Yuan Luo

Email: yuan@yuanluo.net, Homepage: yuanluo.net Cell: (812)272-0208

Education

Indiana University Bloomington Ph.D. in Computer Science *2015*
Jilin University (China) B.S. and M.S. in Computer Science and Technology *2005, 2008*

Career Interests

Big Data, Data Engineering, Platform, Growth and Product Innovation

Selected Technical Skills

Java, C/C++, Shell Script, Python, ETL, Hadoop, Spark, Hive, HBase, MongoDB, RabbitMQ, REST, AWS, RDBM, HPC and other skills related to Cloud Computing, Distributed Systems, Service Oriented Architecture and Big Data.

Experience

Data Architect, Edmodo, Inc. *09/2015 - present*

- Data Platform: Architected data pipeline workflow management system, simplified legacy ETLs, defined and implemented new ETLs. Architected hybrid data platform to enable Data Science as a Service, easy and fast consumption of 70 million user data for internal use and for various analytics-based product offerings.
- Product: Designed and implemented Edmodo's social recommendation system to increase user engagement and onboarding.
- Email and Push Notification Platform: Enhanced the platform to handle recipients and contents preparation and delivery for millions of users on regular basis.

Co-founder, Chinese Online Fitness Initiative (now FitTime) *01/2014 – 01/2015*

- Co-founded a Chinese online fitness initiative which later became the well-known FitTime 即刻运动, helped the initiative grow user base, architect web platform, launch mobile apps, define and implement online marketing strategies.

Research Assistant, Indiana University Data to Insight Center *08/2009 – 08/2015*

- Created Hierarchical MapReduce (HMR), an extension to MapReduce framework that runs MapReduce jobs across aggregated resources from multi-cloud environment at worldwide scale. The HMR papers have been selected by several professors as teaching materials in their graduate courses in Indiana University and the University of Minnesota Twin Cities. The HMR work has also been cited by at least 6 patents from top tech companies.
- Authored Virtual Cluster Controller, a service that creates virtual clusters across multiple cloud platforms, builds a network overlay, and manages the virtual cluster lifecycle in a fault-tolerant manner.
- Architected PRAGMA Cloud at Indiana University, as part of National Science Foundation (NSF) funded PRAGMA project. PRAGMA Cloud then had 13 VM hosting clusters, spanning 11 institutes and 7 different countries/regions over 4 different VM management platforms.
- Authored a matchmaking system, a standalone service that dynamically loads user requirements in JSON, generate java classes and instantiate POJOs without pre-defined schema, and leverages Drools rule engine to make matchmaking decisions. Matchmaker is a key component of NSF DataNet funded Sustainable Environments-Actionable Data (SEAD) project.
- Developed NSF funded Karma Provenance Collection Tool, a standalone tool that can be added to existing cyber-infrastructure for collection and representation of provenance data.
- Led the development of NASA funded InstantKarma project, collect provenance within the NASA Earth Science community.
- Authored a hybrid workflow system. Qualitative and quantitative study of the tradeoffs of a hybrid workflow solution that utilizes multiple workflow systems and solutions to execute a single workflow.
- Developed BPEL-based workflows for NSF funded Linked Environments for Atmospheric Discovery (LEAD) project.

Research Intern, IBM T.J. Watson Research Center *07/2012 - 10/2012*

- Built an infrastructure for running both transactional and analytics workloads. Authored a consistency mechanism syncs data between a NoSQL store (HBase by default) and a distributed transactional cache (IBM Websphere eXtreme Scale).

Research Intern, National Biomedical Computation Resource, University of California, San Diego *06/2009 - 08/2009*

- Extended Opal Toolkit, wrapper for scientific applications as web services on Grid and Cloud resources. Integrated Opal with CSF4 meta-scheduler so that Opal jobs can be scheduled onto heterogeneous HPC clusters managed by batch schedulers.

Researcher, National Biomedical Computation Resource, University of California, San Diego *07/2006 - 09/2006*

- Developed MyWorksphere project, which served as one of the NSF funded TeraGrid science gateway. It was an end-to-end SOA environment that allows existing applications to run transparently on the Grid.

