

Using Red Hat Enterprise Linux AS to Achieve Highly Available, Load-Balanced Clusters

The Red Hat® Enterprise Linux® AS operating system integrates Cluster Manager and IP Load Balancing, features that improve cluster functionality. This article describes how these features can be combined with Dell™ PowerEdge™ servers and other components to achieve high availability and high performance in clusters.

BY DANNY TRINH

In response to growing demands for scalability, reliability, and serviceability, system administrators frequently must provide high-availability, high-performance clustering solutions for their existing networks. The Red Hat® Enterprise Linux® AS (formerly Red Hat Linux Advanced Server) operating system includes two types of integrated clustering functionality: Cluster Manager for high-availability clusters and IP Load Balancing for redundant, load-balancing clusters.¹ Administrators can combine these components with Dell™ PowerEdge™ servers, Dell PowerVault™ storage, Dell PowerConnect™ switches, Dell Fibre Channel switches and host bus adapters (HBAs), and high-speed network interface cards (NICs) to build clusters that meet the scalability and availability requirements of high-end, enterprise-class applications.

Using Cluster Manager to implement high-availability clusters

The minimum hardware requirements for a high-availability cluster of Dell PowerEdge servers running on Red Hat Enterprise Linux AS include:

- Two PowerEdge servers (acting as virtual servers), each with one NIC and one Fibre Channel HBA
- One Dell Fibre Channel array and one Fibre Channel switch
- One PowerConnect switch

Red Hat Enterprise Linux AS includes Cluster Manager and all other software and utilities required to implement high-availability clusters, such as the two-node failover cluster shown in Figure 1. High-availability

¹IP Load Balancing is often known by its project name, Piranha.

clusters primarily use shared storage; therefore, both nodes connect to a Fibre Channel storage array. Fibre Channel offers superior performance and reliability for high-availability computing. To achieve failover, administrators can configure nodes as active/active, in which both nodes host application services, or as active/passive, in which one node hosts application services and the other is a backup waiting for failover.

Configuration of active/passive and active/active failover modes

In active/passive mode, the sole function of the backup node is waiting to take over when the primary services node fails. Although not necessary for cluster operation, ideally all nodes will have identical hardware so that clients cannot detect a difference after failover.

In active/active mode, both nodes provide services to clients, but not the same services. For example, given two network services such as Network File System (NFS) and Server Message Block (SMB) file sharing, node 1 might provide NFS and node 2 might provide SMB. If node 1 fails, node 2 takes over NFS file sharing while continuing to provide SMB. In active/active mode, services experience a performance drop when one node fails because the other node must perform twice the work.

Prevention of false failover

Red Hat Cluster Manager uses quorum partitions on shared storage as the primary mechanism for recording the state of cluster nodes.

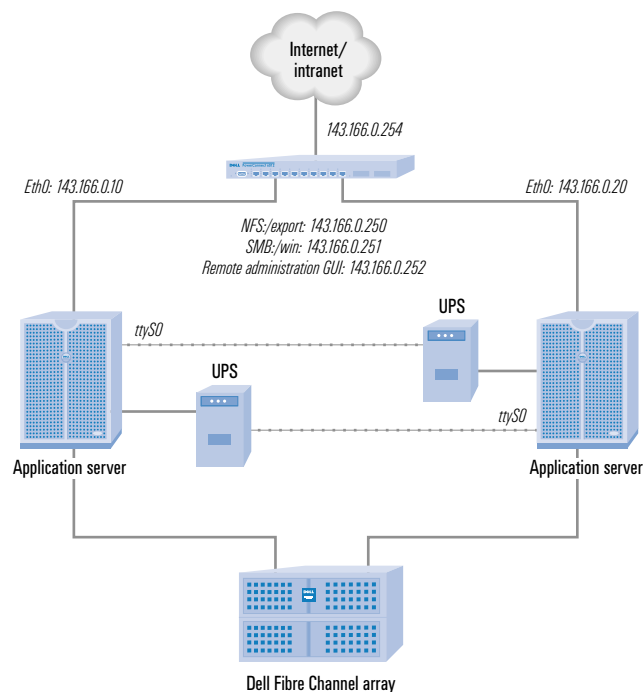


Figure 1. Two-node high-availability computing cluster architecture

Red Hat Cluster Manager uses quorum partitions on shared storage as the primary mechanism for recording the state of cluster nodes.

