

On the Scalability of Multidimensional Databases

István Szépkúti

It is commonly accepted in the practice of on-line analytical processing databases that the multidimensional database organization is less scalable than the relational one. It is easy to see that the size of the multidimensional organization may increase very quickly. For example, if we introduce one additional dimension, then the total number of possible cells will be at least doubled.

However, this reasoning does not take into account that the multidimensional organization can be compressed. There are compression techniques, which can remove all or at least a part of the empty cells from the multidimensional organization, while maintaining a good retrieval performance.

Relational databases often use B-tree indices to speed up the access to given rows of tables. It can be proven, under some reasonable assumptions, that the total size of the table and the B-tree index is bigger than a compressed multidimensional representation. This implies that the compressed array results in a smaller database and faster access at the same time.

This paper compares several compression techniques and shows when we should and should not apply compressed arrays instead of relational tables.