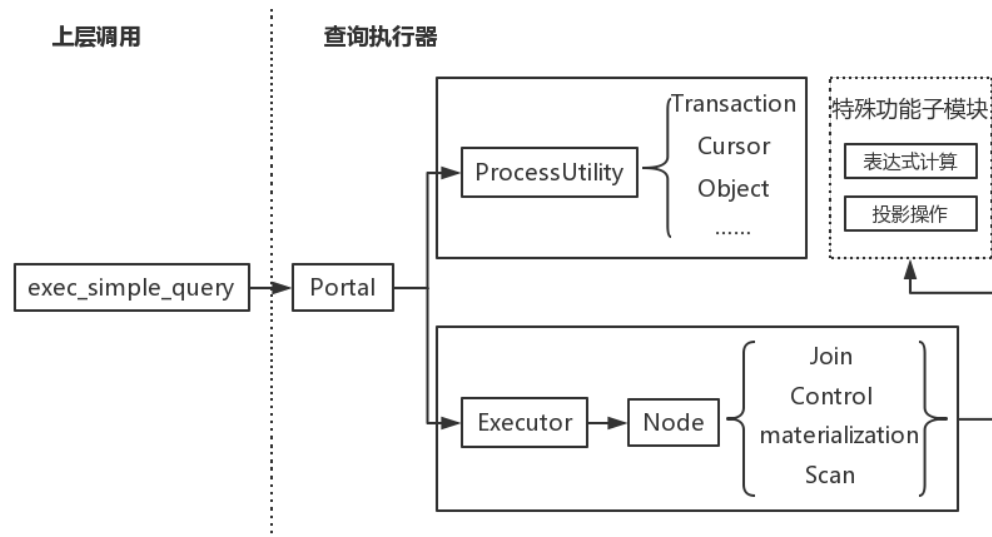


第六章 查询执行

书稿内容分析：

6.1 PostgreSQL 中的查询执行



PostgreSQL 的查询执行器有四个主要的子模块：Portal、ProcessUtility、Executor 和特定功能子模块。

6.2 Greenplum 中的查询执行

本地 聚合操作

异地 广播

重分布

窗口函数

6.3 ORCA 查询优化器

Orca: A Modular Query Optimizer Architecture for Big Data (SIGMOD

2014)

