IBM @server

Linux Clusters
White Paper

July 6, 2004
Norm Snyder
ndsnyder@us.ibm.com
Contents:

Introduction..................................................................................................................3

Why IBM is Doing Linux Clusters.............................................................................4

Linux Cluster Offerings from IBM..............................................................................5
    Hardware
    Software
    System Design and Specification
    Diagram of a Large Linux Cluster

IBM @server Cluster 1350 Overview............................................................................12
    Hardware
    Software
    Future Generations

Advantages of Linux Clusters from IBM.................................................................14
    IBM Solution Differentiators
    Systems Management
    RAS of IBM Components
    IBM’s Packaging Advantage
    Service Processor Network

Conclusion...................................................................................................................16
Introduction

Clustered computing has been with us for several years. It represents an attempt to solve larger problems, or to solve problems in a more cost effective manner, than the more conventional systems of the time.

Greg Pfister, in his wonderful book In Search of Clusters, defines a cluster as “a type of parallel or distributed system that: consists of a collection of interconnected whole computers, and is used as a single, unified computing resource.” Clusters have been devised, formally or informally, from many types of systems. On one extreme, the IBM® zSeries®, when configured in a Sysplex, represents a cluster. One could argue that zSeries with z/VM™ and tens or hundreds of Linux® images might be considered a cluster as well.

The IBM RS/6000® SP™, which has evolved into the IBM® pSeries® 655 may be viewed as a cluster of RS/6000 servers (or RS/6000 server technology) packaged in a proprietary frame or rack. Both of these examples include highly functional software and hardware interconnects that allows the cluster to be viewed and administered as a single system. I would assert that any cluster of interest should have this characteristic as well.

At the other end of the power spectrum, many organizations have assembled Intel® (and other) processor-based servers into clusters of various types. Proprietary solutions such as Microsoft® Wolfpack and HP/Compaq Alpha TruCluster compete for mind share with the generic Beowulf clusters on Linux Intel Architecture boxes. Beowulf, in fact, was the outgrowth of NASA researcher Donald Becker’s solution to the problem of creating a supercomputing resource without having a supercomputing budget.

It is the latter type of cluster that will be addressed in this paper.

One may also characterize clusters by their function:

- **High Availability (HA):** Redundancy and failover for fault tolerance.
- **High Performance:** Lots of systems working together on a single problem - a FLOP farm.
- **Server Consolidation:** Central management of resources dedicated to disparate tasks.

Initial efforts in Linux clustering have been in the high performance computing (HPC) area, and the current Intel Xeon™ high clock rate servers (with as much as twice the floating-point capability of the Pentium® 3 predecessor) will continue to make this area vital. High-availability solutions for Linux clusters are now available from several Independent Software Vendors (ISVs), and span a range of capability and complexity. HA clusters are expected to become common in the near future. Server consolidation should solve a management
problem, not a technical problem. It should no doubt become important as Linux clusters become mainstream.

One final disclaimer: IBM does not distribute Linux. Linux clusters described in this paper are comprised of IBM hardware and middleware, using standard Linux distributions from non-IBM sources as well as Open Source software.

为什么 IBM 要做 Linux 集群

IBM 已经以一种让许多行业观者惊讶的热情接受了 Linux。虽然一个真正的 Linux 福音传教士可能会认为这只是对不可避免的事情的承认，但也有人可能会认为这里面有更多的东西，也许认为这里面有比它更重要的。

快速浏览一下操作系统市场份额和增长的统计数据，就会让人相信 Linux 运动是真实的。Linux 代表了一个重要的商业机会，而 IBM 对这次市场机会的反应是 2001 年估计投入超过 10 亿美元，之后也继续投入大量资金。

计算架构已经经历了许多重要的范式转移。CISC 到 RISC（复杂指令集计算：大量的指令，大多数指令在多个机器周期内执行。精简指令集计算：更少的指令，每条指令在单个机器周期内执行）的一个例子。超级计算从矢量架构转向并行架构的转移是另一个。许多其他例子可以举例但与当前问题不太相关。

