

## Using the Dell OpenManage Deployment Toolkit to Automate Configuration of HPC Cluster Nodes

High-performance computing clusters comprising hundreds of nodes require one-to-many management utilities to configure the nodes with customized BIOS, baseboard management controller, and Dell™ Remote Access Controller settings and to deploy Dell Utility Partitions in a reasonable amount of time. Administrators can use the Dell OpenManage™ Deployment Toolkit to perform these one-to-many operations across multiple nodes simultaneously.

BY SHRANKHLA UPADHYAY, MEGHANA BHAT, BARIS GULER, AND ARUN RAJAN

### Related Categories:

High-performance computing (HPC)

System deployment

Systems management

Visit [www.dell.com/powersolutions](http://www.dell.com/powersolutions) for the complete category index.

High-performance computing (HPC) clusters combine several computer systems with a high-speed interconnect to achieve raw output comparable to traditional supercomputers but at a lower cost. HPC clusters divide large individual tasks among several computers, regulate the processes for high throughput, and present the results to a client as if processed by a single system. Such clusters can handle a broad range of numeric- and data-intensive applications efficiently and are finding favor with organizations in many fields because of their cost-to-performance ratio.

HPC cluster architecture features a parallel computing system consisting of one or more master nodes and one or more compute nodes interconnected by a private network system. The master node serves to synchronize the compute nodes and maintain general cluster health; the compute nodes divide and process the computational tasks. The size of a cluster can range from a few hundred to a few thousand nodes, depending on the organization's performance requirements.

The Dell OpenManage suite offers command-line interface utilities to configure BIOS, baseboard management controller (BMC), Dell Remote Access Controller (DRAC), and RAID settings across supported cluster nodes through the network using the Preboot Execution Environment (PXE) or a CD. However, these utilities can be used only after an OS is deployed, and preconfiguring system settings before OS deployment can often be beneficial. Traditional methods for pre-OS configuration include using KVM (keyboard, video, mouse), console redirection, or a crash cart to manually modify the system settings through the boot menu options one server at a time—a process that can be time-consuming, repetitive, and tedious.

Administrators can use the Dell OpenManage Deployment Toolkit (DTK) to perform one-to-many pre-OS operations across multiple Dell PowerEdge™ servers simultaneously. The DTK consists of command-line utilities to configure system features and provides sample scripts to perform common deployment tasks. The DTK utilities can be integrated with the master node's cluster

management software and then pushed onto individual client nodes without configuring each node individually. This article describes how the DTK can be used in a Linux® OS environment for typical configuration tasks performed on HPC clusters before the deployment of the cluster package as well as during maintenance tasks. It also discusses techniques for configuring BIOS, BMC, and DRAC settings and installing the Dell Utility Partition; these techniques can be scaled to include other DTK configuration features.

### Using the Dell OpenManage Deployment Toolkit in HPC clusters

The life cycle of a typical HPC cluster can be divided into several phases, with different management tasks in each phase. The deployment phase involves remote power management, which is provided by the BMC and requires setting up parameters in the BMC options menu such as IP address, netmask, user ID, and password. The operational phase involves cluster health monitoring, which requires both in-band and out-of-band mechanisms be in place. To enable these mechanisms, BMC options such as username, password, and LAN alert destination and BIOS options such as boot sequence, snoop filter, Demand-Based Switching, logical processor, and console redirection settings must be configured.

On ninth-generation servers, the DTK allows administrators to modify most configuration parameters of the remote system, specifically the BIOS, BMC, and DRAC 5 settings. Administrators can use the DTK syscfg tool to capture and replicate these settings. Alternatively, administrators can use an automated script to capture a complete list of BIOS and BMC settings and use this list as a comparison when troubleshooting problems such as failed nodes and performance declines. The DTK can help provide a safe preinstallation OS environment for performing necessary operations during the deployment of an HPC cluster, helping reduce the time spent manually configuring each node in the cluster. Figure 1 outlines some DTK features and associated commands relevant to cluster deployment.

### Setting up the Dell OpenManage Deployment Toolkit environment

The DTK is available in two versions: a Microsoft® Windows® Preinstallation Environment (WinPE) self-extractable zip file and a Red Hat® Enterprise Linux 3 ISO image. Administrators can extract the WinPE tools and scripts obtained from the zip file and build them into a WinPE image. Administrators can also extract the tools and scripts from the Linux ISO image and use them in a customized ISO image, or use the image provided with the DTK. The information provided in this article refers to the Linux version of the DTK.

Administrators can set up the DTK environment in a cluster using the PXE mechanism available on the master node. The PXE environment allows servers (in this case, the compute nodes) to

boot over a network connection. As each compute node boots into the lightweight Linux environment provided by the DTK through PXE from the master node, it executes the automated scripts that perform the required operations on the system.

