

Data Grid and Gridflow Management Systems

Arun Jagatheesan and Reagan Moore
San Diego Supercomputer Center
University of California, San Diego
La Jolla, CA 92093
{arun, moore}@sdsc.edu

Abstract

The “Grid” is an emerging infrastructure providing coordinated and consistent services to access distributed, heterogeneous computational and information storage resources amongst autonomous organizations. Data grids are being built across the world as the next generation inter-organizational data management systems for coordinated sharing of data as well as storage resources. A data grid is a location independent logical namespace consisting persistent global identifiers for digital entities, storage resources and users in an autonomous inter-organizational collaboration. Data Grid Management Systems (DGMS) provide services on the data grid’s logical name space for management and organization of digital entities and resources. These data-intensive environments involve long run processes, which could be aggregated as Grid Workflows (Gridflows). Gridflows are executed as peer-to-peer workflows in the data grid infrastructure by the Gridflow Management Systems.

The tutorial will introduce the grid, provide use-cases in large projects, examine requirements, introduce some possible solutions for managing data in the grid and using a services-oriented infrastructure. It will take real-world examples from current systems, in particular the Storage Resource Broker and the SDSC Matrix project.

About the Presenters

Arunswaran Jagatheesan (“Arun”) is an Adjunct Assistant Researcher (OPS faculty member) at the University of Florida, and a Visiting Scholar at the San Diego Supercomputer Center (SDSC) at University of California, San Diego. His research interests include Data Grid Management, Peer-to-peer Computing, and Workflow Management Systems. He is the founder and technical lead of the SDSC Matrix Project on Gridflow Management Systems. He is a co-chair of the Grid File System Working Group at the Global Grid Forum, and is involved in research and development of multiple datagrid projects at the San Diego Supercomputer Center.

Dr. Reagan Moore is Co-Program Director for Data and Knowledge Systems at the San Diego Supercomputer Center. He coordinates research efforts on digital libraries, data grids, and persistent archives for projects with NSF, NASA, DOE, NARA, NHPRC, and the Library of Congress. Moore has a Ph.D. in plasma physics from the University of California, San Diego, (1978) and a B.S. in physics from the California Institute of Technology (1967). Moore has worked at the San Diego Supercomputer Center since 1986, as manager of the Cray Time Sharing System, and then as manager of all production software services. Moore currently is co-PI for SDSC participation on 13 research grants ranging from the NSF National Virtual Observatory, to the NSF Southern California Earthquake Center, to the DOE Particle Physics Data Grid, and the NARA Prototype Persistent Archive.

References

- [1] Moore, R.W., Jagatheesan, A., Rajasekar, A., Wan, M. and Schroeder, W., “Data Grid Management Systems,” Proceedings of the 21st IEEE/NASA Conference on Mass Storage Systems and Technologies, April 13-16, 2004, College Park, Maryland, USA
- [2] SDSC Matrix Project, <http://www.npaci.edu/DICE/SRB/matrix/>
- [3] The Global Grid Forum, <http://www.ggf.org/>