



ISSN 1000-1239
CODEN JYYFEY

Journal of Computer Research and Development

计算机研究与发展

第51卷增刊I 2014年9月

Vol.51 Suppl. I Sep. 2014

主办 中国科学院计算技术研究所 中国计算机学会 出版 科学出版社

中国计算机学会会刊

从基于位置的社会网络中发现消费型交通拥堵结点	陈芳芳	马宇驰	刘东权	(212)		
基于 KL 距离的离散概率分布相似性连接	金鑫	孙建伶		(220)		
社会网络中节点间真实影响力的挖掘	蒋恒	王轶彤		(226)		
基于 MYCIN 方法的用户信任关系判别	左祥麟	姚远	王泊	王俊华	左万利	(233)
一种基于评分策略和置信区间的代价敏感决策树	冯少荣	叶林宝	林子雨	赖永炫		(239)
基于加权序列的微博文本特征研究	徐严康	向阳	张波	张骥		(250)

数据隐私与安全

基于位置的社会网络的 k -连接安全分组匿名方法	郝雪燕	朱怀杰	杨晓春	王斌	(256)
基于区间查询概率的差分隐私直方图发布算法	黄臻卿	康健	吴英杰	黄泗勇	叶少珍 (265)
数据共享中面向个性化隐私保护的数据匿名化模型	王茂异	黄云南	王智慧	汪卫	朱拯 (272)

物联网和云计算环境中的数据管理

用于复杂车联网的空时上下文感知技术	郑笛	王俊	黄可荣	(281)	
智慧城市中基于情景感知的分布式配置管理机制研究	魏铁	刘奎恩	贺也平	(290)	
基于范围划分的多 θ 连接算法	张莹	李红燕	王腾蛟	雷凯	(300)

其他

路网环境下的最长频繁路段序列挖掘算法	杨恺希	章翰元	张雨薇	吴昊	孙未未	(307)
一种基于规则的消重方法	钟超玮	刘奇志	胡伟			(317)

系统演示

RouteMiner: 基于频繁模式和兴趣度的旅游路线挖掘系统	乔少杰	王峻毅	韩楠	(325)			
基于高性能地理计算和通信顺序进程的应用构建框架	张冰	吴秋云	杨岸然	陈苹	(329)		
大数据挖掘平台的构建及其在医疗领域的应用	陈磊	于广军	程思远	王飞	熊赞	刘卉	(334)
以路径为中心的大规模图数据处理系统 TripleGraph	谢昌凤	李方旸	张文姬	袁平鹏	(338)		
APS: 一个基于用户浏览行为的广告推送系统	朱珂	肖迎元	许庆贤	艾鹏强	崔文相	(342)	
PuMA: 面向微博话题分析的舆情数据服务平台	杨林瑞	陈磊	熊赞	先梦涵	朱扬勇	(346)	
基于时空立方体数据聚合的移动对象检索服务原型系统	许弘琛	陈苹	熊伟	吴秋云	刘露	(351)	
基于多层次过滤算法的中文指代消解	张尚宏	郝晓燕	(356)				
数据库基准测试可视化工具的设计实现及其应用	李梁	吴刚	刘辉林	王国仁	(360)		
CMMDI: 一个中医多元数据整合平台	孙婧	熊赞	(364)				
微博小医生: 基于新浪微博的医疗建议系统	魏延杰	王宏志	李言路	李建中	高宏	(368)	
基于评论分析的酒店推荐系统	张凯	王科强	王晓玲	金澈清	周傲英	(372)	
基于数据质量评估的数据市场系统	丁小欧	王宏志	朱鎔	李建中	高宏	(377)	
Track-Match: 一个车辆轨迹地图匹配工具	刘骁	金澈清	王晓玲	周傲英	(382)		
基础研究人员社会关系网络查询系统	李东	郝艳妮	何贤芒	周艳春	(387)		
关于数据密集型的众包清洗平台	李可利	王宏志	叶晨	郭欣彤	李建中	高宏	(391)
一站式服务推荐系统	高伟璠	郑芷凌	余文喆	程文亮	张蓉	(395)	

