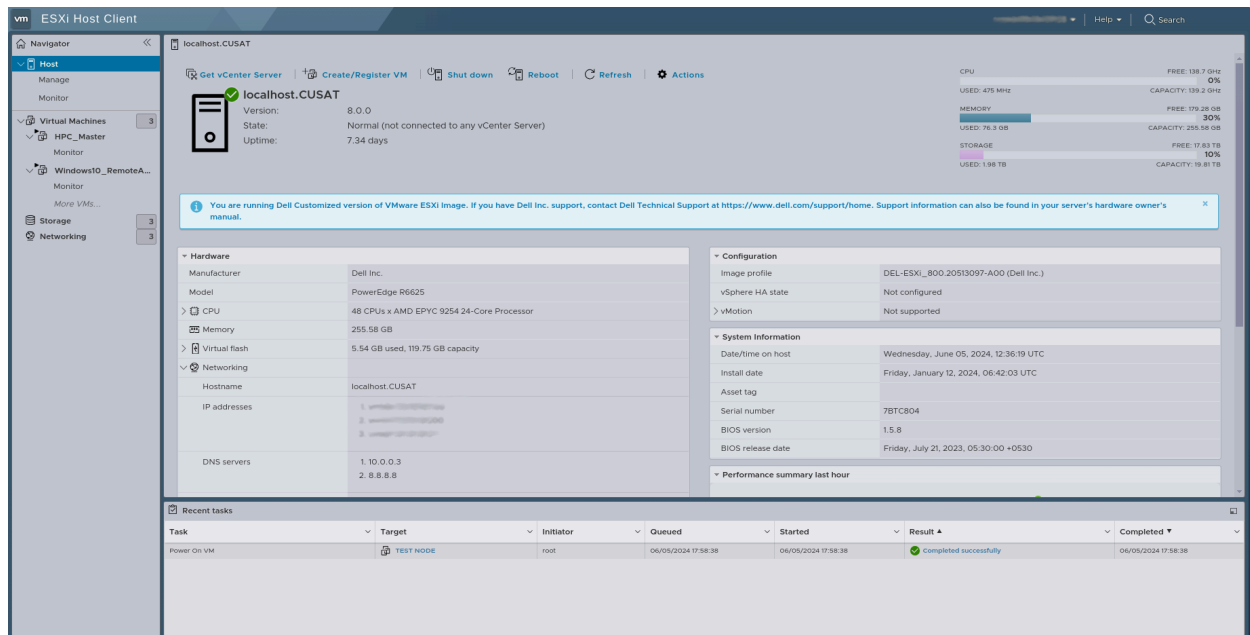


Fig: Snapshot of the UI of the VMware ESXi client.

Virtualization Architecture of the Kuria Cluster Master Node

The master node in the Kuria cluster contains three virtual machines hosted on VMware ESXi. Below are the details of each VM:

