

INVITED PAPER FOR PLENARY KEYNOTE SPEECH AT WMSCI 2013:

Computational Dimensionalities of Global Supercomputing

Dr. Richard S. SEGALL
Arkansas State University
Department of Computer & Information Technology
College of Business
State University, AR 72467-0130 USA
E-mail: rsegall@astate.edu

ABSTRACT

This Invited Paper pertains to subject of my Plenary Keynote Speech at the 17th *World Multi-Conference on Systemics, Cybernetics and Informatics* (WMSCI 2013) held in Orlando, Florida on July 9-12, 2013. The title of my Plenary Keynote Speech was: “*Dimensionalities of Computation: from Global Supercomputing to Data, Text and Web Mining*” but this Invited Paper will focus only on the “Computational Dimensionalities of Global Supercomputing” and is based upon a summary of the contents of several individual articles that have been previously written with myself as lead author and published in [75], [76], [77], [78], [79], [80] and [11]. The topics of these of the Plenary Speech included Overview of Current Research in Global Supercomputing [75], Open-Source Software Tools for Data Mining Analysis of Genomic and Spatial Images using High Performance Computing [76], Data Mining Supercomputing with SASTTM JMP[®] Genomics ([77], [79], [80]), and Visualization by Supercomputing Data Mining [81].

OVERVIEW OF CURRENT RESEARCH IN GLOBAL SUPERCOMPUTING

The objective of this research is to present the concepts of supercomputing, explore its technologies and their applications, and develop a broad understanding of issues pertaining to the use of supercomputing in multidisciplinary fields. It aims to highlight the historical and technical background; architecture; programming systems; storage, visualization, analytics, state of practice in industry, universities and government; and numerous applications ranging from such areas as computational earth and atmospheric sciences to computational medicine.

This paper discusses the definition of supercomputers, current status of world supercomputers, an overview of applications of their use in research, contents of a forthcoming co-edited book on supercomputing, and references for additional research.

DEFINITION OF SUPERCOMPUTER

A supercomputer is a computer at the frontlines of current processing capacity and speed of calculations. Supercomputers were introduced in the 1960s. The supercomputers of the 1970s used only few processors, and in the 1990s machines with thousands of processors began to appear, and by the end of the 20th century supercomputers with massively parallel computing systems composed of tens of thousands of processors. Supercomputers of the 21st century can use over 100,000 processors including those with graphic capabilities.

Supercomputers are used today for highly-intensive calculation tasks for projects ranging from quantum physics, weather forecasting, molecular modeling, and physical simulations. Supercomputers can be used for simulations of airplanes in wind tunnels, detonations of nuclear weapons, for splitting electrons, and helping researchers study how drugs help against the swine flu virus.

Supercomputing can be in the form of grid computing in which the processing power of a large number of computers is distributed, or in the form of computer clusters in which a large number of processors are used in close proximity to each other.

CURRENT STATUS OF WORLD SUPERCOMPUTERS

Tianhe-2, a supercomputer developed by **China’s National University of Defense Technology**, is the **world’s new No. 1 system** with a performance of 33.86 petaflop/s on the Linpack benchmark, according to the 41st edition of the twice-yearly TOP500 [101] list of the world’s most powerful supercomputers. The list was announced June 17 during the opening session of the 2013 International Supercomputing Conference in Leipzig, Germany. Figures 1(a) and 1(b) list top 10 supercomputers in the world as of June 2013.

Rank	Site	System
1	National University of Defense Technology China	Tianhe-2 (MilkyWay-2) - TH-IVB-FEP Cluster, Intel Xeon E5-2692 12C 2.200GHz, TH Express-2, Intel Xeon Phi 31S1P NUDT
2	DOE/SC/Oak Ridge National Laboratory United States	Titan - Cray XK7, Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA K20x Cray Inc.
3	DOE/NNSA/LLNL United States	Sequoia - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom IBM
4	RIKEN Advanced Institute for Computational Science (AICS) Japan	K computer, SPARC64 VIIIx 2.0GHz, Tofu interconnect Fujitsu
5	DOE/SC/Argonne National Laboratory United States	Mira - BlueGene/Q, Power BQC 16C 1.60GHz, Custom IBM

Figure 1(a): Top Supercomputers in the World as of June 2013 – Number 1st to 5th

Source: <http://www.top500.org/lists/2013/06/>

The Cray XK7 "Titan", completed in 2012 and installed at the U.S. Department of Energy Oak Ridge National Labs, is currently the 2nd fastest supercomputer in the world at 17.59 petaflops, consuming 8209 kilowatts of power or 17.5 million billion mathematical calculations per second — on the Linpack benchmark tests to qualify for the list. The Cray machine reportedly has a peak capability of more than 27 petaflops.

Rank	Site	System
6	Texas Advanced Computing Center/Univ. of Texas United States	Stampede - PowerEdge C8220, Xeon E5-2680 8C 2.700GHz, Infiniband FDR, Intel Xeon Phi SE10P Dell
7	Forschungszentrum Juelich (FZJ) Germany	JUQUEEN - BlueGene/Q, Power BQC 16C 1.600GHz, Custom Interconnect IBM
8	DOE/NNSA/LLNL United States	Vulcan - BlueGene/Q, Power BQC 16C 1.600GHz, Custom Interconnect IBM
9	Leibniz Rechenzentrum Germany	SuperMUC - iDataPlex IX360M4, Xeon E5-2680 8C 2.70GHz, Infiniband FDR IBM
10	National Supercomputing Center in Tianjin China	Tianhe-1A - NUDT YH MPP, Xeon X5670 8C 2.93 GHz, NVIDIA 2050 NUDT

Figure 1(b): Top Supercomputers in the World as of June 2013: Numbered 6th to 10th

Source: <http://www.top500.org/lists/2013/06/>

Read more:

http://www.wptv.com/dpp/news/science_tech/titan-cray-xk7-supercomputer-oak-ridge-national-laboratory-has-worlds-fastest-computer#ixzz2E7QurrUT

The Indian government has stated that it has committed about \$940 million to develop what could become the world's fastest supercomputer by 2017, one that would have a performance of 1 exaflop, which is about 61 times faster than today's fastest computers (Reference: http://articles.timesofindia.indiatimes.com/2012-09-17/hardware/33901529_1_first-supercomputers-petaflop-fastest-supercomputer).

