Curriculum Vitae

Sangmi Lee Pallickara

Computer Science Department,
1873 Campus Delivery,
Fort Collins, CO 80523-1873
Email : sangmi@cs.colostate.edu
URL: http://www.cs.colostate.edu/~sangmi

1   Research Interests

My research interests are in the area of Big Data for the sciences; specifically, issues related to modeling, storage, analytics, integration, metadata, and visualization.

2   Education

Aug 2003	**Ph.D.** Computer Science, *Florida State University*, Tallahassee, FL
May 2000	**M.S.** Computer and Information Science, *Syracuse University*, Syracuse, NY
Feb 1993	**B.S.** Physics, *Sookmyung University*, Seoul, Korea

3   Professional Experience

July 17 --	Associate Professor, *Computer Science Department*, Colorado State University
Jan 13 –Jun 17	Assistant Professor, *Computer Science Department*, Colorado State University
Jan 10 – Dec 12	Research Scientist, *Computer Science Department*, Colorado State University
Aug 07 – Dec 09	*Research Technologies*, Indiana University
Sep 04 – Jul 07	Post Doctoral Fellow, *Department Of Computer Science*, Indiana University With Professor Beth Plale
Sep 03 – Aug 04	Post Doctoral Fellow, *Community Grids Lab*, Indiana University
May 00 – Aug 03	Research Assistant, *Department of Computer Science*, Florida State University
Sep 01 – Aug 04	Visiting Research Associate, *Community Grids Lab*, Indiana University
May 99 – Apr 00	Research Assistant, Northeast Parallel Architectures Center, *Department of Electrical Engineering & Computer Science*, Syracuse University

4   Awards

National Science Foundation CAREER Award, 2016.

Best Paper Award. *IEEE/ACM Symposium on Big Data Computing* [22% acceptance rate] with Jared Koontz (graduate advisee), and Matthew Malensek (graduate advisee). 2014.

Best Paper Award. *IEEE/ACM Conference on Utility and Cloud Computing* [27% acceptance rate] with

Matthew Malensek (graduate advisee) and Shrideep Pallickara. 2012.

5 Research Funding

My research is funded by the National Science Foundation, the Department of Homeland Security, the Environmental Defense Fund, Colorado School of Public Health, Google, Amazon and HP.

[R10] CAREER: A Framework for Ad Hoc Model Construction in Data Streaming Environments. $491,243 5/2016-5/2021. **Sangmi Lee Pallickara** (PI). Faculty Early Career Development (CAREER) Award. National Science Foundation. ACI-1553685.

[R9] Methane Mapping at Colorado State University: Phase-3, $77,553 10/2016 - 9/2017. **Sangmi Lee Pallickara** (Co-PI) with Joe von Fischer (PI), Environmental Defense Fund.

[R8] Integration and Analytics of Public Health Big Data for Identifying Risk Factors for Obesity. $19,850 7/2016-6/2017. **Sangmi Lee Pallickara** (Co-PI) with Kaigang Li (PI). Colorado School of Public Health.

[R7] Geo for Google Cloud Credits Beta Program. $11,000. Google. 2016. **Sangmi Lee Pallickara** (PI).

[R6] Assessing Epidemiological and Economic Impacts of Countermeasures and Vaccination Strategies in Disease Outbreaks at the National Scale. $2,046,914   9/2015–9/2018. **Sangmi Lee Pallickara** (Co-PI) with Shrideep Pallickara (PI), Jay Breidt, Neil Harvey, Mike Sanderson, and Dustin Pendell. Science and Technology Directorate. Department of Homeland Security.

[R5]   Methane Mapping at Colorado State University: Database Supporting Scalable Analysis & Visualization of Multidimensional, Time-Series Gas Leak Data. $68,246   2015-2016, **Sangmi Lee Pallickara** (PI) with Joe von Fischer (Co-PI), Environmental Defense Fund.

[R4] Coalition for Development and Implementation of Sensor Systems. $180,000   2014-2016, **Sangmi Lee Pallickara** (Co-PI) with David Dandy (PI), Ken Reardon, Chuck Henry, Melissa Reynolds, Anura Jayasumana, Tom Chen, Rick Lyons, Jennifer Mueller, Lori Peek, and John Volckens, The CSU Catalyst for Innovative Partnerships program.

[R3] *Big Data Analytics of Epidemic Outbreaks*. $1,176,860   9/2013-7/2017. **Sangmi Lee Pallickara** (Co-PI) with Shrideep Pallickara (PI), Neil Harvey (Co-PI), and Jay Breidt (Co-PI). LONG RANGE PROGRAM, Science and Technology Directorate. Department of Homeland Security.

[R2] *Fast and Accurate Approximations of Data Distributions in Voluminous Datasets*. $2,500   2013 **Sangmi Lee Pallickara** (PI), Undergraduate research project, Hewlett Packard Inc.

[R1] *Exploring Performance Hotspots for Query Evaluation Algorithms Over Voluminous Time-Series Data. Infrastructure Grant* (Access to 2,000 virtual machines, $7,500)   6/2013- 6/2015, **Sangmi Lee Pallickara** (PI), Amazon Web Services in Education Research Grant Award.   Amazon Inc.