Each node periodically sends update information to quorum partitions that tells the other node it is still up and running.

Additionally, each node sends a *heartbeat*—a periodic signal that indicates it is still running. Cluster Manager uses the heartbeat as a delay mechanism to prevent false failover. For example, if a cluster node experiences kernel panic, the node cannot send update

information or heartbeat signals—a legitimate case for failover. However, if a cluster node is so busy with I/O-intensive tasks that it cannot send update information as normal, Cluster Manager uses the heartbeat signal—which the busy node can still send—to determine whether failover should wait for a few more seconds. These few seconds can make the difference between false failover and continuous service.

Cluster Manager provides two additional methods for preventing false failover and safeguarding data integrity: Shoot The Other Node In The Head (STONITH) and the watchdog timer. Although very unlikely, a failed node may continue to write data even though it no longer sends a heartbeat signal. Only one node can mount a local file system read-write at a given time; therefore, the working node must shut down the failed node to prevent any data corruption. STONITH allows a node to reset the power on a malfunctioning node, forcing it to reboot by turning off the errant node’s uninterruptible power supply (UPS). Turning off the UPS prevents the dead node from sending update information or a heartbeat signal.

The watchdog kernel module (softdog.o) can be loaded when configuring Cluster Manager. With this module loaded, the system attempts to write to /dev/watchdog at certain polling intervals. If the system fails to do so, the kernel sends a power cycle message signal to the BIOS. The server then reboots itself when an administrator presses the reset power button.

Using IP Load Balancing to implement redundant, load-balancing clusters

The minimum hardware requirements for a load-balancing cluster of Dell PowerEdge servers running on the Red Hat Enterprise Linux AS operating system include:

- Two PowerEdge servers, each with two NICs, to act as active and backup routers
- Three PowerEdge servers, each with one NIC, to act as application servers
- Two PowerConnect switches

To implement IP load-balancing clusters, no additional software is needed; Red Hat Enterprise Linux AS has all necessary software and utilities, such as IP Load Balancing. This utility is an enhancement of the Linux Virtual Server (LVS), which uses the Network Address Translation (NAT) routing mechanism to deliver high availability and scalability for applications and services. Direct routing and IP encapsulation (tunneling) are alternative routing mechanisms for LVS, but Red Hat Linux does not support them.

A cluster configured for IP Load Balancing has two layers (see Figure 2). The first layer, which provides high availability, contains an active and a backup router. These routers are connected to two separate networks: public and private. The active router serves as a NAT router. All network traffic to and from both Internet and intranet users must pass through the active router on its way to and from the application servers (also known as real servers). Clients contact the IP Load Balancing cluster using the public virtual IP address (in Figure 2, 143.166.0.1) to request network services. At any one time, only one of the load-balancing routers is active, and it has both virtual IP addresses assigned to it. When the active router fails, the backup router detects this, takes over the virtual IP

A cluster running on Red Hat Enterprise Linux AS that combines Cluster Manager and IP Load Balancing functionality can provide excellent network services to clients.

addresses, automatically promotes itself to NAT router, and becomes the active router.

Administrators can specify a scheduling algorithm for the load-balancing router to divide the network load among application servers. By default, IP Load Balancing uses a weighted least-connection scheduling algorithm. However, administrators can choose among many scheduling algorithms for one that best fits their specific network requirements. Both load-balancing routers also have unique IP

addresses assigned to their public and private interfaces so they can monitor each other's health.

The second layer in this configuration contains from 2 to 32 application servers connected to a private network. These servers provide such functions as Web and FTP serving. The application servers use one private virtual IP address—192.168.0.254 in the example shown in Figure 2—as their default route to send responses back to the clients; thus, the cluster appears as one server. Each application server must have a unique private IP address.