OBJECTIVE OF THE RESEARCH

The objective of this research is to present the concepts of supercomputing, explore its technologies and their applications, and develop a broad understanding of issues pertaining to the use of supercomputing in multidisciplinary fields. It aims to highlight the historical and technical background; architecture; programming systems; storage, visualization, analytics, state of practice in industry,

universities and government; and numerous applications ranging from such areas as computational earth and atmospheric sciences to computational medicine.

This invited paper not only presents an overview of global supercomputing but also of topics for future research in supercomputing. The study of supercomputing not only entails a study of the background and applications of supercomputing, but also the architecture and peripherals of clouds, clusters and grids, programming systems, storage, visualization and analytics, and state of practice of supercomputing around the world.

RESEARCH AREAS OF GLOBAL SUPERCOMPUTING

Research areas of global supercomputing can be focused on the structure, practice, and applications of supercomputing such as represented by the following research areas. Possible topics for research pertaining to supercomputing may include but are not limited to the following as discussed in Segall et al. ((2013)[75]):

Research Area 1: Background of Supercomputing

- History of supercomputing
- Fundamentals of supercomputing
- Supercomputing hardware
- Supercomputing organizations
- Supercomputing centers and clusters

Research Area 2: Supercomputing Architecture

- Processor architecture
- Chip multiprocessors
- Interconnect technologies
- Quality of service
- Network topologies
- Congestion management of supercomputers
- Power-efficient architectures
- Embedded and reconfigurable architectures
- Innovative hardware/software co-design
- Parallel and scalable system architectures
- Performance evaluation of systems architectures

Research Area 3: Clouds, Clusters, and Grids

- Cloud supercomputing
- Grid supercomputing
- Virtualizations and overlays
- Data management and scientific applications
- Storage clouds architectures
- Programming models
- Tools for computing on clouds and grids
- Quality of service
- Tools for Integration of Clouds, Clusters, and Grids
- Management and monitoring

- Security and identity management
- Scheduling, load balancing, and resource provisioning

Research Area 4: Programming Systems for Supercomputers

- Compiler analysis and optimization
- Libraries
- Parallel application frameworks
- Software engineering for parallel programming
- Solutions for parallel programming challenges
- Parallel programming for large-scale systems
- Performance analysis
- Productivity-oriented programming environments & studies

Research Area 5: Storage, Visualization, and Analytics for Supercomputers

- Databases for high performance computing
- Data mining for high performance computing
- Visualization for modeling and simulation
- Parallel file, storage, and archival systems
- Storage systems for data intensive computing
- Storage networks
- Visualization and image processing
- Input/output performance tuning

Research Area 6: State of Practice of Supercomputing

- Experiences of large-scale infrastructures and facilities
- Comparable benchmarks of actual machines
- Facilitation of “big data” associated with supercomputing
- Multi-center infrastructure and their management
- Bridging of cloud data centers and their management
- Procurement, technology investments, and acquisition of best practices
- Infrastructural policy issues and international experiences
- Supercomputing at universities
- Supercomputing in China, India, and around the world
- Supercomputing data mining

Research Area 7: Applications of Supercomputing

- Bioinformatics and computational biology
- Computational earth and atmospheric sciences
- Computational materials sciences and engineering
- Computational chemistry, fluid dynamics, physics, etc.
- Computational and data-enabled social science
- Computational design optimization for aerospace, manufacturing, and industrial applications
- Computational medicine and bioengineering

OPEN-SOURCE SOFTWARE TOOLS FOR DATA MINING ANALYSIS OF GENOMIC AND SPATIAL IMAGES USING HIGH PERFORMANCE COMPUTING (HPC)

High performance computing (HPC) utilizes supercomputers for data intensive computing of large-scale and distributed databases. This paper discusses an overview of open-source software tools that are available as free downloadable on the web for visualization of data at both the genomic and larger image formats.

SAN DIEGO SUPERCOMPUTING CENTER (SDSC)

San Diego Supercomputer Center (SDSC) houses the Triton Resource, “Dash”, and Protein Data Bank. The Triton Resource is currently under construction but when completed will provide as stated on SDSC [73] webpage supercomputing capabilities in three key areas: a large-scale disk storage facility called Data Oasis; a data analysis facility for petascale research; and a shared research cluster; all of which will be connected by a high-speed 10 gigabit network on the UC-San Diego campus.

Some of these open-source software at SDSC for data visualization are: (1.) NCL (NCAR Command Language) [57] that can also read in binary and ASCII files, (2.) ParaView [63] that is a multi-platform extensible application designed for visualizing large data sets, (3.) RasMol for molecular graphics visualization, (4.) Tecplot [89] which as a single tool allows one to analyze and explore complex data sets, arrange multiple XY, create 2-and 3-dimensional plots with animation, (5.) VAPOR (Visualization and Analysis Platform for Ocean, Atmosphere, and Solar Researchers) [95], (6.) VisIT[96] that allows interactive parallel visualization and graphical analysis for viewing scientific data, (7.) VISTA [97] that is a multi-threaded, platform independent, scalable and robust volume renderer, and (8.) VTK, which provides 3D computer graphics, image processing, and visualization. The reader is referred to Segall et al. ((2010)[76]) for comparisons for some of the open-source visualization software available at SDSC.

