INTERNATIONAL JOURNAL OF
RECENT TRENDS IN MULTIDISCIPLINARY RESEARCH
January-February 2022, Vol 2(01), 01-09. ISSN No: 2583-0368

Current Trends in Cluster Computing

Suneeta naik[1], Rakshita naik[2], Ananya naik[3], Subramanya G. Bhagwath[4]
[1,2,3] UG Students ,Dept. of Computer Science Engineering, Anjuman Institute of Technology and Management, Bhatkal, Karnataka, India.
[4]Assistant Professor, Dept. of Computer Science Engineering, Anjuman Institute of Technology and Management, Bhatkal, Karnataka, India.

Article Type: Research

OPEN ACCESS
Article Citation: Suneeta naik[1], Rakshita naik[2], Ananya naik[3], Subramanya G. Bhagwath[4], "Current Trends in Cluster Computing" International Journal of Recent Trends In Multidisciplinary Research, January-February 2022, Vol 2(01), 01-09.

Accepted date: Feb 25, 2022

Published date : Feb 28, 2022

Published by 5th Dimension Research Publication.

Abstract

High availability becomes a necessity for a growing number of business web applications. Clustering emerges as a natural solution to bring high availability to a large number of users. In this integration paper solutions for business web applications are updated. Achieving high availability and high capacity by using load balancing collections at the application server level and server clusters (including the RAC concept) at the website level is introduced. Collections are often used to improve performance and / or availability beyond what a single computer offers, while they are often more expensive than a single computer with high speed or availability. The main purpose of this collection is to use a large number of process nodes so when you finish your assigned task with a very small amount of your time by manual labor in the trailer. The most important strategy needed to achieve that goal is to move hundreds more from busy areas to inactive areas. Collection implementation involves setting up a server called MCLUSTER with configuration, resetting the collection. Generation of mobile application code and its distribution in appropriate client locations. Customer code receives and generates mobile code that describes the distributed activity sent to the MCLUSTER server and responds to the results. Integrated applications also help administrators remotely control and monitor the performance of each node and the general environment clusters may be a factor in their performance. There are a good number of merger applications on the market today. Most applications are geared towards a particular type of integration while there are those that are flexible enough to adapt to different types of integration. Choosing the right integration app depends on the need of the business.

1 Introduction

Cluster computing is a very important part in mainstream computing. Today's businesses are increasingly relying on widely available web applications for a large number of users. High availability of applications is required for critical business applications (applications found to be important for business operations: banking, finance, online marketing, booking) but also for business application available to a large number of users (each application is not a significant business, but a very large number of users make recreational costs are not acceptable).[1]One way to overcome this limitation is to improve the speed of processors and other components so that they can exceed the power required by computer applications.[2]A computer collection is a set of connected computers, which work closely together in many ways to type a single computer. The elements of the collection are notably noticeable, yet they are not always the same, connected in all the variations by the fast-paced traditional space networks. Cluster computing is an excellent feature because it combines a variety of ready-to-wear items with hardware integrated with hardware, networks, and computer code to behave like a single portable computer.[4]The most efficient set will usually have physically connected nodes. Nodes must be positioned to achieve maximum output.[5]

2. Cluster History

Cloud computing is associated with all computer rental services.This concept came in the early 1950s. Making clouds As of today, 5 technologies have played the main character. These are distribution systems and Peripherals, Virtualization, Web 2.0, Service Orientation and Utility computing [4]

