CS434: Introduction to Parallel and Distributed Computing

Ndze'dzenyuy, Lemfon K.

20th February 2021

Lab 1

1 Part 1

1.1 The Global top 5 HPCs

As the scientific community looks to computing to speed-up our ability to better understand the world and solve long-standing problems, supercomputers have become very crucial for research. Supercomputers are heavily employed in the fields of mining, pharmaceutical research, weather forecasting, epidemiology, protein folding, and theoretical physics, amongst many others. As one begins the study of Parallel and Distributed Systems and High-Performance Computing, it is vital to have at least a basic knowledge of the world's top computing systems, how they are ranked, and their basic design architectures and principles. In what follows, I present short descriptions of the top five supercomputing systems in the world, as ranked by TOP500. TOP500 ranks and details the world's 500 most powerful supercomputers. These rankings were started in 1993 and are currently in their 56th edition. We also consult the GREEN500, another ranking that ranks supercomputers by power efficiency. After looking at the top 5 supercomputing systems globally, we look at the top-ranked supercomputer in Africa and discuss possible areas in Africa that may apply supercomputing.

1.1.1 Fugaku

The Japanese supercomputer, Fugaku, is today the world's highest-ranked supercomputer. It is found in the Riken Center for Computer Science in the Japanese city of Kobe. Fugaku was named after an alternative name for Mount Figi and is estimated to have cost a billion dollars. The Fugaku programme was initiated in 2014 by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) in 2014 to develop a flagship supercomputer for Japan that will succeed the K computer. At the time of the inception of the Fugaku project, the expectation was that it would take Fugaku only a couple of days to solve problems that routinely took K a year. The program hoped that the computer would help Japanese researchers focus on priority issues as selected by MEXT. With 7, 630, 848 cores and 5, 087, 232 GB of memory using the MPP Memory Architecture, and a Linpack performance of 442, 010 Tflops Fugaku is almost three times faster than the second-ranked computer. Its theoretical peak, at 537, 212 Tflops, makes it even a more unique computer. Fugaku uses the A64FX 48C 2.2GHz processor types, with the Tofu Interconnect D, and runs the Red Hat Enterprise Linux Operating System. The accompanying FUJITSU Software Technical Computing Suite V4.0 serves simultaneously as a compiler, Math library and MPI. With an optimized power consumption of 26, 248.36 KW and power efficiency of 15.42 Gflops/Watts, Fugaku ranks 10th on the Green500, a ranking for the most power-efficient supercomputers around the world. Very recently, Fugaku has been used in research on masks related to the COVID-19 pandemic. The Fugaku URL is https://www.r-ccs.riken.jp/en/fugaku/project.

1.1.2 Summit (OLCF-4)

Currently ranked the second most powerful supercomputer, IBM-manufactured supercomputer Summit was previously ranked first from its inception in 2018 till the appearance of Fugaku on the rankings in 2020. The Summit supercomputer was realised as a result of a contract between the US Department of Energy and IBM and is housed at the Oak Ridge National Laboratory in Tennessee. Summit was designed as a successor for its predecessor computer Titan. Summit was eight times better with regards to computational power at its completion, while only using 4,608 nodes compared to Titan's 18,688 nodes. Using the IBM Power9 22C 3.07GHz processor and the NVIDIA Volta GPU, Summit has 2, 414, 592 cores with a total memory of 2, 801, 664 GB. Each node has two CPUs and 6 GPUs, and a coherent memory addressable by all CPUs of 600GB, and uses the Cluster Memory Architecture. With a Linpack performance of 148 600 Tflops and a Theoretical Peak of 200 795 Tflops, Summit is almost three times faster than the higher-ranked Japanese supercomputer, Fugaku. Like, Fugaku, Summit runs a version of Red Hat Enterprise Linux but uses the Dual-rail Mellanox EDR Infiniband interconnect, an XLC compiler, Spectrum MPI, and ESSL and CUBLAS 9.2 as Math Libraries. With a power consumption of 10,096.00 KW and a 14.72 Gflops/Watts Summit is the 11th most efficient supercomputer globally. It is important to note that Summit is used mainly for civilian research. Thus far, it has been instrumental in earthquake simulation, extreme weather simulation with AI, material science, and Genomics. Two teams working on Summit recently won the highly converted Gordon Bell prize for remarkable work in High-Performance Computing. The Summit URL is: http://www.olcf.ornl.gov/olcf-resources/compute-systems/summit.

