谢丹丹

6年工作经验 danxiew@163.com 15652767753

2006/9–2010/7 本科 - 电子科技大学 - 信息管理与信息系统

# § 意向岗位

数据工程师

# § 工作经历

## 2013.11-now 多盟无限科技有限公司 系统研发工程师

* 多盟数据平台核心开发，参与整条数据流过程。负责数据平台的搭建和维护。工作期间观察到交互式数据查询的缺失，提出并独自开发完成OLAP离线数据平台搭建，大幅度提升了业务人员的数据查询效率。同时提出了面向主题的数据仓库给上层业务，减少了各种Dashboard和报表开发工作
* 负责数流式数据管道ETL处理和离线日志处理的开发和维护，并作为实时统计模块的第二负责人
* 负责多盟多个在线广告平台投放服务组件开发，完成了重要组建广告事件通知服务，对服务平稳运行有丰富的经验
* 负责RTB平台的媒体、广告尺寸等业务数据抽取，支持了日常的精准投放业务

## 2011/6-2013/07 北京触控科技有限公司 系统研发工程师

* 负责游戏中心平台设计与开发，运营1个月内流水突破100万，注册用户截至达130万。对接了《我叫MT》等休闲类游戏，主要负责设计用户行为实时统计和支付系统
* 负责手机应用后台开发，完成一款拥有200+万用户的火车票买票应用和其他2个APP项目
* 打造了统一后台管理平台，抽出用户行为统计、留言、消息通知等公用功能，增加功能复用度，大幅度降低开发成本
* 承担一部分运维工作，负责搭建和维护MySQL、SVN、PHP、Redis等环境搭建

## 2008/7-2011/5 电子科技大学星辰工作室 联合创始人&系统研发工程师

* 负责开发维护和银行对接的学生贷款还款系统，提供全校学生合同生产、校验打印服务
* 负责开发学校各种信息发布网站，一共完成了6个，其中主导了4个
* 和校友组件创业团队，开发了承接了成都十所项目管理系统和全国高等学院招生管理项目

# § 项目经历

#### 数据平台（2013/11 － now） 多盟

* 使用**Storm**流式计算框架编写Java程序，搭建了广告用户事件流的实时ETL处理管道，处理每日3T数据的清洗、转换、过滤工作，平均消息延迟3s
* 使用Redis作为核对仓库，构建了数据流的基本anti服务和数据匹配补全服务，使得核对率达到98%
* 构建了以Hive为基础的数据仓库，设计了其中Kafka流式消息转成Rcfile逻辑并且处理了其中较难的时间窗口问题
* 开发了原始日志文件备份服务，使用Python实现了收集散落在各个机器上的nginx日志到HDFS上并利用任务队列的方式设计了备份失败重试特性

#### OLAP系统（2014/5 - now） 多盟

* 引进和推动saiku整套OLAP解决方法，替换了原来简陋的数据查询系统，向各种层次业务人员提供友好的系统，同时减轻了技术开发工作量
* 探索面向分析的列式引擎，从Mysql改为面向分析型数据库（infobright、vertica），可行查询数据容量扩大了至少10倍，并且实现了整个多维度数据制作模块。
* 开发了一款sqlproxy工具，实现了在saiku现有的框架下一个cube可以使用多个数据引擎功能，从而能发挥出各种引擎的特点
* 对mondrian修改并新加了诸多features，以满足了广告大维度数据的展现。比如：1. 加入like、limit语意 2. 增加log4j新接口，支持模糊查询维度列表需求
* 设计了元数据管理系统。利用python构建了整个程序，实现了各种数据库的schema自动生成，同时搭配钩子设计实现了维度表的数据搜集

#### 广告事件消息通知系统（2015/10 - now） 多盟

* 采用生产者消费模型和线程池技术用python构建了qps达1400的消息发送系统，以kafka做输入源和中间任务队列，以redis和mysql分层作为统计服务储存
* 优化了消息任务排队策略，改变连续重试策略成延后重试的方法，有效防治了个别通知目标服务器宕机时候阻塞整个系统的延迟
* 重现设计了更符合业务需求的数据监测点位，简化了系统状态监测和问题排查速度，并且构建了完成的系统性能监测和状态统计管理系统
* 在此模块的python线程池加入监控，防止线程池线程泄露导致服务性能下降

#### 实时用户数据统计系统（2013/1 – 2013/7） 触控

* 使用php构建系统，每日处理的日志量达到3G。数据层完全利用redis做为db，以达到最快速度的数据统计处理速度
* 实现了完备的日志备份系统。利用redis list作为任务队列，首次统一备份一遍原始用户事件日志。当事件经过日志解析器处理后，会根据其事件类型和版本进行分类二次备份。

#### 火车票移动应用（2011/6 – 2013/1） 触控

* 提供准确、实时的火车站点、线路、票务查询和订购业务。使用Redis建立数据缓存，仅用一台机器支撑了2300,000用户
* 利用Dijkstra算法一次性产生所有线路最短路，并通过定时更新策略提供中转查询服务。其中通过水平分表方法存放9,000,000多条线路记录在mysql中

### 科大学生贷款还款系统 （2008 - 2010） 星辰工作室

一个网络桌面软件，使用了C#网络框架和事件驱动并发模型开发，提供全校学生使用

# 技能特长、编程语言和软件

* 熟练使用Java,PHP,Python,C,Shell语言
* 掌握各种数据储存引擎原理，拥有SQL和NoSQL使用和调优经验。实践过MySQL,Hive,Infobright,Vertica,Redis
* 熟练掌握Kafka、Storm、Hadoop、Thrift的设计原理和使用方法
* 对OLAP的理论知识和实践技巧有丰富的经验