CSE427s - 9 Lab9 (merged)

100% (8/8)

✓ **1. PageRank computes the importance of a particular webpage (say *page A*) based on...**

- (A) the number of pages *page A* links to
- (B) the number of pages that link to *page A*
- (C) the importance of the pages *page A* links to
- (D) the importance of the pages that link to *page A*
- (E) the content of the webpage

✓ **2. In the Spark implementation of PageRank these operations require (re)partitioning (create a new stage):**

- (A) loading the data in (pageID, pageID) format
- (B) creating the links RDD
- (C) joining ranks and links
- (D) propragating the ranks
- (E) aggregating the ranks

✓ **3. This data changes during the PageRank *iterations***

- (A) web graph RDD in (pageID,pageID) format
- (B) links RDD
- (C) ranks RDD
- (D) none of the above

✓ **4. Our Spark application (PageRank.pyspark / PageRank.scalaspark) performs *teleportation.***

- (A) True
- (B) False

✓ **5. How do we make the link RDD called *links* persistent? Provide the spark command using the RDD identifier *links.***

links.persist()

✓ 6. Using default settings persisted data is stored...

(A) in HDFS

(B) in memory

(C) on disk

✓ 7. After executing the PageRank computation on the *pagelinks* data. Which page is the most important?

(A) page1

(B) page2

(C) page3

(D) page4

✓ 8. After executing the PageRank computation on the *pagelinks* data. Which page is the least important?

(A) page1

(B) page2

(C) page3

(D) page4