Best Practices for Deploying Hadoop Server on CentOS/RHEL 7 – Part 1

Mohan Sivam | Last Updated: February 17, 2021 | Read Time: 4 mins | CentOS, Hadoop, RedHat | Leave a comment

In this series of articles, we are going to cover the entire **Cloudera Hadoop Cluster Building** building with **Vendor** and **Industrial** recommended best practices.

> Part 1: Best Practices for Deploying Hadoop Server on CentOS/RHEL 7

> Part 2: Setting Up Hadoop Pre-requisites and Security Hardening

> Part 3: How to Install and Configure the Cloudera Manager on CentOS/RHEL 7

> Part 4: How to Install CDH and Configure Service Placements on CentOS/RHEL 7

> Part 5: How to Set Up High Availability for Namenode

> Part 6: How to Set Up High Availability for Resource Manager

> Part 7: How to Install and Configure Hive with High Availability

> Part 8: How to Install and Configure Sentry (Authorization Tool)

> Part 9: How to Install Kerberos (Kerberising the Cluster) for Hadoop Authentication

> Part 10: How to Tune Cluster (Yarn Tuning) on CentOS/RHEL 7

OS installation and doing OS level Pre-requisites are the first steps to build a **Hadoop Cluster**. **Hadoop** can run on the various flavor of Linux platform: **CentOS**, **RedHat**, **Ubuntu**, **Debian**, **SUSE** etc., In real-time production, most of the **Hadoop Clusters** are built on top of RHEL/CentOS, we will use **CentOS 7** for demonstration in this series of tutorials.

In an Organization, OS installation can be done using **kickstart**. If it is a 3 to 4 node cluster, manual installation is possible but if we build a big cluster with more than 10 nodes, it's tedious to install OS one by one. In this scenario, the Kickstart method comes into the picture, we can proceed with the mass installation using kickstart.

Achieving good performance from a **Hadoop Environment** is depends on provisioning the correct Hardware & Software. So, building a production **Hadoop cluster** involves a lot of consideration regarding Hardware and Software.

In this article, we will go through various Benchmarks about OS installation and some best practices for deploying **Cloudera Hadoop Cluster Server** on **CentOS/RHEL 7**.

Important Consideration and Best Practices for Deploying Hadoop Server

The following are the best practices for setting up deploying **Cloudera Hadoop Cluster Server** on **CentOS/RHEL 7**.

- Hadoop servers do not require enterprise standard servers to build a cluster, it requires commodity hardware.

- In the production cluster, having 8 to 12 data disks are recommended. According to the nature of the workload, we need to decide on this. If the cluster is for compute-intensive applications, having 4 to 6 drives is best practice to avoid I/O issues.
- Data drives should be partitioned individually, for example – starting from **/data01** to **/data10**.
- RAID configuration is not recommended for worker nodes, because Hadoop itself providing fault-tolerance on data by replicating the blocks into 3 by default. So **JBOD** is best for worker nodes.
- For Master Servers, **RAID 1** is the best practice.
- The default filesystem on **CentOS/RHEL 7.x** is **XFS**. Hadoop supports XFS, ext3, and ext4. The recommended file-system is ext3 as it is tested for good performance.
- All the servers should be having the same OS version, at-least same minor release.
- It is best practice to have homogeneous hardware (all worker nodes should have the same hardware characteristics (RAM, disk space & Core etc).
- According to the cluster workload (Balanced Workload, Compute Intensive, I/O Intensive) and size, resource (RAM, CPU) planning per server will get differ.

Find the below Example for Disk Partitioning of the servers of 24TB storage.

Partition Information	OS Disks (2 x 1TB RAID 1)	Worker Node A)DataNode Disks (22 x 1TB JBOD)		Master Node B)NameNode Disks (22 x 1TB RAID 1 / JBOD)		Cloudera Manager C)CM Disk (22 x 1TB RAID 1)		Edge Node D)Edge Disk (22 x 1TB RAID 1)	
Partition	Size	Partition	Size	Partition	Size	Partition	Size	Partition	Size
swap	16GB	/data01	1TB	/nn (namenode)	8 TB (Useable: 4TB - **RAID 1**)	/	20 TB (Useable: 10TB)	/	22 TB (Useable: 11TB)
/boot	1GB	/data02	1TB	/zk (zookeeper)	2 TB JBOD	/sql_data (CM Metadata)	2 TB (Useable: 1TB)		
/home	50GB	/data03	1TB	/jn (journal node)	2 TB JBOD				
/opt	100GB	/data04	1TB	/	Remaining				
/tmp	100GB	/data05	1TB						
/usr	50GB	/data06	1TB						
/var	100GB	/data07	1TB						
/var/log	100GB	/data08	1TB						
/var/log/audit	100GB	/data09	1TB						
/	All Remaining	...../data22	1TB						

