Supercomputer using Cluster Computing

Karan Kantharia[1], Adarsh Chaturvedee[2], Manas Nikam[3], Rujul Shringarpure[4], Disha Bhosle[5]

Department Of Electronics Engineering[12345]
Atharva College of Engineering[12345]
Mumbai, India[12345]

Abstract— **The supercomputers are known to operate at their highest operational rate and are faster than other computers at the same time. There have been many approaches to build a supercomputer such as grid and distributed computing, but the most efficient and economical way is by using cluster computing. Computer clusters have emerged as a result of convergence of a number of computing phenomena which include high speed networks, the availability of low-cost microprocessors and software for high-performance distributed computing. Computer clusters are used for computation-intensive purposes, rather than handling IO-oriented operations. In computing, a cluster is made of many individual computers that work collectively to solve a problem at hand, the computing nodes communicate over a fast network, but they share the same memory. These very tightly coupled configurations are designed for work that may approach supercomputing. Due to varied applications, a supercomputer using cluster computing can be beneficial to needs of larger processing speeds and higher data handling capabilities.**

Keywords— Cluster computing, supercomputer.

I. INTRODUCTION

The history of early computer clusters and supercomputing is more or less directly tied into the history of early networks goes back to the 1960s, as one of the primary motivations for the development of a network was to link computing resources, creating a de facto computer cluster. At that time, It had been stated that clusters were not invented by any specific vendor but by customers who could not fit all their work on one computer, or needed a backup. The first production system designed as a cluster was the Burroughs B5700 in the mid1960s. This allowed up to four computers, each with either one or two processors, to be tightly coupled to a common disk storage subsystem in order to distribute the workload. These used innovative designs and parallelism to achieve superior computational peak performance.

A computer that performs calculations at speeds that are currently the highest achievable rates of operations is known as a supercomputer. They are generally used for large scale calculations, thus their operating speeds are considerably higher than the conventional 'interactive' computers that are in widespread use today [1]. Such computers normally take one of two paths for processing, distributed computing or cluster computing. Cluster computing comes in handy when computation-oriented jobs are to be done instead of handling the input-output operations. Such clusters are useful to perform computations at a higher speed and efficiency. A cluster would ideally consist of tightly bound computers to perform tasks in conjunction with each other. They are cost-effective too when compared to a computer with similar specifications and computational ability. It was the result of a number of trends in the field of computers which made such clusters an integral part of a wide range of operations such as small business network clusters or one of the fastest computers such as IBM's Sequoia.

II. LITERATURE SURVEY

A. Cluster Computing

A cluster is a type of parallel or distributed computer system, which consists of a collection of inter-connected standalone computers working together as a single integrated computing resource. One of the issues in designing a cluster is to know how tightly coupled the individual nodes may be. It may be possible that a single computer job may require frequent communication among nodes or it implies that the cluster shares a dedicated network and is densely located, and probably has homogeneous nodes. The other extreme possibility is where a computer job uses one or few nodes, and needs little or no inter-node communication, approaching grid computing.

The activities of the computing nodes are monitored by using a clustering middleware. It is a software layer that sits atop the nodes and allows the users to treat the cluster as by and large one cohesive computing unit such as a single system image concept. The different types of middleware are message based, remote procedure call based, object request broker, OLE/COM and internet based. A key component in cluster architecture is the selection of interconnection technology. It may be classified into four categories depending on whether the internal connection is from the system bus or the I/O bus, or whether the communication between the computers is performed primarily using shared storage or messages.

The operating system in the individual nodes of the cluster provides the fundamental system support for cluster operation. The desired features of a cluster operating system includes manageability, stability, performance, extensibility, scalability, support and heterogeneity. Cluster Management Software (CMS) is primarily designed to administer and manage application jobs submitted to workstation clusters. The task can be a sequential or parallel application that needs to run interactively or in the background. This software encompasses the traditional batch and queuing systems. CMS can be used to help manage clusters by optimizing available resources, prioritize the usage of resources, stealing the CPU cycles, check pointing, task migration and to ensure that the task are completed successfully.

