SORS: Data Centric Debugging: Scaling to Infinity and Beyond?

Objectives

Abstract: Debugging software is a critical activity, but there are very few software tools that address the problem well. The most widely available tools are very low level, and have hardly changed in years. These traditional tools allow a programmer to control the execution of each process (or thread) and to examine and manipulate the state of every thread. However, modern scientific codes have thousand of independent threads, and they manipulate enormous data structures.

Some years ago we pioneered data centric debugging. In data centric debugging a user reasons about an application's state, regardless of how many threads are involved, and how they are mapped onto the machine. Our most mature implementation of this is embodied in Cray's CCDB, which allows a programmer to debug a new version of a code against a reference version. However, a more general form of data centric debugging allows a user to assert statistical tests on data structures, and these can be used to detect anomalies as they arise. Any wonder that debuggers are not widely used - traditional tools simply don't scale to meet the needs of modern supercomputing. As we move to the exa-scale, this can only get worse.

Some years ago we pioneered data centric debugging. In data centric debugging a user reasons about an application's state, regardless of how many threads are involved, and how they are mapped onto the machine. Our most mature implementation of this is embodied in Cray's CCDB, which allows a programmer to debug a new version of a code against a reference version. However, a more general form of data centric debugging allows a user to assert statistical tests on data structures, and these can be used to detect anomalies as they arise.

Importantly, data centric debugging can scale with both the problem size and the machine because we have parallelised the debugging operations themselves. Data centric debugging also addresses another issue that arises in exa-scale - namely that computations will be no longer be bit-wise repeatable - a long held tenet of scientific computing. Data centric thinking is statistical, and thus outliers are detected rather than small changes in numeric values.

In this talk I will introduce data centric debugging and discuss various implementation issues.

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 ccdb.

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.

Speakers

Professor David Abramson
Director, Research Computing Centre
Barcelona Supercomputing Center - Centro Nacional de Supercomputación

---

Source URL (retrieved on *16 Abr 2024 - 11:26*): https://www.bsc.es/es/research-and-development/research-seminars/sors-data-centric-debugging-scaling-infinity-and-beyond