Problem 1

1.1

Size = Header + name + student + birthday + State

$= 4 + 1 + 1 + 8 + 1 = 15B$

1.2.

$4+1+1+8+1 = 15B$

↓ word-allignment

$(4+1+1+2)+(8)+(1+7) = 24B$

At page level:

slot count + free space pointer $= 8+8 = 16B$

For each record:

record pointer + record length $= 8+8 = 16B$

Totally we have:

$16+(24+16)x = 16+40x \ (x \in N)$

$\because 16+40x \le 2048 \ (x \in N)$

$\therefore x \le 50.8 \ (x \in N)$

$\therefore$ max $x = 50$

$\therefore$ maximum number $= 50$

1.3. reorder:

birthday + Record Header + name + student + state + padding

$= (8) + (4+1+1+1+1) = 16B$

$\therefore 16+(16+16)x \le 2048$

$x \le 63.5$

$\therefore$ max $x = 63$

Problem 2.

2.1

DSM: for size: 100 Pages

for credits: 100 Pages

$\therefore$ total: 200 pages

NSM: read both size and credits: all tuples are needed

$\because$ worst cases

$\therefore$ 500 pages.

2.2.

DSM: min: 1 for id 3 for name instructor size = 4 pages

max: 100 for id 2x3 for name instructor size = 106 pages

NSM: min 1 for all = 1 pages

max: worst cases: 500 pages.

Problem 3: Bloom Filter:

3.1:

insert 7:

$h_1(7) = 7$ 0.7 mod 10=

$h_2(7) = 0$

set 0th and 7th to 1:

[1 0 0 0 0 0 0 1 0 0]

insert 27

$h_1(27) = 7$

$h_2(27) = 2$

set 7th and 2nd to 1:

[1 0 1 0 0 0 0 1 0 0]

3.2

Because: Hash functions can collide: different element may map to same bits

When number of elements increases, more bits are set to 1. Causing bit overlaps.

3.3.

set : the false positive probability p for a Bloom Filter

Assume that a hash function selects each array position with equal probability.

$\therefore$ the probability that a certain bit is not set to 1 for 1 hash function is

$1-\frac{1}{m}$

$\because$ we have k hash functions and each has no significant correlation between each other

$\therefore$ the probability that a certain bit is not set to 1 is

$(1-\frac{1}{m})^k$

$\because \lim_{m\to\infty}(1-\frac{1}{m})^m=\frac{1}{e}$

$\therefore (1-\frac{1}{m})^k = ((1-\frac{1}{m})^m)^{k/m} \approx e^{-k/m}$

$\because$ insert n elements

$\therefore (1-\frac{1}{m})^{kn} \approx e^{-\frac{kn}{m}}$

$\therefore$ the probability that it is 1 is

$1-(1-\frac{1}{m})^{kn} \approx 1-e^{-\frac{kn}{m}}$

$\therefore p \approx (1-e^{-\frac{kn}{m}})^k$

where m is the size of the bit array
n is the number of elements inserted
k is the number of hash functions

we choose $k=\frac{m}{n}\ln 2$

$\Rightarrow p \approx (\frac{1}{2})^k < 0.01$

$\Rightarrow k > \log_2(100) \approx 6.64 \quad k \in N^+$

$\therefore$ set $k=7$

$\therefore m \approx \frac{kn}{\ln 2} = \frac{7000}{\ln 2} \approx 10100$

$\therefore$ we can choose $m \approx 10000$. and $k=7$