

The Horizontal Cluster!

Written by Muhammad Kamran Azeem

Friday, 05 June 2009 22:13 - Last Updated Sunday, 14 June 2009 12:14

The Horizontal Cluster!

Note: This article is mainly addressed to the young IT generation of my home country, Pakistan. But, anyone from any part of the world can read this and benefit from it. The ultimate objective though is to do something useful for the world at large, without any classification, discrimination of nationality, gender, race or religion.

Salam to all,

In my previous article, "***A supercomputer on every desk!***" ***Videos released!*** , I pointed out that the how can you have a multi-node 'virtual' HPCC / *Beowulfish* cluster on your own desk. The videos released with this article showed how can you build one, by following simple steps and using open-source technologies.

After that, article, I received a lot of criticism and complaints. People mentioned (angrily) (to me), that it would be better if I can answer these questions first: (a) "*What is a super computer/HPCC?*" , (b) "*What is the use of an HPC Cluster/Super-Computer?*" , and (c) "*Why should we bother?*" . And I admit that is truly my fault. I should have explained them in the same article, before going on and trying to open my heart out on the Horizontal Cluster.

Let me start by answering the first question before I go on explaining the second question.

Question: "*What is a super computer/HPCC?*"

The Horizontal Cluster!

Written by Muhammad Kamran Azeem

Friday, 05 June 2009 22:13 - Last Updated Sunday, 14 June 2009 12:14

Super computer in simple terms is a computer, which can execute more number of calculations in a unit time, compared to a commodity PC.

In old times, PCs were kind of non-existent. Main frames, and later “Minis”, ruled the world. Thanks to the advancement in technology, especially in the field of Processor manufacturing (electronics / hardware) and IT (software), the big bulky supercomputers are not common anymore. Instead, a new breed, known as Beowulf clusters have taken their place. (Well, almost). Beowulf was a project (turned successful), started in 1994, to execute programming code in parallel, on a group (cluster) of PCs, connected to each other over a TCP/IP network. Linux, which was very new at that time, was used for this. What turned out, was surprisingly impressive performance, even though they were using 10Mbps network for cluster communication! Beowulf cluster, are now also known as High Performance Computing (sometimes Compute) Clusters. Many “religious” “beowulf advocates” may disagree and debate. But lets save that for another time and remain focused on what I have to tell you.

In the IT world, there is some in-formal difference between Beowulf clusters and High Performance Compute Clusters, or HPCC. As said earlier, Beowulf clusters normally consists of ordinary PCs. However, HPCC, being a Beowulf cluster, *in principal of operation*, consists of the expensive equipment all over. HPCC are used by big guys, like Oil and Gas industry, Space exploration, etc, who want the fastest machines, and they are ready to spend millions on the equipment and setup. The interconnect in such situations can be (very) costly Myrinet or Infiniband, or a bit cheaper Gig-E. All placed in a high profile data center. As you can see, this “does” cost “too much”. But the reward you get is performance measured in Tera Flops, and more recently in Peta FLOPS! . To let you compare a bit, let me tell you that Cray-1, in 1976 was just 250 Mega FLOPS. Cray-2 (1985-1989) was 3.9 Giga FLOPS! Nowadays, a few hundred dollar, desktop machine/PC is now able to achieve, more than that. Again, please keep the advancement of technology, in mind.

A Beowulf cluster, or an HPCC cluster, can not do magic alone. The magic is to break the application program code into such a way, which can be executed in parallel on multiple machines. Most of the programmers, write serial code (along being buggy and in-efficient). That is, program code will do all tasks in a serial fashion (steps performed / executed one after the other). Such kind of programs “can not” give you performance, no matter what size of cluster you throw at it. They do get executed faster on big iron machines though. But again, that is limited by the size of the machine (and your pocket). So the magic is to use PVM or MPI. PVM is Parallel Virtual Machine, and MPI is Message Passing Interface. Both are popular technologies to write and execute program code written in parallel, while MPI being more popular nowadays.

The Horizontal Cluster!

Written by Muhammad Kamran Azeem

Friday, 05 June 2009 22:13 - Last Updated Sunday, 14 June 2009 12:14

And that was, *what a Super-Computer is*, in detail.

Question: “What is the use of an HPC cluster/super computer?”

Though not limited to the scope and usage, HPCC is normally used for the following:

-

Quantum mechanical physics,

-

Weather forecasting,

-

Climate research,

-

Molecular modelling (computing the structures and properties of chemical compounds, biological macromolecules, polymers, and crystals),

-

Simulation of air planes in wind tunnels, and creation of different flight / combat situations

-

Simulation of the detonation of nuclear weapons,

The Horizontal Cluster!

Written by Muhammad Kamran Azeem

Friday, 05 June 2009 22:13 - Last Updated Sunday, 14 June 2009 12:14

-

Research into nuclear fusion,

-

Cryptanalysis,

-

Oil Well drilling and capacity calculations

-

Reservoir (Oil) simulations

-

Seismic studies

-

Deep see exploration

-

Space exploration

-

Weather forecasting, earthquake prediction, storms, tsunamis and tornado prediction

-

Banking, trend analysis, etc, stock market predictions, etc, etc.

-

The Horizontal Cluster!

Written by Muhammad Kamran Azeem

Friday, 05 June 2009 22:13 - Last Updated Sunday, 14 June 2009 12:14

High performance cars: design and testing / simulations (Formula 1, etc)

-

Professional sports equipment design validations through simulations,

-

Creating and testing various vaccine designs for different diseases,

-

Testing time travelling theories, such as Einstein's Theory of Relativity, etc

-

Decryption of Ciphers and keys

-

Study of human mind, neural networks and designing Artificial Intelligence.