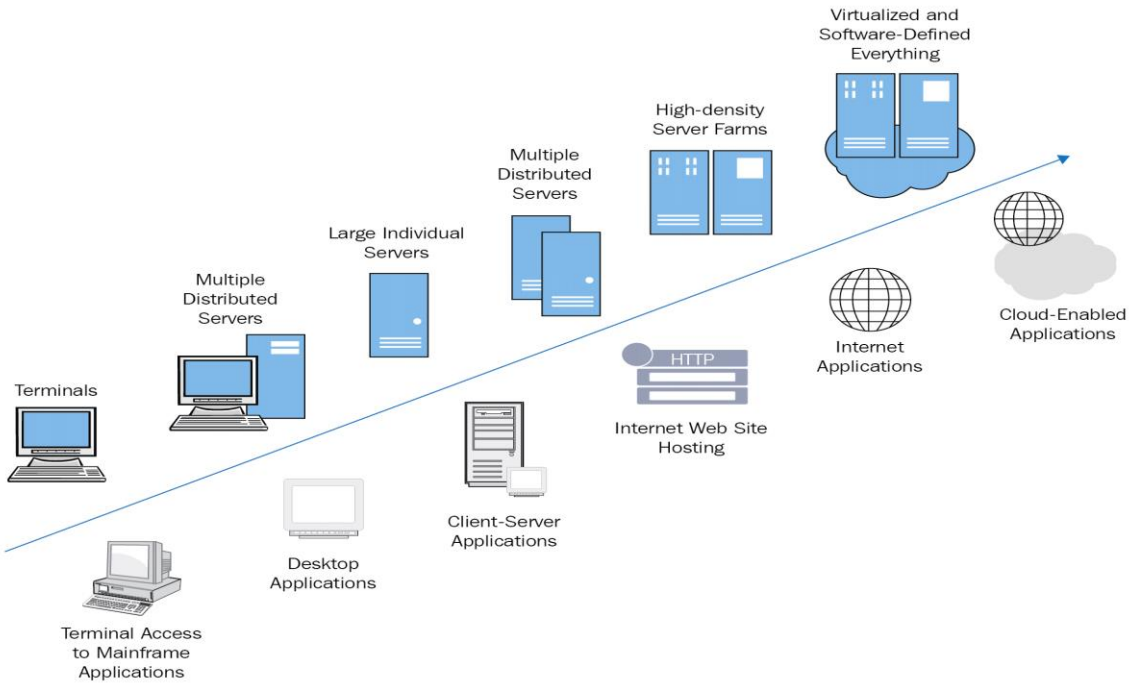

Figure 1: Evolution of cloud computing

3. Cluster Architecture

A computer node can be a single system or multiprocessor (PCs, workstations, or SMPs) with memory, I / O resources, and operating system. A collection usually refers to two or more computers (nodes) connected together. The nodes can be in a single cabinet or physically separated and connected via LAN. A cluster of connected computers (LAN-based) may appear as a single system for users and applications. shared memory programs in relation to. The typical structure of a collection is shown in Figure 2. [2]

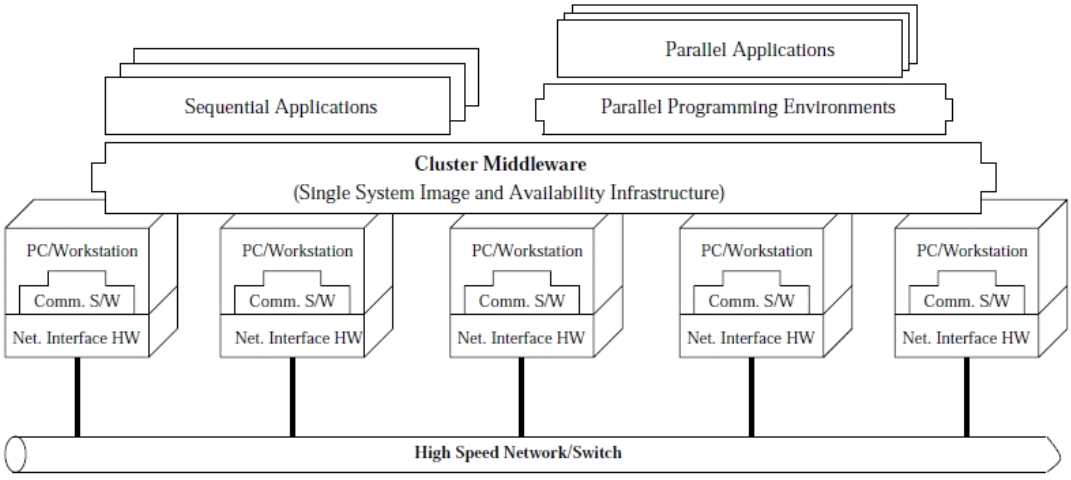

Figure 2: Cluster computer architecture

The communication code should provide quick and reliable means of communicating information between nodes and which may be out of the collection. for example, collections with a special network such as Myrinet use communication rules such as Active Messages to instantly communicate between their nodes. This hardware link outperforms the software and provides direct user-level access to the network, thus eliminating the importance of critical communication. [4]

