

# 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. Eager to learn new technologies and tools to grow professionally.

## EDUCATION

### University of Arizona

M.S. IN COMPUTER SCIENCE, GPA: 3.18

Tucson, Arizona USA

Aug. 2014 - May. 2017

### University of St. Thomas

M.S. IN SOFTWARE ENGINEERING, GPA: 3.8

St. Paul, Minnesota USA

Aug. 2011 - May. 2014

### Harbin Institute of Technology

B.E. IN ELECTRICAL ENGINEERING

Harbin, Heilongjiang China

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.

### 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; proposed an improved graph algorithm design pattern in MapReduce.
- 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 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

<b>Database Systems Implementation</b>	Implemented Heap File, Buffer Manager and B <sup>+</sup> Tree components of Minibase in C++.
<b>Principles of Computer Networking</b>	Implemented a virtual router in C that runs PWOSPF routing protocol and routes real IP packets.
<b>Principles of Compilation</b>	Implemented C- - (a subset of C) compiler (both code generation and optimizations) in C.
<b>Computer Security</b>	Implemented a secure chat program using a hybrid protocol (DES + RSA) in Java.
<b>Parallel and Distributed Programming</b>	Implemented a MPI critical-path profiler and Redundant MPI protocols in C.
<b>Introduction to Machine Learning</b>	Implemented a sparse autoencoder in Python, trained with MNIST handwritten digit dataset.

### UNIVERSITY OF ST. THOMAS

<b>Big Data Architecture</b>	Developed a new design pattern for improving the performance of graph processing in MapReduce.
<b>Data Visualization</b>	Explored what are the key factors that make a championship team with NBA dataset using Python, SQLite and R.
<b>Information Retrieval</b>	Built a search engine in Python (web crawler + indexer + query processor) for Shakespeare's whole collection.
<b>Software Engineering</b>	Built a genetic programming system in C for automatically solving linear regression problem.

## SKILLS

---

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