aha products group

Networking and Information Technology R&D (NITRD)
Big Data Senior Steering Group (BDSSG)

Re: RFI Big Data R&D Strategic Plan

November 14, 2014

Big Data is one of the biggest challenges in this century. The benefits of leveraging new sources of data have created many opportunities and serve as the impetus to usher in the next era of computing. I commend the BDSSG for assuming leadership and the arduous task of guiding our nation to tackle this exciting new problem. Identifying a strategic direction is critical in directing research efforts, which will form the fallow ground for new technologies to be born. Additionally, I am appreciative for the opportunity to provide feedback to this strategic plan and hope that the ideas I present to you in this letter create value in your effort.

Before beginning my feedback, I would like introduce my company and myself. My name is Juan Deaton, PhD. I'm a research scientist for AHA Products Group, which has created data compression and coding products for over two decades. AHA has noticed the many new challenges with storage, processing, and transmitting Big Data. One of our major goals has always been to create new hardware that will increase performance and energy efficiency of complex algorithms.

As you may know, we have reached a critical point in large scale computing. Given the trends outlined in 2009, large scale computing is consuming approximately 3% of the US total power[1], more power than the entire paper industry.[2] In fact, power is the single largest expense in any data center, attributing to over 30% of operating costs. Big Data presents us with critical energy challenges that have real environmental consequences. Given the explosion of information generated by electronic and social media, today's technologies simply cannot keep up with the rate that data is increasing. About 80% of the world's data was created in last four years and by 2020 there is a predicted 40x increase in data production.[3] For these reasons, I would urge the BDSSG to include investments in foundational research and cyberinfrastructure for higher performing and energy efficient to address storage, processing, and transmission of data.

We've found, that an increase in performance and energy efficiency is often achieved simultaneously when approached with the correct technology. This especially true for data compression. With data compression, systems can can increase throughput and storage capacity by representing the information with fewer bits. However, while CPUs have formed the basis for modern computing and software, they are not as suitable for

---

[1] J. Koomey. Growth in data center electricity use 2005 to 2010. Oakland, CA: Analytics Press. August, 2011

[2] J. Glanz. Power, Pollution and the Internet. The New York Times, 2012.

[3] The Rapid Growth of Data, http://assets1.csc.com/insights/downloads/CSC_Infographic_Big_Data.pdf

1126 Alturas Drive ▪ Moscow, ID 83843-8331 ▪ tel: 208.892.5600 ▪ fax: 208.892.5601 ▪ www.aha.com

all computing tasks. In fact, many data centers do not use compression because CPUs can consume more energy and have slower performance when performing compression. However, when data compression is implemented onto specialized hardware, performance and energy efficiency accelerates by a factor of 10x over CPUs. When used in a data center, compression accelerators can increase storage and throughput by factors of three or more. With accelerator hardware, applied to applications such as compression, data centers can realize critical performance increases with little additional energy cost. We believe there are other opportunities in big data applications where accelerators can increase efficiency and performance of these systems.

While specialized hardware will enable short-term performance increases, new approaches to storage, processing, and transmission of data are also necessary as well. I've outlined some future technologies below in addition to my short-term recommendations.

Short Term Recommendations.

- Support data centers research evaluating and developing accelerators, such as data compression or encryption, which simultaneously increase efficiency and performance.
- Support data center research on storage, transmission, and/or computational efficiency using J/TB (Joule per Terabyte) as a metric.
- Support top 10 most energy efficient data center list/competition with recognition and financial motivation.
- Support research for developing hardware accelerators for big data applications and networked hardware accelerators for tera-scale operations.

Long Term Recommendations:

- Support research that will create new energy efficient technologies that can store, process, and transmit exabytes of data using the same amount of energy today's technologies.
- Support research in large-scale data compression, encryption, and erasure coding techniques that operate on terabytes of data.

Thank you for your time. It is my hope that these recommendations will create research to insure our energy security, a clean environment, and advancements in technology for the betterment of mankind.

Juan D. Deaton

Juan D. Deaton, PhD

1126 Alturas Drive ▪ Moscow, ID 83843-8331 ▪ tel: 208.892.5600 ▪ fax: 208.892.5601 ▪ www.aha.com