数据库基准测试可视化工具的设计实现及其应用

李 梁 吴 刚 刘辉林 王国仁

(东北大学信息科学与工程学院 沈阳 110819)

(wugang@ise.neu.edu.cn)

The Design and Implementation of a Database Benchmark Visualization Tool and Its Application

Li Liang, Wu Gang, Liu Huilin, and Wang Guoren

(College of Information Science and Engineering, Northeastern University, Shenyang 110819)

Abstract TPC-C benchmark is the professional standard of testing database performance in online transaction processing (OLTP). But it lacks open source visual tool supports. According to TPC-C benchmark, this paper introduces the design and implementation of an open source benchmark visualization tool, VisualDBBench. Considering that Main-Memory databases have better performance in OLTP processing comparing with those traditional databases, the proposed VisualDBBench tool was exploited to conduct TPC-C benchmark and verified performance superiority of several main-memory databases.

Key words database system performance evaluation; main-memory database; TPC-C benchmark; VisualDBBench tool; OLTP

摘 要 TPC-C 规范作为测试数据库性能的行业标准,已被广泛使用在测试联机事务处理(OLTP)性能,但目前仍缺乏开源的可视化工具作为支持。根据 TPC-C 规范,介绍了一个开源 VisualDBBench 工具的设计与实现,并且使用 VisualDBBench 工具进行测试,验证了针对内存数据库在 OLTP 相对传统数据库具有较优的性能优势。

关键词 数据库系统性能评测;内存数据库;TPC-C 规范;VisualDBBench 工具;联机事务处理

中图法分类号 TP311

TPC-C^[1] 是一种旨在衡量 OLTP 系统性能与可伸缩性的行业标准基准测试项目。这种基准测试项目将对包括查询、更新及队列式小批量事务在内的广泛数据库功能进行测试。目前,虽然存在开源的 TPC-C 测试工具,但直观易用的可视化工具还相当缺乏。本文以开源系统 BenchmarkSQL^[2] 为基础,采用可插拔的设计方式,实现了能够进行多个数据库之间性能测试对比、可视化显示 TPC-C 性能结果的工具。

1 关键技术

1.1 TPC-C 测试基准

测试基准是指根据明确定义的测试设置对一个或多个系统进行系统比较的测试过程。TPC-C 是专门用于测试数据库系统 OLTP 性能的测试基准。它的处理过程由读操作和更新事务操作频繁交互完成。

收稿日期:2014-06-25

基金项目:国家自然科学基金项目(61332006,61370154,61025007,61328202,61100024);国家“九七三”重点基础研究发展计划基金项目(2011CB302200-G);国家“八六三”高技术研究发展计划基金项目(2012AA011004)

TPC-C 测试规范中模拟了一个比较复杂并具有代表意义的 OLTP 应用环境:假设有一个仓库和订单管理系统,它拥有若干个分布在不同区域的仓库;通常每个仓库供货覆盖 10 家地区分公司;每个分公司销售点给 3 000 个客户提供服务;平均每个客户的一个订单有 10 项产品;所有订单中约 1% 的产品在其直接所属的仓库中没有存货,需要由其他区域的仓库来供货.图 1 所示为 TPC-C 商业模型的逻辑结构^[3-5].

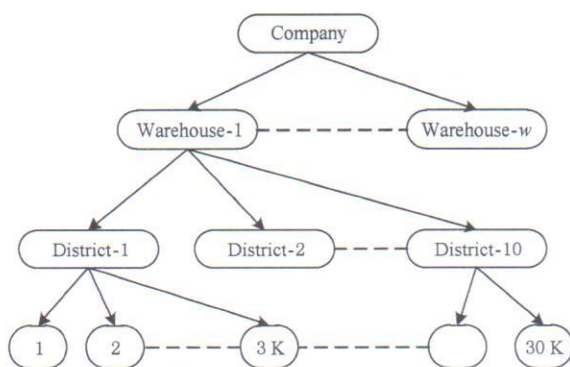


图 1 TPC-C 商业模型的逻辑结构

TPC-C 共定义了 5 种测试事务,在这个环境中,以一个新订单事务为核心活动,并辅以执行其他 4 种事务:支付操作事务、订单状态查询事务、发货事务、库状态查询事务.整个测试流程就是不断地执

行事务.

