MILITARY COLLEGE OF SIGNALS
FINAL EXAM
BESE 15-A
CPS 480 Database Systems

Instructor: A/P Dr. Imran Siddiqi **Time: 2.5 Hours**
Max Marks: 50

Note: This question paper comprises **3** pages.

(3+6)

1. **a.** Consider the following relation.

X	Y	Z
x1	*y1*	*z1*
x1	*y1*	*z2*
x2	*y1*	*z1*
x2	*y1*	*z3*

List all the functional dependencies that this instance satisfies.

The following functional dependencies hold over *R*: $Z \rightarrow Y$, $X \rightarrow Y$, and $XZ \rightarrow Y$

b. Consider a relation R with four attributes ABCD. For each of the following sets of functional dependencies, identify the candidate key(s) for R and state the best normal form that R satisfies (1NF, 2NF, 3NF, or BCNF). (Note: All three parts below are independent of each other and you need to attempt them separately)

i. $C \rightarrow D, C \rightarrow A, B \rightarrow C$
(a) Candidate keys: *B*
(b) *R* is in 2NF but not 3NF.

ii. $B \rightarrow C, D \rightarrow A$
(a) Candidate keys: *BD*
(b) *R* is in 1NF but not 2NF.

iii. $A \rightarrow B, BC \rightarrow D, A \rightarrow C$
(a) Candidate keys: *A*
(b) *R* is in 2NF but not 3NF (because of the FD: $BC \rightarrow D$).

(4+4+2+6)

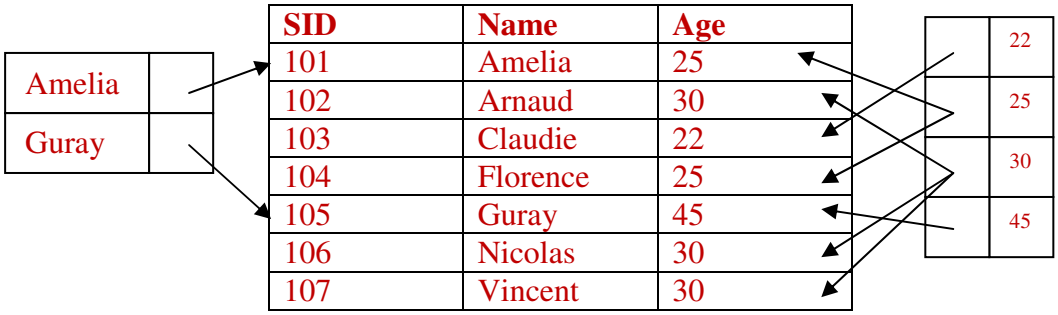
2. **a.** Consider the *Student* relation shown in the following.

SID	Name	Age
101	Amelia	25
102	Arnaud	30

103	Claudie	22
104	Florence	25
105	Guray	45
106	Nicolas	30
107	Vincent	30

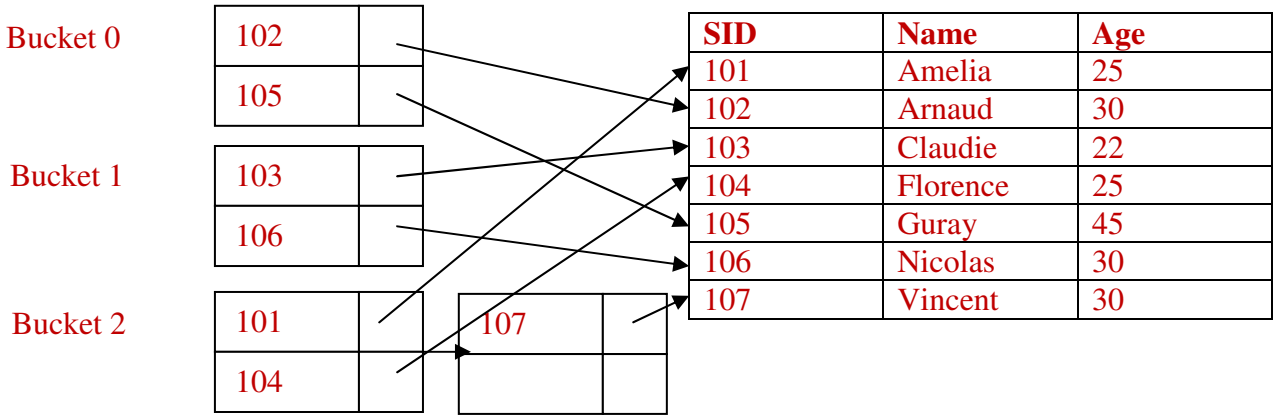
i. Show the sparse index on the field '*Name*' where the index file contains entries '*Ameila*' and '*Guray*'

ii. Show the secondary index on the field 'Age'.



