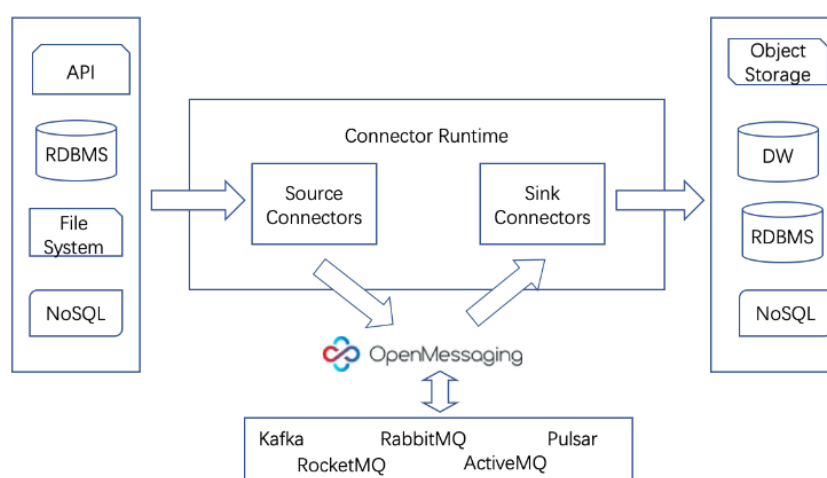


# Apache RocketMQ Connect Apache Doris 实现

## 第1章 介绍

### 1.1 RocketMQ Connect

消息路由主要用于不同地域之间消息的同步，保证地域之间的数据一致性和复制备份。RocketMQ 的消息路由特性打通了多个集群之间的消息互通，高效地实现了分布在不同地域之间的 RocketMQ 集群的消息同步复制。



Messaging Connector 架构图

Source Connector 负责从其它系统获取数据。

Sink Connector 负责从 Producer 中消费消息，将数据写入到另外的系统。

Runtime 是 Source，Sink connector 的运行时环境，负责加载 Connector，提供 RESTful 接口，启动 Connector 任务，集群节点之间服务发现，配置同步，消费进度保存，故障转移，负载均衡等能力。

OpenMessaging 是一套消息中间件领域的规范。OpenMessaging Connect 是 Connect 方面的 api，OpenMessaging 还有其他各种消息中间件领域的规范，例如 mq 客户端规范，存储规范，实现 OpenMessaging 规范可以做到厂商无关。具体的 Connector 可以基于这套规范实现。

### 1.2 Apache Doris

Doris 是一个基于 MPP 的交互式 SQL 数据仓库，用于报告和分析。它的原名是 Palo，是在百度开发的。捐赠给 Apache 软件基金会后，更名为 Doris。

Doris 提供高并发低延迟点查询性能，以及即席分析的高吞吐量查询。

Doris 提供批量数据加载和实时小批量数据加载。

Doris 提供高可用性、可靠性、容错性和可扩展性。

Doris 的主要优点是简单（开发、部署和使用）并在单个系统中满足许多数据服务要求。

---

## 第2章 项目详细方案

本项目旨在根据 Connect Api 实现 RocketMQ 到 Apache Doris 数据同步的 Connector。

项目期间，我将会参考 Kafka 和 RabbitMQ 的设计思想，同时学习 Apache RocketMQ Connect，详细了解 RocketMQ Connect 原理、使用、参数等实现 RocketMQ 到 Apache Doris 数据同步的 Connector。

## 第3章 项目开发时间计划：

A. 第 1 周（2022 年 7.1 - 5 日）

- 学习 Apache RocketMQ Connect，详细了解 RocketMQ Connect 原理、使用、参数等

B. 第 2-3 周（2022 年 7.6 - 19 日）

●根据 Connect Api 实现 RocketMQ 到 Apache Doris 数据同步的 Connector，同时积极与导师进行沟通及汇报任务。

C. 第 4-5 周（2022 年 7.20 - 7.31）

- 开始编制中期报告。
- 编写计方案文档

E. 完成（2022 年 7.31 -8.28）

- 测试功能
- 提交代码
- 优化代码

H. 闭幕（2022 年 9 月 1 日）

- 可交付成果
- 在社区中提交 Issue，完成相关代码

## 第4章 个人简历

我是李秋亮，上海海事大学计算机专业的研一学生。这是我第二次参加开源活动。我一直热爱开源，崇拜每一个为开源而无私奉献的人，自己也常在博客、Github 发表自己的项目及技术总结。同时，我对编码充满热情，力扣刷题坚持了两年。

