Enterprise Sustainability for High Performance Linux Clusters

By Justin Nemmers, RHCE

Abstract

Linux HPC clusters are here to stay. Increases in performance and reliability will ensure that the next generation of supercomputing clusters will be larger and more powerful, while continuing to solve once-intractable problems. This paper examines the advantages of deploying enterprise-level technologies, how Red Hat services can help in all stages of an open source deployment, and why a migration to Red Hat Enterprise Linux might be easier than you think.

November 2004

Table of Contents

Introduction

It was once unthinkable to try and solve intractable scientific and business problems without purchasing or leasing expensive computing systems. Traditional supercomputers--monolithic symmetric multiprocessor (SMP) machines using customized processors or vector-processors--had price tags that matched their size and abilities.

Enter the notion of cluster-based high-performance computing (HPC). By combining relatively inexpensive commodity hardware with various software and hardware interconnects, it is possible to create powerful systems at a fraction of the cost of larger SMP systems. Significant demand for commodity hardware and stiff price competition among OEM vendors made clustered systems a compelling approach for the HPC industry. Additionally, the advent of more affordable, industry-standard 64-bit computing makes it easy to create extremely powerful clustered systems that not only run toe-to-toe with the more expensive SMP and vector machines, but often beat them in almost every common supercomputing benchmark.

Clustered systems are gaining acceptance for both business and scientific use.[1] As clusters move from the workbench to the machine room floor and from development and quality assurance into 24x7 production environments, a proven, vendor-supported enterprise Linux operating system (OS) is increasingly important.

In addition to cluster-based HPC, an enterprise Linux OS should also support larger, commodity-processor-based SMP machines since there are applications that run better on those systems.

The enterprise Linux OS should also be scalable. Independent Software Vendors (ISVs) and hardware manufacturers require an operating system to tie together large numbers of computers so that they function as one entity.

For users in search of these qualities, Red Hat Enterprise Linux is accepted in the HPC industry, powering such clusters as Lawrence Livermore National Laboratory's Thunder (http://www.llnl.gov/linux/thunder/thunder_configuration.pdf). Thunder placed in the top 5 of the TOP500.

At Red Hat, an enterprise Linux OS is just the beginning. HPC customers also realized increased managability with Red Hat Network, improved scalability with Red Hat Global File System, and gained expertise with enterprise support services.

1 Clustered systems are well-represented on the TOP500 (http://www.top500.org), the industry's list of most powerful computers.

Red Hat Enterprise Linux and HPC Clusters

By combining affordability and the rapid innovation rate of the open source community with reliability and the accountability of a major commercial software company, Red Hat helps the high performance community achieve a sustained, standards-based, commodity clustering environment.

Certified Hardware

Red Hat Enterprise Linux supports a large breadth of hardware, providing flexibility in choice of architecture. A list of all hardware currently certified for Red Hat Enterprise Linux can be found at http://hardware.redhat.com/hcl/. This list continues to grow as new and updated drivers are made available through supported updates from Red Hat. Red Hat Enterprise Linux's ability to run on so many different hardware platforms with full Red Hat support provides a level of choice not often available with other Linux platforms.

In comparison to older versions of Red Hat Linux, Red Hat Enterprise Linux includes native driver support for newer motherboard chipsets, IDE, SATA, and SCSI disk controllers, network cards, and Fibre Channel (FC) HBAs. Newer drivers often provide performance and stability increases compared to older Red Hat Linux distributions. In order for a piece of hardware to be certified, Red Hat installs and tests it under Red Hat Enterprise Linux to ensure compatibility. Certified hardware platforms currently in use for HPC include Dell PowerEdge 1750, HP Integrity rx2600, and SGI Altix--all hardware platforms recognized by the TOP500. Additionally, new products such as IBM Cluster 1350--which offers a rack-optimized fully-integrated solution--are also fully certified to work with Red Hat Enterprise Linux.

Hardware Platform Support

Though many current clusters are comprised of 32-bit x86 hardware, 64-bit x86 computing is rapidly gaining industry traction. A unique feature of x86 64-bit computing is that it allows older 32-bit code to run in parallel with 64-bit code. This is particularly important when facing porting and recoding issues. A migration to full 64-bit computing can be done gradually as opposed to an all-or-nothing approach.

