Vijay S Kumar

Dept. of Computer Science and Engineering
395 Dreese Labs, 2015 Neil Ave
Columbus, OH 43210
+1 614 284 6447
vijayskumar@bmi.osu.edu
http://www.cse.ohio-state.edu/~kumarvi

Research Interests

large-scale data management, parallel and distributed systems, scientific workflows, multidimensional data analysis, databases and query processing, knowledge representation

Education

2003–Present **Ph.D. candidate, Computer Science and Engineering**, *Ohio State University*, Columbus, OH.
`Major`: Software Systems, `Minors`: Computer Networks, Databases (GPA: *3.61/4.0*)
`Thesis`: Supporting Analysis and Querying in Scientific Big Data Applications
`Advisors`:Dr. Joel H. Saltz and Dr. P. Sadayappan
`Expected Graduation`: March 2010

1998–2003 **B.E.(with Honors), Computer Science**, *Birla Institute of Technology and Science*, Pilani, Rajasthan, India., (GPA: *9.14/10.0*).
`Thesis`: Simulating Distributed Load-Balancing Algorithms on the C-DAC PARAM-10000 Supercomputer
`Advisor`: Dr. Chittaranjan Hota

M.Sc.(with Honors), Chemistry, *Birla Institute of Technology and Science*, Pilani, Rajasthan, India (*under a dual-degree scheme*)

Professional Experience

Sep 08–Present **Graduate Research Associate**, *Ohio State University*, Columbus, OH.
Department of Computer Science and Engineering
`Advisors`: **Dr. P. Sadayappan** and **Prof. Joel H. Saltz**

Sep 03–Sep 08 **Graduate Research Associate**, *Ohio State University*, Columbus, OH.
Multiscale Computing Lab, Department of Biomedical Informatics
`Advisor`: **Dr. Joel H. Saltz**

Jul 07–Sep 07 **Student Research Scholar**, *Lawrence Livermore National Laboratory*, Livermore, CA.
Institute for Scientific Computing Research (ISCR)
`Supervisor`: **Dr. Ghaleb Abdulla**

Jan 03–Jun 03 **Student Intern**, *i2 Technologies, India Pvt. Ltd.*, Bangalore, India.
Quote Optimizer team
`Supervisor`: **Nayan Jadeja**

Journal Publications

IJHPCA'09 **HPC and Grid Computing for Integrative Biomedical Research**.
Tahsin Kurc, Shannon Hastings, Vijay S. Kumar, Stephen Langella, Ashish Sharma, Tony Pan, Scott Oster, David Erwin, Justin Permar, Sivaramakrishnan Narayanan, Yolanda Gil, Ewa Deelman, Mary Hall, Joel Saltz.
In *International Journal of High-Performance Computing Applications*, Special Issue on Clusters and Computational Grids for Scientific Computing, August 2009.

Computer'08 **Analysis and Semantic Querying in Large Biomedical Image Datasets**.
Vijay S. Kumar, Sivaramakrishnan Narayanan, Tahsin Kurc, Jun Kong, Metin Gurcan, Joel Saltz.
In *IEEE Computer*, Special Issue on Data-Intensive Computing, April 2008. (*Cover Feature*)

TITB'08 **Large-scale Biomedical Image Analysis in Grid Environments**.
Vijay S. Kumar, Benjamin Rutt, Tahsin Kurc, Ümit Çatalyurek, Tony Pan, Sunny Chow, Stephan Lamont, Maryann Martone, Joel Saltz.
In *IEEE Transactions on Information Technology in Biomedicine*, Special Issue on Bio-Grid, March 2008.

