**XX**

27岁 | 男 | 统招本科 | 4年经验 | CET4、CET6

手机：

邮箱：

**个人优势：**

1. 做过多个项目，能对新项目快速上手
2. 热爱技术，工作认真、严谨，具备较强的学习能力和责任心。抗压能力强，能自我激励，善于沟通与团队协作
3. 具备扎实的 Java 相关知识，熟练使用 Java 和 Scala 语言编程
4. 掌握 Spark 及其组件 SparkCore、Spark SQL、SparkStreaming 的使用
5. 掌握 Hadoop 的分布式文件系统及其组件 HDFS、MapReduce、Yarn 的使用。熟悉 Hadoop 集群的搭建
6. 掌握 Kafka 原理，能够和 Spark、Flink 整合实现流式数据的处理和分析
7. 掌握 Hbase、Redis 等 NoSQL 数据库
8. 掌握 Hive 的工作原理，数据仓库的建立，以及使用 HQL 完成对数据主题抽取、多维分析、调优
9. 掌握 Linux 操作系统，可以编写 shell 脚本
10. 掌握 MySQL 数据库的使用及调优
11. 熟悉 Kettle、Sqoop、Impala、Azkaban 等大数据辅助工具的使用
12. 熟悉 Flink 的基本原理，能够使用 Flink 对数据进行实时处理
13. 了解机器学习算法库 SparkMLlib 以及机器学习的相关算法
14. 了解 Python 基础和 Python 的基本使用

**期望职位：**

全职 大数据开发工程师 北京 薪资面议

**工作经历：**

**单位名称：**

职位名称：大数据开发工程师

工作时间：2017年11月---2019年10月

工作描述：主要从事大数据开发工作，包括离线统计、实时计算

**单位名称：**

职位名称：大数据开发工程师

工作时间：2015年10月---2017年10月

工作描述：主要从事大数据开发工作，包括离线统计、实时计算

**项目经历**

**项目一：在线教育用户画像平台**

**项目架构** : SpringBoot + Vue + Flume + Hadoop + Hive + Spark + Hbase + Phoenix + Azkaban

**项目描述** :

随着互联网服务业务的蓬勃发展，教育信息化也越来越受到人们的重视。学习者面对庞大复杂的线上教育资源无从下手。该项目主要是针对用户访问在线教育网站时，通过点击的内容类别、购买的内容类别、在某一页面的停留时长，以及听课后对课程（老师）的评价等角度构建用户画像模型。基于大数据平台采集分析，分别从用户类别、订购内容、行为特征及业务场景等多方面进行数据标签配置，实现模型与应用场景数据共享，采用千人千面等方法进行 UI 数据可视化展现，实现精细化运营及精确营销服务。

**责任描述** :

* 参与项目的前期架构设计与分析
* 将 Hive 中的数据导入到 Hbase 中，实现解耦合
* 通过查询 MySQL 中的数据源信息，进一步获取 Hbase 中数据，再根据标签规则使用 Spark 完成部分匹配型标签的开发，如性别、民族、籍贯、政治面貌、职业、教育程度、就业状况等
* 完成如购课次数、听课频率、购课频率、消费频率、消费周期等部分统计型标签的开发
* 使用 SparkMLlib 根据业务规则完成部分挖掘型标签的开发。如 RFM 模型、RFE 模型和 PSM 模型，通过 KMenas 算法完成不同用户的消费能力，活跃度以及价格敏感的画像

**项目二：在线教育离线 + 实时数仓项目**

**软件架构**：Nginx + Flume + Hadoop + Kafka + Spark + Flink + Hbase + Redis + Superset + Azkaban

**项目描述**：

教育一直与数据密切相关，通过对遍布教、学、研多层面的数据进行整合与大数据技术的有效利用，可以从根本上给教育带来全方位的提升。这就需要一种能够将日常业务处理中所收集到的各种数据转变为具有商业价值信息的技术，而传统的数据库系统已经无法承担这一责任。本项目主要通过搭建离线数仓 + 实时数仓，将维度表放到Redis中，通过 Spark + Flink + Hbase 大数据技术完成了业务数据离线和实时数据统计及大屏展示。结合大数据的综合分析，可以优化招生、教学计划，可以帮助学生改善学习效率，提供符合职业规划的个性化学习服务。同时也能够为管理者提供数据支持和基于数据决策的信息，去帮助管理者更好的决策。