CENTER FOR HIGH PERFORMANCE COMPUTING (CHPC) AT THE UNIVERSITY OF UTAH

The Center for High Performance Computing (CHPC) at the University of Utah [21.] consists of three clusters: Updraft, Arches and Telluride. The Arches Clusters contains the TunnelArch that is the “data mining cluster” for jobs requiring large memory and consists of 48 nodes and 86 processors

with 4 gigabytes of memory node. This “data mining cluster” is smaller than the DelicateArch which consist of 256 nodes and 512 processors and is the “parallel cluster” for “highly parallel” parallel jobs requiring high speed interconnects.

Some of the open-source visualization software that is available at CHPC at the University of Utah are listed in Segall and Zhang ((2010)[76]). AutoDock [1] is a suite of automated docking tools. It is designed to predict how small molecules, such as substrates or drug candidates, bind to a receptor of known 3D structure. BLAST® (Basic Local Alignment Search Tool) [4] is a set of similarity open-source search programs designed to explore all of the available sequence databases regardless of whether the query is protein or DNA. The BLAST programs have been designed for speed, with a minimal sacrifice of sensitivity to distant sequence relationships. The scores assigned in a BLAST search have a well-defined statistical interpretation, making real matches easier to distinguish from random background hits. BLAST uses a heuristic algorithm that seeks local as opposed to global alignments and is therefore able to detect relationships among sequences that share only isolated regions of similarity.

Dalton [15] is a molecular electronic structure program with an extensive functionality for the calculation of molecular properties at the micro-levels of theory. The Dock [18] suite of programs addresses the docking of molecules to each other – (See Figure 2 below) identifying the low energy binding modes of a small molecule within the active site of a receptor molecule. This program can be used to predict binding modes, search databases of ligands for compounds that bind to a particular target receptor, and to examine binding orientations.

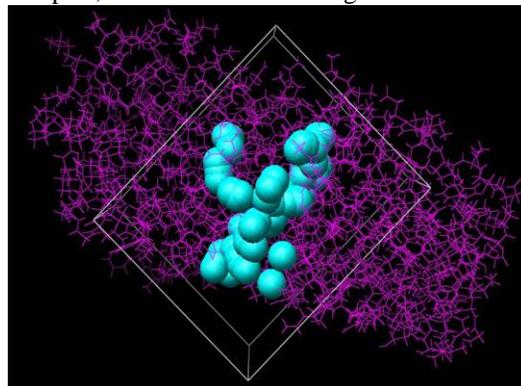


Figure 2: Molecular structure generated using DOCK 6 Software

[Source: http://dock.compbio.ucsf.edu/DOCK6/tutorials/grid_generation/generating_grid.html]

DATA MINING SUPERCOMPUTING WITH SAS™ JMP® GENOMICS

According to NCBI (2007)[56], the detection, treatment, and prediction of outcome for lung cancer patients increasingly depend on a molecular understanding of tumor development and sensitivity of lung cancer to therapeutic drugs.

NCI (2007)[58] states that the application of genomic technologies, such as microarray, is widely used to monitor global gene expression and has built up invaluable information and knowledge that is essential to the discovery of new insights into the mechanisms common to cancer cells, resulting in the identification of unique, identifiable signatures and specific characteristics. According to NCBI (2007)[56] it is likely that application of microarray may revolutionize many aspects of lung cancer being diagnosed, classified, and treated in the near future. NCBI (2007) used microarrays to detail the global gene expression patterns of lung cancer.

The overall design of NCBI (2007)[56] as used in this paper consisted of adjacent normal-tumor matched lung cancer samples that were selected at early and late stages for RNA extraction and hybridization on Affymetrix microarrays. A total of 66 samples were used for microarray analysis in NCBI (2007)[56], including pairwise samples from 27 patients, who underwent surgery for lung cancer at the Taipei Veterans General Hospital, two tissue mixtures from the Taichung Veterans General Hospital, two commercial human normal lung tissues, one immortalized, nontumorigenic human bronchial epithelial cell line, and 7 lung cancer cell lines.

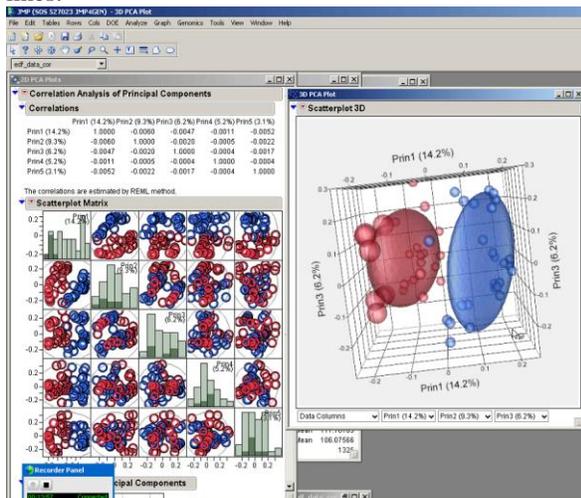


Figure 3: Correlation analysis of principle components for lung cancer data (Source: Segall et al.(2010)[79])

Figure 3 above is computer output using SAS™ JMP® Genomics software on a supercomputer. The reader is referred to Segall et al. ((2010)[79], 2010[80]) for other figures and results for data mining performed using SAS™ JMP® Genomics for Lung Cancer Data.

SAS™ JMP® Genomics has a window called “basic expression workflow” that is the process that runs a basic workflow for expression data used to select variables of interest. The data used for the lung cancer and its associated tumors consisted of over 22,000 rows representing genes and 54 columns representing samples.

BACKGROUND OF SUPERCOMPUTING DATA MINING

According to Wikipedia (2013a)[100], supercomputers or HPC (High Performance Computing) are used for highly calculation-intensive tasks such as problems involving quantum mechanical physics, weather forecasting, global warming, molecular modeling, physical simulations (such as for simulation of airplanes in wind tunnels and simulation of detonation of nuclear weapons).

