

Chunqiu Zeng

Software Engineer

Architecture Group of Data Platform Infrastructure

Department of Data Warehouse

Alibaba Company

TEL: (86)13429633471

EMAIL: grandzeng@gmail.com

HOME PAGE: <http://zengchunqiu.oratec.net>

RESEARCH INTERESTS

Data Mining, Knowledge Management, Machine Learning

EDUCATION

2006-2009 **Sichuan University, Chengdu, Sichuan, China**

Master in Computer Science, Advisor: Changjie Tang GPA:3.6 Rank: top 5% of 200 classmates

2002-2006 **Sichuan University, Chengdu, Sichuan, China**

Bachelor in Computer Science, GPA: 3.42 Rank: top 5% of about 300 classmates

RESEARCH EXPERIENCE

2010.5-Present **Data Warehouse department of Alibaba Company, Hangzhou, Zhejiang, China**

Software Engineer in architecture group of data platform infrastructure

- **Metadata Mining System**, responsible to apply data mining algorithm to discover underlying relationships among large number of tables in data warehouse, in order to optimize data model. In charge of extracting metadata from job log based on SQL parsing for ORACLE, Greenplum, BIEE, MSTR and etc; developing association rule algorithm to mine metadata.

2006.9-2009.7 **Database and Knowledge Engineering Institute in Sichuan University, Chengdu, Sichuan, China**

Research Assistant

- **National Birth Defects Supervision Project**, supported by the 11-th Five Years Key Programs for Science & Technology. Development of China under grant NO.2006BAI05A01. In charge of OLAP module providing ways for rolling up, drilling down, data slicing and data visualization related to OLAP operations.
- **Crime Miner: A criminal data mining system based on artificial intelligence**, supported by Foundation of Innovation Software Engineering for Young People in Sichuan. In charge of architecture design, implementation of system framework, layer of data access and classifying module based on criminal data.
- **National Earthquake Detection and Reporting System (China)**, directed by Seismology Bureau of Sichuan Province in 2007. In charge of seismic waveform user graphic interface module, instant event inspecting module, query cache module for waveform data from database and parameter configuration module.

WORK EXPERIENCE

2009.7-Present **Alibaba Company, Hangzhou, Zhejiang, China**

Data Warehouse Engineer

- **Job Scheduling System**, responsible to schedule more than 10,000 ETL jobs each day rationally and efficiently, on kinds of system resources like RAC, Greenplum, Hadoop/Hive and etc, in a way with high resource utilization and large job throughput. **Main contribution**: split all the resources into different groups; designed and implemented resource scheduling algorithm to strike the balance among resource groups and rationally preempt resource in different groups, intending to fully and efficiently utilize all the resources available. Using SQL, SHELL, and PYTHON developing language.

- **ADS (Automatic Deploying System)**, doing a great contribution to alleviate the manpower consuming in deploying source code with an automatic mode and can deploy hundreds of projects each day. **Main contribution:** architecture design, the whole project implementation including: Adsor main module, Crawler module, Putter module, Executor module and DB connection pool module. Using Python, SQL developing language.
- **DHW (Data High Way)**, responsible to transmit data with high speed among heterogeneous data sources like RAC, Greenplum, HDFS, MySQL, SQLServer and etc. **Main contribution:** architecture design, implementation of distributing framework for DHW based on thrift (open source framework for scalable cross-language service), designed and implemented the distributing balance algorithm for transmitting data on the cluster of DHW hosts , I/O module with HDFS. Using Python, JAVA developing language.
- **Hotspot**, responsible to discover hot tables visited frequently according to the ETL job log and to analyze the hot tables to optimize the data model for data warehouse. **Main contribution:** extracted tables from SQL log based on SQL parsing; viewed a SQL statement as a transaction, using association rule algorithm and etc, to mining the latent relationship among a large number of tables. Using C, Python developing language.
- **Alihive**, a library, responsible to access hive server and record the log for each SQL statement. **Main contribution:** designed and implemented the API to submit job to hive server. Using PERL developing language.

2007.8-2008.3

Alcatel-Lucent Company, Chengdu, Sichuan, China

Intern in MSGUI group as software engineer

- **MAPIM**, used to manage network element like Nodes, Link and etc. **Main contribution:** according to the business logic of network management, completed data design including: conceptual data model, logical data model, physical data model, designed and implemented the DBI(database access interface), and constructed simulative data for database by developing CORBA application using Name Service and Notification Service, mainly using C PLUS PLUS developing language.
- **GUI Server**, responsible to provide service of user graphic rendering. **Main contribution:** customized graphic rendering service in GUI Server by XML; participated to develop the module of responding the request commands from GUI Render Client; implemented CORBA communication module between GUI Server and MAPIM using JacORB, mainly using JAVA developing language.

PUBLICATION

1. 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
2. LI Chuan, TANG Changjie, **ZENG Chunqiu**, WU Jiang, CHEN Yu, QIU Jiangtao, ZHU Jun, DAI Li, JIANG Yongguang. "Discovering Multi-dimensional Major Medicines from Traditional Chinese Medicine Prescriptions", in *Proceedings of 2008 International Conference on BioMedical Engineering and Informatics (BMEI)*, pp. 260 - 264, 2008
3. 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
4. Liang Tang, Changjie Tang, Yexi Jiang, Chuan Li, Lei Duan, **Chunqiu Zeng**, Kaikuo Xu. "TROADGrid: An Efficient Trajectory Outlier Detection Algorithm with Grid-based Space Division", in *Proceedings of the National Database Conference (NDBC) (China)*, pp. 185 - 190, 2008

5. Ni Shengqiao, Tang Changjie, Wang Yongwei, Li Chuan, Zhang Yue, **Zeng Chunqiu**, Tang Liang. "**Boost Gene Expression Programming based on GPU**", in *Proceedings of the National Database Conference* (NDBC) (China), pp. 227 - 233, 2008
6. ZHANG Yue, TANG Changjie, LI Chuan, ZHU Jun, **ZENG Chunqiu**, TANG Liang, LIU Xianbin. "**Mining Naive Intervention Rules in Birth Defect Data**", *Journal of Frontiers of Computer Science and Technology* (China). VOL.1 No.2 ,2009.02, pp.188 - 197, 2009

AWARDS & HONORS

Contest Awards:

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.

Industrial Awards:

1. 07/2005, Excellent intern of Hwadee Company, as a leader of data design team.

Scholarships & Honors:

1. 1/2009, Excellent Graduated Graduate Student of Sichuan University in 2009.
2. 12/2008, Excellent Graduate Student Award of Sichuan University in 2007-2008.
3. 6/2006, Excellent Final Thesis of Bachelor Degree of Sichuan University.
4. 11/2005, Admitted by Sichuan University for graduated study exempting from examination.
5. 2002-2003, 2003-2004, Excellent Undergraduate Student of Sichuan University.
6. 2002,2003,2004,2005, Undergraduate Student Scholarship of Sichuan University.
7. 2002-2003, Awarded Scholarship of WESTERN CHINA METROPOLIS DAILY.

SKILLS

English Proficiency

GRE:	Verbal: 420	Quantitative: 750	Analytic Writing: 3.0		
TOEFL:	Total: 83	Reading: 26	Listening: 23	Speaking: 17	Writing: 17

Programming Skill

C/C++, Java, Python, Shell

Other Skill:

Database: Oracle, MYSQL, MS-SQLServer, Postgresql

Distributing Database: Greenplum

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