

基本信息

姓 名	祁超	生 日	1988-12-29
邮 箱	xxqi1229@126.com	电 话	18610961116
学 历	湖北工业大学(本科)	专 业	网络工程
意 向	JAVA/大数据		

工作经历

时 间：2017-09 ~ 至今

公 司：武汉镜像科技有限公司

部 门：数据中心

工作内容：大数据平台搭建与维护；数据中心 web 系统开发；

主要技能：大数据：flume、kafka、hadoop、spark、hbase、elasticsearch、zookeeper；
Java Web：nodejs、express、springMVC、iBatis、mysql、mongodb、nginx；

时 间：2014-12 ~ 2017-09

公 司：百纳（武汉）信息技术有限公司

部 门：CommonService

工作内容：大数据平台维护；运营与客服 web 系统开发；

主要技能：Java Web：angularjs、nodejs、netty、iBatis、mysql、hbase、nginx；
大数据：flume、kafka、hadoop、spark、hive、hbase、zookeeper、zakaban、canal、
kylin

时 间：2013-12 ~ 2014-12

公 司：搜狐视频（北京）

部 门：移动开发事业部

工作内容：移动视频 Html5 的功能开发和维护（m.tv.sohu.com）

主要技能：jsp、servlet、redis、mysql、mongodb、hibernate、nginx

时 间：2012-02 ~ 2013-12

公 司：人人网（北京）

部 门：安全中心

工作内容：运营平台(admin.renren.com)和安全中心(safe.renren.com)的开发和维护工作

主要技能：jsp、springMVC、mysql、memcached、thrift、zookeeper、elasticsearch、nginx

个人技能

熟练掌握：Java 语言：集合、IO/NIO、并发与锁、线程池、内存模型、GC 原理与优化等。熟练使用 mysql 和 mongodb 数据库，了解事物和索引优化，有处理大数据的经验。熟练使用 flume、kafka、spark、hadoop、hbase、Zookeeper 等大数据技术，阅读过部分源码。熟练使用 redis 和 elasticsearch。熟练使用 springMVC、iBatis。熟悉常用的数据结构和算法。

基本掌握：python、JavaScript、Scala 语言。有使用 hive、azkaban、sparkstreaming 和 sparksql 的经验。有 netty 开发经验。有使用 nodejs 和 express 搭建 web 的经验。熟悉 jquery、html、css 等前端开发。

基本了解：了解 spring Boot 和 Spring Cloud。有使用 nodejs+angularjs 搭建 web 前端的经验。

项目经验

项目名称：数据中心的建立（镜像）

项目描述：数据中心主要包括：数据收集、数据分析、数据展示三部分。数据收集：flume+kafka+hadoop；数据分析：spark+mongodb+elasticsearch；数据展示：bootstrap+echats+nodejs+express+springMVC+mysql。

主要职责：

- a) 独自完成了大数据平台的搭建和 web 系统架构搭建；
- b) 基于 webSocket 的客服聊天系统开发；
- c) 分别实现了基于 mongodb 和 ES 的大数据查询；数据写入/查询优化；
- d) mongodb 插入、查询优化等工作；
- e) spark 任务编写，web 需求开发；

技能列表：flume+kafka+hadoop+spark；nodejs+redis+pringMVC+mybatis+mongodb

项目名称：运营&客服平台的建立（百纳）

项目描述：为了游戏精细化运营，每天会生成多维度报表，给游戏运营和制作人提供数据支持；为客服提供客服后台对用户进行管理。报表用 spark 任务生成数据，azkaban 负责调度 spark 任务，结果保存在 mysql 和 hbase；实时需求采用 kafka+sparkstreaming 写 mysql 或者 elasticsearch 供查询使用；前端展示采用 nodejs + angularjs+echats，后端接口调用采用 ajax + http restful api 的形式。基于 netty 构建自己的 http 服务器提供 web 服务。

