Parallelizing Bitmap Operations

CSE 6441 Term Project Proposal

Omid Asudeh

Proposal:

Bitmap indexes, supported by fast bitwise operation, are a widely used structure in the database area from query optimization to approximate aggregate queries. A challenge is that as the number of elements with in the the bitmap increases, the length of its bitvectors also increase, so at a certain time it cannot fit in the memory. To remedy this issue there has been several compression methods studied in the literature among which the Word Aligned Hybrid, WAH [1], method is quite common. An advantage of WAH compression approach is that the bitwise operations like bitwise_and and bitwise_or can be done on them without uncompressing. At this project, I am going to study the common bitmap operations such as bitwise_and, bitwise_or, bit_count, compress, and uncompress and try to take advantage of parallelism in order to increase their performance. There has been a previous work [2] on parallelizing the compress and uncompress methods on the GPUs, yet there is no work on parallelizing the bitwise_and, bitwise_or and bit_count *without uncompressing the bitvectors*.

Initial list of references:

[1] Wu, Kesheng, et al. *Notes on design and implementation of compressed bit vectors*. Technical Report LBNL/PUB-3161, Lawrence Berkeley National Laboratory, Berkeley, CA, 2001.

[2] Andrzejewski, Witold, and Robert Wrembel. "GPU-WAH: Applying GPUs to compressing bitmap indexes with word aligned hybrid." In *International Conference on Database and Expert Systems Applications*, pp. 315-329. Springer, Berlin, Heidelberg, 2010.

[3]Chen, Zhen, et al. "A survey of bitmap index compression algorithms for big data." *Tsinghua Science and Technology* 20.1 (2015): 100-115.