1. If you have difficulties understanding *bash*, make sure to go over this refresher .
2. What version of Java is required for your version of Hadoop?
   - Java 1.8.0
3. What does the example job that you run throughout this tutorial is supposed to do?
   - Copies the unpacked conf directory to use as input and then finds and displays every match of the given regular expression.
4. What is the purpose of PATH environment variable?
   - Do Hadoop available available to any directory
5. Where the output of the test job stored?
   - In the directory called Output in a file called _SUCCESS
6. What is DFS replication factor?
   - It is the number of copies of a block that must be there in the cluster.
7. What is passphraseless ssh, and why do you need it in this lab?
   - Authentication can be automatically negotiated using a public and private key pair. To connect the main machine to secondary machines, because Hadoop treats the machines as a remote one.
8. How does one upload files to HDFS?
   - hdfs dfs -put etc/hadoop/*.xml input
9. Where does Hadoop store log files by default?
   - Home directory of the user
10. How does one verify that the job finished successfully?
    - A file called_SUCCESS will appear in the output directory.
11. What is stored in the file ~/.ssh/id_rsa?
    - The public key
12. What is stored in the file ~/.ssh/authorized_keys?
    - Copy the public key to the list of autohorized keys
13. How are these two files used for passphraseless ssh?
    - Allow to stablish a safe connection without required password.
14. What is the default user in Docker's ubuntu image?
    - By default docker containers run as the root user.
15. How does one specify an environment variable in Dockerfile?
    - In the docker file after the word ENV
16. What is the purpose of this command chmod +x run.sh?
    - Make the script executable.
17. What is the purpose of this command /etc/init.d/ssh start?
    - It is neccesary to run a script called /etc/init.d/ssh to start the OpenSSH server.
18. When creating Dockerfile, what is the difference between ENTRYPOINT and CMD?
    - CMD is an instruction that is best to use if you need a default command which users can easily override. ENTRYPOINT is preferred when you want to define a container with a specific executable.
19. How does one execute a bash command in a docker container?
    - docker exec -it <container> bash

20. Assume you created a docker container and executed several commands. These commands created new files in the container's file system. How can you access thee files after the container has stopped?

    - Having the ID of the stopped container, we can create a new Docker image. The resulting image will have the same state as the previously stopped container. At this point, we use docker run and overwrite the original entrypoint to get a way into the container.

21. What is the difference between docker run and docker container exec?

    - The difference between "docker run" and "docker exec" is that "docker exec" executes a command on a running container. On the other hand, "docker run" creates a temporary container, executes the command in it and stops the container when it is done.

22. Why we call it pseudo-distributed mode, and not simply distributed?

    - In Pseudo-distributed Mode we also use only a single node, but the main thing is that the cluster is simulated.