

Spring, 2009 CIS 455 / 555

Internet and Web Systems

Final Examination

120 minutes, 7 questions, 120 pts

Try to allocate no more than 1 minute per point, and be concise with your answers.

Your name: _____

1. **Suppose, 3 years from now, we have a new 4G cell phone and wish to use it as a Twitter-like server: we publish our GPS location and an occasional “tweet” like message. What role should replication play in keeping this service available, and which type of consistency should be used on the replica(s)? (20pts)**

- 2. Suppose we want to provide Web Services over a distributed hash table like Pastry. Explain how this can be done, including messaging at both the RPC layer and the interface definition layer. (15pts)**

3. (a) Which of the ACID properties does *two-phase commit* address, and how? (3 pts)

(b) Which of the ACID properties does *two-phase locking* address, and how? (3 pts)

(c) What is the role of logging in ensuring ACID properties? (4 pts)

- 4. Outline the steps for taking a keyword query of the form “k1 k2” and producing ranked results in a partitioned architecture like that of Google or your crawler. In particular, note that k1 and k2 may be indexed at different nodes. (20pts.)**

5. Write the map and reduce functions that, given a distributed log of document requests (each request of the form “GET <doc URL>”), count the number of references to “index.html”. (15pts.)

- 6. Outline the roles of the parties (nodes) and of cryptography in the SSL protocol. What does each node do, and how does it employ cryptography to ensure authentication? (20pts.)**

7. Given an XML version of an inverted index over English documents, `inverted.xml`, of the form:

```
<index>
  <word><text>...</text><id>...</id></word>*
  <entry><wordId> ... </wordId>
    <foundInDocId>...</foundInDocId>*
  </entry>*
</index>
```

and an XML version of the term frequencies:

```
<docFreq>
  <doc><id>...</id>
    <word><id>...</id><count>...</count></word>*
  </doc>*
</docFreq>
```

Write an XQuery that returns, for the word “exam,” all document occurrences and term frequencies. Your output should have the following format:

```
<occurrences>
<document>
  <docId>...<docId>
  <count>...</count>
</document>*
</occurrences>
```

(20pts)