Many of the clusters in the top 25 of the TOP500 are comprised of Itanium or Opteron processors--which Red Hat Enterprise Linux natively supports. To gain even more performance, some HPC clusters use more specialized hardware such as IBM's POWER series, zSeries, and Intel Itanium, also natively supported by Red Hat Enterprise Linux.

Despite numerous hardware differences between supported architectures, Red Hat Enterprise Linux is compiled for each one

based on a common source tree. Deploying a common OS across multiple architectures eases training, administration, code porting, and serves to further commoditize computing hardware. Several examples of non-commodity TOP500 supercomputers running Linux include IBM's Blue Gene series and NASA's recently announced SGI Altix-based Columbia. Red Hat works closely with partners such as NEC, SGI, and Fujitsu to ensure that Red Hat Enterprise Linux is properly tuned to work well on larger commodity SMP systems. Even more architectures will be natively supported in the next major release of Red Hat Enterprise Linux, along with better support for Non Uniform Memory Access (NUMA) interconnects found in SGI products. These features are currently available in the Fedora Project.

Product Choices

Red Hat Enterprise Linux is available in a variety of configurations:

- Red Hat Enterprise Linux WS is the ideal platform for compute nodes. WS is fully supported across various architectures, and up to two CPUs per node.
- Red Hat Enterprise Linux ES is ideal for edge-of-network workloads like web, file, and print serving.
- Red Hat Enterprise Linux AS is intended for use in systems with large amounts of memory or multiple CPUs, such as a head node or file server. AS is available with 24x7 support, ensuring that the most critical systems are always available.

It is important to note that AS, ES and WS are all based on Red Hat Package Manager (RPM) system. This means an application developed on a WS system requires no modifications when deployed on an ES or AS system and vice versa. In terms of libraries and other base packages, AS, ES and WS are identical.

Stability Across Patches

Red Hat ensures that the various Application Programming Interfaces (API) and Application Binary Interfaces (ABI) remain unchanged between minor kernel revisions and updates (i.e. Red Hat Enterprise Linux v.3 U1 to U2). This reduces the likelihood that a kernel or a C-library update will break application code. With regard to system libraries, this stability across revisions helps ensure that a scientific application will continue to provide consistent answers without the need for recompilation. Additionally, each major Red Hat Enterprise Linux release is supported by Red Hat for a full seven years from the date of general availability. Bug fixes, security patches, and most importantly, a clear upgrade path are, for the first time, available for Linux in cluster-based HPC.

Stable and Reliable

Red Hat Enterprise Linux is a proven stable platform. The Red

Hat Enterprise Linux 2.4.21-based kernel contains several important back-ported features from the maturing 2.6 kernel tree. These include the Native POSIX Thread Library (NPTL), which allows for easier porting to Linux, and higher-performing, multi-threaded applications. 4GB-4GB memory support gives processes direct access to larger portions of memory, allowing for larger datasets. The maximum amount of addressable system memory has also been increased to 64 GB on 32-bit x86 systems.

Red Hat chose to back-port these updates into the 2.4.21 Linux kernel because it is a proven, mature, extremely stable kernel. By back-porting some of the best features of the 2.6 kernel, the 2.4.21 kernel was strengthened and the 2.6 kernel has more time to mature before being available for enterprise computing in Red Hat Enterprise Linux.[2]

Software Support
ISV support has never been more important in HPC. Without products like high-performance C and FORTRAN compilers (available from vendors such as Portland Group and Intel), it would be difficult to squeeze more power out of existing hardware. Additionally, various parallel computing debuggers help developers code for the unique environment HPC clusters present. Many of these developer products are supported on Red Hat Enterprise Linux. Red Hat Ready ISVs certify their software on Red Hat Enterprise Linux, and a list of vendors who have provided compatibility information can be found in the Red Hat Ready ISV Catalog (http://www.redhat.com/apps/isv_catalog/).