Installation prerequisites are as follows:

- DTK 2.x CD (created by downloading the DTK ISO image from support.dell.com and burning the image on a CD)
- Master node running Linux or a Linux-based OS
- Compute nodes connected to the master node and ready to install

### Setting up the DTK environment on the master node

A wide variety of cluster management software is available, and administrators can use the DTK to configure basic BIOS, BMC, DRAC, and RAID settings on different types of Dell Linux clusters. Administrators must configure the Trivial FTP (TFTP) server on the master node; the directory used to serve files is typically the /tftpboot directory. The DTK kernel file (SA.1) and ramdisk file (SA.2), available with the DTK download, must be copied to the master node /tftpboot directory.

After the DTK CD is mounted on the server, the files can be copied to the /tftpboot directory using the following commands:

```
mount /dev/cdrom /mnt/cdrom
cp /mnt/cdrom/isolinux/SA.1 /tftpboot/SA.1
cp /mnt/cdrom/isolinux/SA.2 /tftpboot/SA.2
umount /mnt/cdrom
```

Administrators must also create a folder on the Network File System (NFS) share of the master node. In the example scripts shown in this article, /export/home/install/dtk carries the default sample scripts provided with the DTK as well as any custom scripts written for specific server configurations.

Finally, administrators must place an entry for the DTK in the standard master node PXE configuration file for all the compute nodes to obtain the DTK kernel and ramdisk images over PXE. Figure 2 shows an example PXE configuration file. The IP address of the master node (for the private network of the cluster) is set to 10.1.1.1. Additional examples for different types of shares are available in the isolinux.cfg file provided with the DTK Linux ISO image.

### Installing the Dell Utility Partition

The Dell Utility Partition is a bootable partition on the hard drive that contains diagnostic and system utilities. The partition is typically between 32 MB and 60 MB. Installing the partition requires the latest Dell Utility Partition images, which are provided by Dell OpenManage Server Assistant (available on the Dell Installation and Server Management CD that ships with PowerEdge servers).

DTK option	Get example	Set example	Description
<b>Asset</b>	<code>syscfg --asset</code>	<code>syscfg --asset=node1</code>	Sets the asset tag value, which can only be assigned programmatically; Dell OpenManage Server Assistant and the DTK enable administrators to assign this information
<b>Boot sequence</b>	<code>syscfg --bootseq</code>	<code>syscfg --bootseq=3,2,1</code>	Sets the node boot sequence
<b>LCD</b>	<code>syscfg --lcd</code> <code>syscfg --lcd1</code> <code>syscfg --lcd2</code>	<code>syscfg --lcd=user</code> <code>syscfg --lcd1="Change HDD"</code>	Sets the PowerEdge server LCD to reflect information about the server
<b>Intel® Hyper-Threading Technology</b>	<code>syscfg --logicproc</code>	<code>syscfg --logicproc=enable</code>	Enables or disables the logical processor
<b>Power button</b>	<code>syscfg --powerbutton</code>	<code>syscfg --powerbutton=disable</code>	Enables or disables the node power button
<b>USB ports</b>	<code>syscfg --usbports</code>	<code>syscfg --usbports=disable</code>	Enables or disables USB ports
<b>Adjacent cache line prefetcher</b>	<code>syscfg --adjcacheprefetch</code>	<code>syscfg --adjcacheprefetch=enable</code>	Enables or disables cache line prefetch settings to help enhance system performance
<b>Hardware prefetcher</b>	<code>syscfg --hwprefetcher</code>	<code>syscfg --hwprefetcher=disable</code>	Enables or disables hardware prefetch settings to help enhance system performance
<b>LAN</b>	<code>syscfg lancfgparams</code>	<code>syscfg lcp</code> <code>--alertdest=1</code> <code>--destipaddr=192.168.0.101</code> <code>--nicselection=dedicated</code> <code>--dnsserver1=192.168.0.100</code> <code>--dnsrcname=rac-dell</code>	Configures and displays LAN configuration parameters
<b>LAN users</b>	<code>syscfg lanuseraccess</code>	<code>syscfg lua</code> <code>--userid=2</code> <code>--usrprivlmt=administrator</code>	Configures the privilege level and channel accessibility associated with a specified user ID for the LAN channel
<b>Platform Event Filter (PEF)</b>	<code>syscfg pefcfgparams</code>	<code>syscfg pcp</code> <code>--filter=powerfail</code> <code>--filteralert=enable</code>	Configures and displays PEF configuration parameters
<b>Demand-Based Switching (DBS)</b>	<code>syscfg --dbs</code>	<code>syscfg --dbs=enable</code>	Enables or disables DBS; can help minimize power and cooling requirements
<b>Network interface card (NIC)</b>	<code>syscfg --embnic1</code>	<code>syscfg --embnic1=on</code> <code>syscfg --embnic2=onnopxe</code>	Configures the embedded NIC ports to enable or disable PXE and turn NIC ports on or off
<b>Snoop filter</b>	<code>syscfg --snoopfilter</code>	<code>syscfg --snoopfilter=enable</code>	Enables or disables the snoop filter option on supported systems