Education Grants

[E5] Google Cloud Platform Education Award (for the graduate course, CS535 Big Data, $1,700), 9/2016 – 12/2016. Google Cloud Platform, **Sangmi Lee Pallickara** (PI).

[E4] AWS Gift (for the senior undergraduate course, CS435 Introduction to Big Data, $4,550), 1/2016 – 5/2016 Amazon Web Services in Education, Educator Award. **Sangmi Lee Pallickara** (PI).

[E3] AWS Gift (for the graduate course, CS535 Big Data, $1,950), 8/2015 – 12/2015 Amazon Web Services in Education, Educator Award. **Sangmi Lee Pallickara** (PI).

[E2] *AWS Faculty Educator Award* (for the senior undergraduate course, *CS480A2 Introduction to Big Data*, $7,000), 1/2015 – 5/2015 Amazon Web Services in Education, Educator Award. **Sangmi Lee Pallickara** (PI), Amazon Inc.

[E1]   *AWS Faculty Educator Award* (for the graduate course, *CS581 Big Data*, $3,500), 8/2014 – 12/2014 Amazon Web Services in Education, Educator Award. **Sangmi Lee Pallickara** (PI), Amazon Inc.

6   Publications

Listings with the symbol * identify my graduate advisees.

6.1   Journals

[J14]   Matthew Malensek*****, **Sangmi Lee Pallickara**, and Shrideep Pallickara. Hermes: Federating Fog and Cloud Nodes to Support Query Evaluations in Continuous Sensing Environments. *IEEE Cloud Computing*. Vol. 4(2) pp 54-62. 2017.

[J13]   Walid Budgaga*****, Matthew Malensek*****, **Sangmi Lee Pallickara**, and Shrideep Pallickara. A Framework for Scalable Real-Time Anomaly Detection over Voluminous, Geospatial Data Streams. Concurrency and Computation: Practice & Experience. Vol. 29(12) pp 1-16. John-Wiley. 2017

[J12]   Matthew Malensek*****, **Sangmi Lee Pallickara**, and Shrideep Pallickara. Fast, Ad Hoc Query Evaluations over Multidimensional Geospatial Datasets. *IEEE Transactions on Cloud Computing*. Vol. 5(1) pp 28-42 2017.

[J11]   Cameron Tolooee*****, Matthew Malensek*****, and **Sangmi Lee Pallickara** A Scalable Framework for Continuous Query Evaluations over Multidimensional, Scientific Datasets. *Concurrency and Computation: Practice and Experience*. 28(8): pp. 2546-2563. 2016.

[J10]   Matthew Malensek*****, **Sangmi Lee Pallickara**, and Shrideep Pallickara. Analytic Queries over Geospatial Time-Series Data using Distributed Hash Tables. *IEEE Transactions on Knowledge and Data Engineering*. Vol. 28(6): pp.1408-1422. 2016.

[J9]   Matthew Malensek*****, **Sangmi Lee Pallickara**, and Shrideep Pallickara. Autonomous Data Management and Federation to Support High-throughput Query Evaluations over Voluminous Datasets. *IEEE Cloud* Computing. Vol. 3 (3): pp. 40-49. 2016.

[J8]   Walid Budgaga*****, Matthew Malensek*****, **Sangmi Lee Pallickara**, Neil Harvey, Jay Breidt, and Shrideep Pallickara. Predictive Analytics Using Statistical, Learning, and Ensemble Methods to Support Real-Time Exploration of Discrete Event Simulations. *Future Generation Computer Systems*. Elsevier. Volume 56, March, pp. 360–374. 2016.

[J7]   Matthew Malensek*****, **Sangmi Lee Pallickara**, and Shrideep Pallickara. Minerva: Proactive Disk Scheduling for QoS in Multi-Tier, Multi-Tenant Cloud Environments. *IEEE Internet Computing*. Vol. 20 (3): pp. 19-27. 2016.

[J6]   Matthew Malensek*****, **Sangmi Lee Pallickara**, and Shrideep Pallickara. Geometry and Proximity Constrained Query Evaluations over Large Geospatial Datasets Using Distributed Hash Tables. *IEEE Computing in Science and Engineering (CiSE)*. Special Issue on Extreme Data. Vol. 16(4): pp. 53-60. 2014.

[J5]   Matthew Malensek*****, **Sangmi Lee Pallickara**, and Shrideep Pallickara. "Exploiting Geospatial and Chronological Characteristics in Data Streams to Enable Efficient Storage and Retrievals," *Future Generation Computer Systems*. Vol. 29(4): pp. 1049-1061. Elsevier. 2013.

[J4]   **Sangmi Lee Pallickara**, Shrideep Pallickara, and Milija Zupanski, "Enabling Efficient Data Search and Subsetting of Large-scale Atmospheric Datasets". *Future Generation Computer Systems*, Vol. 28(1): pp. 112-118. Elsevier. 2012.

[J3]	Marlon Pierce, Xiaoming Gao, **Sangmi Lee Pallickara**, Zhenhua Gau, and Geoffrey Fox, "QuakeSim Portal and Services: New Approaches to Science Gateway Development Techniques," *Concurrency & Computation: Practice & Experience*. 22(12): pp.1732-1749. 2010.

