Answer questions 1-3 below based on the following three relational schemas:

employee (*ENO*, *NAME*, *ADDRESS*) *works* (*ENO*, *PNO*, *HOUR*) *project* (*PNO*, *LEADER*)

The relation **employee** has three attributes: ENO for the employee identifier, NAME for the name, and ADDRESS for the address. The relation **project** has the attribute PNO for the project number and LEADER is the name of the person who leads the project. The relation **works** describes the project (PNO) which an employee (ENO) works for, and the working hours for that project (HOUR).

1. (5%) Give the relational algebra expression for the following question:
   Find the names and addresses of those employees whose leader is "John".

2. (20%) Write SQL expressions for the following questions, respectively.
   (a) Find the employees (ENO) who work for the project with PNO = "P1" or PNO = "P2".
   (b) Find the employees (ENO) whose average working hours are more than the average working hour of the project with PNO = "P1".
   (c) Create a view called 'LC', which outputs the names of the leaders and the number of employees led by each leader.
   (d) Delete those projects which the employee named "Mary" works for. (You only need to delete the data in the *project* relation.)

3. (10%) Consider the following relational-algebra expression:
   $$\Pi_{\text{NAME, HOUR}} (\sigma_{\text{ENO} = \text{"E1"}} (\text{employee} \bowtie \text{works}))$$
   (a) Represent the relational-algebra expression as an **expression tree**.
   (b) Give an **evaluation plan** based on your answer of (a).

4. (15%) Draw an ER diagram to design the database for an online store, which needs to satisfy all the following descriptions and constraints:
   - We want to record the information about each customer, such as his email, name, address, etc.
   - This store sells books. For each item, we need to know its ISBN, title, authors and price.
   - We want to record the order of each customer. Note that each order belongs to only one customer.
   - Each order can contain many different books.

   For each entity set, design appropriate attributes and primary keys for them by yourselves. You can also assign attributes to relationship sets.

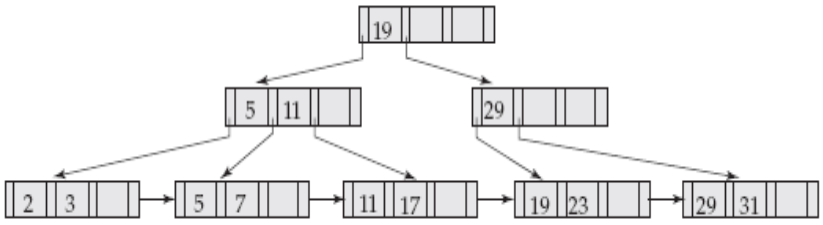

Figure 1

5. (10%) Consider the B^+-tree in Figure 1, where the number of pointers that fit in one node is **four**.
   (a) Show the tree after "deleting 31".
   (b) Based on your answer of (a), show the tree after "inserting 12".
   (c) Based on your answer of (b), show the tree after "inserting 13".

6. (10%) Let relations $r1(A,B,C)$ and $r2(C,D,E)$ have the following properties: $r1$ has 20,000 tuples and occupies 800 blocks; $r2$ has 45,000 tuples and occupies 1500 blocks. Estimate **the number of block transfers** required, using each of the following two join strategies for $r1 \bowtie r2$, respectively. Let us assume M pages of memory, where $M < 800$ pages.
   (a) Nested-loop join
   (b) Merge join (assume that the two relations are already sorted on the join keys)

7. (10%) Suppose there are 3000 transactions in a clothes shop. Suppose 1500 transactions purchase *jeans* and 1000 transactions purchase *T-shirts*. Also suppose that half of the transactions that purchase *jeans* also purchase *T-shirts*. That is, 750 transactions purchase both *jeans* and *T-shirts*. For the following association rule:
   jeans => T-shirt
   (a) What is the confidence?
   (b) What is the support?

8. (10%) What are the differences between the QuadTree index structure and the R-tree index structure? List at least two.

9. (10%) What are the differences between an XML file and a relational schema? List at least two.