**大数据开发方面技能：**

1、熟练使用Linux常用的操作命令，熟悉shell脚本编程

2、掌握hadoop体系架构，理解MapReduce的工作原理，以及HDFS原理，Yarn原理

3、熟悉Hadoop的调优

3、熟悉Hive和HBase的系统架构，能使用Hive进行海量数据的统计分析，能对Hive、HBase进行调优

4、精通Spark体系架构，熟悉SparkCore，SparkSQL，SparkStreaming，SparkMLlib，读过spark的部分核心源代码

5、熟悉Spark的调优

6、熟练掌握Storm集群架构及其工作原理，能够使用storm+kafka对数据进行清洗和实时统计分析

7、熟练使用flume和sqoop技术。做到日志采集分析，同步关系型数据库和hbase上的数据

1. 熟悉常用的机器学习算法(例如逻辑回归、朴素贝叶斯)，熟练运用Mahout，SparkMllib中提供的机器学习算法
2. 掌握常用的推荐算法（例如KNN、基于内容的推荐、基于隐语义模型的推荐），对推荐系统有较深的理解

**Java开发方面技能：**

1. 掌握JavaSE和JavaEE相关知识，并且熟悉常用的Java设计模式(例如工厂、单例、适配器等设计模式)
2. 熟练使用常见的JavaEE开源框架，如Struts2、Spring MVC、Spring、Hibernate、Mybatis
3. 熟练使用Oracle和MySQL数据库，熟练写出各自的分页语句，熟练编写SQL的增删改查语句
4. 深入理解java虚拟机

**工作性质**：全职 **目标职位**：大数据工程师 **目标地点**：北京 **目标薪资**：面议

**2014.06-2016.06 中国联合网络通信有限公司河南省分公司 Java大数据工程师**

工作描述：从事后台开发。参与Hadoop，Spark和机器学习相关大数据项目的开发

**2016.07-2017.12 去哪儿科技有限公司 Java大数据工程师**

工作描述：主要参与医疗大数据项目的需求分析和开发

xx

**性别**：男 **现居**：北京海淀

**学历**： 本科 **专业**：

**生日**：199  **英语**：CET-4

**电话**：（北京） **邮箱**：

PERSONAL RESUME

专业技能

工作经历

求职意向

**项目二**

**项目名称：**日志分析  
**软件架构：**Flume+Hadoop+Hive+Hbase+Mysql+Sqoop  
**开发工具**：IntelliJ IDEA

**开发时间**：2015.06-2015.12

**项目描述：**

通过分布式Hadoop集群处理网站产生的大量日志分析出网站的流览量PV、访客数UV（包括新访客数、新访客比例）、

访问的IP数、跳出率、平均访问时长/访问页数、重点用户等KPI供网站管理者分析网站使用

**责任描述：**  
 1、 主要负责MapReduce的开发

2、 参与数据同步的技术的研发  
 3、 参与在测试集群上调试项目  
**技术描述：**

1、使用Flume收集日志数据，通过HDFS把日志数据导入到Hdfs中

2、通过MapReduce得到想要的数据存储在Hdfs上

3、使用Hive进行数据的多维分析

4、把Hive分析结果通过Sqoop导入到MySQL中

5、提供视图工具供用户使用，指标查询MySQL、明细查询Hbase

**项目一**

**项目名称：**个性化推荐系统

**软件架构：**Flume+Kafka+HDFS+Hive+Hbase+SparCore+SparkMLlib+Redis+Dubbo

**开发工具**：IntelliJ IDEA

**开发时间**：2016.01-2016.06  
**项目描述：**

商都网有很多历史数据，包括用户数据、 用户浏览数据、 用户位置轨迹信息。这些数据极具价值，对其进行挖掘可以做到更好的用户体验。 故而决定利用这些数据做一个个性化推荐系统。

**责任描述：**  
 1、数据清洗

1. 将用户的历史数据转换成逻辑回归需要的训练集
2. 解决数据量过大，导致计算时间过长的问题
3. 推荐系统的调优：性能调优，算法调优

**技术描述：**

1、利用Hive进行数据清洗。

2、利用Spark将用户的历史数据构建为逻辑回归需要的训练集

3、利用sparkMLlib离线计算逻辑回归模型，Redis进行在线的topN的个性化推荐

4、针对推荐的准确率，召回率等指标对推荐系统进行算法调优和性能优化

PERSONAL RESUME

项目经验

项目经验

**项目四**

**项目名称：**管理系统

**软件架构：**SpringMVC+Maven+Mysql+Angularjs+BootStrap  
**开发工具**：IntelliJ IDEA

**开发时间**：2016.07-至今

**项目描述：**

现在医院在治疗病人同时已经开始重视病人的营养介入治疗的问题，包括新的三甲评审及医改当中也涉及到了营养介入治疗的要求，所以有条件的医院已经设立了相应的营养科室来进行相关的营养介入治疗。为了能够让营养科能够方便的监控病人的营养状况和具体的指标情况，因此根据临床营养的流程开发此系统辅助营养师监控病人的营养情况，为后期做临床营养医疗大数据做平台准备

**责任描述：**  
 1、负责临床营养辅助诊断系统的前后台开发，临床营养大数据的项目调研和需求分析  
**技术描述：**

1、后台使用SpringMVC处理业务逻辑

2、前台使用Angularjs，BootStrap进行前后台交互和展示

3、临床营养大数据的可行性分析

**项目三**

**项目名称：**交通智能化综合管控平台（二期）

**软件架构：**Flume+Kafka+Spark Streaming+HBase+Redis  
**开发工具**：IntelliJ IDEA

**开发时间**：2015.01-2015.05

**项目描述：**

随着城市交通信息化建设，信息采集设备的增加，交通基础数据量呈爆炸性增长，以地市为例，每月文本数据量达TB级。交通部门为实现对海量数据的采集、传输、存储、分析和挖掘，进而提高交通管理能力，提出此项目需求。项目二期，主要涉及以下模块：图像识别模块、实时路况模块、轨迹追踪预警模块、路况预测模块、交互查询模块

**责任描述：**  
 1、负责实时路况模块从Kafka导入数据，Spark Streaming计算，结果输出数据库的全部代码的实现，以及本模块项目的调试  
**技术描述：**

1、数据中心将图像数据转换成结构化的文本数据，存到HDFS中

2、通过Kafka 提供API将数据导入到Kafka中

3、通过Spark Streaming从Kafka获取数据，编写代码，实现实时路况的分析

4、将分析结果通过API写入HBase，并缓存到Redis中

5、前台获取Redis中的数据，展示实时路况

PERSONAL RESUME

教育经历

我的优势:

我的劣势:

我喜欢的书籍:

技术类：等

其他：《必然》

我喜欢的社区:

图灵社区

我对自己的定位:

我对技术的态度:

我怎样克服困难:

熟练地阅读源码、手册、规范

怎样保持自己的视野:

个人博客:

微信公众账号:

PERSONAL RESUME

自我评价

**2010.09—2014.7 大学 学士学位**