

Objective: A software development engineer position to satisfy a passion of continuous learning.

Summary: Eight years of enterprise software engineering experience building industry-standard platform systems. Demonstrated ability to quickly pick up new and complex projects.

Languages: Java (primary), Python

Work Experience:

Software Engineer at Cloudera (September 2013 - Present)

- Apache Hive committer and PMC member
 - Lead for joint Hive-on-Spark Join development effort across Cloudera, Intel, and MapR, architected Reduce, Map, Bucket, SMB, Skew join physical implementations on Spark.
 - Maintain Hive PTest precommit infrastructure for community, continuously make infrastructure and framework improvements to maintain the fast pace of the project.
 - Implement various features such as Metrics framework, Parquet timestamp and other integration, and defining File-Permission behavior of Hive data files
- Cloudera Hive team
 - Lead sub-team to improve reliability and supportability of Metastore and HiveServer2, delivering Metrics, WebUI, and scale test coverage so far
 - Resolve high-priority Hive production issues as feature-expert, including multiple customer on-sites to Cloudera's largest customer to ensure operation success.
 - Integrate Hive with complex Cloudera stack, including driving one disruptive rebase.
 - Drive improvements in Hive quality via random query generator in collaboration with Impala team: <https://github.com/cloudera/Impala/tree/cdh5-trunk/tests/comparison>
 - Winner of 2015 Cloudera Hackathon with Cloudera's Kafka Team: Hive on Kafka <https://github.com/HiveKa/HiveKa>
- Talks/Presentations:
 - Hive on Spark at Berlin Buzzwords and various HUG's in US: <http://www.slideshare.net/SzehonHo/hive-on-spark-berlin-buzzwords-53265886>
 - Hive Join Optimizations at Hive User Group: [https://cwiki.apache.org/confluence/download/attachments/27362054/Joins in Hive.pptx](https://cwiki.apache.org/confluence/download/attachments/27362054/Joins%20in%20Hive.pptx)
 - Hive on Kafka at Hive User Group: <http://www.slideshare.net/SzehonHo/hive-on-kafka-53265979>
 - Validating Hive with Random Query Generator: [https://cwiki.apache.org/confluence/download/attachments/27362054/Random Query Gen- Hive Meetup.pptx](https://cwiki.apache.org/confluence/download/attachments/27362054/Random%20Query%20Gen-%20Hive%20Meetup.pptx)

Principal Software Engineer at Informatica, Core Technology Group, Platform Services (June 2007-September 2013):

- Promotions from Software Engineer (April 2009) and Senior Software Engineer (April 2013)
- Lead Developer for Data Integration Service: Main Informatica application service, includes runtime mapping execution, web and RPC clients, metadata deployment, data/metadata caching, and monitoring.
 - Led team of 3 developers in Grid project to increase scalability of application to multiple nodes. Project included state-sharing and synchronization of metadata, caches, mapping execution load-balancing, log and statistics aggregation.
 - Improved startup to near constant time, by implementing background metadata loading during startup, synchronizing against incoming runtime requests.
 - Developed application for PC-Express product line: Designed new components for state management for persisting trial quota consumption and new sequence-state mappings.

- Team Lead for new Notification Service: New component for cross-tool collaboration.
 - Learned agile/project management by leading agile team involving 2 developers, 2 QA, PM and UI.
 - Designed new SDK to meet internal-customer needs, prototyped pub-sub delivery using newly-acquired UltraMessaging low-latency messaging technology.
- Member of Metadata Repository Service: Hibernate-based repository service, backed by database, stored metadata for all Informatica's tools and allow cross-tool metadata reuse.
 - Designed and implemented Hibernate Object-Relational Mapping component (generation and persistence of OR-Mapping using single table per class hierarchy strategy).
 - Optimized metadata deletion: implemented direct SQL for metadata delete, 5x performance improvement over Hibernate-driven delete for bulk object graphs.
 - Optimized metadata browse, save, and backup/restore of search indices for similar improvements.
 - Implemented Catalog storing meta-metadata about partitions, users, etc.
 - For first project, wrote efficient graph algorithms for generic metadata-model processing (copy, partitioning based on relation, de-serialization from Hibernate).

Previous Work:

- Reader for CS61C at UC Berkeley (Fall 2006)
- Lab Assistant CS61C at UC Berkeley Summer Sessions (Summer 2006)
- Physics Tutor at UC Berkeley Student Learning Center (Spring 2004, Fall 2005- Spring 2007)

Education History:

BS Degree at UC Berkeley (2003-2007), Electrical Engineering and Computer Sciences Major

- GPA 3.816/4.0

Relevant Coursework:

- | | |
|--|----|
| • Efficient Algorithms and Intractable Problems (CS170) | A |
| • Introduction to Database Systems (CS186) | A |
| • Introduction to Computer Graphics (CS184) | A |
| • Computer Architecture and Engineering (CS152) | A |
| • Introduction to Networking (EE122) | A- |
| • Operating Systems (CS162) | A |
| • Components and Design Techniques for Digital Systems (CS150) | A |

Relevant University Projects:

- CS184 Graphics Final Project (Procedural Modeling of Buildings- Generated randomized buildings using splitting rules to make an entire city)
- CS152 Computer Architecture Final Project (Created CPU on Calinx board that supports a subset of MIPS assembly language, 5-stage pipeline, and memory cache)
- CS162 Project (Implements parts of Nachos OS simulator, to allow thread synchronization, multiprogramming, demand paging, and reliable data transfer over network.
- CS150 Class Project (Created working FPGA Paint Engine that interfaced with SDRAM, N64 DQ, and AVEncoder).

Academic Achievements

- Phi Beta Kappa Honor Society Member from 2006
- Valedictorian: Cupertino High School Class of 2003
- National Merit Scholarship Finalist 2003
- AP Scholar With Honor 2003

Languages

- Fluent in Mandarin and Cantonese