Data Structures and Algorithms (11)

Instructor: Ming Zhang
Textbook Authors: Ming Zhang, Tengjiao Wang and Haiyan Zhao
Higher Education Press, 2008.6 (the "Eleventh Five-Year" national planning textbook)

https://courses.edx.org/courses/PekingX/04830050x/2T2014/

Content

11.5 Bit Indexing

- B tree is suitable for retrieving small amount of records
- B tree has three major flaws for interactive inquiries of complex database :
  1. B tree is almost useless when dealing with data that are rarely unique.
  2. Constructing and maintaining index is expensive in a database.
  3. Not competent for complex inquiries with grouping demands.

Bit Indexing for a table of a database

date	store	state	class	sales
3/1	32	NY	A	6
3/1	36	AL	A	9
3/1	38	NY	B	5
3/1	41	AK	A	11
3/1	43	NY	A	9
3/1	46	AK	B	3

State = NY

1
0
1
0
1
0

Class=A

1
1
0
1
1
0

state=AK	state=AL	…	state=NY
0	0		1
0	1		0
0	0		1
1	0		0
0	0		1
1	0		0

A set of n-dimensional bit vector (n is the number of records)

Signature File

Signature file

- File30: foo, bar, baz
- File40: baz, bar
- File50: foo

record	bar	baz	foo
30	1	1	1
40	1	1	0
50	0	0	1

Properties of Bit Indexing

1. Storing by "columns".
2. Column data is more easier for compressing than row data, and can save memory by 50%.
3. Memory space needed for indexing is smaller than that of an B tree.

Discussion

- Investigate bit indexing in a column database.

Data Structures and Algorithms

Thanks

the National Elaborate Course (Only available for IPs in China)
http://www.jpk.pku.edu.cn/pkujpk/course/sjjg/

Ming Zhang, Tengjiao Wang and Haiyan Zhao

Higher Education Press, 2008.6 (awarded as the "Eleventh Five-Year" national planning textbook)