Disk Partitioning

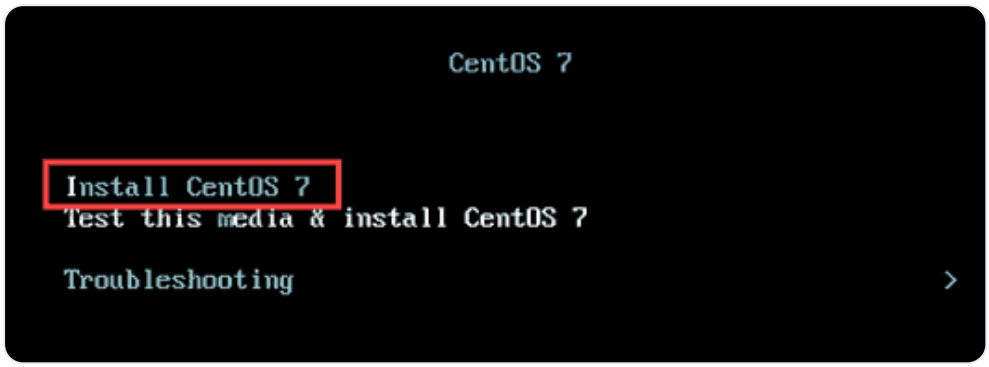
Installing CentOS 7 for Hadoop Server Deployment

Things you need to know before installing CentOS 7 server for Hadoop Server.

- Minimal installation is enough for Hadoop Servers (worker nodes), in some cases, GUI can be installed only for Master servers or Management servers where we can use browsers for Web UIs of Management tools.
- Configuring networks, hostname, and other OS-related settings can be done after OS installation.
- In real-time, server vendors will be having their own console to interact and manage the servers, for example – Dell servers are having iDRAC which is a device, embedded with servers. Using that iDRAC interface we can install OS with having an OS image in our local system.

In this article, we have installed OS (CentOS 7) in VMware virtual machine. Here, we will not be having multiple disks to perform partitions. CentOS is similar to RHEL (same functionality), so we will see the steps to install CentOS.

1. Begin by downloading the CentOS 7.x ISO image in your local windows system and select it while booting the virtual machine. Select 'Install CentOS 7' as shown.

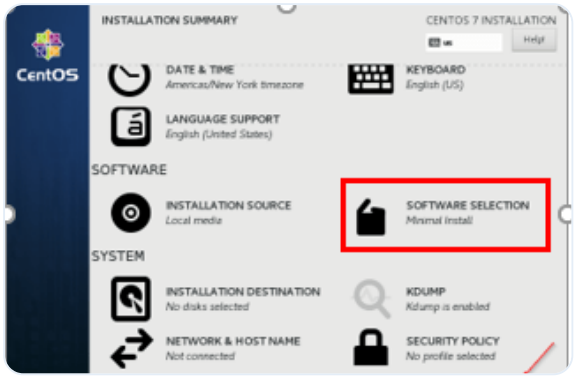


Install CentOS 7 Boot Menu

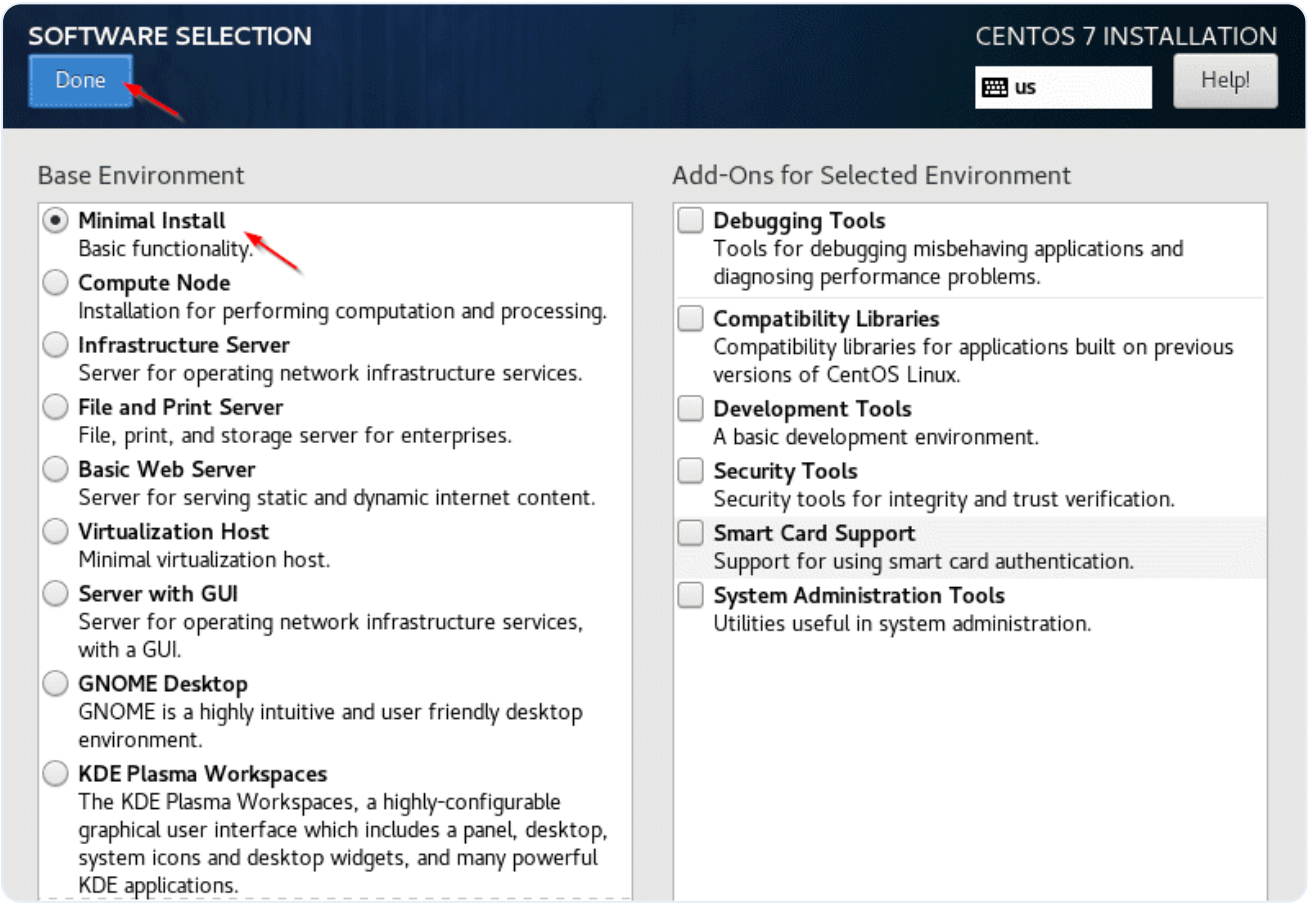
2. Select the Language, default will be English, and click continue.

