ACCOMPLISHING HIGH PERFORMANCE DISTRIBUTED SYSTEM BY THE IMPLEMENTATION OF CLOUD, CLUSTER AND GRID COMPUTING

Er. Vijay Malav[1], Er. Govind Dev Lodha [2], Dr. Amit Sharma [3]

[1,2,3] Associate Professor Department of Computer Science & Engineering ,
Vedant College of Engineering & Technology ,Bundi ,Rajasthan,(India)

ABSTRACT

In this research Paper I am focusing on the To expand the effectiveness of any assignment, we require a framework that would give superior along adaptabilities and cost efficiencies for client. Disseminated figuring, as we are all mindful, has turned out to be extremely well known over the previous decade. Appropriated figuring has three noteworthy sorts, to be specific, group, network and cloud. Keeping in mind the end goal to build up an elite circulated framework, we have to use all the aforementioned three sorts of processing. In this paper, we should first have a presentation of all the three sorts of dispersed figuring. Along these lines analyzing them we might investigate inclines in processing and green supportable registering to improve the execution of a disseminated framework. At long last introducing the future degree, we close the paper proposing a way to accomplish a Green superior circulated framework utilizing group, lattice and distributed computing.

Keywords-component: Cluster Computing, Cloud Computing, Distributed Computing, Grid Computing, High Performance Computing.

I. INTRODUCTION

The paper basically focuses on the various factors that would provide high performance computing environment in distributed systems. High performance computing is narrowly defined as the development and use of the fastest and most powerful computing systems i.e., potential computing. It covers technological, political and economic features of the distributed computing enterprise. The major findings and recommendations on the subject have been summarized in the later sections of the paper.

1.1 Cluster Computing

All along the archetypal generation of computing, projects prescribing immense estimations and ample processing were dependent on jurisdiction or sprinkling corporative. Such hulking supercomputers and frameworks were exorbitant for singleton. Though expenses of PCs are thumbing down, supercomputers are still out of sight. As a result of which Donald Becker and Thomas Sterling imported Beowulf clustering in1993 [1] which lit off the counter computers, building a cluster that emulated dormant supercomputers. The paramount behind this is to create a computing arrangement for providing the necessary processing power at a nominal cost. As the nodes are repository of processors, security is absolutely airy and thence alertness in confining interconnected networks from external networking. Admitting considerable expedient in computing power, clustering certainly has hitches and hesitancies as a comparably newfangled technology. Distributed computing administers to beset a distended sphere of clustering by permitting the nodes to prevail all over the world and also be multiuse machines. Distributed computing has an analogous notion as clustering, allowing many nodes to work on large problems in parallel after breaking them into smaller units. Innumerably work units are distributed several times to too many nodes, curbing the probabilities of processing lapses and narrate for processing done on tedious CPUs. The client supervises the data resurgence and capitulation laps along with the code essential to order the CPU how to routine the work unit.

1.2 Grid Computing

Amalgamation of computers from different managerial realms to achieve a bourgeois objective, coined as grid computing, can be mapped to late 1980‟s and early 1990‟s[2]. Adoption of middleware, software that associates software peripherals and venture utilities, is a preeminent scheme of grid computing to segment and allocate fragments of a program amid numerous computers. It embraces computation in a distributed manner, gathering colossal clusters. Harmonizing applications on grids can be a convoluted job, chiefly while managing stream of instructions across distributed computing assets. Grid workflow systems are refined as a functional form of workflow management system framed distinctively to construct and accomplish an array of ciphering or data handling steps.

1.3 Cloud Computing

Genesis of the term „cloud computing‟ is ambiguous, although it sounds to be borrowed from the habit of employing sketches of clouds to symbolize networks. The custom of catering remote connection to computing activity through networks contributed to prevalent usage of this caption. Cloud computing cite to an exemplary of network computing where a program or utility executes on a connected servers instead of confined computing apparatus. Corresponding to the conventional client-server or mainframe model, a node associates with a server to accomplish a job.

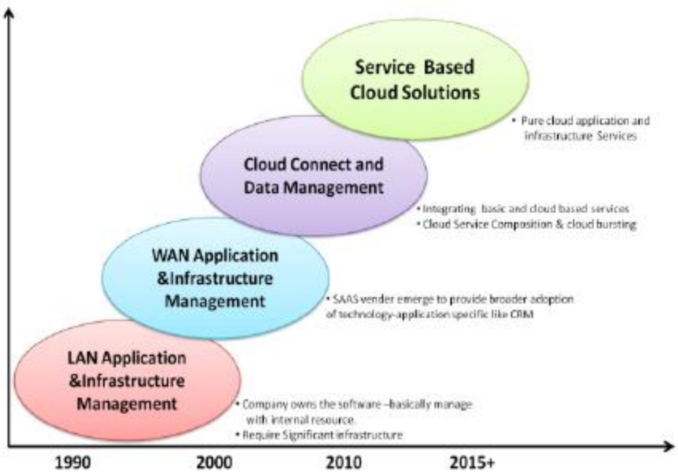

Figure 1: Year wise expansion of Cloud

The digression with cloud computing is that the computation may be executed on a single or many linked nodes at the same instance, applying the notion of virtualization. Virtualization allows multiple servers to be designed and distributed among several autonomous „virtual‟ servers, operating separately seeming to the node to be a single device. These virtual servers are core, extensile, mountable and un-mountable, un-influencing the nodes.

