10. Processing Big Data

10.1. Big Data Definition
   a) What is Big Data?
   b) Discuss the challenges that arise from the characteristics of big data.

10.2. Map/Reduce
   a) Explain the principle of MapReduce.
   b) The basic use of grep is to search for a specific string in the specified document and print all lines in which the string occurs. Propose a solution for implementing the problem of grepping a pattern from a large document corpus based on Map/Reduce.

10.3. Apache Hadoop
   a) What are the purpose and main characteristics of HDFS? What advantage and disadvantages result from these characteristics in comparison to other storage technologies?
   b) Describe how resources are managed between a set of applications in Hadoop with YARN?

10.4. Stream Processing
   a) What is the major difference between batch processing and stream processing?
   b) Describe the basic structure of an application in a Stream Processing Engine.
   c) Discuss the problem of event time windowing in stream processing applications by referring to the Yahoo! Finance Application.