ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

Survey on High-performance Cluster Computing

Prof. Vinayak Shinde[1], Prajakta Gosavi[2] , Ankita Doiphode[3]

[1]Assistant Professor, Dept. of Computer Engineering, L. R. TIwari College of Engineering, Mira Road (E), Thane, Mumbai University, India

[2,3]Research scholars, Dept. of Computer Engineering, L. R. TIwari College of Engineering, Mira Road (E), Thane, Mumbai University, India.

ABSTRACT: A computer cluster is a collection of linked computers, working together so that they form a single computer. The components of a cluster are frequently, but not always, connected to each other through local area networks. Clusters are formed to improve performance and/or availability provided by a single computer. The main objective in the cluster is utilizing a group of nodes so as to complete the assigned job in a minimum amount of time by working cooperatively. The main strategy to achieve such objective is by transferring the extra loads from busy nodes to idle nodes.

KEYWORDS: Cluster computing, parallel computing, cluster architecture, nodes, tightly coupled.

I. INTRODUCTION

Parallel computing has become the paradigm of choice for executing large-scale science which seen many changes since the days of the highly expensive as well as proprietary super computers. The changes and improvements in performance seen in the area of mainframe computing for lots of environments. But these compute environments may not be the cost effective as well as flexible solution for a problem. Over the past decade, cluster technologies have been enhanced that allow include several part of slow cost computers to work in a coordinated fashion to process applications.

In the economics, performance and flexibility of compute clusters which makes cluster computing an attractive alternative to centralized computing models as well as the attendant to cost, inflexibility, and capability issues inherent to these models or systems. Many enterprises are now onward looking at clusters of high-performance[1], low cost computers which provide increased application performance, high availability, and ease of scaling within the data center deployment of computer clusters has been driven by the increase in the performance of off-the-shelf commodity computers, high-speed, low-latency network switches i.e. how much time it takes for a particular packet of data to get from one designated point to the another point and the maturity of the software components. Application performance continues to be of significant concern for various entities including governments, military, education, scientific and now enterprise. This document provides a review of cluster computing, the various types of clusters and their associated applications. Basically it contents a high level informational document; it does not provide details about different cluster implementations and applications.

II. RELATED WORK

GPFS is IBM's parallel, shared-disk file system for cluster computers, available on the SP parallel supercomputer and on Linux clusters. GPFS is used on many of the largest supercomputers in the world. GPFS was built on many of the ideas that were developed in the academic community over the last several years, particularly distributed locking and recovery technology. To date it has been a matter of conjecture how well these ideas scale. While in many cases existing ideas scaled well, new approaches were necessary in many key areas. Paper by Frank Schmuck and Roger Haskin describes GPFS, and discusses how distributed locking and recovery techniques were extended to scale to large clusters[2].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

Cluster computing is effective within the scientific community it is essential that there are numerical libraries and programming tools available to application developers. Cluster computing may mean a cluster of heterogeneous components or hybrid architecture with some SMP nodes. This is clear at the high end, for example the latest IBM SP architecture, as well as in clusters of PC-based workstations. It is observed by Browne, Dongarra & Trefethen that these systems may present very different software environments on which to build libraries and applications, and indeed require a new level of flexibility in the algorithms if they are to achieve an adequate level of performance[3].

III. ARCHITECTURAL RESULT OF CLUSTER COMPUTER

In Parallel Computing world, a computer in the form of cluster consists of a group of loosely or tightly connected computers that work together so that, they can be states as a single system. As compare to grid computers, clusters in the form of computer have each node set to perform the similar task, controlled and scheduled by software. Loose coupling of system means reducing interdependencies between component of system so if change in one component system doesn't affect in another component of system, this feature increase the flexibility of systems whose connected in clusters and tightly coupling means A Tightly Coupled of system, is an object that needs to know quite a bit about other objects and are basically highly dependent on each other's interfaces.

A cluster is a kind of parallel or distributed computer system, which consists of a collection of inter- connected stand-alone computers which are working together as a single integrated computing resource. The typical architecture of a cluster is depicting as in Figure 2. The key components of a cluster include multiple standalone computers (PCs, Workstations, or SMPs), operating systems, high-performance interconnects, middleware, parallel programming environments, and applications.