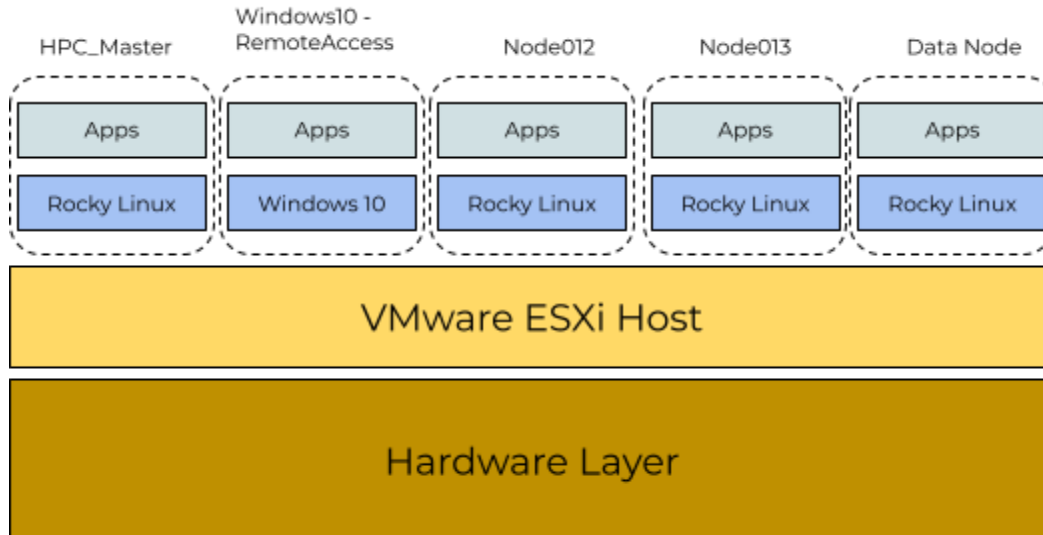


Fig: Virtualization Architecture of the Master node

VM Name	CPUs	RAM	Storage	Function
<i>HPC_Master</i>	8 Cores	64 GB	39.51 TB (currently)	Acts as the master node for the Kuria cluster. Users access the cluster via SSH through this VM.
<i>Windows10 - RemoteAccess</i>	4 cores	8 GB	76.47 GB	Used for remote desktop access (e.g., AnyDesk, Ultraviewer) for external access to the cluster.
<i>TestNode01</i>	8 cores	9 GB	32.09 GB	For testing codes before submitting to the compute nodes
<i>TestNode02</i>	8 cores	9 GB	32.09 GB	For testing codes before submitting to the compute nodes
<i>DataNode</i>	4 cores	8 GB	16.08 GB	Node specialised for data transfer.

VM Access

- **VM #1 (HPC_Master)**: When accessing the cluster via SSH, you connect to this VM, which acts as the central control point for the cluster.

- **VM #2 (Windows10-RemoteAccess):** Primarily used for remote desktop services like AnyDesk or Ultraviewer, allowing administrators to manage the cluster from outside the CUSAT network.
- **VM#3-4 (TestNode01-02):** Used for testing codes before submitting them to the compute nodes.
- **VM#05 (DataNode):** Node reserved for data transferring tasks.

Taking a Snapshot in VMware ESXi

A snapshot in VMware ESXi is a powerful feature that captures the state of a VM at a specific time. It includes the VM's memory, disk, and other settings. Snapshots are useful for backup, testing, troubleshooting, and recovery scenarios. By creating a snapshot, you can roll back a VM to a previous state if needed without affecting the current configuration of the VM.

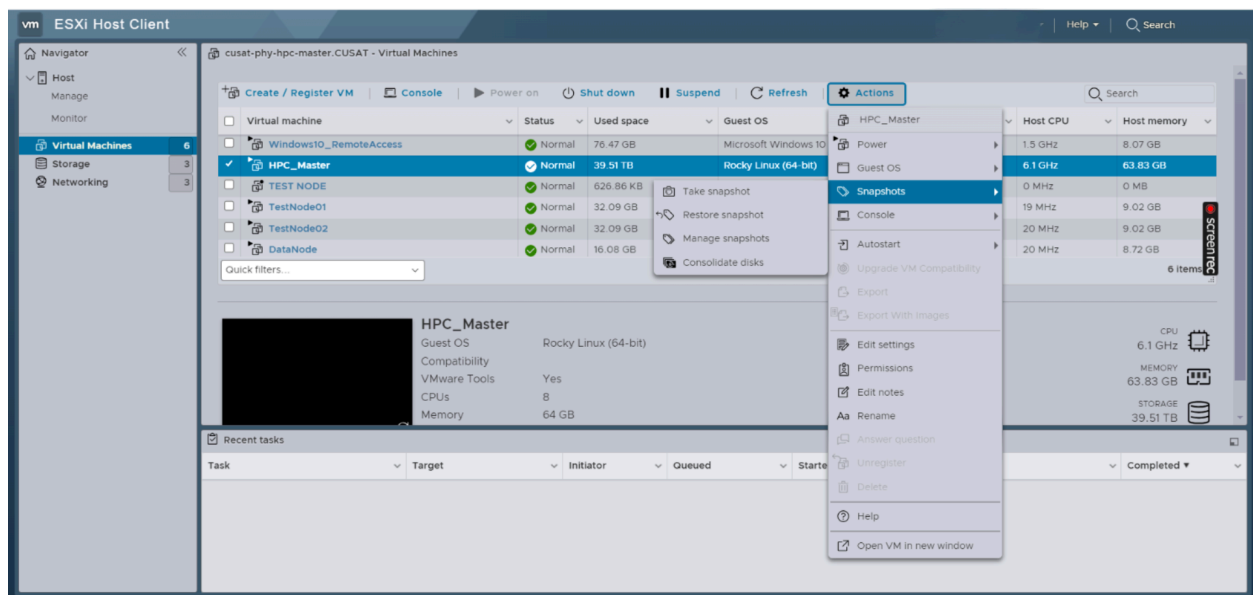


Fig: Taking a snapshot of the Cluster data using VM ESXi Host Client

How to Take a Snapshot in VMware ESXi

Follow these steps to create a snapshot of a VM in VMware ESXi:

1. Access VMware ESXi Host Client
 - *Open a web browser and connect to the ESXi host using the provided IP address.*
 - *Log in using your administrative credentials.*
2. Navigate to the Virtual Machine List
 - *In the VMware ESXi dashboard, locate the "Virtual Machines" section.*
 - *Select the virtual machine for which you want to take a snapshot.*
3. Initiate the Snapshot
 - *Click on the VM's name to open the VM management page.*
 - *In the "Actions" drop-down menu, select Snapshots, then choose Take Snapshot.*
4. Configure Snapshot Settings
 - *Name the Snapshot*
 - *Description (optional): Add any relevant details about why the snapshot is being taken.*
 - *Snapshot Options:*
 - ***Snapshot the VM's memory:*** *Captures the current state of the VM's memory, allowing you to revert to the exact moment of the snapshot. It is useful if you want the VM to return to running.*
 - ***Quiesce guest file system:*** *This ensures that the file system is consistent before the snapshot is taken. It's beneficial for databases and transactional applications.*
5. Take the Snapshot
 - *After configuring the snapshot settings, click **OK** to create the snapshot.*
 - *The snapshot process may take a few seconds to a few minutes, depending on the size of the VM and the options selected.*
6. Monitor the Snapshot Progress
 - *Once the snapshot is initiated, you can monitor the progress in the Recent Tasks section of the ESXi client interface.*
 - *After completion, the snapshot will be available for the selected VM under the "Snapshots" tab.*

Managing Snapshots

- **Reverting to a Snapshot:** If you need to revert to a previous snapshot, go to the "Manage Snapshots" section for the VM, select the desired snapshot, and choose **Revert to**.
- **Deleting Snapshots:** To free up storage, delete snapshots once they are no longer needed. Select the snapshot in the "Manage Snapshots" section and choose **Delete**. Deleting a snapshot merges the changes made after the snapshot into the current VM state.

iDRAC

Integrated Dell Remote Access Controller (iDRAC) is a management solution built into Dell servers, designed to provide administrators with remote access to manage and monitor the server, regardless of the server's state or operating system condition. iDRAC allows system administrators to configure, monitor, and update Dell servers without being physically near the machine.

Key Functionalities of iDRAC:

- **Remote Monitoring & Management:** Access system health, hardware status, and environmental metrics.
- **Power Control:** Remotely power on/off and reboot servers.

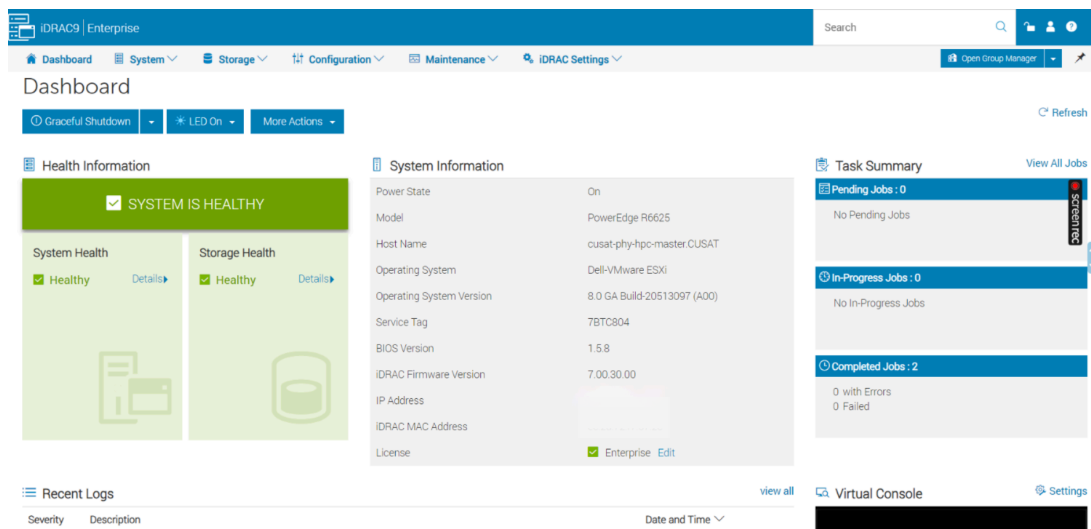


Fig: Snapshot of UI of iDRAC of Master Node

- **Virtual Console & Media:** Remotely manage server inputs and mount media (ISO files).
- **Hardware Configuration & Alerts:** Configure BIOS and RAID and receive real-time alerts related to server power and temperature.
- **Firmware Updates:** Perform remote firmware and software updates.
- **Out-of-Band Management:** Manage servers even if the OS is down or uninstalled.
- **Lifecycle Controller Integration:** Streamline provisioning, diagnostics, and deployment.
- **Security:** Supports role-based access control (enabling administrators to define different user permission levels), secure communications (SSL/TLS), and multi-factor authentication.
- **Automation:** Redfish API and RACADM enable scripting and automation.