[J2]	Beth Plale, Dennis Gannon, Yi Huang, Gopi Kandaswamy, **Sangmi Lee Pallickara**, and Aleksander Slominski, "Cooperating Services for Managing Data Driven Computational Experimentation," *IEEE Computing in Science and Engineering (CiSE)*, (Vol. 7, No. 5) pp. 34-43. 2005.

[J1]	Geffrey Fox, Sung-Hoon Ko, Marlon Pierce, Ozgur Balsoy, Jungkee Kim, **Sangmi Lee**, Kangseok Kim, Sangyoon Oh, Xi Rao, Mustafa Varank, Hasan Bulut, Guruhan Gunduz, Xiahong Qui, Shrideep Pallickara, Ahmet Uyar, "Grid Services for Earthquake Science," *Concurrency and Computation: Practice and Experience in ACES Special Issue*, 14(6-7): pp. 371-393, 2002

6.2   Book Chapters

[B3]	**Sangmi Lee Pallickara**, Matthew Malensek*****, and Shrideep Pallickara. "On the Processing of Extreme Scale Datasets in the Geosciences," Handbook of Data Intensive Computing. Springer. 2012.

[B2]	**Sangmi Lee Pallickara**, Shrideep Pallickara and Marlon Pierce, "Scientific Data Management in the Cloud: A Survey of Technologies, Approaches and Challenges," Chapter 22: pp.517-534, Handbook of Cloud Computing. Springer. ISBN: 978-1-4419-6523-3. 2010.

[B1]	Geoffrey Fox, Dennis Gannon, Sung-Hoon Ko, **Sangmi Lee**, Shrideep Pallickara, Marlon Pierce, Xiaohong Qiu, Xi Rao, Ahmet Uyar, Minjun Wang, Wenjun Wu Book chapter on "Peer-to-Peer Grids," *Grid Computing: Making The Global Infrastructure a Reality*, John Wiley & Sons, ISBN 0470853190, 2003

6.3   Refereed Conference Proceedings

[C31]	Johnson Charles Kachikaran Arulswamy*****, and **Sangmi Lee Pallickara,** Columbus: Enabling Scalable Scientific Workflows for Fast Evolving Spatio-Temporal Sensor Data. *(To appear) Proceedings of the 14th IEEE International Conference of Service Computing (IEEE SCC)*, Honolulu, Hawaii, USA, 2017 [20% acceptance rate]

[C30]	Naman Shah*****, Harshil Shah*****, Matthew Malensek*****, **Sangmi Lee Pallickara**, and Shrideep Pallickara. Network Analysis for Identifying and Characterizing Disease Outbreak Influence from Voluminous Epidemiology Data. Proceedings of the *IEEE International Conference on Big Data* (*IEEE BigData*). Washington D.C., USA. 2016. [18.68% acceptance rate]

[C29]	Cameron Tolooee*****, **Sangmi Lee Pallickara** and Asa Ben-Hur. Mendel: A Distributed Storage Framework for Similarity Searching over Sequencing Data. Proceedings of the 30th IEEE International Parallel & Distributed Processing Symposium (IPDPS). pp. 790-799. Chicago, USA, 2016. [23% acceptance rate].

[C28]	Matthew Malensek*****, **Sangmi Lee Pallickara** and Shrideep Pallickara, Alleviation of Disk I/O Contention in Virtualized Settings for Data-Intensive Computing. Proceedings of the *IEEE/ACM International Symposium on Big Data Computing*. pp. 1-10. Cyprus. 2015. [16% acceptance rate]

[C27]	Jared Koontz*****, Matthew Malensek*****, and **Sangmi Lee Pallickara**. GeoLens: Enabling Interactive Visual Analytics over Large-scale, Multidimensional Geospatial Datasets. Proceedings of the *IEEE/ACM Symposium on Big Data Computing*. pp. 35-44. London, UK, 2014. [22% acceptance rate]
	** Best Paper Award

[C26]   Matthew Malensek*****, Walid Budgaga*****, **Sangmi Lee Pallickara**, Neil Harvey and Shrideep Pallickara, Using Distributed Analytics to Enable Real-Time Exploration of Discrete Event Simulations. *Proceedings of the IEEE/ACM International Conference on Utility and Cloud Computing*. pp. 49-58, London, UK, 2014. [19% acceptance rate]

[C25]   Cameron Tolooee*****, Matthew Malensek*****, and **Sangmi Lee Pallickara**. A Framework for Managing Continuous Query Evaluations over Voluminous, Multidimensional Datasets. Proceedings of *the IEEE Cloud and Autonomic Computing Conference*. London, UK. 2014.