超级计算，或 HPC，往往是范式转移开始的地方。这是因为 HPC 市场细分有三个关键特征：

1. 性能，而不是可靠性，是最重要的。如果你可以走得足够快，偶尔的故障是可以接受并可以管理的，只需重新启动问题。
2. 价格/性能是第二个需求。HPC 通常由大型政府或研究机构完成。虽然预算很大，但预算监督也很重要。
3. 文化接受新颖的架构和解决方案。"流血边缘"先于"前沿技术"，而这个群体往往是早期采用者。

Linux 集群（基于 Intel 架构的处理器）在学术界变得非常流行。它们是计算机科学问题，包括并行编程、并行工具的开发以及分布式系统的管理，廉价的平台。同时，Intel（或 Intel 架构）处理器为基础的 Linux 集群也开始出现在政府和工业研究安装中，最初作为概念系统，现在作为生产性机器用于科学研究。
Today’s traditional HPC installation is running a large parallel RISC-based UNIX® system. In fact, the RS/6000 SP is still prominent on the TOP500 list (http://www.top500.org/) of the world’s fastest supercomputers. Parallel programming is achieved primarily by using an industry standard message passing programming model, called MPI (Message Passing Interface). Another alternative, PVM (Parallel Virtual Machine), is in use by some clients but has largely been superseded by MPI. The UNIX is standard, but includes vendor extensions (AIX 5L™, for IBM systems), and the systems management software layer is proprietary - Parallel Systems Support Program (PSSP) for AIX®.

Linux clusters represent a better price/performance opportunity that takes advantage of a familiar architecture and programming model. Linux is a well understood UNIX variant that, because of its Open Source connection, will probably not be diverged into vendor-specific versions. An Open Source version of MPI, called MPICH, is available for download and correctly implements the standard. Parallel programs using the MPI model can therefore run on Linux clusters. Tools and math libraries are available in the Open Source community, and middleware, like the IBM General Parallel File System (GPFS) and DB2® have Linux versions.

IBM’s interest in Linux clusters on Intel Architecture platforms is simple. Our clients are interested in Linux clusters on Intel Architecture processors. IBM brings skills and knowledge that no other vendor can bring to the table. IBM is uniquely qualified to design and install large Linux super clusters, with the entire infrastructure that is needed to make them systems, as opposed to collections of independent parts.

As did parallel RISC systems in HPC, Linux clusters are moving from the early adopter stage into the mainstream stage. This should be enabled by technologies such as high-performance file systems, high-availability software, and the like. As the Linux cluster systems become mainstream, the business opportunities should increase as well.

**Linux Cluster Offerings from IBM**

Beginning in 2000, IBM created Linux operating system-based cluster systems from IBM®xSeries® (or the older IBM Netfinity®) rack-mounted systems, integrating them with appropriate networks, a systems management layer (hardware and software), and necessary services. These initial offerings consisted of custom configured hardware and software to meet the clients’ needs, coupled with appropriate services to do a custom installation, and necessary support. They were sold as special bid systems, as opposed to formal products.

The IBM®@server Cluster 1300, announced in 2001, represented a formalization or productization of the previous custom-built cluster technology. Like the custom offerings, Pentium 3 technology was used. Custom offerings addressed configuration requirements that were not part of the Cluster 1300 line, such as graphics capability on the compute nodes. In October 2002, IBM announced the IBM®@server Cluster 1350. Similar in architecture to the Cluster 1300, the Cluster 1350 initially used the Intel Pentium 4 (Xeon) processor. Minor changes in the product occurred as the Intel products evolved; faster clock speeds, for
example. In May and November of 2003, the Cluster 1350 added new node types, giving more configuration options. While the configuration options for the Cluster 1350 are much richer than the Cluster 1300, specialized requirements may still be met with custom configurations.

