OpenPOWER ADG Webinar No. 3/2016

ExtraV: Boosting Out-of-Memory Graph Processing with a Coherent Accelerator

Speakers: Jinho Lee (Seoul National University/IBM)

Date: Tuesday, April 26, 2016, 2:00-3:00 pm GMT/UTC

Abstract

Large-scale graph processing is one of the most important applications in big data analytics. Social network services, biomedical problems such as DNA sequencing and analysis, computer networking and even VLSI design and automations are famous examples where the construction and structure of graphs is the key to understanding the complexity of the problem domain.

The out-of-memory graph processing, one of the techniques to process large graphs, usually uses a single server machine with a large storage device (i.e., hard disk) that stores the graph data. Recent research works demonstrated that the out-of-memory system can match the performance of clustered server systems with significantly smaller costs. This is therefore considered a promising alternative. However, existing out-of-memory approaches suffer from low disk bandwidth.

In this talk, we propose the ExtraV framework for out-of-memory graph processing. It is based on a novel concept of graph virtualization and cache coherent hardware acceleration of data compression, filtering and graph traversal. ExtraV consists of four main components: 1) host processor, 2) main memory, 3) AFU (Accelerator Function Unit) and 4) storage. The AFU, a hardware accelerator, executes the general graph processing functions as well as data compression and filtering while the program running on the host processor (called host program) manages the overall execution of graph algorithm and executes more application-specific tasks. Graph virtualization, i.e., a high-level programming model of graph processing, allows designers to focus on algorithm-specific functions. The accelerator, which realizes graph virtualization, gives host programs an illusion that the graph data reside on the main memory in a layout that fits with the memory access behavior of host programs even though the graph data are actually stored in a compressed form on the storage. In order to make best use of the limited disk bandwidth, we also propose a novel method of compressing the edge lists of graph data and we also propose an expand-and-filter technique. Both are implemented on the hardware accelerator. Our experiments with real-life graphs and algorithms running on a real system prototype offer up to more than 10 times speedup compared to state-of-the-art software-only implementations.

About the speaker

Jinho Lee received the B.S. degree in 2009, from Seoul National University, Seoul, Korea, where he received M.S. degree in 2011 and received Ph.D. degree in February 2016. He joined IBM Austin Research Lab as a postdoc researcher in mid-March.

His research interests include interconnection network, memory system, and domain-specific accelerators.

ExtraV: Boosting Out-of-Memory Graph Processing with a Coherent Accelerator

Tuesday, April 26, 2016

10:00 am | Eastern Daylight Time (New York, GMT-04:00) | 1 hr

Join WebEx meeting

Meeting number: 631 155 395

Meeting password: mZpNrPi3

Join by phone

1-866-469-3239 Call-in toll-free number (US/Canada)

1-650-429-3300 Call-in toll number (US/Canada)

Access code: 631 155 395

Global call-in numbers | Toll-free calling restrictions

Add this meeting to your calendar. (Cannot add from mobile devices.)