

CS 3335
Exam III
Nov. 24, 2008
100 points

Name: _____

1. Using a B+-Tree with 3 values per node, insert the following data points into the tree. (20)
Show the resulting tree at every split. Data: {100, 50, 150, 75, 25, 67, 200, 250, 10, 0}.

2. What four things must be defined in a configuration file in order to map and store objects with Hibernate? (8)

3. An IR system reports that it achieves 100% recall at 72% precision. What does this mean? (4)

4. What are the four isolation levels defined in the SQL standard? (8)

5. What is the criteria for the best split possible for a decision tree classifier? Give as much detail as possible! (4)

6. Rewrite the following SQL query into the most efficient one possible. Use the video rental schema from class, repeated on the last page for your convenience. (12)

```
(Select c.customerid, t.titleid
 From Customers c, Titles t, Tapes v, Orders o
 Where c.customerid = o.customerid and o.tapeid = v.tapeid and
       v.titleid = t.titleid and rating = 5)
UNION
(Select c.customerid, t.titleid
 From Customers c, Titles t, Tapes v, Orders o
 Where c.customerid = o.customerid and o.tapeid = v.tapeid and
       v.titleid = t.titleid and rating = 4)
```

7. Define the 2PL protocol, including the two rules which must be followed for serializability to be enforced and the lock compatibility matrix. (10)

8. Provide the k-means algorithm for clustering. Be sure to include the inputs!

(10)

9. Define each part of the ACID acronym. (8)

10. Define support and confidence for association rules. (8)

11. List the groupings formed by the SQL query – (8)

```
Select region, state, storeid, count(sales)
From dataStore
Group by cube(region, state, storeid);
```