ISSN No. 0976-5697

Volume 2, No. 6, Nov-Dec 2011

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

High Performance Computing Solution in Bangladesh through Grid Computing

Husnain-Al-Bustam*
Department of Electrical and Electronic Engineering
Islamic University of Technology
Board Bazar, Gazipur, Bangladesh
husnailb@gmail.com

Saif-Un-Noor
Department of Computer Science and Engineering
American International University, Bangladesh
Dhaka, Bangladesh
saif009.prottoy@gmail.com

Md. Thesun-Al-Amin
Department of Electrical and Electronic Engineering
Islamic University of Technology
Board Bazar, Gazipur, Bangladesh
thesun2144@gmail.com

Abstract: This paper gives us an idea of a good High Performance Computing solution through grid computing in Bangladesh. Here we have used BOINC as our primary tool of grid computing. In Bangladesh if we want to do a good number of research works in chemical and pharmaceutical studies, agricultural field, energy research geophysics and oil applications, weather predictions and some other fields the researchers and the faculty members need a good source of high computing and processing power and here we are dealing with such an idea which will be a good solution of high performance computing (HPC) in Bangladesh. Although our system is a good solution of HPC in Bangladesh it may have resource scheduling problem in large area communication but we have to make the trade-off.

Keywords: Grid Computing, High Performance Computing, BOINC, Research in Bangladesh, Performance and Cost Analysis

I. INTRODUCTION

At present Grid computing has become a good solution for providing a powerful and on demand computing by using resources from multiple administrative domains.[9] Mainly grid computing is an enhanced form of distributed computing in which an organization or a particular sector uses their existing computing resources to solve its own large computing problems. It is a form of parallel computing that relies on completely computers connected to a network. In Grid computing many computers are connected to a network to perform a very complex task together. It is of great use in weather forecasting, pharmaceutical research works, seismic analysis, and many applications of the computational sciences.

As a form of commodity computing, grid computing ensures us the greatest amount of useful computation at a very low cost. In this sense grid computing is much cheaper than supercomputers. The term grid computing was first coined in early 90's. It's like electric power grid as its computational resources are geographically apart. The main features of grid computing includes that the computing resources that are used in computing can be trusted that is one can assume that the PC's don't return the result that are harmful or wrong. But one thing is that due to lack of central control of hardware nobody can guarantee that the nodes will not drop out of the network at random times. Many middleware projects have created a good infrastructure to allow many organizations to setup a platform for grid computing. BOINC (Berkley Open Infrastructure for Network Computing) is a common one for various academic projects.[1][2][5] current grid computing system can be divided into two basic models. The first model is called the service grid and another is called the Desktop grid. Service grid is the classical model of grid computing which is mainly used by the scientific user community and it provides a rich set of facilities. In desktop grid the total performance can be obtained from the unused cycles of the desktop computers. Although a single desktop computer may not be capable of doing that, but when all the available desktops are united in a desktop grid it provides a massive power and a great computing resource. In service grid model the total computing resources are in under control of the participating institutions. On the other hand in desktop grid computers are volunteered by public resources. In our system we are focusing on the combination of service grid and desktop grid so that it remains cheap and at the same time the participating institutions have the full control over the whole system.[5][6]

As our research paper focuses on the technique of developing a High Performance Computing (HPC) solution in our country. High Performance Computing mainly uses cluster of computers to solve many computational problems. High Performance Computing combines computational techniques with various fields like electronics, computer networks, programming languages, algorithms, parallel programming etc. Now a day's High Performance Computing (HPC) is not bounded to supercomputers only. Grid computing as well as cluster of computers is a great solution of High Performance Computing. So our main goal is to establish such a grid infrastructure that will act as a precious and interactive research and development tool for the Bangladeshi researchers.

The total system is designed in such a way where users submit their jobs request and they wait for the notification of the job completion. On the other hand current grid architectures are designed for batch applications. In these kinds of systems user don't get the full control of the system but he or she has the full ability to request for a job and after the system completes it with the aid of the current grid system, he will be notified and then he will find that his requested job is completely done. In these types of systems security issues are involved in large scale and dynamic scheduling is a great factor as the user who has requested the job first, he or she will get the highest priority among the other users who have requested. For the total solution of our proposed grid system for High Performance Computing (HPC) we have used the Berkley Open Infrastructure for Network Computing (BOINC) as tool. BOINC is an open source middleware system for grid computing. The purpose of the BOINC is to provide massive processing power of the personal computers around the world to the researchers. Mainly BOINC is software that can use the unused cycles of CPU on a computer to do scientific calculations and other works related to them.

