Association for Information Systems
AIS Electronic Library (AISeL)

SAIS 2014 Proceedings | Southern (SAIS)

4-14-2014

INDEPENDENT PROJECT BASED LEARNING: APPLYING KNOWLEDGE BY CREATING A LOW-COST SUPERCOMPUTER WITH RASPBERRY PI

Shaun Franks
Middle Georgia State College, shaun.franks@mga.edu

Johnathan Yerby
Middle Georgia State College, johnathan.yerby@mga.edu

CREATING A LOW-COST SUPERCOMPUTER WITH RASPBERRY PI

Shaun Franks
Undergraduate student
Middle Georgia State College
Shaun.Franks@mga.edu

Johnathan Yerby
Middle Georgia State College
Johnathan.Yerby@mga.edu

ABSTRACT

This article was written to detail the process of one-on-one project based learning while creating a supercomputer using Raspberry Pi. Traditionally computing is executed one instruction at a time, usually using a single processor, but in the cluster created using open source software is capable of parallel processing and distributed storage. The speed in which instructions are completed depends on how fast data moves through the hardware. Parallel computing is a faster way to process instructions by breaking the large task into smaller tasks using a coordinated effort to process data simultaneously. Parallel computing is typically handled by supercomputers that range in costs from the hundreds of thousands to over a billion dollars. This paper details the design, challenges, and successes of creating a low-cost supercomputer and possibilities for the creation to be used for industry and continued learning.

Keywords

Celer-Pi, High performance computing, Raspberry Pi, supercomputer, technology demonstration

INTRODUCTION

This case study uses ten Raspberry Pi computers that cost $35 each to create a low cost supercomputer capable of distributing the processing and storage across the affordable small devices. Cost to create high performance parallel computing devices is one of the largest obstacles that prevent educational institutions from taking on projects like this one (Pounds, Nalluriu, & Coleman, 2005). Software resources required for the project included open source (free) software OS Debian Wheezy, Raspbian. Nginx was used to manage bandwidth and latency. MPICH2 was used to manage the cluster, coordinate the processing and distribute storage. An external hard drive, switch, router, cabling, power supply, and a custom built case for the creation were the remaining costs. The Raspberry Pi Supercomputer, which was named Celer-Pi, had a total cost of $900. The name Celer-Pi was created from the Latin word “celer” which means fast, combined with “Pi” from the Raspberry Pi device. One of the lowest cost supercomputers available for retail starts at $500,000. Although the processing and storage capabilities of the $500,000 machine from Cray exceeds the Celer-Pi, the concept and operations are comparable (Gohring, 2013). The creation reached performance of at least 900MHz while running a data-driven website.

The ARM CPU onboard the Celer-Pi device is power efficient and fast. The processor has been used in a number of recent mobile devices, from the iPhone to netbook computers such as the Google Chrome book. With a growing number of mobile devices being sold to consumers that contain ARM processors, the software is being researched with efforts to gain insight to the processing capabilities that it can provide for a low cost and low resource consumption.

The particular setup utilizes a 900MHz processors unified as one without regard to IRQ request conflicts. Since IRQ slots are limited in number, but basically are the direct manager of what gets processed. Less interrupt requests per processor means more time calculating what we are interested in. Instead of each processor requesting a slot on a single motherboard, as in the case of say, an 8 core processor; each processor has its own motherboard to interrupt with approximately the same number of request slots as the integrated multi-core processor.

This project demonstrates how this affordable project can provide new capabilities to educational institutions seeking to provide opportunities to students to learn how to create and work with applications that they will encounter in their future. The independent project-based nature of the project proved to be a key factor to the successful design and implementation of the project. The student had a personal interest in the project, the student was able to work independently to apply existing knowledge to a new complex challenge, and the student was also financially motivated to succeed at this project since the funding was all provided by the learner. As an independent special project the student was able to use the experience as an elective that was used towards graduating with his Bachelor of Science degree in Information Technology.

PROJECT-BASED LEARNING

The nature of this technology demonstration paper is framed within project-based learning. This paragraph will provide brief context to further explain the catalyst for this undergraduate student project. Real world problems that learners are personally interested in or intrinsically motivated to learn about are great opportunities to allow students to learn and apply new knowledge to complex tasks. Project-based learning (PBL) is a model that allows students to explore, inquire, design, and make unique decisions to create their best possible solution. PBL uses authentic driving questions, a community of inquiry, and the use of technological based tools to create the atmosphere where students are more engaged and have the possibility to learn more than a traditionally structured course (Krajcik, Blumenfeld, Marx, & Soloway, 1994).

