

在XML資料倉儲中群組之研究

余政憲、邱紹豐

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

摘要

資料倉儲目前已成為企業進行資料分析與查詢的平台，而在眾多的軟體供應商也相繼開發出許多相關的軟體，來提供企業建構資料倉儲，但大多是以關聯式資料為主。在現今XML的興起，特別適合在網際網路和全球資訊網的環境中流通傳輸，因此也漸漸廣泛的被企業所採用取代電子資料交換系統（EDI：Electronic Data Interchange）。由於XML資料並沒有固定的架構，在關聯式資料庫所使用的資料倉儲模型無法整合在XML資料上，因此針對XML資料來建構資料倉儲更顯得額外困難。資料倉儲中包含許多的資料立方體，在實體化資料立方體如同在關聯式資料庫中表格做資料分群與聚集運算結果。在針對XML資料做Group By時需去掃描XML文件，但當有多個資料立方體需實體化時，需搭配多個XQuery語法文件來累次執行，如此一來對XML資料的掃描次數也相對的增多。在本篇研究中我們採用XCube架構來開發一套以XML資料為基礎的資料倉儲，並透過我們開發的線上分析系統工具來做查詢。在實體化資料立方體方面，我們開發一套演算法，透過該演算法來收集要實體化資料立方體資訊，來達到只要掃描XML文件一次便可實體化所有的資料立方體。

關鍵詞：資料立方體；線上分析系統；群組與聚合

目錄

封面內頁 簽名頁 授權書.....	iii	中文摘要.....	iv	ABSTRACT.....	v	誌謝.....	vi
目錄.....	vii	圖目錄.....	ix	表目錄.....	xi	第1章 前言.....	1
1.1 研究動機.....	1	1.2 研究目的.....	2	1.3 論文架構.....	3	第2章 相關研究.....	4
2.1 XML Data.....	4	2.2 資料立方體.....	6	2.3 線上分析系統操作.....	9	2.4 XCube架構.....	11
2.5 分群演算法.....	17	第3章 系統架構與建構群組.....	20	3.1 系統架構與流程.....	20	3.2 改良式PT、WT與CT.....	21
3.3 資料分群演算法.....	24	第4章 系統實作.....	28	4.1 條件定義.....	30	4.2 線上分析系統查詢綱要.....	36
4.3 線上分析系統之上捲操作.....	38	4.4 線上分析系統之切片操作.....	39	4.5 線上分析系統之上捲切片操作.....	40	第5章 效能評估與實驗.....	42
5.1 測試條件.....	42	5.2 實驗結果與效能評估.....	43	第6章 結論與未來研究方向.....	45	參考文獻.....	47

參考文獻

- [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 Hummer, 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.