Universities involved in research of our country can also use BOINC to create Virtual Campus Supercomputing center (VCSC) for themselves. It will provide the researchers with great computational power like a supercomputer within a small amount of money. The total computation power is provided by the universities resources like faculty computers, laboratory computers etc. All these things provide a total solution for high computational power for the whole university. In our system we are doing the same approach that will help us to build such a system of our own which will work for both the universities and for the people involving in research activities.[10]

Main research activities are going in our country is in the area of agricultural sectors and pharmaceutical sectors. There are a little research activities in other fields. We have agricultural universities, rice research institutes where biological research activities are going on a regular basis. Researchers involved in those activities need a good source of computing power in order to solve large biological problems using various computational model [i.e. DNA sequencing]. We also have to do research activities pharmaceutical sectors and many Universities where the researchers have to deal with large computational problems frequently. So we have come up with an idea of providing them a good source of computational processing power, so that they can continue their respective works without any obstruction and complication.

II. PHILOSOPHY AND MOTIVATION

Scientific research activities are getting complex and harder day by day. If we think of a small scientific data processing we will find that the size and complexity of those data are so high to deal with normally. Suppose in a biological research a scientist needs to examine the DNA sequence. Well DNA sequencing is not that an easy task. So what the modern countries do they deal these things with a high performance computing solution that is they use a great source of processing power or they use the supercomputers. Other critical thing in the arena of complex computational activities includes optimization and simulation problems like various machine learning approaches, Markov chain Monte Carlo sampling, and dealing with large data in various data mining activities.

Research and various activities are going so fast for grid computing that it has helped the researchers to do complex research works in the field of computational biology and bioinformatics. As these kinds of research works deals with the problems of molecular evaluation and genetics the researchers related to these research works often deals with various high computing resources. As these kinds of computing problems are increasing day by day according to our demand. So the researchers picked grid computing as a good solution of their computing resources. And the need for computer resources for these kinds of works led us to work for the development of grid computing and many commodity tools related to it. BOINC and GLOBUS are common toolkit for grid computing. There are also some voluntary projects like lattice project.

In our country these kinds of research activities are not so high. But we are trying hard to contribute in these kinds of scientific research activities. The main reason behind the lack of research activities is the proper amount of fund needed to conduct a research work. And a vital reason is the lack of resources needed for a research work. So if we can provide them with these kinds of facilities then we can hope there will be some good research works in Bangladesh especially in agricultural sector.

So we made a research work so that we can provide good computing resources for high performance computing to our local researchers and to introduce Bangladesh a new type of system which deals with high performance computing at a very low cost which is our main goal.

III. RELATED WORKS AND TECHNOLOGY

In order to provide the High Performance Computing(HPC) solution to our country through grid computing we are going to discuss here the basic functionalities of our system as well as the total mechanism of the system with the basic tools. The main tool we have used for our system is BOINC (Berkley Open Infrastructure for Network Computing). In our system we are involving some universities (15 universities from all over the country) of Bangladesh as a formation of voluntary computing. Mainly the universities are the main resource providers of our solution. Every university in our system will provide thirty (30) personal computers as dedicated servers. And our other source of resources will be the lab computers, faculties and staffs personal computers. BOINC will be installed in each and every computer of those universities (computers employed for our system). There will be three main servers' one master grid server & two slave grid servers. The main responsibility of the servers will be the combination performance and the requested work to be done. All the universities will have VCSC (Virtual

Campus Supercomputing Center).[10,16] Through VCSC they will be connected to each other using the BOINC tool and the whole campus researchers will get the computational power of a large clusters or supercomputers. And those resources (Resources from lab computers, faculty computers) will be connected to a central server of the university. This system will be applicable for all the universities. .Then the central server will be connected to salve grid2 that means the grid of the BOINC VCSC. All the main servers of the universities will be connected to the slave grid2 thorough internet and this slave grid2 will be connected to the master grid server.

On the other hand the dedicated computers of the universities will follow the same rule like the existing university resources but there will be a separate main server for the dedicated servers in a university. Those dedicated servers will be connected to the slave grid1thorugh internet which will be connected to the master grid server. So the slave grid1 and slave grid2 are mainly responsible for all the job scheduling.

Main server will centralize the total combined power. And it will be the source of the whole computational power. There will be a job scheduler who will accept the submitted jobs and schedule them properly. BOINC is the tool which will manage all these basic activities from scheduling to job submission. When the job is done the system will forward the completed task or result to the user who will request for the job to be done.

