

Assignment 6

CIS451/551, Fall 2006

due 5:00pm Monday, December 4. If you submit it, the lowest homework among 6 assignments will be thrown out.

1. Create a schema in XML Schema corresponding to the following DTD:

```
<!DOCTYPE bib [  
  <!ELEMENT book (title, author+, publisher, keyword+)>  
  <!ELEMENT publisher (pub-name, pub-branch) >  
  <!ELEMENT title ( #PCDATA )>  
  <!ELEMENT author ( #PCDATA )>  
  <!ELEMENT keyword ( #PCDATA )>  
  <!ELEMENT pub-name( #PCDATA )>  
  <!ELEMENT pub-branch( #PCDATA )>  
]>
```

2. Given relations $r(A, B, C)$ and $s(C, D, E)$, which have the following properties: r has 40,000 tuples, s has 30,000 tuples, 50 tuples of r fit on one block, and 15 tuples of s fit on one block. Estimate the number of block transfers and seeks required, using each of the following join strategies for $r \bowtie s$:

- Nested-loop join
- Block nested-loop join
- Indexed nested-loop join (suppose there is a primary B+ tree index with height 4 on the join attribute (C).)
- Merge join
- Hash join

3. Suppose that a B+-tree index on $(branch-name, branch-city)$ is available on relation $branch$. What would be the best way to handle the following selection?

$\sigma_{(branch-city < "Brooklyn") \wedge (assets < 5000) \wedge (branch-name = "Downtown")}(branch)$

4. Exercise 15.6. (15.10 in old version). Also, suggest to either add a new edge (i.e., $\langle T1, T3 \rangle$) or remove an edge that would change the story (reverse your conclusion of conflict serializable or not).

5. Consider the following schedule:

Step	T ₀	T ₁	T ₂
1	Write(A)		
2		Write(A)	
3			Write(A)
4	Write(B)		
5		Write(B)	

Is this schedule allowed in timestamp-ordering protocol? Explain why?