Sanchez (1996)[71] cited the importance of data mining using supercomputers by stating “Data mining with these big, super-fast computers is a hot topic in business, medicine and research because data mining means creating new knowledge from vast quantities of information, just like searching for tiny bits of gold in a stream bed”.

According to Sanchez (1996)[71], The Children’s Hospital of Pennsylvania took MRI scans of a child’s brain in 17 seconds using supercomputing for that which otherwise normally would require 17 minutes assuming no movement of the patient. Researchers at the University of Pennsylvania received the Supercomputing ’95 Data Mining Research Award.

Other applications of supercomputers for data mining include that of Davies (2007)[16] using Internet distributed supercomputers, Seigle (2002)[81] for CIA/FBI, Mesrobian et al. (1995)[50] for real time data mining, King (1999)[40] for Texas Medicaid fraud case investigation, and Curry et al. (2007) for detecting changes in large data sets of payment card data.

Grossman (2006)[25] as Director of the National Center for Data Mining at the University of Illinois at Chicago cited in a report of funding acknowledgements several funded proposals by the National Science Foundation (NSF) such as a proposal entitled “Developing software tools and

network services for mining remote and distributed data over high performance networks” and another proposal entitled “Tera Mining: A test bed for distributed data mining over high performance SONET and Lambda Networks.” Other contributions to the development of data mining supercomputing include DMMGS06 (2006)[17] which conducted a workshop on data mining and management on the grid and supercomputers in Nottingham, UK, Grossman (2007)[26] who wrote a survey of high performance and distributed data mining, Sekijima (2007)[83] on application of HPC to analysis of disease related protein, Tekiner et al. (2007)[89] who co-edited a special issue on data mining applications on supercomputing and grid environments, and a entire book on large-scale parallel data mining by Zaki (2000)[107].

DATA VISUALIZATION USING Avizo®

Avizo® software is a powerful, multifaceted tool for visualizing, manipulating, and understanding scientific and industrial data. Wherever three-dimensional datasets need to be processed, Avizo® offers a comprehensive feature set within an intuitive workflow and easy-to-use graphical user interface. (VSG, 2009)[97]).

Some of the core features of Avizo® include advanced 3D visualization by surface and volume rendering, scientific visualization of flow data and processing very large datasets at interactive speed, and 3D data exploration and analysis by displaying single or multiple datasets in a single or multiple viewer window, and navigate freely or around or through these objects. Avizo® can also perform 3D reconstruction by employing innovative and robust algorithms from image processing and computational geometry to reconstruct high-resolution 3D images generated by CT or MRI scanners, 3D ultrasonic devices, or confocal microscopes (VSG, 2009)[97]).

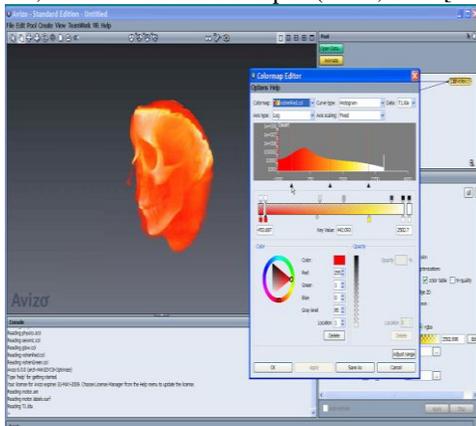


Figure 4: 3D visualization generated by Avizo® of a human skull with the color map editor window [Source: Segall et al. (2009)[78]]

The reader is referred to Segall et al. (2009)[78] for other screen shots of 3D visualizations generated by Avizo®.

CONCLUSIONS AND FUTURE DIRECTIONS OF RESEARCH

The conclusions of the research discussed in the plenary keynote address at WMSCI 2013 and overviewed in this invited paper indicate the value of supercomputing to analyze complex problems with very large scale databases of all types of data. An example of this is the Human Brain Project (www.humanbrainproject.eu) that was started in 2012 and is currently being conducted using supercomputers by over 200 researchers in over 10 European countries and other countries in the world with \$1.6 billion in funding. Supercomputers have increased the speed of calculations from months to seconds for huge data sets such as cancer genome analysis (Bowman (2012) [110]).

The future direction of this research is to investigate the other areas of supercomputing such as listed in the list of areas of proposed areas for the co-edited book on Global Supercomputing. This would also include the applications of other software for supercomputing data mining and to include such areas such as image data processing and other 3D data exploration and analysis. Also the list of over 100 References below indicate a starting point for additional research on global supercomputing and its applications.

REFERENCES

- [1.] Autodesk (2010), <http://autodesk.scripps.edu>
- [2.] Bagnall, A. (2004), Super-Computing Data Mining, Proposal to Engineering and Physical Sciences Research Council, <http://gow.epsrc.ac.uk/ViewGrant.aspx?GrantRef=GR/T18479/01>
- [3.] Barak, S. (2008), NASA shows off new star supercomputer, November 18, <http://www.theinquirer.net/inquirer/news/1049617/nasa-shows-star-supercomputer>
- [4.] Blast (2007), http://blast.ncbi.nlm.nih.gov/blast_overview.shtml
- [5.] Bryant, R. E. and Digney, J. (2007), Data-Intensive supercomputing: The case for DISC, PDL Packet, Fall 2007, Parallel Data Laboratory, Carnegie-Mellon University, <http://www.pdl.cmu.edu>
- [6.] Buchen, L. (2009), Record amount of supercomputer time means new science, Wired Science, April 30, <http://www.wired.com/wiredscience/2009/04/supercomputers-2/>