基准的性能度量用每分钟新增订单事务数 ($tpmC$) 表示,为了与 TPC-C 标准一致,TPC-C 结果的所有参考值必须包括 $tpmC$ 值、与之相关的每个 $tpmC$ 的价格和给定系统价格的可用日期.

2 VisualDBBench 系统结构

2.1 需求分析

仔细分析工具的功能需求,具体来讲,本系统主要有以下 6 个功能:

- 1) 配置待测数据库的个数;
- 2) 配置单个数据库的参数,主要包括数据库连接所用到的 URL 参数和用户名以及密码等;
- 3) 测试数据库:该模块主要包括创建数据库表、加载基本测试数据和创建数据库索引并开启事务测试;
- 4) 实时显示测试结果:为了直观地反映测试结果,必须有实时的可视化显示界面;
- 5) 加载 SQL 语句:这个模块主要负责配置系统执行时需要用到的 SQL 语句;
- 6) 生成多个数据库对比结果:以条形图的形式对比展示多个数据库的性能结果.

如图 2 所示为系统用例图:

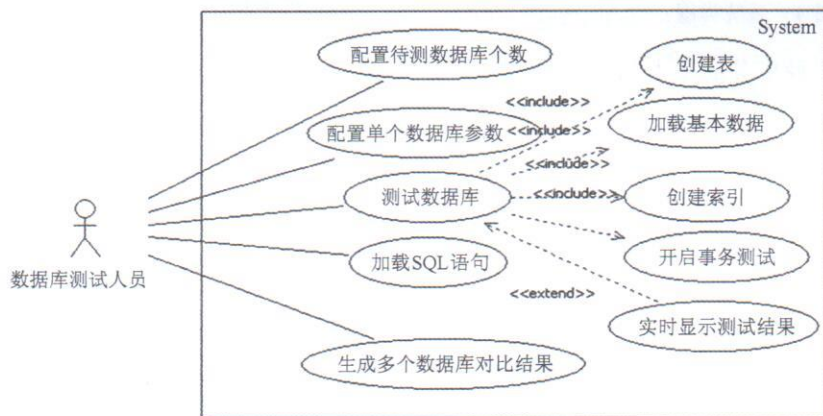


图 2 系统用例图

2.2 测试流程设计

为了实现自动化的 TPC-C 测试,VisualDBBench 需要具有配置参数、创建数据库装载数据、提交数据库事务和可视化显示结果的功能.

图 3 所示为系统的整体处理流程.

2.3 程序结构

工具按面向对象设计,采用 Java 编程.工具的类型图设计如图 4 所示.

程序中主要类结构说明:

1) jTPCC 类. 承担用户与 VisualDBBench 的交互,它包括标签、按钮、文本区等响应事件.用户可以在这里设置参数,如 DBMS 的类型、数据库的连接信息、TPC-C 测试终端和数据仓库的个数以及程序测试时间和 5 种测试事务所占权重.界面中可以动态显示测试进度信息和测试结果信息.

2) jTPCCTerminal 类. 根据用户的参数配置,

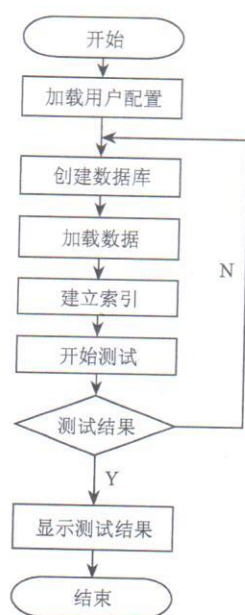


图3 系统处理流程图

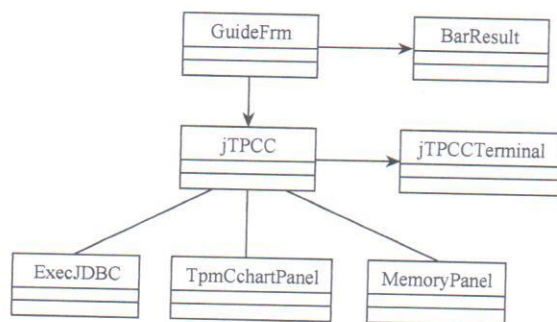


图4 系统类图

模拟每个终端开启 5 种事务的基本测试。

3) ExecJDBC 类. 连接数据库, 运行特定 SQL 语句文件, 考虑到程序通用性的问题, SQL 脚本语言定义在工程目录下的文本文件中, 对于特定的数据库可以使用不同的脚本。

