22

ADDITIONAL READING

In this chapter I list useful reference material on grid computing in general, including data management in the grid, and what has brought us to this point in grid computing and the direction(s) in which it is headed. The research material is broken down by category ranging from research and standards bodies that are rich in information (some of which is specifically referenced here) and technology topics. I encourage the reader to regularly track the activities of the standards bodies such as The Global Grid Forum, IEEE, and W3C and follow the publications of the "father of grid computing," Ian Foster.

One of the key objectives of a data grid is to unite the locality of where work is done and where the data naturally reside. Terms used throughout this book include the minimization of data movement and data affinity. It is imperative for anyone involved in data management for grid computing to understand data affinity, including its benefits, and how to best achieve it. The papers referenced here describe in detail the fundamentals of data affinity, and others describe advanced methods of data distribution policy and data synchronization policy that will be helpful in achieving data affinity.

Happy reading!

USEFUL INFORMATION SOURCES

- IEEE Distributed Systems Online (http://dsonline.computer.org)
- John Narghton, *A Brief History of the Future*, Overlook Press, May 2000

Distributed Data Management for Grid Computing, by Michael Di Stefano

WHITE PAPERS

Grid Computing

- Ian Foster, Carl Kesselman, and Steven Tuecke, *The Anatomy of the Grid Enabling Scalable Virtual Organizations* (available online at http://www.globus.org/research/papers/anatomy.pdf)
- Michael Di Stefano and Steve Yalovitser, "Grid Computing with a Data Grid Plane," September 27, 2002
- Ian Foster, Carl Kesselman, and Steven Tuecke, "The Anatomy of the Grid, Enabling Scalable Virtual Organizations," *Int. J. Supercomput. Appl.* (2001) (available online at http://www.globus.org/research/papers/anatomy.pdf)
- Ian Foster, Jens Vöckler, Michael Wilde, and Yong Zhao, "The Virtual Data Grid: A New Model and Architecture for Data-Intensive Collaboration," *Proceedings of the 2003 CIDR Conference* (available online at (http://www.griphyn.org/chimera/papers/CIDR.VDG.crc.submitted.pdf)
- Reagan W. Moore (San Diego Supercomputer Center), Scott Studham (Pacific Northwest National Laboratory), Arcot Rajasekar (San Diego Supercomputer Center), Chip Watson (Jefferson National Laboratory), Heinz Stockinger, and Peter Kunszt (CERN), "Data Grid Implementations," February 19, 2002 (available online at http://www.ppdg.net/docs/WhitePapers/Capabilities-grids.v12.pdf)
- Ann Chervenak, Ian Foster, Carl Kesselman, Charles Salisbury, and Steven Tuecke, "The Data Grid: Towards an Architecture for the Distributed Management and Analysis of Large Scientic Datasets" (available online at http://www.globus.org/documentation/incoming/JNCApaper.pdf)

GridFTP

- The Globus Project, "GridFTP Universal Data Transfer for the Grid," September 5, 2000 (copyright 2000, The University of Chicago and The University of Southern California) (available online at http://www.globus.org/datagrid/deliverables/C2WPdraft3.pdf)
- Bill Allcock, Lee Liming, and Steven Tuecke (ANL), and Ann Chervenak (USC/ISI), "GridFTP: A Data Transfer Protocol for the Grid, Grid Forum Data Working Group on GridFTP" (available online at http://www.sdsc.edu/GridForum/RemoteData/Papers/gridftp_intro_gf5.pdf)

Distributed File Systems

- *Distributed File System: A Logical View of Physical Storage*, Microsoft
- Peter J. Braam (School of Computer Science, Carnegie Mellon University), *The Coda Distributed File System* (available online at http://www.coda.cs.cmu.edu/)

- M. Satyanarayanan, "Coda: A Highly Available File System for a Distributed Workstation Environment" (available online at http://www-2.cs.cmu.edu/afs/cs/project/coda/Web/docdir/ieeepcs95.pdf)
- See http://www.eecs.harvard.edu/~vino/web/push.cache/node7.html

