Dennis Zang

Moustafa Alzantot

Lab 1

29 November 2017

Reducing the Massive Energy Appetite of Data Centers

In the recent decades, information technology has become an integral part of our economy, for it is ubiquitous in several fields of business, from storing videos on Youtube to storing private medical records of patients. However, as useful it may be for businesses to hold information, it is important to keep the costs of such a commodity at a minimum. In his article "Reducing the Massive Energy Appetite of Data Centers", Charles G. Choi emphasizes the significance of such costs and brings up a few sample solutions researchers have been coming up with recently to lower the staggering levels of power consumption.

Before we discuss the problem of and solutions to the power consumption of data centers, we should first generally look at what a database server is, what functions it needs to fulfill, and what components it has that allow for such functions. A data center is a location in which database servers are held and maintained, while a database server can be described as a computer that specializes in storing, retrieving, and computing with the information in its memory. Generally, the database server must be able to store the information of its clients, find and access the information quickly, and be able to serve multiple clients at a time 24/7. In other words, the database server must have enough processing power (from the CPUs) to work with a large number of clients and have the necessary storage to keep all of the information while maintaining a large amount of RAM (particularly DRAM for main memory) to allow I/O of data for its clients.

When considering the maintenance of such databases, we see that a costly limitation of keeping the database server functional is its consumption of power to operate it constantly and to keep it cool. Unlike normal desktop computers, a server must be able active at all times in order to service its clients. In addition to always being active, it is expected to access such data for multiple clients quickly, meaning that it must make use of a lot of RAM (particularly DRAM) for constantly caching information (especially for “in-memory databases”, in which main memory is primarily relied on for data storage). Both running the CPU 24/7 and having the RAM work require a constant source of power. As for cooling, some means of making sure the CPU and relate components don’t overheat is necessary, otherwise problems like thermal noise, leakage currents in semiconductors, transistor failures, and slowing performance. In fact, it is estimated that about 40% of the total power to maintain a database server goes to simply cooling down the computers themselves.

In total, the costs from the constant use of power can be staggering. All of the data centers in the United States uses about 2% of the total energy generated, or, in other words, 91 billion kWh. Given that there are about 3 million data centers in the United States, a single data center would use about 30 thousand KWh annually, which translates to about $3000 annually per data center (given the average cost of electricity is $0.1042 in the United States) and about $9.48 billion in total for all annually in 2015. Given that the cloud computing industry will still be growing, it would be lucrative to find solutions in making database servers more energy efficient, reducing the costs of maintenance.

Choi brings up three approaches to reduce energy expenditures in his article: design a more efficient cooling system, change the hardware that database servers use to manage memory, or make the actual CPU performance more efficient. As mentioned before, cooling requires a significant amount of power in order to keep the CPUs functional, but a large part of regulating CPU temperature also involves intelligent design and other outside factors. Using different implementations of memory-related hardware is plausible, as different types of hardware have different power requirements. In addition, bettering the CPU's efficiency can increase performance and decrease the amount of power and time the CPU needs for its computations.

The typical approach to cooling the hardware in a database server is to utilize some configuration of air-conditioning, heat sinks, and water evaporating systems to rid the system of excess heat. To create effective systems, one must consider various outside factors like local climate, fluid dynamics, and the actual architecture of the data center itself. Of course, there are other possible methods to cool down the hardware in the data centers. For instance, Google has recently started projects to cool down the database hardware using different means like pumping seawater into its cooling systems and using the DeepMind AI to manage the datacenter's cooling system. Microsoft is also working on a similar project to build the datacenters underwater.

One of the problems of limiting the energy consumption of the hardware itself is achieving satisfactory performance at the cost of power. Because database servers must have good performance in servicing multiple clients and working with a large amount of data, they typically store all the data in main memory using DRAM, which must be kept powered at all times or else the data is lost. In comparison with HDD, DRAM is much faster (about 10^5 to 10^6 times faster). For many applications, caching must be fast enough, so there must be a significant amount to DRAM to avoid cache misses, as mentioned before. One alternative to DRAM is flash memory. MIT researcher Arvind brought up a strategy known as BlueCache, which employs a combination of DRAM and flash memory, as flash memory, albeit slower than DRAM by a factor of about 10, is still a viable option for main memory caching. Also, the somewhat recently announced Micron’s 3D XPoint also serves as a faster alternative to HDD, serving as an SSD by using Phase Change Memory (PCM) material technology. This new type of SSD can allow for less DRAM, as it allows for less of a performance penalty for cache misses.

Improving CPU efficiency is an obvious method of decreasing power consumption while increasing performance. A specific case that Choi brings up in his article is the microchip design, Piton, developed by researchers at Princeton. In layman’s terms, the Piton architecture is a 25-core processing unit that specializes at performing the same task in parallel. In addition to threading in parallel, the chip is compactly designed to take advantage of locality to increase performance even further. Doing so can allow for the energy efficiency to increase by about 20%.

In general, to reduce the power consumption of data centers, the various factors that costs electricity (cooling, memory storage, and CPU performance) ought to be considered. All in all, there are numerous different implementations that one can decrease the power consumption of data centers.

Works Cited

Arvind, et al. “BlueCache: A Scalable Distributed Flash-based Key-value Store”. 2016. http://www.vldb.org/pvldb/vol10/p301-xu.pdf.

Barroso, Luiz André, et al. The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines. Morgan & Claypool Publishers, vol. 2, 2013. http://www.morganclaypool.com/doi/pdf/10.2200/S00516ED2V01Y201306CAC024.

Choi, Charles Q. “Reducing the Massive Energy Appetite of Data Centers.” *Inside Science*, 18 Sept. 2017, www.insidescience.org/news/reducing-massive-energy-appetite-data-centers.

Dean, Jeff. “Challenges in Building Large-Scale Information Retrieval Systems”. https://static.googleusercontent.com/media/research.google.com/en//people/jeff/WSDM09-keynote.pdf.

Delforge, Pierre. “New Study: America's Data Centers Consuming -- and Wasting – Growing Amounts of Energy.” *NRDC*, 15 Dec. 2016, www.nrdc.org/experts/pierre-delforge/new-study-americas-data-centers-consuming-and-wasting-growing-amounts-energy.

“Intel and Micron Increase 3D XPoint Manufacturing Capacity with IM Flash Fab Expansion.”*Intel Newsroom*, 13 Nov. 2017, newsroom.intel.com/news/intel-micron-expand-3d-xpoint-manufacturing-capacity-im-flash-fab-expansion/.

Mellor 21 Apr 2016 at 08:58 tweet_btn(), Chris. “Storage with the Speed of Memory? XPoint, XPoint, That's Our Plan.” *The Register® - Biting the Hand That Feeds IT*, 21 Apr. 2016, www.theregister.co.uk/2016/04/21/storage_approaches_memory_speed_with_xpoint_and_storageclass_memory/.

Patel ,D. Chandrakant and Amip J. ShahCost. “Model for Planning, Development and Operation of a Data Center". Hewlett-Packard Development Company, L.P. 2005. http://www.hpl.hp.com/techreports/2005/HPL-2005-107R1.pdf.

"U.S. Energy Information Administration - EIA - Independent Statistics and Analysis." *State Electricity Profiles - Energy Information Administration*, U.S. Energy Information Association, 17 Jan. 2017, www.eia.gov/electricity/state/.