[C24]   Matthew Malensek*****, **Sangmi Lee Pallickara** and Shrideep Pallickara. Polygon-Based Query Evaluation over Geospatial Data Using Distributed Hash Tables. Proceedings of the *IEEE/ACM* Conference on Utility and Cloud Computing. pp. 219-226. Dresden, Germany. 2013. [24% acceptance rate]

[C23]   Matthew Malensek*****, **Sangmi Lee Pallickara** and Shrideep Pallickara. "Autonomously Improving Query Evaluations over Multidimensional Data in Distributed Hash Tables." Proceedings of *the ACM Cloud and Autonomic Computing Conference*. Miami, USA. 2013. [35% acceptance rate]

[C22]   Matthew Malensek*****, **Sangmi Lee Pallickara**, and Shrideep Pallickara. Expressive Query Support for Multidimensional Data in Distributed Hash Tables. Proceedings of *the IEEE/ACM Conference on Utility and Cloud Computing*. pp. 31-38. Chicago, USA. 2012. [27% acceptance rate]
   ** Best Paper Award

[C21]   Matthew Malensek*****, **Sangmi Lee Pallickara**, and Shrideep Pallickara. "Galileo: A Framework for Distributed Storage of High-Throughput Data Streams," Proceedings of the *IEEE/ACM Conference On Utility and Cloud Computing*. pp. 17-24. Melbourne, Australia. 2011. [26.7% acceptance rate]

[C20]   **Sangmi Lee Pallickara**, Matthew Malensek***** and Shrideep Pallickara. "Enabling Access to Time-Series, Geospatial Data for On Demand Visualization," Proceedings of the *IEEE Symposium on Large-Scale Data Analysis and Visualization*, Providence, Rhode Island. 2011.

[C19]   **Sangmi Lee Pallickara**, Shrideep Pallickara, Milija Zupanski, and Stephen Sullivan, "Efficient Metadata Generation to Enable Interactive Data Discovery over Large-scale Scientific Data Collections," Proceedings of the IEEE International Conference on Cloud Computing Technology and Science, Indianapolis. November 2010 [24.9% acceptance rate]

[C18]   **Sangmi Lee Pallickara**, Marlon Pierce, Chin Hua Kong, and Qunfeng Dong, "Enabling Large Scale Scientific Computations for Expressed Sequence Tag Sequencing over Grid and Cloud Computing Clusters," Proceedings of the PPAM 2009 Eighth International Conference on Parallel Processing and Applied Mathematics, Wroclaw, Poland, September 13 – 16, 2009

[C17]   **Sangmi Lee Pallickara** and Marlon Pierce, "SWARM: Scheduling Large-scale Jobs over the Loosely-Coupled HPC Clusters," Proceedings of the IEEE International Conference on e-Science. Indianapolis, 2008, December 7-12 2008 [29.1% acceptance rate]

[C16]   Marlon Pierce, **Sangmi Lee Pallickara**, Yu Ma, Mike Lowe, Qunfeng Dong and Samy Meroueh, "Indiana University TeraGrid Gateway Support," Proceedings of The *International Conference for High Performance Computing, Networking, Storage and Analysis* (collocated with *SuperComputing '08*), November, 2008.

[C15]   Yogesh Simmhan, **Sangmi Lee Pallickara**, Nithya N. Vijayakumar, and Beth Plale, "Data Management in Dynamic Environment Driven Computational Science," Proceedings of The *IFIP International Federation for Information Processing*, Vol. 239, Grid-Based Problem Solving Environments, P.W. Gaffney and J.C.T. Pool. Eds. Spring Boston. pp.317-333, 2007

[C14]   **Sangmi Lee Pallickara**, Beth Plale, Liang Fang, Dennis Gannon, "End-to-End Trustworthy Data Access in Data-Oriented Scientific Computing," Proceedings of The *IEEE Cluster Computing and Grid (CCGRID)*

2006: 395-400. [23.7% acceptance rate]

[C13]	Yiming Sun, Scott Jensen, **Sangmi Lee Pallickara**, and Beth Plale, "Personal Workspace for Large-scale Data-driven Computational Experimentation," Proceedings of *the 7th IEEE/ACM International Conference on Grid Computing (Grid'06)*, Barcelona. 2006. [20.2% acceptance rate]

[C12]	Dennis Gannon, Beth Plale, Marcus Christie, Liang Fang, Yi Huang, Scott Jensen, Gopi Kandaswamy, Suresh Marru, **Sangmi Lee Pallickara**, Satoshi Shirasuna, Yogesh Simmhan, Alex Slominski, and Yiming. Sun, "Service Oriented Architectures for Science Gateways on Grid Systems," Proceedings of The *International Conference on Service Oriented Computing 2005*, B. Benatallah, F. Casati, P. Traverso (Eds.), LNCS 3826, pp. 21-32, 2005. Springer-Verlag Berlin Heidelberg 2005

[C11]	**Sangmi Lee Pallickara**, Beth Plale, Scott Jensen, Yiming Sun, Short Paper: "Monitoring Access to Stateful Resources in Grid Environments," Proceedings of the *IEEE International Conference on Services Computing*, Orlando, Florida, pp. 343-346, 2005.

[C10]	Shrideep Pallickara, Geoffrey Fox, Beytullah Yildiz, **Sangmi Lee Pallickara**, Sima Patel and Damodar Yemme, "On the Costs for Reliable Messaging in Web/Grid Service Environments," Proceedings of the *IEEE International Conference on e-Science & Grid Computing*, Melbourne, Australia. 2005. [31.6% acceptance rate]

[C9]	Shrideep Pallickara, Geoffrey Fox, and **Sangmi Lee Pallickara**, "An Analysis of Reliable Delivery Specifications for Web Services," Proceedings of the *IEEE ITCC Conference on Information Technology*. Las Vegas. 2005.