4. Cluster Application

Collections have proved to be effective in a variety of data mining applications. The information mining approach involves each activity and the deep performance of the data. Collections offer 2 basic roles: Data collections and Integrated collections. [4]

Integrated applications also help administrators remotely control and monitor the performance of each node and the general environment clusters may be a factor in their performance. There are a good number of merger applications on the market today. Most applications are geared towards a particular type of integration while there are those that are flexible enough to adapt to different types of integration.

A. MPI as the Standard Clustering Application : Its simplicity and adaptability to different programming language has made MPI a highly favored clustering application.

B. Linux Clustering Applications : The open source community has also developed specific applications for clustering. Some of the recognized clustering applications are the following: • Linux-HA

• Beowulf
• MOSIX
• Kerrighed
• OpenSSI
• Linux Virtual Server

C. Windows Compute Cluster Server: The software giant Microsoft Corporation has also developed their own clustering application. This type of clustering application is closely linked to another product from the same company simply called Windows server.

D. Choosing the Right Application: Choosing the right clustering application really depends on the business need. However, administrators have to make sure that the clustering application should not be too much for the nodes to handle or would have functions that could be worthless. [5]

Compatible applications reflect a wide range of communication behaviors and place different needs on a basic network. This may vary from program to program, or program category depending on the requirements of the calculation procedures. The types of applications are: [3]

*Compute Intensive Applications
*Data or I/O Intensive Applications
*Transaction Intensive Applications

Cluster is also used in following applications: Web workers; Search motors; Email; Security; Proxy; Database workers. [4]

5. Cluster Classifications:

Clusters offer the following features at a relatively low cost:

*High Performance
* Expandability and Scalability
* High Throughput
* High Availability

Cluster technology allows organizations to expand their processing capabilities using standard technologies (hardware hardware and software components) that can be purchased at a relatively low cost. This provides an extension - an affordable development path that allows organizations to maximize their computing capacity - while maintaining their existing investment and without incurring many additional costs. Application performance is also improved with the support of expandable software environment. Another advantage of combining the power of a failover that allows a backup computer to take over failed computer functions found in its collection.

Clusters are classified into many categories based on various factors as indicated below.

A. Application Target - Computational science or mission critical applications.

*High Performance(HP)Clusters
* High Availability(HA)Clusters

The main focus of this book is on the HP collections and the technologies and conditions needed to be used on the same computer. However, we also discuss the issues involved in building HA collections with the aim of integrating performance and availability into a single system.

B.Node Ownership : Owned by an individual or dedicated as a cluster node.
*Dedicated Clusters
* Nondedicated Clusters

The difference between the two scenarios is based on the identity of the nodes in the collection. In the case of dedicated clusters, one person does not own the workspace resources are distributed so that parallel computing can be done throughout the cluster. Another unspecified case is that people who own workstations and applications are killed for stealing idle CPU cycles. The reason for this situation is based on the fact that many CPU cycles of the workplace are not used, even at very high hours. Parallel computing in a flexible set of unstructured applications is called adaptive parallel computing.

C.Node Hardware :PC, Workstation, or SMP.
*Clusters of PCs(CoPs)or Piles of PCs(PoPs)
*Clusters of Workstations(COWs)
*Clusters of SMPs(CLUMPs)

D.Node Operating System : Linux, NT,_ Solaris,_ AIX,_ etc.
*Linux Clusters (e.g., Beowulf)
*Solaris Clusters (e.g., Berkeley NOW)
*NT Clusters (e.g., HPVM)
*AIX Clusters (e.g., IBM SP2)

E. Node Configuration : Node architecture and type of OS it is loaded with.
*Homogeneous Clusters: All nodes will have similar architectures and run the same OSs. *Heterogeneous Clusters: All nodes will have di_erent architectures and run di_erent OSs.

