

一种异构数据集成中的动态查询优化方法*

An Approach of Dynamic Query Optimization in Heterogeneous Data Integration

李瑞轩, 卢正鼎, 吴 炜, 肖卫军

LI Rui-xuan, LU Zheng-ding, WU Wei, XIAO Wei-jun

(华中科技大学计算机科学与技术学院, 湖北 武汉 430074)

(School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei 430074, P.R.China)

摘 要: 查询优化是异构数据集成中需要解决的关键问题之一, 但环境的分布性、异构性以及局部数据源的自治性使得异构数据集成中的查询优化变得非常困难。通过对异构数据集成中查询后处理调度的分析, 给出了查询后处理的全局查询图表示, 并给出了基于全局查询图的查询后处理并发调度方法, 最后给出了一种基于统计推理的查询后处理动态优化策略, 它可用于基于全局查询图的查询调度中, 实现查询后处理的动态优化。

Abstract: Query optimization is one of the key issues in heterogeneous data integration. But it is very difficult to design and implement query optimization because of the distribution and heterogeneity of the computing environment and the autonomy of multiple local data sources. This paper firstly analyzes the scheduling of post query processing in heterogeneous data integration and introduces the global query graph for post query processing. Then, it gives an approach of parallel scheduling of post query processing based on global query graph. Finally, a dynamic query optimization strategy based on statistical reasoning is presented. It can be used to optimize the post query scheduling in heterogeneous data integration.

关键词: 异构数据集成; 查询优化; 查询后处理; 动态优化; 全局查询图

Key words: Heterogeneous Data Integration; Query Optimization; Post Query Processing; Dynamic Optimization; Global Query Graph

中图分类号: TP311.13

文献标识码: A

稿件编号: 23929

1 引言

目前, 随着数据处理技术的快速发展, 网络上存在着大量非常有价值的信息, 这些信息以各种形式存储在各种各样的数据源中。随着应用需求的不断增加, 越来越多的用户希望能够透明地访问和处理来自多个数据源中的数据。然而, 这些数据源可能在数据模型、存储方式、数据语义、查询语言、访问控制策略等方面存在着很大的差异。传统的数据集成技术, 已无法适应人们及时获取更多更新更全面数据的需要。因此, 研究分布环境下多种异构数据源的集成成为一种必然趋势^[1]。

查询优化一直是数据处理中非常重要而又非常复杂的问题。开发数据集成系统的目的就是为了更有效地查询分布在各个数据源中的数据, 而查询优化是影响查询处理性能的关键因素。与传统的分布式数据管理系统中的查询优化比较起来, 异构数据源集成系统中的查询优化要复杂得多。不同的局部数据源可能使用不同的查询语言、查询处理算法和性能评价方法, 这使得不同系统之间的互操作及全局优化之间的合作更加困难。在全局优化中需要的局部数据源的重要信息(例如局部代价参数)可能无法得到或者只能得到

*收稿日期: 2003-10-08; 修订日期: 2004-03-。

资助项目: 国家科技攻关计划项目资助(2002BA103A04)

作者简介: 李瑞轩(1974-), 男, 湖北宜昌人, 博士生, 讲师, 研究方向为分布异构系统集成; 卢正鼎, 教授, 博士生导师, 研究方向为 CIMS 及分布式系统。

通讯地址: 430074 湖北省武汉市华中科技大学计算机学院; Tel: (027)87544285; E-mail: rxli@public.wh.hb.cn

Address: School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei 430074, P.R.China