Chunqiu Zeng

School of Computing and Information Sciences.
Florida International University.
Miami, FL 33199, USA .

TEL: 786-527-0301
EMAIL: grandzeng@gmail.com
HOMEPAGE: http://users.cs.fiu.edu/~czeng001

RESEARCH INTERESTS

Large Scale Data Mining; System Management; Machine Learning;

EDUCATION

2011.12 - Present Ph.D student in the school of computing and information sciences,Florida International University, Miami, Florida 33199, USA. Advisor: Tao Li

2006-2009 Master degree in the school of computer science, Sichuan University, Chengdu, Sichuan, China. Advisor: Changjie Tang

2002-2006 Bachelor degree in the school of computer science, Sichuan University, Chengdu, Sichuan, China

RESEARCH PROJECTS

2011.12-present **Knowledge Discovery Research Group, Florida International University, Miami, Florida, USA (**Research Assistant**)**

- **LSDA(Large Scale Data Analysis)**. This project aims at building a large scale data analysis system with the following functionalities: (1) Integrate complex data mining algorithms and exchange intermediate results among sub tasks (rather than simple data processing in Hadoop); (2) Monitor the system resource consumption in real time based on the trace data; and (3) Balance the workload of nodes in a cluster based on monitored utilization data.
- **System Resource Monitoring and Mining**. This project aims at building an integrated framework to monitor the resource utilization of nodes in a cluster. Besides the monitoring, the framework also supports real time queries over the continuously monitored log streams and provides mining modules to discover useful patterns for problem determination.

INDUSTRIAL PROJECTS

2009.7-2011.12 **Alibaba Company(the largest e-commerce company in China), Hangzhou, Zhejiang, China (**Data Warehouse Engineer**)**

- **Metadata Mining System** is to apply data mining algorithms to discover underlying relationships among a large number of tables in data warehouse to optimize data models. **Main contribution:** extract metadata from job logs based on SQL parsing for DB2, ORACLE, Greenplum, BIEE, MSTR, etc.; and develop association rule algorithms to mine metadata.
- **Job Scheduling System** is to schedule more than 10,000 ETL jobs each day efficiently on multiple system resources such as DB2, RAC, Greenplum, Hadoop/Hive, etc. The system has high resource utilization and large job throughput. **Main contribution**: split all the resources into different groups; design and implement resource scheduling algorithm to achieve the balance among resource groups and rationally preempt resources among different groups, for comprehensively and efficiently utilizing all the available resources.
- **Alihive** is to access a hive server and records the log for each SQL statement. **Main contribution**: design and implement the API to submit job to the hive server. Using PERL.

- **ADS (Automatic Deploying System)** is to alleviate the manpower consumption in deploying source codes to production servers automatically (on the scale of hundreds of projects each day). **Main contribution**: architecture design and the whole project implementation using Python, SQL, DB2.
- **DHW (Data High Way)** is to transmit data with a high speed among heterogeneous data sources like DB2, RAC, Greenplum, HDFS, MySQL, SQLServer, etc. **Main contribution**: architecture design, implementation of the distributing framework for DHW based on thrift, design and implementation of the distributing balance algorithm for transmitting data on the cluster of DHW hosts , I/O module with HDFS using Python, JAVA.
- **Hotspot** is to discover hot data tables which are visited frequently according to the ETL job log and to analyze the hot data tables to optimize the data model for data warehouse. **Main contribution**: extract tables from SQL log based on SQL parsing; use an association rule algorithm, etc to mine the latent relationships among a large number of tables.

2007.8-2008.3 **Alcatel-Lucent Company, Chengdu, Sichuan, China (**MSGUI group as software engineer**)**

- **MAPIM** is to manage network elements such as Nodes and Links. **Main contribution**: according to the business logic of network management, complete data design including the conceptual data model, the logical data model, the physical data model; design and implement the DBI(database access interface), and construct simulative data for database by developing CORBA application using Name Service and Notification Service.
- **GUI Server** is to provide service of user graphic rendering. **Main contribution**: customize graphic rendering service in GUI Server by XML; develop the responding module for the request commands from GUI Render Client; implement CORBA communication module between GUI Server and MAPIM using JacORB, JAVA.

SELECTED PUBLICATION

1. **Chunqiu Zeng**, Tao Li. **Automatically Discover SQL Statement Templates From Query Logs to Match Procedures**.(in preparation), 2013.
2. **Chunqiu Zeng**, Tao Li. **Multiple Clustering Views Meet Uncertain Data.**(in preparation),2013.
3. Li Zheng, Chao Shen, Liang Tang, Chunqiu Zeng, Tao Li, Steve Luis, Shu-Ching Chen and Jainendra K. Navlakha. **Disaster SitRep - A Vertical Search Engine and Information Analysis Tool in Disaster Management Domain.** The 13th IEEE International Conference on Information Integration and Reuse (IRI), pp.457 - 465, 2012
4. **Chunqiu Zeng**, Jie Zuo, Chuan Li, Kaikuo Xu, Shengqiao Ni, Liang Tang, Yue Zhang, Shaojie Qiao. "MPSQAR: **Mining Quantitative Association Rules Preserving Semantics**", *in Proceedings of the International Conference on Advanced Data Mining and Applications* (**ADMA**), pp. 572 - 580, 2008
5. Liang Tang, Changjie Tang, Lei Duan, Chuan Li, Yexi Jiang, **Chunqiu Zeng**, and Jun Zhu. "**MovStream: An Efficient Algorithm for Monitoring Clusters Evolving in Data Streams**",*in Proceedings of the 2008 IEEE International Conference on Granular Computing* (GrC), pp. 582 - 587, 2008

AWARDS & HONORS

1. 12/2006, The 2nd National Open Source Software Competition (China), Award finalists.
2. 06/2006, Amway College Student's computer works competition, Second class award.
3. 2004, Second Prize in Sichuan Contest District for National Undergraduate Mathematical Contest in Modeling.

PROGRAMMING SKILLS

Developing Language: familiar with Java(JDBC), C/C++, Python, SQL;

Database: DB2,Oracle, MYSQL, Postgresql; Greenplum(**Distributed Database**);

Distributing System: Hadoop/Hive, familiar with Map/Reduce Framework