-

Study of green house effect and global warming

-

Agriculture research,

-

Water storage / dam building,

-

Electricity generation planning and operation,

The Horizontal Cluster!

Written by Muhammad Kamran Azeem

Friday, 05 June 2009 22:13 - Last Updated Sunday, 14 June 2009 12:14

-

Oil exploration,

-

Military planning,

-

New generation of more complex war games,

-

, and many.... many more.

If you want to categorize them industry-wise, it would be:

-

Science: Astronomy, Physics, Biology, Chemistry, Earth Sciences, Health

-

Engineering: Automotive, Aerospace, Marine, Defence, Oil and gas

-

Others: Finance, Internet, Media, etc.

The Horizontal Cluster!

Written by Muhammad Kamran Azeem

Friday, 05 June 2009 22:13 - Last Updated Sunday, 14 June 2009 12:14

Also, read this article at Forbes.com : [American Business's Secret Competitive Weapon: HPC](#) . That will probably help you understand my point. You will see for yourself, that how technology / HPC is helping *even the sports industry*

There is an extension to the answer to the question asked. The uses of super-computer are limited only by your imagination. In simple words, whatever you can put this technology to work, is its yet another use. There are no formal hard drawn lines here.

Now, let's deal with the last question question:

Question: “ *Why should we bother?*”

Well, if you have read the previous article and you are an IT enthusiast, then probably you already have an answer. Nevertheless, I will try to explain it to everyone, as below. Remember, the main target audience of this article are people from third world and under-developed countries. Read the text in that context. Again, it is still usable for the developed and developing countries.

You should bother, because something needs to be done in all of the fields of science and technology, listed above, in your country / community. But unfortunately, you have few resources available to you. You cannot just go ahead and buy an HPCC or super computer for yourself, or your company, organization or research institute. So in that case, what would you do? Or, what can be done? There are chances, that your fellow countrymen/community members are already started studying parallel programming and they need heavier computing entities / real clusters to solve some of the most un-heard computing problems. There is also a chance that you are already in-search of a readily available magical solution, without spending anything extra. These are some of the clear reasons, to: Why you should be bothered.

The Horizontal Cluster!

Written by Muhammad Kamran Azeem

Friday, 05 June 2009 22:13 - Last Updated Sunday, 14 June 2009 12:14

The Horizontal Cluster!

Now comes the actual point of this article. If you are wondering that who would deliver these magical clusters to you and where would you place them? Well, the answer is that, you already have them, ... and, in place ! In-fact, you have hundreds of clusters available to you, but you never bothered to look for them. If you wanted someone to come , wake you up and spoon feed you with a cluster, then congratulations! I I have come to you, to do exactly that!

As you know, and have seen with your own eyes, that all computer institutes have a computer lab, with some network infrastructure in place. (Great info, isn't it!) And all offices, now-a-days have computers too! Connected through a network switch. Good. So when these machines are doing nothing in the evening, why not use them as a cluster? Simple! If you are already having Linux on all these PCs, then your task is much simplified. However, If you have that disease (windows) installed on all of your PCs and you don't want to get rid of it too (which is quite easy in-fact!), then you can try installing Linux on a separate partition on those PCs. The plan is, when everyone goes home, you just reboot the machines (it's upto you, how you reboot them), into Linux, and just start submitting your program jobs through a master node, which you have already assigned the role of master. Good. So all of a sudden, you have a full cluster at your disposal, running actual number crunching jobs! You can provide access to this cluster through web or ssh or whatever method you like, to your users. Better!

If you wonder about the electricity bill, (which you are), then, my idea is that you can try selling your cluster time to users outside your organizational hierarchy. People who are not your students or employees. (Or whatever your creative mind can think of.) This way, the cluster will pay for the electricity. At the same time the cost / productivity ratio will become better as the computers will be working most of the time.

There is another solution. If you computer lab / office network already has Linux on it, and your users just do basic office activities most of the time, like email, web-sufing and word processing, then you don't even have to wait for them to go home! You can just schedule your scheduling software to use the unused CPU cycles from all the computers on your network and run compute jobs during work hours as well!

So just think about it for some time. Think of all the IT companies turned into compute clusters

The Horizontal Cluster!

Written by Muhammad Kamran Azeem

Friday, 05 June 2009 22:13 - Last Updated Sunday, 14 June 2009 12:14

all of a sudden, WITHOUT spending an extra penny! The hardware and infrastructure is already there, and you will use complete stack of free / open-source software to setup that. (This will be covered in a future article.)

If you can bear a little more, I would ask you to just think of all these small / individual clusters as power houses. And now the related next idea: **The Grid!** Yessssssss ! The cloud , the grid, (the matrix if you want to call it!) . Connect all , or few of these clusters together over internet , to form a grid. That can easily be done. You can develop trust relationships between different companies and educational institutes in NxN manner for mutual benefit. If you have a program code which needed some extra ordinary compute power to run, and was too big for any small scaled cluster, then just join a few clusters in a grid form and submit your compute job. That is it!. You have a nation wide, region wide, or world wide compute cluster with you, without spending a penny! What else could be better than this?

I hope this gives your thought process, a spin. And if it does, act now.

Regards,

Kamran

Related articles:

["A supercomputer on every desk!" Videos released!](#)

[Who uses a supercomputer anyway?](#)

The Horizontal Cluster!

Written by Muhammad Kamran Azeem

Friday, 05 June 2009 22:13 - Last Updated Sunday, 14 June 2009 12:14
