1. What is a namenode?
   - The NameNode is the centerpiece of an HDFS file system. It keeps the directory tree of all files in the file system, and tracks where across the cluster the file data is kept.
2. What is a datanode?
   - The DataNode is responsible for storing the actual data in HDFS.
3. What is replication factor?
   - The total number of replicas across the cluster.
4. Why did you need ssh-copy-id?
   - Command ssh-copy-id installs an SSH key on a server as an authorized key
5. How is Hadoop configuration different between namenode and datanodes?
   - hdfs namenode -format and hdfs --daemon start datanode
6. How do you check the content of HDFS file system?
   - hdfs dfs -cat <path>
7. How often do you need to run hdfs namenode -format?
   - Everytime you decide to format the file system
8. What is YARN? What are components of YARN?
   - It is the Cluster management component of Hadoop 2.0. YARN has three main components: ResourceManager: Allocates cluster resources using a Scheduler and ApplicationManager.
9. How do you specify the worker nodes for YARN?
   - That property is defined in core-site.xml and in yarn-site.xml
10. How do you list active YARN worker nodes?
    - You can run yarn nodes -list to check
11. Where does Hadoop store log files?
    - `$HADOOP_HOME/logs`
12. What is the purpose of the file /etc/hosts?
    - The /etc/hosts file contains the Internet Protocol (IP) host names and addresses for the local host and other hosts in the Internet network.
13. What is virtual memory and why Hadoop cares about it?
    - Virtual memory is a feature of an operating system that enables a computer to be able to compensate shortages of physical memory by transferring pages of data from random access memory to disk storage.
14. How do you check available storage in HDFS?

    - You can check the free space in an HDFS directory with a couple of commands. The -df command shows the configured capacity, available free space and used space of a file system in HDFS.