Select CentOS 7 Language

3. Software Selection – Select the 'Minimal Installation' and click 'Done'.

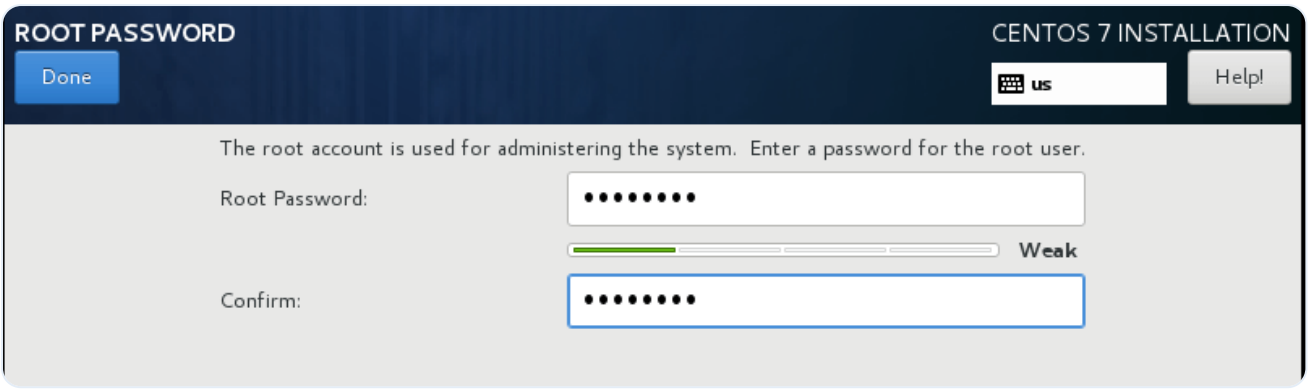


CentOS Software Selection



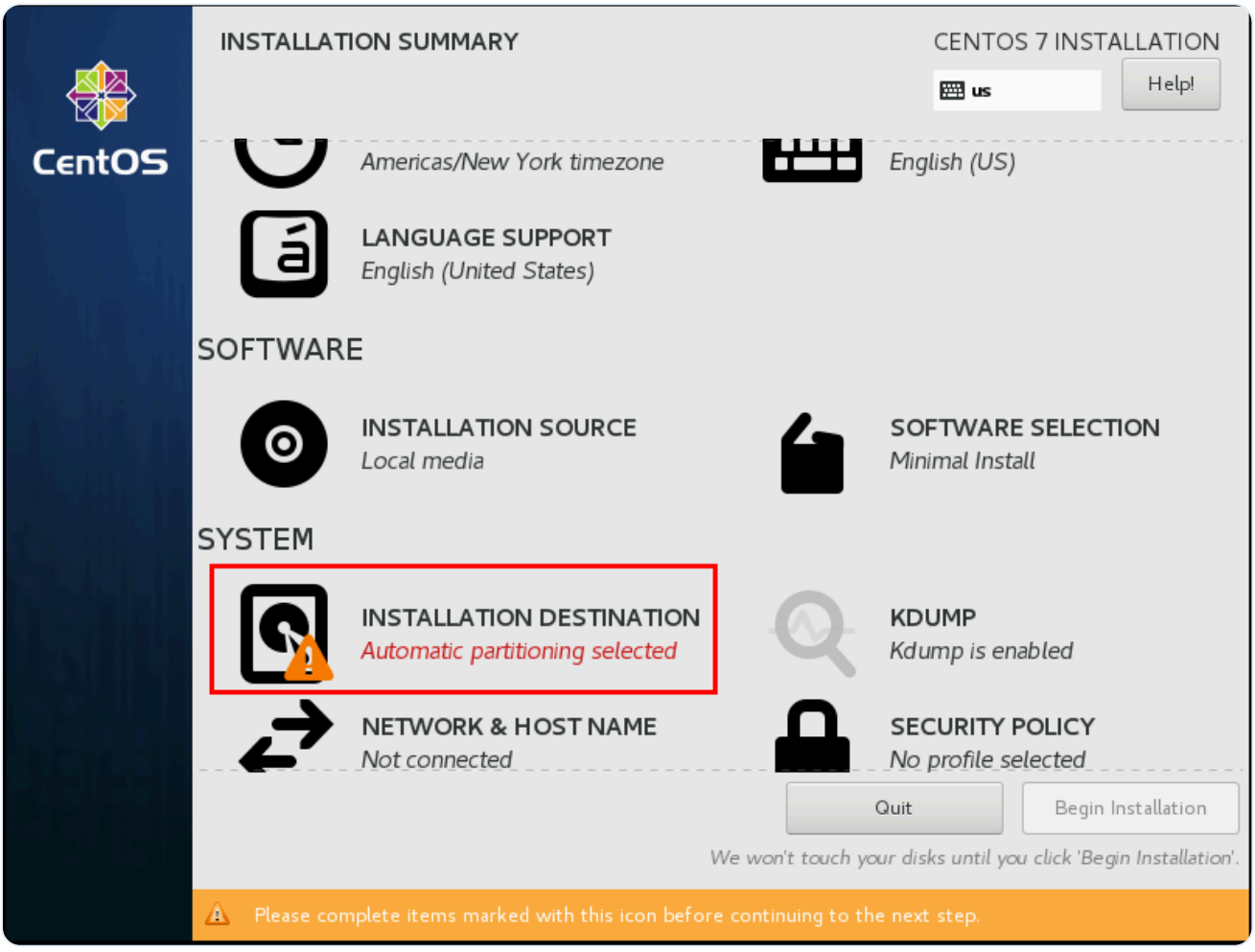
CentOS 7 Minimal Installation

4. Set the root password as it will prompt us to set.

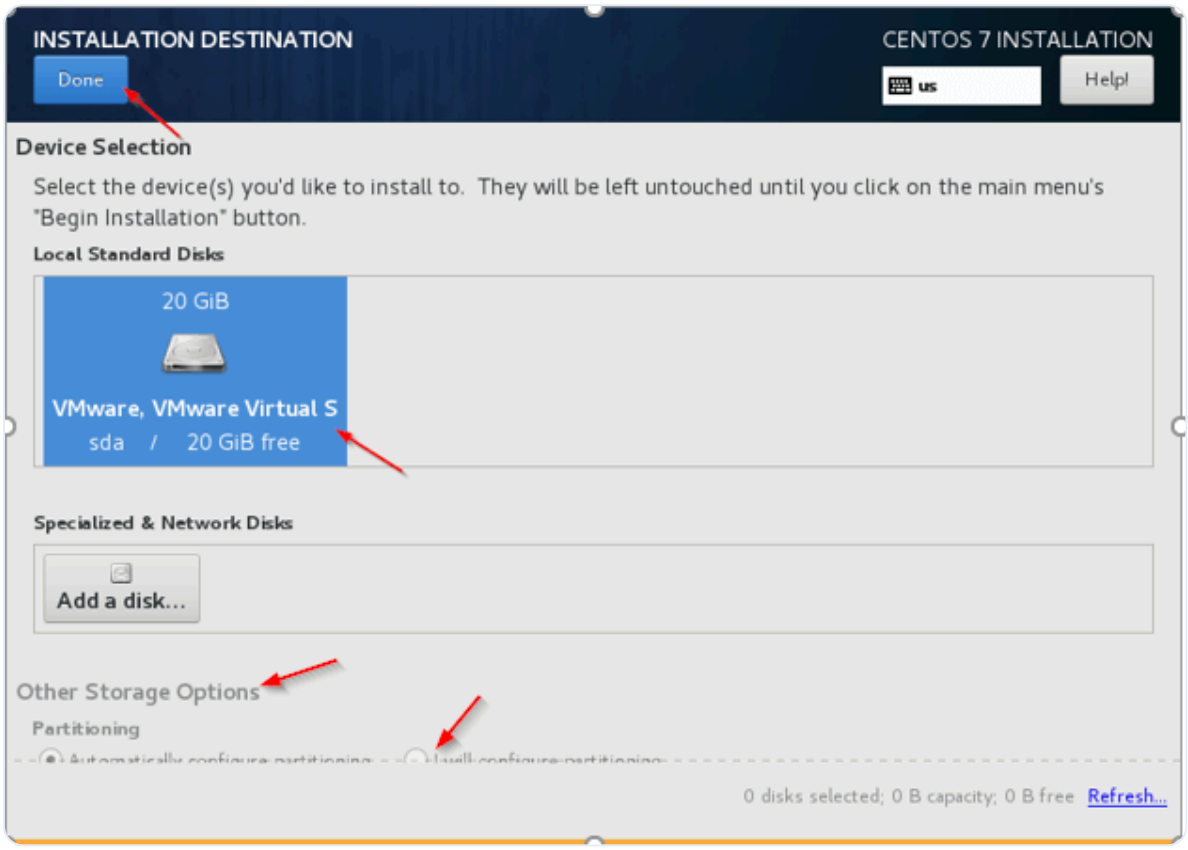


Set Root Password

5. Installation Destination – This is the important step to be cautious. We need to select the disk where the OS has to be installed, dedicated disk should be selected for OS. Click the 'Installation Destination' and select the Disk, in real-time multiple disks will be there, we need to select, preferable 'sda'.