- [7.] Bull, L., (2004), Super-Computing Data Mining, Proposal to Engineering and Physical Sciences Research Council, <http://gow.epsrc.ac.uk/ViewGrant.aspx?GrantRef=GR/T18455/01>
- [8.] Business Wire (2005), "NASA Takes Google on Journey into Space." Business Wire. Retrieved May 16, 2009 from HighBeam Research: <http://www.highbeam.com/doc/1G1-136789727.html>
- [9.] Business Wire (2009), "Pervasive Software Launches Pervasive DataRush Academic Alliance Program." Business Wire. Retrieved May 16, 2009 from HighBeam Research: <http://www.highbeam.com/doc/1G1-198531829.html>
- [10.] Center for High Performance Computing (2010), Software – Alphabetical, <http://www.chpc.utah.edu/docs/manuals/software> .
- [11.] Committee on the Future of Supercomputing, National Research Council (2003), *The Future of Supercomputing: An Interim Report*, ISBN-13: 978-0-309-09016-2, <http://www.nap.edu/catalog/10784.html>
- [12.] Computer Weekly News (2009), "Researchers from University of Manitoba describe findings in supercomputing.(Report)." NewsRX. 2009. Retrieved May 16, 2009 from HighBeam Research: <http://www.highbeam.com/doc/1G1-195175275.html>
- [13.] Cook, Jeffrey S.(2011), "Supercomputers and Supercomputing," Chapter 17 of Zhang, Qingyu and Segall, Richard S. and Cao, Mei, Visual Analytics and Interactive Technologies: Data, Text and Web Mining Applications, IGI Global, Hershey, PA, ISBN 13: 978-1-60-960102-7, pp. 282-294.
- [14.] Curry, C., Grossman, R., Locke, D., Vejcek, S., and Bugajski, J. (2007), Detecting changes in large data sets of payment card data: A case study, *KDD'07*, August 12-15, San Jose, CA.
- [15.] Dalton (2010), <http://www.kjemi.uio.no/software/dalton/dalton.html>
- [16.] Davies, A. (2007), Identification of spurious results generated via data mining using an Internet distributed supercomputer grant, Duquesne University Donahue School of Business, <http://www.business.duq.edu/Research/details.asp?id=83>
- [17.] MMGS06 (2006), Data Mining and Management on the Grid and Supercomputers, Mini-Workshop Call for Participation, Fifth UK e-Science All Hands Meeting (AHM 2006) 18-21 September 2006, East Midlands Conference Centre, Nottingham, UK, http://www.allhands.org.uk/2006/programme/workshops/data_mining.html
- [18.] Dock (2010), <http://www.chpc.utah.edu/docs/manuals/software/dock.html>
- [19.] Dubrow, A. (2009), Unpacking Protein Function, HPC Wire, April 21, <http://www.hpcwire.com/topic/visualization/Unpacking-Protein-Function.html>
- [20.] Economic Times (New Delhi, India), "India Inc. shows signs of IBM Blue Gene," McClatchy-Tribune Information Services. (2006). Retrieved May 16, 2009 from HighBeam Research: <http://www.highbeam.com/doc/1G1-140729995.html>
- [21.] Gara, A. and Kozloski, J.(2007), BlueGene supercomputing for biological systems: Past, present and future, *HPC Workshop: "Hardware and software for large-scale biological computing in the next decade"*, December 11-14, Okinawa, Japan, <http://www.irp.opist.jp/hpc-workshop/slides.html>
- [22.] Gimp (2010), <http://www.gimp.org/>
- [23.] Glover, T. (2006), "Bill Gates brings supercomputing to the masses." Sunday Business (London). July 30, Retrieved May 16, 2009 from HighBeam Research: <http://www.highbeam.com/doc/1G1-148878095.html>
- [24.] Grace (2010), <http://plasma-gate.weizmann.ac.il/Grace/>
- [25.] Gromacs (2010), <http://www.gromacs.org>
- [26.] Grossman, R. (2006), National Center for Data Mining: Funding Acknowledgements, University of Illinois at Chicago, <http://www.ncdm.uic.edu/grantAcknowledgment.html>
- [27.] Grossman, R. (2007), Data grids, data clouds and data webs: a survey of high performance and distributed data mining, *HPC Workshop: "Hardware and software*