[C8]	**Sangmi Lee** and Geoffrey Fox, "Wireless Reliable Messaging Protocol for Web Services (WS -WRM)," Proceedings of the *IEEE 2nd International Conference on Web Services (ICWS 2004)*, pp.350-357, San Diego. 2004. [28.7% acceptance rate]

[C7]	Geoffrey Fox, Sunghoon Ko, Kangseok Kim, **Sangmi Lee**, Sangyoon Oh, "Universally Accessible Collaboration Frameworks for Ubiquitous Computing Environments," Proceedings of the *International Conference on Ubiquitous Computing (ICUC2003)*, Seoul Korea, 2003.

[C6]	**Sangmi Lee**, Sunghoon Ko, Geoffrey Fox, "Adapting Content for Mobile Devices in Heterogeneous Collaboration Environments," Proceedings of the *International Conference on Wireless Networks (ICWN'03)*, pp. 211-217, 2003

[C5]	Geoffrey Fox, Hasan Bulut, Kangseok Kim, Sung-Hoon Ko, **Sangmi Lee**, Sangyoon Oh, Shrideep Pallickara, Xiaohong Qiu, Ahmet Uyar, Minjun Wang, Wenjun Wu "Collaborative Web Services and Peer-to-Peer Grids," Proceedings of the *IEEE Collaborative Technologies Symposium (CTS'03)*, 2003.

[C4]	**Sangmi Lee**, Sunghoon Ko, Geoffrey Fox, Kangseok Kim, Sangyoon Oh, "A Web Service Approach to Universal Accessibility in Collaboration Services," Proceedings of the *International Conference on Web Services (ICWS'03)*, pp. 333-339, USA. 2003. [29.8% acceptance rate]

[C3]	Geoffrey Fox, Sung-Hoon Ko, Kangseok Kim, Sangyoon Oh, **Sangmi Lee**, "Integration of Hand-Held Devices into Collaborative Environments," Proceedings of the *International Conference on Internet Computing (IC'02)* pp.231-250, 2002. USA.

[C2]	**Sangmi Lee**, Geoffrey Fox , Sunghoon Ko, Minjun Wang, Xiaohong Qiu ,"Ubiquitous Access for Collaborative Information System Using SVG," Proceedings of th*e SVG OPEN Conference*. 2002, Zurich, Switzerland

[C1]	Hasan Bulut, Geoffrey Fox, Dennis Gannon, Kangseok Kim, Sung-Hoon Ko, **Sangmi Lee**, Sangyoon Oh, Xi Rao, Shrideep Pallickara, Quinlin Pei, Marlon Pierce, Aleksander Slominski, Ahmet Uyar, Wenjun Wu, Choonhan Youn, "An Architecture for e-Science and its Implications," Proceedings of *the IEEE International Symposium on Performance Evaluation of Computer and Telecommunication Systems*

(SPECTS). July, 2002.

6.4   Refereed Workshop Proceedings

[W5]   **Sangmi Lee Pallickara** and Beth Plale, "Enabling End-to-End Trustworthiness in Data-Oriented Scientific Computing," Proceedings of *the Workshop on Web Services-based Grid Applications (WGSA'06)* in association with International Conference on Parallel Processing (ICPP-06). 2006.

[W4]   Scott Jensen, Beth Plale, Sangmi Lee Pallickara, and Yiming Sun, "A Hybrid XML-Relational Grid Metadata Catalog," Proceedings of the Workshop on Web Services-based Grid Applications (WGSA'06) in association with International Conference on Parallel Processing (ICPP-06). 2006.

[W3]   **Sangmi Lee Pallickara**, Beth Plale, Scott Jensen, and Yiming Sun, "Structure, sharing, and preservation of scientific experiment data," Proceedings of the *IEEE 3rd International Workshop on Challenges of Large Applications in Distributed Environments (CLADE)*, pp 107-114, 2005.

[W2]   Sangyoon Oh, **Sangmi Lee Pallickara**, Sunghoon Ko, Jai-Hoon Kim, Geoffrey Fox, "Cost Model and Adaptive Scheme for Publish/Subscribe Systems on Mobile Environments," Proceedings of the *International Workshop on Active and Programmable Grids Architectures and Components (APGAC05), Lecture Notes in Computer Science, Springer-Verlag* 3516, pp. 275-278, Atlanta, USA, 2005.

