

Chapter 3.11

Generating Join Queries for Large Databases and Web Services

Sikha Bagui

The University of West Florida, USA

Adam Loggins

Zilliant Inc., USA

ABSTRACT

In this data-centric world, as Web services and service oriented architectures gain momentum and become a standard for data usage, there will be a need for tools to automate data retrieval. In this article the authors propose a tool that automates the generation of joins in a transparent and integrated fashion in heterogeneous large databases as well as Web services. This tool reads metadata information and automatically displays a join path and a SQL join query. This tool will be extremely useful for performing joins to help in the retrieval of information in large databases as well as Web services. [Article copies are available for purchase from InfoSci-on-Demand.com]

INTRODUCTION AND RELATED WORKS

As we are working with more and more data, the sizes of databases are getting larger and larger. As businesses are going global, Web services are becoming a standard for sharing data (Srivastava et al., 2006; Resende and Feng, 2007). Enterprises are moving towards service oriented architectures where several large databases may be layered behind Web services, hence databases are having to become adaptable with loosely-coupled, heterogeneous systems (Srivastava et al., 2006) too. In such scenarios of Web services and service oriented architectures, which may be dealing with several loosely coupled heterogeneous large databases, it is no longer humanly possible to have handy all the information on all the tables and primary keys in all the large databases. Although considerable work is

being done on the challenges associated with Web services addressing the problem of multiple Web services to carry out particular tasks (Florescu et. al., 2003; Ouzzani and Bouguettaya, 2004), most of this work is targeted towards work-flow of applications, rather than coordinating how data can be retrieved from multiple large databases in Web services via SQL (Srivastava et al., 2006). In this article we try to address one aspect of this problem of retrieving data from multiple heterogeneous large databases using SQL. Specifically, we present a tool that automatically formulates joins by reading the metadata of databases in the context of very large databases or in the context of Web services which may employ the use of several large heterogeneous databases.

Let us look at an example of a query presented to a Web service: *Suppose a health insurance company needs to verify the salary, health, and travel patterns of a person before determining the amount of health insurance he/she needs to pay. In a Web service, this will require joining of several tables. And, of course, no one person will have knowledge of all the primary key/foreign key relationships between the tables to join in the Web services.*

When databases were smaller, it was possible to have knowledge of most of the tables and primary key/foreign key relationships in databases, and SQL join queries could easily be built by joining tables in databases. But, in large databases layered behind Web services, it will not be possible to have knowledge of all the database schemas.

The join operation, originally defined in the relational data model (Codd 1970, 1972), is a fundamental relational database operation, facilitating the retrieval of information from two relations (tables). Writing efficient joins is simple for small databases since few relations are involved and one has knowledge of the complete database schema. But, writing efficient joins is a challenge in large database scenarios and Web services where it may not be possible to have a complete picture of the database schema and its relations.

Since joins are one of the most time-consuming and data-intensive operations in relational query processing, joins have been studied discussed extensively in the literature. Mishra and Eich (1992) present a very comprehensive study of works that have been done on joins. Query optimization issues in joins, and devising strategies for distributed join processing have also been discussed by many, for example, Kim et al. (1985), Perrizo et al. (1989), Segev (1986), Swami and Gupta (1988), Yoo and Lafortune (1989), and Yu et al (1985, 1987). These works have to be extended in the context of databases for Web services and service oriented architectures. Srivastava, et. al (2006) addresses the problem of query optimization over Web services on a much broader scale.

In this article we present a tool that we have developed that will: (i) read the meta data of databases, that is, search the database model or schema and discover the relationships between the tables using table indexes defined in the database catalogs; (ii) find efficient join paths between the tables to be joined; and, (iii) generate a SQL join query (in ANSI SQL standard).

This rest of the article is organized as follows: Section two briefly describes relational databases with respect to the join operation; section three presents an architectural overview of our tool; section four presents the configuration details of our tool; section five describes how we tested our tool and presents some results; and section six presents the conclusion. Some relevant code portions are presented in the appendices.

RELATIONAL DATABASES AND THE JOIN OPERATION

In relational databases, data is stored in the form of tables or relations. Each table has information on a particular subject or concept and is composed of a group of “related” attributes. The attributes in a table are all “related” in the sense that they

14 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/generating-join-queries-large-databases/37666

Related Content

Content-Determined Web Page Segmentation and Navigation for Mobile Web Searching

(2021). *Result Page Generation for Web Searching: Emerging Research and Opportunities* (pp. 88-108).

www.irma-international.org/chapter/content-determined-web-page-segmentation-and-navigation-for-mobile-web-searching/268298

On the Use of Web Services in Content Adaptation

Khalil El-Khatib, Gregor V. Bochmann and Abdulmotaieb El Saddik (2010). *Web Technologies: Concepts, Methodologies, Tools, and Applications* (pp. 2099-2114).

www.irma-international.org/chapter/use-web-services-content-adaptation/37731

Object Grouping and Replication on a Distributed Web Server System

Amjad Mahmood and Taher S.K. Homeed (2007). *International Journal of Information Technology and Web Engineering* (pp. 17-33).

www.irma-international.org/article/object-grouping-replication-distributed-web/2621

Cloud Adoption in Enterprises: Security Issues and Strategies

Ramanathan Venkatraman and Sitalakshmi Venkatraman (2016). *Web-Based Services: Concepts, Methodologies, Tools, and Applications* (pp. 1848-1872).

www.irma-international.org/chapter/cloud-adoption-in-enterprises/140880

Resource Scheduling and Load Balancing Fusion Algorithm with Deep Learning Based on Cloud Computing

Xiaojing Hou and Guozeng Zhao (2018). *International Journal of Information Technology and Web Engineering* (pp. 54-72).

www.irma-international.org/article/resource-scheduling-and-load-balancing-fusion-algorithm-with-deep-learning-based-on-cloud-computing/204359