IBM offers installation and support services for Linux clusters through IBM Global Services (IGS). IBM Business Partners may offer similar installation services. All solutions discussed represent complex integration into most clients' network environments, and an installation contract is recommended.

The basic components of a Linux cluster, either a Cluster 1350 or a custom offering, are as follows:

**Hardware:** Rack-mounted Intel Architecture systems are the basis of the Linux cluster. Dense packaging is a requirement, with standard 19" racks being the favored package. Racks are available in 42U, 25U, and 11U sizes. Within the racks are found nodes, fast interconnects such as switches or fabrics, network management hardware, terminal servers and the like.

Nodes may be functionally grouped into two categories: (1) Compute nodes perform the computational problem for which the system is designed; (2) Infrastructure nodes, such as head nodes, management nodes and storage nodes, provide systems management and specialized functions needed to make the collection of compute nodes into a system.

Compute nodes need to be as dense as possible, and have few connectivity requirements. The inclusion of a service processor for systems management functions (to be described later) is essential. The IBM @server xSeries 335 was the original compute node for the Cluster 1350, and in its present version remains a popular choice today. It allows one or two Intel Xeon processors with robust (fast and large) memory and on-board disk, in a “1U” form factor. A “U” or EIA Unit is 1.75 inches of height in a standard 19-inch rack package. The x335 has a built-in service processor and two slots for connectivity to the other components of the system. It comes standard with dual Gigabit Ethernet adapters.

With the May 2003 announcement, IBM @server BladeCenter™ HS20 blade may be included in the Cluster 1350 as well, offering very dense packaging for compute nodes. A BladeCenter chassis occupies 7U of rack space, and can hold up to 14 blades. Each blade is a complete system, with 1 or 2 Xeon processors, memory, IDE disk, etc. I/O connectivity, like Ethernet and Fibre, are provided by the chassis, as is power. Like the x335 nodes, each blade runs a copy of the Linux operating system, and is managed by the IBM systems management software.

With the November 2003 announcement, IBM added a compute node based on the AMD Opteron™ chipset. Called the IBM @server 325 (e325), it is similar in packaging to the x335. It is 1U high (1.75 inches) and contains one or two Opteron processors. It also contains
a memory subsystem, disk, a service processor, slots, built-in Gigabit Ethernet ports, all of the normal IBM reliability features, etc.

Clients may choose Intel processor-based x335, Intel processor-based BladeCenter HS20, or AMD Opteron processor-based e325 as compute nodes. Since configurations are customized to the client’s needs, a particular system may have any or all of the three types of compute nodes described.

Head nodes, management nodes, and storage nodes provide special function for management of the cluster (like boot support, hardware management, external I/O, etc.). The IBM xSeries 345, a 2U Xeon processor-based server, is used as the management node in the Cluster 1350, and may also be used as a storage or compute node.

The x345 nodes are optimal for storage nodes, using simple file systems, GPFS or databases. Fibre storage may be attached within the rack, and other Storage Area Network (SAN) or Network Attached Storage (NAS) devices might be chosen.

If a SAN is present, a cluster file system such as Matrix from PolyServe might be chosen. PolyServe is an IBM Business Partner whose Matrix clustered file system provides a symmetric scalable high-performance solution that can be shared by all applications in the cluster or the network.

A High Availability (HA) cluster might include two or more x345 storage nodes with HA software from SteelEye. SteelEye is an IBM Business Partner whose HA software is sold both by SteelEye and by IBM.

Switches or other fabrics are used for interprocessor communication for parallel programming, and for various management functions.

