# **个人简历**

# **个人信息**

|  |  |  |
| --- | --- | --- |
| 姓 名：王冲 | | 性 别：男 |
| 年 龄：29岁 | | 籍 贯：陕西咸阳 |
| 工作经验：6年 | | 学 历：本科 |
| 电子邮件：wangch\_chosen@163.com 移动电话：17791830884  技术博客: <http://www.chongblog.xyz/about/>GitHub: https://github.com/wangchong386  **求职意向** | | |
| 目标地点：西安 | 性 质：全职 | |
| 目标职能：资深大数据开发/分析工程师  **教育背景** | 薪资要求：面议 | |
| 2010年9月 至 2014年7月  **工作经历** | 西安邮电大学/通信工程专业 | |

2018年5月~至今 华为 高级数据/分析开发工程师

2017年5月~2018年5月 百度 高级大数据开发工程师

2014年7月~2017年5月 敦煌网 数据仓库工程师

# **自我评价**

勤奋，踏实，对工作充满热情, 对数据敏感，有较强的逻辑分析能力，有互联网数据仓库开发经验；自学能力与环境适应能力较强，善于分析解决问题，并提出可行性方案；良好的心理素质，具备良好的编码习惯，能够承受高强度的工作压力,具有良好的团队合作精神。**技能**

* 掌握hadoop集群的搭建以及各组件的维护升级，了解数据仓库建设基本思路，有数据仓库建设项目经验，熟悉数据仓库的主题分析.构建DMP大数据平台经验
* 熟练ETL设计与开发，熟悉Oozie,kettle,Rundeck等ETL调度依赖工具
* 语言：熟悉shell，python。能够快速理解JAVA代码，熟悉IDEA等开发环境
* 熟悉 spark内存分布式计算(Spark core,Spark sql,Spark streaming 等组件)
* 熟悉Flume日志获取，Kafka消息中间件，sqoop数据传输,Zeppelin交互式分析查询
* 精通spark-sql或hive sql，编写UDF,有较强的开发调优能力
* 熟悉oracle,mysql,postgresql等主流数据库
* 熟悉使用Smartbi、BO、Saiku报表工具,及集成到项目

**华为（2018年5月~至今）**

* 职位：高级研发工程师/Team Leader
* 部门：消费者云服务
* 产品线：CBG
* 直接下属:12人
* 主要职责：

1. 负责华为搜索，云服务，应用市场等大数据平台，数据仓库，数据建模等实时和离线的搭建

2. 为数据画像，WiseMkt，搜索，推荐，信息流，AI平台，Analytic Kit等20+数据产品，提供数据分析服务

3. 每天涉及1600+数据源，日吞吐量在1.5PB左右

**百度（2017年5月~2018年5月）**

* 职位：高级研发工程师
* 部门：百度大搜-oppd-ubs数据平台
* 产品线：手机百度
* 主要职责：

1.负责 ubs 数据平台 ETL 任务的开发与运维工作，参与流量的反作弊策略以及回溯的运维工作，保证自身原因 0 错误，保障核心数据流的 SLA

2. 负责开发手机百度多维分析模型任务

1. 参与集群迁移，负责手机百度产品线的任务组以及日志数据的迁移工作

**敦煌网（2014年7月2017年5月）**

* 职位：数据开发工程师
* 部门：大数据/基础架构
* 主要职责：

1. 负责公司hadoop集群搭建以及优化。将公司hadoop集群升级到2.0，增加 kafka,spark,sparkstreaming等组件。为流式计算处理提供条件。对Yarn 资源管理系统进行调优。
2. 参与数据开发,将Flume实时采集用户搜索曝光、点击、浏览、收藏等行为数据，使用kafka+sparkstreaming实时处理落到HDFS上，使用hsql/spark sql + shell 进行批处理开发。使用kettle进行调度依赖，将计算结果推送到关系数据库。
3. 负责ETL流程维护以及优化，支撑BI系统及数据挖掘所需数据需求，负责从数据源到数据集市整条链路的打通及故障处理维护工作

**项目经验**

一．手机百度KPI监控系统

* 项目时间：2017年9月至今
* 项目描述：监控手机百度活跃/新增、用户时长以及用户留存的主要KPI指标，为高层决策以及业务分析提供可靠的依据。
* 数据日吞吐量：10PB
* 责任描述：

1. 数据仓库开发：依托于公司开发的bigdata大数据平台，通过hql结合 python、shell等脚本对原始日志解析入库，然后创建多个hql任务查询hive 表，经过日志表–>mart数据表–>udw大表得到数据结果，最后将kpi结果通过 sqoop插件将hdfs结果导入到palo(优化版的mysql)数据表，完成网络库端上用户数据分析和存储.
2. 日志反作弊策略优化：对日志中解析后的From,ip,query,url,ua五大特征，使用反作弊组提供的规则引擎与株连策略的API进行离线计算标记。再通过hql 进行merge入库。供下游表计算KPI使用。

二．数据智囊系统

* 项目时间：2014年9月 - 2017年5月
* 项目描述：敦煌网为卖家量身打造的数据展示和分析工具，可帮卖家时实监控自身店铺经营指标数据；多维解析行业发展趋势；深入分析买家购买行为；时刻更新买家最新搜索习惯等。
* 集群规模： hadoop机器数45个
* 数据日吞吐量：2TB左右，日任务job数10000个，核心任务job数4000
* 责任描述：

1. 将Flume实时采集用户行为数据，使用kafka+sparkstreaming实时处理落到HDFS上，
2. 经过ods基础层–>mds中间层–>sds应用层的ETL处理，将多维数据分析结果通过sqoop导入到Mysql中进行统一管理，ETL使用的hive sql/Pig 计算+shell脚本来实现的。计算近7天，30，自然月的商户流量以及交易情况，并且计算同环比指标。
3. 对整个ETL流程优化，保证数据实时准确完成处理。