Enterprise Support Services

Support is extremely important in any enterprise computing environment. Red Hat provides industry-leading support and professional services for all stages of open source deployment.

Technical Account Management
A Red Hat Global Support Technical Account Manager (TAM) allows for a deep technical relationship with Red Hat. A TAM is your advocate within Red Hat engineering and the Open Source community and is the best way to synchronize development plans with future Red Hat releases or gain early access to beta versions of subscribed software channels. https://www.redhat.com/support/offerings/technical.html

2 It is important to note that Red Hat did not unilaterally make changes to the 2.4.21 kernel tree. Once Red Hat developed and tested the code changes, back-ports were submitted to the managers of the 2.4 kernel tree for acceptance by the open source community. Red Hat maintains the strong position that mainstream kernel acceptance is required for long-term sustainability and compatibility with future versions. Companies who create separate code branches jeopardize sustainability for their customers since the company is solely responsible for producing fixes and new releases without the benefit of the open source community.

Developer Support
Developer support provides development advice and tools assistance for customers who are developing or porting an application to Red Hat Enterprise Linux. Red Hat employs a large number of GCC developers who can be directly tapped in order to help optimize a particular piece of code for maximum performance.
www.redhat.com/support/offerings/developer.html

Consulting
If consulting assistance is required, Red Hat Professional Services provide a range of high caliber technical services to accelerate and enhance Linux and Open Source deployment.
www.redhat.com/services/consulting/

Training
While direct Red Hat support is certainly valuable, training administrators to effectively plan, implement and manage Red Hat Enterprise Linux systems can potentially solve a plethora of issues before they arise. Red Hat Global Learning Services are industry-leading and award winning certification programs. Red Hat Certified Technician (RHCT), Red Hat Certified Engineer (RHCE), and Red Hat Certified Architect (RHCA) are performance-based training programs for IT professionals including administrators, architects, security specialists and developers.
www.redhat.com/training/

Management: A Solvable Problem

As HPC clusters get larger, management of the hardware, software, and OS image on each node gets more complex. Both open source and commercial software/hardware solutions have emerged to help address this issue. Open source projects such as Ganglia (ganglia.sourceforge.net/) and Platform Rocks (www.platform.com/products/Rocks/) simplify cluster installation, management and parallel command execution. Without these management tools, administering large clusters would be a much more difficult task. Aside from health monitoring and command execution, there are often configuration files and OS image issues that require administrative overhead to manage. Red Hat Network (RHN) Satellite offers a user-friendly interface that works as well for ten systems as it does for ten thousand.

Red Hat Network Satellite is a powerful systems management platform for Red Hat-based clusters. It can simplify overall node image management, allow for bare-metal provisioning, and automate routine remote management. It can schedule tasks on all nodes, or on any subset of the cluster nodes. Red Hat Network Satellite allows for the creation of custom software and

configuration channels. Custom channels coupled with Red Hat's up2date and rhn_check utilities, ensure that cluster nodes have the right version of the software and the correct configuration files. The creation of custom software channels lets administrators use RHN Satellite to deploy and manage much more than just Red Hat Enterprise Linux. Specific software channels can be created for particular applications, or even entire distributions, such as Red Hat Linux 7.3. Red Hat Network Satellite can also efficiently deploy and update Red Hat products as Red Hat Global File System (GFS). Because Red Hat Network is highly flexible, one installation of Red Hat Network Satellite can be used to manage more than one environment; a single instance could conceivably service a cluster and other non-cluster-related Red Hat Enterprise Linux installations simultaneously, reducing administrative overhead and training requirements.

A Scalable Solution

Scalability is obviously a key factor in cluster-based HPC. With large clusters often come large network I/O, file system I/O, and data storage requirements. Production clusters require data to be redundantly stored, frequently backed up, and highly available. Designing such a file system architecture is a problem unique to HPC, as even the most demanding business applications rarely require thousands of nodes accessing the same physical file system. Ideally, each cluster node would be attached to a Storage Area Network (SAN) via FC. This is obviously cost prohibitive, as FC ports on a SAN are many times more expensive than the equal bandwidth on an interconnect such as Gigabit Ethernet. Various technologies have been leveraged in order to provide the file system performance that is demanded by massive clusters.

