Last update on January 8, 2016

Saliya Ekanayake

saliya.org | esaliya@gmail.com | 812-391-4914
www.linkedin.com/in/saliya
752 Woodbridge Drive, Bloomington, IN 47408

Summary

Computer scientist with expertise in parallel and distributed computing with the emphasis on high performance big data benchmarking

Education

Indiana University — Bloomington, IN

Ph.D. in Computer Science — *Fall 2009 - spring 2016 (expected)*

Dissertation: *Towards a Systematic Approach to Big Data Benchmarking*

This work investigates the challenges in developing high performance big data applications using a set of Global Machine Learning (GML) applications with significant computation and communication as a yardstick. It presents several optimizations applicable in general and demonstrates the performance and scalability with and without these improvements. Also, it presents a multifaceted and multidimensional classification of existing big data benchmarks. One outcome of this research includes a suite of high performance distributed parallel machine learning applications - SPIDAL

M.Sc. in Computer Science - GPA 4.0 — *Fall 2009 - spring 2011*

Major: Systems - *Distributed Systems, Cloud Computing, Scientific Computing*
Minor: Programming Languages - *Parallelism in Programming Languages, Programming Language Principles, Programming Language Implementation, Programming Language Foundations*

University of Moratuwa — Moratuwa, Sri Lanka

B.Sc. in Computer Science & Engineering - GPA 3.75 — *2004 - 2008*

Experience

Indiana University — Bloomington, IN

Research Assistant — *Fall 2009 – present*

Investigated the challenges in developing high performance big data applications and presented optimizations to improve performance and scalability.

Developed Scalable Parallel Interoperable Data Analytics Library (SPIDAL) in Java - a suite of high performance clustering and multidimensional scaling algorithms

Developed parallel tools and performed large scale analyses on gene sequences as part of a collaborative projects with the Biology Department at Indiana University

WSO2 Inc. — Colombo, Sri Lanka

Senior Software Engineer — *2009*

Developed OSGI components to support proxy services and endpoints in WSO2 Enterprise Service Bus (ESB). WSO2 ESB is the middleware used in ebay for its transactions.

Software Engineer — *2008 - 2009*

Contributed features and bug fixes to Apache Web Services projects such as Apache Axiom, Apache Synapse, and Apache Axis2.

Intern — *2006 - 2007*

Implemented JavaScript based web service deployment support in Apache Axis2 engine. Also, developed a JavaScript client library (WSRequest) to consume web services.

University of Moratuwa. — Moratuwa, Sri Lanka

Visiting Lecturer at the Dept. of Computer Science and Engineering — *2008 - 2009*

Taught Data Structures and Algorithms for 2nd year undergraduates in Computer Science and shared responsibility over syllabus, course structure, and grades with another lecturer

Instructor for "Introduction to Object Oriented Programming with C++" — *Spring 2007*

Assisted students with class projects, assignments, and hands-on sessions

St. Mary's Convent — Matara, Sri Lanka

Assistant Physics Teacher — *2003 - 2004*

Assisted students with freshman level physics assignments and lab work

Awards

Best Intern Presentation awarded by the Institute of Engineers in Sri Lanka (IESL) for implementing and presenting the JavaScript support in Apache Axis2 Web services engine

National Best Quality Software Merit Award for the project Rampart2 - a high performance implementation of web services security specification for Apache Axis2

Skills

Technical Expertise: Well experienced with High Performance Computing (HPC), parallel and distributed computing, high performance optimization techniques for Java, and large scale data analysis

Programming Languages: Proficient in Java, C#, Scheme. Familiar with C, C++, and Python

Distributed and Parallel Runtimes: MPI, OpenMP, Hadoop, Spark

Operating Systems: Windows, Windows Server & HPC Tools, Linux

Projects

Indiana University — Bloomington, IN

SPIDAL is a suite of highly optimized multi-dimensional reduction (MDS) and clustering algorithms designed to analyze big data on large multi-core HPC clusters. Its main improvement is zero intra-node messaging using Java memory map based inter-process communication within a node. Traditional MPI experience significant performance degradation with increasing process counts per node; hence the improvement using memory maps. Other optimizations include zero garbage collection using off-heap buffers and improved cache utilization using reordered loops and 1D arrays. SPIDAL is written in Java, and uses MPI for inter-node communication.

Million Sequence Clustering is a pipeline of applications and tools designed to analyze similarity and structure of large sequence data sets. It includes a number of applications such as data sanitizing, parallel implementations of sequence alignment, MDS and clustering, post processing, 3D results visualization, and job submission.

TryOut is a C# based desktop job submission and monitoring tool for Windows HPC clusters. It allows users to quickly change application (such as MDS or clustering) configuration options and deploy a job to the cluster. It registers handlers to redirect the program output to TryOut. Also, it saves metadata so job monitoring is resilient to network failures and application shutdowns.

Compiler for Scheme Language is a functional compiler with optimizations implemented from scratch to support the Scheme language. It has a Universal Intermediate Language (UIL) layer, which can be extended to support languages other than Scheme as well.

Heat-map Generator is a 2D heat map and histogram generator program for large binary or vector data. It is parallel application written using MPI and produces a script that invokes GnuPlot to generate the graphs.

