PRINCIPLES OF DATA ORGANISATION

Bitmaps

MOTIVATION

- Key, pointer pairs ~ index
  - Abstract definition
- Bit = 1/0
  - True/False properties

BITMAPS

- Indexing of attributes with ‘small’ domains
  - E.g., status (active/inactive), level (low/medium/high)
- Records are stored in a primary file
  - E.g., Person(name:string, age:integer, address:string, gender:Boolean)
- For values having small domains we create a bitmap index (see the next slide) to answer special queries fast
  - E.g., The percentage of female employees
- For each value of the domain a vector of bits is stored telling which objects share the given property → array of bits
  - Size of the bitmap equals the number of records and each record is therefore related to exactly one position in the bit string
    - The position is the same in the primary file and in the bitmap index → we do not need to read the whole primary file to find the full record
  - When a record has a given value, the corresponding bit in the corresponding bitmap is turned on
  - Querying using bitwise logical operations

BITMAPS – EXAMPLE

Who works in research?
Who works in design and marketing?
Who works in research and not design?
Who works in research and is female?

Employees (gender)									
	Janice	Michael	Sharon	David	Kevin	John	Mary	Terry	Jill
Male	0	1	0	1	1	1	0	1	0
Female	1	0	1	0	0	0	1	0	1

Employees (department)									
	Janice	Michael	Sharon	David	Kevin	John	Mary	Terry	Jill
Design	0	1	1	0	0	1	1	1	0
Research	0	0	0	1	1	0	1	0	1
Marketing	1	0	0	0	0	1	0	0	0

BITMAPS – SPACE COMPLEXITY

- Grow linearly with the database size
- Can be read by large blocks
  - Data are saved by lines - one line per file
- Can be compressed
  - E.g., most people work in a single department → lots of zeroes that can be stored efficiently
  - Complication: update

- What is a small domain?
  - Does not have to be 0/1
  - Can be 3, 4, 5 … 10, … 23, …
  - In practice: hundreds of distinct values are still usable

A bitmap file for **2^19** (524,288) records with values 0/1 has 64 KiB.
For domain of size 10 we have 640 KiB.

Who has medium salary?

BITMAPS – EXAMPLE

Person(name:string, age:integer, address:string, gender:Boolean, salary:[low,medium,high])

Employees (salary)									
	Janice	Michael	Sharon	David	Kevin	John	Mary	Terry	Jill
Low	1	0	0	0	0	1	1	0	0
Medium	0	1	1	1	1	0	0	0	0
High	0	0	0	0	0	0	0	1	1

Row = separate file

low = 00, medium = 01, high = 11 (and a free option)

Employees (salary)									
	Janice	Michael	Sharon	David	Kevin	John	Mary	Terry	Jill
1. Bit	0	0	0	0	0	0	0	1	1
2. Bit	0	1	1	1	1	0	0	1	1

We read both files

Remember: we read blocks, not bits!