Utilities to set up IP Load Balancing

The Piranha Configuration Tool and the `iptables` command are the two main utilities needed to set up IP Load Balancing. Key software components of the Piranha Configuration Tool include:

- **pulse:** This controlling process runs on the load-balancing routers, monitors the network heartbeat, and controls health monitoring and router failover.
- **lvs:** This daemon runs on the active router and calls the `ipvsadm` service to keep the LVS redirection table up-to-date.
- **nanny:** This monitoring daemon runs on the active router to monitor health for each application server and virtual service.
- **piranha-gui:** This service starts a Web-based graphical user interface (GUI) for generating LVS configurations and monitoring the state of the IP Load Balancing cluster.

The `iptables` command sets firewall marks on packets for any of the multiports that are required by network services. For example, HTTP and HTTPS must be the same on the application server to which a client connects.

IP Load Balancing can be used for many network services that do not change data frequently, such as static Web sites, read-only

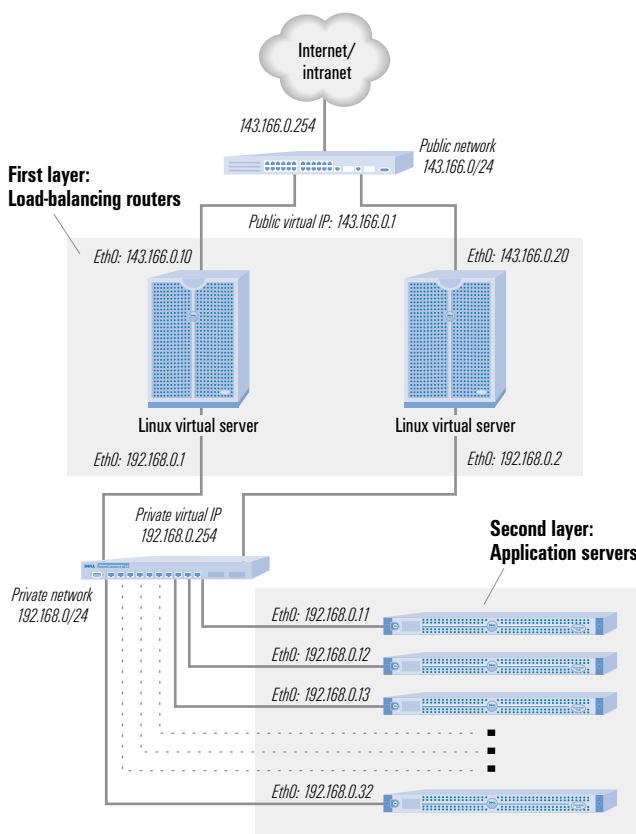


Figure 2. IP Load Balancing cluster architecture

FTP servers, or video-streaming servers. Because each application server has its own storage, all data on one application server must be the same as that on the other application servers. Like Linux, IP Load Balancing does not require specialized hardware; all Dell PowerEdge servers are excellent choices for implementing this functionality.

