

简历

个人信息

姓名:

Git account:

电话:

邮箱:

教育背景

2007.9 - 2010.3 浙江大学计算机科学 硕士

学习方向为计算机应用, 数据挖掘

参与发表 WWW 会议论文<tag oriented document summarization>

校级创新设计大赛三等奖

INVIDIA CUDA 编程比赛一等奖

2003.9 - 2007.6 浙江大学软件工程 学士

浙江大学大学生创新项目比赛三等奖

2005 成绩优秀奖学金

2004 国家二等奖学金

技术特色

熟悉 Clickhouse 数据库源码并做修改, Clickhouse 开源社区贡献者, 与开源社区关系密切

熟悉 hadoop hdfs, mapreduce 源码并做修改

熟悉 spark core, spark sql, spark streaming 源码并做修改

熟悉 kafka, flume 源码并做修改

熟悉 Greenplum MPP 分布式数据库源码, Greenplum 开源社区贡献者

熟悉 Postgress 源代码, 熟悉数据库原理操作等

熟悉 CockroachDB, Druid, CloudFoundry, Hive 等源码

熟悉 LINUX 操作系统, SHELL, 网络 TCP/IP 等基础知识

多年 JAVA 开发经验, 熟悉 JVM, 独立设计完成较大项目

多年 C/C++开发经验, 参与大型项目开发运维

有使用 GO, Docker, Scala, Python, Node.js 等进行开发的经验

基础级机器学习数据挖掘

有带小团队进行技术开发的经验

有数千台机器的运维升级经验

大型项目快速定位问题，排查问题，解决问题的经验
在分布式存储，分布式计算领域经验丰富

踏实认真，热爱技术，对技术精益求精

工作经历

2018.9 - 至今 快手 大数据平台部 资深架构师

从 0 开始搭建 Clickhouse 集群，对公司内部各业务团队提供 Clickhouse 技术支持
目前已有数百节点，上千张表，十多个业务团队在使用我们提供的实时大数据分析服务
存储规模列存压缩后数 P，每天导入上万亿行原始数据，峰值超过每天两万亿
全面负责部署维护 Clickhouse,根据内部业务需求对 Clickhouse 进行定制开发，包括留存，漏斗，路径，AB-TEST 分析，用户画像，用户分群，机密数据库等功能，BUG FIX, 版本升级等
支持从 HDFS, KAFKA 等外部系统自动导入业务数据
快手学院 CLICKHOUSE 技术讲师

2018.6 - 2018.9 神州云联技术有限公司（初创公司） 技术总监

基于 ClickHouse 打造商用分布式数据库
基于 Dgraph, CockroachDB 打造商用多模数据库
数据中台建设
全面负责公司的技术开发，包括技术调研，技术方案制定，难点攻关，日常技术交流等

2017.11 - 2018.5 今日头条数据部 资深架构师

- 负责 druid 对今日头条各产品实时事件数据流进行分析
 - 数百节点，处理数十个数据源，高峰时每秒上百万数据
- 完成新版事件实时分析系统（TEA 系统）的技术选型和系统搭建，使用 clickhouse
 - 邀请 clickhouse 开发团队访问今日头条进行技术交流，现场指导部署
 - 使用 clickhouse 方案 PK 掉优化近一年的 spark 事件分析老系统
 - Clickhouse 已成为内部 TEA 系统的新方案，准备上线数百台机器，并成为 Aeolus, Mophiline 等数据分析系统的实现方案

后续进展：目前头条已经重度使用 Clickhouse 达数千台规模，充分证明当初技术选型的正确性

2016.3 - 2017.11 云狸科技(北京) 初创公司技术总监 首席架构师

负责公司大数据分析平台的设计搭建

- 源码级修改 kafka, flume, spark, hadoop 等大数据组件
添加虚拟路径、访问控制、资源共享、数据处理、传输加密等功能
结合业务对分析组件进行定制化
Spark Streaming, Spark core, Spark sql 模块

Hadoop hdfs, mapreduce 模块

Kafka 通讯模块

Flume - kafka 通讯模块

2. 设计并实现 Unified ETL 模块, 支持常用数据库的存量与增量数据的实时同步, 数据加密, 数据建模等功能 (JAVA)

包括 Mysql, Postgres, Sqlserver, Oracle 等常见数据库, 支持多种数据类型

包括 Hadoop proxy

在某项目中, 采用统一的数据模型将 14 个不同的业务数据库集成在一起, 对数据分析进行支持

对数据传输进行压缩, 多线程并行上传, 断点续传等优化, 性能提升 10 倍以上

使用 Nginx 反向代理, jetty 库, Gruva, Spring 等

3. 编写常用业务代码, 支持业务分析功能

使用 JAVA 编写业务代码

Scala 编写数据处理代码

Jar 包加密

基于 spark streaming 进行预计算

2014.4 - 2016.2 EMC/Pivotal Greenplum 内核团队核心成员 资深工程师

1. 负责 GPDB 日常 bug fix 等基础工作

Gdb 源码调试

2. 学习 Postgres, GPDB, GPFDIST, GPLOAD, GPHDFS 等源码, 参与 GPDB 5.0 的部分开发

熟悉 planer, dispatcher, executor, external table 等模块

添加对数组操作函数, 添加对新 SQL 语句的支持

3. 后期专注于 GPHDFS 模块的设计优化

增加对 Parquet, Avro, Hive 的支持, 并行加载, 类型检查, 类型映射, Schema 更新等

使用 mapreduce 底层代码优化对 hadoop 文件的读写

增加了 GP 表与外部表的 Schema 检测, 需对比 GP 表字段类型和外部数据源字段类型是否匹配

获 2015 年全球(Global)杰出贡献奖

在公司技术人员年度会上宣讲自己的技术成果

在公司内部创新比赛中, 以 GPDB + Redis 整合的方案, 获得一等奖

2011.3 - 2014.4 盛大云计算公司北京研发团队 资深工程师

1. 参与设计开发自研产品 EBS(Elastic Block Service), 核心研发(C++, LINUX, 规模约 50 台)

从 0 开始全程参与, 对业务流程, 内部各模块均有涉猎

负责数据通讯模块的开发, 基于 `epoll`, `pthread` 的高性能异步网络通信
基于 `thrift` 的内部消息通信
异常处理, 故障检测, 曾解决 3 个锁相互锁死的 `bug`
`gdb`, `pstack` 等常用调试工具
运行日志分析, `awk`, `sed`, `shell`
常用小组件, 连接池, 内存池等

因团队扩张进入下个项目

2. 参与开发云引擎, 核心研发, 基于 CloudFoundry (RUBY, JAVA, LINUX)
阅读 Cloudfoundry 源码, 包括 `dea`, `router`, `message queue`, `driver`, `controller` 等的源码
阅读部分 ElasticSearch 源代码并结合业务做小改动
设计实现了应用商店, 应用管理等业务功能(JAVA WEB)
云引擎线上运维

因盛大总公司要求, 随大团队转而支持酷六 CDN, 已有项目仍做维护

3. 参与酷六视频 CDN 的优化, 维护 2000+机器(C, LINUX)
主要参与视频分发模块的开发运维升级等工作
分库分表, 消息队列
对视频分发流程熟悉, 热点视频, 冷点视频分发等

2010.4 - 2011.3 创新工场云计算团队 高级工程师
基于 `hadoop`, `hbase`, `cassandra`, `rabbitmq` 等搭建云平台
对外提供存储, 计算等服务