

zhanlijun

首页新随笔联系订阅管理

随笔 - 49 文章 - 0 评论 - 370

Visitors

CN

177,205

US

15,556

HK

5,780

JP

3,721

TW

3,008

SG

1,356

CA

816

FR

805

GB

775

AU

695

DE

670

KR

336

Pageviews: 377,090

Flags Collected: 83

FLAG

counter



个人经历

2015 至今 阿里巴巴

2013-2015 美团

2010-2013 中科院（硕士）

2006-2010 浙大（本科）

阿里巴巴RDC长期招聘Java研发工程师，有意者站内联系！

昵称：zhanlijun

园龄：4年10个月

粉丝：664

关注：5

+加关注

最新随笔

1. 一个复杂系统的拆分改造实践

2. mysql死锁问题分析

3. 近期code review几处小问题集锦

4. 你应该知道的RPC原理

5. 如何健壮你的后端服务？

6. 如何用消息系统避免分布式事务？

7. 一个故事讲清楚NIO

8. 地图匹配实践

9. 利用模拟退火提高Kmeans的聚类精度

10. 空间插值文献阅读（Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall）

深入浅出空间索引：2

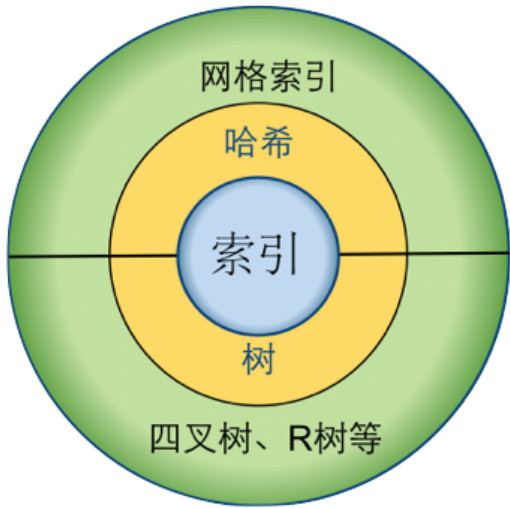
<http://www.cnblogs.com/LBSer/p/3403933.html>

深入浅出空间索引2

第一篇讲到了传统的索引如B树不能很好的支持空间数据，比如点（POI等）、线（道路、河流等）、面（行政边界、住宅区等）。本篇将对空间索引进行简单分类，然后介绍网格索引。（深入浅出空间索引1：<http://www.cnblogs.com/LBSer/p/3392491.html>）

一、空间索引有哪一种？

传统索引使用哈希和树这两类最基本的数据结构。空间索引虽然更为复杂，但仍然发展于这两种数据结构。因此可以将空间索引划分为两大类：1）基于哈希思想，如网格索引等；2）基于树思想，有四叉树、R树等。



二、网格索引

哈希是通过一个哈希函数将关键字映射到内存或外存的数据结构，如何扩展到空间数据呢？

2.1. 网格索引原理

扩展方法：对地理空间进行网格划分，划分成大小相同的网格，每个网格对应着一块存储空间，索引项登记上落入该网格的空间对象。

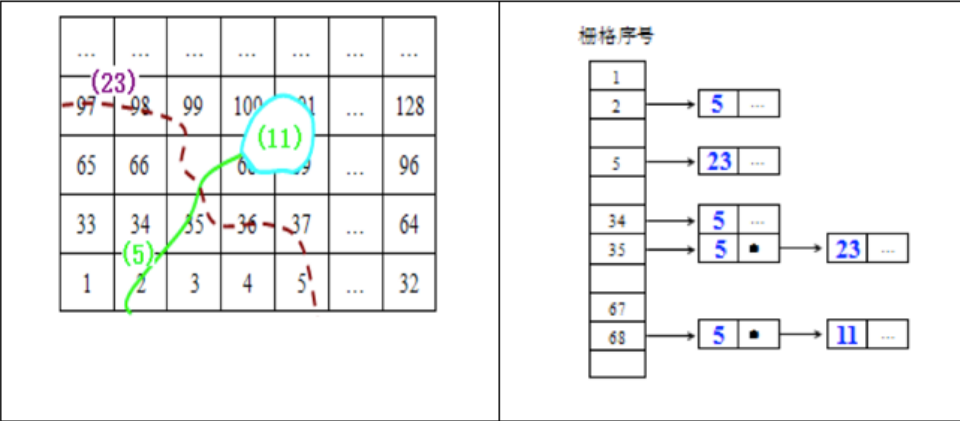
举个例子，我们将地理空间进行网格划分，并进行编号。该空间范围内有三个空间对象，分别是id=5的街道，23的河流和11的商圈。这时候我们可以按照哈希的数据结构存储，每个网格对应着一个存储桶，而桶里放着空间对象，比如对2号网格，里面存储着id=5的空间对象，对35号网格，桶里放着id=5和id=23的空间对象。

