# Lizhen Shi

1128 Ocala Rd, APT G1
Tallahassee, FL 32304

(205) 542-8124
ls15x@my.fsu.edu

## I. INTERESTS

Big Data, Cloud Computing, Biological Computing, Machine Learning, Deep Learning

## II. QUALIFICATION

- In-depth understanding and rich hands-on experience in Hadoop, Spark, Hive, Pig, and Tez frameworks.
- Strong parallel programming experience in MPI, Spark, Hadoop MapReduce, Pig Latin Data Flow, and HiveQL Language.
- Proficient in shell, SQL programming
- Hands-on experience in big data analysis, large scale biological data processing.
- 6 years working experience as a software engineer in multinational project teams.

## III. EDUCATION

- 08/2015 ~ present PhD Candidate Computer Science GPA 4.0 Florida State University, FL
- 08/2014 ~ 08/2015 M.S. Computer Science GPA 4.0 Auburn University, AL
- 08/2004 ~ 06/2007 B.S. Computer Science North China Electric Power University, China

## IV. EXPERIENCE

- Teaching Assistant 08/2015 ~ present Florida State University
- Summer Intern 05/2015 ~ 08/2015 Lawrence Berkeley National Laboratory
- Research Assistant 08/2014 ~ 05/2015 Auburn University
  Thesis: Feature Enhancement and Performance Evaluation of BioPig Analytics

## V. PROJECTS

***At Florida State University:***

- Designed a hybrid clustering algorithm based on LPA for metagenome read clustering
- Developed a scalable sequence clustering tool named SpaRC based on Apache Spark.
- Designed a cross-layer scheduler for improving DAG-structured query processing in MapReduce.

***At Lawrence Berkeley National Lab:***

- Upgraded BioPig toolkit from Hadoop 1 to Hadoop 2 (2.7.0).
- Tuned Hadoop parameters for BioPig performance optimization.
- Implemented K-mer Similarity to extended BioPig toolkit functionality

***At Auburn University:***

- Implemented an MPI program to sort millions of integers using multiple algorithms.
- Conducted research on two big data frameworks: Hadoop and Spark.
- Developed an android project that can support downloading files from top cloud storage providers.

***At PwC, CleNET and Kingbase:***

- Developed/Maintained projects in various domains using C#, Java languages and SQL server databases.

## VI. PUBLICATIONS

- **Lizhen Shi**, Zhong Wang, Weikuan Yu, Xiandong Meng. *Performance Evaluation and Tuning of BioPig for Genomic Analysis*. International Workshop on Data-Intensive Scalable Computing Systems (DISCS) in conjunction with the ACM/IEEE Supercomputing Conference (SC'15), Austin, TX, November 2015.
- **Lizhen Shi**, Zhong Wang, Weikuan Yu, Xiandong Meng. *A Case Study of Tuning MapReduce for Efficient Bioinformatics in the Cloud.* Journal of Parallel Computing (ParCo'16).
- Ji Huang, Stefania Vendramin, **Lizhen Shi**, Karen M McGinnis. *Construction and optimization of a large gene Co-expression network in maize using RNA-Seq data*. Plant physiology (2017)
- **Lizhen Shi**, Xiandong Meng, Elizabeth Tseng, Michael Mascagni, Zhong Wang. *SpaRC: scalable sequence clustering using Apache Spark.* Bioinformatics.

- **Lizhen Shi**, Volkan Sevim, Michael Mascagni, and Zhong Wang. *Leveraging long-read sequencing for cost-effective metagenome clustering*. (To be submitted)
- **Lizhen Shi,** Bo Chen. *A Vector Representation of DNA Sequences Using Locality Sensitive Hashing.* (bioRxiv)