The Study of Optimizing the Dynamic Task Scheduling in the Grid Environment

余政憲、邱紹豐

E-mail: 9601131@mail.dyu.edu.tw

ABSTRACT

The data warehouse is becoming an important platform for decision support processes. Corporations develop their strategies based on the analysis and query results on the data stored in the data warehouse. More and more Application Service Providers （ASP ’ s ） promote their applications for businesses to build their own data warehouses. This new technology is now an necessary tool for corporations to increases their competitiveness. In addition, XML is taken by the scientific and business communities as the major data format, especially for corporations exchanging their data. Since the XML data do not have fixed structures, it is difficult to adapt the existing data warehouse technology developed in the traditional relational database to the data formatted in this standard. In a data warehouse, data cubes, computed by performing the grouping operation, e.g. the GROUP BY instruction in SQL, are pre-computed and stored in the warehouse. If the data are stored in the XML format in the data sources, the regular grouping operation in the relational database is not suitable to aggregate the data. Using the existing XML techniques, it is required to have an XQUERY for each data cube. Thus, to implement a data warehouse with multiple data cubes, more than one scan on the source data is needed. In order to reduce the number of scans on the original data, we develop a data warehouse for the XML data based on the XCube structure. With the data cube stored in this structure, we provide necessary OLAP （On-Line Analytical Processing ） operations on the cubes. In this research, we also develop algorithms to materialize the data cube with the number of scanning the original data minimized. We also experiment our algorithm with numbers of data sets to prove the correctness of our theory.

Keywords : Data Cube ; OLAP ; Aggregation ; Data Warehouse

Table of Contents

REFERENCES

[1] W3C, “ Extensible Markup Language （XML） ”, http://www.w3.org/XML/, 2003.

[2] S. Abiteboul, P. Buneman and D. Suciu, “ Data on the Web, ” Morgan Kaufmann Publishers, 2000.

[3] J. Roy, A. Ramanujan, “ XML: data's universal language, ” IT Professional, Volume 2, Issue 3, pages 32-36 , 5 June 2000.

[4] A. Zisman, “ An overview of XML, ” Computing and Control Engin. J, Volume 11, pages 165-167, Aug 2000.

[5] W3C, “ Overview of SGML Resources ”, http://www.w3.org/MarkUp/SGML/, Nov 1995.

[6] James Clark and Steve DeRose, “ XML path language （XPath）, ” http://www.w3.org/TR/xpath.

[7] S. Boag, D. Chamberlin, M. Fernandez, D. Florescu, J. Robie, J. Simeon, and M. Stefanescu, “ XQuery: A Query Language for XML, ” W3C Working Draft. Available from http://www.w3.org/TR/xquery.

[8] XQuery 1.0 and XPath 2.0 Functions and Operators. W3C Working Draft, 30 April 2002.

[9] XML Pointer Language （XPointer） Version 1.0. W3C Candidate Recommendation , 11 September 2001.

[10] XSL Transformations （XSLT） Version 2.0. W3C Working Draft , 30 April 2002.

[11] Philippe Le He'garet, Ray Whitmer, and Lauren Wood, “ Document Object Model （DOM）, ” W3C DOM IG, 19 January 2005.

[12] Venky Harinarayan, Anand Rajaraman, Jeffrey D. Ullman, “ Implementing Data Cubes Efficiently, ” ACM SIGMOD, 205 – 216, 1996.

[13] Jim Gray, Adam Bosworth, Andrew Layman and Hamid Pirahesh, “ Data Cube: A Relational Aggregation Operator Generalizing

Group-By, Cross-Tab, and Sub-Total, ", ICDE 152-159, 1996.
[14] Surajit Chaudhuri, Umeshwar Dayal, " An Overview of Data Warehousing and OLAP Technology, " SIGMOD Record 26: 65-74, 1997.
[15] Surajit Chaudhuri, Kyuseok Shim, " Including Group-By in Query Optimization, " VLDB : 354 - 366, 1994.
[16] Yannis Sismanis, Antonios Deligiannakis, Yannis Kotidis, Nick Roussopoulos, " Hierarchical Dwarfs for the Rollup Cube, " DOLAP ' 7 November, 2003.
[17] Jiawci Han & Micheline Kamber, " Data Mining Concepts and Techniques, " Morgan Kaufmann Publishers, 2000.
[18] Wolfgang Hu"mmer, Andreas Bauer and Gunnar Harde, " XCube: XML for data warehouses, " Proceedings of the 6th ACM international workshop on Data warehousing and OLAP, 33-40, 2003.
[19] Rajesh R. Bordawekar and Christian A. Lang, " Analytical processing of XML documents : opportunities and challenges, " ACM SIGMOD Record Volume 34, Issue 2, June 2005.
[20] S.Paparizos, S. Al-Khalifa, H.V. Jagadish, L.Lakshmanan, A.Nierman, D.Srivastava, and Y. Wu, " Grouping in XML, " In EDBT Workshop on XML Data Management, 2002.
[21] H. V. Jagadish, Shurug Al-Khalifa, Adriane Chapman, Laks V. S. Lakshmanan, Andrew Nierman, Stelios Paparizos, Jignesh M. Patel, Divesh Srivastava, Nuwee Wiwatwattana, Yuqing Wu and Cong Yu, " TIMBER: A native XML database, " VLDB J: 274-291, 2002.
[22] H. V. Jagadish, Laks V. S. Lakshmanan, Divesh Srivastava and Keith Thompson, " TAX: A Tree Algebra for XML, " DBPL 149-164, 2001.