CSE427s - 5 Lab5

80% (4/5)

✗ 1. What is the output of the first MapReduce Job for the **given data**? Compute the actual keys and values.

Separate multiple key-value pairs by ; E.g. (key1, value1); (key3, value5); etc.

This question will be graded by the TA.

(item1,1);(item2,4);(item3,3);(item4,2);(item5,1);(item6,1);

user-id	item-id
user1	item1
user1	item2
user1	item3
user2	item3
user2	item2
user3	item2
user3	item4
user4	item5
user4	item6
user5	item3
user5	item4
user5	item2

⊘ 2. Now let's assume that our second job runs **two** Mappers.
The first Mapper gets the records for item1, item2, and item6. What is the output of this Mapper for a **top-2 list**?

Separate multiple key-value pairs by ;

This question will be graded by the TA.

(_,(item1,1));(_,(item2,4))

⊘ 3. The second Mapper gets the records for item3, item4, and item5. What is the output of this Mapper for a **top-2 list**?

Separate key-value pairs by ;

This question will be graded by the TA.

(_,(item3,3));(_,(item4,2))

✓ 4. What do the Mappers of job 2 compute?

- (A) Item occurrences
- (B) Item counts
- (C) local top-N lists
- (D) aggregated (global) top-N list
- (E) none of the above

5. Now, let's do the Reduce Tasks. What is your final top-2 list?

Separate key-value pairs by ;

This question will be graded by the TA.

(item2,4);(item3,3)

6. What does the single Reducer of job 2 compute?

- A Item occurrences
- B Item counts
- C local top-N lists
- D aggregated (global) top-N list
- E none of the above

7. Can the Rducer be used as a Combiner? Identify the **true statement** below. Note that Combiners combine the Mapper output of **all** Map Tasks that run on the **same** compute node.

- A The Reducer of job 1 (aggregate item counts) can be used as a Combiner.
- B The Reducer of job 2 (producing top-N-list based on counts) can be used as a Combiner.
- C The Reducers of both jobs can be used as Combiners.
- D The Reducers of none of the jobs can be used as Combiners.

8. Using a **Combiner**...

- A will **speed up** the execution of your entire MapReduce job.
- B will help to resolve **memory issues** of your MapReduce job.
- C will reduce the amount of data being **transfered across the network**