Beowulf Cluster – A Poor Man's Supercomputer*

V. Rajaraman

In this article, I describe the development of the Beowulf cluster, a parallel computer that consists of a network of off-the-shelf personal computers interconnected by Ethernet and uses Linux as its operating system. This style of building parallel computers has become very popular as fairly high-performance computers can be built in a university laboratory at a reasonable cost. Eighty percent of supercomputers listed in the TOP500 list of the fastest computers in the world are clusters.



V. Rajaraman is at the Indian Institute of Science, Bengaluru. Several generations of scientists and engineers in India have learnt computer science using his lucidly written textbooks on programming and computer fundamentals. His current research interests are parallel computing and history of computing.

Keywords
Parallel computer, computer cluster, message passing interface, Beowulf competition.

Prologue

The early 1980s were exciting times for researchers in parallel computers. The Cosmic Cube built at Caltech in 1983 showed that a parallel computer built using off-the-shelf microprocessors could be used effectively for solving compute-intensive problems. Several companies including Intel, nCUBE, Thinking Machines Corporation, Meiko, and Parsytech built parallel computers and successfully marketed them in the early and mid-eighties. In parallel with these developments, Cray also built the XMP and YMP series of expensive parallel supercomputers, which had the advantage of numerous well-tested application programs. However, there were major problems confronting the users of parallel computers in the late 1980s. One was vendor instability. Many parallel computer manufacturers had a short life span. In addition, these parallel computers were incompatible. Even programs written in a high-level language such as Fortran on one machine could not be run on another vendor's machine because the primitives developed by vendors to communicate among processes were proprietary.

*Vol.28, No.2, DOI: https://doi.org/10.1007/s12045-023-1546-z

In 1991, the US congress passed the High-Performance Computing Act, which led to the creation of the High-Performance Computing and Communication Program (HPCC) in early 1992.

Meanwhile, in 1991, the US congress passed the High-Performance Computing Act, which led to the creation of the High-Performance Computing and Communication Program (HPCC) in early 1992. One of the major successes of HPCC was the creation of the Internet. Besides, HPCC also provided funding to many academic institutes in the US to buy supercomputers. Research funding to develop parallel computers was given to universities and government laboratories. The National Aeronautics and Space Administration's (NASA) laboratories were among the laboratories that received funding from HPCC to design and develop parallel computers.

The Genesis of the Beowulf Cluster

As one of the projects of the HPCC program, a Center of Excellence in Space Data and Information Sciences (CESDIS) was set up at the Goddard Flight Center of NASA at Greenbelt, Maryland. One of its goals was to build a supercomputer with a teraflop speed by 1997. James Fischer, who headed the supercomputer project, started a project to build one and invited Thomas Sterling, a scientist at NASA, to join the effort. Coincidentally Donald Becker, a friend and a previous colleague of Sterling, was his house guest waiting to join another government laboratory. Becker was an expert in Linux, an open source operating system. Sterling and Becker informally discussed issues related to the networking of computers and the role of Linux in such networks. Becker's prospective employer did not issue the appointment letter due to some bureaucratic problems. Sterling, who had been grappling with the problem of building a parallel computer, had a flash of intuition and thought it would be possible to build a parallel computer using the processors of PCs and networking them using Linux as the operating system. He felt that Becker's expertise in Linux would be invaluable in this effort. Sterling requested Fischer to recruit Becker to work with him on the supercomputer project. Fischer agreed, and a team of half a dozen scientists was formed with Sterling and Becker in the lead. When the administrator of the NASA laboratory asked Sterling for the

Sterling, who had been grappling with the problem of building a parallel computer, had a flash of intuition and thought it would be possible to build a parallel computer using the processors of PCs and networking them using Linux as the operating system.