Clustered, scalable parallel-access file systems are the answer. These are built by attaching a number of systems to a pool of storage, typically a SAN, allowing each file server full accessibility. Each one of these file servers can then make the data available to the rest of the cluster by some other means--perhaps NFS or an iSCSI variant. Red Hat acquired the Global File System (GFS) from Sistina. Earlier this year Red Hat made GFS available to the Open Source Community under the GPL, and has continued to mature and refine its code.

GFS is a POSIX-compliant linear scalable parallel file system. This means that, as file system I/O requirements increase, SAN-attached GFS nodes can be added to the file server pool in order to increase bandwidth. File system size under the 2.4 kernel is currently limited to between 1 and 2 TeraBytes, but with the next major release of Red Hat Enterprise Linux, file system size limits will drastically increase: 16 TeraBytes on 32-bit hardware, and

up to 8 ExaBytes when using a 64-bit version of Red Hat Enterprise Linux on the appropriate hardware.

Migrating to Red Hat Enterprise Linux

This paper has presented compelling reasons to select Red Hat Enterprise Linux as a platform in a new HPC cluster, but what about upgrading an existing cluster from some other version of Linux to Red Hat Enterprise Linux?

While no major migration is simple, moving a running application from a version of Red Hat Linux to Red Hat Enterprise Linux is made easier by toolsets and compatibility libraries specifically provided by Red Hat to ensure applications compiled on a Red Hat Linux 7.3 or 9 system will run on a Red Hat Enterprise Linux v.3 system. Often, the very fact that the application runs on a newer Red Hat Enterprise Linux system means that code porting and recompilation efforts will be much easier than backporting bug and security fixes into the older Red Hat Linux versions. In terms of library compatibility, Red Hat Enterprise Linux v.3 uses GCC 3.2 and glibc 2.3.2. Although newer versions of libc are mainly backwards-compatible, there are still some applications which require older versions, such as GCC 2.96. For this reason, Red Hat provides five important packages: compat-libstdc++-devel, compat-glib, compat-gcc-c++, compat-gcc, and compat-libstdc++. These five packages go a long way to ensure that legacy applications will work under Red Hat Enterprise Linux v.3. Once the migration to Red Hat Enterprise Linux is complete, there will always be a supported upgrade and migration path to future releases of Red Hat Enterprise Linux. Stability of the platform is key in making Red Hat Enterprise Linux a better choice over other distributions. Reliable, predictable, and heavily tested changes ensure future upgrades and patch installations will not require large amounts of testing.

There are certain instances where upgrading is deemed important. As the patch streams have ended for older versions of Red Hat Linux, systems administrators have been forced to decide either to further restrict local shell access to their systems or manually build and update installed packages from community-provided source code. Although either is certainly a reasonable option, it requires a significant amount of the administrator's time to monitor the various vulnerability and security-related websites available, and then address potential security flaws and bugs as they arise. The Red Hat Enterprise Linux patch stream delivered via Red Hat Network solves these issues.

An additional advantage of Red Hat Enterprise Linux is that it is a familiar and flexible platform. If systems are already running a version of Red Hat Linux, system administrators have likely

added and removed packages as needed to create the smallest and most efficient node image possible. This helps in the imaging of newly added nodes, as well as the resurrection of dead nodes because it speeds up the process of installing the OS image onto the node. Red Hat Enterprise Linux excels in its flexibility. Starting with a minimal Red Hat Enterprise Linux WS install, a powerful, compatible, simple and small node image can be created quickly.

Conclusion

Linux HPC clusters are here to stay. The continued increases in performance and reliability ensure that the next generation of supercomputing clusters will be larger and more powerful, continuing to solve once-intractable problems. Red Hat is dedicated to the HPC community, and with software such as GFS, Red Hat Network Satellite Server, and Red Hat Enterprise Linux, the job of building and supporting these massive compute farms is only going to get easier. Red Hat is confident that Red Hat Enterprise Linux is currently the best platform for HPC clusters, and that its product line and services will continue to exhibit a strong commitment to the HPC market.