- for large-scale biological computing in the next decade", December 11-14, Okinawa, Japan, <http://www.irp.oist.jp/hpc-workshop/slides.html>
- [28.] Grossman, R. L. and Guo, Y. (2002), Parallel methods for scaling data mining algorithms to large data sets, *Handbook on data mining and knowledge discovery*, Oxford University Press (J M Zytkow, Editor), pp. 433-442.
- [29.] Grossman, R. L. and Hollebeek, R. (2002), The national scalable cluster project: three lessons about high performance data mining and data intensive computing, *Handbook of massive data sets*, Kluwer Academic Publishers Norwell, MA, pp. 853-874.
- [30.] GSE7670, National Center for Biotechnology Information, <http://www.ncbi.nlm.nih.gov/geo/>
- [31.] Haskins, W. (2008), The secret lives of supercomputers, Part 1, TechNewsWorld, August 7, <http://www.technewsworld.com/story/64076.html>
- [32.] Haskins, W. (2008), The secret lives of supercomputers, Part 2, TechNewsWorld, August 8, <http://www.technewsworld.com/story/64082.html>
- [33.] Holland, W. (2008), OU's second supercomputer also ranked in top 100 globally, *The Daily*, November 28.
- [34.] Hosch, W. L. (2009), Supercomputers, Encyclopedia Britannica, <http://www.bookrags.com/eb/supercomputer-eb/>
- [35.] Image Spatial Data Analysis Group (2009), National Center for Supercomputing, University of Illinois at Urbana-Champaign, <http://isda.ncsa.illinois.edu>
- [36.] ImageMagick (2010), <http://www.imagemagick.org/script/index.php>
- [37.] Jones, A. (2008), Is supercomputing just about performance?, *ZDNet Asia*, December 15, <http://www.zdnetasia.com/insight/hardware/0.39043471.62048062.00.htm>
- [38.] Jones, A., (2009), Are supercomputers just better liars?, *ZDNet Asia*, April 22, <http://www.zdnetasia.com/insight/hardware/0.39043471.62052736.00.htm>
- [39.] Karin, S. and Bruch, M. (2006), Supercomputers, *Macmillan Science Library: Computer Sciences*, <http://www.bookrags.com/research/supercomputers-csci-01/>
- [40.] Kennedy, K. (2003), Simplifying the development of grid applications, NSF Press Release, NSF PR 03-103, September 17, <http://www.nsf.gov/od/lpa/news/03/pr03103.htm>
- [41.] King, E. (1999), "Supercomputing Goes Mainstream?(Texas Medicaid program targets fraud)(Technology Information)." *Enterprise Systems Journal*, Retrieved May 16, 2009 from HighBeam Research: <http://www.highbeam.com/doc/1G1-53578812.html>
- [42.] Kudal, A. (2007), Super computers in pharmacy, *Pharmainfo.net*, v. 5, n.1, <http://www.pharmainfo.net/reviews/super-computers-pharmacy>
- [43.] Kushner, D. (2009), Game developers see promise in cloud computing, but some are skeptical, <http://www.spectrum.ieee.org/print/8007>, March 3.
- [44.] Lai, E. (2009), Intel to ship Larrabee graphics chip in early 2010, *Computerworld*, April 15, <http://www.computerworld.com/action/article.do?command=viewArticleBasic&articleId=9131651>
- [45.] M2 Presswire (2001), NSF grant extends support for interconnecting national research networks, M2 Communications Ltd. 2001. Retrieved May 16, 2009 from HighBeam Research: <http://www.highbeam.com/doc/1G1-68874263.html>
- [46.] M2 Presswire (2009), San Diego Supercomputer Center chooses HPC SAN solution from Sun Microsystems; Solution meets the demands of SDSC's data-intensive national research center by enabling seamless data sharing across multiple platforms at over 2GB/sec., M2 Communications Ltd. 2001. *HighBeam Research*. 16 May. 2009 <<http://www.highbeam.com>>.
- [47.] Malon, D., Lifka, D., May E., Grossman, R., Qin, X., and Xu, W. (1995), Parallel Query Processing for Event Store Data, *Proceedings of the Conference on Computing in High Energy Physics 1994*, edited by S. C. Loken, pp. 239-240.

- [48.] Mao, Y., Gu, Y., Chen, J., Grossman, R.L. (2006), FastPara: A high-level declarative data-parallel programming framework on clusters, Proceedings of Parallel and Distributed Computing and Systems (PDCS 2006), November 13-15, Dallas, TX.
- [49.] Merrion, Paul (1997). "Power up: research center bytes off bigger mission, more bucks, corporate partners. (National Center for Supercomputing Applications)." Crain's Chicago Business. Crain Communications, Inc. 1997. Retrieved May 16, 2009 from HighBeam Research: <http://www.highbeam.com/doc/1G1-19598061.html>
- [50.] Mesa_3D (2010), <http://www.mesa3d.org/>
- [51.] Mesrobian, E. , Muntz, R., Shek,E., Mechoso,, C. R., Farrara, J.D., Spahr, J.A., Stolorz, P.(1995), Real time data mining, management, and visualization of GCM output, IEEE Computer Society, v.81, http://dml.cs.ucla.edu/~shek/publications/sc_94.ps.gz
- [52.] Molden (2010), <http://www.cmbi.ru.nl/molden/molden.html>
- [53.] Munger, Frank (2012), Titan Cray XK7 supercomputer: Oak Ridge National Laboratory has world's fastest computer, November 12, 2012, knoxnews.com, http://www.wptv.com/dpp/news/science_tech/titan-cray-xk7-supercomputer-oak-ridge-national-laboratory-has-worlds-fastest-computer
- [54.] National Center for Biotechnology Information, <http://www.ncbi.nlm.nih.gov/geo/>
- [55.] NCBI (2006), "Gene Expression Profiling in Breast Cancer: Understanding the Molecular Basis of Histologic Grade to Improve Prognosis, Gene Expression Omnibus (GEO), Series GSE2990,
- [56.] NCBI (2007), "Expression data from Lung Cancer", Gene Expression Omnibus (GEO), Series
- [57.] NCL (2010), <http://www.ncl.ucar.edu/>
- [58.] Network World Middle East (2009), "Consortium tackles cloud computing standards." Al Bawaba (Middle East) Ltd. 2009. Retrieved May 16, 2009 from HighBeam Research: <http://www.highbeam.com/doc/1G1-191788297.html>
- [59.] NewsWire (2002), February 12, <http://www.govexec.com/dailyfed/0202/021202gsn1.htm>
- [60.] Northwest Alliance for Computational Science and Engineering (NACSE) (2009), High Performance Computing Sites (U.S.), <http://www.csi.uoregon.edu/nacse/hpc/>
- [61.] Novakovic, N. (2009), Supercomputing in the Financial Crisis, HPC Wire, May 11, <http://www.hpcwire.com/topic/visualization/Supercomputing-in-the-Financial-Crisis.html>
- [62.] NWChem (2010), <http://www.emsl.pnl.gov/capabilities/computing/nwchem/>
- [63.] Paraview (2010), <http://www.paraview.org/>
- [64.] Pettipher, M. (2004), Super-Computing Data Mining, Proposal to Engineering and Physical Sciences Research Council, <http://gow.epsrc.ac.uk/ViewGrant.aspx?GrantRef=GR/T18462/01>
- [65.] PR Newswire (2001), "Compaq Unveils the AlphaServer ES45, Industry's Most Powerful Mid-Range Server," PR Newswire Association LLC , Retrieved May 16, 2009 from HighBeam Research: <http://www.highbeam.com/doc/1G1-79162470.html>
- [66.] PR Newswire (2005), "Ohio Supercomputer Center (OSC)-Springfield Deploys Ammasso 1100 RDMA Over Ethernet Adapters for Distributed Computing Over Wide Area Network (WAN) Connections." PR Newswire Association LLC. 2005. Retrieved May 16, 2009 from HighBeam Research: <http://www.highbeam.com/doc/1G1-134959564.html>
- [67.] PR Newswire (2009), "Purdue University Explores SGI RASC Technology to Lower Costs in Scientific Production Environments." PR Newswire Association LLC. 2006. Retrieved May 16, 2009 from HighBeam Research:

