Pinglei Guo

Santa Cruz, CA 831-295-1214

Email: piguo@ucsc.edu GitHub: https://github.com/at15 Linkedin: https://linkedin.com/in/at1510086

EDUCATION

MS. Computer Science University of California Santa Cruz GPA 4.0 Sep. 2016 – Mar. 2018

BS. Material Science Shanghai Jiao Tong University GPA 3.3 Sep. 2012 – June 2016

WORK EXPERIENCE

PayPal: Software Engineer Intern San Jose, California June 2017 - Sep. 2017

- Built API gateway for internal **container orchestration** platform using **Go**. Deployed in **production** on **Mesos** using **Docker**.
- Enhanced dashboard using **Angular 4**, used by operation team to troubleshoot Java, Node application and manage cluster itself.
- Introduced full text search using **Solr**, wrote and **open sourced** Go client with enhanced performance and SolrCloud support.
- Furthered continuous integration (**CI**) and deployment (**CD**) pipeline using **Jenkins** and **Docker**.

GitCafe: Software Engineer Intern Shanghai, China Jan. 2015 – Mar. 2015

- Reduced **Ruby on Rails** application load time by 5% through optimizing regular expression in markdown parser.
- Fixed user subsubcription logic, solved 20% pricing related user tickets.

Dongyue Web Studio: (Part-time) Full stack web developer & Tech lead Shanghai, China Sep. 2013 – Jan. 2016

- Led web and mobile team. Refactored online ticket booking application **tongqu.me** in 3 month, used by 20,000 students .
- Built high traffic website and REST API using **PHP** and **MySQL**, reduced database contention using job queue and cache.
- Utilized **Redis** as cache and rate limiter, increased QPS by 120%, reduced database load by 40%, filtered out most bot traffic.
- Refactored **jQuery** codebase using **AngularJS**, increased homepage loading speed by 60% using Ajax and pre-render.

PROJECT EXPERIENCE

Distributed Time Series Database github.com/xephonhq/xephon-k UCSC Nov. 2016 – Present

- Implemented a distributed time series database on top of **Cassandra** in **Go**. Support both **JSON** and **Protobuf** via HTTP/2.
- Designed a **columnar format** modeled after **Parquet** and **InfluxDB** with higher compression and less write amplification.
- Created **benchmark suite** for Xephon-K, OpenTSDB, KariosDB, InfluxDB and a consistent client interface for different TSDB.

GPU accelerated in-memory time series processing github.com/at15/ts-parallel UCSC Apr. 2017 – June 2017

- Expanded **benchmark suite** for different **C++** GPU computing framework on **CUDA** and OpenCL, Thrust, Boost, ArrayFire.
- Implemented OLAP queries like top-K, group by for multi dimensional time series data on both CPU and GPU backends.
- Initiated a in memory **column store** with run length and dictionary encoding, saved 90% space for regular time series data.

Distributed systems monitoring prototype Shanghai Jiao Tong University Mar. 2015 – Jan. 2016

- Enhanced monitoring system for distributed system using **Cassandra** and **MongoDB** written in **Java** and **C++**.
- Deployed in China Telecom's **Kafka** cluster, detected **anomaly** in disk and memory usage, improved **capacity planning**.

B+ index for Hive Shanghai Jiao Tong University Nov. 2015 - Jan. 2016

- Implemented B+ index for **Hive**, index is generated using **MapReduce** and stored in **HDFS**.
- Accelerated point and range query using in memory LRU cache, supporting external cache like **Memcached** and **Redis**.

SKILLS

Language	Go, Java, PHP, JavaScript, TypeScript, Python, SQL, C++, Scala, Shell
Database	Cassandra, Solr, MySQL, MongoDB, Redis, Elasticsearch, KairosDB, OpenTSDB, InfluxDB, Prometheus, Graphite
DevOps	Docker, Vagrant, Ansible, Mesos, Aurora
Framework	Angular, Laravel, Spring, Dropwizard, Express, Rails, CUDA, Hadoop