4) BarResult 类. 以柱状图的形式显示对比每个数据库的性能测试结果。

5) MemoryPanel 类. 以仪表盘的形式实时显示 JVM 内存使用情况。

6) TpmCchartPanel 类. 获取测试过程中 tpmC 的数值, 实时显示给用户。

2.4 工具适用性和可插拔性实现

工具的目的是要对多个数据库进行对比测试, 这就要求工具必须能够是可插拔的。VisualDBBench 工具有丰富的配置文件, 可以非常灵活地连接上多种数据库进行 TPC-C 测试。

虽然各大数据库厂商都采用了 SQL 语言标准, 但是它们又都不是完全的统一, 只能看作是标准

SQL 语言的方言。鉴于此, 工具为了能最大限度地支持多个数据库, 并没有把 SQL 脚本语言内嵌到代码内, 而是另外存放特定的文本文件中, 遇到不兼容的情况, 可以适当地更改 SQL 脚本。

3 VisualDBBench 工具测试过程演示

3.1 工具演示环境

本次实验主要目的在于测试传统数据库和内存数据库在 OLTP 方面的性能差异, 因此实验中选取了几个被广泛使用且具有代表性的数据库作为测试对象。表 1 所示为工具开发与运行环境:

表 1 工具开发与运行环境

运行平台或工具	版本号或型号
Windows	XP sp3
CPU	CoreI i3120
RAM	DDR3 2GB
JDK	1.8.0
NetBeans	7.4
MySQL	5.0
H2	1.3.175
JFreeChart	1.0.13

3.2 配置文件

为了实现工具的通用性、灵活性和可插拔特性, 在工具的外部用一个 conf.txt 的配置文件来管理需要测试的多个数据库。另外, 在每一个数据库对应的配置文件内, 记录测试的 4 个属性, 分别是待测数据库的名字、JDBC 连接的驱动类、JDBC 连接数据库对应的 url 以及连接数据库的授权用户名和密码。

3.3 TPC-C 基准测试步骤

1) 双击 run.bat 文件即可启动程序主界面, 工具会自动加载相应的配置文件。另外, 还可以通过界面左侧的 Option 中的按钮来配置“测试模拟终端个数”和“事务百分比”等参数。本次演示, 程序会有 3 个向导型界面, 依次可用来测试 H2 数据库、MySQL 数据库和 H2 内存模式数据库。

2) 点击左侧 Option 分面板中 Control 按钮, 此时界面如图 5 所示, Control 分面板内有 7 个功能按钮。

3) 依次单击 Create Table, Load Data, Create index, Create Terminal, Start Transaction 按钮, 分别执行数据库表的创建、数据加载、创建索引、创建测试模拟终端和开启事务测试过程。

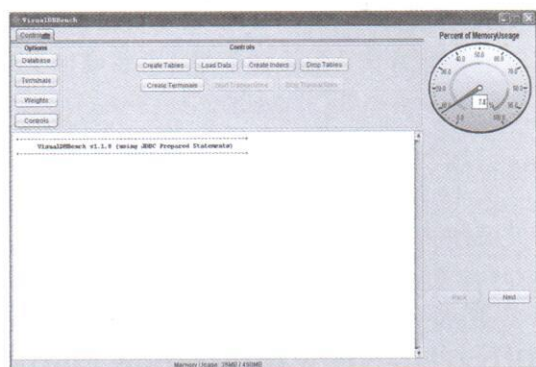


图5 程序运行主界面

4) 单个数据库测试结束,会在界面左下角的区域输出提示,此时单击 next 按钮,重复上述过程对下一个数据库进行 TPC-C 测试。

3.4 测试中实时获取 tpmC 参数值

在对每个数据库的测试过程中,可以在 tpmC result 选项卡页面实时观测 tpmC 变化情况,工具可实时显示当前 tpmC 值和 tpmC 的平均值,图 6 所示为数据库实时 tpmC 的测试结果。

