简历

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 姓名 | 张三 | | 年龄 | 28 |  |
| 性别 | 男 | | 民族 | 汉 |
| 籍贯 | 江苏常州 | | 学历 | 本科 |
| 英语能力 | 英语四级 | | 专业 | 计算机工程 | |
| 工作经历 | xxxx.xx – xxxx.xx xxxxxx公司 软件工程师  xxxx.xx – xxxx.xx xxxxxx公司 大数据工程师  xxxx.xx – xxxx.xx xxxxxx公司 大数据分析师 | | | | |
| 个人技能 | 1）熟悉Hadoop生态圈相关技术，如Hbase，Hive，Zookeeper，Sqoop等，有着丰富的故障处理和运维经验。  2）熟悉数据仓库的模型和设计架构，了解数据仓库建模阶段划分  3）熟悉Spark运行机制，研读Spark streaming部分内核源码  4）熟悉Linux系统的基本操作和集群环境的搭建和部署  5）熟练掌握Java、Scala等编程语言，熟悉程序设计和架构的基本思想  6）具备良好的需求调研、需求分析、模块设计、系统设计、开发流程、技术文档编写能力，英语六级，可比较熟练阅读英文技术文档、  7）使用python  8）熟练运用eclipse、Intellij IDEA开发工具进行大数据项目开发和调试  9）熟练使用maven | | | | |
| 项目经验 | | | | | |
| 项目一 | 项目名称： | **中搜聚力联盟Plus广告换量系统** | | | |
| 项目描述： | 聚力移动联盟是中搜面向APP应用开发者提供的，以品牌互推、流量互导、广告分账为主的一站式服务平台。 | | | |
| 个人职责： | 1. 参与需求分析、技术方案选型以及集群搭建及日常维护 2. 用Flume周期性收集日志 3. Kafka Consumer进行分析过滤,将分析过的数据存放到HDFS上 4. 通过Sqoop把数据导入到Hive中并用Hive SQL对数据进行分析、最后使用Sqoop将数据导入到Mysql数据库存储。 | | | |
| 技术要点： | 1. 使用分布式Kafka集群处理对Flume获取的数据信息（包括但不限于CPM千次展示、CPC千次点击、CPV富媒体展示等） 2. 通过MapReduce编程实现对数据进行过滤,后再存储到HDFS上。 3. 采用Spark Streaming对用户积分进行实时计算结果存入Hbase数据库中 4. 使用Spark SQL进行数据的多维分析并把分析结果导入MySQL中进行统一管理。 5. 对于报表我们则采用了Hive进行查询将需求信息反馈给用户系统，并在web页面以图表的形式展现给客户。另外报表中常常会设计很多复杂的查询为了能更好更方便的查询出需求信息我们采用了Hive SQL+MapReduce这样可以满足对复杂报表进行处理。 | | | |
| 项目二 | 项目名称： | 中搜云商城大数据推荐系统 | | | |
| 项目描述： | 该项目是分布式大型综合性的B2C移动电商平台，类似京东商城、天猫商城等。依托中搜搜悦移动互联网生态圈，辐射全国十几家分公司周边商户，致力于供应商与卖场分离，无缝对接京东自营、当当网所有商品，为广大消费者提供一站式移动购物平台。推荐系统是云商-电商大数据技术部核心业务系统，通过数据挖掘以及机器学习两种技术手段对用户进行个性化的推荐，以帮助云商城为顾客购物提供完全个性化的决策支持和信息服务。 | | | |
| 个人职责： | 参与大数据平台的搭建、优化及计算性能提升，负责数据仓库模型的设计与开发工作，协助数据分析体系的数据建模、统计。 | | | |
| 技术要点： | Nginx集群会收集用户的行为事件（我们以浏览某件商品为例），将数据发送到 Kafka消息队列，后端的Spark Streaming实时服务会从Kafka拉取数据进行消费，将数据读出来并进行实时计算分析，将结果写入Redis，可以实时获取用户的行为数据，同时将原生数据存入HDFS备份，周期性创建Hive分区表，再将原生数据导入Hive数据仓库中，以供离线综合统计分析 | | | |
| 项目三 | 项目名称： | 某电信公司通话记录改造 | | | |
| 项目描述： | 将通话记录数据由原来的oracle系统改造成使用大数据架构解决方案。主要使用hbase做通话数据的存储方案。需要将原有oracle数据导入到hbase中，以及新生成数据通过flume收集到kafka，再通过消费者存储到hbase数据库。 | | | |
| 个人职责： | 参与项目的改造设计  按照文档要求，实现hbase协处理器处，以及rowkey的盐析处理  按照部分统计分析设计要求，实现业务代码，例如按季度、年份、月份实现通话数据量的查询分析统计 | | | |
|  | 技术要点： | 使用flume实时收集生成的通话数据  将flume收集的log输入发送到kafka集群  使用kafka消费者程序抽取通话记录到hbase表中  将Spark部署成分布式查询引擎，使用thrfitserver服务器完成和web前端交互，提高在线分析的响应事件。 | | | |
| 自我评价 |  | | | | |