

Keynote Talk

Algorithms and Application for Grids and Clouds

Geoffrey C. Fox

School of Informatics and Computing and Digital Science Center
Indiana University
Bloomington IN
gcf@indiana.,edu

Abstract

We discuss the impact of clouds and grid technology on scientific computing using examples from a variety of fields -- especially the life sciences. We cover the impact of the growing importance of data analysis and note that it is more suitable for these modern architectures than the large simulations (particle dynamics and partial differential equation solution) that are mainstream use of large scale "massively parallel" supercomputers. The importance of grids is seen in the support of distributed data collection and archiving while clouds are and will replace grids for the large scale analysis of the data.

We discuss the structure of algorithms (and the associated applications) that will run on current clouds and use either the basic "on-demand" computing paradigm or higher level frameworks based on MapReduce and its extensions. Looking at performance of MPI (mainstay of scientific computing) and MapReduce both theoretically and experimentally shows that current MapReduce implementations run well on algorithms that are a "Map" followed by a "Reduce" but perform poorly on algorithms that iterate over many such phases. Several important algorithms including parallel linear algebra falls into latter class. One can define MapReduce extensions to accommodate iterative map and reduce but these have less fault tolerance than basic MapReduce. We discuss clustering, dimension reduction and sequence assembly and annotation as example algorithms.

Categories & Subject Descriptors: D.1.1.3,G.3, J.3

General Terms: Algorithms, Performance

Keywords: MPI, Grids, Clouds, Data Deluge, MapReduce, Life Sciences, Dimension Reduction, Clustering, BLAST

Bio

Fox received a Ph.D. in Theoretical Physics from Cambridge University and is now professor of Informatics and Computing, and Physics at Indiana University where he is director of the Digital Science Center and Associate Dean for Research and Graduate Studies at the School of Informatics and Computing. He previously held positions at Caltech, Syracuse University and Florida State University. He has supervised the PhD of 61 students and published over 600 papers in physics and computer science. He currently works in applying computer science to Bioinformatics, Defense, Earthquake and Ice-sheet Science, Particle Physics and Chemical Informatics. He is principal investigator of FutureGrid – a new TeraGrid facility to enable development of new approaches to computing. He is involved in several projects to enhance the capabilities of Minority Serving Institutions. His website is <http://www.infomall.org>.