6.4 Greenplum 源码修改实例

查询优化器进行改进：

1. select-from 子句去除冗余表。
2. 新的多表连接算法——贪心算法

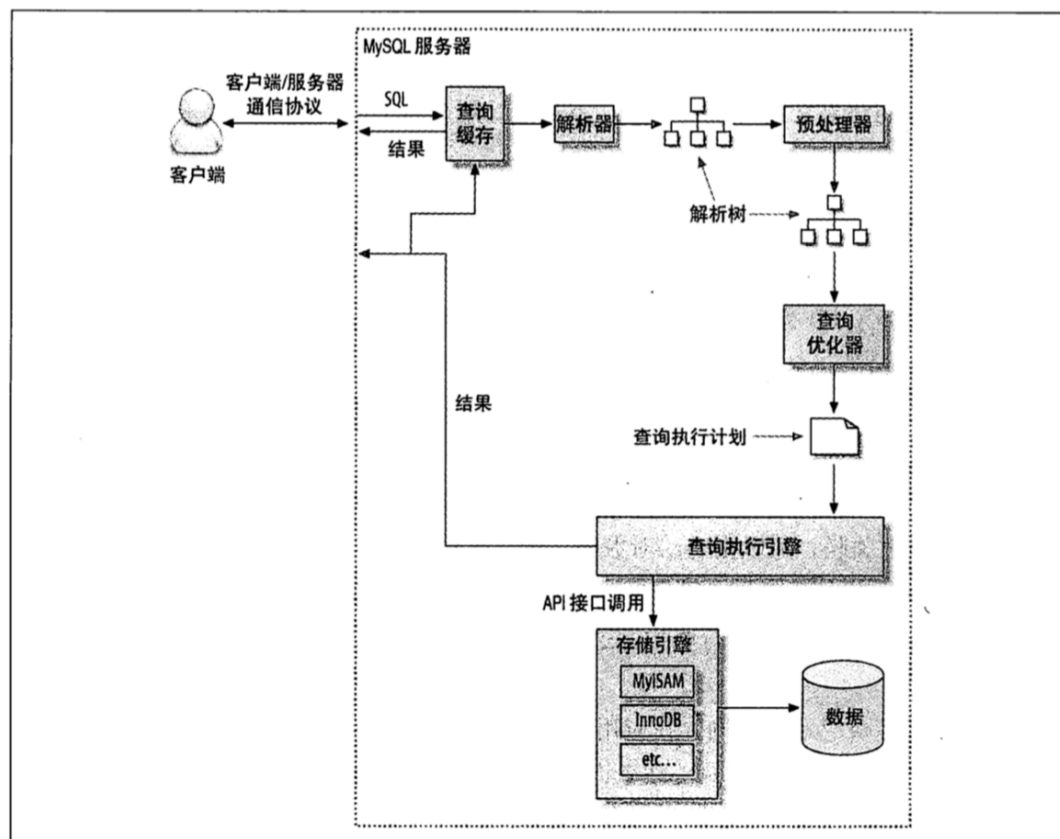
6.5 查询执行实验导引

一些实验介绍

2.拟加入的内容

1. Mysql 查询执行流程对比

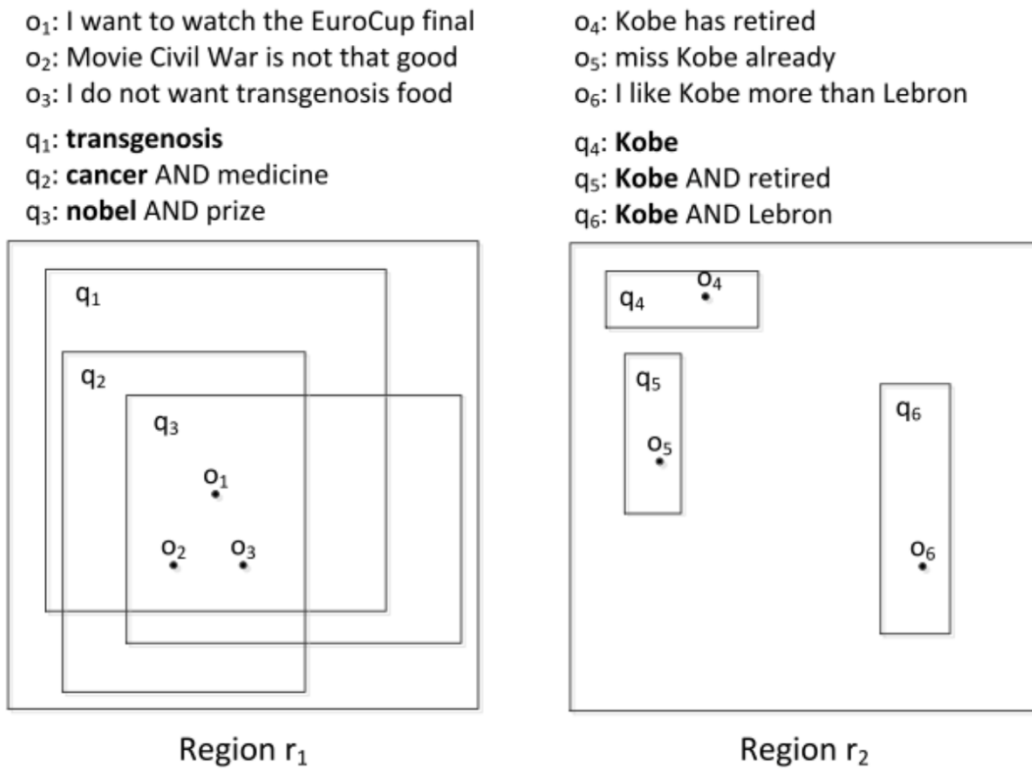
——高性能 MYSQL



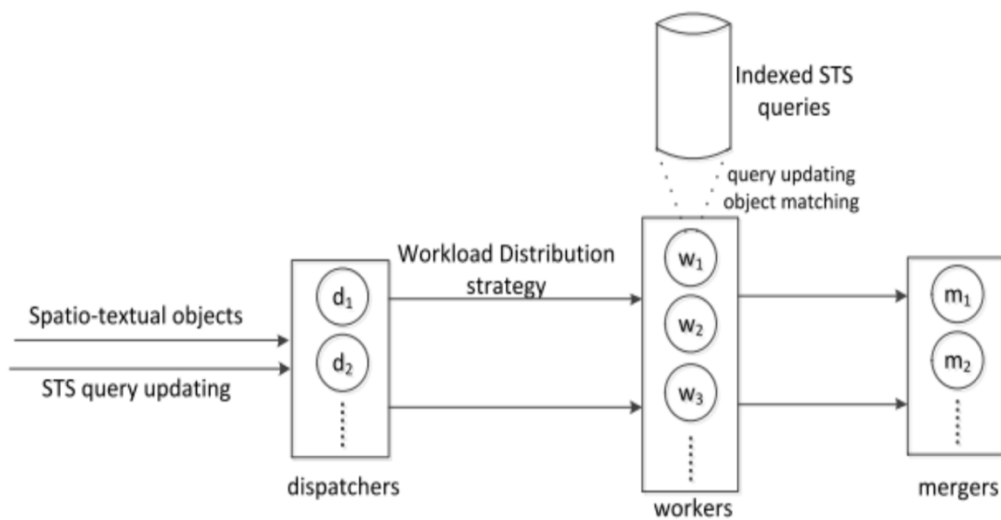
2. Distributed Publish/Subscribe Query Processing on the Spatio- Textual Data Stream --ICDE 2017

应用场景：随着装备了 GPS 的移动计算设备以及社交网络服务的普及，同时包含着文本和位置信息的数据正以前所未有的速度产生。数以亿计的用户可以在微博 上发布包含地点信息的信息，在 Facebook 上创建包含位置信息的事件；用户也可能对某个区域的事情表示兴趣，并订阅一些位置或者关键词。发布/订阅系统为上述用户需求提供了基础技术，用户可以通过系统中的 Query 来从所有用户的发布内容中，获取满足一定条件的内容。

现状：对于 Spatio-Textual 数据，有两种分类方法，一是根据空间区域划分，另一种是根据文本划分。已有的工作中对以上两种方法都进行了尝试，但在本论文的情境下，两种方法很难满足总负载最小和负载均衡条件。



实现： 提出并设计了一种分布式发布/订阅 查询处理系统 (PS2Stream)



主要思想总结：

首先用空间划分将 STS queries 和 Spatio-Textual objects 拥有较大相似度的空间分开。然后再通过评估划分后的系统负载情况，采用空间划分 或者文本划分方法进一步处理，如果任何一种划分方法都不能使得总负载减小且当前划分区间不小于 worker 数量，则停止划分。

3.需要讨论的问题

- 1.扩展的论文资料是否需要加入查询优化部分**
- 2.如果需要的话，本章节的内容和结构是否需要修改**