For parallel programming InterProcess Communication (IPC), a commonly used switch is the Myrinet™-2000 switch from Myricom, Inc. The Myrinet switch is a fast, scalable high bandwidth switch. Like all true scalable switches, as the number of nodes attached to the switch increases, the aggregate bandwidth goes up proportionately, and the latency remains constant. Said another way, the bandwidth on each path is the same and the number of paths depends on the number of nodes, with each node having a path to all other nodes regardless of the cluster size. Traditional per path bandwidth is approximately 250 MB/sec. each direction (500 MB/sec bidirectional) with a latency in the four to six microsecond range. Communication is user space to user space, and may be over IP or GM, Myricom’s user-level software. With the July 2004 announcement of the Myrinet “E” adapter, bidirectional bandwidths in the 930 to 950 MB/sec. and latencies of 3.5 microseconds are achievable.

If the parallel programming environment requires less interprocessor communication, a less robust interconnect fabric such as Ethernet (at lower prices) might be substituted. GigaNet,
Quadratics, SCI, or ServerNet might also be chosen for a custom offering. InfiniBand from Topspin® was included in the standard product offering in July, 2004.

Switches are used to construct an internal network for systems management and to interface external networks. Various switches from Cisco may be used, including the 6503 and 6509 series and the 3750G-24T-S. Switches such as those from Extreme Networks, Force 10 (and others) provide alternative solutions in a custom offering.

Terminal servers provide remote access to the OS consoles of the nodes through a serial network. Additional functionality added by a KVM (Keyboard, Video, Mouse) switch and the C2T Interconnect™ (Monitor/Keyboard/Mouse Cable Chaining) technology allow association of a single KVM with any node on the systems management network. A terminal server and KVM switch are included with the Cluster 1350.

External I/O, such as SCSI RAID devices should typically also be in the racks with the nodes, switches, and so on.

**Software:** The basic software for the Linux cluster is the Linux operating system, which is installed on each node in the cluster. Fortunately, for large clusters, the systems management software allows the installation to be cascaded through the management networks. The Cluster 1350 can be configured with a Red Hat Linux distribution, a SUSE LINUX distribution, and possibly other distributions in the future.

While Linux clusters are sometimes called Beowulf clusters, this is an oversimplification. Beowulf clustering software was the first real attempt to combine Intel Architecture systems into a single entity, Beowulf itself is not a “thing”, but rather a collection of software that must be downloaded and integrated by the installer. It has recently been reported that tutorials on Beowulf installation have expanded to fill two days.

A more mature systems management tool, called xCAT (for eXtreme Cluster Administration Tool), was developed by IBM for the deployment of the custom solution clusters. The following functions are supported by xCAT:

- Remote hardware and software reset (Ctrl+Alt+Del)
- Remote OS/POST/BIOS console
- Remote vitals (fan speed/temp/etc.)
- Remote power control (on/off/state)
- Remote hardware event logs
- Remote hardware inventory
- Parallel remote shell
- Command line interface (no GUI)
Single operations can be applied in parallel to multiple nodes
Network installation (PXE)
Support for various user defined node types
SNMP hardware alerts

An IBM Redbook details the installation considerations for an xCAT cluster. It is available at: http://publib-b.boulder.ibm.com/Redbooks.nsf/65f0d9cea6e0ab57852569e0007452bb/e0384f6e6982b28986256a0f005b7ba4?OpenDocument
Note that xCAT uses the xSeries service processor and that xCAT is not in Open Source. It is supplied at no charge with custom Linux cluster solutions from IBM as an “as-is” software component. It is written in script; as a result source code is included.

Cluster Systems Management (CSM) for Linux, an IBM Licensed Program Product, provides systems management function in a fashion similar to the PSSP systems management layer on the IBM @server Cluster 1600 / RS/6000 SP. CSM is the standard system management software for the Cluster 1350.

CSM is now also available to manage clusters of xSeries, BladeCenter and e325 servers running Linux, pSeries servers running AIX 5L, and pSeries servers running Linux, or a combination of the three. CSM for Linux is intended for clients who want robust cluster systems management on an open, Intel Architecture processor-based server.

The IBM Redbook for CSM may be found at: http://www.redbooks.ibm.com/pubs/pdfs/redbooks/sq246601.pdf.

