

NATIONAL UNIVERSITY OF SINGAPORE

School of Computing

C S S E M I N A R

Title: **Efficient Cloud-Based Systems for Big Data and Analytics**

Speaker: Dr. Peter Hofstee
Distinguished Research Staff Member
IBM Austin Research Laboratory

Date/Time: 15 October 2015, Thursday, 02:00 PM to 03:30 PM

Venue: Executive Classroom, COM2-04-02

Chaired by: Dr Mitra, Tulika, Professor, School of Computing
(tulika@comp.nus.edu.sg)

Abstract:

In this talk we explore the current state, and likely evolution, of systems to manage and analyze the vast amounts of digitized data we collect from the real world. We begin by taking a look at a prototypical Hadoop-based storage-centric Big Data system, and we discuss some ways in which such a system can be improved by leveraging heterogeneous compute and the changing ratios between HDD-based storage and networking costs. We make some observations about the similarities and differences between the typical analytics infrastructure and how it is likely to evolve and computational models for simulation-based HPC and illustrate some of the weaknesses of the Hadoop computing paradigm. Next we look at in-memory Big Data (Redis and Spark) and explore how such systems can also be made more cost-effective by leveraging heterogeneous elements in combination with high-IOP non-volatile memory. We explore the "Netezza" data warehouse appliance, and we show how its fundamental architecture of near-data compute can likely be extended to include analytics of unstructured data and data organized as graphs. Throughout the talk we use technologies from the OpenPOWER consortium to make our points concrete, and we also use gene sequencing as a concrete example of a Big Data application that can benefit from the proposed improvements. Besides sharing my perspective on systems for Big Data and Analytics I hope to convince my audience that in spite of the concerns about semiconductor technology scaling, there are many system enhancements coming our way that have significant potential benefit, but that will also require very active involvement of the software community to efficiently leverage.

Biodata:

H. Peter Hofstee (Ph.D. California Inst. of Technology, 1995) is a distinguished research

staff member at the IBM Austin Research Laboratory, USA, and a part-time professor in Big Data Systems at Delft University of Technology, Netherlands. Peter is best known for his contributions to heterogeneous computer architecture as the chief architect of the Synergistic Processor Elements in the Cell Broadband Engine processor, used in the Sony Playstation3 and the first supercomputer to reach sustained Petaflop operation. After returning to IBM research in 2011 he has focused on optimizing the system roadmap for big data, analytics, and cloud, including the use of accelerated compute. His early research work on coherently attached reconfigurable acceleration on Power 7 paved the way for the new coherent attach processor interface on POWER 8. Peter is an IBM master inventor with more than 100 issued patents and a member of the IBM Academy of technology.