Data Reorganization Techniques to Improve Compression of Bitmap Indices

“Lead Inventors: John R. Kender Ph.D. & Hassan H. Malik Ph.D.

Compressed Bitmap Indices in Commercial Databases Take Space, Time to Query:
Many commercial database systems use Run-Length Encoding (RLE) based compressed bitmap indices (such as Oracle’s BBC, or WAH) to execute queries. A compressed bitmap index is a special kind of index that stores the bulk of its data as bitmaps, compressed in a way that allows performing logical operations without first requiring to decompress the bitmaps. Queries are executed in most cases by performing bitwise logical operations on these indices. The storage space and time needed to execute queries depends directly on the sizes of the indices used to execute the query.

The sizes of these compressed bitmap indices are driven by the availability of long consecutive 0 or 1 bits in the original uncompressed bitmap (i.e., a large number of consecutive 0 bits can be compressed in a single “run”, same for 1 bits). If a large number of consecutive 0 or 1 bits are not found, the indices consume a lot more space and the queries take more time to execute. The default ordering of transactions may not be optimal and may not result in smallest possible index bitmaps. Reorganizing bitmaps to reduce the number of bit shifts, and thereby increase compressibility (via RLE), would directly improve storage efficiency and query execution performance. As databases power many transactions over the internet, this boost in efficiency would be a welcome addition to database product lines.

Reorganizing Bitmaps Increases Compressibility, Improve Data Storage, and Query Execution:
This technology is an algorithm to improve the compressibility of bitmap indices. It focuses on maximizing the overall similarity of neighboring rows, instead of processing on a column-by-column basis as most methods do. It makes use of Hamming distances, which can be computed efficiently. The first algorithm, HDO (Hamming Distance Order) finds the most suitable candidate rows for each position in the final re-ordered dataset, and then applies local tie-breaking criteria to select the best candidate. The second, aHDO, is a linear approximation to HDO and reorders the whole index in a user-defined number of iterations. These steps can be performed as a pre-compression step, which will improve the performance of existing commercial database systems.

Applications:
• Provides a storage and performance benefit to databases using RLE compression on bitmap indices
• Can also be used in other contexts where RLE compression is used, such as images and sound data (not just limited to databases)

Advantages:
• Substantially decreases the storage required for bitmap indices, provides between 5% and 400% space savings
over the original ordered compressed bitmap, resulting in similar query processing performance improvements

- Provides a fast, linear-time implementation that can be used for large databases
- Does not require modifying the query execution engine, or the code for compression. Process can be added as a pre-compression plug-in, which will be more palpable to IT managers

Patent Status: Copyright

Licensing Status: Available for Licensing and Sponsored Research Support

Inventors

Hassan Haider Malik