



A Study on Cluster Computing

Kanika Jethwani*, Surbhi Gaur

IT, GGSIPU, New Delhi,
Delhi, India

Abstract— A computer cluster is a group of computers that are linked and they work together closely, so that it appears as a single computer. The connection of components of a cluster among themselves is through fast local area networks. Each computer in the cluster has an operating system. A cluster improves performance and/or availability over that provided by a single processor. They are much more cost-effective than other processors of comparable speed or availability. Multiple computers consists of multiple storage devices form clusters (typically PCs and UNIX work stations) which appears as a single highly available system to the users. Cluster computing can be used for load balancing and high availability. The main concept of cluster computing is that, the cluster appears as a single system to the users. Computer clusters have applicability from small business clusters with few nodes to some of the fastest supercomputers in the world. In this paper authors have tried to explain the concept of Cluster computing (Brief History, Architecture, Types of Clusters, Features, Example of Cluster Computing Advantages and Disadvantages)

Keywords—Cluster, Hardware, Software, System, Applications

I. INTRODUCTION

Cluster computing can be described as an amalgamation of the fields of parallel, high-performance, distributed, and high-availability computing. Many academic and industry people like system designers, algorithm developers, network developers, and language designers, standardizing forums, faculties and graduate students are doing lot of research on Cluster Computing. Its applicability is not just limited to the areas of engineering; other areas are also benefited from this technology.

Clusters are based on the communication between nodes and designing fast and low latency networks is a must for clusters to become the configuration of the future. Plenty of work being done in this are in both hardware (Infiniband, SCI, Myrinet, QSnet) and software (Low latency protocols such as VIA). Allowing heterogeneity in both hardware and software (OS) is extremely important and is becoming one of the key issues in cluster research. This heterogeneous feature allows a better scalability and expandability (i.e. use old and new components) and has to be addressed at many levels. Tools are needed that allow users to develop applications for these configurations, and scheduling systems, which offer mechanisms for exploitation of heterogeneity and which delivers better application performance. File systems are also of great interest as clusters are being used for I/O intensive applications such as e-commerce, WEB servers or databases. Clusters also provide an excellent platform for solving a range of parallel and distributed applications in both scientific and commercial areas. Clusters can be used in scientific or supercomputing applications, such as complex crystallographic, earthquakes or hurricanes prediction, relativistic quantum chemistry of actinides and global climate modeling [2].

For the commercial applications, cluster can be best used in e-commerce as superserver, which consolidates web server, ftp server, e-mail server, database server, etc. Clusters can also be used in data mining applications to provide the storage and data management services for the data set being mined and computational services required by the data filtering, preparation and mining tasks. Other commercial application includes image rendering, network simulation, etc.

II. BRIEF HISTORY OF CLUSTER COMPUTING

The first clustering product that was commercially available was ARCnet, developed by Datapoint in 1977. It was not a commercial success. After this, DEC released their VAXcluster for the VAX/VMS operating system in 1984. The ARCnet and VAXcluster products supported parallel computing, shared file systems and peripheral devices. The idea was to provide the advantages of parallel processing, while maintaining data reliability and uniqueness. VAXcluster, now VMScluster, is still available on OpenVMS systems from HP running on Alpha and Itanium systems.

III. ARCHITECTURE OF CLUSTER COMPUTING

A Cluster is a type of distributed or parallel processing system that consist of a group of computers that are interconnected, working together as a single computing resource.

A computer can be a single processor or multiprocessor system (Workstation, PCs) consisting of all the facilities such as storage space, I/O and operating system. The nodes of a cluster can be physically separated and connected though a LAN or grouped in a single storage.