随笔分类(57)
java(3)
LBS(10)
paper阅读笔记(2)
大数据(6)
定位原理/算法(3)
发表的SCI/SSCI(4)
服务治理(4)
空间索引原理(7)
数据库(5)
推荐相关(1)
线上问题定位及解决(2)
消息系统(2)
信息检索算法/实践(6)
应用服务器(2)

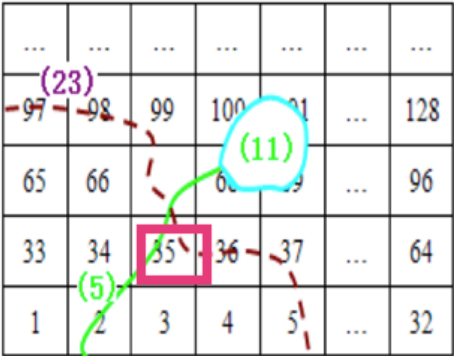
积分与排名
积分 - 115075
排名 - 2612

最新评论
1. Re:如何设计实现一个地址反解析服务？  如果仅仅是为了将用户坐标解析到道路级别的话，也未必需要用栅格。对于任意一条道路，根据历史记录，可以得到定位于这条道路的所有点，根据这堆点可以得到一个外包多边形，以后所有落在这个多边形内的点都可以认为是.....  --张可纯biubiu
2. Re:GeoHash核心原理解析  lucene里面使用了geohash，但是计算距离的时候貌似还是用经纬度计算距离，那使用geohash还有什么意义呢？  --casterQL

阅读排行榜
1. GeoHash核心原理解析(43980)



假如我们要查询某一空间范围内有哪些空间对象，比如下面的红框就表示空间范围，我们可以很快根据红框的空间范围算出它与35号和36号网格相交，然后分别到35号和36号网格中查找空间对象，最终找出id=5和id=23的空间对象。



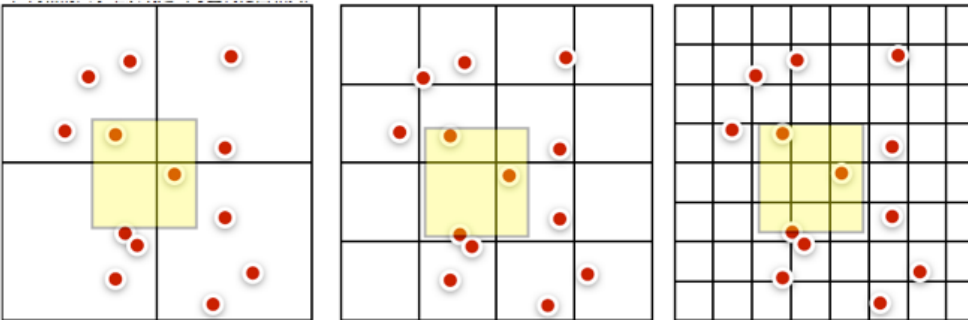
2.2. 网格索引缺点

1) 索引数据冗余

网格与对象之间多对多关系在空间对象数量多、大小不均时造成索引数据冗余。比如11号商圈这个空间对象在68, 69, 100, 101这4个网格都有存储，浪费了大量空间。

2) 网格的大小难以确定

网格的划分大小难以确定。网格划分得越密，需要的存储空间越多，网格划分的越粗，查找效率可能会降低。对于图a，这个查询需要查询4个网格，由于4个网格覆盖了整个空间，因此这个查找其实是将空间范围内所有的点数据都遍历一遍，失去了索引的意义。

