To what extent is distributed computing an effective replacement for supercomputing?

IB Extended Essay

Session: May 2021

IB Subject: Computer Science

Personal Code:

Word Count: 3858

Table of context

Introduction

Supercomputing has existed for as long as computers have. The first computer was designed to perform calculations. Supercomputing follows the same principles, but as time has passed these calculations have become infinitely more complex, and as a result, computers have evolved to meet those requirements. Supercomputers perform calculations that can lead to anything from possible cancer cures, to better predicting the weather. Supercomputers lead to innovation in computers that often work their way down to personal computers.

Supercomputers are simply very powerful computers. They can perform calculations in seconds that would take a normal computer years. This is often due to these computers being far more powerful and having more processors. A processor is the “brain” of the computer, it performs most of the calculations that the computer requires. The processor, also known as the central processing unit (CPU), is made up of a control unit that contains the memory address register (MAR) and memory data register (MDR). The part of the CPU that performs calculations is known as the Arithmetic Logic Unit. The speed of a CPU is measured by the number of cycles it can perform per second. Modern CPUs can perform billions of cycles per second(GHz). The fastest supercomputers can have millions of processors when taking into account the graphical processing unit(GPU) processors. GPUs have the same components as a CPU, but have a far simpler structure. Where CPUs will have multiple levels of cash (very fast memory) to reduce the latency as much as possible. GPUs tend to have faster graphics DRAM, known as GDDR, and just one or two levels of cash. This is because CPU's have low latency tolerance, and GPUs have a higher latency tolerance. GPUs do not include many of the more complex parts of CPUs such as branch predictor units, speculative units, or Out-of-order execution units. This frees up the space to include a lot more ALUs. More AlUs means that more threads of computation can be run at the same time. This allows GPUs to have high throughput which is a necessity when processing millions of pixels per second. GPUs are mostly used for applications that have data parallelism, such as graphics, physics, and scientific computing. Supercomputers can also have multiple GPUs. This combination makes supercomputers hundreds of thousands of times more powerful than most personal computers (PCs)("Shingari").

Computer performance is often measured in FlOPS or Floating Point Operation per Second. While not always a perfect representation of real-world performance for personal computers, for supercomputers it is the only comparable measure of processing power. FLOPS are especially good at measuring how powerful that computer is in fields of scientific calculations that make heavy use of floating-point calculations. This is also a measurement that is used on other devices such as gaming consoles, and computers. Modern supercomputers can perform thousands of floating-point operations per second.

Origins

Computers were always used to perform complex calculations, and as computers became more common, their use cases increased. There started to be a divide between computers that were used for scientific use cases and personal computers. Scientific use cases became more common and required more and more processing power. In the 1960s this market for more powerful computers was mostly controlled by Remington Rand and IBM, who also produced less powerful computers. The man largely credited as the creator of the supercomputer, Seymour Cray, had worked for Remington Rand but left to work for Control Data Corporation. His company was dedicated to creating very fast (powerful) scientific computers for a small lucrative market. They largely succeeded in their goal, making the CDC 1604 which was very popular in scientific labs. This was in part due to the use of transistors instead of vacuum tubes. In response, IBM made the 7030 also known as the sketch. While IBM was able to produce the most powerful computer in the world, IBM's vacuum tube-transistor hybrid had few clients causing them to lose 20 million dollars and leading them to withdraw from the market for some time. CDC quickly retook the title of the fastest computer in the world with the CDC 6600's 3 million FLOPs (floating-point operations per second). The CDC 6600 was also coined the term supercomputing. Afterward, Cray went on to start his own supercomputer company in which he had a pivotal role in every part of the creation of the supercomputer. This company pioneered the technology that is now used in almost every computer around the world ("Hosch").

Distributed computing systems started popping up from 1970-1977. At first, this was only done in labs where all the computers were owned by the same company as the supercomputer itself. These were mostly academic tests, but by the 1990s the Great Internet Mersenne Prime Search changed that. GIMPS was the first distributed computer to use volunteer computing. Volunteer computers presented a new set of challenges, mainly revolving around reliability and security. Volunteer computing allowed distributed computers to grow from a few hundred computers to a few hundred thousand (“Beberg”).