- <http://www.highbeam.com/doc/1G1-154447464.html>
- [68.] Ramsey, D. (2009), Supercomputing to help UC San Diego researchers visualize cultural patterns, <http://ucsdnews.ucsd.edu/newsrel/general/01-09CulturalPatterns.asp>
- [69.] Rees, CA, Demeter, J., Matese, JC, Botstein, D., and Sherlock, G. (2004), GeneXplorer: an interactive web application for microarray data visualization and analysis, BMC Bioinformatics, v. 5, n.141, www.biomedcentral.com/1471-2105/5/141
- [70.] Rich, D. (2008), Semantic Supercomputer reaps competitive advantage from patient data, HPC Wire, September 3, <http://www.hpcwire.com/topic/systems/SemanticSupercomputingReapsCompetitiveAdvantagefromPatentData.html?viewAll=y>
- [71.] Sanchez, E. (1996), Speedier: Penn researchers to link supercomputers to community problems, The Compass, v. 43, n. 4, p. 14, September 17, <http://www.upenn.edu/pennnews/features/1996/091796/research>
- [72.] SAS (2009), JMP® Genomics 4.0 Product Brief, http://www.jmp.com/software/genomics/pdf/103112_jmpg4_prodbrief.pdf
- [73.] SDSC (2006), Cyberinfrastructure, San Diego Supercomputer Center, <http://www.sdsc.edu/about/Infrastructure.html>
- [74.] SDSC (2010), Visualization Software Installed at SDSC, San Diego Supercomputer Center, <http://www.sdsc.edu/us/visservices/software/installed.html>
- [75.] Segall, Richard S.; Zhang, Qingyu and Cook, Jeffrey S.(2013), "Overview of Current Research in Global Supercomputing", *Proceedings of Forty-Fourth Meeting of Southwest Decision Sciences Institute (SWDSI)*, Albuquerque, NM, March 12-16, 2013.
- [76.] Segall, Richard S. and Zhang, Qingyu (2010), "Open-Source Software Tools for Data Mining Analysis of Genomic and Spatial Images using High Performance Computing", *Proceedings of 5th INFORMS Workshop on Data Mining and Health Informatics*, Austin, TX, November 6, 2010.
- [77.] Segall, Richard S., Zhang, Qingyu and Pierce, Ryan M.(2010), "Data Mining Supercomputing with SAS JMP® Genomics: Research-in-Progress", *Proceedings of 2010 Conference on Applied Research in Information Technology*, sponsored by Acxiom Laboratory of Applied Research (ALAR), University of Central Arkansas (UCA), Conway, AR, April 9, 2010.
- [78.] Segall, Richard S., Zhang, Qingyu and Pierce, Ryan M.(2009), "Visualization by Supercomputing Data Mining", *Proceedings of the 4th INFORMS Workshop on Data Mining and System Informatics*, San Diego, CA, October 10, 2009.
- [79.] Segall, Richard S., Zhang, Qingyu, and Pierce, Ryan (2010), "Data Mining Supercomputing with SAS JMP® Genomics", *Proceedings of 14th World Multi-Conference on Systemics, Cybernetics and Informatics: WMSCI 2010*, Orlando, FL, June 29-July 2, 2010
- [80.] Segall, Richard S., Zhang, Qingyu, and Pierce, Ryan (2010), "Data Mining Supercomputing with SAS JMP® Genomics", *Journal of Systemics, Cybernetics and Informatics (JSCI)*, Vol. 9, No. 1, 2011, pp.28-33.
- [81.] Segall, RS, Zhang, Q., and Pierce, RM (2009), Visualization by supercomputing data mining, *Proceedings of the 4th INFORMS Workshop on Data Mining and System Informatics*, San Diego, CA, October 10, 2009.
- [82.] Seigle, G. (2002), CIA, FBI developing intelligence supercomputer, *Global Security*
- [83.] Sekijima, M. (2007), Application of HPC to the analysis of disease related protein and the design of novel proteins, *HPC Workshop: "Hardware and software for large-scale biological computing in the next decade"*, December 11-14, Okinawa, Japan, <http://www.irp.oist.jp/hpc-workshop/slides.html>
- [84.] Sekijima, M. (2007), Application of HPC to the analysis of disease related protein and the design of novel proteins, *HPC Workshop: "Hardware and software for large-scale biological computing in the next decade"*, December 11-14, Okinawa, Japan, <http://www.irp.oist.jp/hpc-workshop/slides.html>