Constructivism plays an important role in independent project based learning that was used in the Raspberry Pi Supercomputer project because it allowed the student to use prior knowledge learned through his previous course-work and his experiences using technology and apply what he already knew to the complex challenge of creating a machine capable of parallel processing and storage. Constructivism and PBL allow learners to learn new basic skills that fill in gaps or tie together their prior knowledge and experiences (Diehl et. al., 1999). Project-based Learning affords a unique opportunity where the student that is intrinsically motivated to work autonomously over extended periods of time and have frequent but short directed meetings with the instructor overseeing the student's teaching (Thomas, 2000).

In the Celer-Pi project the student was interested in creating a machine capable of spreading the processing instructions across multiple devices. The student read a 2012 article where a team of engineers at the University of Southampton, UK built a supercomputer called, "Iridis-Pi" for approximately $4031.00 plus switches and cables. The Southampton project utilized 64 Raspberry Pi devices and Lego building blocks to create the case (Sharma & Williams, 2012; Cox et. al., 2013). The student author of this paper approached his professor to get some basic advice about managing a cluster using a Windows Enterprise version. Through the first one-on-one meeting the student and professor guided the learner to change his initial idea. The new idea was to find a different method and different software to have roughly the same computing ability as the Iridis-Pi. The initial goal was to keep the entire cost of the device at about 10 percent of the Southhampton build, which would have the Celer-Pi creation with a target price of approximately $450.

In exchange for the large dedication to the project the student would be able to use the project as an upper-level special project to be used toward graduation. The Raspberry Pi is an inexpensive credit card size device that can plug into a television and computer keyboard to provide an avenue for basic tasks that would usually require a personal computer, like word processing, spreadsheets, watching video, and playing games. The Raspberry Pi Foundation also wishes to encourage children across the world to learn programming by providing open source code and a very low purchase price of $25-$35(Raspberry Pi Foundation, 2014). The idea of using Raspberry Pi devices for more than their intended use of being a stand-alone inexpensive computing device is very popular right now as shown in the weblog 40+ Cool Ideas for your Raspberry Pi Project (Tom, 2012). Using the Pi devices would allow processing and storage to be spread across nine devices with one of the Pi devices serving as the master to coordinate the computing. There were other supplies needed to complete the build including a switch, cabling, memory cards, and some basic hardware to complete the build as detailed in Table 1.

Item Description	Quantity	Cost each
50-Pack RJ-45	1	3
Adafruit 16x2 LCD	1	20
Cisco SR2024CT 24 port 10/100/1000 Compact Switch	1	178
Cisco-Linksys WRT54G Wireless-G Router	1	40
Ethernet Cable CAT6 – 100FT	1	9
Static IP Address Lease	1	15
Raspberry Pi Model B Revision 2.0 (512MB)	10	35
Sony 32GB SDHC/SDXC Memory Card	10	15
Hardware for custom built rack		135
	Total	**$900**

Table 1. Total Project Expense

Problems facing parallel computing today continue to prove the validity of Amdahl's Law originally presented at AFIPS Spring Joint Computer Conference of 1967 in Sunnyvale, California (Smith, 1984). At some point according to Amdahl law, more hardware does not equal more performance. At some point the value of adding more cores to fix problems levels off.

According to this law IT personnel cannot just throw 1000 processors at a job and expect it to automatically give each node 100% of the work load. Software has to be written to take advantages of the architecture of the system. Environments not optimized for a parallel system will not use all of the available resources of the system. Gene Amdahl stated at the conference of 1967, that "The only difficulty is involved in knowing how to exploit this additional hardware" (Amdahl, 1967).

New parallel computing techniques are constantly being developed by skilled programmers using proprietary and Open Source material. According to Amdahl law, software must be applied in order to gain advancements in parallel computing (Amdahl, 1967). Multi-processing software is currently being engineered with MPICH2, IBM, MIT, Hadoop, Cuda, and Debian/Linux systems. Updated versions of the software provide important software security updates and generate increased use of each core processor in the multi core environment.

The ability of users to keep up with technology that grows at an ever expanding rate is impossible even for the technically minded. Applied knowledge is only correct if the system researched is understood and correct standards are applied. Due to this great technological separation, educated users must continue to cipher through material that supports the validity of certain tasks they are trying to accomplish with computers. Being able to reference valuable material and deciding cognitively why it is needed is essential to implementing the correct process. Applied knowledge is then correctly administered using commands initiated in standard operating procedures format.

