

谢丹丹 电话:15652767753 Email:danxiew@163.com

---

## § 工作经历

### 2013.11-now 多盟无限科技有限公司 系统研发工程师

#### 多盟数据仓库 (2013.11 – now)

keyword: Java, shell, python, kafka, hive, vertica, infobright

##### 分层事实数据系统

- 设计开发了面向广告、媒体和用户画像主题等的立方体数据模型。
- 使用了各种不同引擎来的速度和容量来放入不同颗粒度数据，最求更小的颗粒度更快的查询速度，并不断探寻更合适引擎同时淘汰不适合引擎。
- 开发了每日20亿条（1T）从storm流出的操作型历史原始数据收集、入hive和其中的数据清洗
- 构建了部分外部点到点的数据质量端到端一致性监测，并且开发了多种分维度层级检测空白异常监测

##### 维度数据收集同步系统

- 实现了中心式异构数据结构和数据同步，借鉴淘宝的datax异构数据引擎插件设计思想，利用python构建了整个程序，解决维度数据不再出现不一致性。并通过搭配多盟的任务调度平台，实现了定时收集业务数据更新功能

##### sqlproxy

针对不断各种异构的数据引擎，开发了这款借助于mondrian的聚集表路由功能的隐藏数据分层工具。对mondrian层面完全透明，其相当于正常访问一种新的关系型数据库。

- 遵循JDBC接口，提供面向java的一个通用接口。
- 通过mondrian用户的查询sql，自动解析出表,再根据表库配置，自动实现信息自动匹配跳转。

#### 多盟olap系统 (2014.5 - now)

keyword: Java, shell, mondrian, saiku

独自基于开源saiku整套解决方案搭建了多盟OLAP系统。

- 使用新系统更替原来整套面向人员查询的数据系统。引擎从Mysql改为面向分析型数据库，查询数据容量扩大了至少10倍，同时采用成熟的cube数据模型，更适合从各种角度批量获取数据
- 对mondrian修改并新加了诸多feature，以满足了广告大维度数据的展现

## 多盟点击事件消息通知系统（2015.12 - now）

keyword: python, php, kafka, redis, thrift, mq, logtailer, dm303, yama, qmm（后5个是多盟自有基础平台）

通知点击事件给广告主和第三方数据监测平台，整个系统分为click collector and sender、click monitor和统计后台三个部分。

- 改进了消息发送机制，采用回归式队列代替原来低效的线性队列形式。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吞吐量
- 重新设计了click monitor模块的数据监测点位，简化了系统状态监测和问题排查速度。其中思路借鉴于fb303，使得统计形成一个中心存取方式，从而使得数据和应用分离，形成各种展现集中统计可视化平台

## 2011.6-2013.07 北京触控科技有限公司 软件工程师

### 用户数据统计和分析系统（2013.1 - 2013.7）

- 开发了api有用户注册、登录、好友关系、支付、数据、日志统计和报表
- 设计了实时日志分析系统：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

### 火车票数据服务项目（2011.6 - 2013.1）

keywords: php, redis, mysql

- 构建了准确、实时的火车站点、线路、票务查询和订购业务
- 建立了数据源层、数据库层、内存缓存层、客户端层四层数据模式，仅用两台服务器来承受日900,000次请求，并且可是实现在线火车时刻表数据更新
- 设计了中转模块，通过Dijkstra算法预先生成所有线路最短路并通过分表技术存放9,000,000多条线路记录在mysql中，保证了最短路径的中转方案和查询性能

### 精品应用推荐（2012.3 - 2012.5）

- 设计实现了从appstore，appshopper等多方渠道的信息抓取器
- 开发了记录用户详细的操作行为日志收集模块——安装次数，击行为和下载行为，以此为根据来动态更改推荐排序

### § 教育背景

2006.9 – 2010.7 电子科技大学 信息管理与信息系统

### § 技能特长、编程语言和软件

- Java(expert),PHP(expert),python,Linux C,Shell
- Linux,hive,hadoop,infobright,vertica,kafka,Redis,Mysql