Refereed Conference/Workshop Publications

HPDC'09 **An Integrated Framework for Parameter-based Optimization of Scientific Workflows**.
Vijay S. Kumar, Tahsin Kurc, Gaurang Mehta, Karan Vahi, Varun Ratnakar, Jihie Kim, Ewa Deelman, Yolanda Gil, P. Sadayappan, Mary Hall, Joel Saltz.
In *ACM International Symposium on High Performance Distributed Computing* (HPDC), June 2009, Munich, Germany. (*Acceptance Rate: 19%, Invited to Special Issue on HPDC in the Cluster Computing Journal*)

IPDPS'09 **Architectural Implications for Spatial Object Association Algorithms**.
Vijay S. Kumar, Tahsin Kurc, Ghaleb Abdulla, Scott R. Kohn, Joel Saltz, Celeste Matarazzo.
In *IEEE International Parallel and Distributed Processing Symposium* (IPDPS), May 2009, Rome, Italy. (*Acceptance Rate: 23%*)

NGS'08 **Designing and Parameterizing a Workflow for Optimization: A Case Study in Biomedical Imaging**.
Vijay S. Kumar, Mary Hall, Jihie Kim, Yolanda Gil, Tahsin Kurc, Ewa Deelman, Joel Saltz.
In *NSF Workshop on Next Generation Software* (NGS), held in conjunction with IPDPS, April 2008, Miami, Florida.

Cluster'07 **Performance vs. Accuracy Trade-offs for Large-scale Image Analysis Applications**.
Vijay S. Kumar, Tahsin Kurc, Jun Kong, Ümit Çatalyurek, Metin Gurcan, Joel Saltz.
In *IEEE International Conference on Cluster Computing* (CLUSTER), September 2007, Austin, Texas. (*Acceptance Rate: 39%*)

NGS'07 **Intelligent Optimization of Parallel and Distributed Applications**.
Bhupesh Bansal, Umit Catalyurek, Jacqueline Chame, Chun Chen, Ewa Deelman, Yolanda Gil, Mary Hall, Vijay S. Kumar, Tahsin Kurc, Kristina Lerman, Aiichiro Nakano, Yoon-ju Lee Nelson, Joel Saltz, Ashish Sharma, Priya Vashishta.
In *NSF Workshop on Next Generation Software* (NGS), March 2007, Long Beach, California.

SC'06 **Large Image Correction and Warping in a Cluster Environment**.
Vijay S. Kumar, Benjamin Rutt, Tahsin Kurc, Ümit Çatalyurek, Sunny Chow, Stephan Lamont, Maryann Martone, Joel Saltz.
In *ACM/IEEE Conference on Supercomputing: International Conference on High Performance Computing, Networking, Storage and Analysis* (SC), November 2006, Tampa, Florida. (*Acceptance Rate: 23%*)

Cluster'05 **Distributed Out-of-Core Preprocessing of Very Large Microscopy Images for Efficient Querying**.
Benjamin Rutt, Vijay S. Kumar, Tony Pan, Tahsin Kurc, Ümit Çatalyurek, Yujun Wang, Joel Saltz.
In *IEEE International Conference on Cluster Computing* (CLUSTER), September 2005, Boston, Massachusetts. (*Acceptance Rate: 32%*)

ParCo'05 **Supporting Large Scale Medical and Scientific Datasets**.
Umit Catalyurek, Shannon Hastings, Kun Huang, Vijay S. Kumar, Tahsin Kurc, Stephen Langella, Sivaramakrishnan Narayanan, Scott Oster, Tony Pan, Benjamin Rutt, Xi Zhang, Joel Saltz.
In *International Conference on Parallel Computing: Current and Future Issues of High-End Computing* (ParCo), September 2005, Malaga, Spain.

HPC'05a **A Runtime Framework for Partial Replication and its Application for On-Demand Data Exploration**.
Sivaramakrishnan Narayanan, Ümit Çatalyurek, Tahsin Kurc, Vijay S. Kumar, Joel Saltz.
In *High Performance Computing Symposium* (HPC), part of the Spring Simulation Multiconference, April 2005, San Diego, California.

HPC'05b **Design and Implementation of a Data Server using a Peer-to-peer Storage System**.
Vijay S. Kumar, Tahsin Kurc, Ümit Çatalyurek, Joel Saltz.
In *High Performance Computing Symposium* (HPC), part of the Spring Simulation Multiconference, April 2005, San Diego, California.

Research Reports

Architectural Implications for Spatial Object Association Algorithms.

