Sample solutions to Midterm #2 from January-April 2006

Question 1

Some good examples are:

- duplicate elimination
- join operations
- bulk loading of a B+ tree
- creating an index
- sorting rid's during intersection of where clause conjuncts

Question 2

000 → 0, 16, 8
001 → 1
10 → [Next] 2, 6, 10 → 14
11 → 3, 7
100 → 4, 12, 28
101 → 5

Question 3

00 → 4, 16
01 → 1, 17
10 → 18, 26
11 → 15, 23, 3
And, all nubs have 2's in them.

Question 4(a)

Number of Pages for...
i) HockeyPlayer = floor(4096 bytes/page / 400 bytes/page) = 10 tuples/page => ceil(1200 tuples / 10 tuples/page) = 120 pages
ii) HockeyFan = floor(4096/200) = 20 tuples/page => ceil(3,000,000/20) = 150,000 pages

SMJ:
#I/Os = 120*2*2 + 150,000*2*3 + 120 + 150,000 = 1,050,600

Question 4(b)

30+10 = 40 bytes per data entry

floor(4096 bytes/leaf page / 40 bytes/data entry) = 102 data entries / page

Therefore: ceil(3,000,000 data entries / 102 data entries/page) = 29,412 leaf pages

ceil(29,412 / (102+1) pointers) = 286 pages at level 2

ceil(286 / 103) = 3 pages at level 1
ceil(3/103) = 1 page for the root

Question 4(c)

From part (a), HockeyPlayer has 120 pages. Scan this relation. Write only the Canucks tuples, that is, 120 pages * 1/30 reduction factor = 4 pages ... for the Temp table.
(The 1/30 comes from the fact that the HockeyTeam relation has 30 tuples.)
Therefore, so far, we've done 120 + 4 = 124 I/Os.

INL uses the clustering index; the hash index won't be of any value to us for this problem.

There are 1200 players and 3,000,000 fans; so this works out to 3,000,000/1200 = 2500 fans per player. We will use the clustering index to get back a whole bunch of HockeyFan records. Recall from Question 4(b), that there are 29,412 leaf pages. We need to read the leaf pages in order to determine what their row IDs are. Once we know this, we get the data pages from disk (which store the relevant tuples).

#I/Os for doing the INL join = 4 pages (to read in the Temp pages) + (4 pages * 10 tuples/page) * ((2.5 – 1) probe I/Os + ceil(1/1200 * Leaf Pages) + (1/1200 * Data Pages)). The –1 is to avoid double-counting the first leaf page—this calculation is similar to some of the examples in class.
= 4 + 40*(1.5 + ceil(29412/1200) + ceil(150,000/1200))
= 4 + 40*(1.5 + 25 + 125)
= 6064 I/Os

Grand total= 124 I/Os + 6064 I/Os = 6188 I/Os.