Researcher, Platform Computing (now IBM) Grid Technology Research Center at Jilin University, China *09/2005 - 06/2008*

- Developed the Community Scheduler Framework 4 (CSF4), the first WSRF compliant grid meta-scheduler that was released as an execution management service of Globus Toolkit 4. Developed Platform Enterprise Grid Orchestrator (EGO), the first and then-only grid platform that delivers virtualization, automation and sharing of all resources to applications.

Teaching Experience

Associate Instructor, School of Informatics and Computing, Indiana University Bloomington *08/2009-05/2010; 08/2014-12/2014*

- Graduate Course: Management, Access, and Use of Big and Complex Data, with Prof. Beth Plale, Fall 2014
  Designed projects that emphasis on (Twitter) data analytics and performance evaluation of data storage (MongoDB) and queries.
- Graduate Course: Distributed Systems, with Prof. Beth Plale, Spring 2010.

Course project: Parallel ray tracing application using Amazon Elastic MapReduce and Web Services.
- Graduate Course: Algorithms Design and Analysis, with Prof. Paul Purdom, Fall 2009.
  Course project: Write a fast routine that multiplies large non-negative integers, code must be correct up to a million bits.
- Undergraduate Course: Introduction to Computers and Computing, with Charles E. Pope, Fall 2009.

Honors

K. Jon Barwise Fellowship, Indiana University School of Informatics and Computing *2008-2009*

Visiting Scholar, University of California, San Diego, hosted by Dr. Peter Arzberger, and Dr. Wilfred Li *2006*

Excellent undergraduate thesis (design) 2005, Jilin University; Excellent Student Leadership 2001-2002, Undergraduate Student Award 2003-2005, Graduate Student Award 2005-2006, Jilin University; Graduate Student Fellowship with full tuition waiver and monthly stipend 2005-2008, Ministry of Education, China

Selected Professional Services and Activities

- Technical Program Committee Member, The 3rd, 4th and 5th IEEE International Conference on Cloud Networking (2014, 2015 and 2016)
- Workshop Co-chair, The 1st - 4th PRAGMA Students Workshop, 10/2012, 03/2013, 10/2013 and 04/2014
- Session Chair and Program Committee Member, The 23rd - 27th PRAGMA Workshop at Seoul, Korea; Bangkok, Thailand; Beijing, China; Tainan, Taiwan; Bloomington, Indiana, USA, 10/2012, 03/2013, 10/2013, 04/2014 and 10/2014
- Journals Reviewer, Concurrency and Computation: Practice and Experience (2010 -); Scalable Computing (2013); IEEE Systems Journal (2013); Journal of Parallel and Distributed Computing (2014); Frontiers of Information Technology & Electronic Engineering (2016); International Journal of Computing Science and Mathematics(2016); International Journal of Big Data Intelligence(2016)
- Conference Reviewer: IEEE/ACM CCGrid (2012); IEEE CloudNet (2014-2016); IEEE eScience (2014)
- Co-founder of PRAGMA Students Steering Committee, PRAGMA Students, 2012-2014

Presentations and Invited Talks (excludes conference paper talks)

[1] Cross-Institute Virtual Cluster Management in PRAGMA, *PRAGMA 27 Workshop,* Bloomington, Indiana, USA, Oct 15-17, 2014
[2] User-level controllability of virtual clusters using HTCondor, *PRAGMA 26 Workshop,* Tainan, Taiwan, April 9-11, 2014
[3] Hierarchical MapReduce: Towards Simplified Cross-Domain Data Processing, Invited talk at Jilin University, China, October 22, 2013
[4] A Hierarchical MapReduce Framework, Invited talk at IBM Student Workshop for Frontiers of Cloud Computing 2012, IBM's Thomas J. Watson Research Center in Hawthorne, New York, July 30-31, 2012
[5] Hierarchical MapReduce, *Invited talk at Cloud Computing Lecture*, Indiana University, Oct 12, 2011.
[6] Opal-Sigiri: Software as a Service on PRAGMA Testbed, *PRAGMA 20 Workshop*, Hong Kong, China, March 2-4. 2011.
[7] Metascheduling using the Community Scheduler Framework (CSF4), *NBCR Summer Institute 2009*, UCSD, 2009.
[8] My WorkSphere: Integrated and transparent access to Gfarm computational data grid through GridSphere portal with meta-scheduler CSF4, *Invited talk at NBCR Special Seminar*, UCSD, Aug 28th 2006.
[9] Cluster and Grid Computing: Transparent access and workflow management, *NBCR Summer Institute 2006*, UCSD, 2006.

