1. What is the only data storage structure allowed in BigTable? (10)

2. (a) Explain hubs and authorities in web search. (5)
   (b) Explain tf-idf. (5)

3. (a) If a dataset has 1 large 5-item set, what is the minimum number of large 4-item sets it must contain? Why? (5)
   (b) If a dataset has 1 large 5-item set, what is the maximum number of large 6-item sets it must contain? Why? (5)
4. Given the Java class and Hibernate mapping file provided, provide the table diagram that would be created by Hibernate.

```java
public class Person {
    private Long id;
    private int age;
    private String name;
    private Date birthday;

    public Person() {}

    ...
}

<hibernate-mapping>
    <class name="Person" table="PERSON">
        <id name="id" column="PERSON_ID">
            <generator class="native"/>
        </id>
        <property name="age"/>
        <property name="name" column="PERSON_NAME"/>
        <property name="birthday" type="timestamp"/>
    </class>
</hibernate-mapping>
```
5. Consider the following table. Find the 2 nearest neighbors to the first row (other than the row itself).

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
</tr>
</tbody>
</table>

6. Provide four example applications, one for each of association, classification, clustering and collaborative filtering. Defend your choices.
7. For each of the schedules below, indicate whether or not the schedule is recoverable, cascadeless, strict, serializable and serial. Indicate all that are true by placing an X in the blank.

<table>
<thead>
<tr>
<th></th>
<th>Recoverable</th>
<th>Cascadeless</th>
<th>Strict</th>
<th>Serializable</th>
<th>Serial</th>
</tr>
</thead>
<tbody>
<tr>
<td>t1</td>
<td>R(X)</td>
<td>W(X)</td>
<td>R(Y)</td>
<td>W(Y)</td>
<td>abort</td>
</tr>
<tr>
<td>t2</td>
<td></td>
<td></td>
<td></td>
<td>R(X) W(X) R(Y) W(Y)</td>
<td>commit</td>
</tr>
</tbody>
</table>

(b) __________ Recoverable __________ Cascadeless __________ Strict __________ Serializable __________ Serial  

<table>
<thead>
<tr>
<th></th>
<th>Recoverable</th>
<th>Cascadeless</th>
<th>Strict</th>
<th>Serializable</th>
<th>Serial</th>
</tr>
</thead>
<tbody>
<tr>
<td>t1</td>
<td>R(X)</td>
<td>W(X)</td>
<td>R(Y)</td>
<td>W(Y)</td>
<td>abort</td>
</tr>
<tr>
<td>t2</td>
<td>R(X)</td>
<td>W(X)</td>
<td>R(Y)</td>
<td>W(Y)</td>
<td>commit</td>
</tr>
</tbody>
</table>

8. Define all parts of the ACID acronym.  

(10)
9. Assume a unique $B^+$-tree such that a node has three pointers, the tree has three levels, and the values in the root node are 50 and 100.

(a) Provide a possible (correct) full subtree for the last pointer in the root node.

(b) Indicate which of your nodes would be accessed to answer the SQL query

```
SELECT *
FROM T
WHERE A > 150
```
10. Provide the nested-block loop join algorithm.