CSE427SLAB1

CSE427s - 4 Lab4

80% (8/10)

✗ 1. **Job** is the primary interface to describe a MapReduce job to the Hadoop framework for execution.

- (A) True
- (B) False

✓ 2. The name of a MapReduce job is defined by the name of the driver class.

- (A) True
- (B) False

✓ 3. Why should you always implement ToolRunner in your MapReduce programs?

- (A) It speeds up the actual job execution.
- (B) It allows you to run your job locally using the command line.
- (C) It allows parameters to be passed via the command line.
- (D) It is required to run your job locally via Eclipse.
- (E) It saves time when testing different configuration settings (such as number of reducers, etc).

✗ 4. What are the various ways to configure MapReduce jobs? Name **two** possible ways. (This answer will be graded by the TA.)

command line
configuration file

✓ 5. Which ways of setting job configurations overwrite others and when are default values considered?
Identify the correct order according to the **priority** (weakest to highest).

- (A) command line - driver - Hadoop configuration files - defaults
- (B) defaults - command line - driver - Hadoop configuration files
- (C) driver - command line - Hadoop configuration files - defaults
- (D) defaults - Hadoop configuration files - command line - driver

✓ 6. Files provided via DistributedCache are stored in HDFS in the same way as other data (i.e., when executing hadoop fs -put).

- (A) True
- (B) False

7. **Which data can be provided via DistributedCache?**

- [x] A look-up tables
- [ ] B job statistics (that will be updated by the currently running job)
- [ ] C detailed transaction or customer information for every data record
- [x] D java libraries that are used in Mappers or Reducers
- [ ] E entire log file dumps

8. **The output of each Reduce Task is re-sorted by the Reducer output key.**

- [ ] A True
- [x] B False

9. **In the output of a MR job the values are sorted.**

- [ ] A True
- [x] B False

10. **In a secondary sort implementation, which key is used to determine which Reducer an entry goes to?**

- [x] A primary key
- [ ] B secondary key
- [ ] C composite key

11. **Why do we need to implement the sort comparator for the secondary sort example in this lab? (This question will not be graded, so you may answer it for yourself after finishing the lab.)**

.

12. **Why do we need to implement the group comparator for a secondary sort such as the one discussed in this lab? (This question will not be graded, so you may answer it for yourself after finishing the lab.)**

.