Volunteer computing was the next big step in distributed computing. Volunteer computing allowed the public to volunteer their personal computers to a distributed computer network. This means that the organization running the distributed computing network only needs to invest in servers and architecture. The computing part is done by the public. This limits the computational power of computers to the number of people that volunteer their computers (“Beberg”).

In 2000 the folding at home project started. This distributed computing network is dedicated to folding protein simulations. These simulations focus mostly on the misfolding of proteins, which can lead to many diseases. These simulations require a tremendous amount of computing power due to the nature of protein folding. Being a volunteer-based distributed network, the main limitation of this is the number of volunteers it has and the computers those individuals have. Folding at home is currently the most powerful distributed computing network, at times even being the most powerful computer network in the world-beating out modern traditional supercomputers (Beberg).

Many colleges and universities performed significant amounts of research about distributed computing, in both volunteer and closed network environments. Berkeley University's Berkeley Open Infrastructure for Network Computing (BIONIC), this infrastructure has been used by dozens of organizations, for both volunteer and closed networks. Originally released in 2004, it is still widely used today.

Supercomputers started with a goal, unlike distributed computers and their lab experiments, but as time went on distributed computers found their own applications. Computing applications grew and both supercomputers and distributed computers started to be more common, similarly to the household computer they became a part of everyday life.

Architecture

Early on supercomputers started using multiple processors. At first this was used to connect hundreds of cheaper chips together to make the computer faster. This was the case for the connection machines CM-1, which had over 65 thousand cheap single-bit processors that operated in groups of 16 to make 16-bit chips. This was a cheaper alternative to using complex expensive processors. At first, the computing tasks were distributed and organized by a single processor, This limited the speed of the computer to the speed of one single processor. In response to this bottleneck, supercomputers went in favor of decentralized controls, which were also used in the connection machines CM-1. This multiprocessing approach allowed the computer to perform multiple tasks simultaneously rather than switching between the two on a single processor. Once decentralized controls matured the limitations of supercomputers became how many processors it had. The top supercomputers have thousands of processors all working together (“Hosch”).

Distributed computers have a similar architecture to older supercomputers. Where the computational tasks are assigned to different processors in older supercomputers, here they are assigned to different computers by a server. Except where modern supercomputers switched to a decentralized control structure, distributed computers did not. By continuing to have a centralized control structure the distributed computing network could be limited to the speed of that server and the number of tasks it sends out. There is also time lost due to the nature of distributed computing being spread out, the tasks are sent over the internet and have to be downloaded which can take time. When the tasks are completed they have to be sent back to the server (Beberg).

The folding at-home project uses a client-server architecture similar to many other volunteer-based distributed computing networks. In this situation, the client is installed by the volunteer on their personal computer. That client then requests a task from the assignment server, which assigns it to a work server. The client communicates with the work server and requests a work task. The client downloads the work task sent by the work server. Then the volunteer’s personal computer performs the task, and the client sends it back to the work server and logs it accordingly. Folding at home also has a credit system to give credit to the volunteers that donate the most computational resources, and has a leaderboard on their website. The servers also generate the work task that the clients will complete (Beberg).

Supercomputer Fugaku based out of Japan is an ARM-based supercomputer instead of using more traditional x86 processors. This type of processor is physically used in mobile, and other small lower power devices. Fugaku is the first example of a high-performance ARM computer. Supercomputers are one of the best use cases for ARM processors as programs have to be coded specifically for ARM and x86 programs often don't run properly. In supercomputers, the programs are custom coded for that specific supercomputer and as a result, the programs can be designed for ARM processors.