Figure 1. Thomas Sterling standing in front of a Naegling 10 GFLOPS Beowulf cluster at Caltech. (Photo courtesy: https://spinoff.nasa.gov/Spinoff2020/it_1.html)

name of the project, he had not thought of one. On the spur of the moment, he remembered that he had an old English epic belonging to his mother (who was a scholar in English), which had Beowulf as its hero and gave Beowulf as the name of the project [1]. In the epic, the warrior Unferth doubts the ability of Beowulf to slay the monster Grendel who had been tormenting the Danes for twelve years. To everyone's surprise, Beowulf slays it. However, when colleagues and others asked "Why Beowulf?" Sterling's standard reply was "an underdog challenging a formidable opponent", meaning the established parallel computer vendors.

The First Beowulf Cluster [2]

The team led by Sterling decided to use motherboards with an Intel 40486 (model DX4) with a clock speed of 100 MHz and a 16 MB RAM as the main computing engine. The model DX4 was one of the fastest in the x486 family with 16 KB primary cache and 256 KB secondary cache memory. It had a built-in math co-processor. A 500 MB disk drive was attached to each board. Sixteen of these computers (without keyboard and display) were networked using two Ethernet cables, each with a peak speed of 10 Megabits per second (Mbps). Two Ethernets were used to increase the speed of communication between processors to around

The team led by Sterling decided to use motherboards with an Intel 40486 (model DX4) with a clock speed of 100 MHz and a 16 MB RAM as the main computing engine.

20 Mbps per second. This enabled faster message-passing between processes running on remote computers. Higher speed Ethernets were not available when this cluster was built in 1993. A special device driver was written to manage the communication of packets in the network with a dual Ethernet connection. This was possible as the system used a Linux operating system whose source code was available. A PC with a display and a keyboard managed the networked computers, called a cluster.

Beowulf cluster was used to solve many benchmark problems such as a 2-dimensional fluid dynamics code and a gravitational n-body simulation.

This cluster was named Beowulf and was used to solve many benchmark problems such as a 2-dimensional fluid dynamics code and a gravitational n-body simulation. Parallel programs were written for these problems using standard message-passing primitives. A 16-computer cluster was able to attain a speed of 60 MFLOPS. A Cray CRI T3D parallel computer with 16 processors ran the same program 2.5 times faster. Experiments showed that problems requiring a large number of inter-computer communication did not perform very well. The beauty of the system was that it was built using commodity off-the-shelf (COTS) components with not a single custom-made component. It also used an open source operating system. This made it a "make-it-yourself" parallel supercomputer.

Programming Beowulf Cluster

An effort was initiated under the leadership of Jack Dongarra in 1992 and resulted in a draft Message Passing Interface (MPI) Standard.

In the early 1990s, programming distributed memory parallel computers was not standardized. Each vendor had a set of primitives which were added to Fortran or C to enable communication among processes running on remote processors. An effort was initiated under the leadership of Jack Dongarra in 1992 and resulted in a draft Message Passing Interface (MPI) Standard. It was presented at the Supercomputing '93 Conference in 1993, and comments were invited. (Supercomputing is a conference on Supercomputers organized by the Association for Computing Machinery (ACM) and the Institute of Electrical and Electronics Engineers (IEEE) annually in the US). Based on the feedback, the first version of the MPI standard was released in June

1994 by a body of professionals known as the *MPI Forum* (`www.mpi-forum.org`). MPI specifies a public-domain, platform-independent, message-passing library of routines for several languages, including Fortran and C, that can be used to write programs for distributed memory parallel computers. Programs for Beowulf clusters are written in any one of the popular high-level programming languages, such as Fortran, C, or C++, using the MPI library of routines for inter-process communication. This makes the programs portable.

Programs for Beowulf clusters are written in any one of the popular high-level programming languages, such as Fortran, C, or C++, using the MPI library of routines for inter-process communication.