LLNL-Preprint'09 Vijay S. Kumar, Tahsin Kurc, Joel Saltz, Ghaleb Abdulla, Scott R. Kohn, Celeste Matarazzo.
Preprint LLNL-CONF-410283, Lawrence Livermore National Laboratory, Feb 2009.

LLNL-Preprint'08 **Evaluation of Different Database/Data Management Architectures for Spatial Object Association**.
Ghaleb Abdulla, Celeste Matarazzo, Vijay S. Kumar, Tahsin Kurc.
LLNL-TR-404171-DRAFT, Lawrence Livermore National Laboratory, May 2008.

Selected Projects

IMPACT **Improving Multidimensional Data Analysis through Performance Accuracy Trade-offs**.
An integrated framework designed to support parameter-based optimization of multidimensional data analysis workflows where users can trade accuracy of analysis for performance gains. Users express their high-level requirements semantically. The framework translates these semantic descriptions of queries and applications into appropriate system-level configurations (e.g., workflow and batch scheduling, dataflow streaming) of a distributed runtime environment, such that performance improvements transpire. (*Team of 12; system development involves a collaboration between Ohio State University, University of Southern California and University of Utah*)

OCVM **Out-of-Core Virtual Microscope**.
Capable of supporting complex analysis and querying of potentially terabyte-scale out-of-core image datasets in large-scale heterogeneous cluster environments. OCVM builds on DataCutter, which provides a *filter-stream* programming model. (*Team of 3; in collaboration with scientists from the National Center for Microscopy and Imaging Research, San Diego*)

PerfAcT **Performance Accuracy Trade-off module**.
An extension to OCVM to support application-level quality-of-service by trading accuracy of data analysis for performance in large-scale image analysis applications. (*Team of 2; in collaboration with imaging experts from the Ohio State University*)

xLSST **Fast Crossmatch for the LSST**.
A hybrid system that combines the useful features of a massively parallel system (Netezza Performance Server) and a network database system (MySQL Cluster) to provide improved responses to a class of real-time data analysis (spatial crossmatch) queries in a cluster environment. (*Team of 4, as part of the petabyte-scale Large Synoptic Survey Telescope (LSST) project; in collaboration with data management experts at the Lawrence Livermore National Laboratory*)

MUD-Pond **Multidimensional Data Server atop P2P storage**.
Designed and implemented a multidimensional-data server using Pond, a peer-to-peer storage system prototype developed at University of Berkeley. The server supports subsetting and filtering queries against large multidimensional scientific datasets.

Awards and Honors

* Research abstract selected for presentation at the *Doctoral Showcase* for Ph.D. students at ACM/IEEE Conference on Supercomputing (SC), November 2009. (One of 12 selected entries)
* *Outstanding Paper Award*, IEEE International Conference on Cluster Computing, 2005.
* Member, Upsilon Pi Epsilon Honor Society, Alpha Chapter of Ohio.
* Graduated with *distinction* from the Birla Institute of Technology and Science, 2003.

Professional Activities

* Student Member, IEEE
* External reviewer for the following conferences

  International Conference on High Performance Computing (HiPC 2009)

  International Conference on High Performance Computing and Communications (HPCC 2008)

International Conference on Distributed Computing Systems (ICDCS 2005)
IEEE/ACM International Workshop on Grid Computing (Grid 2005)

Technical Skills

Concepts	Component-based programming, Grid workflows, stream processing, Hadoop/MapReduce, application QoS
HPC	MPI
Languages	C, C++, Java, SQL, Unix shell scripts, Python, RDF, OWL, LISP
Build tools	make, Ant, SCons
Version control	Subversion, CVS
OS	Linux (Red Hat/CentOS, Ubuntu, Debian)
Databases	MySQL, MySQL Cluster, PostgreSQL

Strengths

* Experience in understanding and modeling real-world applications in various scientific domains including astronomy and bio-medical informatics
* Fostering close collaborations with scientists and actual users of large-scale applications, and gathering their high-level objectives
* Rapid development of system prototypes in response to user requirements
* Experience with integration of middleware and systems at various levels to provide end-to-end solutions for performance improvements

References

Available upon request.