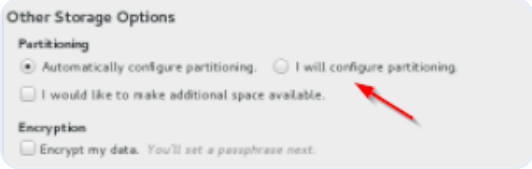


Select Installation Destination



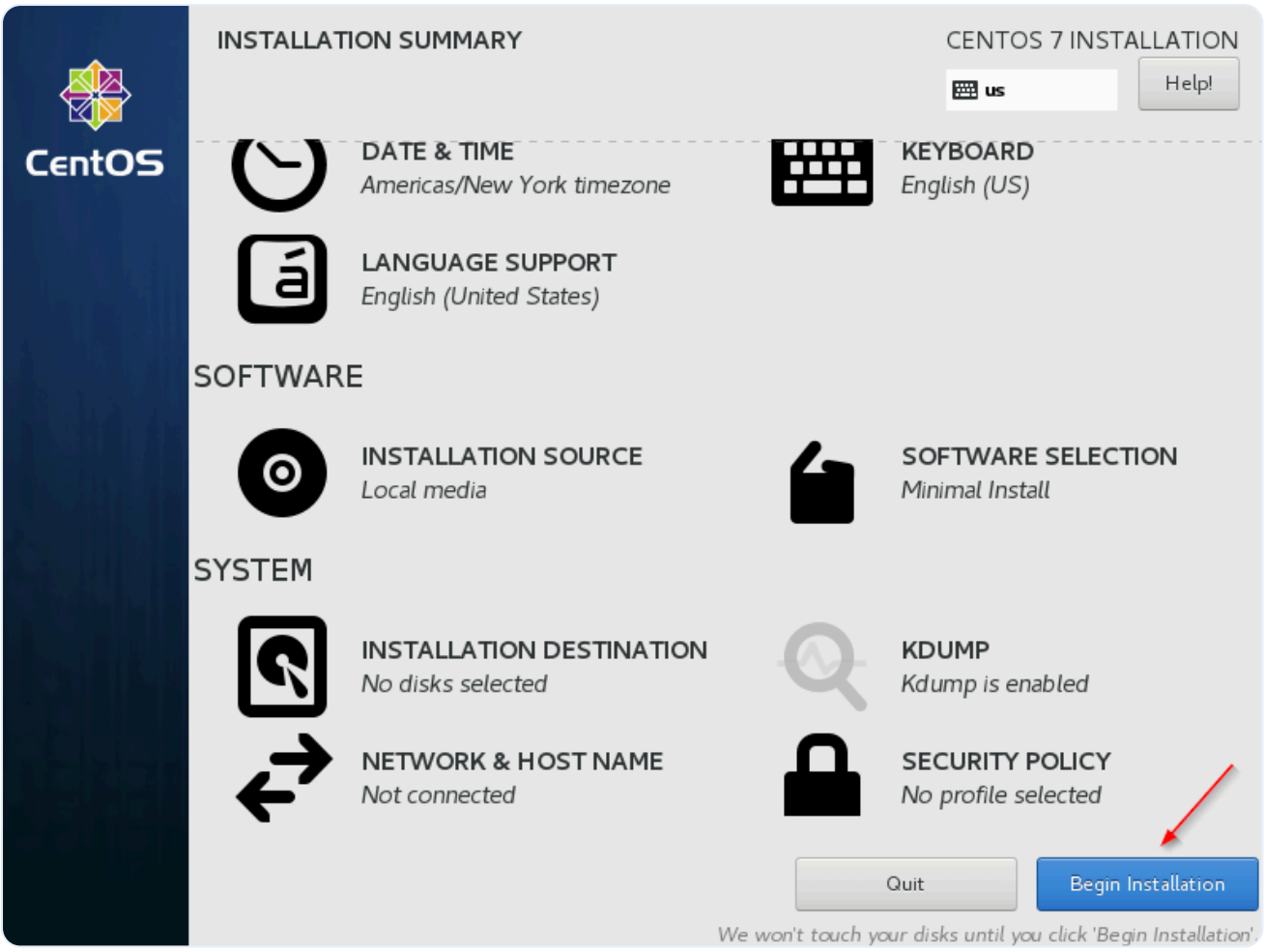
Select Disk for CentOS Installation

6. **Other Storage Options** – Choose the second option (I will configure partitioning) to configure OS related partitioning like /var, /var/log, /home, /tmp, /opt, /swap.