Suppose a researcher from a university requests a job to the central server of the university he is connected with. That server will forward the task to the main server. And the main server will forward that task to that grid from which it can does the job within a short period of time employing a good number of resources. One thing is to be noted that the master grid will use the unused cycles of the computers that is employed for BOINC VCSC and the available cycles of the dedicated servers that is the cycle of high performance clusters.

Here job scheduling is a great issue. If the system finds the total system is busy then it will notify the user and then the Users may have to wait for a certain period of time. The job submission scheduling will follow the FIFO structure that is the First In – First Out structure that the job requested first will be done first.

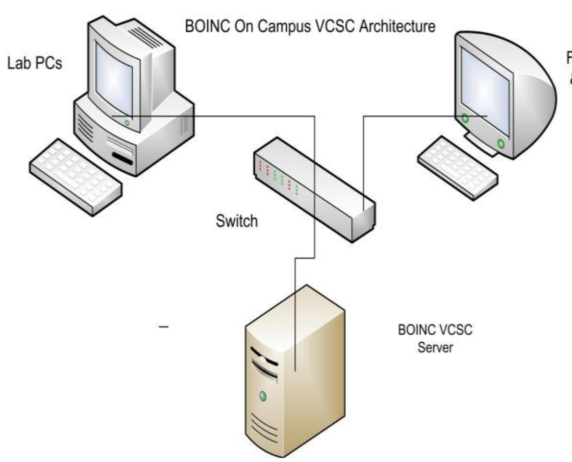

Figure. 1 BOINC on campus VCSC architecture

A. *Setup and Internal Mechanism of BOINC:*

This section will mainly describe the internal mechanism of BOINC and how the user's job is completed by BOINC.

After installing BOINC in a user pc the user pc gets some tasks from the grid server. Here is a thing that the server will never give any task that requires more powerful hardware than the user has. For example the central server will never give such task that requires minimum 2 GB RAM if the user pc has 1 GB RAM.

The computer will download all executable and input files from the server and then it will run all the executable files on its own system that means the computers which is currently employed for the job and produce the output. Later it will upload the output files to the server that means the requested grid.

The internal architecture of BOINC includes a core client, a BOINC manager (GUI interface) and a screensaver which communicates with other clients using TCP protocol and it will be activated when the user is away from his computer means the user is not using his computer for any purpose.

The client communicates with other servers using HTTP protocol and the BOINC manager provides the GUI interface to the user.

This is how the total system basically works. Here one thing is to be noted that BOINC keeps record of the amount of project done by a computer with the aid of gird servers connected to it. Basically this is called CREDIT. When a computer completes the given task CREDIT is earned. [10]

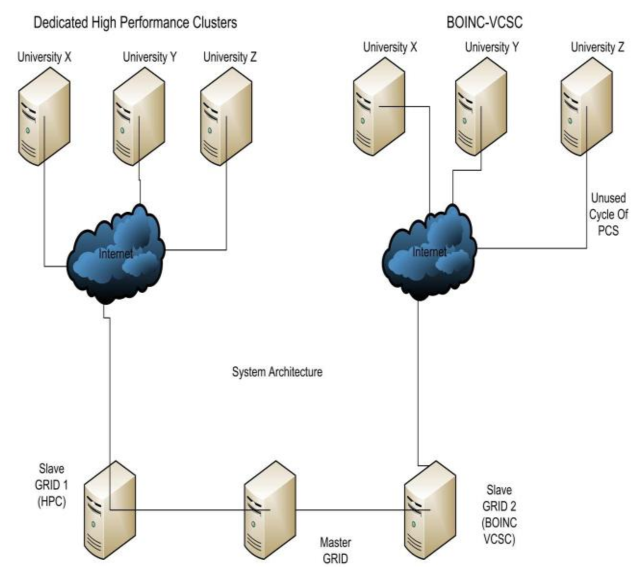

Figure. 2 Total System architectur

IV. RESULT ANALYSIS

Previously in the tools section we have mentioned that we have two slave grids in the system & one is the grid for high performance clusters & another is for the BOINC VCSC. We have considered fifteen universities for modeling the system & each of them will have clusters that comprising of 30 nodes & for the BOINC – VCSC the number of computational resources are mentioned in the following table.