我选择 RocketMQ 是因为本身对 Java 语言及生态圈组件足够熟悉，同时对 Kafka, RabbitMQ 非常熟悉，通读过经典书籍，这段时间也一直在学习 RocketMQ。以下是我的详细个人简历。

---

# 李秋亮

联系电话: 15687176251

电子邮箱: 779147708@qq.com

博客: [https://blog.csdn.net/weixin\\_44484668](https://blog.csdn.net/weixin_44484668)

## 教育经历

上海海事大学 (硕士)	计算机技术	2021.09-2023.06
新生入学奖学金、中国研究生数学建模竞赛三等奖、PAT 程序测试能力甲级竞赛		
西南林业大学 (学士)	木材科学与工程	2015.09-2019.06
top1 (202)、托福 (85)、GRE (149+167+3.5)、国家奖学金、国家励志奖学金、省级优秀毕业、省三好		

## 专业技能

1. 熟练掌握 Java 基础, socket 网络编程, 常用集合及源码, 熟悉常见设计模式 (单例模式, 工厂模式)。
2. 熟悉 JVM 类加载机制, 内存区域及垃圾回收机制。
3. 熟练使用 MySQL, 熟悉 MySQL 索引, 事务, 存储引擎, 锁机制, MVCC。
4. 熟悉 Redis 底层数据结构, 了解源码, 了解网络模型, 内存策略。
5. 熟悉 Kafka 原理, 了解源码, 了解 RabbitMQ。
6. 熟悉数据结构与算法, 操作系统原理, 常见网络协议。
7. 熟悉 Java 高并发, Synchronized 锁升级, Lock, 信号量机制, Volatile 及 CAS 原理, 死锁, 线程池。
8. 了解 SSM, SpringBoot 等常用框架, 目前在研究 Spring 源码。

## 实习经历

上海慧泉软件有限公司	2021/05—2021/9	Java 开发
<ol style="list-style-type: none"><li>1. 按需求手册, 在内部框架 (公司内部整合了 Spring 和 Hibernate 的框架) 上进行功能开发, 负责项目为各大高校财务管理系统中的日常报销模块。</li><li>2. 参与 2021 年上海成人自考正式报名系统及上海高考模拟报名系统的维护, 接触到了高并发场景。</li><li>3. 在 SQL Server 后台编写 SQL 语句对业务进行验证, 复现客户 (高校) 使用过程中出现的问题并 debug。</li><li>4. 收获: 经历了从新人到带新人的阶段; 对于上线项目从 0 到 1 的开发、测试、运维有了自己的理解。</li></ol>		

## 项目经历

Java 分布式卡包系统	SpringBoot+Kafka+Redis+MySQL+Hbase
<ol style="list-style-type: none"><li>1.功能描述: 实现仿小米卡包的一个在线卡券存放中心。商户投放卡券和 Token, 用户领取使用。</li><li>2.项目特点: 1) 考虑到项目解耦, 分为商户和用户两个子系统。2) 考虑到未来数据量增长, 用户数据存放在 HBase 并使用 Kafka 进行系统间通信。3) 为提高 IO 性能, 将频繁用于验证信息的 token 存放到 Redis 中。4) 为了简化任务量, 使用 Postman、TextLab 进行接口单元测试。</li><li>3.收获: 了解到小米卡包的业务流程, 明白了 SpringBoot 在项目开发中的自动化配置作用, 引起了我对 Kafka, Redis, MySQL 底层学习的兴趣。</li></ol>	

网站维护+数据仓库	SpringBoot+Redis+Kafka+Flume+Hadoop
昆山人才网: <a href="http://www.kshr.com.cn/">http://www.kshr.com.cn/</a>	
<ol style="list-style-type: none"><li>1.功能描述: 项目分为两部分, 昆山人才网的维护及数据仓库的搭建。ES 实现模糊查询, Redis 缓存数据, Logstash 同步 Sql server 和 ES 数据。网站埋点数据采集到 Hive 仓库中进行数据分析, 结果展示。</li><li>2.项目特点: 1) 考虑到大数据量, 对 Kafka 进行压测后部署到 3 台服务器, Flume 使用 Kafka sink。2) 负责搭建日志采集平台, 流程: 日志服务器-&gt;Flume-&gt;Kafka-&gt;Flume-&gt;Hadoop-&gt;Hive-&gt;MySQL(结果展示)。</li><li>3.收获: 学习到如何保证项目实现高可靠, 高并发, 高可用; 技术选型的重要性; 团队协作的重要性。</li></ol>	