

# Quiz 03: HDFS & MapReduce Revision

52100674@student.tdtu.edu.vn [Switch account](#)



Draft saved

\* Indicates required question

Email \*



Record 52100674@student.tdtu.edu.vn as the email to be included with my response

Student ID \*

52100674

Full Name \*

Trần Thị Vẹn

Group \*

6



Q1: Which command is used to list a directory content in HDFS? \*

- ☐ bin/hdfs dfs -sl
- ☐ bin/hdfs dfs -sl <dir>
- ☐ bin/hdfs dfs -ls /
- ☒ bin/hdfs dfs -ls <dir>

Q2: Which command is used to copy a directory from HDFS to the local machine? \*

- ☐ bin/hdfs dfs -put <path> <path>
- ☒ bin/hdfs dfs -get <path> <path>
- ☐ bin/hdfs dfs -copyFromLocal <path> <path>
- ☐ bin/hdfs dfs -cp <path> <path>

Q3: Which is **NOT** a major stage in the MapReduce model? \*

- ☐ Map
- ☐ Group by key
- ☐ Reduce
- ☒ Combine by key



Q4: The output of the Map function is a set of \*

- ☐ 0
- ☐ 1
- ☐ key-value pairs
- ☒ All of them

Q5: Output items of the Map stage are sorted in which stage? \*

- ☐ Map
- ☒ Group by key
- ☐ Reduce
- ☐ None of them

Q6: The MapReduce model has a bottleneck in which stage? \*

- ☐ Map
- ☒ Group by key
- ☐ Reduce
- ☐ None of them



Q7: Given a map function and a reduce one as below. If the input document is "a b \*  
c a a", then what is the output set after the Group by key stage?

**map(key, value):**

```
# key: document name; value: text of the document
for each word w in value:
    emit(w, 1)
```

**reduce(key, values):**

```
# key: a word; value: an iterator over counts
result = 0
for each count v in values:
    result += v
emit(key, result)
```

- ☐ (a, 1) (b, 1) (c, 1) (a, 1) (a, 1)
- ☐ (a, 3) (b, 1) (c, 1)
- ☒ (a, [1, 1, 1]) (b, 1) (c, 1)
- ☐ (b, 1) (c, 1) (a, [1, 1, 1])

Q8\*: In Hadoop, which class is used to reduce the volume of data transferring between Map and Reduce? \*

- ☐ Mapper
- ☐ Grouper
- ☐ Reducer
- ☒ Combiner

Submit

Clear form



Never submit passwords through Google Forms.

# Google Forms