- **Health Diagnostics:** Perform detailed hardware diagnostics and access event logs.

The screenshot displays the iDRAC9 Enterprise web interface. The top navigation bar includes links for Dashboard, System, Storage, Configuration, Maintenance, and iDRAC Settings. The main content area is divided into two sections: Recent Logs and Virtual Console.

Recent Logs: A table listing system events with columns for Severity, Description, and Date and Time.

Severity	Description	Date and Time
✓	The system inlet temperature is within range.	Mon Sep 30 2024 14:25:09
⚠	The system inlet temperature is greater than the upper warning threshold.	Mon Sep 30 2024 14:15:37
✓	The system inlet temperature is within range.	Fri Sep 27 2024 09:22:16
⚠	The system inlet temperature is greater than the upper warning threshold.	Fri Sep 27 2024 09:15:36
✓	The system inlet temperature is within range.	Fri Sep 20 2024 09:45:50
⚠	The system inlet temperature is greater than the upper warning threshold.	Fri Sep 20 2024 09:43:04
✗	The system inlet temperature is greater than the upper critical threshold.	Fri Sep 20 2024 08:57:44
⚠	The system inlet temperature is greater than the upper warning threshold.	Fri Sep 20 2024 08:49:58
✓	The input power for power supply 1 has been restored.	Fri Sep 13 2024 19:00:12
✓	The power supplies are redundant.	Fri Sep 13 2024 19:00:09

Virtual Console: A section on the right side of the interface, currently showing a black screen with a 'Start the Virtual Console' button at the bottom.

Fig: Log and Virtual Console in iDRAC

Group Manager in iDRAC provides a convenient way to manage multiple Dell servers from a single interface, simplifying overseeing different servers. The group manager can be launched directly after logging into the master node. From the group manager, the iDRAC of individual nodes can be launched without authentication.

The screenshot displays the iDRAC9 Group Manager web interface. The top navigation bar includes links for Grouped Servers, Summary, Discovered Servers, and Jobs. The main content area is divided into two sections: Summary and Discovered Servers.

Summary: A section on the left side of the interface, showing a donut chart for Compute (12) and a table for Top 2 System Model(s).

Compute (12): A donut chart showing the status of 12 servers: 0 Critical, 0 Warning, 12 Healthy, and 0 Unknown.

Top 2 System Model(s): A table listing the top 2 system models.

Server Model	Mix	Number
PowerEdge R6625	91.67%	11
PowerEdge R7625	8.33%	1

Discovered Servers: A table listing discovered servers with columns for Health, Host Name, iDRAC IP Addresses, Service Tag, Model, iDRAC Firmware Version, and Last Status Update.

Health	Host Name	iDRAC IP Addresses	Service Tag	Model	iDRAC Firmware Version	Last Status Update
✓			G9TC804	PowerEdge R6625	7.00.30.00	Wed 25 Sep 2024 10:50:08
✓			6BTC804	PowerEdge R7625	7.10.30.00	Tue 24 Sep 2024 10:19:05
✓			F9TC804	PowerEdge R6625	7.00.30.00	Wed 25 Sep 2024 10:49:52

System Information: A section on the right side of the interface, showing details for the selected server (PowerEdge R6625).

System Information: A table listing system information for the selected server.

System Information	Value
iDRAC Connectivity	Online

Fig: Snapshot of Group Manager UI

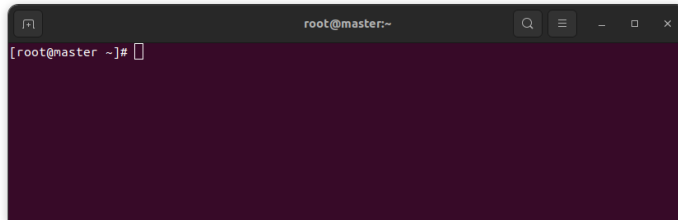
More information about iDRAC and its functionalities can be found [here](#).