II. CURRENT DEVELOPMENTS AND LIMITATIONS

In this section, we describe components of the Grid infrastructure that address the management of application data and resource information, as they are the focus of our discussion in the rest of this paper. The infrastructure that focuses on management of distributed application data is commonly labeled a *Data Grid*. An increasing number of scientific disciplines manage large data collections generated by measurements and derivation of measurement data. As a result, many Data Grids are currently being deployed.

Infrastructure targeting resource information is often referred to as a *Grid Information Service.* A number of research groups have designed and prototyped components for collecting, indexing, and publishing Grid information. The problems of indexing, discovering, and accessing such "Grid information services" is in some respects quite similar to those encountered when indexing, discovery, and accessing other data sources. However, we will see in the rest of this paper that both infrastructures raise a number of distinct research questions from a distributed computing perspective. For both infrastructures, appropriate *data schemas* must be defined so that information can be encoded, stored, and searched in an efficient manner. A number of recent developments have made contributions in that area. In the Data Grid context, the Chimera system targets a data schema that can be used to establish a *virtual data catalog* that describes all ways in which data in the catalog has been derived. This is a generic solution that should be applicable to many different VOs and has been demonstrated for high-energy physics and astronomy applications. In the context of Grid Information Services, schemas are being developed for various Grid resource types as part of the GGF activities in the Grid Information Services working group . Commonalities with Common Information Model (CIM) are also being explored.

The definition of schemas is an important, but in some sense mundane, issue. More challenging is the design and implementation of a distributed system that implements mechanisms to publish information, disseminate information, notify participant of information changes, locate information, and retrieve information. Initial Grid infrastructure efforts have engineered software solutions for those mechanisms.

Worldwide, issues such as scalability and usability are becoming a near-term concern. These issues are being increasingly recognized by the Grid computing community and recent work explores avenues of research that are strongly connected to distributed systems and distributed computing research questions.

In that sense, Grid computing presents a key opportunity for distributed systems and distributed computing researchers. Grid developers are implementing large scale infrastructures such as GriPhyn as this paper is being written, and those infrastructures provide a great "playground" to explore research issues in a concrete setting that will have a major impact on disciplinary science. Furthermore, information dissemination techniques developed in the distributed systems community (e.g. wide-area group communications)

III. TRENDS IN HIGH PERFORMANCE COMPUTING

Since 60 years, the field of computing has undergone rapid transformations. Despite this, the long term evolution of performance seems to be stagnant. Massively Parallel Processor (MPP) systems are being accepted for engineering as well as for new commercial applications. At the onset of 1990s, MPP systems came into the market claiming vector multiprocessors as pushovers. Top500 list, to provide a more reliable statistics on high performance computers, in June 1993 declared that 156 MPP systems were employed already.

Based on the present Top500 data and assumption that the current rate of performance improvement would continue in future, we can hypothesize the observed conduct and analyze these values with the intention of government programs such as the department of Computing and Communications, and PetaOps initiative. Considering that in 2005, no small system made it to the Top500 ranking. First PetaFlop/s were available around 2009 and rapid changes were adopted in technology used in high performance computing devices, but still there is currently no reasonable image possible for architecture of the PetaFlops systems at the end of the decade.

IV. DESIGN TECHNIQUES FOR GREEN SUSTAINABLE COMPUTING

Many advanced methodologies for enhancing energy efficiency of IT(Information Technology) and making it more feasible involve the need to dynamically accustom computation to the appropriate energy profile. Complex distributed computing environments provide an array of opportunities for managed adoption among multiple nodes and at multiple levels. Cluster EAC (Energy Adaptive Computing) is a model that requires a significant computation employing multiple servers before response for a request submitted by a client can be returned. This implies client‟s role is minimal and adaption of energy is of prime concern in data center infrastructure. In EAC cluster, the adaption of energy happens at multiple levels, with power limit that a level needs to adapt. Client-server EAC requires to be handled with a well-managed end-to-end adaptation including the client, server and the interceding network. The motive of management is to elevate the client experience within the energy limits. As the client becomes more mobile and demand richer capabilities, the limited battery capacity gets in the way causing hindrance to energy adaptations Cloud techniques have been proposed to outsource mobile computation to cloud platform that can make the required resources on demand available . Energy adaption in P2P (Peer-to-Peer) environment requires cooperation among peers. This issue has been examined and the solution proposed is an energy adaptive version of Bit Torrent protocol . The various issues for realization of EAC are: Hierarchical Power Control, Demand side adaption, supply side adaption, and QoS (Quality of Service) aware scheduler.

