

基本信息	
姓 名：	籍 贯：
性 别：	学 历：
年 龄：	专 业：
工作时间：4 年	毕业院校：
电 话：	英语水平：
居 住 地：	邮 箱：

求职意向		
期望职位：大数据开发工程师	期望地点：北京	期望薪资：面议

技术能力	
<div>1. 熟练使用 scala , java 语言编程</div> <div>2. 掌握 SparkCore,SparkSQL,SparkStreaming</div> <div>3. 深入理解 spark 工作机制。掌握 spark 任务的执行流程, 可以对出现的问题进行排查, 性能调优</div> <div>4. 理解 Hadoop 的分布式文件系统 HDFS</div> <div>5. 理解 MapReduce 原理, 实现编码</div> <div>6. 理解 Hadoop 的资源调度框架 yarn</div> <div>7. 熟悉 Hive 的工作原理, 了解数据仓库建立, 完成对数据主题抽取和多维分析</div> <div>8. 掌握 Kafka、Flume 数据采集工具的原理和使用</div> <div>9. 掌握 Kafka 和 Spark 的集成</div> <div>10. 掌握 SparkSQL 和 Hive 的集成</div> <div>11. 掌握 Hadoop 和 hive 的集成</div> <div>12. 掌握 Redis 内存数据库的基本原理</div> <div>13. 熟悉 java 基础, jsp ,servlet ,javascript ,Extjs ,POI , jQuery</div> <div>14. 熟悉 hibernate ,struts , spring ,ibatis , springmvc</div> <div>15. 熟悉基于 CXF , axis2 的 webService,</div> <div>16. 熟悉 mysql,oracle,sqlserver 数据库的使用</div> <div>17. 熟悉 Tomcat,Jboss Web 服务器的使用</div> <div>18. 能阅读英文技术文档。具备良好的文档写作能力</div> <div>19. 熟练使用 Linux 操作系统, 编写 shell 脚本</div>	

工作经历		
------	--	--

2014 /8-2017/8	XX 科技 (中国) 有限公司	大数据开发工程师(Spark/MapReduce)
----------------	-------------------	---------------------------

兼 Java 开发工程师

2013 /2--2014 /7 北京市 XXX 技术股份有限公司

JAVA 软件研发工程师

项目经验

项目一：电信 CRM 大数据项目

项目名称：基于 Spark 交互式的用户行为分析系统

开发环境：IDEA +maven+ Java-jdk + scala-sdk

系统架构：hadoop+ookeeper+flume+Spark+hive+nginx+kafka+mysql+Oracle

项目描述：

基于 Spark 交互式的用户行为分析系统主要是针对用户访问电信应用所产生的用户行为数据的分析系统。主要分为会话分析模块，跳出率分析模块，各个区域热门销售品和产品的统计模块，广告流量实时统计和订单监控模块。会话分析模块主要利用 SparkCore 进行用户访问 session 聚合统计，将结果保存到 MySQL。比如：总会话长度、总会话数量、总无效会话数量，分时间区间来计算各个区间的会话数量/会话长度/无效会话数量，按照时间比例抽取 session（数据抽样），获取点击、下单、支付次数排名前几的产品或者销售品（套餐或者可选包）。跳出率分析模块主要是 SparkCore 计算两个页面之间的转换率、页面的跳出率、整个系统的跳出率等。各个区域热门销售品和产品的统计模块主要利用 SparkSQL 统计每天每个区域热销的销售品和产品，卡等。广告流量实时统计和订单监控模块主要是用 SparkStreaming 对实时收集得到的用户点击广告数据，和订单数据进行实时数据分析，然后将结果保存到 MYSQL 中

责任描述：

1.使用 SparkCore 开发会话分析模块，这个模块主要是读取 Hive 分区表的数据进行会话分析，然后将分析结果保存到 MySQL 数据库中。主要分析指标有总会话长度、总会话数量、总无效会话数量，分时间区间来计算各个区间的会话数量/会话长度/无效会话数量，按照时间比例抽取 session（数据抽样），获取点击、下单、支付次数排名前几的产品或者销售品

2.使用 SparkStreaming 开发广告流量实时统计和订单监控模块,这个模块主要是从 Kafka 中实时拉取广告点击数据和订单数据进行实时的分析，然后将分析结果保存到 MySQL。主要功能有广告黑名单的过滤，黑名单的更新，实时统计每天、每个广告、每个省份、每个城市下的广告的点击数量，获取各个省份的热门广告数据。实时统计最近一段时间的广告点击量。实时统计每天每类订单的数量做订单监控，每个省份各种类型的单子数量