Science on Dynamic Virtual Clusters is a proof of concept on dynamic virtual clusters for Super Computing 2009 conference. Implemented a graphical monitoring tool to show performance statistics and usage on dynamically changing clusters. Used Java, JFreeChart, and NaradaBrokering.

University of Moratuwa — Moratuwa, Sri Lanka

Rampart2 is a project to develop a high performance Web services security module for Apache Axis2 and I implemented the XML Signature specification.

XML C14N (Google Summer of Code 2007) is an implementation of the Canonical XML and Exclusive XML Canonicalization specifications for Apache Axis2.

TCP Monitor plug-in for Eclipse implements Apache TCPMon utility as a plug-in for Eclipse

Ency is Java based steganography application that encrypts text and stores in .bmp images. It can also retrieve and decrypt text from such images.

Publications

In Publication

- **S. Ekanayake**, S. Kamburugamuve, and G. Fox, "SPIDAL: High Performance Data Analytics with Java and MPI on Large Multicore HPC Clusters", submitted to the 24th High Performance Computing Symposium (HPC 2016), Pasadena, CA

Peer-reviewed Papers

- F. Fox, S. Jha, J. Qiu, **S. Ekanayake**, "Towards a Comprehensive Set of Big Data Benchmarks", in Proceedings of the International Advanced Research Workshop on High Performance Computing, Cetraro, Italy, July, 2014.
- R. Yang, G. L. House, **S. Ekanayake**, U. Schutte, J. D. Bever, T. Haixu, and G. Fox, "Integration of Clustering and Multidimensional Scaling to Determine Phylogenetic Trees as Spherical Phylograms Visualized in 3 Dimensions." pp. 720-729.
- Y. Ruan, **S. Ekanayake**, M. Rho, H. Tang, S.-H. Bae, J. Qiu, and G. Fox, "DACIDR: deterministic annealed clustering with interpolative dimension reduction using a large collection of 16S rRNA sequences", in Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine, Orlando, Florida, 2012, pp. 329-336.
- J. Qiu, S. Beason, S.-H. Bae, **S. Ekanayake**, and G. Fox, "Performance of Windows Multicore Systems on Threading and MPI", in Proceedings of the 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, 2010, pp. 814-819.
- A. Hughes, Y. Ruan, **S. Ekanayake**, S. H. Bae, Q. Dong, M. Rho, J. Qiu, and G. Fox, "Interpolative multidimensional scaling techniques for the identification of clusters in very large sequence sets", BMC Bioinformatics, vol. 13 Suppl 2, pp. S9, 2012.
- J. Qiu, J. Ekanayake, T. Gunarathne, J. Y. Choi, S. H. Bae, H. Li, B. Zhang, T. L. Wu, Y. Ruan, **S. Ekanayake**, A. Hughes, and G. Fox, "Hybrid cloud and cluster computing paradigms for life science applications", BMC Bioinformatics, vol. 11 Suppl 12, pp. S3, 2010.
- L. Stanberry, R. Higdon, W. Haynes, N. Kolker, W. Broomall, **S. Ekanayake**, A. Hughes, Y. Ruan, J. Qiu, E. Kolker, and G. Fox,"Visualizing the protein sequence universe", in Proceedings of the 3rd international workshop on Emerging computational methods for the life sciences, Delft, The Netherlands, 2012, pp. 13-22.

Book Chapters

- J. Qiu, J. Ekanayake, T. Gunarathne, J. Y. Choi, S.-H. Bae, Y. Ruan, **S. Ekanayake**, S. Wu, S. Beason, G. Fox, M. Rho, and H. Tang, "Data Intensive Computing for Bioinformatics", Bioinformatics: Concepts, Methodologies, Tools, and Applications, pp. 287-321: IGI Global, 2013.

Technical Reports

- **S. Ekanayake**, "Evaluation of Java Message Passing in High Performance Data Analytics", 2014
- **S. Ekanayake**, "Survey on High Productivity Computing Systems (HPCS) Languages", Pervasive Technology Institute, Indiana University, Bloomington, 2013.
- **S. Ekanayake**, "Study of Biological Sequence Clustering", Pervasive Technology Institute, Indiana University, Bloomington, 2013.
- **S. Ekanayake**, "Sequence Clustering Tools", Pervasive Technology Institute, Indiana University, Bloomington, 2013.
- A. L. Huges, J. Qiu, **S. Ekanayake**,T. Gunarathne, S. T. Wu, H. Li, J. Y. Choi, S. H. Bae, Y. Ruan, "Parallel Applications And Tools For Cloud Computing Environments", in 2nd IEEE International Conference on Cloud Computing Technology and Science (CloudCom 2010), Indianapolis, USA, 2010.
- A. Hughes, **S. Ekanayake**, J. Qiu, and G. C. Fox, "Service-Oriented Architecture for Biosequence Analysis Workflow", Technical Report, 2010

Presentations

- **S. Ekanayake**, "Twister K-Means Clustering", in 2nd IEEE International Conference on Cloud Computing Technology and Science (CloudCom 2010), Indianapolis, USA, 2010.