Since the release date of the 32 bit Pi for education purposes in 2011 the company has steadily paced a rate of sale at 1 million per year. On the 17 of November 2013 Raspberry Pi sales crossed the 2 million mark. The Pi is becoming ever more popular with users and Linux distributions following suit. The credit card sized device offers great capability for the creative minded.

Having access to a relatively low retail device gives a student the ability to study new areas of technology. The Pi was first considered for the potential possibilities Pi could offer for web development and advertising internet market purposes. Using Linux Open source software the project had the capability of being fully functional when properly configured.

The technology industry provides many open source repositories from which to choose an operating system or web server from. During the process of the build several multi operating systems and web servers were researched then implemented. In search of a Linux operating system built for the Raspberry Pi ARM processor Debian software was considered. Debian Squeeze software was found to not yet offer specific updates for the Pi HF processor. Do to this interoperability issue Debian Squeeze was deemed less dependable.

The Raspbian software chosen has direct support from the Debian Wheezy library. This insures the OS is constantly updated directly from the Debian repositories increasing its stability and security. Information sent to the Pi was created by Debian Wheezy to provide updated code specifically for the Arm HF processor on the Pi. The updates are tested at Debian before being sent to the Pi cluster system. This effectively reduces the possibility of conflicts with the cluster or system wide crashes. The processors benefit from updates received, gaining increased operability and run capacity.

Apache2 and Nginx were both considered and implemented into the Celer-Pi. Apache2 contains more modules than Nginx yet, both work well with large scale clusters. Apache is seen more in major transactions and Nginx is light on resources when dealing with quick small file transfers. The main development computer had sufficient 300GB SSD memory and Duo AMD Turion X2 64 processing capability and received Apache2.

To respond and deliver information from client requests Nginx 1.4 server software was implemented into the Pi cluster. The web server is lightweight yet robust containing powerful code that operates popular sites such as Facebook, Hulu, Pinterest, Dropbox, and Word Press to name a few. Since Nginx is a relatively newcomer in the market the web servers software requires knowledge to be obtained daily and immediate implementation of security of data applied. Securing the server software is of vital importance in the overall health of the system.

Open SSH version 6.0 was initialized to place the main computer into master mode and allow connection to the network. SSH applies proven encryption methods transferring data securely and protecting the integrity data. Open SSL was installed to use the cryptography tool kit manager on the Session layer of the OSI model. SSL sits between http and TCP providing security protection. The software provides many key cryptographic functions such as Time Stamp requests that support generation and verification. Open SSL 1.01 E was installed to support a number of different cryptographic algorithms such as Ciphers, hash functions, Public-key, and Apache licensing.

The distributed file system is accomplished by the MPICH2 version of the MPI standard. MPICH2 was chosen for its interoperability with a wide range of devices. It is a high performance and lightweight alternative to the original MPICH software build from National Argonne Labs. MPICH2 software adds capability of controlling multiple processors from one master location through the combined use of cluster architecture and networked nodes. Using MPICH2 software the total environment can be controlled by the master Pi or independent environments can be controlled separately.

Once the initial build configuration was complete, initiation of communications using MPICH2 build data was calculated based on number of subsystems. Using the master Pi a calculation command was then sent to the subordinate Pi systems. Data received at the subordinate Pi cluster is included instructions for portions of data sets to calculate and send results back to the master Pi. The master Pi receives information from the ongoing processes of each sub machine and displays the data.

