

分布式数据库项目

——大规模信息系统构建技术导论课程项目

个人设计报告——王伊斓

指导教师：鲍凌峰

组长：马良

组员：王伊斓、罗依民

日期：2023 年 5 月 21 日

1 引言

1.1 系统目标

本项目是《大规模信息系统构建技术导论》的课程项目。项目内容为：在大二春夏学期《数据库系统》课程的基础上，结合《大规模信息系统构建技术导论》所学知识，实现一个分布式关系型简易数据库系统，实现数据分布、集群管理、分布式查询、副本管理、容错容灾、负载均衡等功能。

1.2 设计说明与任务分工

本系统由小组 3 位成员合作编写完成,使用 Java 开发,VScode 作为集成开发环境, Github 进行合作开发，每个组员都完成了目标任务，具体的分工安排如下：

表格 1 成员信息及分工

成员姓名	学号	分工职责
马良	3200105143	Master Server、Client、Region Server 之间的 Socket 通讯 Client、Master Server 代码 最终程序的集成与功能调试 视频录制
罗依民	3200105307	Region Server 代码 最终程序的集成与功能调试 视频录制
王伊斓	3200105318	Master Server 中的 Zookeeper 部分，负载均衡，FTP 操作 报告撰写、汇报 PPT 制作

2 个人负责模块设计

2.1 Master Server 中的 Zookeeper 部分

Master Server 中的 Zookeeper 主要负责完成 Master Server 中的信息管理功能，包括监听管理 Region 节点，Region 节点上线目录，维护管理节点与对应 ip、对应表的信息。

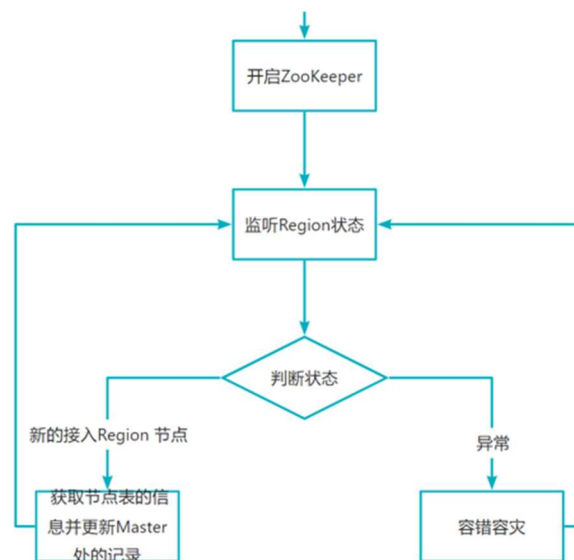


图 1 Master Zookeeper 工作流程

2.2 负载均衡

负载均衡指的是：接收 client 创建表的指令信息，根据当前各节点的使用情况，返回合适的 Region Server ip。选取最佳 Region 的方法为查看 Master 处维护的 Region Server 与所属表的结构，查找表数目最小 Region 并返回 IP，让 Client 与该 Region 建立连接。

2.3 FTP 操作

FTP 操作主要运用于 Region Server 处数据库信息文件的备份。

对于数据表的增加或修改，Region Server 进行相应处理后，应该将处理后的数据文件进行备份。因此引入了 FTP 上传操作，参数即为文件名称。

FTP 使用的另一个场景为容错容灾，当检测到一个 Region Server 异常，应调用负载

均衡，找到最空闲的 Region Server，并将异常 Region Server 在 FTP 中的数据备份文件下载到负载均衡中所选择的 Region Server。因此引入了 FTP 下载目录中所有文件的操作。

3 个人总结

本次实验中基本功能都已经实现，但部分功能模块还能够进一步完善改善，如副本管理时，仅使用了一个 FTP 服务器，而没有考虑到该 FTP 服务器出现错误运行状态后的处理方式；使用更大规模的测试数据也是测试系统性能的一个重要方法。

此次项目中，我熟悉了 Java 项目的结构，锻炼了编码能力，同时在撰写总体设计报告时，增强了对项目以及分布式数据库的理解，收获颇多。