项目二：电信 CRM 大数据项目

项目名称：数据同步与稽核系统

开发环境：IDEA +maven+ java-jdk + scala-sdk

系统架构：hadoop+ookeeper+Spark+hive+mysql+Oracle+ftp+spring+springmvc+springQuartz+jquery

项目描述：

数据同步与稽核系统的主要功能是把关系型数据中的需要稽核的表的数据利用 SparkSQL 同步到 Hive 分区表(按省)中，然后利用 SparkSQL 对每个分区的数据进行数据稽核，将生成的稽核文件保存到 HDFS 上的不同的目录。然后利用 ftp 的 client 将生成的稽核文件上传到稽核平台，供对端稽核系统获取

数据进行稽核。数据同步与稽核系统的主要分为 任务调度，数据同步，数据稽核几个部分。

任务调度：用户根据不同的菜单创建不同的数据同步与稽核任务，系统将创建的任务保存到 oracle 中的 task_Info 报表中。当用户点击启动任务的时候，将任务动态的添加到 SpringQuartz 中的任务列表中进行任务调度。

数据同步：数据同步部分的主要功能是，根据用户创建的数据同步定时任务，利用 SparkSQL 将 Oracle 中的数据同步到 hive 分区表中，供后续稽核使用。用户行为分析使用。

数据稽核：根据用户创建的定时任务，利用 SparkSQL 将 Hive 表中的数据进行稽核任务，然后将稽核的结果以 text 文件保存到 HDFS 上的不同的目录下。然后后台的定时任务定时去指定目录下载文件，并根据规则重新命名，最终将结果上传到对端的 ftp 目录

责任描述：

1. 根据稽核业务的不同类型，梳理需要进行数据同步的表。比如：订单信息稽核，销售品月稽核，产品日稽核，卡资源稽核，产品月稽核等等
2. 在编写 hive 脚本模型
3. 分稽核类型开发 SparkSQL 稽核处理代码
4. 项目开发工程的框架搭建，以及公共代码的编写

项目三：集团 CRM 协议转换平台

项目名称：集团 CRM 协议转换平台

开发环境：myeclipse+oracle+maven+jdk+svn

系统架构：springmvc+spring+ibatis+java+extjs+axis (webservice) + redis

项目描述：

集团 CRM 协议转换平台分为核心和本地，核心主要负责集团向外提供服务，开放业务，做一些集团到省份和省份到集团的报文转换业务，所有外部系统想要接入集团业务，必须在核心模块注册，通过核心模块接入，用到的技术和中间件：Spring，SpringMVC，Ibatis，anaxis2,redis，Oracle，集团的一些框架和日志平台，集团 CRM 协议转换平台本地包主要负责集团向外开放接口业务，做协议转换业务，细化集团接口业务

责任描述：

集团 CRM 协议转换平台项目：负责集团进行服务能力开放，外围系统接入，和协议转换业务。这个项目我主要负责需求的接收，安排开发，和进度追踪，再就是相应的需求设计和开发（担任项目小组组长）。

项目四：联通一级电子运维系统

项目名称：联通一级电子运维系统

开发环境：myeclipse+oracle+maven+jdk+svn

系统架构：springmvc+spring+hibernate+java+extjs+cx (webservice)

项目描述：

联通一级电子运维系统 Ucloud 项目，此项目包含电路调度模块，局数据模块，重保管理模块，设备评价，宽带拆装等模块，这些模块为全国的联通公司服务。这个项目主要用到了 springMVC,Extjs Hibernate,普元的流程引擎，基于 Cxf 的 WebService.

责任描述：

先后参与了联通一级电子运维系统的电路调度模块，局数据模块，重保管理模块的研发，在电路调度模块中参与与一站式，集中集客的 ESB 接口开发，和所有和电子运维有关的 ESB 接口。在局数据，重保管理模块中负责前台页面设计到后台的数据存储的开发。

项目五：ATM 和公司 OA 加班调休管理系统

项目名称：ATM 和公司 OA 加班调休管理系统

开发环境：myeclipse+oracle+jdk+svn+mysql

系统架构：spring+springmvc+hibernate+mybatis+extjs+axis2

项目描述：

ATM 项目的主要任务就是利用 WebService 进行数据的采集和存储,定时将数据同步到相应的数据库（主要是 mysql 和 oracle）。加班调休管理系统是统计公司人员加班和调休状况的,员工可以在这个系统上提交加班申请，然后由上级领导审批。也可查看自己加班调休情况。

自我评价

参与过多个项目的开发，熟悉常见业务处理，能快速融入团队；

逻辑思维能力强，思路清楚，学习能力强，对新技术有着强烈的好奇心；

乐于与用户以及同事和领导沟通，以便快速解决项目遇到的问题；

具有良好的英语阅读能力，能阅读英文资料、技术文档等；

喜欢这个行业，并打算长期做下去。