Shuo Yang

952-297-6289 | imsure95@gmail.com | imsure.github.io | imsure | shuo-yang-12ab7047

Summary

8 years of programming experience with C/C++, Python and Java, specializing in system programming and backend/infrastructure. Proficient in Relational Database Management Systems, especially MySQL. Experienced with Big Data technology such as Hadoop and Hive. Good exposure to NoSQL databases such as MongoDB and HBase. Understanding of fundamentals in data science.

Education

University of Arizona — *Tucson, Arizona USA*

M.S. in Computer Science, GPA: 3.18 — *Aug. 2014 - May. 2017*

University of St. Thomas — *St. Paul, Minnesota USA*

M.S. in Software Engineering, GPA: 3.8 — *Aug. 2011 - May. 2014*

Harbin Institute of Technology — *Harbin, Heilongjiang China*

B.E. in Electrical Engineering — *Aug. 2004 - May. 2008*

Academic Experience

University of Arizona — *Tucson, Arizona USA*

Research Assistant with Dr. Richard Snodgrass on ANTARES project — *Aug. 2014 - Aug. 2015, Aug. 2016 - Mar. 2017*

- Designed and implemented the Python API for astronomical alert data manipulation, using MySQL as DB backend.
- Worked on various components of ANTARES system, including data injection, alert simulator and alert packet format specification.
- Managed the dedicated CentOS cluster; designed and built the autoconfiguration and bootstrap system using Puppet and Vagrant.
- Analyzed and improved the performance of ANTARES data processing pipeline.
- Researched data provenance and designed provenance framework for ANTARES to answer provenance questions that astronomers would ask.

University of Arizona — *Tucson, Arizona USA*

Independent Study with Dr. Beichuan Zhang on Named Data Networking (NDN) project — *Spring 2016 & Fall 2016*

- Researched consumer-driven congestion control mechanisms in NDN.
- Implemented TCP-like congestion control algorithms (RENO, CUBIC and VEGAS) in NDN consumers (in C++) to lay the groundwork for comparison with more advanced congestion control algorithms for NDN.

University of Arizona — *Tucson, Arizona USA*

Teaching Assistant with Dr. Lester McCann on CS460: Database Systems — *Fall 2015 & Spring 2016*

- Held office hours; graded programming and written assignments; prepared solutions for written assignments.
- Designed the final project "Database-driven Web Application" using Oracle, Tomcat and JSP; evaluated students' design and implementation.

University of St. Thomas — *St. Paul, Minnesota USA*

Student Researcher with Dr. Brad Rubin and Dr. Jadin Jackson on project: Neural Modeling in Hadoop — *Aug. 2013 - May. 2014*

- Researched graph processing in Hadoop.
- Implemented a large scale basal ganglia neural network model in MapReduce.
- Implemented the model with Apache Giraph, improved performance by 60% compared to the MapReduce implementation.

Work Experience

Danfoss Power Solutions — *Plymouth, Minnesota USA*

Software Engineer Intern — *June. 2012 - May. 2013*

- Migrated the legacy software (written in C) to the new hardware platform with other team members and conducted unit testing.
- Improved overall software quality using static code analyzer FlexeLint, identified and corrected several vulnerabilities existed in the legacy code.
- Built a tool (in Python) for summarizing, indexing and querying large volume of warning messages produced by FlexeLint.

Beijing Farsight Technology and Information — *Beijing, China*

Embedded Software Engineer & Training Assistant — *July. 2009 - July. 2011*

- Developed Linux device drivers and applications (in C) for various ARM platforms (ARM 9, 11 & Cortex-A8).
- Assisted training instructors in preparing training content and developed technical how-to documents for ARM and Embedded Linux training.
- Instructed trainees in their lab sessions and final design project.

Platomix Technologies. (Start-up company) *Beijing, China*

Software Engineer *Feb. 2009 - July. 2009*

- Responsible for the development on RTL device, which was part of the initial prototype for Samsung's Remote Test Lab (RTL).
- Implemented the communication protocol (in C++) between RTL device (LiMo platform) and RTL proxy.
- Developed a multi-threaded daemon process (in C++) for RTL device to handle requests from proxy, in order for RTL client to remotely control the device.

Publication

Yang S, Spielman ND, Jackson JC, Rubin BS. Large-scale neural modeling in MapReduce and Giraph. In IEEE International Conference on Electro Information Technology. IEEE Computer Society. 2014. p. 556-561. 6871824. Available from, DOI: 10.1109/EIT.2014.6871824

Selected Coursework

University of Arizona

Database Systems Implementation (CS560) - Minibase *Spring 2016*

- Worked on a team of two, we developed core components for Minibase to make it function correctly.
- Implemented Heap File, Buffer Manager and B^+ Tree in C++.

Principles of Computer Networking (CS525) - Virtual Router *Fall 2015*

- Implemented a virtual router in C that runs PWOSPF routing protocol and routes real IP packets.

Principles of Compilation (CS553) - C- - Compiler *Spring 2015*

- Implemented code generator for C- - (a subset of C) compiler in C that converts abstract syntax tree (AST) into MIPS assembly code.
- Implemented various O1, O2 and O3 optimizations.

Parallel and Distributed Programming (CS522) - MPI Profiler & Redundant MPI *Fall 2014*

- Implemented a profiler (in C) for MPI that generates a critical-path profile for MPI programs.
- Implemented the redundant MPI protocol (in C) by which an MPI program runs redundantly such that if one MPI rank fails, the program will continue with the replica rank.

University of St. Thomas

Big Data Architecture (SEIS736) - Design patterns for graph algorithms in MapReduce *Fall 2013*

- Researched design patterns for graph algorithms in MapReduce.
- Developed an improved design pattern to avoid writing entire graph structure onto disk when the graph structure remains unchanged during processing.

Data Visualization (SEIS785) - Who's going to win 2013 NBA Championship ? *Spring 2013*

- Developed ETL process to pull NBA datasets from multiple sources into a data warehouse using Python and SQLite.
- Visualized and analyzed data to explore what are the key factors that make a championship team using R.

Information Retrieval (SEIS731) - Shakespeare Search Engine *Fall 2012*

- Built a search engine from scratch in Python for Shakespeare's whole collection, including a web crawler, an indexer and a query processor.
- The indexer supports both inverted index and positional index and the query processor supports both free text queries and phrase queries.

Skills

Languages	C/C++, Python, Java, SQL, Shell, HTML, R, MIPS Assembly
Database Systems	MySQL, SQLite, Oracle, MongoDB, HBase
Operating Systems	Linux/Unix, Mac OSX
Frameworks	Hadoop, Hive, Giraph, Pthreads, MPI, Django, Numpy, Pandas
Tools	LaTeX, Emacs, Eclipse, Git, Vagrant, Puppet, Chef, Jupyter