Jianfeng Jia

jianfeng.jia@gmail.com (949)678-9893
Seeking a Full-time Big Data Software Engineer Position

EXPERIENCE

Ph.D. Student UC Irvine Sept.2012 - Dec.2017

Big Data Management for Interactive Analytics.
Creator of UCI Cloudberry, committer of Apache AsterixDB, Hyracks, and Pregelix.

Software Engineer Intern Sumo Logic, Redwood City,US Jun.2015 - Sept.2015

Designed and implemented the lookup operator for a scalable streaming platform.

Software Engineer Intern SRCH2.com, Irvine, US Jun.2014 - Sept.2014

Led the Android development group.

Research Engineer Sogou.com, Beijing, China Jul.2008 - Jul.2012

Led a research group of 5 engineers to improve the precision of the Sogou Chinese Input Method product which was used by 300 million people.

EDUCATION

Ph.D. Candidate University of California, Irvine Sept.2012 - Present

Research topic:*Big data management, Large scale data analytics and visualization*

M.S. Computer Science Xiamen University, China Sept.2005 - Jul.2008

B.S. Computer Science Xiamen University, China Sept.2001 - Jul.2005

SKILLS

Scala, Java, C++, Android, Python, Hadoop, Spark

SELECTED PROJECTS

Cloudberry System, http://cloudberry.ics.uci.edu UCI

Build a middleware system on top of a parallel database that supports efficient interactive analytics and visualization on billions of records.

Lookup Operator for Large-Scale Streaming data Sumo Logic

Developed a lookup operator that can join the large-scale fast streaming data with the static data in databases.

Big-Object-Aware Memory Manager in Apache AsterixDB Apache AsterixDB

Designed a big-object-aware memory manager to support the Big-Object feature in AsterixDB.

Android SDK for SRCH2 C++ library SRCH2.com

Developed the local and the server version of the Android Search SDK which used the SRCH2 C++ search engine library under the hood. Implemented an error-tolerant search app using this SDK.

Big Graph Computing System Evaluation UCI

Built a benchmark to evaluate the performance of popular Graph Computing Systems(Pregelix, GraphX, Giraph, Graph Lab) .

Genome Assembling using Hyracks UCI

Used the Hyracks platform to build the genome graph of billions of nodes. Achieved 30% performance improvement comparing to Hadoop solutions.

Dashboard Data Flow System using HBase and Pig Sogou.com

Built an data flow processing reporting system that can generate the key metrics of the product every day. The project used HBase to store 15T data (30G per day) and used Pig scripts to analyze and explore the global user behavior to the online dashboard.

Large-scale Language Model(LM) for Cloud IME Sogou.com

Built an automatic module for the LM building process updated weekly from the 500G corpus. The system was built on top of Hadoop platform.

Automatic New Word Detection Sogou.com

Developed a New Word Detection system which was implemented on top of Hadoop platform.

Dependency Treelet Based Chinese-to-English SMT System Xiamen University

Implemented a dependency grammar structure based statistical machine translation system.

HONORS AND AWARDS

Google Graduate Student Award in ICS, UC Irvine, 2017
Best Data Visualization Award in Data in UCI Data Science Hackathon, UC Irvine, 2016
Rhinoceroses prize for improving 20% precision rate in the IME product, Sogou Research, 2009

PUBLICATIONS

Caching Geospatial Objects in Web Browsers, T.Kim, V.Thirumaraiselvan, **J.Jia**, C.Li, ACM SIGSPATIAL 2017.
Use of Twitter Data to Predict Zika Virus Cases in the United States during the 2016 Epidemic, S.Masri, **J.Jia**, C.Li, G.Zhou, M.Lee, G.Yan, J.Wu. PlosONE 2017.
Twitter Coverage of Climate Change and Health before and after the 2016 US Presidential Election, S.Hopfer, M,Runnerstrom, **J.Jia**, T.Kim, C.Li. APHA 2017.
Towards Interactive Analytics and Visualization on One Billion Tweets, **J.Jia**, C.Li, X.Zhang, C.Li, M.Carey and S.Su, ACM SIGSPATIAL 2016.
Pregelix: Big(ger) Graph Analytics on A Dataflow Engine, Y.Bu, V.Borkar, **J.Jia**, M.Carey, T.Condie, VLDB 2015.
Dependency-Based Chinese-English Statistical Machine Translation, X.Shi, Y.Chen, **J.Jia**, CICLing 2007.