Zhao Zhang

Room 3.226, Advanced Computing Building
10100 Burnet Road
Austin, Texas 78758
zzhang@tacc.utexas.edu

CURRENT POSITION

Texas Advanced Computing Center July 2016 - present
Research Associate, Data Intensive Computing Group

EDUCATION

University of Chicago September 2009 - June 2014
Ph.D. Computer Science
Advisor: Ian T. Foster

University of Chicago January 2007 - December 2007
M.S. Computer Science

Beijing University of Posts and Telecommunications August 2002 - June 2006
B.E. Software Engineering

EXPERIENCE

Postdoc Researcher and Data Science Fellow July 2014 - June 2016
AMPlab and Berkeley Institute for Data Science Berkeley, CA

- Machine Learning Pipeline Diagnostic System
  Design a fine-grained lineage capturing and serving system to enable interactive machine learning pipeline diagnostics. Integrate the diagnostic system with Key- stoneML, a distributed machine learning framework built on top of Apache Spark.
- Kira: Flexible Astronomy Image Processing in Clouds
  Build a programmable astronomy image processing toolkit with Apache Spark. Evaluate the application performance tradeoffs with various software and hardware configurations in both batch and streaming mode.

Graduate Student Researcher September 2009 - June 2014
Department of Computer Science, University of Chicago Chicago, IL

- AIMES Application Skeleton
  Build a distributed application skeleton framework, which generates pseudo distributed applications that mimic the computational and I/O behavior of real applications that are parallelized with Bash, MPI, Pegasus or Swift.
- ExM: System support for extreme-scale, many-task applications
  Profile the I/O and data flow behavior of a set of many-task applications. Propose a file system benchmark as the capability envelope of a supercomputer to support many-task applications and quantitatively evaluate the performance bottlenecks of existing parallel file systems. Design a distributed in-memory file system that leverages temporal locality in RAM and collective data movements to speedup many-task applications 10-100x faster compared to the peripheral parallel file system on Blue Gene/P supercomputer. Propose a concise programming interface by extending Bash scripts to enable parallel scripting on

supercomputers.

Scientific Application Specialist January 2008 - July 2009
Computation Institute, University of Chicago Chicago, IL

- Falkon: A Fast and Light-weight Task Execution Framework
  Build multi-layer task scheduler that scales up to O(100, 000) CPU cores on supercomputers.

Senior Technical Associate July 2006 - December 2006
Lucent Technology Beijing, China

- Super Distributed Home Location Register
  Develop and test customized distributed database migration products.

GitHub https://github.com/zhaozhang

SELECTED PUBLICATIONS

- Y. You, **Z. Zhang**, J. Demmel, K. Keutzer, C. Hsieh.
  ImageNet Training in Minutes
  CoRR, abs/1709.05011 (2017)
- **Z. Zhang**, W. Xu, N. Gaffney, D. Stanzione.
  Early Results of Deep Learning on the Stampede2 Supercomputer
  The Intel Xeon Phi Users Group, 2017, DOI: 10.13140/RG.2.2.36806.78404
- **Z. Zhang**, E. Sparks, M. J. Franklin.
  Diagnosing Machine Learning Pipelines with Fine-grained Lineage
  In Proceedings of the 26th international symposium on High-performance parallel and distributed computing (HPDC' 17)
- **Z. Zhang**, K. Barbary, F. A. Nothaft, E. Sparks, O. Zahn, M. J. Franklin, D. A. Patterson, S. Perlmutter.
  Kira: Processing Astronomy Imagery Using Big Data Technology.
  To appear, IEEE Transactions on Big Data, Volume: 3. 2016
- D. S. Katz, A. Merzky, **Z. Zhang**, S. Jha.
  Application Skeletons: Construction and Use in eScience.
  In Press, Corrected Proof, Future Generation Computer Systems (FGCS), 2015.
- D. Crankshaw, P. Bailis, J. E. Gonzalez, H. Li, **Z. Zhang**, M. J. Franklin, A. Ghodsi, M. I. Jordan.
  The Missing Piece in Complex Analytics: Low Latency, Scalable Model Management and Serving with Velox.
  7th Biennial Conference on Innovative Data Systems Research (CIDR), 2015.
- F. A. Nothaft, M. Massie, T. Danford, **Z.Zhang**, U. Laserson, C. Yeksigian, J.Kottalam et al.
  Rethinking data-intensive science using scalable analytics systems.
  In Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, pp. 631-646. ACM, 2015.
- **Z. Zhang**, D. S. Katz.
  Using Application Skeletons to Improve eScience Infrastructure.
  In e-Science (e-Science), 2014 IEEE 10th International Conference on, vol. 1, pp. 111-118. IEEE, 2014.
- **Z. Zhang**, D. S. Katz, T. Armstrong, J. Wozniak, I. Foster.
  Parallelizing the Execution of Sequential Scripts.

In Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis (Supercomputing), p. 31. ACM, 2013.

- **Z. Zhang**, D. S. Katz, M. Wilde, J. Wozniak, I. Foster.
  MTC Envelope: Defining the Capability of Large Scale Computers in the Context of Parallel Scripting Applications.
  In Proceedings of the 22nd international symposium on High-performance parallel and distributed computing (HPDC), pp. 37-48. ACM, 2013.
- **Z. Zhang**, D. S. Katz, J. Wozniak, A. Espinosa, I. Foster.
  Design and Analysis of Data Management in Scalable Parallel Scripting.
  In Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis (Supercomputing), p. 85. IEEE Computer Society Press, 2012.
- I. Raicu, **Z. Zhang**, M. Wilde, I. Foster, P. Beckman, K. Iskra, B. Clifford.
  Toward Loosely Coupled Programming on Petascale Systems.
  In Proceedings of the 2008 ACM/IEEE conference on Supercomputing, p. 22. IEEE Press, 2008.

ACADEMIC TALKS

- Scientific Computing Meets Big Data Technology: An Astronomy Use Case, Big Data, October 2015, San Jose, CA
- Flexible Astronomy Image Processing in Clouds, SLAC National Accelerator Lab- oratory, Stanford University, July 2015, Palo Alto, CA
- Kira: Flexible Astronomy Image Processing in Clouds, AMP Retreat, May 2015, Santa Cruz, CA
- Parallelizing the Execution of Sequential Scripts, Supercomputing13, November 2013, Denver, CO
- Enabling Parallel Scripting on Large Scale Computers, AMPLab, UC Berkeley, August 2013, Berkeley, CA
- MTC Envelope: Defining the Capability of Large Scale Computers in the Context of Parallel Scripting Applications, HPDC13, June 2013, New York, NY
- Parallel Programming on Clouds, Grids, and Supercomputers, Center for Earth System Science, Tsinghua University, December 2012, Beijing, China
- Design and Analysis of Data Management in Scalable Parallel Scripting, Supercomputing12, November 2012, Salt Lake City, UT

PROFESSIONAL SERVICE

- **Technical Program Committee Member**, International Symposium on Big Data Computing, 2014
- **Guest Co-editor**, Future Generation Computer Systems Special Issue on eScience Applications and Infrastructure, 2014.
- **Organizer**, Weekly System Research Seminar, UChicago Systems Group, 2010-2014.
- **Publicity Chair**, 5th Workshop on Many-Task Computing on Grids and Super- computers (MTAGS) 2012, November 2012, Salt Lake City, UT.
- **Proceedings Chair**, IEEE International Conference on eScience, October 2012, Chicago, IL.
- **Organizer**, 1st Greater Chicago Area System Research Workshop, May 2012, Chicago, IL.

REFERENCE

- **Niall Gaffney**
  Director of Data Intensive Computing, Texas Advanced Computing Center
  Email: ngaffney@tacc.utexas.edu

- **Michael J. Frankin**
  Chair, Department of Computer Science, UChicago
  Email: mjfranklin@cs.uchicago.edu

- **Saul Perlmutter**
  Director, Berkeley Institute for Data Science, UC Berkeley
  Franklin W. and Karen Weber Dabby Professor of Physics, UC Berkeley
  Email: saul@lbl.gov

- **Ian T. Foster**
  Director, Computation Institute, UChicago & ANL
  Arthur Holly Compton Distinguished Service Professor of Computer Science, UChicago
  Email: foster@anl.gov

- **Daniel S. Katz**
  Assistant Director, Scientific Software and Applications, National Center for Supercomputing Applications (NCSA)
  Email: d.katz@ieee.org