Special Issue - 2017

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICIATE - 2017 Conference Proceedings

The scalability of an application indicates its ability to effectively use additional resources as the size of the cluster and the application grow [2]. A cluster can potentially provide a large data storage capacity since each node in the cluster will typically have at least one disk attached. These disks can produce a large, common storage subsystem. This storage subsystem can be more effectively accessed using a parallel I/O system. The general design goals of a parallel I/O system are to increase the bandwidth, maximize read and write operations, minimize unnecessary and costy communication, and to maximize the hit-ratio.

B. Supercomputer

The systems with a large number of processors generally take one of two designated paths. One such approach is grid computing approach in which the processing power of many computers is organized as distributed, diverse administrative domains, is used whenever a computer is in its idle state. In the second approach, a large number of processors are used in close proximity to each other, a computer cluster. In such a centralized vastly parallel system the flexibility and speed of the interconnections become very important parameters. The use of multi-core processors combined with centralization is an emerging field.

Supercomputers generally aim for the maximum in capability computing as compared to capacity computing [6]. Using the maximum computing power to solve a single large problem in the shortest amount of time is the major aspect of capability computing. Often a capability system is able to solve a problem of a size or complexity that no other computer can. On the contrary, Capacity computing can be thought of as using efficient cost-effective computing power to solve a fewer large problems or collectively many small problems. Architectures and specifications that lend themselves to supporting many users for routine everyday tasks may have a lot of capacity, but they are not considered supercomputers for the reason being that they do not solve a single very complex problem.

In general, the speed of supercomputers is measured and benchmarked in "FLOPS" which stands for FLoating point Operations Per Second. These measurements are commonly used with an SI prefix such as tera and peta. A typical supercomputer consumes large amounts of electrical power which is converted into heat, and thus requires a lot of cooling. Heat management is a major issue in complex electronic devices, and affects powerful computer systems in various ways. The thermal design power and CPU power dissipation issues in supercomputing surpass those of traditional computer cooling technologies. The energy efficiency of computer systems is generally measured in terms of "FLOPS per watt".

C. Methodology

In order to get more computing power and better reliability we orchestrate a number of low-cost commercial off-the-shelf computers which leads to a variety of architectures and configurations.
The computer clustering approach connects a number of readily available computing nodes like the personal computers used as servers on a fast local area network [4]. Computer clustering is of a far more distributed nature unlike the peer to peer or grid computing systems. It relies on a centralized management approach which makes the nodes available as orchestrated shared servers.

A computer cluster may be a simple two-node system which just connects two personal computers, or may be a very fast supercomputer. The Beowulf cluster is a basic approach to building a cluster which may be built with the help of a few personal computers to produce a cost-effective alternative to traditional high performance computing [5]. Although a cluster may consist of just a few personal computers connected by a simple network, the cluster architecture may also be used to achieve very high levels of performance.
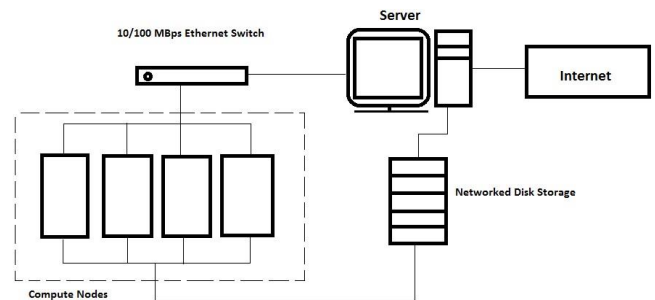

Fig. 1. *Block Diagram of a cluster computer in accordance with Beowulf*

A Beowulf cluster is simply a collection of inexpensive commercial off the shelf (COTS) computers networked together running Linux and parallel processing software [7]. Raspberry Pi is a single-board Linux-powered computer and it provides a unique feature in which they have external low level hardware interfaces for embedded systems use and are cheaper for implementing a super computer using cluster computing.