V. PROJECT ANALYSIS

In order to understand and achieve desired high performance system, we must be aware of features of specified computing techniques. To make study and analysis of these systems more understandable and communicable, this section describes and examines some famous projects in these fields. As per the framework of the paper, the discussion begins with the exploration of THE LATTICE PROJECT . The Lattice project being a research project in Grid computing, targets unification of computing resources into a grid computing system to make the

International Journal of Advance Research in Science and Engineering
Vol. No.5, Issue No. 04, April 2016
www.ijarse.com
IJARSE
ISSN 2319 - 8354

resources systematically accessible and recognizable. To achieve something high we must recognize the reason as to why that height is required. In case of Lattice, the ever increasing complexity and size of data calls for increase in computational efficiency. To attend this call, lattice project requires amalgamated computational resources. Since divergent computational resources are used, software development is in a dire need of this project which it slakes by employing open source equipments. The development of user interface along with increased scalability engrossed humongous effort. Lattice project engaged features of grid for accessing large-scale resources and modeled it over the scope of personal computers. Architecture born out of fusion of Globus and BOINIC was adopted in various applications and projects like, BLAST (Basic Local Alignment Search Tool), ClustalW, IM (Isolation with Migration) Et. Al. Researchers in Asia –Pacific region have been trying to solve numerous issues and problems linked to cluster computing, like employment of multithreaded DSM runtime systems; reduction in network overheads and communication patterns; development of realist communication models; distributed and parallel file systems. Various researches and papers presented in this regard have been recorded marvelously by Mark Baker, RajkumarBuyya, Ken Hawick, Health James and Hai Jin in their paper Cluster Computing R&D in Australia.

Google App Engine and Amazon Web Service are two leading cloud platforms. The comparative study of these platforms by Chao He under the guidance of Prof. Raj Jain revealed that neither of them was significant in successful round trip time and throughput facets of cloud computing. Also it was concluded that none of these two was superior to the other. Cloud computing can get judicious performance in comparison to the traditional web servers depending upon the service delivered. The experiments conducted by them provided a better insight on the methodology to construct the cloud computing infrastructure and platforms.

VI. FUTURE PROSPECTS OF DISTRIBUTED COMPUTING

Increasing power and speed data centre is not always efficient and sometimes leads to an additional cost, so one should not expect to increase the efficiency more than a required limit. Distribution of data centers and use of closest data centers is a better and a far more optimal choice. It has been predicted that storage and computing on personal computers will be forgotten and transferred into distributed clouds. Therefore, architecture and evaluation of data centers should be performed for future of computing through suitable prediction. According to review and evaluation performed in the field of high performance computing, high performance distributed computing through grid, cluster and cloud still has a shortage in performance evaluation and special measures are required for this work. It is better to consider delay in evaluations or implement a criterion for evaluation of service level agreement because these agreements are most important for the users and one can present more accurate evaluation in future by specifying type of user‟s requests or specifying and distinguishing all users. High performance embedded computing (HPEC) systems are amongst the most challenging systems in the world to build. The primary sources of these difficulties are the large number of constraints on an HPEC implementation:

I **Performance:** latency and throughput.

II **Efficiency:** processing, bandwidth, and memory.

III **Form Factor:** size, weight, and power.

IV **Software Cost:** code size and portability.

Thus in future, we hope to achieve high performance distributed system by combining best features of grid, cluster and cloud computing as well as reconfigurable computing. Besides the aforementioned requirements, the emergence of Jungle Computing has given a boost to the field of Distributed Computing. It uses a system which is distributed, is highly diverse and provides computing at very high speeds. But the fact that it is highly non-uniform is viewed as a hindrance if not handled properly. There is an urgent need for easy and efficient Jungle Computing in scientific practice, by exploring a set of state-of-the-art application domains. Thus, the need of an hour is a system which not only combines the features of grid, cloud and cluster computing but goes beyond it to incorporate efficient jungle computing, thus providing an easier and faster system.

VII. CONCLUSION

We discussed origin of cluster, grid and cloud computing and studied their architecture, characteristic features and discussed their current applications and fields of implementation. Further we had an overview on trends of computing and glimpse of green sustainable computing which allowed us to create intent of developing a high performance distributed system which would meet the aim of green sustainable computing and would combine best features of all the available computing models, especially the most popular ones as per trends in computing. In the nutshell, we conclude that by extrapolating trends in high performance computing we draw the conclusions that parallel computing is the core mechanism by which computer performance can cope up with the predictions of Moore's law in the face of increasing influence of performance and the architecture of HPC will continue to develop at quick rates. Thus, it would be increasingly important to find paths to motivate scalable parallel programming without compromising with transportability. Such a challenge could be defeated by evolution of software systems and algorithms that support portability besides relaxing burden of program design and implementation. Table 1 enable us to compare and achieve this high performance distributed system using grid, cluster and cloud computing.