Other software, Open Source and proprietary, may be selected and tailored to the client’s needs, and may be installed as part of a complete system solution. Examples of this software include the Portable Batch Scheduler (PBS) and the Maui Scheduler, both made available as Open Source, providing sophisticated job and resource scheduling, and accounting data. Other examples include MPICH for parallel programming, many math libraries, parallel debug and performance tools, and many ISV applications. IBM provides tools for CSM to work with the popular OSCAR Open Source package through its alphaWorks® site (http://www.alphaworks.ibm.com/tech/ect4linux).

**System Design and Specifications:** Design of a Linux cluster or super cluster is not trivial. Several layers of abstraction must be addressed, and the level of complexity increases with the cluster size, in some cases super linearly. Proposals for Linux cluster solutions need to be generated by those with the necessary understanding of the issues. Simple aggregation of the necessary parts is almost never successful.

Some basic design principles and best practices are fundamental.
Although the number of nodes needed for the problem or workload at hand is fairly straightforward, more (infrastructure) nodes will be needed:

- Each 32 to 64 compute nodes may require a head node that is a management node where compilers, job schedulers and the like are found.

  *If a head node will ever serve as a compute node, it must be similar in configuration.*

- Each system will require a management node that may or may not be one of the head nodes.

- External I/O will require one or more storage nodes (and may require network switches).

  *And when can one type of node serve double duty?*

Three functional networks must be included:

- A network for interprocess communication (IPC) is needed. Its speed depends upon the problems to be addressed.

- A network is needed for file I/O. If an IPC network is present, it might also be the I/O network.

- A network is needed for systems management. Often 10/100 Ethernet, it depends on cascaded topologies using head nodes, management nodes, etc., connected by Ethernet switches.

A terminal server must be included.

Network topographies, network addresses, and cable schemes (to include cable lengths) must be addressed. While a rack will theoretically hold 42 1U boxes, space is often left in the rack for expansion, and/or to manage power and heat issues.

And finally, the physical characteristics (power, heat, floor loading, service clearances, tip factor, etc.) must be addressed. Traditional machine rooms with raised floors may well be needed, as may adequate conditioned power and adequate cooling, both in BTUs and in Cubic Feet per Minute of cool air distribution.

The Cluster 1350 addresses these issues. Best practices are observed by design.
Diagram of a Large Linux Cluster

Logical Network View

- Myrinet

IPC Network - Myrinet

Terminal Servers

Console

Myrinet

38.4K Serial

x345

Front-end or Head Node

x345 Storage Node

x345 Management Node (2nd optional)

Service Processor Network

Backbone Networks

Network Public

I/O and Management Network

Secure Management Network

Public Network

IPS

IP-PCI Network

10/100/1000 mbit

10/100/1000 mbit

10/100 mbit

RS-485

10/1000 mbit

100/1000 mbit

100/1000 mbit

100/100 mbit

10 mbit

100/1000 mbit

100/1000 mbit
IBM @server Cluster 1350 Overview

**Hardware:** The Cluster 1350 is made up of e325, x335, x345 and BladeCenter HS20 nodes, in IBM 19” standard 11U, 25U, or 42U racks. Switches (Myrinet) for interprocess communication are optional. 10/100 Mbps and Gigabit Ethernet switches and terminal servers are included. This is conceptually the same as custom-built solutions that were previously delivered.

A complete description of the Cluster 1350 may be found at:  

A major difference is the integration. While the custom solutions may be integrated in any fulfillment center, or even on the client’s floor (not recommended), the Cluster 1350 hardware is manufactured (i.e. integrated) and tested at an IBM plant. Many advantages are realized from the integration and manufacturing processes:

1. The manufacturing facility orders and manages parts inventory, including non-IBM parts.
2. Normal forecasting and supply demand planning for component parts follow naturally.
3. Component and assembly level testing is included.
4. Engineering design considers issues such as packaging, cooling, efficient assembly, shipping and in-field servicing.
5. Compliance to governmental emissions regulations is tested.