User Creation & Management

Creating a user with Administrative Privileges:

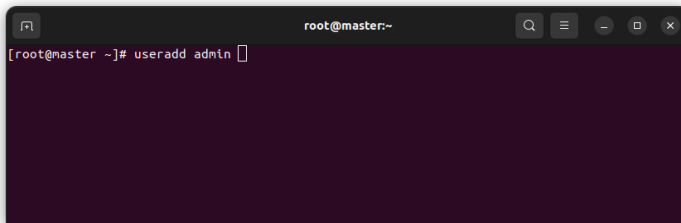
Step 1: Login as root user

- SSH into the admin console: `ssh root@<ip_address>`



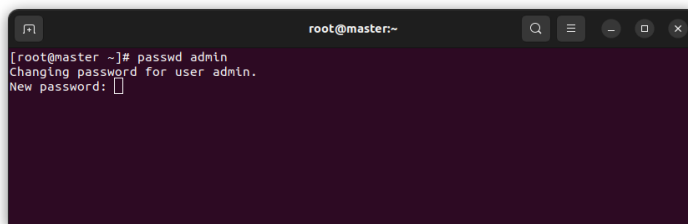
Step 2: Add the user

- Add user by: `useradd <username>`



Step 3: Set the password

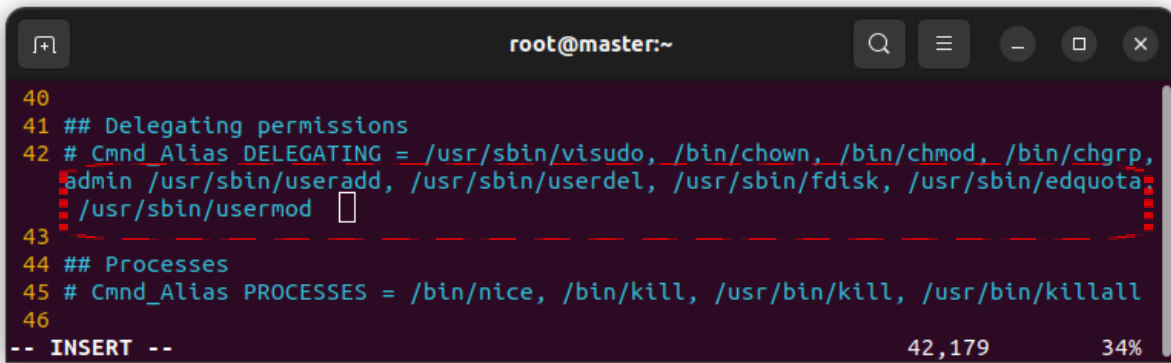
- Password can be set by the command: `passwd <username>`



Step 4: Update “Sudoers” to provide permissions for the admin user

- The super-user (root user) should grant the admin user special privileges.
- Special privileges can be granted by editing the `sudoers` file.
- Editing the `sudoers` file can be done using the Vim editor.
 - `vim /etc/sudoers/`
- To grant permissions, the `sudoers` file must be edited at two locations.

- Location-1: Line 42

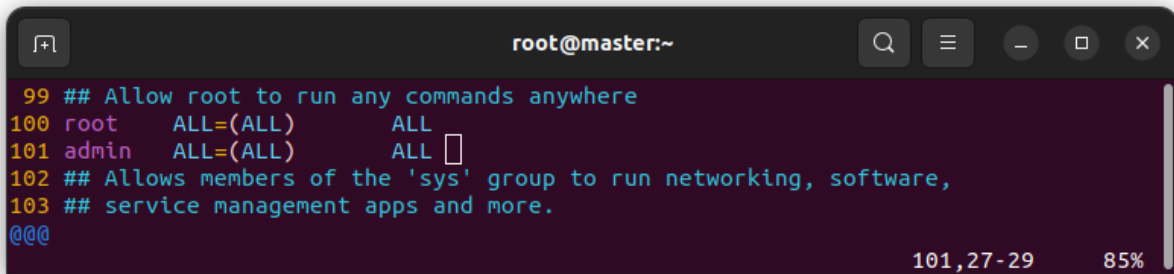


```
root@master:~  
40  
41 ## Delegating permissions  
42 # Cmnd_Alias DELEGATING = /usr/sbin/visudo, /bin/chown, /bin/chmod, /bin/chgrp,  
admin /usr/sbin/useradd, /usr/sbin/userdel, /usr/sbin/fdisk, /usr/sbin/edquota  
/usr/sbin/usermod  
43  
44 ## Processes  
45 # Cmnd_Alias PROCESSES = /bin/nice, /bin/kill, /usr/bin/kill, /usr/bin/killall  
46  
-- INSERT -- 42,179 34%
```

