QUIZ:

1. The files a user receives with the <glite-wms-job-output> command when the job is 'Done' should be specified in:
   a. the InputSandbox attribute
   b. the OuputSandbox attribute
   c. the working directory of the Grid Worker Node
   d. the working directory of the UI
Correct Answer is b.

2. Which of the following about MapReduce is true?
   a. MapReduce is designed for high performance computing.
   b. MapReduce is more suitable for batch processing than real-time processing.
   c. MapReduce is based on publications by IBM.
   Correct Answer is b.

3. Indicate which of the following sentences is false for Multicore Scheduling on the Grid:
   a. the application would get roughly equivalent speedup as the number of assigned cores
   b. it can be used for applications that require more memory than allows a single core job
   c. the application can get a better I/O performance
   d. it ensures that the WMS reserves an amount of cores on the same node, even if this is not specified in the JDL file
Correct Answer is b.

4. Collections and Parametric jobs:
   a. can reduce the overhead for job submission and thereby the load on WMS
   b. ensure that child jobs will land on the same Worker Node
   c. generate the same job ID both for the parent and corresponding children jobs
   d. can be submitted from any machine which runs a linux distribution
Correct Answer is a.

5. Which of the following statements is correct about HDFS?
   a. the Hadoop Filesystem is no different than traditional shared file systems
   b. redundancy on HDFS is handled by storing the same data on multiple machines.
   c. when reading or writing, all data passes through the NameNode
Correct answer is b.

6. The chances of your job to run successfully on a remote Worker Node are larger when:
   a. your program is first tested and/or compiled by logging in one of the Worker Nodes
   b. your program is first tested and/or compiled on a UI
   c. your program is first tested and/or compiled on the Grid Gate
   d. your program does not use the Sandbox for data transfer
Correct Answer is b.

7. Which of the following statements about Hadoop is correct?
    a. the Grid infrastructure can be replaced by a Hadoop cluster.
    b. Hadoop runs only on the cloud.
    c. The Hadoop framework is open source and can be used free of charge
    d. the only way to use a Hadoop cluster is by using the MapReduce programming model directly
Correct answer is c.

8. To execute multicore jobs on the Grid:
    a. the user creates a Parametric job for each thread
    b. the user submits a different jdl for each thread
    c. the user issues 'CPUNumber' and 'SmpGranularity' attributes in the jdl file
    d. the user should set the correct environment only on the UI
Correct Answer is c.

9. Using the lcg-infosites command (available in your VM), find out the batch queue that does *not* belong to 'gb-ce-wur.els.sara.nl' grid gate:
    a. cream-pbs-medium
    b. cream-pbs-long
    c. cream-pbs-express
    d. cream-pbs-tutor
Correct answer is d. In your VM terminal execute <lcg-infosites --vo tutor ce | grep gb-ce-wur.els.sara>. The 'cream-pbs-tutor' is not listed in the batch queues.

10. Which of the following statements about Big Data is false?
    a. All data sets larger than 1TB are Big Data by definition.
    b. Big Data is not just about the size, but also the nature of the data.
    c. Big Data can be defined in terms of the four V's: Volume, Variety, Velocity and Veracity
Correct answer is a.