b. Organize the *Student* relation shown above using hash indexing on the search key *SID*. The following parameters are given:

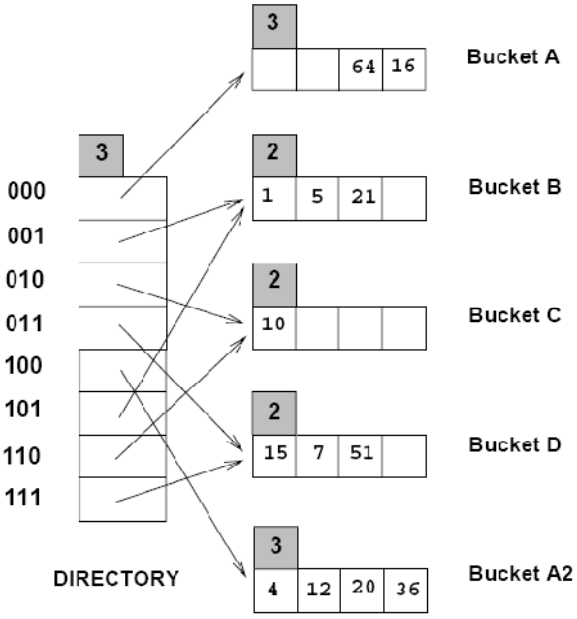
Hash Function = (Sum of Digits in SID) mod 3
Number of buckets = 3
Capacity of each bucket = 2



c. How the hash indexing you performed in part *b* (above) is different from hash file

In hash indexing the buckets contain the key attributes and point to the records while in hash file the buckets contain the actual records.

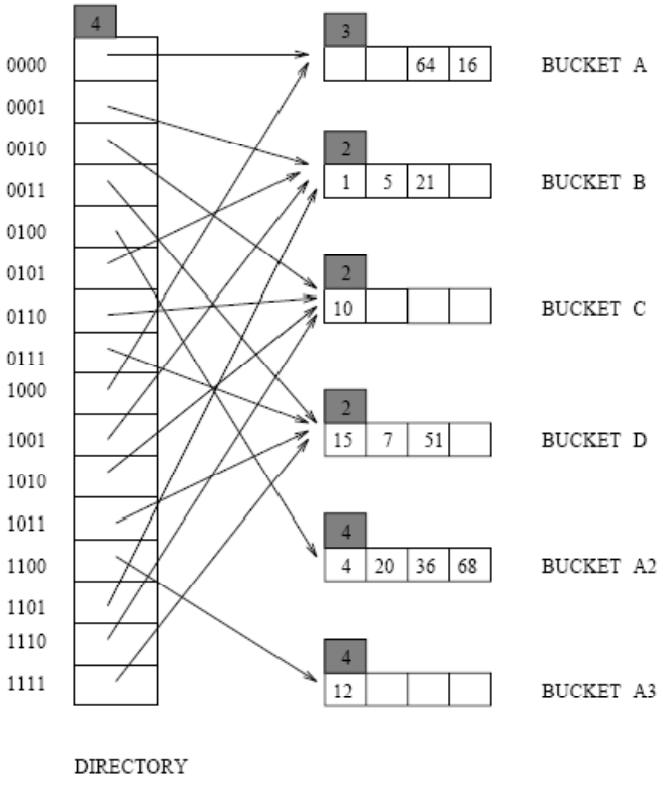
d. Consider the extendible hashing index shown in the following.

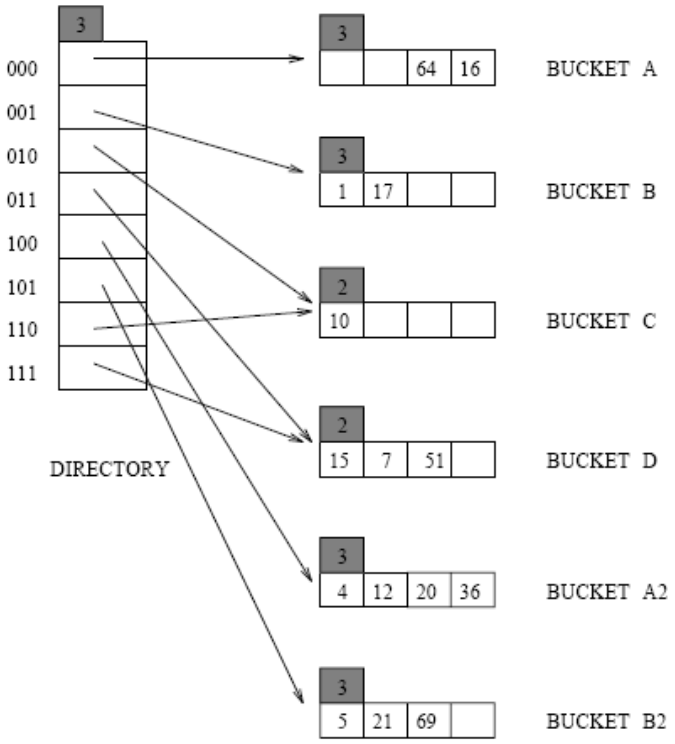


i. Show the index after inserting an entry with hash value 68.

ii. Show the index after inserting entries with hash values 17 and 69 to the original index.

