

Evolution of Data Cube Computational Approaches

Rebecca Boon-Noi Tan
 Monash University, Australia

INTRODUCTION

Aggregation is a commonly used operation in decision support database systems. Users of decision support queries are interested in identifying trends rather than looking at individual records in isolation. Decision support system (DSS) queries consequently make heavy use of aggregations, and the ability to simultaneously aggregate across many sets of dimensions (in SQL terms, this translates to many simultaneous group-bys) is crucial for Online Analytical Processing (OLAP) or multidimensional data analysis applications (Datta, VanderMeer, & Ramamritham, 2002; Dehne, Eavis, Hambrusch, & Rauchaplin, 2002; Elmasri & Navathe, 2004; Silberschatz, Korth & Sudarshan, 2002).

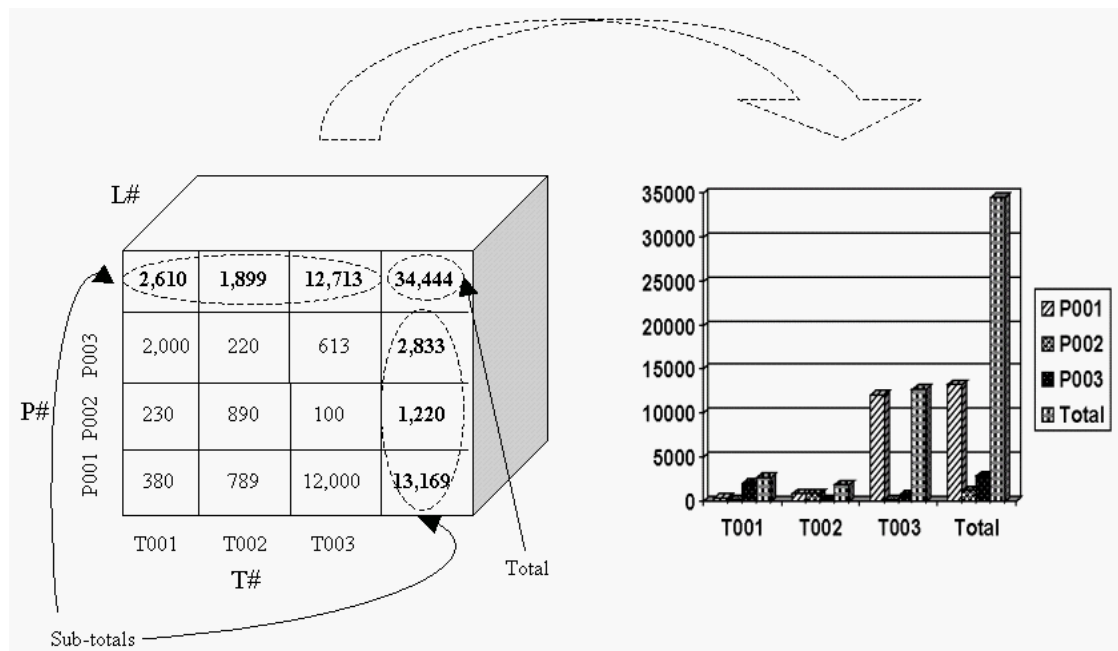
BACKGROUND

Although aggregation functions together with another operator, Group-by in SQL terms, have been widely used

in business application for the past few decades. Common forms of data analysis include *histograms*, *roll-up totals* and *sub-totals for drill-downs* and *cross tabulation*. The three common forms are difficult to use with these SQL aggregation constructs (Gray, Bosworth, Lyaman, & Pirahesh, 1996). An explanation of the three common problems: (a) *Histograms*, (b) *Roll-up Totals and Sub-Totals for drill-downs*, and (c) *Cross tabulation* will be presented.

Firstly the problem with *histograms* is that the SQL standard Group-by operator does not allow a direct construction with aggregation over computed categories. Unfortunately, not all SQL systems directly support histograms including the standard. In standard SQL, histograms are computed indirectly from a table-valued expression which is then aggregated. However, the cube-by operator is able to have direct construction. The histogram can be easily obtained from the data cube using the cube-by operator to compute the raw data. The histogram on the right-hand side of Figure 1 shows the sales amount as well as the sub-totals and total of a range of the

Figure 1. An example showing how a histogram is formed from results obtained from a data cube



6 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/evolution-data-cube-computational-approaches/10643

Related Content

General Model for Data Warehouses

Michel Schneider (2005). *Encyclopedia of Data Warehousing and Mining* (pp. 523-528).

www.irma-international.org/chapter/general-model-data-warehouses/10653

Development of Control Signatures with a Hybrid Data Mining and Genetic Algorithm

Alex Burns, Shital Shah and Andrew Kusiak (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications* (pp. 2226-2247).

www.irma-international.org/chapter/development-control-signatures-hybrid-data/7757

Data Mining with Incomplete Data

Hai Wang and Shouhong Wang (2005). *Encyclopedia of Data Warehousing and Mining* (pp. 293-296).

www.irma-international.org/chapter/data-mining-incomplete-data/10610

Web Mining in Thematic Search Engines

Massimiliano Caramia and Giovanni Felici (2005). *Encyclopedia of Data Warehousing and Mining* (pp. 1201-1205).

www.irma-international.org/chapter/web-mining-thematic-search-engines/10780

Mining Clinical Trial Data

Jose Ma. J. Alvir, Javier Cabrera, Frank Caridi and Ha Nguyen (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications* (pp. 3675-3693).

www.irma-international.org/chapter/mining-clinical-trial-data/7857