F. Levels of Clustering : Based on location of nodes and their count.
*Group Clusters(#nodes:2-99): Nodes are connected by SANs(System Area Networks) like Myrinet and they are either stacked into a frame or exist within a center.
* Departmental Clusters(#nodes:10s to 100s)
*Organizational Clusters(#nodes: many 100s)
*National Metacomputers(WAN/Internet-based):(#nodes: many departmental / organizational systems or clusters)
*International Metacomputers (Internet-based):(#nodes: 1000s to many millions)[2]

6. Cluster Security

Collection safety is one of the most important things to consider during compilation. Node integration attacks can come in a variety of ways - it can be as simple as a virus where its sole purpose is to destroy files or it can be a powerful spy that can easily hijack node controls for malicious purposes. Only one error is needed to prevent the whole system from being corrupted. Whenever a network opens a connection to its controller, it automatically opens up to different types of attacks. This is also possible for users who are trying to access nodes and store data.

A. Network Tool Base: Software operators such as Microsoft Corporation and Sun Microsystems and Open Source communities have network tools that are suitable for use in a variety of integration applications. If a new network tool is used for integration, security may not be possible immediately. But with any network tool used, it is important to have the latest version. Different types of attacks were attacked in the last year only that an update is always needed to achieve complete protection.

B. Local Security Competitive Domain: With cluster-based security, the administrator can easily manage clusters as security is online. However, domain-based security can be easily compromised. Local security, on the other hand, prides itself on complete security by making administrator credentials accessible. It requires a lot of resources especially when they are configured for the first time. [5]

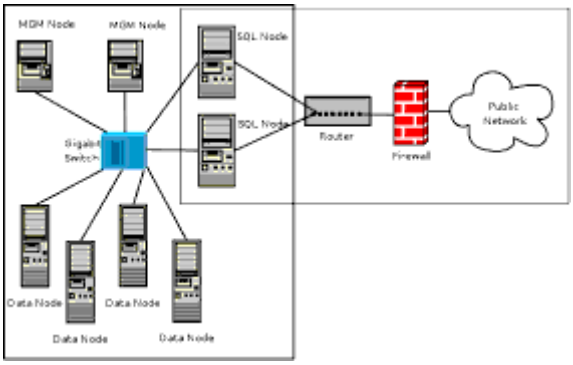

Figure 3:clustering security

7. Virtual Clustering

Virtual clustering is done by connecting Virtual Machines (VMs) into a cluster, Fig. 4. Virtual machine is a layer of software that simulates the operation of a particular machine or processor on a target machine. The advantage of using virtual nodes is that during hardware upgrades, all cluster nodes are always available. This can be achieved by moving live Virtual Cluster Node one from the Hosting System to be upgraded to another Hosting Program. The main features of virtual machines are VMware ESX, [20], Windows Virtual Server 2005 and the latest Oracle VM. Oracle VM is a virtualization server software that fully supports both Oracle and non-Oracle applications. Many Oracle applications are certified to run on Oracle VM. VM is also another important grid technology, [1].

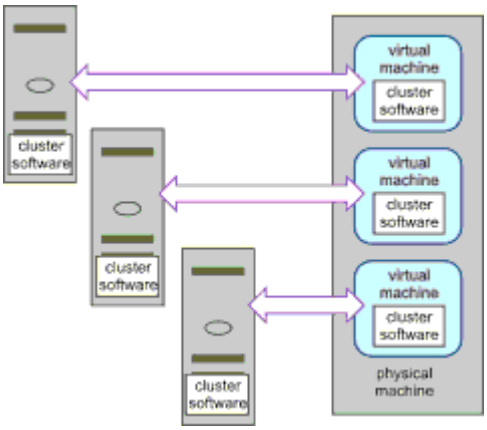

Figure 4: virtual clustering

8.Review Of Cluster Computing For High Available Business Web Application:

I. High Availability

Following are basic concepts related to availability, [1]:

A. Reliability:

The probability that a system will perform its required functions under stated conditions for a stated period of time at a given confidence level.

