Fast Sparse Data Cube Computation

Technology #ms97-03-27b

“Lead Inventor: Kenneth A. Ross, Ph.D.

Data Cube Query Limited by Sparse Data:
Datacube queries compute aggregates over database relations at a variety of granularities, and they constitute an important class of decision support queries. An example of a data cube query is “Broken down by supplier, part and month, find the total sales in 1996, including all subtotals across each dimension.” Real-world data is frequently sparse, and hence efficiently computing datacubes over large sparse relations is important. Current techniques for computing data cubes over sparse relations do not scale well with the number of attributes, especially when the relation is much larger than the memory.

Data Cube Queries Aggregated for Sparse Data Sets:
The technology is a novel algorithm for fast computation of data cubes over sparse relations. The computation of data cubes is dominated by the I/O operations. When the relation fits in the memory, the technology does not incur any overhead besides the input of the relation and output of the data cube itself. When the relation does not fit in the memory, the algorithm uses a divides the problem of computing data cube into small problems, computes data cube over them, and then combines the result.

Applications:
• Computing aggregate queries over multidimensional sparse data sets
• The sparse and multidimensional data sets are found in various applications such as supply chain, retail, geographical information systems, and imaging

Advantages:
• An efficient algorithm for computing data cubes when the relation fits within the memory
• A divide and conquer approach when the relation does not fit within the main memory. The approach can be parallelized on multiple machines.

Patent Status: Patent Issued (US 5,987,467) ~ see link below.

Licensing Status: Available for Licensing

Inventors

Divesh Srivastava