

# 公子

- **概况**：6年工作经验 | 男 | 1988年1月 | 本科 - 计算机科学与技术 | 未婚 | 175 cm |
- **电话**：138-0013-8000 | **邮箱**：10086@qq.net | **地址**：北京市
- **公司**：北京市xyz网络科技有限公司 | **行业**：互联网/电子商务 | **职位**：大数据研发工程师

## 自我评价

6年互联网行业工作经验。2年基础后台研发，4年数据平台研发管理经验。从工作之初，参与基础平台搭建，到目前负责数据仓库团队，同数仓团队一起紧跟公司业务发展，支撑业务的数据需求；协同产品梳理数据需求，引导业务方按照合理的方式提出数据需求。负责数据模型、数据报表的设计和开发。熟悉Java/Shell/Scala开发，熟悉Hive/Hadoop/Spark/HUE等开源大数据工具，熟悉MySQL/MongoDB等数据库；保持持续学习的心态，具有较强抗压能力，良好的沟通能力、团队精神和执行能力。

## 求职意向

- **目标职能**：大数据研发工程师 | **期望薪资**：面议 | **到岗时间**：一个月内 | **目标地点**：北京市

## 工作经验

2015.10-至今 北京市xyz网络科技有限公司 | <=1000人 | 在职 2年5个月 | BI-大数据研发工程师

开发工具：Hadoop/Hive/Spark/Scala/MySQL/Java/Hue

- 负责诸葛IO数据分析工具的接入。
- 数据模型设计、开发和维护；Hive UDF的开发和维护。
- 对接外部需求，安排数仓组的工作，支撑全线业务的取数、报表、邮件等形式的数据需求。
- 推进数据采集、数据同步规范，推动业务产品、业务开发、数据分析人员按照规范执行。
- 数据运维，确保数据数据质量；维护数据指标，确保指标的明确性。
- 对运营、市场、商务等业务人员以及数据分析师分别进行培训，提高数据使用方的数据意识。
- 数仓重构，根据维度建模理论对数仓的业务模型重新梳理，为新业务的接入和建模提供参考。
- 编写和维护数仓的文档。

2014.08-2015.10 深圳abc网络科技有限公司 | 互联网/电子商务 | >=3000人 | 在职 1年1个月 | 技术中心-后台研发工程师

开发工具：Hadoop/Hive/Hue/MySQL/Java

- 土巴兔装修日记服务的重构和维护。
- 土巴兔业主后台、图片服务等基础服务的开发和维护，解决了微信、PC、App 端各自开发造成的投入高、维护难、流程乱、易冲突的问题。
- 协助搭建土巴兔首个Hadoop集群及维护，保证整个集群的稳定；指导数据分析师优化Hive。

2012.08-2014.04 上海zzz信息科技有限公司北京分公司 | 互联网/电子商务 | <=100人 | 在职 2年 | 技术中心-数据平台负责人

开发工具：Python/MySQL/MongoDb

抓取线上微博、大众点评数据，包括位置微博、微博用户、点评网商家、点评网用户、点评网用户评论，清洗、分析和挖掘，找出对应关系，帮助商家精准营销，并协助公司App的推广。基于这些数据和关系

- 用户建模：用户的活动范围、路径、消费偏好、整体消费水平、常去的位置。
- 商家建模：商家的热度、影响力、同业客流排行、消费水平排行、消费者评价排行、老客户、潜在客户。

通过该平台，协助运营部进行产品推广，推进市场部开发线下商户，助力商务部展开商业地产合作。

## **2012.03-2012.08 北京oooo科技有限公司 | 互联网/游戏 | 100-150人 | 在职 5个月 | 技术中心-软件开发工程**

开发工具：Python/MySQL/Hadoop/Hive

负责报表管理平台的设计和搭建；负责为广告部、运营部等相关部门设计业务报表；在公司内部分享自动化报表管理，推广报表管理平台；广告后台日志清洗和分析，统计日报和周报。

## **项目经验**

### **诸葛IO数据分析工具的接入 2017.12-2018.03**

负责诸葛IO数据分析工具的接入，将app业务模块重新梳理，设计符合诸葛IO规范的埋点，并推动产品、开发进行埋点，并在线上后持续测试数据，验证数据的准确性。

数据分析师、市场人员、运营人员等可以通过该工具进行实时趋势分析，分析关键页面的转化漏斗、关键操作的留存等，为业务的发展规划提供决策参考。

### **OneData指标管理 2017.09-2017.10**

负责xyz网的数据指标的管理，推动业务方明确各个指标的业务定义，然后数仓给出指标的技术定义。通过统一指标的口径，避免数据烟囱。

数仓参与开发OneData指标管理工具，为业务产品、数据分析师提供唯一的指标查询入口。数仓持续维护指标，根据业务发展，调整指标定义或者新增指标。

### **实时路径模型开发和重构 2016.08-2017.05**

负责开发用户路径模型、成单路径、实时成单路径。

路径模型是成单路径的基础，采用 MapReduce 开发。在使用过程中多次重构和优化。存储上采用 ORC格式，磁盘占用仅为普通格式的十分之一，极大提供磁盘使用效率。

实时成单路径是先通过Spark Streaming 消费 Kafka，将结果load至 Hive 外表，后续的计算基于 Hive，路径通过路径模型计算。

### **土巴兔业主后台 2014.12-2015.10**

负责整理业主后台的流程；把控项目的研发进度。

土巴兔的业主等登陆业主后台，实时查看自己的装修进度，及时参与到装修过程中的准备、水电、泥木、油漆、竣工这5大关键阶段，发现问题可以及时投诉和维权。

### **土巴兔图片服务 2014.10-2014.11**

负责项目的图片服务的研发和测试。

公司业务迅速发展，为了方便、更高效的存储文件（图片占80%），急需开发一套公用的、对外提供接口、各端接入的服务，便有了以Java开发、MongoDB存储的设计方案。

## **土巴兔大数据平台 2014.08-2014.12**

Hadoop集群搭建、维护和优化；指导BI组Hive优化。

刚刚入职时，Hadoop集群只有雏形，机器资源紧张，BI人手短缺，从最初搭建Hadoop的同事手上接过的集群只有3台机器，存储、计算都是问题。接手后，一方面优化Hadoop的配置文件，降低并行计算时资源浪费，提高内存的使用率；一方面根据数据特性，对新人的Python、Hive脚本进行优化。

## **语言能力**

英语CET六级

## **兴趣爱好**

羽毛球