IBM facilitates service by acting as a single point-of-contact for the client. While the OEM hardware components carry their own brand, IBM acts as a single point of contact for warranty and maintenance service for the entire cluster, including the OEM content.

While there are many economies accrued from a manufactured product, such as lower costs and higher quality, not all configuration issues are addressed with this approach. Requirements that fall outside the product boundaries may occur, and special bids will be entertained.

**Software:** The software for the Cluster 1350 is essentially that of the customized solution. The Cluster 1350 runs the Linux operating system from Red Hat or SUSE LINUX and CSM for Linux is included with the Cluster 1350 as the basic management software. Clients must obtain the version of the Linux operating system appropriate to the configuration chosen, and may choose to install Linux and CSM themselves or have this service done by IBM or a qualified business partner. Specialized software, Open Source or proprietary, as well as GPFS for Linux, may be included in the installation as part of an installation services contract.
Software from IBM ISV partners may be included. SteelEye for High Availability has been noted earlier in this paper. Matrix from PolyServe (also noted earlier) provides an innovative SAN solution that is accessible from other (non-Linux) platforms and provides significant performance advantages in Linux environments.

An additional objective is to be able to deliver application specific software stacks. A commercial software stack, for example, might include WebSphere®, DB2, MySQL, SteelEye, Matrix, etc. An HPC stack might include MPICH, PVM, the Maui Scheduler, math libraries, compilers, and programming/profiling tools. Another application area that lends itself to this approach is bioinformatics.

**Future Generations:** Linux Clusters began as a simple proposition; rack mounted Intel servers running Linux as the operating system. Recent developments provide an interesting insight toward trends and directions.

Clusters now use Intel Architecture components in addition to those from Intel itself. AMD Opteron is a popular choice for some applications.

Blades have become compute nodes in clusters, allowing for increased density.

The recently announced IBM BladeCenter JS20 blade uses an IBM Power Architecture™ processor and runs Linux.

IBM® pSeries products support Linux. The Cluster 1600 uses POWER™ technology with Linux or AIX 5L as the operating system.

The cluster management software, CMS, is now supported for Linux nodes, both Intel Architecture and POWER. It also supports POWER nodes running AIX 5L, becoming a heterogeneous cluster management system.
Advantages for Linux Clusters from IBM

IBM Solution Differentiators: Linux clusters from IBM may be considered to be the most full function Linux clusters available. Given that clusters are available from both “white-box” and other name brand vendors, it is important to understand the functionality of the IBM added value systems.

Systems Management: Systems management is considered to be the single most important attribute of a large cluster. While small clusters can manage with minimal systems management, the need sometimes goes up super-linearly (more than linearly, less than exponentially) with the cluster size. Clients with experience in large systems, especially the RS/6000 SP or Cluster 1600, already understand this.

Consider the tasks needed to manage a cluster: each node must be booted. Software changes/fixes must be applied to each node. The cluster must have a place, preferably a single place, for users to log in, run compiles, access system resources, etc. The systems administrator must have a way to control access to systems resources (security) and to monitor and manage the health of the system.

There are two major components of systems management: the service processor and the software management layer.

The Service Processor: All nodes offered in Linux clusters today have an embedded service processor. It is included in the price of the system and does not occupy a slot. It allows remote power on/off, monitors hardware health and provides alerts for various modes of failure detection. There is no standard for service processors; the implementation of each vendor (if it exists) is unique.

The Systems Management Software: Linux clusters from IBM, whether controlled by xCAT or by CSM, have a very high function systems management layer. Both are dependent on the service processor (sometimes called SP, not to be confused with the RS/6000 SP). Neither would necessarily be expected to function correctly on another vendor’s cluster. CSM is currently supported only on IBM clusters, and xCAT is distributed as-is.

