

# NATIONAL UNIVERSITY OF SINGAPORE

School of Computing

## PH.D DEFENCE - PUBLIC SEMINAR

**Title:**           **Big Data Processing with Peer-to-Peer Architectures**

Speaker:       Mr Goh Wei Xiang

Date/Time:    9 September 2014, Tuesday, 10:00 AM to 11:30 AM

Venue:         MR1, COM1-03-19

Supervisor :   Dr Tan Kian Lee, Professor, School of Computing

### Abstract:

Recent develop

ments have brought about the introduction of, what some may classify as, disruptive technologies into the peripheral of both researchers and developers alike; we frequently hear the mention of the Big Data paradigm, or the deployment with cloud computing, or the NoSQL movement, or the MapReduce framework. While some may have their reservations on these new concepts, their continual widespread adoption in the industry undoubtedly indicates previously unsatisfied needs for certain systemic providence from the solutions of yesteryear. Three such desirable qualities of a system architecture can be identified: massive horizontal scalability, elastic resource consumption and robust distributed processing.

Currently, the predominant architecture adopted for modern data processing system is that of the master/workers architecture; this is said to be for the simplicity of the system design. However, it is perhaps profitable to investigate more elaborated alternatives, especially if systemic qualities may thus be enhanced. Extrapolating from the desirables, it appears that structured peer-to-peer (P2P) overlays present as a good match to the said conditions. In this seminar, two of the three identified dimensions of desirable systemic qualities are explored for the feasibility of adopting a structured P2P overlay in the context of modern data processing.

On horizontal scalability, work has been done to develop a generalized data processing framework, much like the MapReduce framework except that the programming model and the system architecture are completely decentralized. The Katana framework builds on the algebraic structure exhibit by many structured P2P overlays to materialize its programming model, which encompasses the expressiveness of the MapReduce programming model. Experimental results indicate that the augmented expressiveness, coupled with the decentralization of control, provides performance improvement in execution over widely scaled clusters.

In terms of robust processing, research has been conducted to investigate the incorporation

of the decentralized fault-tolerance of structured P2P overlays into modern data processing system. In particular, the Katana framework is extended to incorporate robust system-wide operations. Experimental studies indicate that the overhead incurred by the extended Katana framework, called hardened Katana framework, is comparable to, if not lesser than, that of the MapReduce framework.