STANDARDS BODIES

Globus—Data Grid

- See http://www.globus.org/datagrid/
- GridFTP Universal Data Transfer Protocol for the Grid: The University of Chicago and The University of Southern California, September 5, 2000 (available online at http://216.239.51.104/search?q = cache:b0sOC7xSLh8J:www.globus.org/datagrid/deliverables/C2WPdraft3.pdf + GridFTP&hl = en&ie= UTF-8)

Global Grid Forum

- See http://www.ggf.org/L_About/about.htm
- *GridFTP*: http://www.gridforum.org/6_DATA/gridftp.htm

W3C

- Definition of Web Services: http://www.w3.org/TR/2003/WD-ws-arch-20030808/

PUBLIC AND UNIVERSITY GRID EFFORTS

- NASA; *Information Power Grid* (available online at http://www.ipg.nasa.gov/ipgusers/globus/6-globus.html)
- The DataGrid Project: "DataGrid is a project funded by European Union. The objective is to build the next generation computing infrastructure providing intensive computation and analysis of shared large-scale databases, from hundreds of TeraBytes to PetaBytes, across widely distributed scientific communities" (project report available online at http://eu-datagrid.web.cern.ch/eu-datagrid/)
- *In-Vigo*, University of Florida (available online at http://invigo.acis.ufl.edu/docs/aboutInVigo.html)
- *Virtuoso: Resource Management and Prediction for Distributed Computing Using Virtual Machines* (available online at http://www.cs.northwestern.edu/~plab/Virtuoso)
- *OceanStore*: http://oceanstore.cs.berkeley.edu/info/overview.html

SCIENTIFIC RESEARCH USE OF GRID COMPUTING

- *Grid Embedded Optimization and Design Search for Engineering*: http://www.geodise.org/
- *NEESgrid* (http://www.neesgrid.org/index.php): *Earthquake Engineering*
- *FusionGRID* (http://www.fusiongrid.org/index.html)
- Ayon Basumallik, Seung-Jai Min, and Rudolf Eigenmann, "Towards OpenMP Execution on Software Distributed Shared Memory Systems," School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN (available online at http://www.ece.purdue.edu/ParaMount)
- Lee Liming of the George E. Brown, Jr., "Network for Earthquake Engineering Simulation (NEES). The MOST Experiment: Earthquake Engineering on the Grid," presentation at The Application Research Group of the Global Grid Forum, meeting on Case Studies on Grid Applications in Munich, Germany, March 2004 (available online at http://www.zib.de/ggf/apps/index.html)
- Kate Keahey at the Argonne National Laboratory, "The National Fusion Collaboratory Project: Applying Grid Technology for Magnetic Fusion Research," presentation at The Application Research Group of the Global Grid Forum, meeting on Case Studies on Grid Applications in Munich, Germany, March 2004 (available online at http://www.zib.de/ggf/apps/index.html)
- Professor Simon Cox of Geodise (www.geodise.org), "Geodise: Grid Enabled Design Optimisation and Design Search, Applications and Testbeds Working Group Workshop," presentation at The Application Research Group of the Global Grid Forum, meeting on Case Studies on Grid Applications in Munich, Germany, March 2004 (available online at http://www.zib.de/ggf/apps/index.html)
- Alex Rodriguez, Dinanath Sulakhe, Elizabeth Marland, Natalia Maltsev, Ian Foster, Michael Wilde, and Veronika Nefedova, "Grid Enabled Server for High-throughput Analysis of Genomes," presentation at The Application Research Group of the Global Grid Forum, meeting on Case Studies on Grid Applications in Munich, Germany, March 2004 (available online at http://www.zib.de/ggf/apps/index.html)
- Dr. F. M. Brochu at the University of Cambridge (UK), "Running MadGraph on the LHC Computing Grid (LCG)," presentation at The Application Research Group of the Global Grid Forum, meeting on Case Studies on Grid Applications in Munich, Germany, March 2004 (available online at http://www.zib.de/ggf/apps/index.html)

WEB SERVICES

- Gunjan Samtani and Dimple Sadhwani, "Web Services and Straight Through Processing, Web Services in the Financial Industry," June 26, 2002 (available online at http://www.webservicesarchitect.com/content/articles/samtani06.asp)