[W1]   Sangyoon Oh, **Sangmi Lee Pallickara**, Sunghoon Ko, Jai-Hoon Kim, Geoffrey Fox, Publish/Subscribe Systems on Node and Link Error Prone Mobile Environments," Proceedings of the *Wireless and Mobile Systems Workshop (ICCS 2005), Lecture Notes in Computer Science, Springer-Verlag*, 3515, pp. 576-584

6.5   Ph.D. Dissertation

Sangmi Lee. August 2003.
Title: A Modular Data Pipelining Architecture for Enabling Universal Accessibility in Peer to Peer Grids.
Advisors: Prof. Geoffrey Fox and Prof. Gordon Erlebacher

6.6   Technical Reports and Posters (Not Peer-Reviewed)

[T6]   Duck Keun Yang*****, Zach Cutler*****, Jared Koontz*****, **Sangmi Lee Pallickara**, Adam Gaylord, Joe von Fischer, "Enabling Active Data Collection and Dissemination of Methane Concentrations," Research Poster at Celebrating Undergraduate Research and Creativity (CURC), Colorado State University, April 2015

[T5]   Kong, C.H., **Sangmi Lee Pallickara**, and Marlon Pierce, "Fault Detection of TeraGrid Resources Using Inca," June, 2009. Poster at *TeraGrid* '09

[T4]   **Sangmi Lee Pallickara**, Marlon Pierce, "Orienting Scientific Data Management to Harness the Data Cloud," Digital Science Center, Indiana University, Dec, 2008, Technical Report.

[T3]   Geoffrey Fox, **Sangmi Lee**, Sung-Hoon Ko, Kangseok Kim, Sangyoon Oh, "CAROUSEL: Universally Accessible Web Service Architecture for Collaborative Applications," Community Grids Lab, Indiana University, 2002

[T2]   Geoffrey Fox, Sung-Hoon Ko, Kangseok Kim, **Sangmi Lee**, Sangyoon Oh, "Status of Hand-Held Interfaces to the Garnet Collaborative Environment," Community Grids Laboratory, Indiana University, 2002

[T1]   **Sangmi Lee**, "Overview of the Virtual Collaborative Network (VNC)," Department of the Computer Science, Florida State University, 2000

7   Research Projects

Nov 14 –   GLEAN
The Glean project focuses on performing analytics at scale over Big Data. The datasets we consider are in the order of Petabytes and encompass billions of files representing trillions of observations, measurements, or simulation datapoints. Glean achieves this by combing innovations in large-scale storage systems, cloud computing, machine learning, and statistics. A particular focus of this effort is to perform analytics in real-time over streaming data representing time-series observations.

Nov10 –   GALILEO
This geospatial data storage system is designed to provide efficient access to time-varying geospatial datasets generated by networked observational instruments such as radars and satellites. The storage system is designed to enable large-scale visualizations and processing of geospatial datasets. http://galileo.cs.colostate.edu

Jan 13 –   GEOLENS
GEOLENS is an interactive visual analytics framework that supports fast and expressive interactions with voluminous geo-spatial datasets. It incorporates support for advanced visual analytics techniques such as brushing and linking.

Apr 10 –   ADDS (ATMOSPHERIC DATA DISCOVERY NETWORK SERVICE)
Dec 12   This project investigates methods to efficiently discover data that are part of large, binary collections published by major data hosting services. ADDS crawls such servers and builds metadata collections that can be queried to access relevant portions of the dataset. The service supports the BUFR binary data format for observational datasets and the netCDF format that is used to encode simulation output data. This research is based on a collaborative effort with the Cooperative Institute for Research in Atmosphere (CIRA) at Colorado State University and the University Corporation for Atmospheric Research (UCAR) at Boulder.

Aug 07 –   TERAGRID project
Dec 09   This open scientific discovery infrastructure combines resources at eleven partner sites to create an integrated, persistent computational resource. Using high-performance network connections, the TERAGRID integrates high-performance computers, data resources and tools.

As part of my work on this project, I designed the SWARM meta scheduling system. SWARM leverages Condor and Globus technologies to harness traditional high throughput computing clusters. It is not unusual for these high-throughput jobs to run for several days at a time. SWARM is presently deployed in a gene sequencing application developed by the *Center for Genomics and Bioinformatics* at Indiana University.

Aug 07 –   POLAR GRID project

Dec 09	The NSF-funded POLAR GRID project aims to understand the impact of rising sea levels and how they relate to global climate change. As part of this effort I worked on building science gateways that support data collection, examination and analysis for ice sheet research that help scientists. My research within the project included issues relating to metadata and storage.
Sep 04 – Jul 07	LEAD (Linked Environments for Atmospheric Discovery) project: The NSF-funded LEAD project makes meteorological data, forecast models, and analysis and visualization tools available to anyone who wants to interactively explore the weather as it evolves http://lead.ou.edu. My research on this project has included work on data cataloging, transferring and storing schemes based on Service Oriented Architecture (SOA) principles. These issues were investigated within the context of the MYLEAD system, which I designed.
Sep 01 – Aug 04	CAROUSEL The CAROUSEL project focused on developing an environment for supporting ubiquitous accesses to real-time collaborative applications in Grid settings. Devices that were supported include portable devices, such as 3G SmartPhones and 801.11b equipped PDAs and conventional desktop PCs. As part of this project, I designed a data pipelining architecture that was formally verified using Perti Nets. I also developed a protocol for reliable communications between these pervasive devices in wireless settings.

