

# **Big Data Technologies**

## **Polytech Nice Sophia – 2024/2025**

### **Solution of Lab 2**

#### **Questions:**

1. You see the three HDFS processes: Namenode, DataNode, SecondaryNameNode
2. /opt/hadoop/etc/hadoop
3. /tmp/hadoop-root/dfs/name in file “hdfs-site.xml”. The “fsimage” files and the “edits” files are stored in this directory
4. /tmp/hadoop-root/dfs/data in file “hdfs-site.xml”. The data blocks are stored in the subdirectories of /tmp/hadoop-root/dfs/data
5. See property “dfs.replication” in file “hdfs-site.xml”. It takes on the value “1” since there is only one datanode in the virtual machine.
6. The “http-address” is given in “hdfs-site.xml” (property: dfs.namenode.http-address). The namenode activity is described on <http://0.0.0.0:9870/>
7. The information on the Datanodes are linked to the namenode webpage.
8. The property “fs.defaultFS” defines the entry point to the HDFS storage system. We can read/write data on HDFS via the URL (Uniform Resource Locator) hdfs://0.0.0.0:9000

Note: many properties by default are defined in the directory

\$HADOOP\_HOME/share/doc/hadoop/hadoop-project-dist/hadoop-hdfs/hdfs-default.xml

#### **Perform:**

1. hdfs dfs -mkdir /lab2
2. Just download the file by using wget  
wget <http://www.umich.edu/~umfandsf/other/ebooks/alice30.txt>
3. Assume the file "alice30.txt" is in your current directory:  
hdfs dfs -put alice30.txt /lab2/alice.txt
4. hdfs dfs -ls /lab2
5. To see the size: hdfs dfs -du /lab2/alice.txt
6. hdfs dfs -cat /lab2/alice.txt | head -n 25
7. hdfs dfs -cp /lab2/alice.txt /lab2/aliceHdfsCopy.txt
8. hdfs dfs -get /lab2/alice.txt aliceCopy.txt
9. hdfs fsck /
10. hdfs dfs -rm /lab2/alice.txt
11. hdfs dfs -rm -r /lab2