

大数据系统软件 Lab1 实验报告

组员

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

实验内容

根据框架提供的分布式数据库系统，实现该系统对一组节点集群上的关系进行分区的功能。每个分区都是原关系的子集，每个节点只负责几个分区的放置和获取。根据分区规则将一张数据规模较大的关系表结构划分为若干个分区关系表结构。对于集群写入的每一条数据，根据分区规则将其划分到若干个节点的不同分区关系中，以完成分区功能。

代码架构

Cluster.BuildTable

通过解析 json 格式的数据，得到原表名以及每个节点创建的子表的列，再得到每列的所有 predicates。在解析过程完成后，通过新增的 Node.CreateTableRPC 接口将建表模式传给对应节点

Cluster.FragmentWrite

直接将待写入的数据行通过新增的 Node.InsertRPC 接口进行传输，在节点处进行 predicate 检查，并提取出该节点需要写入的列进行写入

亮点

实际运用场景下，常常无法预知用户行为，这些行为可能对系统造成危害。因此我们小组构想了许多可能出错的场景，并在代码中添加了对应的报错信息，增强系统鲁棒性。

此外，用户可能使用 TypeString 输入一系列 TypeBoolean 的数据。因此我们小组在本次实验中配备了完善的输入类型转化机制，会自动转换输入的数据类型，并写了测试代码 lab1_transform_test.go