The top supercomputers have started to reach limitations in the speed that data can be transmitted, with IBM’s summit moving memory and storage physically closer to the CPU’s and GPU’s in an attempt to achieve better performance. When Summit was built IBM chose to have more GPU than most typical supercomputers have because of the workloads that they wanted to focus on, ending up having 3 GPUs per processor, to accelerate some tasks. This improves the overall speed and processing power of the supercomputer. (“Shankland”).

Processing power

The raw processing power of supercomputers is difficult to comprehend. The processing power is measured in Flops, but modern supercomputers are measured in trillions of flops. While the top supercomputer is constantly changing, the current top computers have over a trillion FLOPS (teraFLOPS). The top supercomputer at the time of writing this is the Supercomputer Fugaku based in Japan. It has over 7 million cores (CPU & GPU). The raw computing power of these computers is always increasing as the technology behind it evolves, whether it is improvements in the physical size of the chips, or it is improvements to cooling, every iteration of the supercomputer is faster. Most of the top supercomputers are from the United States and China, with the US currently holding the number 2 and 3 spots. Japan's Fugaku has 415,530 TeraFlop/s of processing power. That is over double the amount of the next supercomputer on the list, named Summit. IBM's summit has a comparatively low 148,600 teraFlop/s with 2,414,592 cores ("Top 500 June 2020."). But that is still significantly higher than the next supercomputer on the list. Computers are built for different tasks and thus focus on different hardware. These supercomputers also have different focuses, some choose to have many more graphical processing units to accelerate tasks or focus on rendering.

Calculating the performance of a distributed computing network can be more complicated than calculating the performance of a traditional supercomputer. If it is a volunteer-based distributed computer network (which most are) the performance can vary greatly from month to month, year to year, based on the public's perspective on the research the distributed computing network is working on.

The folding at home computing network became much more powerful during the covid-19 pandemic. The public wanted a way to help covid-19 research, and Folding at home was working on covid-19 research. The increase in users was so great that the

Folding at Home team started being limited by their servers. At the time of writing Folding at home has 17,278,972 CPU cores, and 774,919 teraFlops of computing power. This number is constantly changing as more people volunteer their computing resources. Folding at home releases the number based on the computers that have returned work units within the last 50 days. As the covid-19 pandemic has progressed the number of volunteers has decreased and folding@home has lost some processing power (“Folding@home Stats”).

Applications

The computing power of a supercomputer is often based on what the end goal is for that computer. Since the beginning of computers, they have been used by the military to conduct different research. That hasn’t changed, and probably won’t. At first, cryptography was most common, along with a plethora of other national security projects. Now nuclear research is a particularly common use for supercomputers along with chemical research and cryptology. Outside of military applications, supercomputers are still used but often don’t require the same level of computing horsepower. The aerospace, petroleum, and automotive industries all use supercomputers as part of their research and development.

Distributed computing has similar applications to traditional supercomputing and they often end up working on very similar tasks. Volunteer distributed computing is where these applications stray more, due to the nature of volunteer computing. Volunteers prefer to know what their computers are working on as it is their personal property. As a result, most military applications are not suitable for volunteer computing. The public tends to prefer to support causes that are close to them, or have a direct impact on them or their family. This all depends on the public's reaction to the distributed computing network, because without volunteers there would be no computing power.

One of the original volunteer computing networks is SETI, also known as the search for extraterrestrial intelligence. It was founded in 1999 and operated for 21 years. SETI uses the Berkeley open infrastructure for network computing (BOINC) which is also used by many other distributed computing networks. Rosetta@home is another distributed computing network that is a part of BOINC, its focus is similar to that of folding at home’s it focuses on predicting and designing protein structures to researching cures to major human diseases. BOINC based programs send the volunteer's computer instructions, and download the necessary applications. Then the volunteer’s computer performs the computations based on what is sent in the instructions and uploads files to the project servers. Similarly to folding at home, BOINC also has a credit system to reward volunteers for their computations. This credit has no monetary value, it is simply proof of volunteering.Bionic is responsible for dozens of volunteer based distributed computing networks.

