

NATIONAL UNIVERSITY OF SINGAPORE

School of Computing

PH.D DEFENCE - PUBLIC SEMINAR

Title: **Continuous Optimization of Distributed Stream Programs**

Speaker: Mr Sumanaruban Rajadurai

Date/Time: 15 September 2016, Thursday, 10:00 AM to 11:30 AM

Venue: MR3, COM2-02-26

Supervisor : Dr Wong Weng Fai, Associate Professor, School of Computing

Dr Peh Li Shuian, Professor, School of Computing

Dr Saman Amarasinghe, Professor, Massachusetts Institute of Technology

Abstract:

Stream programs are an important class of applications in computer science, spanning several domains such as financial market, multimedia, and radio astronomy. Due to the recent advancements in collecting and storing data in massive scale, stream programs need to distribute and scale over dozens of nodes to attain high-performance, necessitating specialized system support. As the modern high performance hardware environments are highly dynamic, the system must also provide a lifetime support to the programs to ensure desired quality-of-service.

However, building a highly optimizing domain specific compiler for stream programming comes with several inherent challenges. Specifically, it has become less viable for heuristics based traditional compilers to generate optimal code for each and every hardware platform. Further, maintaining desired quality-of-service during dynamic changes needs advanced features.

In this thesis work, we have designed and implemented a distributed compiler and runtime system, StreamJIT, for high-performance stream processing. StreamJIT can automatically distribute and scale programs over cluster nodes per program's needs. Further, StreamJIT continuously optimizes the programs via various advanced features such as cluster-wide dynamic recompilation, downtime-free live reconfiguration, and online autotuning. The cluster-wide dynamic recompilation makes all ahead-of-time domain specific optimizations available to a running program. The downtime-free live reconfiguration recompiles and redistributes program instances across cluster nodes on-the-fly. The online autotuning, unlike the previous offline autotuning systems, continuously optimizes the code in the background. Series of experiments show StreamJIT is high performance, scalable, and resilient to dynamic changes.