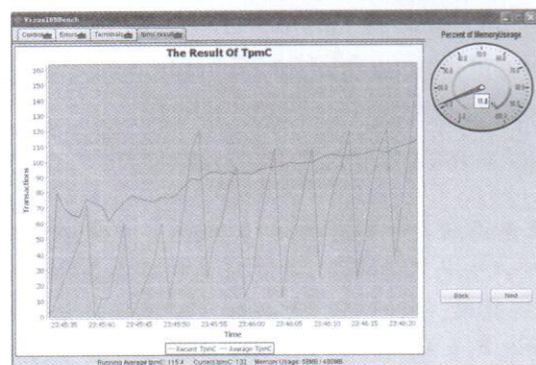


图6 实时显示 tpmC 数值

3.5 获取多个数据库的比较结果

向导界面的最后会将 3 次测试结果做对比,图 7 所示为 H2 数据库、MySQL 数据库和 H2 内存

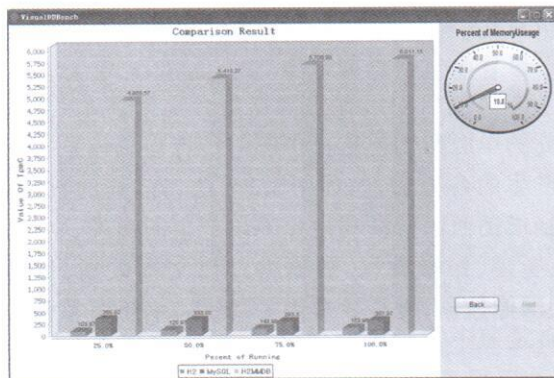


图7 多个数据库性能结果对比

数据库的 TPC-C 测试结果对比。

H2 数据库和 MySQL 数据库的性能结果并没有显著的差别,但相比内存模式的 H2 数据库,内存数据库在单位时间内处理事务的能力则出现百倍级的提升。

文献[4]中指出,内存数据库的性能通常为传统基于磁盘的数据库性能的百倍以上,从图 7 可以看出,实验结果基本与其吻合。

4 结束语

本文的研究目的是设计并开发一个自动化的数据库基准测试工具,使测试者无需深入研究 TPC-C 测试规范,不需要开发测试程序,测试过程也无需用户干预。

此外,通过测试发现,内存数据库在事务处理方面的能力显著优越于传统基于磁盘的数据库。

参 考 文 献

- [1] Transaction Processing Performance Council. TPC-C benchmark, Version 5.11. [2010-05-20]. <http://www.tpc.org/tpcc/default.asp>
- [2] BenchmarkSQL, Version 2.3.3. [2014-06-03]. <http://sourceforge.net/projects/benchmarksql/files>
- [3] 王欣晖,于戈,王国仁,等. TPC-C 在面向对象数据库上的设计与实现 //数据库研究进展 97——第十四届全国数据库学术会议论文集(上). 成都:四川联合大学,1997: 5
- [4] 于戈,王欣晖,王国仁,等. TPC-C 测试标准及其在面向对象数据库上的设计与实现. 计算机科学,1998(3): 57-60
- [5] 王良,蔡荣. 基于 TPC-C 标准的自动化测试工具 TPCCLoader. 计算机工程与应用,2005(25): 102-105,115
- [6] Plattner H, Zeier A. In-Memory Data Management. Berlin: Springer, 2009

李 梁 男,1992 年生,本科生,主要研究方向为内存数据库。

吴 刚 男,1978 年生,副教授,主要研究方向为语义 Web、分布式数据库、并行数据库、内存数据库。

刘辉林 男,1964 年生,教授,主要研究方向为内存数据库。

王国仁 男,1966 年生,教授,主要研究方向为不确定数据管理、数据密集型计算、内存数据库。