DreamLabs is another Medical research-based distributed computing network. But unlike Folding at home DreamLabs network mostly consists of mobile devices. Dreamlabs is also affiliated with Vodafone, which helps them have a reach around the world. DreamLab focuses mostly on cancer research but during the Covid-19 pandemic, they like many other distributed computing networks shifted to help find a cure for Covid-19. DreamLabs had over 500 thousand downloads on android alone. While mobile devices tend to have less powerful processors, there are more of them around the world, they can make up for the lack of processing power by having more devices ("DreamLab”).

Cost

Supercomputers are very expensive to plan, build and operate. As a result they often have sponsors that help fund their operations. In return, the sponsors can use the supercomputer for their own research. For example, IBM’s summit was sponsored by the U.S Department of Energy. Summit is said to have cost 200 million to build. In return for their sponsorship summit researched alternative energy and energy efficiency. These sponsorships do limit the use cases of supercomputers. Due to their inherent high cost many smaller projects simply can’t afford to use a supercomputer. Military applications are very common because militaries often sponsor supercomputers, and as a result can take advantage of their computing power for months or even years. Supercomputers require a lot of power to operate and produce a lot of heat. Most supercomputers operate 24/7, only stopping if a component is broken. Even then the broken component would be isolated, and the rest of the computer would continue working. Summit uses somewhere between 10 and 15 megawatts of power to operate, depending on the computing task placed on the machine. All computers require some form of cooling, while modern processors have thermal limits to stop them from overheating and breaking. Properly cooled processors can run at much faster speeds and be far more efficient. Most laptops and personal computers simply use a heatsink and some fans. A heatsink is a device that absorbs heat produced often made of copper as it is a good heat conductor. In computers heatsinks are generally placed on the CPU and GPU to absorb the heat produced. To dissipate this heat, heatsinks use metal fins to increase their surface area. And use a fan to pass cool air through the fins, cooling the CPU or GPU. , More powerful computers require more cooling. This is accomplished by using liquid cooling (water cooling). The water is pumped through a heatsink on the CPU and GPU, then through a radiator or heat exchanger to dissipate heat and cool the processors. IBM's Summit uses liquid cooling to cool all their CPUs and GPUs. To do this 4 thousand gallons of water are pumped through per minute ("Rosenfield").

Supercomputer Fugaku's ARM approach to supercomputing has a lot of advantages when it comes to the cost of the supercomputer. ARM processors are lower power, lower cost in comparison to x86 processors. While this lowers the cost of each individual processor the total development cost totaled about 1 billion dollars. This total includes 6 years of technology development and design. This high cost is partially balanced out by the lower operating cost of ARM processors, as they are less power-hungry, and produce less heat.

Distributed computing costs are far lower than those of supercomputers. This is largely thanks to not having to pay for thousands of processors and the high operating costs associated with them. Servers are the only cost for distributed computing. The number of servers required is based on the number of volunteers the network has, and more servers can be added at any time if needed. With fewer servers, there is a lower operating cost. The volunteers end up paying most of the operating costs simply by being the ones owning the computers. Most operating costs come in the form of electricity consumption, heat output, and possibly reduced life of the CPU and GPU. All of these costs are dealt with by the volunteer. But generally, these are lower power computers that use less electricity, and with adequate cooling, it would have a very small impact on the lifespan of the CPU and GPU.

Conclusion

Supercomputers are a far more mature technology that has been tried and tested. Distributed computing as we know it today has only existed in the last 20 or so years. And in those 20 years, distributed computing made a lot of progress. But now computers are more common than ever before. The potential of distributed computing increases every day, every hour as a new computer is made. Phones are computers. There are more “smart devices” than ever before, each of them has a processor. Powerful or not, combine all of those devices together and the computing power would be unrivaled. More than a billion devices are running the Windows operating system, another 2+ billion running Androids, and over a billion running Apple’s IOS. These numbers increase exponentially every year. And this is not including all the other devices that now have microprocessors. From fridges to door locks, everything has a microprocessor. A lot of these devices spend the majority of their lifespans doing fractions of what they are capable of. Simply taking advantage of the unused processing time would create a distributed computer network faster than anything else on Earth. But that is not likely to happen, for the same reasons more people don't donate their computing time to Folding at home or any number of other distributed computing networks. People are often worried about the security risks of allowing their computers to perform tasks without knowing what the tasks are.