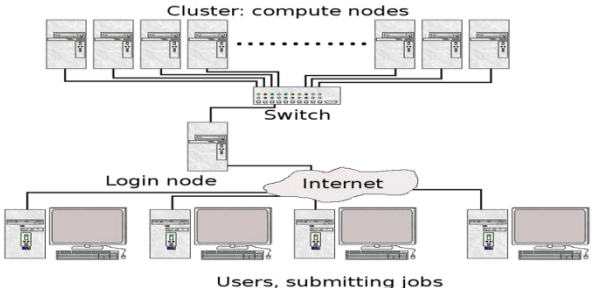

Figure 1 Cluster example diagram[4]

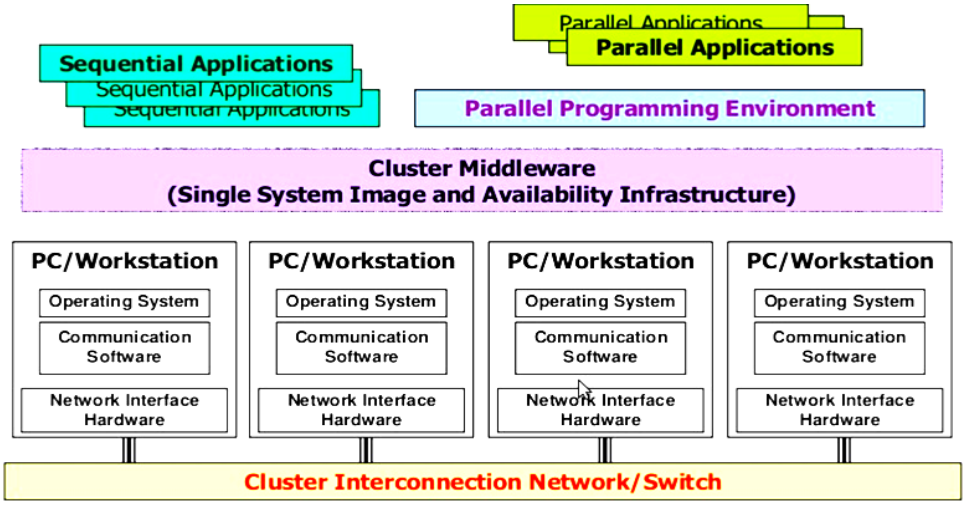

Figure 2 Cluster Architecture[5]

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

IV. ADVANTAGES OF CLUSTER COMPUTING

(1) Manageability: It takes so much effort, cost and money to manage a large number of components. But, with cluster, collection components are combined to work as a single entity. So, management becomes easy.

(2) Single System Image: In this user gets the feel that he is working with a single system, but actually he is working with a large number of components. There is no need to worry about different components, he only needs to manage a single system image.

(3) High Availability: As we know components are replicas of each other, so if one component goes down because of ay technical issue, then some other component can takes its place, and user can continue to work with the system.

V. DISADVANTAGES OF CLUSTER COMPUTING

(1) Programmability Issues: If the components are different in terms of software from each other, and then there may faces some problems when combining all of them together as a single entity.

(2) Problem in Finding Fault: As we are dealing with a single entity, problem may arise when finding out fault that which of the component has some problem associated with it.

(3) Difficult to handle: As cluster computing involves merging different or same components with different programmability, so a non-professional user may find it difficult to manage[3].

VI. CLUSTER APPLICATION

Clusters have been paradigm of choice as an execution platform for a range of application classes, ranging from super-computing and database-based ones. This is due to Clusters are being used as execution environments for huge Challenge Applications such as weather forecasting , automobile crash simulations, life sciences, computational fluid dynamics, nuclear simulations, data mining etc. Currently these applications considered intractable without the use of state-of-the-art parallel supercomputers.

The availability of their resource requirements, such as processing time of system, memory, and communication which needs distinguishes huge challenges applications from other applications. For example, the executing scientific applications used in assuming a serious turn of situations such as earthquakes or hurricanes requires enormous computational power as well as storage resources.

