NOTE: These are practice questions only. The actual exam questions are different. For optimal preparation, please refer to your lecture notes, assignments, and readings.

Question 1. Assume you have an undirected graph describing friendship relations. The graph is stored on an HDFS file called **follower.txt**. To understand the content, a tiny sample of the file is as follows:

```
1   2,3,4
2   1,5,6
3   1,5
4   1
5   2,3
6   2
```

Here, user 1 is friends with users 2, 3, and 4.

Write a MapReduce program that calculates the number of common (shared) friends for all pairs of friends. The output for the previous example would look like this (friends pair, count):

```
(1,5)  2
(1,6)  1
(2,3)  2
(2,4)  1
(3,4)  1
(5,6)  1
```

Question 2. Hadoop

a. Explain the execution flow of MapReduce. Use Question 1 as a guiding example.
b. What is the main shortcoming of Hadoop 1.0 that led to the development of Hadoop 2.0?
c. What is the major architectural change introduced to solve that? describe in how it addressed that shortcoming.

Question 3. DSMS

a. Explain the specific data processing issues which are specific to data processing on streams.
b. Explain why are DSMSs designed to only do one pass on the data in order to evaluate a query.

Question 4. On Kafka. Give a brief overview of the problem that Kafka solves, and describe the idea of its core functionality

Question 5. On Raft

a. Briefly explain the following concepts in the context of the Raft algorithm:

   i. Leader Election
   ii. Log Replication
   iii. What role do the AppendEntries RPCs play in the Raft algorithm? How do they contribute to log consistency?

b. Consider the following scenario: Assume a Raft cluster with 5 nodes: S1-5. If Node S1 is the leader and it crashes, describe step by step what happens in the cluster. How is a new leader elected? What happens if Node A comes back online?

c. Consider the following state of our 5 nodes cluster (see figure). If S1 crashes, and S4 times out first, and calls for an election, can it become a leader? why?