Solution





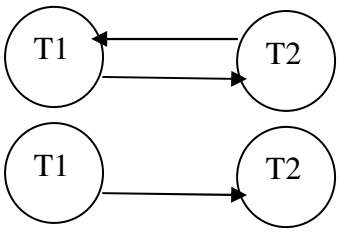
(3+3)

3. **a.** Using precedence graph, find if the following schedules are conflict serializable or not.

T1: READ(A) READ(B) WRITE(B)
T2: READ(A) READ(B) WRITE(B)

T1: READ(A) WRITE(A) READ(B) WRITE(B)
T2: READ(A) WRITE(A) READ(B) WRITE(B)

Solution



b. Given two transactions *T1* and *T2* and two data objects *A* and *B*, give an example schedule which results in:

i. A *write-write* conflict
ii. A *read-write* conflict

Solution is not unique. One example could be the schedule one in part a above.

(4+6)

4. **a.** Assume you are given a document database of six documents containing the text as indicated in the following.

Document	Text
0	pease porridge hot pease porridge cold
1	pease porridge in the pot
2	nine days old
3	some like it hot some like it cold
4	some like it in the pot
5	nine days old

Show the results of creating an inverted file index on the given documents.

Word	Documents
pease	{D0,0},{D0,3}, {D1,0}
porridge	{D0,1}, {D0,4}, {D1,1}
hot	{D0,2}, {D3,3}
cold	{D0,5}, {D3,7}
in	{D1,2}, {D4,3}
the	{D1,3}, {D4,4}
pot	{D1,4}, {D4,5}
nine	{D2,0}, {D5,0}
days	{D2,1}, {D5,1}
old	{D2,2},{D5,2}
some	{D3,0},{D3,4},{D4,0}
like	{D3,1},{D3,5},{D4,1}
it	{D3,2},{D3,6},{D4,2}

b. Consider the four documents below:

Document	Words
0	Apple, Cell
1	Doll, Goat
2	Elephant, Ink
3	House, Fruit

Let the hashing function be the 5 bit binary representation of the first character of each word. For simplicity, assume A=1, B=2, C=3 and so on.

i. Show the result of creating a signature file for the above documents.
ii. What documents are retrieved if a user queries the term '*Elephant*'.

Hash(Apple) = 00001
Hash(Cell) = 00011
Hash(Doll)= 00100
Hash(Goat) = 00111
Hash(Elephant) = 00101
Hash(Ink) = 01001
Hash(House)=01000

Hash(Fruit) = 00110

Signature D1 = 00001 OR 00011 = 00011
Signature D2 = 00100 OR 00111 = 00111
Signature D3 = 00101 OR 01001 = 01101
Signature D4 = 01000 OR 00110 = 01110

Elephant = 00101

Elephant AND D1 = 00001
Elephant AND D2 = 00101
Elephant AND D3 = 00101
Elephant AND D4 = 00100

D2 and D3 are retrieved.

(4+3+2)

5. **a**. Assume that MCS has a total of 500 students. It is known that:

i. 300 students play cricket
ii. 375 students live in hostels
iii. 200 students both play cricket and live in hostels

	Cricket	NO Cricket	Sum
Hostel Living	200	175	375
Out Living	100	25	125
Sum	300	200	500

Complete the above table and find the support and confidence of the following association rules.

i. Cricket => Hostel Living
Support: 40%, confidence: 66.7%

ii. Cricket => Out Living
Support: 20%, confidence: 33.3%

b. Given the following transactions:

t1: Beef, Chicken, Milk
t2: Beef, Cheese
t3: Cheese, Boots
t4: Beef, Chicken, Cheese
t5: Beef, Chicken, Clothes, Cheese, Milk
t6: Chicken, Clothes, Milk
t7: Chicken, Milk, Clothes

State three association rules which satisfy *minsup* = 30% and *minconf* = 80%.

Clothes → Milk, Chicken
Clothes, Chicken → Milk
Milk→Chicken etc.

c. What is the difference between data mining and data warehousing?

Data mining consists of finding interesting trends or patterns in large datasets, in order to guide decisions about future activities. Data warehousing is the process of transforming data into information and making it available to users in a timely enough manner to make a difference. A data warehouse provides an enterprise with the *memory* while data mining provides it *intelligence*.

++++++++++++++ Bon Courage ++++++++++++