D. Hardware and Software

The main computing device used for this project is Raspberry Pi 3.It Raspberry Pi acts as a single computing node that is connected to a main server or centralized system. It is a lowcost stand-alone computer. It the newest edition after Raspberry Pi 2 Model B. It consists of a 1.2GHz 64-bit quadcore ARMv8 CPU, 1GB RAM, 802.11n Wireless LAN and Bluetooth version 4.1 for establishing communication between devices. MicroSD card, compatible with Raspberry Pi 3, is used as for data storage. The SD cards used for this project have a capacity of 8GB. A NETGEAR 5-Port Switch is used for interconnecting LAN. It provides a 10/100 Mbps fast Ethernet connection. Besides these, Anker Premium 3ft Micro USB cables for connecting the Raspberry Pi development boards, a USB Hub for multiple USB support and Belkin RJ45 CAT 5e 3ft Cables are some other components used.

Raspbian OS is an operating system designed and optimized for Raspberry Pi. It is a free operating system based on Debian. It is a pre-compiled software bundled in a standard and user-friendly format for hassle-free installation. For efficient Cluster computing, an SMP-based approach is required. OpenMP is a standard used for this

Special Issue - 2017

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICIATE - 2017 Conference Proceedings

project, which offer a mixed-mode programming approach which helps to achieve adequate levels of performance. For cluster applications, many important numerical libraries have been used, that were originally developed for parallel supercomputer. They provide a good starting point for clusters. MPI (Message Passing Interface) is used as a communication protocol for parallel computing [3,8]. A Network Mapper software is used to map and configure different devices connected to a network. Besides these, some middleware software like MPI4PY have been used to interface Raspberry Pi with communication software like MPI. It helps to make the supercomputer available to people who are not well versed with PYTHON but must have a good command over the C language.

E. Applications

Supercomputers have the capacity to perform highly intensive computations with ease and thus they can be made useful in a variety of ways. Automatic number plate recognition with the help of digital image processing. It will be helpful to track the vehicle in case of theft or while the vehicle is parked in parking lots the information of the owner will be available instantaneously. A cluster can be used to create a perfect base to build highly scalable Internet services. Their inherent working principle makes them ideal constituents for parallel processing. Applications that need to be executed in the same manner but with different initial conditions can be performed by parametric computing. Petroleum reservoir simulations to better understand and manage the resource can be done with cluster computing. Weather forecasting uses many variables for mathematical calculations and perform high level of computations using a supercomputer. Similarly, earthquake simulations based on multiple factors can be performed to be forewarned and gauge the impacts of such calamities.

III. CONCLUSION

The ability of cluster computers to process the larger quantities of data in a much shorter period of time has found its way into various applications ranging from a web server, an audio processing system, data mining, network simulations, to image processing. There is even much more emphasis placed on using the commodity-based hardware and software components to achieve high performance and scalability and at the same time keep the ratio of price versus performance low. A number of software solutions for providing a single system image for clusters are emerging, yet there are still many opportunities for integrating various cluster tools and techniques and making them work together and so help create more usable and better unified clusters. Eliminating the severe drawbacks of supercomputers which include portability and heating, supercomputers using cluster computing is the answer to the needs of the future generations to come.

REFERENCES

1. R. Buyya (Ed.), High Performance Cluster Computing: Systems and Architectures, Vol. 1, 1st Edition, Prentice-Hall, Englewood Cliffs, NJ, 1999.
2. "The scaling of many-task computing approaches in python on cluster supercomputers" by Monte Lunacek Res. Comput., Univ. of Colorado Boulder, Boulder, CO, USA at 2013 IEEE International Conference on Cluster Computing.
3. "MPI based cluster computing for performance evaluation of parallel applications" by B R Nanjesh Dept. of Comput. Sci. & Eng., Adichunchanagiri Inst. of Technol., Chikmagalur, India at Information & Communication Technologies (ICT), 2013 IEEE Conference.
4. "An Introduction To Cluster Computing Using Mobile Nodes" by Priti Sharma ,Lingaya's Univ., Faridabad, India at 2009 Second International Conference on Emerging Trends in Engineering & Technology
5. A. J. van der Steen, "An Evaluation of Some Beowulf Clusters," *Cluster Computing*, vol. 6, no. 4, Oct. 2003, pp. 287-297.
6. G. Pfister, In Search of Clusters, 2nd Edition, Prentice-Hall, Englewood Cliffs, NJ, 1998.
7. IEEE Task Force on Cluster Computing. http://www.ieeetfcc.org.
8. The Beowulf Cluster site, http://www.beowulf.org Message Passing Interface (MPI) Forum, http://www.mpi-forum.org