8   Teaching

Department of Computer Science, Colorado State University

Term	Course	Enrollment	Audience
Spring 2017	CS435: *Introduction to Big Data*	57	Senior undergraduate course
Fall 2016	CS535: *Big Data*	20	Graduate Course
Spring 2016	CS435: *Introduction to Big Data*	60	Senior undergraduate course
Fall 2015	CS535: *Big Data*	22	Graduate course
Spring 2015	CS480A: *Introduction to Big Data* *	39	Senior undergraduate course
Fall 2014	CS581/CS535: *Big Data*	25	Graduate course
Fall 2013	CS581: *Big Data**	22	Graduate course
Spring 2013	CS200: *Algorithms and Data Structures*	79	Core undergraduate course
Spring 2013	CS480: *Principles of Data Management*	18	Senior undergraduate course
Fall 2012	CS200: *Algorithms and Data Structures*	100	Core undergraduate course
Spring 2012	CS480: *Principles of Data Management**	17	Senior undergraduate course
Spring 2012	CS200: *Algorithms and Data Structures*	57	Core undergraduate course
Fall 2011	CS200: *Algorithms and Data Structures*	72	Core undergraduate course

* Offered for the first time

Department of Computer Science, Indiana University

Co-Instructor	_Fundamentals of Computer Networks_, Spring 2006. (Undergraduate course)	
Instructor	Indiana University Science Summer Camp, Summer 2005. (High School level)	
Guest Instructor	_Distributed Systems_, Spring 2005. (Graduate course)	
Guest Instructor	_Fundamentals of Computer Networks_, Spring 2005. (Undergraduate course)	
Supervision of Graduate Students	Sep 08 –Dec 09	Chin Hua Kong: Fault detection in Grid Computing Environments
	Jul 05–Aug 06	Suba Periyasami: MVC based Portal interfaces for Web-based data sharing.
	Sep 05–Aug 06	Ning Liu: Trustworthy scientific data management
NSF-REU Program	Aug 06- Jul 07	Supervised 2 undergraduate students (Christina Hoffa and You-Wei Chaeh) as part of the NSF's REU program,

SWiFT: Education Outreach for Female High School Students [2016-2021]

I organize a weeklong summer educational camp for 12 rising 11th grade female students from the Rocky Mountain High School and the Fort Collins High School in Fort Collins, Colorado. _SWiFT: Summer STEM Camp for Women in Fort Collins_ (http://swift.cs.colostate.edu), was held in the Computer Science Department at Colorado State University. This is a 5-year effort. The camp provides activity-based learning and hands-on experiences to demonstrate how mathematical concepts that students learn in school are applied to solve real-world problems using data science.

9   Professional and Academic Service

Member of Editorial Board

Journal of Big Data. Publisher: Springer New York/Heidelberg, Germany. 2013 – present

Technical Program Committees: Conferences, Symposia, and Workshops

Program Committee Member: IEEE/ACM International Conference on Utility and Cloud Computing, Shanghai, China, 2017.
Program Committee Member: International Workshop on _Autonomic Management of Grid and Cloud Computing_, Co-located with the ACM Cloud & Autonomic Computing Conference. 2017.
Program Committee Member: ACM Symposium on High Performance Parallel and Distributed Computing (HPDC). Washington D.C., USA. 2017.
Program Committee Member: IEEE/ACM International Symposium on Big Data Science, Engineering and Applications, Shanghai, China, 2016.
Program Committee Member: IEEE/ACM International Conference on Utility and Cloud Computing, Shanghai, China, 2016.
Program Committee Member: International Workshop on _Autonomic Management of Grid and Cloud Computing_, Co-located with the ACM Cloud & Autonomic Computing Conference. 2016.
Program Committee Member: ACM Symposium on High Performance Parallel and Distributed Computing (HPDC). Kyoto, Japan. 2016.
Program Committee Member: ACM Symposium on High Performance Parallel and Distributed Computing (HPDC). Vancouver, Canada. 2015.
Program Committee Member: International Workshop on _Autonomic Management of Grid and Cloud Computing_, Co-located with the ACM Cloud & Autonomic Computing Conference. 2015.

Program Committee Member: IEEE/ACM International Conference on Utility and Cloud Computing. Cyprus. 2015
Program Committee Member: ACM Symposium on High Performance Parallel and Distributed Computing (HPDC). 2014.
Program Committee Member: IEEE/ACM International Conference on Utility and Cloud Computing. London, UK. 2014.
Program Committee Member: International Workshop on *Autonomic Management of Grid and Cloud Computing*, Co-located with the ACM Cloud & Autonomic Computing Conference. 2014.
Program Committee: IEEE/ACM International Conference on Utility and Cloud Computing. Dresden, Germany. 2013
Program Committee: International Workshop on *Autonomic Management of Grid and Cloud Computing*, Co-located with the ACM Cloud & Autonomic Computing Conference. 2013.
Program Committee: IEEE/ACM International Conference on Utility and Cloud Computing. Chicago, USA. 2012
Program Committee: IEEE/ACM International Conference on Utility and Cloud Computing. Melbourne, Australia, 2011.

NSF Panel
2013, 2016

Journal Reviewer
I also review manuscripts for IEEE Transactions on Cloud Computing, IEEE Transactions on Big Data, IEEE Internet of Things Journal, IEEE Transactions on Service Computing, *Future Generation Computer Systems*, Journal of *Geographical Information Science*, International Journal of Geographical Information, Springer *Cluster Computing* and International Journal of *Distributed Sensor Networks*.