A more realistic option would be for companies that already deploy hundreds or thousands of computers to their employees that just sit in their offices unused for hours to use those in a distributed computing network. In this instance, the company could use this computing power to do anything from market research to research and development of future products. The company would also not have to pay for the uses of an external supercomputer. But this would only work effectively for companies with thousands of employees as these computers are often not very powerful. This would also not be an option for a lot of medical research and private research as those often don't have hundreds or thousands of computers laying around. Volunteer computing could be an option for either of those as a more cost-effective solution than a supercomputer.

Supercomputers still have their advantages over distributed computing, mainly because for distributed computing all tasks (work files) have to be sent and downloaded onto the volunteer's computer. That takes time. While the large distributed computing networks can make up for this with raw computing power, it is an inefficiency that can not be solved yet. Supercomputers such as Summit made strides in reducing the inefficiencies from storage/memory transfer rates. Working with invidia, IBM's Research team created an NVLink between the CPUs and GPUs for Summit. A proprietary interface that allows CPUs and GPUs to share data up to 4X faster than x86-based systems. To make the system more efficient at moving data IBM put in extremely high-speed connections between the 4608 individual CPU/GPU compute nodes. Summit is also an example of a more specialized supercomputer, by having 3 GPUs per CPU it can accelerate specific tasks, and become more efficient at computing those workloads. The military application would also not be viable for distributed computing as most of that information is classified (“Rosenfield”).

Future supercomputers are always in the works, with years of planning and designing. The US is planning to finish a supercomputer capable of more than an exaflop per second. China is also working on an exaflop capable supercomputer. There is a new wave of supercomputing just over the horizon. These computers will be faster and more efficient than anything before them (“Khalili”).

While supercomputers have some advantages to distributed computing, distributed computing is far from where it could be. More and more distributed computing networks are coming out focusing on different topics. As mentioned earlier there are more devices with processors than ever before and there will continue to be more and more. And distributed computing will always be more cost-effective than building a supercomputer from the ground up.

Works Cited

Agliamzanov, Ramil, Muhammed Sit, and Ibrahim Demir. "Hydrology@ Home: a distributed volunteer computing framework for hydrological research and applications." *Journal of Hydroinformatics* Accessed 26 Oct. 2020.

Beberg, Adam L., et al. "Folding@ home: Lessons from eight years of volunteer distributed computing." *2009 IEEE International Symposium on Parallel & Distributed Processing*. IEEE, 2009. Accessed 26 Oct. 2020.

Curtis, Vickie. "Patterns of participation and motivation in Folding@ home: The contribution of hardware enthusiasts and overclockers." *Citizen Science: Theory and Practice* Accessed 26 Oct. 2020.

Clark, Don. "Intel Slips, and a High-Profile Supercomputer Is Delayed." *The New York Times*, The New York Times Company, 27 Aug. 2020, www.nytimes.com/2020/08/27/technology/intel-aurora-supercomputer.html. Accessed 26 Oct. 2020.

"Computer." *Britannica School*, Encyclopædia Britannica, 22 Sept. 2020, school.eb.com/levels/high/article/computer/117728. Accessed 26 Oct. 2020.

"Distributed Computing." *Britannica School*, Encyclopædia Britannica, 31 Oct. 2012, school.eb.com/levels/high/article/distributed-computing/471610. Accessed 26 Oct. 2020.

"DreamLab app helps fight cancer and coronavirus while users sleep." *Vodafone*, Vodafone Group, www.vodafone.com/about/vodafone-foundation/focus-areas/dreamlab-app. Accessed 26 Oct. 2020.