Figure 1. System parameters configurable with the DTK

```
label dtk
kernel SA.1
append initrd=SA.2 ramdisk_size=36817 Stage3_type=cdrom DEBUG=0 quiet
ide=force_255_heads share_type=nfs share_location=10.1.1.1:/export/home/install/dtk share_script=dtk.sh
share_opts= nolock
```

Figure 2. Example PXE configuration file

```
svctag='syscfg --svctag | awk -F= '{print $2}''
sysname='syscfg --sysname | awk -F= '{print $2}' | cut -c11-14'
./upinit.sh --overwrite -disk=/dev/sda -size=32 -file=../../systems/UP/2800/upimg.bin
```

Figure 3. Sample dtk.sh script

```
syscfg --bootseq=cdrom.emb.0,floppy.emb.0,hdd.emb.0,nic.emb.1
syscfg --logicproc=disable
syscfg --snoopfilter=enable
syscfg --dbs=enable
syscfg --embnic1=on
syscfg --embnic2=onnopxe
```

Figure 4. Sample dtk.sh script snippet for BIOS configuration

```
syscfg lancfgparams --ipaddress 10.0.0.3 --subnetmask 255.0.0.0
syscfg lancfgparams --alertdest=1 --destipaddr=10.0.0.20
syscfg lancfgparams --nicselection=dedicated
```

Figure 5. Sample dtk.sh script snippet for BMC and DRAC configuration


The compute nodes must install the partition before installing the HPC cluster management software. Therefore, during installation, the preferred label in the PXE configuration file, `default`, under the master node `/tftpboot/pxelinux/pxelinux.cfg` folder is changed to `dtk`, causing the compute nodes to boot from the DTK image for the installation of the Dell Utility Partition. Administrators can make the compute nodes boot from the master node PXE setup by pressing F12 as the nodes start to boot.

The `dtk.sh` script, a sample of which is shown in Figure 3, is used to install the Dell Utility Partition; this script should be placed in the NFS share at `/export/home/install/dtk`. Once the Dell Utility Partition installation is complete on all compute nodes, the master node label is changed back to the original label, and the standard cluster installation can proceed normally.

### Capturing and modifying system configuration settings

Administrators can capture or modify BIOS, BMC, and DRAC settings using DTK commands such as those shown in Figure 1. To capture all system configuration settings accessible by the DTK, administrators can use the `syscap.sh` script, available at `/export/home/install/dtk/template/scripts`, as a default template. Figures 4 and 5 show example script snippets that administrators can incorporate into the main `dtk.sh` automated script file.

### Configuring cluster nodes with the Dell OpenManage Deployment Toolkit

The Dell OpenManage Deployment Toolkit can help administrators perform a wide range of server configuration tasks on large HPC clusters. Using DTK features and its preconfigured scripts in a pre-OS environment can help simplify the setup and configuration of critical server parameters in both homogenous and heterogeneous HPC clusters, with the ultimate goal of helping improve data center deployment, management, and administration, and eventually total cost of operation. 

**Shrankha Upadhyay** is an engineering analyst in the Scalable Systems Group at Dell. Her current interests include HPC cluster management, operating systems, and cluster file systems. She has a B.E. in Computer Science from Birla Institute of Technology and Science, Pilani.

**Meghana Bhat** is a senior engineer in the Deployment team in the Dell OpenManage Product Group. Her areas of interest include open source software and standards. She has worked on Dell OpenManage Server Assistant and is currently working on the Dell OpenManage Deployment Toolkit. Meghana has a B.E. in Computer Science from Vishwakarma Institute of Technology, Pune.

**Baris Guler** is an HPC application specialist in the Scalable Systems Group at Dell. His current research interests are parallel processing, diskless HPC clusters, performance benchmarking, reservoir engineering and simulation, and numerical methods. Baris has a B.S. in Petroleum and Natural Gas Engineering (PNGE) from the Middle East Technical University and an M.S. in PNGE from the Pennsylvania State University.

**Arun Rajan** is a systems engineer in the Scalable Systems Group at Dell. His current interests and responsibilities include HPC cluster management, cluster computing packages, performance benchmarking, and product development. He has a B.E. in Electronics and Communications Engineering from the National Institute of Technology, Tiruchirappalli, and an M.S. in Computer and Information Science from the Ohio State University.