Yuan Luo, Ph.D.

yuan@yuanluo.net | yuanluo.net

Curriculum Vitae

Education

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

Experience

Data Engineering, Facebook, Inc. — *07/2017 - present*

- Drive data engineering and analytics effort in the following areas:
  - Early stage products in emerging markets;
  - Search engine optimization.

Data Architect, Edmodo, Inc. — *09/2015 – 07/2017*

- Product, User Acquisition and Engagement:
  - Initiated and led "data to action" project and built a system (Datacow) that consumes massive data and triggers actions such as email. The system has been widely used in Edmodo, from user onboarding and content digests, to user reactivation and marketing. Datacow is the central engine to trigger all the non-transactional emails and push notifications and has contributed greatly to the growth of active users.
  - Initiated and led the development of Edmodo's social recommendation system (PYMK), to increase user acquisition and engagement: a) user acquisition: recommend users to invite people from their contacts list, and b) user engagement: recommend users to connect with peers on Edmodo. Data shows that PYMK achieved the highest user acquisition rate among all viral invite loops.
- Data Engineering:
  - Completed redesigned the entire data platform, including architected and led the development of data pipeline system, ETL process, and hybrid data stores. The new data platform (Redmodo), eases and expedites the consumption of 80+ million user data for internal use (eg, Datacow) and for various product (eg, PYMK), while significantly increased the reliability and reduced the maintenance cost.
- Push Notification and SMS Platform:
  - Enhanced the platform to handle recipients and contents preparation and delivery with personalized user experience for millions of users on regular basis (tens of millions of deliveries per day).

Co-founder, Stealth startup (now FitTime) — *01/2014 – 01/2015*

- Lost 40 lbs of fat in a 4-month period in early 2013. Shared fitness experience with hundreds of thousands of followers on weibo.com. My body transformation story and diet recipes have received hundreds of millions of cumulative views on Internet and impacted millions of people to keep fit and stay healthy. Co-founded in late 2013 a stealth startup which later became the well-known FitTime 即刻运动, helped to grow user base from 0 to 2M+, architect web platform, launch mobile apps, define and implement product (eg., social network, fitness programs via videos), marketing (eg., social media) and monetization (eg., premium program subscription, fitness store) strategies. Left the startup to focus on PhD in early 2015 while the other co-founder went back to China and has grown the business into a 10M+ user startup (http://fit-time.com, 2017)

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 top research universities, and cited by at least 7 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.
- 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).

Researcher, 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.

Visiting Scholar, University of California, San Diego *07/2006 - 09/2006*

- Contributed to MyWorksphere project, which is a SOA environment served as one of the NSF funded TeraGrid science gateway, allowing existing applications to run transparently on the Grid.

Researcher, Platform Computing (IBM) Grid Technology Research Center at Jilin University *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.

Selected Technical Skill

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

Teaching Experience

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

- 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 - 6th IEEE International Conference on Cloud Networking (2014 - 2018)
- 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

Google Scholar shows that the publications were cited over 180 times. Especially, [10] was cited by 10+ patents from IBM and EMC, and [5] was cited by 2 patents from IBM.

PhD Thesis

[1] **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

[2] **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
[3] 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
[4] 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.
[5] 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.
[6] 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

[7] 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.
[8] 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.
[9] **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.
[10] **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.
[11] 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.
[12] 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.

Technical Reports

[13] B. Plale, E. Withana, C. Herath, K. Chandrasekar, **Y. Luo**, F. Terkhorn, Strengths and weaknesses of sub-workflow interoperability, *Indiana Unversity Computer Science Tech. Rep. TR700*

Dataset

[14] **Y Luo**, B Plale, S Jensen, YW Cheah, H Conover Provenance of AMSR-E Data from the National Snow and Ice Data Center (NSIDC), - OPM XML Ver, 2012. dx.doi.org/10.5967/M0F47M2D

Posters

[15] A Personal Cloud Controller, *PRAGMA 26 Workshop*, Tainan, Taiwan, April 9-11, 2014
[16] Network Transfer over Pacific Rim on PRAGMA Cloud: Performance and Tuning, *PRAGMA 25 Workshop*, Chinese Academy of Science, Beijing, China, Oct. 16-18, 2013
[17] Vortex2 Metadata Management on PRAGMA Cloud: A GeoPortal Experience. *PRAGMA 22 Workshop*, Monash University, Melbourne, Australia, April 17-19, 2012.
[18] A Hierarchical MapReduce Framework, *PRAGMA 22 Workshop*, Monash University, Melbourne, Australia, April 17-19, 2012.
[19] Improving Twister Messaging System Using Apache Avro, *The 2nd IEEE International Conference on Cloud Computing Technology and Science (CloudCom 2010)*, Indianapolis, USA, Nov 30 - Dec 3. 2010.
[20] Karma: Provenance Aggregation Across Layers of GENI Experimental Networks, *PRAGMA 19 Workshop*, Changchun, China, Sept 13-15. 2010.
[21] GDIA: A Scalable Grid Infrastructure for Data Intensive Applications, *National Biomedical Computation Resource Summer Institute 06*, San Diego, Aug. 2006.
[22] My WorkSphere: Integrated and Transparent Access to Gfarm Computational Data Grid through GridSphere Portal with Metascheduler CSF4, *3rd International Life Sciences Grid Workshop*, Yokohama, Japan, 2006; Yokohama, Japan, 2006.

❖ **Short version (RESUME) is also available.**