**责任描述**：

* 参与项目的前期架构设计与分析，得出运营及网站的关键性能指标
* 对于 MySQL 中的订单数据，负责使用 Kettle 对其进行抽取，完成基本的数据预处理，转化和装载到 Hive 中
* 负责在 Hive 数仓中使用 Spark SQL 进行分析，对业务场景不同维度的业务字段进行分析统计，包括课程种类、课程成交量、课程成交额，地区分布，教师排行等统计指标。使用 Sqoop 将数据导出到 Hbase 数据库
* 使用 Canal 服务器实时监听 MySQL 的 Binlog 日志，放到到消息队列 Kafka 中
* 使用 Flink 消费 Kafka 中的消息，完成实时 ETL 处理，并计算用户访问指标如 PV、UV、访客页面停留时间、平均访问频度、深度、时长，跳出率等指标，将结果数据存储到 Hbase 中

**项目三：电商平台离线数仓项目 + BI 展示**

**软件架构**：Nginx + Flume + Kettle + Hadoop + Hive + Spark + MySQL + Kylin + Superset

**项目描述**：

随着技术的飞速发展，经过多年的数据积累，互联网公司已经保存了海量的原始数据和各种业务数据，所以数据仓库技术是各大公司目前都需要着重发展投入的技术领域。数据仓库是面向分析的集成化数据环境，为企业所有决策制定过程，提供系统数据支持的战略集合。通过对数据仓库中数据的分析，可以帮助企业改进业务流程、控制成本、提高产品质量等。该项目主要针对各种原始数据进行分析，统计结果，最后展示出来，为决策者提供数据支持，为企业的走向提供方向。

1. 将 MySQL 中的数据使用 Kettle 进行 ETL，装载到 Hive 中
2. 使用 Flume 采集 Nginx 日志，下沉到 HDFS
3. 基于 Spark 和 Hive 构建一套 Hive + SparkSQL 的数据仓库
4. 通过 Spark 对数据仓库中的数据拉宽、指标计算、汇总等处理
5. 将分析的数据通过 Kettle 导入到 MySQL 中
6. 将 MySQL 中的数据通过 Superset 展示，便于管理者观看与决策

**职责简述**：

* 参与项目整体架构的分析，以及业务的设计
* 参与数仓的搭建
* 负责处理离线数据部分指标的分析统计，如活跃用户的统计、付费趋势的统计、采购流程的转化、新用户的留存等
* 协助 BI 工程师使用 Apache 的 BI 工具 Superset 进行数据的可视化展示

**项目四： 煤炭行业大数据分析报告**

**软件架构**： Hadoop + Hive + Sqoop + SSM + MySQL + Azkaban

**项目描述**：

该项目主要是对煤炭行业数据的综合分析，数据来自不同的行业，数据格式繁多，数据量大，公司的业务是对行业数据提取分析出有价值的信息，提供行业分析报告，例如分析化工、房产、钢铁、电网等行业煤电用量和消费信息，全国煤产量及消耗量，电厂数据，煤炭进出口数据，煤炭价格，主焦煤，动力煤运量，价格等，来自不同行业的数据经过 Hive 做 ETL 工作之后做具体的指标维度分析，存入数据仓库，Hive 中的数据通过Sqoop 导出到业务 MySQL 中，使用 JavaEE 相关技术实现报表展示。

**技术描述**：

1. 使用 Hadoop 作为大数据平台基础架构
2. 使用 HDFS 存储公司采集到的海量行业数据
3. 使用 MapReduce 对不同的数据源做预处理工作
4. 使用 Hive 构建数据仓库，通过 HQL 进行指标计算
5. 使用 Sqoop 将 Hive 中的数据导出到 MySQL 中，提供数据查询
6. 使用 SSM 框架搭建数据可视化平台

**责任描述**：

* 编写 MapReduce 程序对业务人员采集到的数据做预处理
* 将清洗后的数据上传到 HDFS
* 使用 Hive 做指标计算
* 参与数据可视化开发

**项目五：电商日志分析系统**

**软件架构**：Hadoop + Flume + Hive + Sqoop + Mysql + Echarts + Azkaban

**项目描述**：

通过分布式 Hadoop 集群处理网站产生的大量日志，来挖掘其中有用的数据信息。Web 日志包含着网站最重要的信息，通过日志的分析，我们可以知道网站的访问量，哪个页面访问人数最多，哪个页面最有价值等。进而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。该项目的主要目的是，将采集的用户行为信息在数仓中进行数据的处理，落地到 MySQL，最终为不同指标的数据查询提供支持，为企业决策提供依据。

**技术描述**:

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
5. 数据展示，使用 Echarts 将数据进行展示在页面上

**责任描述**：

* 参与前期项目分析，设计系统整体架构
* 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
* 负责使用 Flume 采集日志信息