In the decades, these applications would have been run on parallel supercomputers which were costing millions of dollars in order to calculate assumption in advance of the actual events. Such applications can be migrated to run on commodity off-the-shelf-based clusters and deliver comparable performance at a much lower cost. In reality , in different situation expensive parallel supercomputers have been replaced by low-cost commodity Linux clusters in order to reduce maintenance costs as well as increase overall computational resources.

Clusters are increasingly being used as only for running commercial applications. In a business environment, for example, suppose in a bank, many of its activities are performed simultaneously. However, a problem will occurred if the server that is handling customer transactions fails. The bank's activities could come to halt as well as customers unable to deposit or withdraw money from account. Such situations can be responsible for a great deal of inconvenience which results in loss of business and confidence in a bank.

This is where clusters can be very much useful. A bank could continue to operate even after the failure of a server by automatically isolating failed components and migrating activities to alternative resources (i.e. virtual server) as a means of offering an uninterrupted service. Clusters are basically used to host lots of new Internet service sites. For example, free email sites like Hotmail, and search sites like Hotbot use clusters. Cluster-based systems can be used to execute as many Internet applications: Web servers, Search engines, Email, Security, Proxy and Database servers.

In the commercial area these servers can be used to create what is known as an enterprise server. The servers can be optimized, tuned, as well as managed for better increased efficiency and responsiveness depending on the workload through various load balancing techniques. A large number of binary language machines (PCs) can be clustered along

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

with storage and applications for the purpose of scalability, huge availability, and better performance. The leading companies building these systems are Compaq, Hewlett-Packard, IBM, Microsoft and Sun. The Linux Virtual Server is nothing but a cluster of servers[6], connected by a fast network. From the prospective of end user it provides a viable platform for building scalable, cost-effective and a more reliable Internet service than a tightly coupled multi-processor system since any failed components can be easily isolated and the system can continue to operate without any disruption.

Another example, The Linux Virtual Server access clients' network connection requests to the different servers as they are scheduled in sequence and makes the parallel services of the cluster appear as a single virtual service with a single IP address. Prototypes of the Linux Virtual Server have already been used to modeled many sites that cope with heavy loads, such as Linux portal. Client applications interact with the cluster as if it were a single server.

The clients doesn't affected by the interaction with the cluster and do not need any modification. The applications performance and capability is enhanced by adding one or more nodes to the cluster, by automatically detecting node which cause in failures and by re configuring the system appropriately to achieve high availability.

REFERENCES

1. Yeo, Chee Shin, et al. "Cluster computing: high-performance, high-availability, and high-throughput processing on a network of computers." *Handbook of nature-inspired and innovative computing*. Springer US, 2006. 521-551.
2. Frank Schmuck and Roger Haskin, "GPFS: A Shared-Disk File System for Large Computing Clusters", *Proceedings of the Conference on File and Storage Technologies (FAST'02), 28–30 January 2002, Monterey, CA, pp. 231–244. (USENIX, Berkeley, CA.).*
3. Browne, S., Dongarra, J., & Trefethen, A. (2001). Numerical libraries and tools for scalable parallel cluster computing. *Int. J. High Perform. Comput. Appl*,*15*(2), 175-180.
4. Kaur, Kiranjot, and Anjandeep Kaur Rai. "A comparative analysis: Grid, cluster and cloud computing." *International Journal of Advanced Research in Computer and Communication Engineering* 3.3 (2014): 5730-5734.
5. Jin, H., Buyya, R., & Baker, M. (2001). Cluster computing tools, applications, and Australian initiatives for low cost supercomputing. *MONITOR Mag.(The Institution of Engineers Australia)*, *25*(4).
6. Ross, Robert B., and Rajeev Thakur. "PVFS: A parallel file system for Linux clusters." *Proceedings of the 4th annual Linux Showcase and Conference*. 2000.

BIOGRAPHY

Prof. Vinayak Shinde is Assistant Professor in Computer engineering Department, L. R. Tiwari College of Engineering, Thane. His research interest is in Parallel Computing.

Prajakta Gosavi is a Research scholar in the Computer engineering Department, L. R. Tiwari College of Engineering, Thane. Her research interest is in parallel, cluster computing, grid computing etc.

Ankita Doiphode is a Research scholar in the Computer engineering Department, L. R. Tiwari College of Engineering, Thane. Her research interest is networking, database etc