

李宜润

17838190419 376942103@qq.com



基本信息

英语水平: CET-6 政治面貌: 中共党员
计算机等级: 二级 (C++) 学历: 工学硕士
期望岗位: 大数据开发工程师、Java 开发

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。

工作经历

2017.07-2018.07	华图教育	数据开发部	大数据开发
2018.08-至今	新华三大数据技术有限公司	中台开发部	大数据开发、Java 开发

工作技能

熟练使用 Java 编程语言, 理解面向对象设计思想;
熟悉常用的 linux 命令, 能在 linux 系统下搭建应用环境;
熟悉 shell 脚本, 能在原有 shell 脚本基础上发现 bug 并修改;
熟悉 kafka 消息队列的使用, 实现流式数据的过滤和分析;
熟悉 springboot 及 mybatis 组件的使用。
熟悉 idea、eclipse 编程工具的使用。
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;
理解 hdfs 分布式文件系统存储结构和高可用原理;
理解 Zookeeper 分布式服务框架, 理解 HA 高可用集群;
理解 Hbase 的存储原理, Hbase 存储架构;
理解 hadoop mapreduce 计算框架编程, 对 yarn 的资源调度, 作业监控有一定认识;
了解 Spark 相关组件, 了解 Python、Scala 编程语言, 了解 spark RDD, spark streaming 编程。
能阅读英文技术文档。具备良好的文档写作能力;

项目经历

项目名称: 安徽农信双集群 HA (2019.07—2019.09)
项目描述: 安徽农村信用社在使用 H3C 行业建模引擎内容管理模块近一年时, 提出主备双集群两中心的需求, 为提高集群稳定性和高可用, 设想再搭建一套集群作为备集群, 在主集群因意外原因停止服务时可以对外提供服务。项目大体步骤为: 客户点击操作将数据存储到 Hbase 和 ES 的同时, 将请求转化为元数据写入 PG 数据库中, 后将该元数据发送到 kafka 集群中, 再通过消费者消费 kafka 集群中的元数据, 解析提取后, 去相应 hbase 表中获取数据, 最后使用 feign 构建内部请求调用备集群接口将数据同步到备集群中。项目补充说明: 程序中的三个定时任务, 考虑到

———态度决定人生, 细节决定成败!

单次数据消费失败的情况，程序会启动一个定时任务来重发未被消费成功的数据到 kafka 中；考虑到数据写 ES 是一个异步的过程，可能会有失败的情况，故启动一个定时任务将同步 ES 失败的数据读取后再次调用 dataSearch 相关接口；定时删除任务是在数据完全被同步之后，将数据从 PG 数据库中删除。

责任描述：负责 kafka 发送者和消费者的开发工作；

负责元数据的解析和 feign 的调用；

负责定时补偿发送等三个定时任务的开发工作；

负责应用环境的搭建部署、安装脚本修改和功能测试工作；

项目名称：Elasticsearch 中间件（2019.03—2019.06）

项目描述：该项目实现了 ES 的全文检索功能，用户上传元数据和文档之后可以根据相应字段和文档内容进行模糊搜索。用户上传元数据和文档时，会先存到 Hbase 中，同时构造元数据写入到数据库中，另起 kafka 发送者线程读取数据库中的元数据发送到 kafka 中，消费者从 kafka 中读取元数据，解析后去 Hbase 中取相应元数据和文档存储到 ES 中，文档采用 pipeline 方式进行传输。该中间件给接口调用方提供 client 包，客户端和服务端采用 GRPC 框架通信。

责任描述：负责元数据和文档的上传、更新和删除接口的开发工作；

负责 kafka 发送者和消费者的开发工作；

负责 snake 项目的安装部署脚本的开发工作。

项目名称：Elasticsearch 中间件（2018.12—2019.02）

项目描述：该项目通过 ES 完成 Hbase 的二级索引功能，增加 Hbase 进行模糊搜索的能力，将 Hbase 中需要进行索引的列存储到 ES 中，其中 Hbase 的 rowkey 作为 ES 的 id，搜索时先查 ES，将查到的索引名、id 和需要的返回列一起返回，再拿返回的 id 作为 rowkey 去相应的 Hbase 表中查找数据，项目给调用方提供 client 包，调用方通过调用 client 包的相关接口实现需求，客户端与服务端采用 GRPC 框架进行通信。