- Type the username followed by the location of the permissions
 - In the figure, we give permissions to a user named “admin”, highlighted in the Redbox.
 - Common permissions required for admin:

Permission	Location	Use
useradd	/usr/sbin/useradd	To create a new user (standard).
userdel	/usr/sbin/userdel	To delete a user.
fdisk	/usr/sbin/fdisk	To partition the hard drive.
edquota	/usr/sbin/edquota	To set up a disk quota. Disk quota allows administrators to control the number of files and data blocks that can be allocated to groups or users.
usermod	/usr/sbin/usermod	Usermod command or modify user is a command in Linux that is used to change the properties of a user Eg: password, username etc.

- Location-2: after **Line 100**

A terminal window titled 'root@master:~' with a dark background. It shows a configuration file with the following lines: 99 ## Allow root to run any commands anywhere, 100 root ALL=(ALL) ALL, 101 admin ALL=(ALL) ALL, 102 ## Allows members of the 'sys' group to run networking, software, 103 ## service management apps and more. The prompt '###' is visible at the bottom left. The bottom right corner shows '101,27-29' and '85%'.

```
root@master:~
99 ## Allow root to run any commands anywhere
100 root  ALL=(ALL)    ALL
101 admin ALL=(ALL)    ALL
102 ## Allows members of the 'sys' group to run networking, software,
103 ## service management apps and more.
###
101,27-29 85%
```

- The 100-th line grants permission for the root user to execute any command (**ALL**) as any user (**ALL**) on any host (**ALL**).
- The syntax can be broken down like the following:

user_or_group **host=(run_as_user)** **commands**

- **user_or_group**: This is the user or group to which the sudo permissions apply.
 - **host**: This specifies the host or hosts on which the sudo permissions apply. If you see (ALL), it means any host.
 - **(run as user)**: This specifies the user to whom the commands will be run. If you see (ALL), it means any user.
 - **commands**: This specifies the commands that the user or group is allowed to run with sudo privileges. If you see (ALL), it means any commands
- Similar permissions should be given to the admin user, as shown in the 101-th line.

Note:

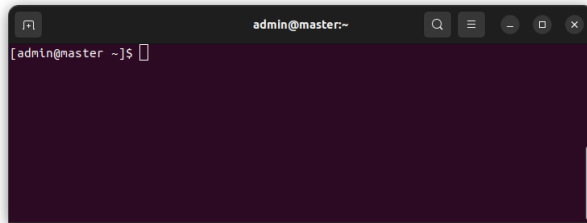
- To execute commands whose location is in **/usr/sbin** the admin needs to enter their password, while the super-user can execute them without a password.
- To execute **sbin** commands without entering the password every time we can use,

admin ALL=(ALL) NOPASSWD:ALL

Creating a standard user:

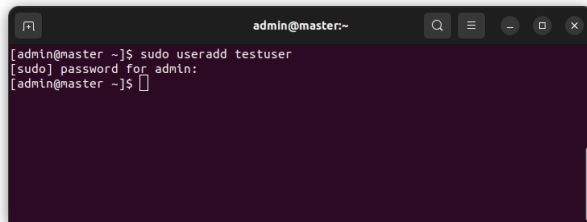
Step 1: Login as Administrator

- SSH into the admin console: `ssh admin@<ip_address>`

A terminal window titled 'admin@master:~' with search, menu, and window control icons. The prompt is '[admin@master ~]\$' followed by a cursor.

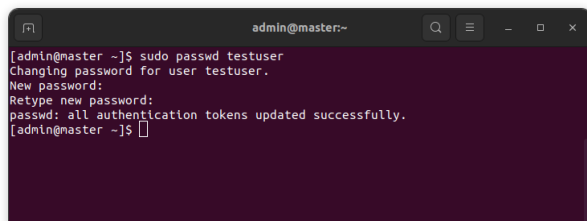
Step 2: Add the user

- Add user: `sudo useradd -u XXXX -d /home/user/user_id <user_id>`

A terminal window titled 'admin@master:~' with search, menu, and window control icons. The prompt is '[admin@master ~]\$'. The command 'sudo useradd testuser' is entered. The prompt changes to '[sudo] password for admin:'. The command 'admin@master ~\$' is entered again, followed by a cursor.