- [85.] Sharma, M. and M. Dhaniwala (2009), Asia's fastest supercomputer props up Indian animation, *HPC Wire*, May 4 <http://www.bollywoodhungama.com/features/2009/05/04/5125/>
- [86.] Sotiriou C, Wirapati P, Loi S, Harris A et al. (2006), Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J National Cancer Institute*, February 15; volume 98, number 4, pp. 262-72. PMID: 16478745
- [87.] Strait Times (2009), Supercomputers to fight flu, http://www.straitstimes.com/Breaking%2BNews/Tech%2Band%2BScience/Story/STIStory_386257.html
- [88.] Su LJ, Chang CW, Wu YC, Chen KC et al. (2007), "Selection of DDX5 as a novel internal control for QRT-PCR from microarray data using a block bootstrap resampling scheme", *BMC Genomics* 2007 June 1; volume 8, number 140. PMID: 17540040
- [89.] Tecplot (2010), <http://www.tecplot.com/>
- [90.] Tekiner, F., Mettipher, M., Bull, L., and Bagnall, A. (2007), *Mediterranean Journal of Computers and Networks, SPECIAL ISSUE on Data Mining Applications on Supercomputing and Grid Environments* Volume 3, No. 2, April 2007.
- [91.] The University of Utah (2010), The Center for High Performance Computing, www.chpc.utah.edu
- [92.] Thoennes, M.S., and Weems, C.C. (2003), "Exploration of the performance of a data mining application via hardware based monitoring", *The Journal of Supercomputing*, volume 26, pp. 25-42.
- [93.] U.S. Department of Defense (2009), DoD Supercomputing Resource Center (DSRC), U.S. Army Research Laboratory (ARL), <http://www.arl.hpc.mil/Applications/index.php>
- [94.] US Fed News Service, Including US State News (2006), "Renaissance computing institute, University of North Carolina researchers collaborate to develop better bioinformatics tools for understanding melanoma", HT Media Ltd. 2006. Retrieved May 16, 2009 from HighBeam Research: <http://www.highbeam.com/doc/1P3-1154370261.html>
- [95.] VAPOR (2010), <http://www.vapor.ucar.edu/>
- [96.] VisIT (2010), <https://wci.llnl.gov/codes/visit/>
- [97.] VISTA (2010), <http://www.sdsc.edu/us/visservices/software/vista/>
- [98.] VSG Visualization Sciences Group (2009), Avizo The 3D visualization software for scientific and industrial data, http://www.vsg3d.com/vsg_prod_avizo_overview.php
- [99.] Watanabe, T. (2007), The Japanese next generation supercomputer project, *HPC Workshop: "Hardware and software for large-scale biological computing in the next decade"*, December 11-14, Okinawa, Japan, <http://www.irp.oist.jp/hpc-workshop/slides.html>
- [100.] Wikipedia (2013a), Supercomputers, Retrieved July 24, 2013 from <http://en.wikipedia.org/wiki/Supercomputer>
- [101.] Wikipedia (2013b), TOP500, July 24, 2013, http://en.wikipedia.org/wiki/Top_500_Supercomputers
- [102.] Wikipedia (2013c), GMOD, July 25, 2013 <http://en.wikipedia.org/wiki/GMOD>
- [103.] Wilkins-Diehr, N. and Mirman, I. (2008), "On-Demand supercomputing for emergencies", *Design Engineering Technology News Magazine*, February 1, <http://www.deskeng.com/articles/aaagtk.htm>
- [104.] *Wireless News* (2005), From Jets to Potato Chips, Q-Bits to Hobbits, the Experts Take It On at SC05, Retrieved May 16, 2009 from HighBeam Research: <http://www.highbeam.com/doc/1P1-114593315.html>
- [105.] Workshop on Data Mining and System Informatics (2009), San Diego, CA, October 10, 2009.
- [106.] *Xinhua News Agency* (2008), "China Exclusive: China builds supercomputer rivaling world's 7th fastest." *Xinhua News Agency*, Retrieved May 16, 2009 from HighBeam Research: <http://www.highbeam.com/doc/1P2-16788194.html>
- [107.] Zaki, M.J. (2000), *Large-Scale Parallel Data Mining*, Lecture Notes in Artificial Intelligence, Springer Press, ISBN 3-540-67194-3.
- [108.] Zaki, M.J., Ogihara, M., Parthasarathy, S., and Li, W. (1996), "Parallel data mining for association rules on shared-memory multi-processors", *Proceedings of the 1996*

ACM/IEEE Conference on Supercomputing,
Pittsburgh, PA, Article Number 43, ISBN 0-
89791-854-1.

- [109.] Zhang, Qingyu and Richard S. Segall
(2008), “Visual Analytics of Data-Intensive
Web, Text and Data Mining with
Multiprocessor Supercomputing”, funded
proposal for \$15,000 by Arkansas State
University College of Business Summer
Research Grant, Jonesboro, AR, 2008.
- [110.] Bowman, Dan, (2012), New
supercomputer speeds cancer genome
analysis to seconds, October 3,
[http://www.fiercehealthit.com/story/new-
supercomputer-speeds-cancer-genome-
analysis-seconds/2012-10-03](http://www.fiercehealthit.com/story/new-supercomputer-speeds-cancer-genome-analysis-seconds/2012-10-03)

Research and Applications in Global Supercomputing. IGI Global, 2015. 56-81. Web. Due to remarkable advances in computer technology, scientists are looking for better digital supercomputing tools to deal with the complexities of the datasets. Supercomputers are the fastest computers we know of. They are characterized by a very high computational speed and an immense number of processors (Chinta, 2013). They are usually seen in corporations, etc filling a big room. The speed of supercomputers are measured in FLOPS (Chinta, 2013). Simply put, Floating point operations means computations that involve very large decimal numbers, usually 300 digits in a single number. A computationally simple method of dimensionality reduction that does not introduce a significant distortion in the data set would thus be desirable. In random projection (RP), the original high-dimensional data is projected onto a lower-dimensional subspace using a random matrix whose columns have unit lengths. RP has been found to be a computationally efficient, yet sufficiently accurate method for dimensionality reduction of high-dimensional data sets. In this paper we give experimental results on using RP as a dimensionality reduction tool on high-dimensional image and text data sets. In both application areas, random projection is compared to well known dimensionality reduction methods. Using a large, shared memory supercomputer called Nautilus, Leetaru has analyzed the tone and geographic dimensions of a 30-year archive of global news to produce real-time forecasts of human behavior such as national conflicts and the movement of specific individuals. A range of advanced analysis techniques were used to produce a network 2.4 petabytes in size containing more than 10 billion people, places, things, and activities connected by over 100 trillion relationships—more data than any current computing system can handle. By leveraging advanced supercomputers like Nautilus, Leetaru is a