B. Mean time between failures (MTBF) :

The median time between failures is defined as the sum of the working life of an item during a particular measurement period, divided by the total number of failures between the population during that interval. The MTBF can be interpreted as the expected length of the system that will work between failures.

C. Mean time to repair (MTTR):

The current correction time is the total amount of repair time collected over a period of time divided by the total number of repair actions completed at the same time.

D. Availability:

Availability (natural), AI, the chances that the system will work properly over time; this calculates system operating time and configuration and does not include downtime related to block maintenance,

configuration, and administration. Leisure time includes both scheduled and unplanned periods. Scheduled downtime includes scheduled maintenance tasks, hardware additions and software to improve software availability and system development. Random downtime includes system failure, operating system crash, application failure due to application interruptions, unforeseen events such as viruses or power outages and human errors such as unintended application or system reset.

E. Five levels of high availability :

In Fig. 5,is presented stair-step solutions model to describe high availability across a range of levels, . Solutions model ranges from simple to complex, from single systems to multiple systems, from single sites to multiple sites.

- Level 1 - Single system (ECC memory, Hot plug, redundant NICs)
- Level 2 - Single system (RAID Controllers and Drive Array)
-Level 3 - Cluster Fault Resilient (failover cluster)
Level 4 - Cluster Shared Services (eg. parallel database, Real Application Cluster - RAC)
- Level 5 - Multi-Site Cluster (geographically dispersed cluster)

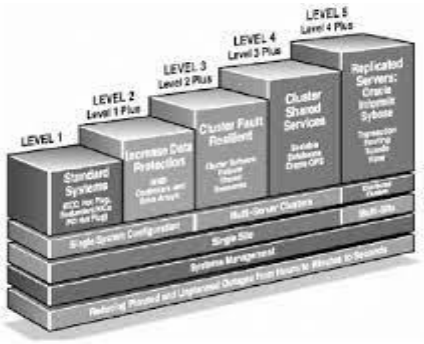

Figure 5:Five levels of high availability

II. High Capacity

In addition to providing high quality available collections that are supposed to provide the service with a large number of users. Determining which hardware configuration will best meet the requirements of your application is a process known as power setting. Collecting precise performance requirements is an important part of the energy planning process. The goal of energy planning is to provide satisfactory service levels to users in a cost-effective manner. One must predict in advance the maximum number of fellow users and the acceptable response time.

A. Number of servers in cluster :

When using cluster computing total number of users is generally sum of users per servers.

B. Response time :

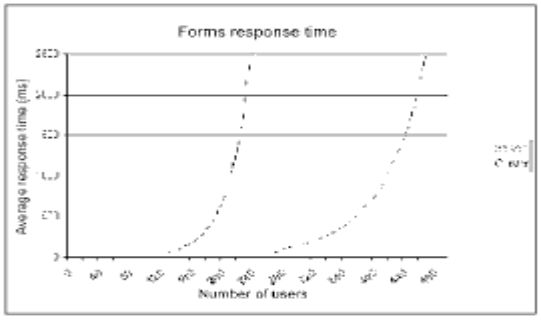

Figure 6: Response time for user load, on single server and cluster

Response time is a performance measure defined as the round-trip delay to process a client request. An acceptable average response time may be defined as twice the average response time for one user.

C. Behavior under load of users :

Operation usually remains unchanged until the machine is full: the “hockey stick” or “knee” point on the graph. Once the filling point is reached, performance decreases significantly during an unacceptable user response time. When specifying a number of servers one should consider the maximum acceptable number of users for each server with a valid user response time and sit to the left of the "hockey stick", Figure 6.

