性别：男

地址：

邮箱：

年龄：岁

电话：

政治面貌：中共党员

学历：本科

求职意向：大数据开发工程师

工作经历

2017.7-2018.12 北京贝思科技术有限公司 大数据开发工程师

1. 负责公司海量日志数据进行批量、实时处理，针对不同的业务逻辑设计对应的解决方案；
2. 参与大数据项目数据开发工作，参与业务系统数据收集、处理、存储方案的设计开发；
3. 参与公司大数据基础设施的研发与运维，提升运行效率、稳定性和可用性；
4. 协助对业务数据进行分析，对大数据实施过程中相关技术问题解决；
5. 负责数据收集、清洗和规约等工作，完成数据指标的统计分析；

2016.10-2017.6 北京天亿电联科技有限公司 java大数据实习生

1. 协助大数据平台搭建及仓库建模；
2. 参与修改并优化spark、hadoop生态圈等组件的运行；
3. 利用分布式计算集群实现对数据的分析、挖掘、处理；

职业能力

1.具有扎实的Java、Scala

3.熟悉Hadoop集群安装配置及相关技术的使用，如：hdfs、mapreduce、yarn、kafka、flume、sqoop、hive

4.掌握Spark集群的安装配置并具有相应的使用经验，如：

3.熟练掌握flume、sqoop等数据传输框架；

6.熟练掌握javaweb开发，掌握Spring、。。。框架和springboot框架等主流开发框架，

熟练使用idea、maven、svn等开发及版本控制；

2.熟练使用Linux，以及Linux命令；shell基础

7.熟练掌握mysql、redis、hbase等主流关系型和非关系型数据库；

10.了解storm以及Flink进行流式实时数据分析；

11.了解python基础和python的基本使用；

13.接触过机器学习和深度学习，目前感兴趣的方向；

项目经验

项目一：沧州市车流量系统 2018.2-2018.8

项目架构：hadoop+flume+kafka+spark+hbase+redis+mysql

开发工具：Idea+Maven+svn

项目简介：

随着城市化的建设，信息采集设备的增加，交通基础数据量呈爆炸性的增长，以城市化为例，每月文本数据量达到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。主要实现以下模块：

1. 实时路况预测；
2. 跟车分析，计算出目标车是否被跟踪过；
3. 计算出经常高速通过的卡扣，可以在此安装违章拍摄设备。

此项目的实现可以分为：

1. 利用Flume从服务器上实时收集日志数据；
2. 采用Kafka技术，存储数据，缓冲并发的压力，使系统稳定运行；
3. 通过Spark Streaming从kafka中获取数据，进行计算处理；
4. 将分析结果写入HBase，并缓存到Redis中；
5. 前台获取Redis中的数据，进行分析结果的展示

职责简述：

1. 参与卡扣流量单模块需求调研、分析，业务模块划分，数据库设计；
2. 实现道路实时拥堵统计模块；
3. 实现卡扣状态监控模块和卡扣流量分析模块；

项目二：要出发旅游app日志信息分析系统 2017.7 - 2017.12

软件架构：Hadoop+Flume+Kafka+Hive+spark+Sqoop+Mysql

开发工具：Idea+Maven+svn+scala

项目描述：

本项目是为合作的企业进行日志分析处理的系统，项目主体分为3层架构，数据收集层，数据分析层和数据展示层，旨在构建智慧旅游的系统，本系统提供离线的数据分析，如指定路线日（月、年）客流量统计、指定交通方式月（年）人次统计、指定景区的市场分析；以及实时处理的用户推荐转化率；

1. 使用Flume收集各种服务器的数据，将数据实时传输到kafka，收集日志数据；
2. 基于Spark和hadoop构建一套HDFS+Hive+SparkSql的数据仓库；
3. 离线分析时通过mapreduce对数据进行清洗，然后使用Hive表和自定义UDF函数对数据进行离线分析；
4. 将分析的数据通过Sqoop导入到mysql中，以便展示；
5. 实时分析时通过kafka传输的数据直接对接SparkStream，通过算子将数据进行清洗，最后实时分析计算指标，将数据也写入mysql中；
6. 将mysql中的数据通过ssm和页面展示，便于使用者观看；

职责简述：

1. 参与项目整体架构的分析，以及业务的设计；
2. 负责处理离线数据部分指标的分析统计，如交通方式月（年）人次统计、成熟客源市场分析；
3. 负责flume收集日志信息，将日志信息push到kafka消息队列中；

项目三：亚太电商日志分析系统 2016.12 - 2017.04

软件架构：Hadoop+Flume+Hive+Sqoop+Mysql+SSM+echarts

开发工具：Idea +Maven+ svn

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、重点用户等信息供网站管理者分析网站使用，此项目中用到以下技术:

1. 使用Flume收集日志数据，将原始的数据保存到hdfs中;
2. 通过编写指定业务的MapReduce进行日志数据的清洗;
3. 将数据转化为结构化的Hive表，然后使用Hive外部表和自定义UDF函数对数据进行离线分析；
4. 使用Sqoop将Hive分析处理后的数据导入到MySQL中，以便Web server查看；
5. 数据展示，使用ssm框架和echarts将数据进行展示在页面上；

项目职责：

1. 参与前期项目分析，设计系统整体架构；
2. 负责用户基本信息分析模块的编码工作，如网站的PV、UV、转化率等指标；
3. 负责使用flume集群采集日志信息；

个人评价

1. 参与过多个项目的开发，熟悉常见的业务处理,能够快速的融入团队中；
2. 性格乐观向上，待人友善，能够和团队相处融洽；
3. 对待新技术有强烈的好奇心，喜欢钻研新技术，学习能够较强；
4. 对待工作负责，态度端正，能够按时完成领导交代的任务；
5. 个人身体素质还不错，有一定的抗压能力；
6. 热爱编程，热爱生活，愿意和同事功能进步；
7. 有不错的表达能力，可以及时的和领导同事进行沟通；