1. (a) There will be two subsequent passes ($\lceil \log_{200} 40000 \rceil = 2$). Each input pass will take $8,800,000 \cdot (15) = 132,000,000 \mu s = 132,000$ seconds, while an output pass uses $\frac{8,800,000}{20} (14 + 20) = 14,960,000 \mu s = 14,960$ seconds. The total time, in seconds, is

$$2 \cdot (132,000 + 14,960),$$

or about 81.6 hours.

(b) There will be five subsequent passes ($\lceil \log_{10} 40,000 \rceil = 5$). Like the output pass above, both the input and output pass will take $\frac{8,800,000}{20} (14 + 20) = 14,960,000 \mu s = 14,960$ seconds. The total time is

$$5 \cdot 2 \cdot 14,960 = 149,600$$

seconds, or about 41.5 hours.

2. (a) A nested loop join with 102 buffer pages, using the smaller relation department for the outer loop, will result in

$$1000 + \frac{1000}{100} \cdot 6000 = 61,000$$

disk I/Os.

(b) Sorting both the employee and department table, using 101 buffer pages, can be done in two passes (including the initial pass). Thus, employee requires $2 \cdot 2 \cdot 6000 = 24,000$ disk I/Os and department needs $2 \cdot 2 \cdot 1000 = 4000$ disk I/Os to sort. The total used by a sort/merge join algorithm is going to be $24000 + 4000 + 6000 + 1000 = 35,000$ disk I/Os.

3. (a) RA: $\pi_{dname} \sigma_{salary > 40000} (employee \bowtie_{dno = dnumber} department)$

TRC: \{ $P | \exists E \in employee \exists D \in department (P.dname = D.dname \land E.dno = D.dnumber \land E.salary > 40000)$ \}

(b) RA: $\pi_{dependent.name, dependent.bdate} \sigma_{dlocation = "Sugarland"} (dependent \bowtie_{ssn = ssn} employee \bowtie_{dno = dnumber} dept_locations)$

DRC: \{ $< N, B > | < s, N, B, \ldots > \in dependent \land < \ldots, s, \ldots, \ldots, d > \in employee \land < d, 'Sugarland' > \in dept_locations$ \}

4. (a) The analysis phase begins at the checkpoint. The log is scanned forward and the DPT and Xact table are modified. At the end of this phase, the DPT is

$$\{(p1, 30), (p2, 80), (p3, 40), (p4, 70), (p5, 100), (p6, 140)\}$$

and the Xact table is

$$\{(t1, 130, aborting), (t3, 70, active), (t4, 140, active)\}$$

(b) The redo phase begins at LSN 30, the smallest LSN in the DPT.
(c) 150 CLR t4 undo LSN 140
160 CLR t4 undo LSN 80
165 end t4
170 CLR t3 undo LSN 70
180 CLR t3 undo LSN 60
190 CLR t3 undo LSN 40
195 end t3
200 CLR t1 undo LSN 20
210 CLR t1 undo LSN 10
215 end t1