

大数据系统软件 Lab2 实验报告

组员

- 陈荣钊 2018010863
- 伍冠宇 2018010683
- 张后斌 2018010858

实验内容

在 Lab1 的基础上，实现分布式系统中对表的自然连接功能。

代码架构

Cluster.Join

Cluster.Join 依据表的输入顺序，自左向右依次进行 Join。

Join 使用 semi-join 算法执行，master 节点先计算出两表的外键，随后提取出两表的外键和行号计算 Join，最后将得到的行号用于提取原数据并组合。

附加功能

多表连接

多表连接的测试代码见文件 `lab2_multi_table_test.go`。

测试文件将表 `student` 进行混合分区，以验证 DDBMS 的鲁棒性

多外键连接

多外键连接的测试代码见文件 `lab2_multi_foreignKey_test.go`。

测试文件将两张表的每一列都存在不同分区，以验证 DDBMS 的鲁棒性