Table-I

Serial No	Name Of University	BOINC-VCSC Faculty, Staffs, & Lab PCs	High Performance Clusters
1	DU	3300	30
2	BUET	1680	30
3	KUET	875	30
4	RUET	750	30
5	CUET	770	30
6	DUET	650	30
6	SUST	1275	30
8	IUT	675	30
9	AUST	800	30
10	AIUB	850	30
11	NSU	1275	30
12	BRAC	650	30
13	UIU	725	30
14	EWU	875	30
15	IIUC	1150	30
Total		16300	450

a. Cost Calculation of High Performance Clusters: For the cost calculation first we have a look on the specification of the PCs for high performance clusters & servers for the GRID. Specification of computers for high performance clusters is given in Table 2.

Table-II

Serial No	Specification	Price (Approximately)
1	Intel PDC E6500 2.93 GHz	4300 BDT
2	Intel Main board	4100 BDTa
3	2GB DDR3 1333	1600 BDT
4	Casing	1500 BDT
5	DVD Writer 20X	1400 BDT
6	250GB SATA HDD	2600 BDT
Total		15,500 BDT

[a] Bangladeshi Taka

Per computer will cost 15,500 BDT. As we have considered 30 dedicated computers for the high performance clusters, the total cost will be (15,500*30) =465,000 BDT. Others cost like networking accessories & one monitor for each clusters we consider the total cost are about five lacks BDT for each university. As we have considered 15 universities, so the total cost for the high performance clusters (dedicated computers) will be seventy five lacks BDT. We are expecting 150 computers of stated configuration in the table-2 from the government for the cluster. So the price of 150 PCs is nearly twenty five lacks BDT including all other expenses. So for the high performance cluster we need nearly 1 Cr. BDT.

That is equivalent to US$0.135 million. [11].

According to our mechanism of design we need 3 servers. So for the specification of server are given below-
1)Intel BX 5500 Clone Server 2)Processor - Quad Core Intel Xeon E5620 2.4GHz 3)Chipset - Intel S5500BC with Intel 5500 Server Dual Xeon Supported 4)RAM (MB) - 2048 5)HDD (GB) - 500x2 6)Operating system
The price of this server is one lack & ten thousand BDT. So the price of three servers is three lacks & thirty thousand BDT. And for the networking tools, monitor, & other purpose we need approximately five lacks BDT. So finally for the server we need eight lacks & thirty thousand BDT that is equivalent to $11,500 USD. So the grand total for this project will be US$0.1465 million.[11]

b. Cost Calculation for BOINC-VCSC: Budget at least a few person-months for getting a working system running [16].

A. *Performance Analysis:*

In our GRID of BOINC-VCSC we have the participation of 16,300 computers, running an average of 50% of the time. In terms of computing power, this is roughly equivalent to an 8,150 node cluster [16]. And from the high performance clusters we are getting the computational power of 600 PCs constantly. So we can conclude that we will have the computing power of 8,750 PCs. As the computing power is supplied by campus PCs, e.g. computer labs machines, desktop and laptops belonging to faculties and staffs, we have make here approximation considering the number of PCs for BOINC-VCSC.

Table –III [13, 14]

Processor name	Model	Cache L2	Clock Speed (GHz)	GFLOPS
Intel® Pentium® 4 processor	521	1MB	2.80	5.60
Intel® Celeron® processor	440	512KB	2	8.00
Intel® Pentium® dual-core processor	E6300	2MB	2.80	22.4
Intel® Pentium® dual-core processor (HPC)	E6500	2MB	2.93	23.44
Intel® Core™2 Duo processor	E7200	3MB	2.53	20.24
Intel® Core™2 Quad processor	Q8200	4MB	2.33	37.28
Intel® Core™ i3 Desktop Processor	i3-330E	4MB	2.13	17.04

In Table-3 we have mentioned that what type of processor that we will use in our grid. To measure peak performance we have used Intel's product specifications & comparisons of processors. In the table-4 we have shown the total peak performance of the slave grid of the BOINC VCSC & that is 0.195984 PFLOPS. Now question comes how we have calculated this. Consider the following case.

From one Intel® Pentium® 4 processor we are getting 5.60 GFLOPS peak performance. As we have considered approximately 6000 PCs with Pentium 4 processor. So finally we are getting (5.60*6000) =33,600 GFLOPS or 0.0336 PFLOPS peak performance. In this fashion we have calculated the rest.