Outreach Service

Director	2016, 2017 SWiFT Education Outreach for Female High School Students, Fort Collins, Colorado
Committee Member, Judge	2017 CURC (Celebrate Undergraduate Research and Creativity), Colorado State University 2015 CURC (Celebrate Undergraduate Research and Creativity), Colorado State University 2014 CURC (Celebrate Undergraduate Research and Creativity), Colorado State University 2006 Indiana Women in Computing (INWIC) 2005 Women in Science Program (WISP) 7th Annual Research Day
Co-organizer, RAM Welcome	2012 Computer Science, Colorado State University.
Session Instructor	2005 ACM JETT (Java Engagement for Teacher Training), Indiana University
Assistant Session Instructor	2004 ACM JETT (Java Engagement for Teacher Training), Indiana University

Departmental Service

Graduate Committee, Computer Science Department (2014 –)
Operations Committee, Computer Science Department (2013 – 2014)

10 Academic Advising

Current Graduate Advisees

Walid Budgaga, Ph.D. candidate [Co-advised with Shrideep Pallickara]
Saptashwa Mitra, Ph.D.
Chuanqi Huang, M.S.
Yu Qui, M.S.
Maxwell Roselius, M.S.

Current Undergraduate Research Advisees

Scott Magisano (Jan 2017 ~)

Undergraduate Honors Thesis Advisee

Erin Doan (May 2017 -)

Supervised Undergraduate Honors Thesis

Sadie Henry, B.S. (May 2017)

Supervised Graduate Degrees

Matthew Malensek (Ph.D., 2017)   [Co-advised with Shrideep Pallickara]
Dissertation: Low-latency, Query-Driven Analytics over Voluminous Multidimensional, Spatiotemporal Datasets
Currently: Assistant Professor, Computer Science – University of San Francisco

Naman Rajivkumar Shah, M.S. (June 2017)
Johnson Charles Kachikaran Arulswamy. M.S. (March 2017)
Harshil Shah, M.S. (Dec 2016)   [Co-advised with Shrideep Pallickara]
Cameron Tolooee, M.S. (Dec 2015)
Jared Koontz, M.S. (May 2015)
Budgaga Walid, M.S. (Aug. 2014)   [Co-advised with Shrideep Pallickara]

Ph.D. Dissertation Committee Memberships

Thilina Buddhika
Ryan Stern
Sachini Weerawardhana

Ph.D. Dissertation Committee External Committee Memberships

Physics: Matthew Judah

Masters Thesis Committee Memberships

Thilina Buddhika (2015)
Darshan Washimker (2016)
Leo Vigneshwaran Sudalaikkan (2016)

Undergraduate Honors Theses Committee

Amanda Carbonari (2015)
Danielle Alexander (2013)

Masters Thesis External Committee Memberships

Electrical and Computer Engineering: Shibayan Chatterjee (2016)
Physics: Matthew Judah (May 2016)

Masters Project External Committee Memberships

Electrical and Computer Engineering: Tushar Jagtap (Nov. 2014), Marvin Antony Devadass (Apr. 2015)

Past Undergraduate Research Advisees

Maxwell Roselius (July 2016 ~ Dec 2016)
Xuehao Hu (Jan 2015 – May 2016)
Duck Keun Yang (Jan 2015 – May 2016)
Zachary Cutler (Dec. 2014)
Cameron Tolooee (Dec 2013)
Li (Mike) Yang (Dec 2013)

11 Invited Talks

Enabling Petascale Predictive Analytics. Rocky Mountain Advanced Computing Consortium. August 2016.

Scalable Analytics Over Unstructured Multidimensional Time-series Data. National Data Integrity Conference. 2015.

Enabling Scalable Analytics Over Unstructured Multidimensional Time-series Data, Physics Department Colloquium series at Colorado State University, Oct. 2014.

Math in Big Data, Las Chicas de Matematicas, University of Northern Colorado, June. 2014.

Queries to Support Analytics Over Unstructured Multidimensional Time-series Data, CISL Seminar Series, May. 2014, NCAR, Boulder.

Scalable Query Evaluations over a Billion Files, National Science Foundation, EarthCube, *Big Weather* Workshop, Jan. 2014, NCAR, Boulder.

Expressive query evaluations over fast evolving large-scale geospatial datasets, ACM SC13 KIISE-KOSCEA HPC SIG Joint Workshop on High Performance and Throughput Computing at ACM *SuperComputing* Conference, Nov. 2013, Denver.

Supporting Scalable Server-side data preprocessing in distributed data storage systems, ACM SC13 Workflow BoF at the ACM *SuperComputing* Conference, Nov. 2013, Denver.

A Small Window into the World of Big Data, CS department seminar series, Nov. 2013.

Supporting Knowledge Extraction from Unstructured Multidimensional Time-series Data, Computer Science IAB (Industrial Advisory Board) meeting, Oct. 2013.

Support for Expressive Query Evaluations over Unstructured Multidimensional Time-series Data, BMAC Computer Science Department Colloquium series at Colorado State University, Sept. 2013.

Working with Big Data in the Geosciences: Finding the needle in the haystack, ISTeC Big Data Forum, CSU, Apr. 2013.