The architecture of a Cluster is shown in the following figure:

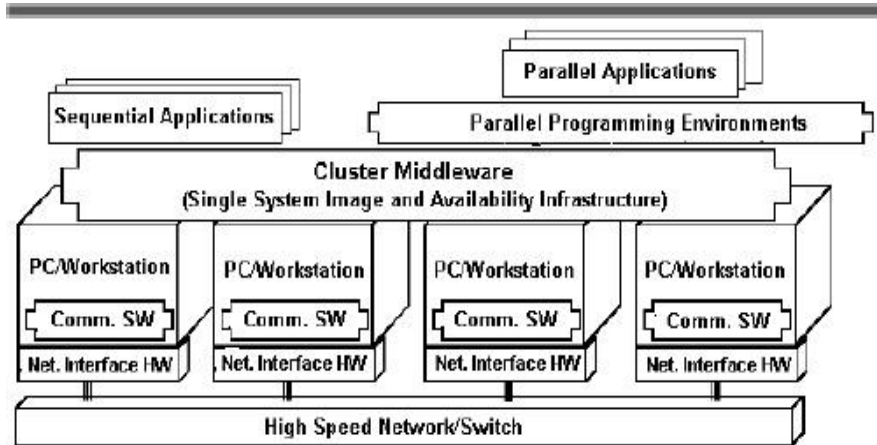


Fig 1: Cluster Architecture

The components of the architecture are discussed below:

- Multiple High performance computers(PCs and Work Stations)
- Operating System
- High performance Networks
- Network Interface Cards (NIC)
- Communication Protocols and Services
- Cluster Middleware (Single System Image and Availability Infrastructure)
 - Hardware
 - Operating System Kernel or Gluing Layer
 - Applications and Subsystems
 - Applications
 - Runtime Systems
 - Resource Management and Scheduling software
- Parallel Programming Environment and Tools [3].

IV. FEATURES OF CLUSTER COMPUTING

- **Stability:** It is the first and foremost characteristic of the cluster computing is its strength and health against any crashing process. Any problem can be recovered by dynamic reconfiguration.
- **Performance:** The performance aspect of the cluster computing mainly depends upon the memory management and processing. The programmer is able to transparently modify the relevant parameters to (fine-tune the OS) his specific demands.
- **Extendibility:** The user can easily integrate cluster-specific extensions; such as adding user-loadable drivers and by increasing the intensity of the documentation of interfaces.
- **Scalability:** The scalability of a cluster is mainly influenced by the properties of the contained nodes, which is dominated by the performance characteristics of the interconnectivity. This also includes the resource limits that an OS exhibits, foremost the maximum size of usable address space.
- **Support:** Conventionally, many intelligent and technically superior approaches in computing failed due to the lack of support such as, hardware drivers and middleware. This support is very much available in the cluster computing. The only drawback is the hardware costs (because usually dozens of nodes are to be installed). But it certainly enhances the cluster applications.
- **Heterogeneity:** Clusters provide a dynamic and evolving environment in that they can be updated with standard hardware just as the user needs to or can afford. Therefore, a cluster environment does not necessarily consist of homogenous hardware which requires the same operating system.
- **Manageability:** The other characteristic of cluster computing is that even after integrating a number of off-the shelf commodity computers it is quite possible to manage the single system unit.

V. TYPES OF CLUSTERS

- **Load Balancing Clusters:** - This type of clusters routes the requests for resources among available nodes that runs same programs or have same content. All the nodes of the clusters have the capability of handling the request for the same content or application. If there is a failure in any node, then the request is again distributed among the remaining nodes. For example-Distribution seen in web hosting environment.
- **Parallel/Distributed Processing Clusters:** - These types of cluster have the characteristics of increasing availability, performance, and scalability for applications, mainly that requires computational tasks or data intensive tasks. To solve a specific computational or data-mining task, a parallel cluster is needed. In parallel computing a request is divided into number of sub requests that are distributed among multiple nodes for processing within a cluster, unlike the load balancing or high-availability cluster.

- High availability clusters (HA- Mission critical application):- These clusters provide the end users with the uninterrupted availability of data or services. The purpose of these clusters is to ensure that a single instance of an application is only ever running on one cluster member at a time but if and when that cluster member is no longer available, the application will switch over automatically to another cluster member.
- High Performance and High Throughput Clusters: They are used for applications which require high computing capability.

