

CS150A Quiz02

File Formats

Assume that each page in our system can hold 128 KB (1 KB = 1024bytes), integers are 32-bits wide, and bytes are 8-bits wide.

Consider the following relation:

```
CREATE TABLE Submissions (  
    record_id integer UNIQUE,  
    assignment_id integer,  
    student_id integer,  
    time_submitted integer,  
    grade_received byte,  
    PRIMARY KEY (assignment_id, student_id)  
);
```

Assume the column record_id corresponds to the row's actual record ID.

Q1: How large (in bytes) is a record?

$$4 \times 4 + 1 = 17 \text{ byte.}$$

Q2: Suppose we begin each page with a 32-bytes header plus a bitmap. At most, how many records can fit in an unpacked page?

Page: $8 \times 128 \times 1024 = 2^{20}$ bit

Record: $17 \times 8 + 1 = 137$ bit

Header: $32 \times 8 = 256$ bit

$$\frac{\text{Page-head}}{\text{Record}} = \frac{2^{20} - 256}{137} = \frac{2^{20} - 256}{137} = 7651 \text{ records.}$$

We add two variable-length fields to our table schema. Now our table looks like this:

```
CREATE TABLE Submissions (  
    record_id integer UNIQUE,  
    assignment_id integer,  
    student_id integer,  
    time_submitted integer,  
    grade_received byte,  
    comment text,  
    regrade_request text,  
    PRIMARY KEY (assignment_id, student_id)  
);
```

We decide to use slotted pages to store the variable length records. Each page begins with a 32-bytes header plus a slot directory. (Assume this header contains information such as the number of valid records in the page.) Each pointer inside the slot directory consumes 20bits/record, while the record header storing field offsets is 32 bits wide.

→ maximum record: all text NULL

Q3: What is the **maximum** number of records that can fit in our slotted pages?

$$\text{Page} = 2^{20} \text{ bit}$$

$$\text{Header} = 32 \times 8 = 256 \text{ bit}$$

$$\text{Record} = 17 \times 8 + 20 + 32 = 188 \text{ bit}$$

$$\text{Number of record} = \left\lfloor \frac{2^{20} - 256}{188} \right\rfloor = 5576$$

CD

Q4: We decide to squash the two text fields together into one field using a semicolon separator character (;), which allows us to shrink the record header from 32 bits to 16 bits at the cost of 8 bits (for the semicolon). For example, the columns ("Submitted late", "Dog ate my homework") get compressed into "Submitted late;Dog ate my homework". Which of the following are true with this new scheme?

Check all that apply.

- A. Professor Gonzales can enter the comment "Fantastic work; good job!"
- B. Fewer records will fit in a page
- C. More records will fit in a page
- D. It is possible for the query "SELECT grade_received FROM Submissions" to finish faster