Contribution ID: **147** Type: **Oral Presentation**

Caches all the way down: Infrastructure for Data Science

Wednesday, March 8, 2017 9:00 AM (45 minutes)

The rise of big data science has created new demands for modern computer systems. While floating performance has driven computer architecture and system design for the past few decades, there is renewed interest in the speed at which data can be ingested and processed. Early exemplars such as Gordon, the NSF funded system at the San Diego Supercomputing Centre, shifted the focus from pure floating point performance to memory and IO rates. At the University of Queensland we have continued this trend with the design of FlashLite, a parallel cluster equiped with large amounts of main memory, Flash disk, and a distributed shared memory system (ScaleMP's vSMP). This allows applications to place data "close" to the processor, enhancing processing speeds. Further, we have built a geographically distributed multi-tier hierarchical data fabric called MeDiCI, which provides an abstraction very large data stores cross the metropolitan area. MeDiCI leverages industry solutions such as IBM's Spectrum Scale and SGI's DMF platforms.

Caching underpins both FlashLite and MeDiCI. In this talk I will describe the design decisions and illustrate some early application studies that benefit from the approach.

Presenter: Prof. ABRAMSON, David (University of Queensland)

Session Classification: Keynote Session II