Clusters can be classified according to the node type as homogeneous clusters and heterogeneous clusters [1].

- Homogeneous clusters: In these types of clusters, all nodes have similar characteristics. Each node is much like any other node. Amount of memory storage available and connections among the nodes are similar.
- Heterogeneous clusters: In these types of clusters, all nodes have different properties, in the terms of amount of memory storage available and interconnect performance.

Clusters may be classified according to the hierarchy they inherit.

- Single level or single-tier clusters: In these types of clusters, there is no hierarchical sequence of nodes defined. Any node may be used for any purpose. The main advantage of the single tier cluster is its simplicity. The main disadvantage is its limit to its expansion [4].
- Multi level or multi-tier clusters: In these types of clusters, there is hierarchical sequence between the nodes. There are node sets, where each set has a specialized function.

VI. EXAMPLES OF CLUSTER COMPUTING

Berkeley Network of Workstations (NOW)

- An important representative of cluster systems is the Berkeley Network of Workstations (NOW). The NOW project achieved over 10 Gflops on the Linpack benchmark, and in 1997 it made to one of the top 200 fastest supercomputers in the world.
- The hardware and software requirements for the project contained 40 SUN Sparcstations running Solaris and 100 SUN Ultrasparcs, PC Unix variant or 35 Intel PCs running Windows NT, and between 500 and 1000 disks, connected by switched network known as Myrinet.
- Sockets, MPI, a parallel version of C, called Split C are the programming environments used in NOW.
- The basic communication primitive in Berkeley NOW are Active Messages. It is a simplified remote procedure call that can be implemented efficiently on a wide range of hardware.

The Beowulf Cluster

- The main concept behind the Beowulf cluster project was to have supercomputer processing power with the help of off-the-shelf commodity machines. The earliest Beowulf clusters used to contain sixteen 100 MHz DX4 processors, connected using 10 Mbps Ethernet. Built in 1995, was the second Beowulf cluster, which used 100 MHz Pentium processors, connected by 100 Mbps Ethernet.
- The third generation of Beowulf clusters was built by different research laboratories. JPL and Los Alamos National Laboratory, each of the labs, built a 16-processor machine incorporating Pentium Pro processors. These machines were combined, to run a large N-body problem.
- The communication between processors in Beowulf was done using the TCP/IP over the Ethernet internal to the cluster. To satisfy higher bandwidth requirements, multiple Ethernets were also used. Channel bonding is a concept that connects multiple Ethernets, to distribute the communication traffic. Channel bonding was successful in increasing the sustained network throughput by 75% when dual networks were used.

FlashMob I

- A large number of computers, via a high-speed LAN, were connected to create a supercomputer that works together as a single supercomputer. This supercomputer is known as FlashMob.
- A FlashMob computer, not like an ordinary cluster, is temporary and organized in *ad hoc* manner for the objective of working on a single problem. It made use of volunteers and ordinary laptop PCs, and was designed, to create a supercomputer in few hours, by anyone.

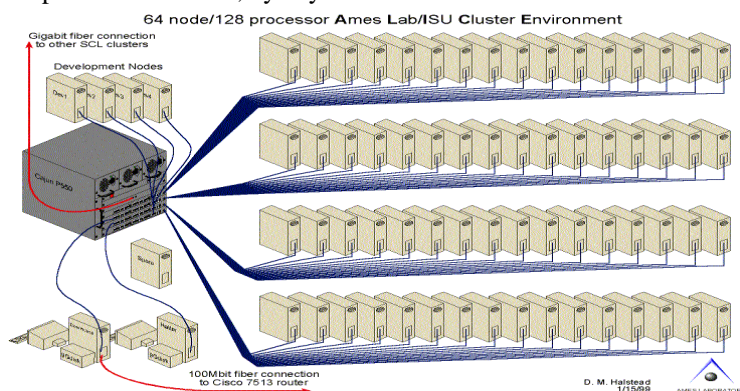


Fig 2: FlashMob I

ALICE

ALICE means Ames Lab - ISU Computing Environment. It is a collection of 64 dual-processor Pentium Pros, working at 200 MHz. All these PCs are connected with a central Fast Ethernet switch, providing a flat network topology. Moreover, there is a file server, a master node and 4 development nodes.

