Yuan Luo, Ph.D.

yuan@yuanluo.net | *yuanluo.net*

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

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 Architect, Edmodo, Inc.
09/2015 - present

- 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 has been reached out to 20+ million users within a few months after launch and contributed greatly to the growth of (weekly) active users. Datacow is becoming the central engine to trigger all the non-transactional emails and push notifications.
  - 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. Tests show 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 workflow management system, ETL process, and hybrid data stores. The new data platform (Redmodo), eases and expedites the consumption of 70 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

- Co-founded a stealth startup which later became the well-known FitTime 即刻运动, helped the startup grow user base, architect web platform, launch mobile apps, define and implement product and marketing strategies. Later left the startup to finish PhD.

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 (now 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.

❖ **Excludes publications (140+ citations), invited talks, and scholarly services. See Curriculum Vitae for more information.**