1.1.3 Sierra

In third place, we have the US supercomputer Sierra. Sierra was developed with similar architecture as the Summit supercomputer and was another product of a working contract between the Department of Energy and IBM, and is housed at the Lawrence Livermore National Laboratory. IBM, NVIDIA, and Mellanox manufactured Sierra. It uses 1,572,480 cores and has a memory of 1,382,400 GB implementing the Cluster Memory Architecture. Sierra also has 4320 Nodes, each equipped with two IBM Power9 22C 3.1GHz processor and four NVIDIA Tesla V100 GPUs. As a successor to the Trinity supercomputer, Sierra is also known as ATS-2 and is used primarily for predictive applications of US nuclear weapons stockpile stewardship, helping to ascertain their safety, reliability, and effectiveness. Sierra has a Linpack Performance of 94, 640 Tflops and a theoretical peak of 125, 712 Tflops. Like all other higher-ranked supercomputers globally, Sierra runs on a version of the Red Hat Enterprise Linux distribution. The Summit supercomputer uses an IBM XLC compiler, an IBM Spectrum MPI, and ESSL and CUBLAS 9.2 as Math Compilers. With a power consumption rate of 7,438.28 KW and power efficiency of 12.72 Gflops/Watts, Sierra is ranked the 17th Most efficient supercomputer by Green500. The Sierra URL is: https://hpc.llnl.gov/hardware/platforms/sierra.

1.1.4 Sunway TaihuLight

When the Chinese supercomputer Sunway TaihuLight debuted on the Tod500 in June 2016, it was ranked the world's most powerful supercomputer. It held on to that position for two years, after which it quickly fell to third and now fourth. The Sunway TaihuLight was designed and manufactured by the National Research Center for Parallel Computer Engineering and Technology. It is housed at the National Supercomputing Center in Wuxi, in the Jiangsu Province in the People's Republic of China. The Sunway TaihuLight is rumoured to have resulted from the US-China trade wars and uses Chinese-developed technologies predominantly. It uses a total of 40,960 Chinese-designed SW26010 64-bit designed using the Sunway architecture, with each processor chip containing 256 cores and an additional four auxiliary cores for system management. In total, Sunway has 10, 649, 600 Cores and a 1, 310, 720 GB Memory, using the MPP memory architecture and running a custom Sunway RaiseOS 2.0.5 Operating System. Sunway TaihuLight uses the Sunway SW26010 260C 1.45GHz processor type and has a Linpack Performance of 93, 014.6 Tflops and a Theoretical Peak of 125, 436 Tflops. With a power consumption rate of 15.371.00 KW and power efficiency of 6.05 Gflops/Watts, the Sunway TaihuLight, at number 40 on Green500, is the lowest-ranked supercomputer among the world's top supercomputers in terms of power efficiency. The Sunway TaihuLight project is estimated to have cost 1.8 Billion Yuan and is used mostly in oil prospecting, pharmaceutical research, weather forecasting and life sciences. The Sunway TaihuLight URL is http://www.nsccwx.cn/wxcyw.

1.1.5 Selene

NVIDIA's in-house supercomputer, Selene, is ranked as the fifth most powerful supercomputer in the world. Manufactured by NVIDIA, Selene uses AMD EPYC 7742 64C 2.25GHz processors and has over 555, 520 cores, with a 1, 120, 000 GB Memory that uses a cluster architecture. Selene uses a Mellanox HDR Infiniband interconnect and runs on the Ubuntu 20.04.1 LTS operating system. It has a Linpack performance of 63, 460 Tflops and a Theoretical Peak of 79, 215 Flops. For a compiler, Selene uses the NVIDIA NVCC V11, Intel Composer 2020.0.166, and the NVIDIA CUDA V11.0.148 and Intel MKL 2020.0.116 as Math libraries while running the OpenMPI 4.0.3 for a message parsing interface. Selene's interesting fact is that although it often takes about 9 – 12 months to build supercomputers, Selene was built in only three weeks at an NVIDIA data centre. At a power consumption rate of 2646.00 KW and power efficiency of 23.98 Gflops/Watt, the Selene is the fifth on the Green500 rankings and the most efficient supercomputer among the world's most powerful five supercomputers. The Selene URL is https://www.nvidia.com/DGXSuperPOD.