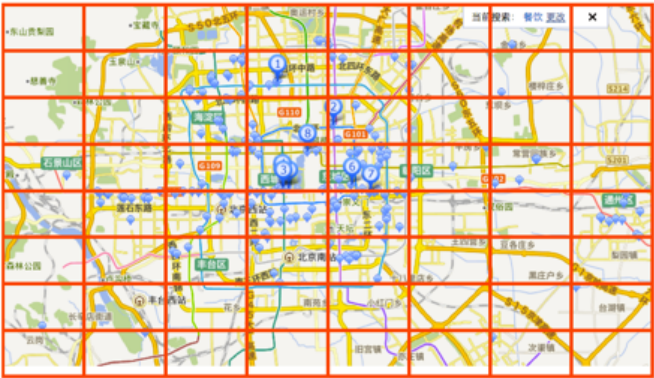


3) 很多网格没有数据

空间数据具有明显的聚集性，比如POI只在几个热点商贸区聚集，在郊区等地方很稀疏，这将导致很多网格内没有任何空间数据。

2. 你应该知道的RPC原理(30598)
3. 如何用消息系统避免分布式事务？(23677)
4. mysql死锁问题分析(22275)
5. 位图索引:原理（BitMap index）(21132)

评论排行榜
1. 地图匹配实践(82)
2. 如何用消息系统避免分布式事务？(42)
3. 你应该知道的RPC原理(23)
4. GeoHash核心原理解析(22)
5. 地理围栏算法解析（Geo-fencing）(20)



下一节将介绍四叉树。

转载请标明源地址：http://www.cnblogs.com/LBSer

分类：[LBS](#),[空间索引原理](#)

好文要顶

关注我

收藏该文



[zhanlijun](#)  
关注 - 5  
粉丝 - 664  
[+加关注](#)

2013-11-02 16:49 0

« 上一篇：[深入浅出空间索引：为什么需要空间索引](#)  
» 下一篇：[GDAL并行I/O](#)

posted @ 2013-11-02 16:49 zhanlijun 阅读(4385) 评论(7) 编辑 收藏

评论列表

#1楼 2014-03-18 13:51 刀尖红叶	四叉树的介绍呢？	支持(0) 反对(0)
#2楼[楼主 ] 2014-03-19 22:33 zhanlijun	@ 刀尖红叶 好久没登录了，马上下篇	支持(0) 反对(0)
#3楼 2015-11-15 23:21 Leon凌霄	没有下篇了啊，结论是啥？使用geohash来索引吗	支持(0) 反对(0)
#4楼 2016-06-27 17:55 qifuguang.me	博主，四叉树呢？	支持(0) 反对(0)
#5楼 2017-05-05 19:18 时光飞仔	兄弟，说好的下一篇呢，现在都2017了，拖得太久了。。。	支持(0) 反对(0)
#6楼 2017-08-09 12:52 1加1equal雨山		

博主 说好的下篇呢

支持(0) 反对(0)

#7楼 2017-09-19 20:20 喵喵de天真

博主说好的下一篇呢，我都跳槽了2个公司了

支持(0) 反对(0)

[刷新评论](#) [刷新页面](#) [返回顶部](#)

注册用户登录后才能发表评论，请 [登录](#) 或 [注册](#)，[访问网站首页](#)。

- 【推荐】超50万VC++源码：大型组态工控、电力仿真CAD与GIS源码库！
- 【缅怀】传奇谢幕，回顾霍金76载传奇人生
- 【推荐】业界最快速.NET数据可视化图表组件
- 【腾讯云】买域名送解析+SSL证书+建站
- 【活动】2050 科技公益大会 - 年青人因科技而团聚



最新IT新闻：

- 美团打车：已拿下上海1/3市场份额
  - 孙宏斌：投资乐视网肯定是失败了，对财务和团队判断有失误
  - 蓝色光标陷劳资纠纷 6年人员成本增逾10倍
  - 投资育碧、腾讯财报 这两件事应该一起看
  - 7小时通宵大搜查！英国隐私监管机构进驻剑桥分析可查服务器
- » 更多新闻...



最新知识库文章：

- 写给自学者的入门指南
  - 和程序员谈恋爱
  - 学会学习
  - 优秀技术人的管理陷阱
  - 作为一个程序员，数学对你到底有多重要
- » 更多知识库文章...