Exercise Sheet 13

Exercise 1 (MapReduce/Hadoop)

1. What is **Hadoop**?
2. Describe the functioning of the **MapReduce programming model**.
3. Explain (*in just a few sentences*) **two examples**, where MapReduce is helpful.
4. Describe the working method of the Google **PageRank algorithm**.
5. What is the advantage of the 64 MB chunk size of the Hadoop Distributed File System (**HDFS**)?
6. What is the drawback of the 64 MB chunk size of the **HDFS**?
7. What kind of data stores the **Namenode**?
8. What kind of data store the **Datanodes**?
9. What is **Pig**?
10. What is **Pig Latin**?
11. What is **Hive**?
12. What is **HBase**?
13. What is **Cloudera**?

Exercise 2 (Hadoop Cluster)

1. Launch a Hadoop Cluster in an infrastructure service like EC2, Google Compute Engine, HP Cloud Compute, or alternatively on your mobile personal computer.
2. Execute the π calculation example, which has been discussed in slide set 11.
3. Find a useful use case for your Hadoop cluster and try it out.
4. Present your use case during the exercise session.