IBM Cluster

The IBM cluster is currently consisting of 22 dual-processor IBM RS/6000 43P-260 workstations and 2 quad-processor 44P-270 workstations, for 52 processors in total.



Fig3: IBM Cluster

22 IBM RS/6000 43P models 260 with:

- 4 MB of Level 2 cache per processor
- 2 Power3 processors running at 200MHz (PPC 64 bit compliant)
- Gigabit Ethernet on a PCI card of 64 bit
- 2 striped 9GB disks for local use (leading to ~16GB for scratch)
- 1 GB of RAM
- Integrated Fast-Ethernet

G4 Cluster

The G4 cluster is made of 16 single processor G4 computers each consisting of 16 dual processor G4s and 512 MB RAM, with 1 GB RAM. Each G4 runs a Black Lab Linux and is connected with Myrinet for communications and via Fast Ethernet for management purposes.



Fig 4: G4 Cluster

VII. ADVANTAGES AND DISADVANTAGES

1) Advantages of Cluster Computing

- **Economical:** Cluster Computing systems are up to 15 times cheaper than traditional supercomputers in respect to the same performance.
- **Scalability:** It is easy to upgrade and maintain.
- **Reliability:** It is reliable to work with, as the Cluster Computing system continues to operate even in case of failures in parts of it.
- **Reduced Cost:** The price of off-the-shelf consumer desktops has reduced, and the reduced cost has corresponded with the increase in their processing power and performance.

- Processing Power: The parallel processing power of a high-performance cluster can prove to be more cost effective than a mainframe with similar power as of cluster, in most cases.
- Improved Network Technology: Clusters are connected through a single virtual local area network (VLAN), and the network treats each computer as a separate node. The information could be distributed throughout these networks with very less delay, making sure that data doesn't stuck between nodes.
- Scalability: Mainframe computers have a fixed processing capacity whereas the computer clusters can be easily expanded as per the requirements change, by adding to the network some additional nodes
- Availability: When a mainframe computer fails, the entire system results in a failure. However, if a node in a computer cluster fails, its operations can be made to transfer to another node in the cluster, ensuring that there is no interruption in service.

2) Disadvantages of Cluster Computing

- It is difficult to manage and organize a large number of computers, for cluster computing
- In the case of non-parallel applications, it has poor performance.
- As compared to the single server, physical space needed is considerably greater.
- As compared to a single server, power consumption increases.
- Software Skew: - The biggest disadvantage is of skewness in software. Skewness means that, if the software configuration on some nodes is not similar to that on others, then these small differences can cripple a program.

VIII. CONCLUSION

High performance computing is a key role of computer science. Now a day, cluster computing is gaining much attention and is the most used high performance computing environment. Even though, the problems and issues related to non-parallelizable are still unsolved by cluster computers, the clusters provide a realizable and achievable solution, to most of the concerns that are related to high computing power.

REFERENCES

- [1] Dilek Demirel, "High Performance Computing Cluster Architectures", CMPE 511.
- [2] Rajkumar Kumar Buyya, Hai Jin, Toni Cortes, "Cluster Computing", Future Generation Computer Systems 18 (2002) v-viii
- [3] Mark Baker and Rajkumar Buyya, "Cluster Computing at a Glance".
- [4] Zheng, Y. H. Guo, W. Gao, B. Xu, Research on Load Balance of Multi Cluster Architecture Based on Business Components Partition, IEEE Proc. of 3rd Int. Conference on Information Technology and Applications, 2005.