Table-IV

Processor Name	Quantity	GFLOPS	TFLOPS	PFLOPS
Intel® Pentium® 4 processor	6,000	33,600	33.6	0.0336
Intel® Celeron® processor	5,000	40,000	40	0.040
Intel® Pentium® dual-core processor	3,200	71,680	71.68	0.07168
Intel® Core™2 Duo processor	1,500	30,360	30.36	0.03036
Intel® Core™2 Quad processor	500	18,640	18.64	0.01864
Intel® Core™ i3 Desktop Processor	100	1704	1.704	0.001704
Total	16,300	195,984	195.984	0.195984

For the High Performance Cluster we have peak performance = (600*23.44) = 0.014064 PFLOPS. The total peak performance mentioned in the table-3 is the peak performance for 16,300 PCs. As these PCs will run approximately for an average of 50% of the time so we will have the peak performance half of the total, that means (195,984/2) =97,992

GFLOPS. That is equal to 0.097992 PFLOPS. We have shown here a linear calculation for the peak performance but in practical it is not linear. Considering this fact and as these PCs are the PCs of computer labs, faculties & staffs and as they will use it so we can expect we will get the peak performance approximately 40 TFLOP from the BOINC-VCSC. Combining this peak performance of BOINC-VCSC's grid & high performance cluster's grid, in the master grid we have total (40+14.064) =54.064 TFLOP or 0.054064 PFLOPS peak performance.

B. Comparison Between IBM Road Runner and Our Proposed System:

The price of IBM Roadrunner is US$133-million & that will give a peak performance of 1.7 PETAFLOPS.[12] As the PCs for BOINC we already have so here we are ignoring those PCs cost. The cost for High Performance Clusters PCs are US$ 0.1465 million & that will give us peak performance 0.054064 PFLOPS. We think so that is pretty enough to serve our purpose of research in Bangladesh.

V. CONCLUSION

According to our performance analysis we can come up with the decision to set up such an efficient system as well as a cost effective system in Bangladesh that will provide a better High Performance Computing solution for our Bangladeshi researchers. We may face some scheduling problem as well as some security problems here but in order to provide such a system in low cost we have to achieve this system as the solution of high performance computing in Bangladesh.

VI. REFERENCES

[1]. Myers, DS, Cummings MP. 2003. Necessity is the mother of invention: a simple grid computing system using commodity tools. Journal of Parallel and Distributed Computing.

[2]. Cummings, MP, Huskamp JC. 2005. Grid computing. Educause Review. 40:116–17

[3]. Bazinet, AL. 2009. The Lattice Project: A Multi-model Grid Computing System.

[4]. Bazinet, AL, Cummings MP. 2011. Computing the Tree of Life - Leveraging the Power of Desktop and Service Grids.. Fifth Workshop on Desktop Grids and Volunteer Computing Systems (PCGrid 2011)

[5]. Grid computing as applied distributed computation: a graduate seminar on Internet and Grid computing Browne, J.C.;Cluster Computing and the Grid, 2004. CCGrid 2004. IEEE International Symposium on 19-22 April 2004 Page(s):239 – 245.

[6]. From Grid Middleware to a Grid Operating System Ali, A.; McClatchey, R.; Anjum, A.; Habib, I.; Soomro, K.; Asif, M.; Adil, A.; Mohsin, A.;Grid and Cooperative Computing, 2006. GCC 2006. Fifth International Conference Oct. 2006 Page(s):9 - 16 Digital Object Identifier 10.1109/GCC.2006.49

[7]. Reliability analysis of grid computing systems Dai, Y.S.; Xie, M.; Poh, K.L.;Dependable Computing, 2002. Proceedings. 2002 Pacific Rim International Symposium on 16-18 Dec. 2002 Page(s):97 - 104 Digital Object Identifier 10.1109/PRDC.2002.1185624

[8]. Quality of service in global grid computing Valcarenghi, L.;High Performance Interconnects, 2005. Proceedings. 13th Symposium on.17-19 Aug. 2005 Page(s):4 - 5 .Digital Object Identifier 10.1109/CONECT.2005.31

[9]. What is the Grid? A Three Point Checklist. I. Foster, GRIDToday, July 20, 2002

[10]. http://boinc.berkeley.edu

[11]. http://www.bangladeshbank.org/econdata/exchangerate.php2

[12]. http://en.wikipedia.org/wiki/IBM_Roadrunner

[13]. http://www.intel.com/support/processors/sb/cs-023143.htm

[14]. http://www.intel.com/support/processors/sb/CS-028241.htm#7

[15]. On the Use of Cloud Computing for Scientific Workflows Christina Hoffa1, Gaurang Mehta2, Timothy Freeman3, Ewa Deelman2, Kate Keahey3, Bruce Berriman4, John Good4

[16]. http://boinc.berkeley.edu/trac/wiki/VirtualCampusSupercomputerCenter

[17]. Investing in crop agriculture for higher growth & productivity & adaption to climate change (http://bangladeshfoodsecurity.files.wordpress.com/2010/05/investing.pdf