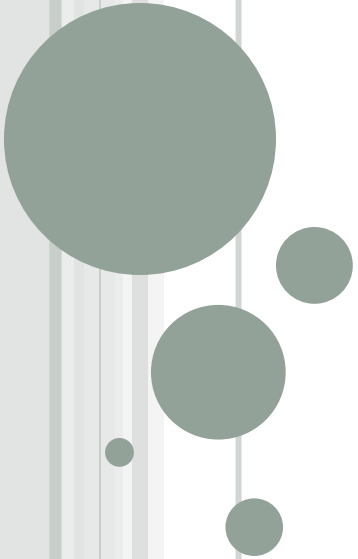

Indexing Methodology: I Discussion Week 6

Database Systems (H)
Dr Chris Anagnostopoulos

Task: Secondary Index

Context:

- Consider a file with b data blocks.
- Build Secondary Index with $m < b$ blocks over the *non-ordering, non-key* attribute department number DNO $\in \{1, 2, \ldots, 10, \ldots\}$
- X: number of employees working on DNO = 10 *per* data block. That is, given a block, there are X employees of Dept. 10.
- $P(X \geq 1) = 0.5$, i.e., the probability that at least one employee works in DNO = 10 is 50% within a block. That is, when we pick up a block at random, the probability of finding therein *at least* one employee of Dept. 10 is 0.5
- We can fit f block-pointers per block, i.e., f is the number of the block-pointers (e.g., physical addresses), which can fit in one block.

Task:

- Which is the expected cost (#block accesses) retrieving the employees of DNO = 10 using the Index? (expressed as a function of b, f, and m)