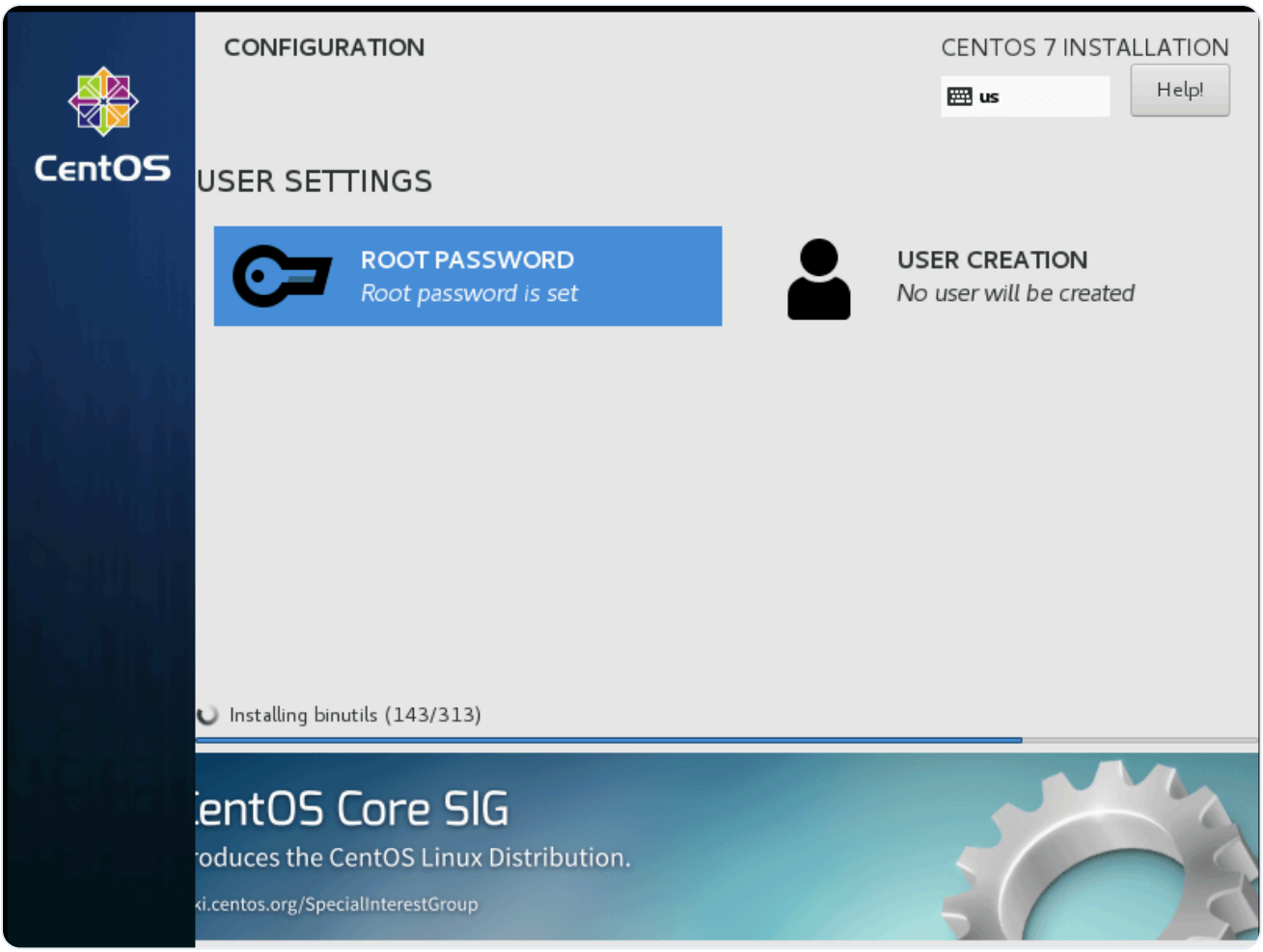


Manual CentOS Partitioning

7. Once done, begin the installation.

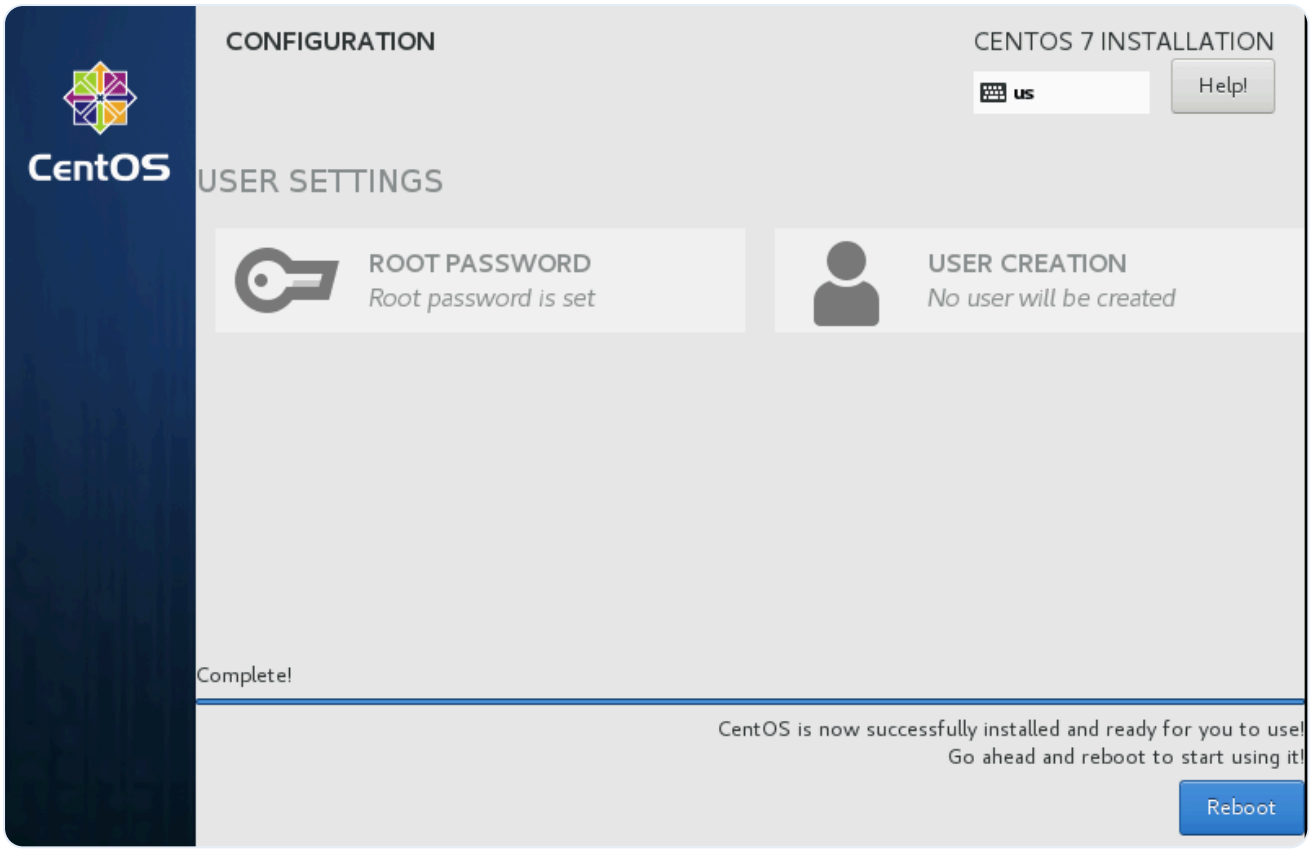


Begin CentOS Installation



CentOS 7 Installation

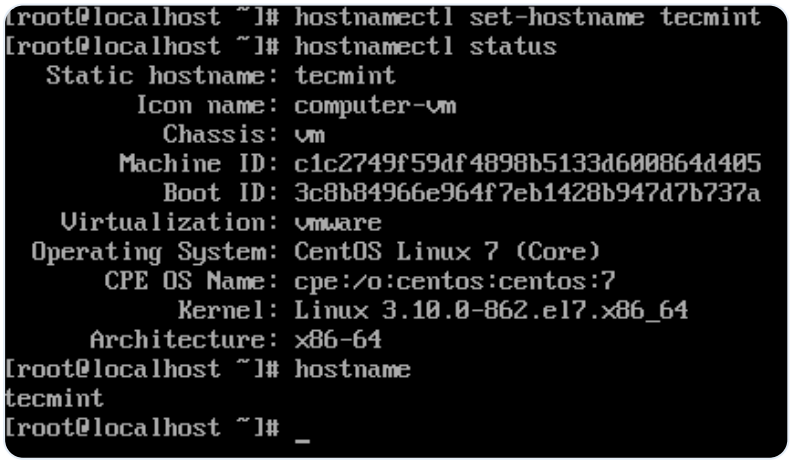
8. Once the Installation completed, reboot the server.



CentOS 7 Installation Complete

9. Login into the server and set the hostname.

```
# hostnamectl status
# hostnamectl set-hostname tecmint
# hostnamectl status
```



Set Hostname on CentOS

Summary

In this article, we have gone through OS installation steps and best practices for filesystem partitioning. These are all general guideline, according to the nature of the workload, we may need to concentrate on more nuances to achieve the best performance of the cluster. Cluster planning is art for the **Hadoop** administrator. We will have deep dive into OS level pre-requisites and security Hardening in the next article.

CentOS Tips, Hadoop Tips, RHEL Tips

PREVIOUS ARTICLE:

How to Monitor Performance Of CentOS 8/7 Server Using Netdata

NEXT ARTICLE:

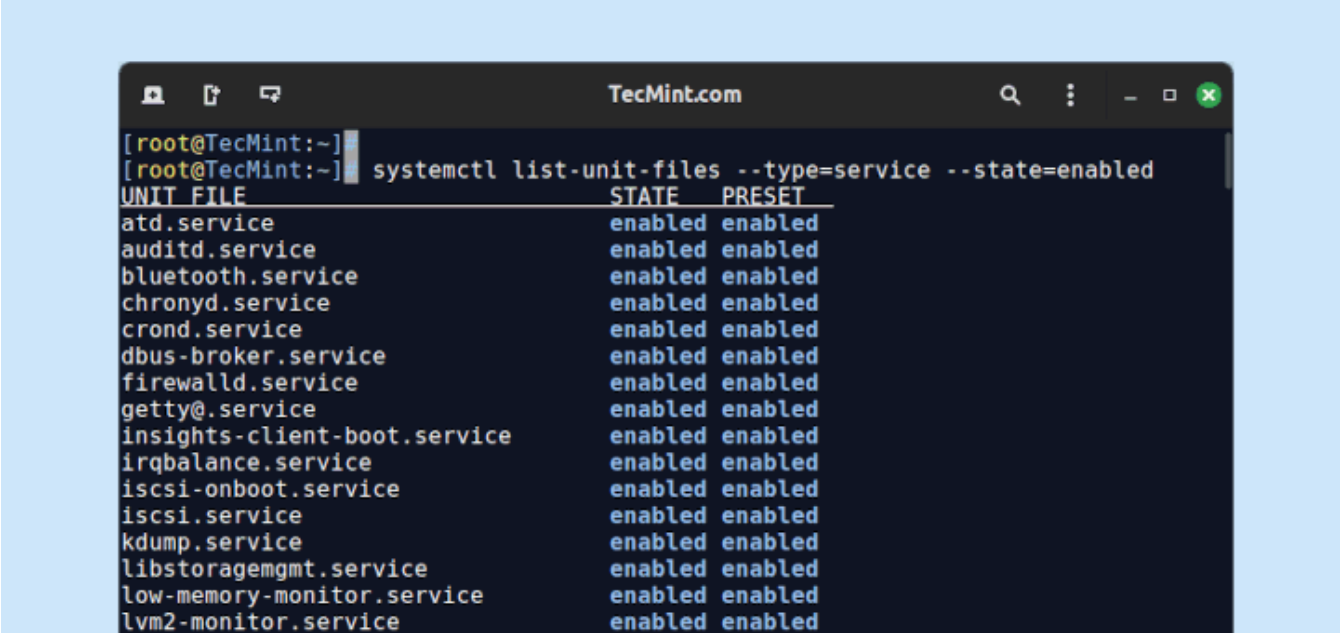
How to Use Platform and Keyword Module in Python

Mohan Sivam

A Solution Architect and Bigdata infrastructure with over 10 years of experience in Information technology.

Each tutorial at TecMint is created by a team of experienced Linux system administrators so that it meets our high-quality *standards*.

Related Posts



How to Disable and Remove Unnecessary Services on Linux

Improve Website Performance – Install Memcached on RHEL 9

How to Migrate CentOS 7 to Rocky Linux 9

How to Reset Forgotten Root Password in RHEL Systems

How to Install and Configure NTP in RHEL Systems

How to Install Terraform (Infrastructure as Code) in Linux

Got Something to Say? Join the Discussion...

Thank you for taking the time to share your thoughts with us. We appreciate your decision to leave a comment and value your contribution to the discussion. It's important to note that we moderate all comments in accordance with our comment policy to ensure a respectful and constructive conversation.

Rest assured that your email address will remain private and will not be published or shared with anyone. We prioritize the privacy and security of our users.

Name *

Email *

Save my name, email, and website in this browser for the next time I comment.

Post Comment

Search...

Do You Enjoy My Blog?

Support from readers like YOU keeps this blog running. Buying me a cup of coffee is a simple and affordable way to show your appreciation and help keep the posts coming!

Buy Me a Coffee

Linux Commands and Tools

How to Undo or Redo a Yum Install on CentOS and RHEL

3 Ways to Extract and Copy Files from ISO Image in Linux

Pscp – Transfer/Copy Files to Multiple Linux Servers Using Single Shell

How to Add Text to Existing Files in Linux

How to Create Hard and Symbolic Links in Linux

Fast – Test Your Internet Download Speed from Linux Terminal

Linux Server Monitoring Tools

Collectl: An Advanced All-in-One Performance Monitoring Tool for Linux

IPTraf-ng – A Console-Based Network Monitoring Tool

How to Install LibreNMS Monitoring Tool on Debian 11/10

How to Monitor Node.js Applications Using PM2 Web Dashboard

All You Need To Know About Processes in Linux [Comprehensive Guide]

5 Tools to Scan a Linux Server for Malware and Rootkits

Learn Linux Tricks & Tips

How to Delete HUGE (100-200GB) Files in Linux

10 Practical Examples Using Wildcards to Match Filenames in Linux

7 Ways to Determine the File System Type in Linux (Ext2, Ext3 or Ext4)

4 Ways to Batch Convert Your PNG to JPG and Vice-Versa

Learn How to Set Your $PATH Variables Permanently in Linux

How to Set and Unset Local, User and System Wide Environment Variables in Linux

Best Linux Tools

6 Most Notable Open Source Centralized Log Management Tools

8 Best Open Source Web Servers in 2024

5 Open Source Log Monitoring and Management Tools for Linux

5 Best Reference Management Software for Linux in 2024

6 Best Whiteboard Applications for Your Linux Systems

Top 5 Open Source Collaboration Platforms for Linux in 2024