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
Group-By, Cross-Tab, and Sub-Total,
ICDE 152-159, 1996.


