李宜润

18103966973 376942103@qq.com

基本信息

籍　　贯：河南省 驻马店市　　政治面貌：中共党员
英语水平：CET-6　　学历：工学硕士
目标岗位：大数据研发工程师 、spark 工程师、Hadoop 工程师

教育背景

2014.09-2017.07	中国石油大学（北京）	地质资源与地质工程	工学硕士
2010.09-2014.07	长江大学	地质学	理学学士

科研成果

以第一作者发表中文核心（科学技术与工程）论文 1 篇，参与发表中文核心 1 篇（四作）。
以第一作者发表会议（第四届非常规油气地质评价学术研讨会）论文 1 篇。

工作技能

理解 hdfs 分布式文件系统存储结构和高可用原理；
熟悉 Zookeeper 分布式服务框架，理解 HA 高可用集群；
掌握 hadoop mapreduce 计算框架编程，对 yarn 的资源调度，作业监控有一定认识；
熟悉 hive 数据仓库工具及 HQL 的书写，能对日志数据进行查询，统计等数据操作；
熟悉 linux 系统，了解常用的 linux 的 shell 命令，能在 linux 系统下搭建开发环境；
理解面向对象设计思想，能够阅读 Java 代码；
熟悉 kafka、flume 数据采集工具的使用，实现流式数据的过滤和分析；
能阅读英文技术文档。具备良好的文档写作能力；
理解 Hbase 的存储原理，Hbase 存储架构，实现数据的毫秒检索；
了解 Spark 相关组件，了解 Storm 运行流程；
熟悉 Python、Scala 语言编程，能运用 Scala 进行 spark RDD，spark streaming 编程。

项目经历

项目名称：纸牌比大小游戏
开发环境：eclipse+jdk
项目描述：先创建一副扑克牌，遍历这副扑克牌将每张牌的花色和大小打印出来，然后对这幅扑克牌进行洗牌，再创建两个玩家，由用户输入，玩家 ID 和姓名。每个玩家发两张牌，比较两个玩家手中最大手牌的大小，哪个玩家最大手牌的点数最大就获胜，点数一样的情况下，按照花色黑红梅方的顺序判定大小，最后打印出两名玩家的手牌。

项目名称：某视频网站运营指标分析项目
开发环境：eclipse+maven+jdk+linux
系统架构：hadoop+zookeeper+hive
需求描述：统计某视频网站的常规指标，各种 TopN 指标：视频观看数 Top10；视频类别热度 Top10；视频观看数 Top20 所属类别包含这 Top20 视频的个数；视频观看

数 Top50 所关联视频的所属类别 Rank；每个类别中的视频热度 Top10；每个类别中视频流量 Top10；上传视频最多的用户 Top10 以及他们上传的视频；每个类别视频观看数 Top10。

项目描述：项目源数据是两个文件，一个是视频表，字段有视频的 ID 标识、视频上传者、视频的类别、视频的观看数、视频流量和视频相关视频的 ID 等。另一个表为用户表，字段有上传者的用户名，上传的视频数等。先使用 MapReduce 对视频表中的数据进行清洗，剔除不合要求的数据。再根据不同的需求，通过 Hive，使用 Hql 统计出各种 TopN 数据。

项目步骤：1、通过 MapReduce 对原始数据进行清洗，生成规范数据文件上传到 hdfs；
2、然后使用 Hive 对数据进行多维分析；
3、再把 hive 分析结果使用 Sqoop 导出到 Mysql 中。

项目名称：Spark Streaming 实时流处理日志项目

开发环境：IDEA+maven+JDK+linux

软件架构：hadoop+zookeeper+flume+ kafka+ Spark+hbase

需求描述：实时（到现在为止）的日志访问统计操作

项目描述：项目数据来源的日志为 Python 脚本产生的，通过 crontab 定时执行 Python 脚本模仿服务器日志的产生，日志包括 ip、time、url、status、referer。然后使用 flume 采集产生的日志数据并 sink 到 Kafka 消息队列中，然后将日志信息传给 Spark Streaming 进行实时数据处理。最后将计算结果写入到 hbase 上。

项目步骤：1、通过 Python 脚本模仿日志的产生；
2、Flume 的选型，在本例中设为 exec-memory-kafka；
3、打开 kafka 一个消费者，再启动 flume 读取日志生成器中的 log 文件，可看到 kafka 中成功读取到日志产生器的实时数据；
4、让 Kafka 接收到的数据传输到 Spark Streaming 当中，这样就可以在 Spark 对实时接收到的数据进行操作了；
5、Spark 中对实时数据的操作分为数据清洗过程、统计功能实现过程两个步骤。其中统计功能的实现基本上和 Spark SQL 中的操作一致，体现了 Spark 的代码复用性，即能通用于多个框架中。
6、计算结果写入到 Hbase。

自我评价

- 乐于沟通，能快速融入团队，具备团队合作精神；
- 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；。
- 对工作尽职尽责，乐于从事有挑战性的工作；
- 具有良好的英语阅读能力，能阅读英文资料、技术文档等；

个人主页

个人主页：http://Larry-Arun.github.io/

————态度决定人生，细节决定成败！