CSE427SLAB1

CSE427s - 3 Lab3 80% (8/10)

✓ 1. When reading data from HDFS, the **number of Map Tasks** is usually driven by the *total size of the input data* and the *HDFS block size*.

- (A) True
- (B) False

✓ 2. The **number of Reduce Tasks** is set by the developper.

- (A) True
- (B) False

✓ 3. The output of every MapReduce job is a single file with key-value pairs sorted by keys in total order.

- (A) True
- (B) False

✓ 4. The output of the Reduce Task is typically stored

- (A) on the compute nodes that executed the Reducers
- (B) in HDFS
- (C) on the client
- (D) in RDMS

✓ 5. The ResourceManager is responsible for executing the **tasks** of a MapReduce job.

- (A) True
- (B) False

✓ 6. The *driver* of a MapReduce program is executed

- (A) on the ResourceManager (MapReduce master node)
- (B) on the compute node that executes the ApplicationMaster
- (C) on the NameNode (HDFS master node)
- (D) on the client

✓ 7. The ResourceManager is the ultimate authority that arbitrates resources among **all applications in the Hadoop cluster**.

- (A) True
- (B) False

✗ 8. How does the Hadoop framework deal with **slow Mappers**?

– if a Map task appears to be running slow a new instance will be started on another machine-> new task attempt for the same task
– the result of the first Map task to be finished is used
– once one Map task is finished all other attempts will be killed

✓ 9. Not setting the number of reducers in the job configuration will turn the job into a **map-only job**.

(A) True

(B) False

✗ 10. What is the benefit of **testing** your MapReduce program **locally** on the client (without using the real cluster or a pseudo cluster)? This question will be graded by the TA.

test incremental changes to code quickly