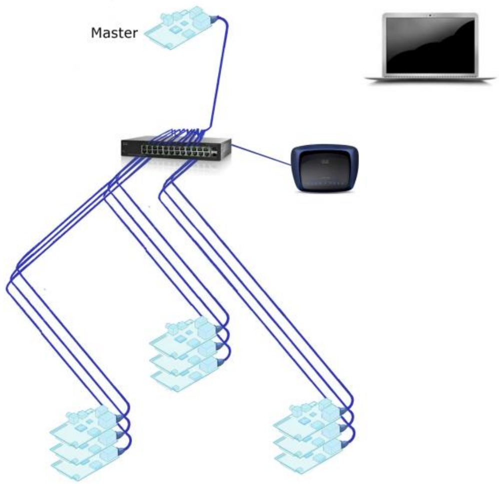

Figure 1. Cluster design

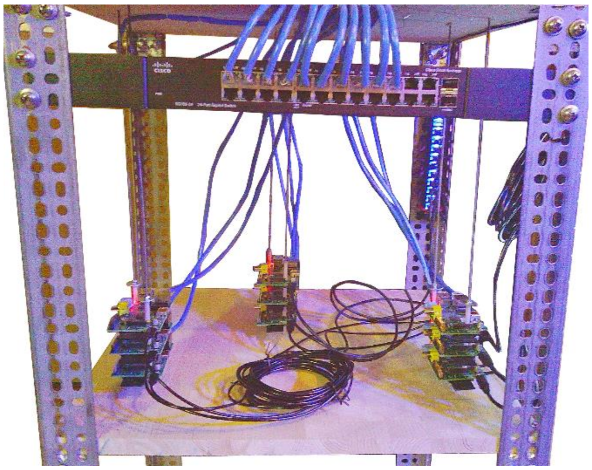

Figure 2. Celer-Pi completed

The system provides the ability to develop, administer, and secure an online website for the recurring cost of a $12 domain name and $15 static IP lease each year. Each Pi device is able to support 8-10 websites with PHP5, MySQL, and email servers in the cluster. The Celer-Pi, shown above in Figure 2, has a small physical footprint, uses a small amount of power, and does not require additional cooling beyond that of a normal office environment. The design of the Celer-Pi allows for scalability to add additional Pi devices with minimal configuration, as well as the ability to quickly and inexpensively replace any single Pi device that may fail.

At the hardware level each computer system comes complete with its own Mac/IP, hardware, core, Solid Drive, and Ethernet 10/100 network connectivity. Only the master Pi, as shown in Figure 1, is able to reach the outside network. Clients are allowed communicate with the master Pi through http port 80. The independent environments can only be accessed through direct communication from the master Pi. The number of users that can be responded to or data that can be analyzed by the master is limited only by the bandwidth and the amount of solid state memory available to the system.

The master Pi to Pi stack relationship allows for the separation of MySQL, PHP5, Nginx, and email servers on individual Pi. This relationship increases response and up time in websites through increased availability of resources and data communications, effectively removing single bottleneck points of failure.

The current system setup allows for quick addition of new Pi and Solid Drive expansion to the distributed file system network. Main memory and backups can be safely stored off the network using Nginx and Solid State Memory. If any computer in the cluster or even the whole cluster crashes Wheezy software images, Nginx web servers, and user data from MySQL will be refreshed using offsite Solid Drive card backups from the main computer. The complete Pi system could be reformatted in 10-15 minutes.

Challenges

Google was chosen to purchase the domain name from due to the proximity of the company's only eastern seaboard data center located 50 miles away in Griffin, Georgia. The purchase of the domain name affords administrative access to Google's cloud DNS and business development tools.

The Google Corporation uses advanced servers and fiber wire for data communications inside the datacenter's network and its networks across the world. The corporation has well documented planned future fiber wire installations in the surrounding major cities within the next few years. This will allow the Pi cluster to have a direct fiber wire link to the facility. Once the connections are made using fiber optics from the datacenter the Pi cluster will be able to respond to a user in any geographic location in the world. The Pi cluster will continually benefit from these high speed data connections that will provide it and its users with high quality data transmission speed, availability, and security.

A static IP was leased to allow any user to find the network that is browsing the internet and wants to view a site being hosted by the Celer-Pi. The router allows for the look up of the master Pi which then finds the subordinate Pi addresses on the personal network and requests data. Under the current hardware setup the Cisco switch can support up to 24 connections simultaneously to individual Raspberry Pi computers. The Pi devices are mounted to the 4mm all thread wire with added plastic spacers separating each individual Pi. This is accomplished using pre-drilled holes from the factory. Each Pi in the cluster is therefore suspended allowing for greater heat exhaustion if needed.

Now that the hardware has been purchased, the software installed, and network configured the system must be constantly maintained and improved upon. The server administrator must continue to apply knowledge of practical standards in the technology industry to keep a constant check on the security health of the system.

Security of the data and ability to monitor the network is a must. Updates can be implemented system wide using scripted commands from the master Pi. A LCD display soldered onto the master Pi provides a continuous live access feed of the network monitoring and communications. Safe back up and security of the data provides enhanced development, stability, and growth.

