EPL646 – Advanced Topics in Databases
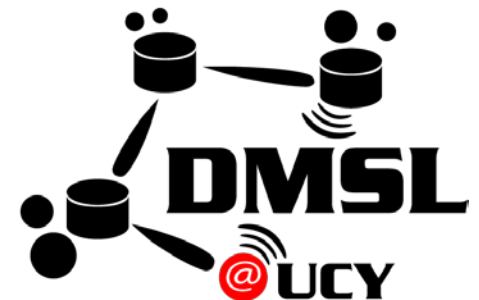# Part A:
Storage and Indexing Overview

http://www.cs.ucy.ac.cy/~dzeina/courses/epl646/labs/lab.html

Question 1

(Exercise 8.3) Consider a relation stored as a randomly ordered file for which the only index is an unclustered index on a field called sal. If you want to retrieve all records with sal > 20, is using the index always the best alternative? Explain.

Question 2

(Exercise 8.9) What main conclusions can you draw from the discussion of the five basic file organizations discussed in Section 8.4? Which of the five organizations would you choose for a file where the most frequent operations are as follows?

- Search for records based on a range of field values.
- Perform inserts and scans, where the order of records does not matter.
- Search for a record based on a particular field value

Question 2

File Type	Scan	Equality Search	Range Search	Insert	Delete
Heap					
Sorted					
Clustered					
Unclustered tree index					
Unclustered hash index					

B = Number of data pages when records are packed onto pages with no wasted space
D = Average time to read or write a disk page
C = Average time to process a record
R = Number of records per page
H = Time required to apply the hash function to a record
F = Fan-out (tree indexes). Typically at least 100

University of Cyprus

DMSL @UCY

Question 3

(Exercise 8.11) Consider the following relations:

          Emp(_eid_: integer, _ename_: varchar, _sal_: integer, _age_: integer, _did_: integer)

          Dept(_did_: integer, _budget_: integer, _floor_: integer, _mgr_eid_: integer)

Salaries range from $10,000 to $100,000, 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. You can assume uniform distributions of values.

For each of the following queries, which of the listed index choices would you choose to speed up the query? 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), how would your answer change? Explain briefly.

Query: Print ename, age, and sal for all employees.

a) Clustered hash index on *(ename, age, sal)* fields of Emp.
b) Unclustered hash index on *(ename, age, sal)* fields of Emp.
c) Clustered B+ tree index on *(ename, age, sal)* fields of Emp.
d) Unclustered hash index on *(eid, did)* fields of Emp.
e) No index.

Query: *Find the dids of departments that are on the 10th floor and have a budget of less than $15,000.*

a)   Clustered hash index on the *floor* field of Dept.
b)   Unclustered hash index on the *floor* field of Dept.
c)   Clustered B+ tree index on *(floor, budget)* fields of Dept.
d)   Clustered B+ tree index on the *budget* field of Dept.
e)   No index.

University of Cyprus

DMSL @UCY

Questions?

http://www.cs.ucy.ac.cy/~dzeina/courses/epl646/labs/lab.html