



# 数据结构与算法（十一）

张铭 主讲

采用教材：张铭，王腾蛟，赵海燕 编写  
高等教育出版社，2008. 6（“十一五”国家级规划教材）

<http://www.jpk.pku.edu.cn/pkujpk/course/sjjg>



# 主要内容

- 基本概念
- 11.1 线性索引
- 11.2 静态索引
- 11.3 倒排索引
- 11.4 动态索引
- 11.5 位索引技术
- 11.6 红黑树



## 11.5 位索引技术

- B 树适合于查找并取回少量记录的情况
- 对于数据仓库的复杂交互式查询，B树有三个缺点：
  1. B 树对唯一值极少的（低基数）数据字段几乎毫无价值
  2. 在数据仓库中构造和维护索引的代价高
  3. 对于带有分组及聚合条件的复杂查询无能为力



## 对于数据库表的位图索引

长为  $n$  的位向量的集合  
( $n$  为文件的记录数)

date	store	state	class	sales	State = NY	Class=A
3/1	32	NY	A	6	1	1
3/1	36	AL	A	9	0	1
3/1	38	NY	B	5	1	0
3/1	41	AK	A	11	0	1
3/1	43	NY	A	9	1	1
3/1	46	AK	B	3	0	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



## 特征文件

Signature file ( 也译为 “签名文件” )

- 文件30 : foo, bar, baz
- 文件40 : baz, bar
- 文件50 : foo

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



## 位图索引特点

1. 按“列”为单位存储数据
2. 列数据比行数据更易进行压缩，  
可节省 50% 的磁盘空间
3. 索引空间比 B 树小



## 思考

- 调研列数据库中的位图索引



# 数据结构与算法

谢谢聆听

国家精品课 “数据结构与算法”

<http://www.jpk.pku.edu.cn/pkujpk/course/sjjg/>

张铭，王腾蛟，赵海燕

高等教育出版社，2008. 6。 “十一五” 国家级规划教材