Selected Publications

PhD Dissertation

Yuan Luo, Virtual Cluster Management for Analysis of Geographically Distributed and Immovable Data, [Bloomington, Ind.] Indiana University, 2015-08. Doctoral Committee: Beth Plale (chairperson), Geoffrey Fox, Judy Qiu, Yuqing Wu, and Philip Papadopoulos.

Journal Papers

[1] **Luo, Y.**, et al. (2014) Hierarchical MapReduce: towards simplified cross-domain data processing. Concurrency Computat.: Pract. Exper., 26: 878–893. doi: 10.1002/cpe.2929
[2] Jensen, S.; Plale, B.; Aktas, M.S.; **Yuan Luo**; Peng Chen; Conover, H., "Provenance Capture and Use in a Satellite Data Processing Pipeline," Geoscience and Remote Sensing, IEEE Transactions on , vol.51, no.11, pp.5090,5097, Nov. 2013, doi: 10.1109/TGRS.2013.2266929
[3] Hongliang Li, Xiaohui Wei, Qingwu Fu, **Yuan Luo**. (2013) MapReduce Delay Scheduling with Deadline Constraint, *Concurrency and Computation: Practice and Experience*, doi: 10.1002/cpe.3050.
[4] Ding, Z., Wei, X., **Luo, Y.**, et al. Customized Plug-in Modules in Meta-scheduler CSF4 for Life Sciences Applications, *New Generation Computing*, Vol.25 No.4 2007.
[5] Ding, Z., Wei, X., **Luo, Y.**, et al. A Virtual Job Model to Support Cross-Domain Synchronized Resource Allocation (In Chinese with English Abstract), *Journal of Jilin University (Science Edition)*, Vol. 46, No.2, Mar 26, 2008.

Conference and Workshop Papers

[1] Peng Chen, Beth Plale, You-Wei Cheah, Devarshi Ghoshal, Scott Jensen, and **Yuan Luo**. Visualization of Network Data Provenance, Workshop on Massive Data Analytics on Scalable Systems, co-located with HiPC, Pune, India, Dec. 18-21, 2012.
[2] Plale, B., Withana, E. C., Herath, C., Chandrasekar, K., **Luo, Y.** Effectiveness of Hybrid Workflow Systems for Computational Science, *International Conference on Computational Science (ICCS)*, Omaha, Nebraska, Jun 4-6, 2012.
[3] **Luo, Y.** and Plale, B. Hierarchical MapReduce Programming Model and Scheduling Algorithms, *Doctoral Symposium of the 14th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid)*, Ottawa, Canada, May 13-16, 2012.
[4] **Luo, Y.**, Plale, B., et al. 2011. A Hierarchical Framework for Cross-Domain MapReduce Execution. In *Proceedings of the Second International Workshop on Emerging Computational Methods for the Life Sciences* (ECMLS '11). ACM, New York, NY, USA, 15-22.
[5] Wei, X., **Luo, Y.**, Gao, J., et al. The Session Based Fault Tolerance Algorithm of Platform EGO Web Service Gateway*, Proceedings of International Symposium on Grid Computing (ISGC2007)*, Academia Sinica, Taipei, Taiwan, March 26-29, 2007.
[6] Ding, Z., **Luo, Y.**, Wei, X., et al. My WorkSphere: Integrative Work Environment for Grid-unaware Biomedical Researchers and Applications, *2nd International Workshop on Grid Computing Environments (GCE06) at SC06*, Tampa, FL. Nov. 12-13, 2006.