

THETA

The Higher Education Technology Agenda

Clusters to Clouds: Pedagogy and Research Training Environment

Educating the next generation of software engineers is essential with the increased move to an Internet-based society. The need to support big data and data analytics are challenging many of the typical scenarios and paradigms associated with software engineering. In the digital age, data is often messy, distributed and growing exponentially. In this context there are swathes of technologies that are shaping the landscape for dealing with these phenomenon. Cluster and high performance computing has been a core approach for processing data, but more recently Cloud computing has gained increasing prominence as the answer. To tackle this wave of approaches, the Melbourne eResearch Group at the University of Melbourne have focused on developing training materials that expose software engineers to latest developments in this space. This paper covers the pedagogy of the course and describes the way in which it utilizes national cloud resources made available through the National eResearch Collaboration Tools and Resources (NeCTAR – <http://www.nectar.org.au>) project and its Research Cloud program, and the Research Data Storage Infrastructure (VicNode – <http://www.vicnode.org.au>) storage resources. Examples of the solutions developed by the students are illustrated to demonstrate the practical experiences in developing Cloud based solutions that focus especially on 'big data' challenges. In particular case studies exploring real-time processing of Twitter data and associated data analytics through use of MapReduce and ElasticSearch algorithms are shown along with use of noSQL technologies such as CouchDB.

Richard Sinnott
University of Melbourne

SHARE THIS:



Loading...

 Follow