III. Cluster Computing

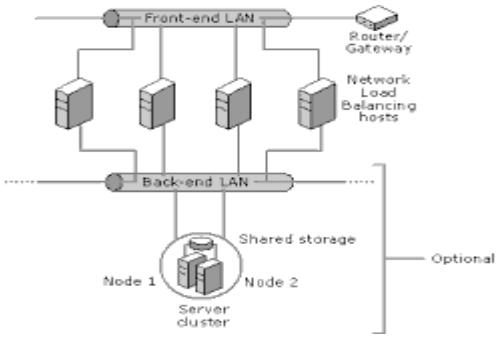

Figure 7 : Clustering applied to second and third tier

Iv. Load Balancing

Load balancing is a method of distributing work between two or more computers in order to obtain full service usage. It supports multiple live components simultaneously and without limiting user uploads provides a way to use the failure method, e.g. the service continues to operate despite the failure of one or more device components. Load scales are available in hardware and software versions.

A. Hardware load balancers :

Hardware load balancers are dedicated piece of hardware for partitioning of network traffic, commonly integrated with switch/router, Fig. 8.

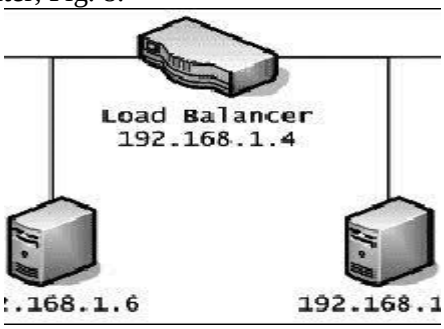

Figure 8: Hardware load balancer

B. Software load balancer :

Software load balancer is either dedicated server with load balancing software or solution built into operating system (eg. Windows NLB).

Windows Network Load Balancing (NLB), Fig. 8, is a clustering technology offered by Microsoft as part of Windows 2000 Advanced Server and Windows Server 2003 family operating systems. To scale performance, Network Load Balancing distributes IP traffic across multiple cluster hosts. It also ensures high availability by detecting host failures and automatically redistributing traffic to the surviving hosts.

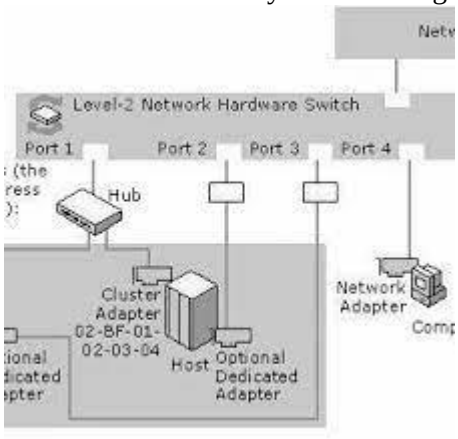

Figure 9: Windows NLB

V. Server Cluster

Server clusters can be realized in two main configurations:

A. Active/passive clustering

One node in the cluster remains inactive, while the other node (or nodes when using Datacenter Server) is active. If the active node fails, the processing of group recognition applications will be switched to passive node (failover). Once the failed node is restored, the application can revert to the original location, becoming the active node again (failback). The main effect of functional / functional groups is the costs associated with having a second system that stays idle.

Microsoft SQL Server is a relational database management system (RDBMS) produced by Microsoft and example of active/passive cluster (Fig. 10].

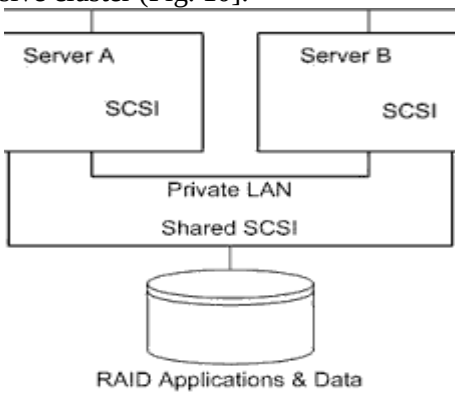

Figure 10: Microsoft SQL Server

B. Active/active clustering :

All servers run their own function simultaneously. Every computer in the collection is available for actual operation (active), and each computer in the collection is also available for restoring the resources and function of any other computer in the collection. There is no need to have a second, idle server waiting to fail. A drawback with active / active clustering is the risk of overloading the node replacing the failed node because it now has to perform its own function and that of the failed node.