"Exploring the GPU Architecture." *Niels Hagoort*, NielsHagoort.com, 12 Mar. 2019, nielshagoort.com/2019/03/12/exploring-the-gpu-architecture/#:~:text=Its%20architecture%20is%20tolerant%20of,the%20retrieve%20data%20from%20memory. Accessed 24 Dec. 2020.

"Folding@home Stats." *Folding@home*, stats.foldingathome.org/os. Accessed 26 Oct. 2020.

Hosch, William L. "Supercomputer." *Britannica*, Encyclopædia Britannica, 28 Nov. 2019, www.britannica.com/technology/supercomputer. Accessed 26 Oct. 2020.

INTRODUCING THE WORLD'S MOST POWERFUL SUPERCOMPUTER. NVIDIA Corporation, 2018.

Khalili, Joel. "I confess, I'm scared of the next generation of supercomputers." *techradar.pro*, Future US, 29 Aug. 2020, www.techradar.com/news/should-we-fear-the-next-generation-of-supercomputers. Accessed 26 Oct. 2020.

Larson, Stefan M., et al. "Folding@ Home and Genome@ Home: Using distributed computing to tackle previously intractable problems in computational biology." Accessed 26 Oct. 2020.

"Multiprocessing." *Britannica*, Encyclopædia Britannica, 2 May 2011, www.britannica.com/technology/multiprocessing. Accessed 26 Oct. 2020.

Perry, James Richard "Rick." "The Future Is in Supercomputers." *The White House*, 3 May 2018, www.whitehouse.gov/articles/the-future-is-in-supercomputers/. Accessed 26 Oct. 2020.

Rosenfield, Michael. "Introducing the world's smartest, most powerful supercomputer." *IBM*, 14 June 2018, www.ibm.com/blogs/research/2018/06/summit/. Accessed 26 Oct. 2020.

Rosetta@home. Baker Lab, boinc.bakerlab.org/rosetta/. Accessed 26 Oct. 2020.

Riabinin, Maksim, and Anton Gusev. "Learning@ home: Crowdsourced training of large neural networks using decentralized mixture-of-experts." *arXiv preprint arXiv:2002.04013* (2020).

Shankland, Stephen. "IBM's world-class Summit supercomputer gooses speed with AI abilities." *CNET*, 8 June 2018, www.cnet.com/news/ibms-world-class-summit-supercomputer-gooses-speed-with-ai-abilities/. Accessed 26 Oct. 2020.

Shingari, Davesh. "GPU Architecture." *Davesh Shingari*, Google Sites, sites.google.com/site/daveshshingari/explorations/computer-architecture/gpu-architecture. Accessed 24 Dec. 2020.

Shirts, Michael, and Vijay S. Pande. "Screen savers of the world unite!." *Science* 290.5498 (2000): 1903-1904. Accessed 26 Oct. 2020.

“Top 500 June 2020.” Top 500, TOP500.org, JUNE 2020, https://www.top500.org/lists/top500/2020/06/. Accessed 26 Oct. 2020.

Vincent, James. "World's fastest supercomputer will be built by AMD and Cray for US government." *The Verge*, Vox Media, 7 May 2019, www.theverge.com/2019/5/7/18535078/worlds-fastest-exascale-supercomputer-frontier-amd-cray-doe-oak-ridge-national-laboratory. Accessed 26 Oct. 2020.

Vodafone. Vodafone Group, www.vodafone.com.au/foundation/dreamlab. Accessed 26 Oct. 2020.

"What is Rosetta@home." *Rosetta@home*, Baker Lab, boinc.bakerlab.org/rosetta/rah/rah_about.php. Accessed 26 Oct. 2020.

Wolfson, Elijah. "The US passed China with a supercomputer capable of as many calculations per second as 6.3 billion humans." *Quartz*, Quartz Media, 9 June 2018, qz.com/1301510/the-us-has-the-worlds-fastest-supercomputer-again-the-200-petaflop-summit/#:~:text=At%20eight%20times%20the%20speed,use%20in%20artificial%2Dintelligence%20applications. Accessed 26 Oct. 2020.