Management functions include hardware control, including power on/off and reset, software reset, OS/POST/BIOS console, vitals like temperature and fan speed, hardware event logs, and hardware inventory, all available to the system administrator through a remote interface. The above functions can also be monitored and generate SNMP hardware alerts. In addition, both software packages provide network installation tools and the ability to apply single operations to multiple nodes in parallel.
While other vendors do indeed provide some systems management function, our experience has been that they have less function than xCAT or CSM. Both IBM solutions are architected for very large clusters; they are scalable. Scalability is a design criterion. Software not designed for scalability sometimes performs exponentially worse as the system size increases. IBM scalability can be demonstrated by the very large Linux clusters deployed by IBM, several over 1000 nodes.

**RAS of IBM Components:** IBM server hardware includes features that make the individual nodes less likely to fail, and to allow predictive maintenance where failures might occur. Built-in hardware redundancy for many functions allows the system to gracefully survive many failures.

**IBM's Packaging Advantage:**

**C2T:** Monitor/Keyboard/Mouse Cable Chaining - C2T is a daisy chain cable arrangement that eliminates the need for one set of KVM cables for each node. That in turn eliminates up to 120 cables per rack, radically decreasing the number of terminal servers required. A 128 node system, for example, would reduce the cable and terminal server cost by a factor of 10 compared to direct cabling.

**Service Processor Network:** Serial cables are daisy chained to the Remote Service Attachment (RSA) card on every 18th (or so) node. This allows one Ethernet cable back to Ethernet switch for 18 nodes, instead of one cable per node. Unlike other offerings, which use a PCI slot for the service processor, in the Linux cluster from IBM, the service processor is built in, and contains its own Ethernet port and associated cable.
Conclusion:

Linux is increasingly popular and increasingly capable. Linux clusters represent a way to horizontally scale beyond the capabilities of any given Linux SMP implementation. The price/performance of the Intel platforms and the enhanced functions of the RISC platforms both provide attractive building blocks for Linux clusters.

IBM has many years of clustering experience on various platforms and brings much of this experience to bear in the Linux cluster environment. The components necessary to successfully build and manage large Linux clusters, or Linux super clusters, include dense packaging, built-in systems management using service processors and highly capable scalable software, and strong systems integration skills.

IBM has built and continues to build some of the world’s largest and most complex Linux clusters in response to our client’s requirements. The intellectual capital that allows successful creation and implementation of the largest clusters is being brought to bear on the emerging commercial market for Linux clusters of all sizes.
This rising interest in clusters led to the formation of an IEEE Computer Society Task Force on Cluster Computing (TFCC1) in early 1999. An objective of the TFCC was to act both as a magnet and a focal point for all cluster computing related activities. As such, an early activity that was deemed necessary was to produce a White Paper on cluster computing and its related technologies. To sum it up, clusters for technical and scientific computing based on Linux and other standalone Unix platforms like AIX are here, and they work. In the area of commercial cluster computing, Linux still lacks essential functionalities which conventional Unix systems and in parts even Windows NT do offer. Setting up a local Linux cluster is useful to quickly test applications targeted for Linux clusters but are developed on a Windows machine. Prerequisites. Linux-based Service Fabric clusters do not run natively on Windows. To run a local Service Fabric cluster, a pre-configured Docker container image is provided. Before you get started, you need: At least 4-GB RAM. Creating a simple Linux cluster. One of the most interesting things about clustering is that you can build Linux-based clusters with minimal effort if you have basic Linux installation and troubleshooting skills. Let's see how this is done. For our cluster we'll use MPICH and a set of regular Linux workstations. As we did when building the Linux cluster, we'll go through the steps to be performed on the head node. OSCAR will configure all other nodes automatically, including the OS installation. And here is an excellent paper by Jack Dongarra on the HPC Challenge benchmark and a thorough introduction to it. The NAS Parallel Benchmarks (NPB) are a small set of programs designed to help evaluate the performance of parallel supercomputers.