

June 14th, 2019

Caches All The Way Down: Infrastructure For Data Science

David Abramson

University of Queensland, Australia

Abstract

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 equipped 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 of very large data stores across the metropolitan area. MeDiCI leverages industry solutions such as IBM's Spectrum Scale and HPE'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.

Short bio



Professor David Abramson, Director, Research Computing Centre, has been involved in computer architecture and high performance computing research since 1979.

He has held appointments at [Griffith University](#), [CSIRO](#), [RMIT](#) and [Monash University](#). Prior to joining UQ, he was the Director of the Monash e-Education Centre, Science Director of the Monash e-Research Centre, and

a Professor of Computer Science in the Faculty of Information Technology at Monash.

From 2007 to 2011 he was an Australian Research Council Professorial Fellow.

David has expertise in High Performance Computing, distributed and parallel computing, computer architecture and software engineering.

He has produced in excess of 200 research publications, and some of his work has also been integrated in commercial products. One of these, Nimrod, has been used widely in research and academia globally, and is also available as a commercial product, called EnFuzion, from [Axceleon](#).

His world-leading work in parallel debugging is sold and marketed by [Cray Inc](#), one of the world's leading supercomputing vendors, as a product called cddb. David is a Fellow of the Association for Computing Machinery ([ACM](#)), the Institute of Electrical and Electronic Engineers ([IEEE](#)), the Australian Academy of Technology and Engineering ([ATSE](#)), and the Australian Computer Society ([ACS](#)). He is currently a visiting Professor in the Oxford e-Research Centre at the University of Oxford.