So far the Pi cluster has been running without power or data interruption since October 12, 2013. The system is designed for addition of more switches and independently added Raspberry Pi nodes to the cluster. The server rack perimeter and computer hardware size calculations suggest that with addition of switches 100 Pi can comfortably reside within the rack.

CONCLUSION

Amdahl was one of the first to realize the power limitations of the single processor design. Applying parallel computing major corporations can solve large scale problems proficiently. An example of multiprocessing used daily is Google's search engine ability to index hundreds of millions of web pages a day and display them to users in under a second. Data grows at such an exponential rate some enterprise companies have recently had trouble keeping up. Companies have had to adapt and employ the use of parallel computing needed for intensive use of the system resources. The technologies in these environments are enhanced infrastructures designed to support increasingly large amounts of data with increased efficiency. Alleviating other resources and increasing revenues for the company. International Business Machines, Oracle, Amazon, Google and Microsoft's heavy investments into multi-processing research helps solidify that the future is infrastructure built with simple components and central processing units combined with software engineered with a main goal to support multiprocessing.

The university was able to allow the student to gain course credit to be used towards graduation for a project that was personally interesting to the learner. Projects that are student-centered complex real-world problems encourage students to engage in life-long learning and students are likely to be very satisfied with the experience. The independent nature of the project-based learning in this particular case allowed the student to conduct in-depth research and work on the project in depth. The college will continue to offer this type of experience to students that are personally interested in a topic that is beyond the scope of the program curriculum, but certainly a valid study for an undergraduate Bachelor of Science in Information Technology student to learn more about. The student in this case will maintain the device that he has created and is looking at opportunities to turn this endeavor into a business idea by hosting websites and potentially recreating the low-cost supercomputer to sell as a whole device.

ACKNOWLEDGMENTS

I want to thank the owner of Dynapp, Wes Byrd and the lead developer Albert Maine for their assistance with obtaining a business internet connection to run the Celer-Pi. Dynapp allowed me to intern where I learned many of the skills that were utilized in this project. I also wish to thank the co-author of this paper, Professor Johnathan Yerby for his guidance and time dedicated to making this project such a success.

REFERENCES

1. Amdahl, G. (1967, April). *Validity of the single processor approach to achieving large scale computing capabilities*. AFIPS '67 spring joint computer conference, New York, NY. Retrieved from http://www-inst.eecs.berkeley.edu/~n252/paper/Amdahl.pdf
2. Cox, S. J., Cox, J. T., Boardman, R. P., Johnston, S. J., Scott, M., & O'Brien, N. S. (2013). Iridis-Pi: a low-cost, compact demonstration cluster. *Cluster Computing*, 1-10.
3. Diehl, W., Grobe, T., Lopez, H., & Cabral, C. (1999). Project-based learning: A strategy for teaching and learning. *Boston: Center for youth development and education*.
4. Gohring, N. (2013, May 7). Own your own cray supercomputer for a mere $500,000. *CNN Money*. Retrieved from http://money.cnn.com/2013/05/07/technology/enterprise/cray-supercomputer/.
5. Krajcik, J. S., Blumenfeld, P. C., Marx, R. W., & Soloway, E. (1994). A collaborative model for helping middle grade science teachers learn project-based instruction. *The elementary school journal*, 483-497.
6. Pounds, A. J., Nalluri, R., & Coleman, B. L. (2005). The development of a Tri-Use cluster for general computer education, high performance computing education, and computationally intensive research. *ACM Southeast Regional Conference*, 345. doi:10.1145/1167350.1167446.
7. Raspberry Pi Foundation. (2014, January). *Faqs - Raspberry Pi*. Retrieved from http://www.raspberrypi.org/faqs.
8. Sharma, A. & Williams, B. (2012, September 16). Build your own supercomputer at home with Raspberry Pi computers. *Mest Education*, Retrieved from http://mesteducation.com/build-your-own-supercomputer-at-home-with-Raspberry-Pi-computers/.
9. Smith, J. E. (1984). Decoupled access/execute computer architectures. *ACM Transactions on Computer Systems (TOCS)*, *2*(4), 289-308.
10. Thomas, J. W. (2000). A review of research on project-based learning. Retrieved from http://www.ri.net/middletown/mef/linksresources/documents/researchreviewPBL_070226.pdf.
11. Tom, R. (2012, December 3). [Web log message]. Retrieved from http://pingbin.com/2012/12/30-cool-ideas-Raspberry-Pi-project/.