

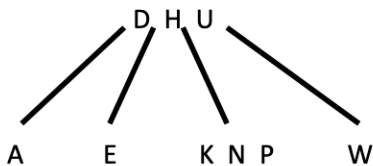
CSE3241 Activity – Indexes

Names	Date

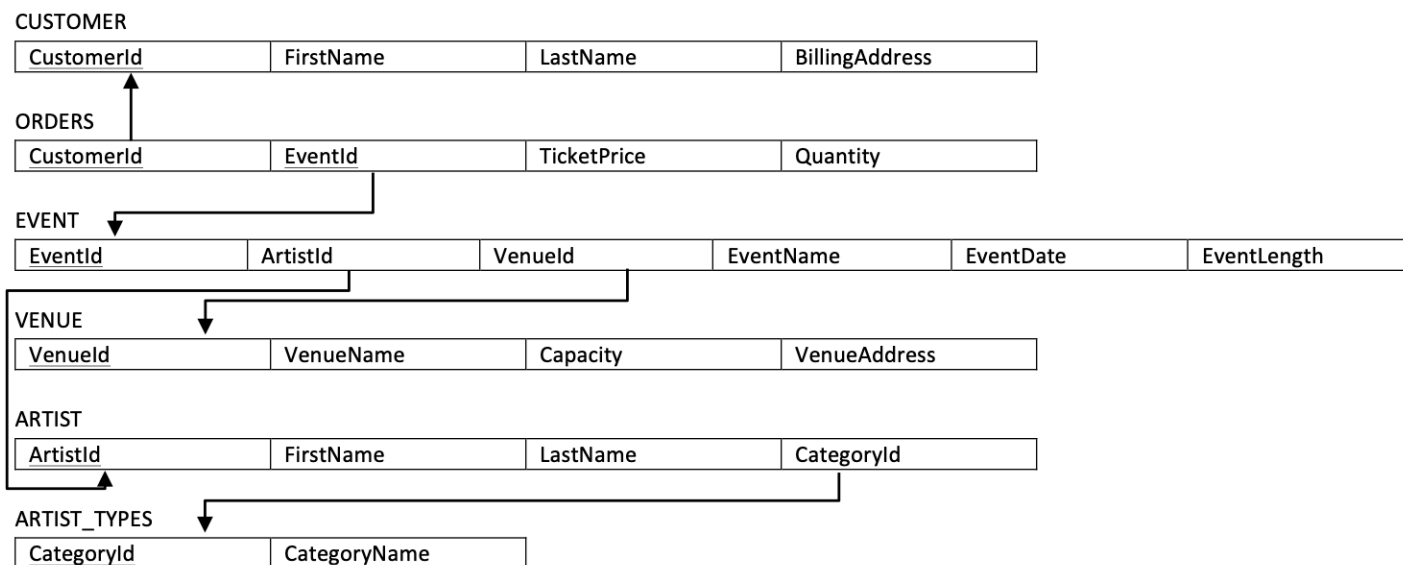
1. For each SQL query below, recommended an index or set of indexes, and explain your decision

<p>SELECT * FROM EMP WHERE Name="John Doe"</p>	<p>Will benefit from</p> <ul style="list-style-type: none"> (a) Tree-index on Name (b) Hash-index on Name (c) Tree-Index on Salary (d) Hash-Index on Salary (e) Both (a) and (b) (f) Both (c) and (d)
<p>Explain</p>	
<p>SELECT * FROM EMP WHERE LName="Smith" AND Salary < 10000</p>	<p>Will benefit the most from</p> <ul style="list-style-type: none"> (a) Tree-index on Name, Salary (b) Hash-index on Name, Salary (c) Tree-Index on Salary (d) Hash Index on Salary
<p>Explain</p>	

2. Consider the following B-tree of order four (i.e., each node has a maximum of four children), and minimum branching factor of two (i.e., a minimum of two children).



Show the B-tree that results from inserting J and then M into the above B-tree.



3. Given the database above, assume that the CUSTOMER table has a million records, the ORDERS table has 10 million records, the EVENT table has 250,000 records, the ARTIST table has 25,000 records and the VENUE table has 10,000 records as you answer the following (12pts total, 2pts each):
 - a. For the following query, provide the pseudocode for the following SQL


```

          SELECT C.CustomerId, C.FirstName, C.LastName, O.TicketPrice, E.EventName
          FROM CUSTOMER as C, ORDERS as O, EVENT as E
          WHERE C.CustomerId = O.CustomerId AND E.EventId=O.EventId;
          
```
 - b. Determine the number of operations it would take to perform this query if there are no indexes at all on the tables above:
 - c. For the query given in part (a), propose a set of indexes that would improve the speed of this query. List these indexes in TABLENAME.(COL1, COL2, ... COLN) format where (COL1,COL2,... COLN) is a list of columns the table is indexed on. Assume worst-case performance for your indexes (hash indexes have a constant time C lookup time, tree indexes have a lookup time $O(\log(n))$). You should also assume that orders are evenly distributed among customers and events (so each customer has on average 10 order records with him/her, and each event has on average 40 order records associated with it.)