Nicole Günther Guerrero

Medium entry

Before trying to solve a big data processing problem using parallel programming models and its deployment on the cloud, its important to know each of the concepts needed to do this. Starting with, what is big data? In simple terms big data refers to large quantities of data that are so large they can't be processed through queries on a regular computer. In order to process these large amounts of data, we usually use Cloud Computing. Cloud computing refers to mechanisms used to remotely take the data to perform operations to it. By using cloud computing we can go through data more quickly than we could on a normal computer. Knowing this we can now explain what a parallel programming model is, it is in simple terms a way to break complex tasks into smaller ones where you run all the smaller tasks simultaneously on multiple CPUs/computers/ virtual machines.

Having all the concepts explained we now can focus on one big data processing problem that would benefit from implementing it using parallel programming model on the cloud. Nowadays we produce approximately 2.5quintillion bytes of data everyday and companies have to do something with all the data. In order to process these quantities of data I would recommend the implementation of big data processing using parallel programming on the cloud. This would allow companies to process and manage the data in a faster manner.

The first steps needed to do this is to extract the data and start storing it on the cloud. In order to start the processing of the data. With the data stored some parallel programming models could be applied, so the data would be divided into smaller quantities and processed all at the same time. Depending on what the companies want to do to the data, the requirements of the parallel programing model will change.

References

Heavy.AI. (n.d.). *What is Parallel Computing? Definition and FAQs | HEAVY.AI*. https://www.heavy.ai/technical-glossary/parallel-computing#:%7E:text=Cloud%20computing%20is%20a%20relatively,typically%20reserved%20for%20large%20enterprises.

Price, D. (2021, December 13). *Infographic: How much data is produced every day?* CloudTweaks. https://cloudtweaks.com/2015/03/how-much-data-is-produced-every-day/#:%7E:text=Today%2C%20our%20best%20estimates%20suggest,a%20staggering%2018%20zeros!).

ScienceDirect. (2016). *Parallel Programming Model*. https://www.sciencedirect.com/topics/computer-science/parallel-programming-model#:%7E:text=A%20parallel%20programming%20model%20is,%2C%20and%20I%2FO%20systems.

Write a Medium like entry describing practically how to solve a (big) data processing problem that implies specifying operations that must be programmed under a parallel programming model and that can be deployed on the cloud.