CS186 Discussion Section Week 2
File Organization and Indexing
Fall 2013

1. What two formats are used for variable length records? What are the pros and cons of each?

2. Why would you use fixed length records? Can you give specific examples? Can you give specific examples of when you would use them?

3. Why would you use variable length records? Can you give specific examples of when you would use them?

4. Given: CREATE TABLE Students (sid VARCHAR(20) not null, name VARCHAR(20), login VARCHAR(10), age INTEGER, gpa FLOAT not null)

What would be the size of a single tuple using fixed length records?

What would be the maximum size of a single tuple using the two types of variable length records? Assume array offsets are stored as ints and field delimiters are one byte.

What would be the minimum size of a single tuple using the two types of variable length records? Assume array offsets are stored as ints.

5. Why do we use slotted pages?

6. Let's say you have 52 bytes of free space in your slotted page. What is the biggest record you can add to the page? Let's assume that the entire page is only 256 bytes and you want to minimize the space the slot directory takes up.

7. Fill in the following table for the I/O costs of operations (B = # disk blocks in file)

Assumptions: Single record insert and delete.

For heap files, insert allows appends to end of file, delete leaves free space in page (don't have to worry about compacting data)

For sorted files, no gaps allowed and pages are fully packed and files are compacted after deletion. We are assuming worst case here.

	Heap File	Sorted File
Scan All Records	B	
Equality Search (1 match)		
Range Search		
Insert		
Delete		

8. What are important factors in determining whether or not you should add an index to add to a table?

9. Consider the table Enrolled(sid, course, grade) with 500 pages, 6,000 tuples and the query SELECT * FROM Enrolled where sid > 4500. Assume SIDs are unique and range from 0 to 6000.

How many I/Os would this query take if the table was stored in a heap file?

How many I/Os would this take if the table was stored in a sorted file sorted by grade?

How many I/Os would this take if the table was stored in a sorted file sorted by SID?

10. True or False? Given the table Students(sid, gpa, age), a hash index on gpa will significantly increase the performance of the following query: SELECT * FROM Students WHERE age > 20;

11. True or False? Given the table Students(sid char(20), gpa float, age integer), a clustered tree based index on gpa will increase the performance of the following query: SELECT * FROM Students where age > 20 AND gpa > 3.5;