A Framework to Process Iceberg Queries using Set-intersection and Set-Difference Operations

Abstract

Many data mining queries are basically identified as iceberg queries. Applications are required to be compute aggregate functions over an interesting attributes to find aggregate values above some specified threshold. Such queries are called as iceberg queries. We propose set operations instead of bitwise-AND operations to evaluate iceberg queries efficiently using very little memory and significantly fewer passes over data, as compared to current techniques that use Dynamic pruning approaches and Vector alignment algorithms. Set operations reduces the execution time and make evaluation process of Iceberg query very effective by reducing the number of bitmaps that are needed. The exhaustive experimentation gives better results than existing strategies.

References

- Bin He, Hui-I Hsiao, Ziyang Liu, Yu Huang and Yi Chen, "Efficient Iceberg Query Evaluation Using Compressed Bitmap Index", IEEE Transactions On Knowledge and Data
A Framework to Process Iceberg Queries using Set-intersection and Set-Difference Operations

Engineering, vol 24, issue 9, sept 2011, pp. 1570-1589


- J. Bae and S. Lee, &quot;Partitioning Algorithms for the Computation of Average Iceberg Queries&quot; ; in DaWaK, 2000.


- M. Jorgens &quot;Tree Based Indexes vs. Bitmap Indexes: A Performance Study&quot; ; In DMDW, 1999.


- Spiegler I; Maayan R &quot;Storage and retrieval considerations of binary databases&quot; ; Information processing and management: an international journal 21 (3): pages 233-54, 1985.

Index Terms
Keywords
Database  Iceberg query  Bitmap vector  Set intersection  set difference and
Threshold