- Service-Oriented Architecture Explained by Sayed Hashimi 08/18/2003 (available online at http://www.ondotnet.com/pub/a/dotnet/2003/08/18/soa_explained.html)
- Hao He, What is Service-Oriented Architecture? September 30, 2003 (available online at http://webservices.xml.com/pub/a/ws/2003/09/30/soa.html)
- Greg Goth, "Web Services Easing toward the Mainstream" (available online at http://dsonline.computer.org/0310/f/d10newp.htm)
- Michael Stevens, "Service-Oriented Architecture Introduction, Parts 1 and 2" (available online at http://www.developer.com/java/ent/article.php/10933_1010451_2)
- "The Benefits of a Service-Oriented Architecture" (available online at http://www.developer.com/tech/article.php/1041191)
- Greg Goth, "Web Services Easing toward the Mainstream," IEEE Distributed Systems Online (http://dsonline.computer.org/0310/f/d10newp.htm)
- Larry Peterson, Tom Anderson, David Culler, and Timothy Roscoe, "A Blueprint for Introducing Disruptive Technology into the Internet," *Proceedings of the 1st ACM Workshop on Hot Topics in Networking* (HotNets), October 2002.

DISTRIBUTED COMPUTING

- Jim Gray, *Distributed Computing Economics*, Microsoft Research, San Francisco, CA, March 2003
- D. Gelernter and A. J. Bernstein, "Distributed Communication via Global Buffer," *Proceedings of the ACM Principles of Distributed Computing Conference*, 1982
- D. Gelernter, "Generative Communication in Linda," *TOPLAS* **7**(1), 80–112 (1985)
- N. Carriero and D. Gelernter, "Linda in Context," *Commun. ACM* **32**(4) (April 1989)

COMPUTE UTILITY

- Vadim Kotov, *On Virtual Data Centers and Their Operating Environments*, Computer Systems and Technology Laboratory, Hewlett-Packard Laboratories Palo Alto, HPL-2001-44, March 8, 2001
- Integrasoft, L.L.C. (www.integrasoftware.com), Platform Computing, Inc. (www.platform.com), Corosoft, Inc. (www.corosoft.com), "Presentation: The Virtual Data Center"
- "On demand business: The new agenda for value creation" (www.ibm.com)
- The Global Grid Forum (http://www.gridforum.org/)
- Duke University Department of Computer Science, *COD*, *Cluster on Demand* (available online at http://issg.cs.duke.edu/cod/)

SERVICE-ORIENTED ARCHITECTURES

- Hao He, "What is Service-Oriented Architecture?" O'Reilly webservices.xml.com, September 30, 2003 (available online at http://webservices.xml.com/pub/a/ws/2003/09/30/soa.html)
- John Fontana, "Resurrecting the Distributed APP Model," *Network World* (September 29, 2003) (available online at http://www.nwfusion.com/buzz/2003/0929soa.html)

DATA AFFINITY

- Research possibilities to pursue in peer-to-peer networking (see http://www.praxagora.com/andyo/professional/p2p_research.html#affinity)
- Large-scale distributed systems (LSDS) (see www.inria.fr/rapportsactivite/RA2003/grand-large2003/LSDS.html)
- Artur Andrzejak, Sven Graupner,Vadim Kotov, and Holger Trinks, *Algorithms for Self-Organization and Adaptive Service Placement in Dynamic Distributed Systems*, Internet Systems and Storage Laboratory, HP Laboratories, Palo Alto, CA, HPL-2002-259, September 17, 2002* (available online at http://www.zib.de/andrzejak/my-papers/HPL-2002-259.pdf)
- Kavitha Ranganathan and Ian Foster, "Decoupling Computation and Data Scheduling in Distributed Data-Intensive Applications"
- Jeremy Stribling, Kirsten Hildrum, and John D. Kubiatowicz, *Optimizations for Locality-Aware Structured Peer-to-Peer Overlays*, Report UCB/CSD-03-1266, Computer Science Division (EECS), University of California, Berkeley, CA, August 2003
- Open MP (available online at http://www.openmp.org/and http://www.hpc.unimelb.edu.au/vpic/omp/perf/4.html)
- Active Harmony (http://www.dyninst.org/harmony/)