主要职责：

- a) spark 任务编写，报表生成和展示；
- b) 基于 Hive 的大数据处理；
- c) 基于 Innodb 分区表的大数据处理以及 TokuDB 引擎优化；
- d) 基于 elasticsearch 的用户聊天与后台搜索相关需求开发；
- e) 基于 Canal 和 Hbase 的用户画像系统；

技能列表：nodejs、echats、angularjs、netty、spark、kafka、mysql、elasticsearch

项目名称：大数据平台 Groot 维护优化（百纳）

项目描述: 为了精细化运营, 需要收集大量的行为数据供运营分析使用。Groot 采用 Flume 收集用户数据, 在 Kafka 中聚合缓存, 通过分布式分拣程序按照用户行为持久化到 hdfs。在 hdfs 上提供 spark 计算框架、hive 数据仓库和 hbase 分布式存储服务。

项目职责: Groot 系统的维护、分布式分拣程序的升级

技能列表: flume、kafka、hadoop、hive、hbase、spark、canal、kylin

项目名称: m.tv.sohu.com 开发维护 (搜狐视频)

项目描述: m.tv.sohu.com 和搜狐视频客户端一样是搜狐视频移动设备的主要入口之一。站点需要提供基本的视频网站的所有功能; 最主要的是需要改变用户的使用习惯, 拉拢用户, 提高搜狐视频的用户量和客户端的下载量。(站点每天的视频播放量 (VV) 达到 4500W; 每天拉起客户端 100W; 每天通过站点下载客户端达到 5W;)

主要职责: 主要是维护和新版发布的工作;

- 1、nginx 缓存升级, 性能调优, varnish 的 http 加速等;
- 2、redis 集群的改造, 包括集群扩展, 哨兵监控等。
- 3、后端编码, 高并发接口的实现; 压力测试

技能列表: nginx、resin、jsp、servlet、java、mysql、redis、varnish、tcpcopy、高并发。

项目名称: 一键删除 & UGC 检索 (人人网)

项目描述: 为了快速获取用户的 UGC, 采用 elasticsearch 对用户的 UGC 进行索引, 供客服和运营查找非法 UGC, 一旦确定, 可以快速删除。

个人职责: 负责 Elasticsearch 集群搭建、索引建立

技能列表: elasticsearch、DelayQueue

项目名称: URL 爬虫 (人人网)

项目描述: 为了增强用户体验, 在人人网上发布的外部链接地址都会转化为 <http://rrurl/xxx> 的短链形式。大量淘宝客到人人网发布推广链接的时候, 严重影响用户体验。爬虫系统会对用户发布的 url 进行爬虫, 找到用户链接的最终目的地, 并判断链接是否非法

个人职责: 全权负责, 主要包括: HTMLParser 页面解析, url 跳转判断

技能列表: HTMLParser、HTTP

项目名称: 任务处理系统 (人人网)

项目描述: 安全中心每天会执行大量的任务 (例如: 删除用户状态、照片, 封禁用户等), 为了防止任务突增带来的接口安全隐患, 将所有任务执行异步化。将任务封装存储数据库, 启动多个 resin 服务, 通过 zookeeper 选举 leader, leader 负责分发任务, follower 负责执行任务, 每个 resin 服务中的任务执行线程可以通过 zookeeper 动态调控。

个人职责: 全权负责, 主要包括: 多 resin 集群的 leader 管理, 动态线程调控等

技能列表: zookeeper、resin、DelayQueue & LinkedBlockingDeque

项目名称: UGC 过滤系统 & 人人小秘书自动问答系统 (人人网)

项目描述: 人人网各业务 (相册、状态、日志...) 将用户行为日志推送到安全中心 kafka, 消费者实时消费数据, 对用户发布内容和样本库中的样本进行比对, 判断相似度, 如果相似度达到阈值, 执行对应的操作。

我的职责：全权负责，主要包括：Lucene 分词、文本相似度 Thrift 服务、用户处理等。

技能列表：kafka、lucene、thrift、mysql

个人评价

热爱技术，勇于挑战；喜欢 trouble shotting 和 bug fix；平时看看 blog 和 github。