Two Interesting Beowulf Clusters

An important development was the demonstration of a Beowulf cluster designed by the Centre for Advanced Computer Research of the California Institute of Technology (Caltech) at the Supercomputing '97. Caltech's Cluster consisted of 138 PCs using a Pentium Pro processor (100 MHz clock with a built-in floating-point processor) and 16 PCs with dual Pentium Pro processors. A gigabit Ethernet switch inter-connected the PCs providing fast communication between them. The front-end computer was a quad Pentium Pro computer with a keyboard and display. The cluster ran a benchmark program called the n-body problem that is used to predict the individual movements of a group of objects attracted by gravitational forces. It attained a speed of 10.9 GFLOPS [3] and won the Gordon Bell Prize for the best performance at the lowest price at Supercomputing '97 [4].

Engineers at the Australian National University designed another interesting Beowulf cluster called Bunyip with 196 computing nodes. The computing node was a dual Pentium III with a 550 MHz clock, built-in floating-point processor, 384 MB RAM, and 13 GB disk [5]. Each node had three Ethernet interface cards running at 100 Mbps. There were two front-end computers Pentium IIIs each with a keyboard, display, hard disk, and CDROM. They also had two spare nodes to substitute for the failed nodes. It ran using the Linux operating system. The entire system, including labour, cost around $151,000. The system ran at 163 GFLOPS

speed for an ultra-large-scale neural network training problem that was demonstrated at Supercomputing '2000 and won the Gordon Bell Prize for the best performance of a high-performance computer at the lowest price at this conference.

Epilogue

The building and demonstration of the Beowulf cluster was a landmark in parallel computing. It heralded a flurry of activity by several universities that started building PC-based clusters in their laboratories.

The building and demonstration of the Beowulf cluster was a landmark in parallel computing. It heralded a flurry of activity by several universities that started building PC-based clusters in their laboratories.

With the performance of PCs and networking speed increasing steadily and the cost per GFLOP of processors going down continually, Beowulf clusters are becoming very fast and cost-effective. Many companies are now custom-building clusters for universities based on the specifications given by them. The Beowulf community is active with a mailing list at `Beowulf.org`.

It is interesting to note that, in the last twenty years, more than 80% of systems in the TOP500 supercomputers list are clusters many of them using commodity off-the-shelf components. (TOP500 is a list published every 6 months of the fastest 500 supercomputers in the world) [6].

Another piece of information of interest to students is that the Supercomputing Conference (SC) runs a student cluster competition every year. A prize is awarded to a student team that designs and builds a small cluster and applies it to optimally solve a scientific application during the two days of the conference [7].

Acknowledgment

I thank Prof. P. C. P. Bhatt, Dr. N. Dayasindhu, Prof. R. Govindarajan, and Prof. S. Ramani for reviewing this article and their valuable suggestions that improved it.

Suggested Reading

[1] **Beowulf clusters make supercomputers accessible, NASA Publication, 2020.** https://spinoff.nasa.gov/Spinoff2020/it_1.html

[2] **Donald J. Becker, Thomas Sterling, Daniel Savarese, John E. Dorband, Udaya A. Ranawake, and Charles V. Packer, Beowulf: A parallel workstation for scientific computation, *Proceedings, International Conference on Parallel Processing*, 1995. (Accessible at** webhome.phy.duke.edu**).**

[3] **Steve McGrath, Supercomputing gets a new hero, *The Free Library*, Nelson Publishing, 1998.** https://www.thefreelibrary.com/Supercomputing+gets+a+new+hero-a021071072

[4] **What is Beowulf?** http://yclept.ucdavis.edu/Beowulf/aboutbeowulf.html

[5] **Douglas Aberdeen, Jonathan Baxter, and Robert Edwards, 92¢/MFlop/s ultra large-scale neural-network training on a P III cluster, IEEE Computer Society.** https://arxiv.org/pdf/1911.05181.pdf

[6] **See** https://www.top500.org/

[7] **See** https://sc22.supercomputing.org/program/studentssc/student-cluster-competition/

Address for Correspondence
Prof. V. Rajaraman
Supercomputer Education & Research Centre
Indian Institute of Science
Bengaluru 560 012, India.
Email: rajaram@iisc.ac.in