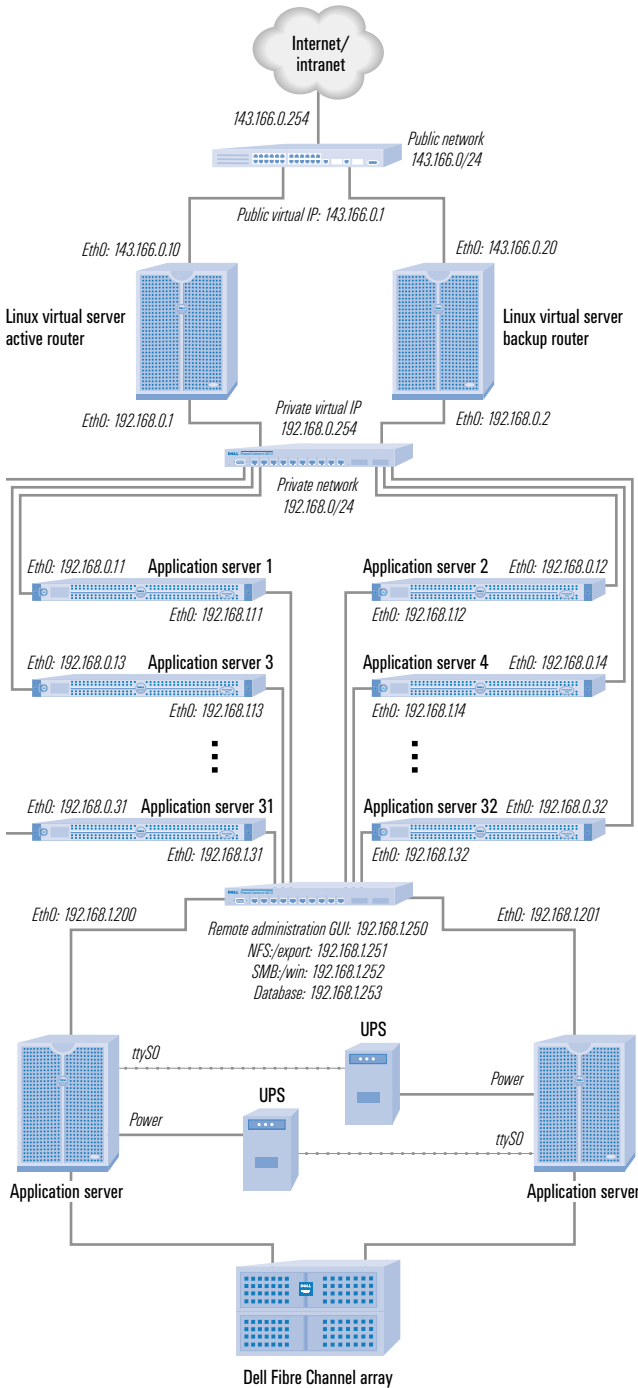


Figure 3. High-availability computing cluster with IP Load Balancing and Cluster Manager

Achieving highly available, scalable environments

High-availability clusters can help decrease downtime, save money, and increase productivity. Load-balancing clusters distribute incoming IP network requests across multiple servers, helping to increase performance, scalability, and fault tolerance of the network. Integrating both approaches into one cluster reaps the advantages of both high availability and load balancing. A cluster running on Red Hat Enterprise Linux AS that combines Cluster Manager and IP Load Balancing functionality can provide excellent network services to clients. Figure 3 shows an example configuration. To system administrators, such a configuration—providing a single place for network services—is easy to maintain. To users, this cluster configuration appears as a single virtual server.

By combining Red Hat Enterprise Linux AS, Dell PowerEdge servers, Dell PowerVault storage, Dell PowerConnect switches, and high-speed NICs, IT organizations can achieve highly available, scalable environments for services such as Domain Name System (DNS), mail, proxy, NFS share, SMB share, and Dynamic Host Configuration Protocol (DHCP). These components also can provide the foundation for clusters running applications that require scalability and high availability, such as e-commerce Web sites, Web farms, and application centers.

Danny Trinh (danny_trinh@dell.com) is a senior analyst in the Linux Development Group at Dell. He tests and certifies all Dell PowerEdge servers and peripherals for compatibility with Red Hat Linux. Danny has an associate degree and is a Certified Novell Engineer® (CNE®), Microsoft® Certified Systems Engineer (MCSE), and Red Hat Certified Engineer® (RHCE®).

Load-balancing clusters distribute incoming IP network requests across multiple servers, increasing performance, scalability, and fault tolerance of the network.

FOR MORE INFORMATION

iptables: <http://www.redhat.com/docs/manuals/linux/RHL-9-Manual/ref-guide/ch-iptables.html>

Linux Virtual Server Project: <http://www.linuxvirtualserver.org>

Piranha Configuration Tool: <http://www.redhat.com/docs/manuals/enterprise/RHEL-AS-2.1-Manual/install-guide/ch-lvs-piranha.html>

Red Hat Cluster Manager Installation and Administration Guide: <http://www.redhat.com/docs/manuals/enterprise/RHEL-AS-2.1-Manual/cluster-manager>

Red Hat Enterprise Linux AS: <http://www.redhat.com/software/rhel/as>