Uni Freiburg, Web Science Group
Prof. Peter Fischer
Systems Infrastructure for Data Science - Winter 2012/13

Exercise Sheet #6: Performance Tuning and Benchmarking in Databases

December 7, 2012

Exercise 6.1 : Index Tuning

Consider the following relations:

$$Emp(\underline{eid}, ename, salary, age, did)$$

$$Dept(\underline{did}, budget, floor, mgr\_eid)$$

The attributes *eid*, *did*, and *mgr_eid* refer to Employee-ID, Department-ID and Manager's ID respectively. Salaries range from 10,000 to 100,000 EUR, ages vary from 20 to 80, each department has about five employees on average, there are 10 floors, and budgets vary from 10,000 to 1 million EUR. Assume uniform distributions of values.

A. Given the query,

> *Print ename, age, and salary for all employees.*

which of the listed index choices would you choose to speed up the query-processing? Would your answer change if your database system does not consider index-only plans (i.e., data records are always retrieved even if enough information is available in the index entry)?

- Clustered hash index on $(ename, age, salary)$ fields of Emp.
- Unclustered hash index on $(ename, age, salary)$ fields of Emp.
- Clustered B+ tree index on $(ename, age, salary)$ fields of Emp.
- Unclustered hash index on (eid, did) fields of Emp.
- No index.

B. Answer the above question for the query:

> *Print did of departments that are on the 10th floor and have a budget of less than 15,000 EUR.*

You have the following index choices:

- Clustered hash index on the $floor$ field of Dept.
- Unclustered hash index on the $floor$ field of Dept.
- Clustered B+ tree index on $(floor, budget)$ fields of Dept.
- Clustered B+ tree index on the *budget* field of Dept.
- No index.

C. Suppose the database has a clustered index on *eid* and an unclustered index on *age* for the relation *Emp*. Give an example of an update on this relation that will be definitely speeded up because of the available indices. Now, give an example of an update that will be definitely slowed down because of the available indices.

Exercise 6.2 : 5-Minute and 1-Minute Rules

	1987	**1997**	**Now**
RAM	5000$/MB	15$/MB	5$/GB
Page Size	1KB	8KB	?
Disk	15000$	2000$	100$
IOPS	15	64	180

Table 1: Changes of disk, memory characteristics over years

	RAM	**Flash disk**	**SATA disk**
Price and capacity	3$/8x64 MBit	200$/256 GB	150$/1024 GB
Access latency		0.1 ms	12 ms
Transfer rate		500 MB/s	150 MB/s
4KB reads/s		9250	83
256KB reads/s		1630	73

Table 2: Prices and performance of flash and disks.

A. Suppose the price of memory falls by half, and the speed of disk access (number of accesses per second) doubles, while all others factors remain the same. What would be the effect of this change on the 5 Minute and 1 Minute rule?

B. Considering Table 1, what should the current page size roughly be for the 5-Minute rule to be still valid?

C. Now consider that we have an enhanced memory hierarchy with flash memories between RAM and disk. Given Table 2, compute the page sizes for which the 5-minute rule holds for the cases of page movements between RAM-SATA, RAM-Flash and Flash-SATA.