Abstract

Performance of the data warehouse depends on physical design. Index selection and storage of multidimensional data bases are important activities of physical designing process. Conventional indexing techniques such as bitmaps, B-trees and hash based indexing systems need large storage space for storing indexes along with data itself. Spelling variants, misspellings and transliteration differences are source of uncertainty in data with in the databases. Misspelled and distorted key values are also hard to map in present indexing systems. In this paper neural network based physical design is suggested, a class of artificial neural network known as self-organizing net is used for indexing data warehouse at physical level. Indexes of active neurons will be used for generating indexes for the data values. In conventional indexing techniques every key value is mapped to a specific point in space, while in neural network based database indexing system, every key value is mapped to a region in space. This region is a class to which the key values of similar type belong. Indexes generated through this method used optimal space for storage, as only final weight matrices after training of neurons are stored. Self-organizing net based indexing is very robust as distorted key values get indexed to right classes. Accuracy of our self-organizing net based indexing system in
mapping key values with distorted keys is found to be high.

Reference

http://dx.doi.org/10.1023/A:1009777418785
system for multidimensional analysis and OLAP, 13th Int'l symposium on parallel and
distributed processing.
aggregation operator generalizing group-by, cross-tabs, and sub-totals. In Proc. of the 12th Int'l
Conference on Data Engineering, pp 152-159.
of Data (Montreal, Quebec, Canada, June 04 - 06, 1996). SIGMOD’96. ACM, NewYork, NY,
205216. DOI=http://doi.acm.org/10.1145/233269.2333333
24-35.
techniques for multi-dimensional interval data. SIGMOD Rec. 20, 2 (Apr. 1991), 138-147. DOI=
http://doi.acm.org/10.1145/119995.115807
426-435. DOI= http://doi.acm.org/10.1145/119995.115861
Conventional to Spatial and Temporal Applications (Data-Centric Systems and Applications).1st
5, 19, October 2006, pp 143-161.
Reading, MA, 1989.
(Jun. 1990), 677-680. DOI= http://doi.acm.org/10.1145/78973.78978
[26] Ramakrishna M.V., Justin, Zobel (1997).,"Performance in Practice of String Hashing
Functions", Proceedings of the fifth International Conference on Database Systems for
Advanced Applications, Melbourne, Australia, April 1-4, 1997.
436-445. DOI= http://doi.acm.org/10.1145/119995.115862
Iceberg Cubes by Top-Down and Bottom-Up Integration : The StarCubing Approach, IEEE
simultaneous multi-dimensional aggregates. In Proc. ACM-SIGMOD International Conferences
on Management of Data, pp 159-170.
[31] Zhao Yihong, Tufte Kristin, Naughton F Jeffrey (1996), On the Performance of an
Array-based ADT for OLAP workloads, Technical Report CS-TR-96-1313, University of
Index Terms

Computer Science  Databases

Key words

Self-organizing net
multidimensional databases
indexing