(1) Google File System

(Continued)

A
/chunks/1
/chunks/2

B
/chunks/1
/chunks/3

C
/chunks/1
/chunks/2
/chunks/3

D
/chunks/2

E
/chunks/3

offline

(a) The master is recovering. Fill in the chunk map after reports from above workers:

<table>
<thead>
<tr>
<th>Chunk Number</th>
<th>Workers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A B C</td>
</tr>
<tr>
<td>2</td>
<td>A C</td>
</tr>
<tr>
<td>3</td>
<td>B C E</td>
</tr>
</tbody>
</table>

(b) what replication operations is the master likely to direct workers to perform?

Copy 2, from C to E

(c) D reboots; what replication operations is the master likely to direct workers to perform?

Delete 2 from C

(2) MapReduce

Describe the purpose of the following MapReduce pseudocode:

function map(Long reqID, Request req):
    if (req.date == "2014-16-10")
        emit(req.type, req.latency)

function reduce(int key, Iterator values):
    sum = 0, count = 0
    for each val in values:
        sum += val, count += 1
    emit (key, sum/value, count)

Get avg. latency of each type of request on specified day.
(3) PageRank

a) Consider the following web graph, where circles are pages and arrows represent hyperlinks. Assuming all PageRank's add to 6, label the PageRank for each page. Assume a simpler version of PageRank that does not model occasional jumps to a random page without a link.

(4) Inverted Index

<table>
<thead>
<tr>
<th>Doc 0</th>
<th>Doc 1</th>
<th>Doc 2</th>
<th>Doc 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>lorem ipsum dolor sit amet</td>
<td>ipsum amet</td>
<td>lorem dolor sed</td>
<td>ipsum dolor lorem</td>
</tr>
</tbody>
</table>

(a) consider the above documents, and generate posting lists for an inverted index:

<table>
<thead>
<tr>
<th>lorem</th>
<th>ipsum</th>
<th>dolor</th>
<th>sit</th>
<th>amet</th>
<th>sed</th>
</tr>
</thead>
</table>

(b) list the unsorted results of each query and give the number of posting lists that must be read:

(i) query: “lorem”

(ii) query: “lorem AND ipsum”

(iii) query: “lorem AND NOT ipsum”