Table 1: High Performance Grid v/s Cluster v/s Cloud Computing

Category	Grid	Cluster	Cloud
Size	Large	Small to medium	Small to large
Resource type	Heterogeneous	Homogeneous	Heterogeneous
Initial Capital Cost	High	Very High	Very low
Typical ROI	Medium	Very High	High
Network type	Private Ethernet based	Private IB or proprietary	Public Internet Ethernet Based
Typical Hardware	Expensive	Very expensive- top of the line	Usually VMs atop of hardware
If I didn't know any better:	Faster workstations	supercomputers	Bunch of VMs
SLA requirement	High	Strict	Low
Security Requirement	High	Very low- but typically high	Low

REFERENCES

[1] Michael Hill, “Distributed Computing: An Unstoppable Brute Force”, SANS Institute Info Sec Reading Room, pp 1-13, November 13, 2003.

[2] Manish Parashar and Craig A. Lee, “Grid Computing: Introduction and Overview”, pp 1-6, 2005, Available at: www.researchgate.net

[3] R. Buyya, “High Performance Cluster Computing: Architectures and Systems”, ISBN-10: 0130137847, Prentice Hall; 1 edition, 21 May 1999.

[4] Albert Y. Zomaya, “Handbook of Nature-Inspired and Innovative Computing integrating classical models with emerging technologies”, EBook Springer, chapter 16, pp 521 – 530, 2006.

[5] Krishna Kant, Muthu kumar Murugan, and David Hung Chang Du, “Energy Adaptive Computing for a Sustainable ICT Ecosystem”,Design Technologies for Green and Sustainable Computing Systems, Springer, pp 59-91, 2013.

[6] Miettinen AP, Nurminen JK, “Energy efficiency of mobile clients in cloud computing”, ACM HotCloud'10 Proceedings of the 2nd USENIX conference on Hot topics in cloud computing, pp 4-11,2010.

[7] Chun B-G, Ihm S, Maniatis P, “Clone: elastic execution between mobile device and cloud”, in the ACM proc of the sixth conference on Computer systems (EuroSys‟11), pp 301-314, 2011.

[8] Curevo E, Balasubramanian A, Cho D-K, Wolman A, Saroiu S, Chandra R, Bahl P, “Maui: making smart phones last longer with code offload”, ACM MobiSys'10, Proceedings of the 8th international conference on Mobile systems, applications, and services, pp 49-62, 2010

[9] Mayank Raj, Krishna Kant, Sajal K. Das, “Energy Adaptive Mechanism for P2P File Sharing Protocols”, Parallel Processing Workshops: Euro-Par 2012, Springer Lecture Notes in Computer Science, Volume 7640, pp 89-99, 2013.

[10] Top500 Super Computer Report, Available at: www.top500.org.

[11] Volodymyr Kindratenko, Pedro Trancoso, “Trends In High Performance Computing”, IEEE Computing, vol. 13 no. 3, pp. 92-95, 2011.

[12] I. Foster and C. Kesselman, “Computational Grids: Blueprint for a new computing infrastructure”, ISBN:1-55860-475-8, ACM book,1998.

[13] Joshy Joseph, Craig Fellenstein, “Grid computing”, report published by IBM Press on December 30, 2003.

[14] Christian Vecchiola, Suraj Pandey and Rajkumar Buyya, “High-Performance Cloud Computing: A view of Scientific Applications”, International Symposium on Pervasive Systems, Algorithms, and Networks, IEEE Computer Society, pp 4-16,2009.

[15] Peter Mell, Timothy Grance, “The NIST Definition of Cloud Computing”, National Institute of Standards and Technology Special Publication no. 800-145, 7 pages report, September 2011.

[16] Ye Xiaotao, LvAili, Zhao Lin, “Research of High Performance Computing with Clouds”, ISBN 978-952-5726-10-7, pp 567-577, 2010.

[17] Cloud Computing Definition, National Institute of Standards and Technology, Version 15, available at:www.csrc.nist.gov.

[18] Jaliya Ekanayake and Geoffrey Fox, “High Performance Parallel Computing with Clouds and Cloud Technologies”, 1st International Conference on Cloud Computing, October, 2010.

[19] High-Energy Grid Development, Available at:www.gridcafe.org.

[20] Michel Dayed, Jack Dongarra, “High Performance Computing for Computational Science”, ISBN 978-3-642-19327-9, Springer, pp 120-121, 2010.