1.2 The top HPC machine in Africa

1.2.1 Toubkal

Toubkal is the fastest supercomputer in Africa and is housed at the Mohammed VI Polytechnic University in Ben Guerir, Morocco. Toubkal was developed through a partnership with experts from the University of Cambridge, Dell and Intel, and the African Supercomputing Center at the Mohammed VI Polytechnic. The goal is to provide cutting-edge supercomputing resources to African researchers and entrepreneurs with promising projects in the fields of Artificial Intelligence, Data Analytics, Genomics, Food Security, Agriculture and Mining. The Toubkal is manufactured by Dell and uses the Xeon Platinum 8276L 28C 2.2 GHz processors with the Mellanox HDR100 interconnect. The computer has a total of $71, 232$ cores with a $244, 224$ GB memory capacity running on a CentOs Scientific-OpenStack operating system. With a Linpack performance of 3158.11Tflops and a Theoretical peak of 5014.73 Tflops, the Toubkal is ranked number 98 globally and the first in Africa. It has a power consumption rate of 2646.00 KW and is ranked the 352 most efficient supercomputer globally. It uses an Intel compiler and the Intel MLK 2020.3 Math library alongside the Intel MPI 2020.2 message parsing interface implementation. The URL for the Toubkal is https://ascc.um6p.ma.

1.3 HPCs in Ghana?

Given that Ghana is only a lower-middle-income economy, one would not be wrong to wonder if it needs HPCs and what their potential uses could be. The debate on whether Ghana needs HPCs could be looked at from many angles. However, we will focus on highlighting a few areas in which HPC, should Ghana eventually have them, could be used in what follows.

1.3.1 Possible Applications

Earthquake Simulation and Prediction: Although being an African country, Ghana is not oblivious to the threat of earthquakes. Since the earliest recorded earthquake in Ghana, which occurred in 1615, Ghana has recorded more than ten more earthquakes. The latest of which was a 4.0 magnitude earthquake that occurred on June 24, 2020. Although the damage caused by earthquakes in Ghana thus far have not been on the magnitude of those witnessed in other parts of the world, it goes without saying that governments still have to prepare for every eventuality. Should Ghana purchase HPCs, they could be used in simulations to predict and help government officials plan accordingly for such earthquakes.

Weather Forecasting: Ghana could also use supercomputers in weather and climate forecasting. This is important and crucial for safety reasons and agricultural reasons. It should be noted that even with the current levels of technological development, the accuracy of weather and climate forecasts often border around 85

Research on pharmaceutical solutions for tropical diseases: As researchers around the world use supercomputers for genomics and other health-related research, Ghana could also employ these computers in research for cures for tropical diseases like Malaria and river blindness to name a few.

Space and Atomic Energy Research: Supercomputers could also be used in Ghana for Space exploration and atomic energy research. Founded in 2012, the Ghana Atomic Energy Commission and the Ghana Space Science and Technology Centre are in charge of nuclear energy usage/research and space exploration/research in Ghana. Researchers at these two institutions could benefit immensely from having access to supercomputers, permitting them to research a world-class standard.

1.3.2 Anticipated challenges of using HPCs in Ghana

A few challenges exist in using supercomputers in Africa. One such challenge is that power is mainly unstable, and the enormous amounts of power consumed by supercomputers may mean that such systems will have to look for stable alternate sources of energy. Another challenge is that there may be a shortage of adequately trained personnel to use the computer appropriately. This means that the computer may be underused or in want of persons to maintain the system properly. Also, most of the components are not produced in Africa, meaning that system failures that require new components will need those components to be bought and shipped from abroad.

2 Part 2

I installed Ubuntu 20.04.2 as a dual boot alongside my Windows 10 operating system. I learned how to use Visual Studio Code, but I am also learning how to use gvim. I wrote a program to calculate the sum of the elements in an array using pthreads. I created two threads, and made each calculate the sum of half of the array, and then I summed the sums that were calculated by the individual threads. I then compiled and ran the *hello_world.c* program that used OpenMP.

2.1 i

As indicated above, I chose to use a dual partition with Ubuntu 20.04

2.2 ii

I installed and learned how to use visual studio code, but also in the process of how to use gvim.

2.3 iii

I wrote a program to calculate the factorial of a number and compiled it using gcc.

2.4 iv

I wrote and ran a program that calculated the sum of the elements in an array using pthreads.

2.5 v

I compiled and run the code presented in the question statement.

2.6 vi

For the program to calculate the factorial of a number

Algorithm 1 Calculate the factorial of a number

1: **procedure** FACTORIAL(n)   ▷ n is the number whose factorial we want to compute.
2:     **if** ($n < 0$) **then**
3:         Tell the user that factorials cannot be computed for negative numbers
4:         Exit program
5:     **if** ($n == 0$ or $n == 1$) **then**
6:         Return 1
7:         Exit program
8:     $temp = 1$
9:     $j = 0$
10:     **while** $j < n$ **do**
11:         $temp = temp * (n - j)$
12:         Return temp
13:         Exit program.