Oracle RAC, Fig. [11], is a collection site with a shared archive that overcomes the limitations of standard non-shared and disk-shared methods to provide the most comprehensive and available web solutions for all your business applications. Oracle RAC is an example of a functional / functional group. It also enables customization by simply adding servers to the collection. [1]

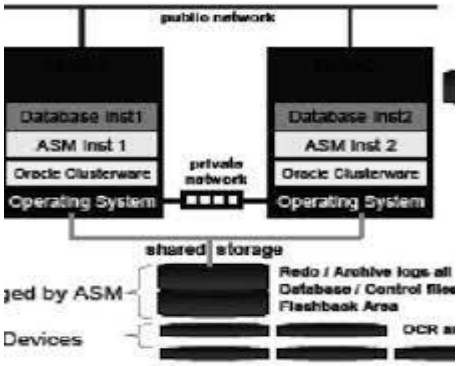

Figure 11: Oracle Real Application Cluster (RAC)

9. Future Cluster Technologies

Emerging hardware technologies and software development tools mean that cluster-based systems quickly close the operating gap with dedicated dedicated computer platforms. Cluster systems eliminating cycles of inactivity on PCs and workplaces will continue to use any hardware and software components available in public workplaces. Collections dedicated to advanced applications will continue to emerge as new and more powerful computers and network connectivity become available in the marketplace. Individual cluster nodes may be SMPs. Currently two PCs and four processors and UNIX workstations are common.

Software that allows SMP nodes to be effectively and efficiently compatible applications will be

upgraded and added to the OS kernel soon. It is likely that there will be widespread use of Gigabit Ethernet and, thus, will be the de facto standard for clusters. In order to reduce the latency of system software applications will bypass the OS kernel, thus avoiding the need for expensive system phones, and utilizing the use of smart network cards. This can obviously be achieved using smart NICs, or alternatively used on chip network interfaces such as those used by the new DEC Alpha 21364.

The ability to provide a rich set of development tools and resources as well as the provision of robust and reliable services will determine the choice of OS to be used in future collections. UNIX-based OS may be very popular, but further development and adoption of Windows NT will mean that it will not be lagging behind.

10. Conclusion

Our need for accounting resources in all fields of science, engineering and commerce measures our ability to meet these needs. The use of computer collections, perhaps, is one of the most promising ways to close the gap between our needs and the resources available. The use of COTS-based collection systems has many benefits including:

* Price / performance compared to a large dedicated compatible computer.
* Increasing growth often accompanied by annual funding patterns.
* Provision of a multi-purpose system: which, for example, may be used for secretarial purposes during the day and as an asset associated with supercomputing at night.

These and other benefits will promote the emergence of cluster computing and its adoption as a means of providing material supercomputing resources. [2]

Cluster computing is undoubtedly gaining importance instead of the more expensive Parallel computers. Cluster computing by properly combining the computational processing power of standalone computers can greatly improve the performance of distributed systems. On the other hand, Cluster Computing may damage the functionality of a problem if the appropriate check is not used for the problem size and component size. However, the Network Congestion problem can be easily solved by using the Mobile Agent concept

Reference:

1. Dubravko Miljković Hrvatska elektroprivreda d.d. (HEP d.d.) , "Review of Cluster Computing for High Available Business Web Applications" , May 2008.

2.Mark Bakery and Rajkumar Buyyaz, Division of Computer Science University of Portsmouth Southsea, Hants, UK, School of Computer Science and Software Engineering Monash University, Melbourne Australlia,"Cluster Computing at a Glance".

3.KUMAR KAUSHIK and COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY KOCHIN-682022,"Cluster Computing",August 2008.

4.Trupti Bhor, Yogeshchandra Puranik ,"International Journal of Trend in Scientific Research and Development (IJTSRD)" , "An Introduction to Cluster Computing ",May-June 2021.

5.Ms. Arshpreet Kaur and Ms. Manpreet Kaur Mann ,"International Journal of Computer Science and Its Applications", "Cluster Computing".