Step 3: Set the password

- Password can be set by the command: `sudo passwd <user_id>`

A terminal window titled 'admin@master:~' with search, menu, and window control icons. The prompt is '[admin@master ~]\$'. The command 'sudo passwd testuser' is entered. The prompt changes to '[sudo] password for admin:'. The command 'admin@master ~\$' is entered again. The output shows: 'Changing password for user testuser.', 'New password:', 'Retype new password:', 'passwd: all authentication tokens updated successfully.', followed by the prompt 'admin@master ~\$' and a cursor.

Step 4: Set up memory quota

- The quota that limits the maximum filesize a user can create can be set by the command: `sudo edquota -u <user_id>`
- It will open up a file in the terminal that needs to be edited in the section given in the red box.

```
admin@master:~  
Disk quotas for user testuser (uid 1006):  
Filesystem      blocks      soft      hard      inodes      soft      hard  
/dev/sdb1        132         0      50G         32         0         0  
-- INSERT --
```

Note:

- To edit the file, enter into INSERT mode by pressing the button “i”
- **hard 50G** indicates 50GB of memory allocated to the user. The maximum file size a user can create is 50 GB.
- To save the file, **press the escape button and type “:wq”**
- To check the quota allocation process is successful, we can use the following command:
> sudo repquota -as

Block size and Inode number

- Disk quotas can be configured by using the **block size** and **Inode number**.
- If we want to control the size of files, we would configure the quota based on block size.
- If we want to control the number of files, we would configure the quota based on the inode number.
- To control both, we need to specify both.
- While setting up the edquota, we can see the following seven columns:

Column	Name	Description
1	Filesystem	Partition where this quota will apply
2	blocks	Number of blocks currently used by this user
3	soft	Soft block size limit for users. It gives a warning sign to the user if the file size exceeds the soft limit.
4	hard	Hard block size limit for users.Sets the maximum size of the file that can be created by the user.
5	inodes	Number of inodes currently used by this user
6	soft	Soft inodes limit for users. Warns the user if the number of files exceeds the soft limit.
7	hard	hard inodes limit for users. Sets the maximum number of files that can be created.

Adding User to Group:

In the Kuria cluster, users are classified at two levels: by their designation and by their Research Group affiliation.

Designation-based classification:

Users are categorised into three types based on their designation:

1. **Faculty**
2. **PhD User**
3. **Student User**

Each user is added to a corresponding group after creation:

- Faculty → **FACULTY-GROUP**
- PhD → **PHD-GROUP**
- Student → **STUDENT-GROUP**

These groups determine the Slurm partitions accessible to the users and their storage limits.

Users can be added to the appropriate group with the following command:

```
> sudo gpasswd -a <user_id> <group_id>
```

Details of Slurm partitions and storage limits:

User-type	Partitions available	Storage Limit
Faculty	normal, gpu, test, faculty	500 GB
PhD	normal, gpu, test, phd	300 GB
Student	normal, gpu, test	200 GB

Partition Access:

Name of Partition	Time Limit	Node Access
normal	1 day	Nodes 01-11
gpu	infinite	Node 11
test	1 hr	Node 12-13
data (**available at request)	infinite	Node 14

faculty	14 days	Nodes 01-11
phd	7 days	Nodes 01-11

Research Group-based classification:

In the Kuria cluster, users are also categorised by their Research Group affiliation, with each group named after the group's Principal Investigator (PI). This classification is used to control and limit resource usage within each research group, ensuring fair distribution of resources among all groups.

Convention followed for usernames:

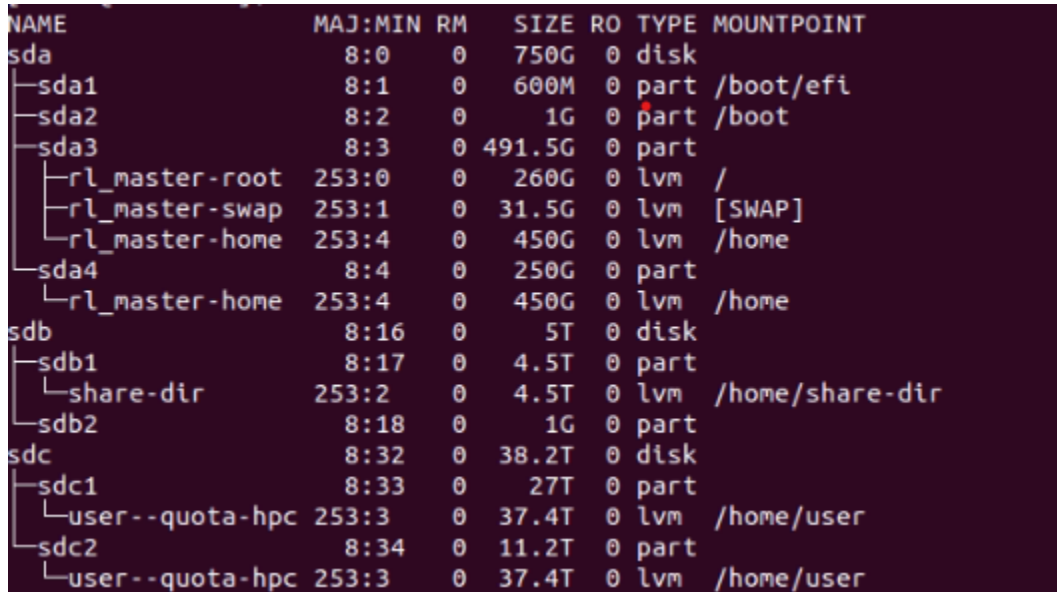
Each user in the Kuria cluster is identified by their research group and department to which they are affiliated. Each research group accessing the Kuria Cluster will obtain 3 user IDs: Faculty, PhD and student ID.

Let a research group led by PI John Doe be affiliated with the Department of Physics. The user IDs that the research group will obtain are given by;

Type of User	User id
Faculty ID	jd_phy
PhD ID	pjd_phy
Student ID	sjd_phy

Filesystem

Kuria Cluster utilizes three main disks, with a combination of standard partitions and Logical Volume Management (LVM) for flexible disk management.



NAME	MAJ:MIN	RM	SIZE	RO	TYPE	MOUNTPOINT
sda	8:0	0	750G	0	disk	
├─sda1	8:1	0	600M	0	part	/boot/efi
├─sda2	8:2	0	1G	0	part	/boot
├─sda3	8:3	0	491.5G	0	part	
│ └─rl_master-root	253:0	0	260G	0	lvm	/
│ └─rl_master-swap	253:1	0	31.5G	0	lvm	[SWAP]
│ └─rl_master-home	253:4	0	450G	0	lvm	/home
└─sda4	8:4	0	250G	0	part	
└─rl_master-home	253:4	0	450G	0	lvm	/home
sdb	8:16	0	5T	0	disk	
├─sdb1	8:17	0	4.5T	0	part	
│ └─share-dir	253:2	0	4.5T	0	lvm	/home/share-dir
└─sdb2	8:18	0	1G	0	part	
sdc	8:32	0	38.2T	0	disk	
├─sdc1	8:33	0	27T	0	part	
│ └─user--quota-hpc	253:3	0	37.4T	0	lvm	/home/user
└─sdc2	8:34	0	11.2T	0	part	
└─user--quota-hpc	253:3	0	37.4T	0	lvm	/home/user

1. Disk sda (750GB):

- **sda1** (600MB): Mounted on **/boot/efi** this partition is for the EFI system and is essential for booting in UEFI mode.
- **sda2** (1GB): Mounted on **/boot**, it contains the kernel and other files required for booting.
- **sda3** (491.5GB): This is split into three logical volumes using LVM:
 - **rl_master-root** (260GB): Mounted as the root (**/**), this contains the operating system and essential system files.
 - **rl_master-swap** (31.5GB): Dedicated for swap space, used for virtual memory management.
 - **rl_master-home** (450GB): Used for storing administrator specific data under **/home**.

2. Disk **sdb** (5TB):

- **sdb1** (4.5TB): This partition is part of the LVM volume **share-dir** (4.5TB), mounted on **/home/share-dir**, for shared data across users and for storing job logs of all users in the cluster.
- **sdb2** (1GB): This partition isn't allocated to any mounted filesystem.

3. Disk **sdc** (38.2TB):

- **sdc1** (27TB): Part of the LVM volume **user-quota-hpc** (37.4TB), mounted under **/home/user**. Used for storing the info about the users in the cluster
- **sdc2** (11.2TB): Also part of the **user-quota-hpc** logical volume, contributing to the 37.4TB user storage.

Storage Configuration

- **LVM (Logical Volume Management):** LVM is used for flexibility and easier disk management, allowing dynamic resizing of partitions. The root filesystem, swap space, home directories, and user quotas are all managed through LVM volumes.
- **User Quotas:**
Disk **sdc** is allocated for user-related data under the LVM volume **user-quota-hpc**.