责任描述：负责对 condition 进行解析，调用方会将 sql 解析为不同的 condition，在查 ES 之前需要将 condition 解析为相应的 querystring；

负责与 ES 交互，通过 querystring 查询方式获取结果。

项目名称：Oozie 中间件（2018.10—2018.11）

项目描述：Oozie 工作流的组装需要繁多的配置，启动调度也只能通过命令行或者直接调用 API 接口，无法做到作业权限的管理，内置作业模板等，但在管理具有复杂逻辑的多 Hadoop 任务方面具有丰富功能的优势。为了屏蔽它在使用上的复杂性，帮助用户快速构建应用，开发一套基于 Oozie 的轻量级任务管理中间件(workflow)来提供更便捷的操作。Oozie 中间件包括 web 前端、web 后端和服务端后端，web 后端使用 springboot 和 mybatis 开发，web 后端与服务端后端采用 GRPC 框架进行通信；用户可以通过界面来上传作业（Java、FS、Sqoop、Hive、Hive2、Shell、MR、subworkflow），并为作业创建工作流和为工作流创建相关调度。项目设置了若干内置作业和内置工作流（将 Hbase 的数据或者 HDFS 上的数据直接拷贝到 ES 中），用户直接在页面填入相关参数点击运行即可到达目的。

责任描述：负责任务调度的 web 后端开发、内置作业及内置工作流的维护工作。

———态度决定人生，细节决定成败！

项目名称：H3C 行业建模引擎数据源模块（2018.08—2018.09）

项目描述：H3C 行业建模引擎包括数据源、数据集和内容管理等模块，数据源模块负责连接多种数据源（MySQL、PG、GP、Oracle、Hive 等），拼接多种数据源不同的 sql 写法，对外提供统一的 API 供使用方调用，支持不同数据源的数据提取为数据集及不同数据源之间表的拷贝等。

责任描述：负责开发及维护数据源模块（DataSource）的功能，修改测试提交的问题单。

项目名称：华图教育用户画像系统（2018.02—2018.07）

项目描述：该项目主要基于 hive 数据仓库之上的数据分析展现模块，其中主要用用户基础信息画像，包括：年龄，地域，性别，职业等基础信息画像。不同网站的流量趋势画像，其中包括：访问趋势，新增访问客，活跃访客，访问量（每日，每周，每月）。访问画像：地域分析，客户端环境，设备属性画像，移动终端，网络连接画像（运营商）。根据画像推送相关的招考信息、网络课程信息、分校信息及考试题库等。

责任描述：负责参与项目 hive 相关表结构设计；

负责定时脚本任务，使用 hive sql 对数据进行分析统计；

负责 hive 的优化，对于复杂逻辑采用 MapReduce 进行统计分析。

项目名称：华图教育网站数据分析统计（2017.07—2018.01）

项目描述：该项目主要对华图教育网站的访问日志进行分析，计算网站关键指标，其中关键性指标有，浏览量（PV）、注册用户数、独立 ip 数及跳出率（跳出数/PV）等，跳出率是非常重要的访客黏性指标，它显示了访客对网站的兴趣程度，跳出率越低说明流量质量越好，这些访客越可能是网站的有效用户（购买课程）。通过对用户点击行为的分析，向用户推送合适的课程，提高有效用户的转化率。项目使用 Flume 监控日志文件，采集用户点击行为产生的日志 sink 到 kafka 集群，Kafka 消费者消费消息队列中的数据，并将数据写入到 HBase 中，通过 mapreduce 对数据进行清洗，使用 hive 进行数据的多维分析，把 hive 分析结果使用 sqoop 导出到 mysql 中从而进行前台展现。

责任描述：负责开发 kafka 消费者将数据写入 hbase；

负责写 mapreduce 任务对数据进行清洗；

负责使用 Hive sql 分析统计日志数据，并将结果导出到 MySQL 中。

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；

———态度决定人生，细节决定成败！