

Exam for the “Big Data Systems” course

Task: Use the following systems:

- a. Hadoop
 - Implement a data analytics application using the MapReduce programming model **(10 points)**
 - Resources: <http://hadoop.apache.org/>
- b. Hadoop Distributed File System (HDFS)
 - Implement a simple application that uses basic HDFS file-system operations **(10 points)**
 - Resource: <http://hadoop.apache.org/>
- c. Apache Pig
 - Implement a query using basic database operators **(10 points)**
 - Resource: <https://pig.apache.org/>
- d. BigTable/HBase
 - Implement a simple application that uses basic BigTable operations/APIs **(10 points)**
 - Resource: <https://hbase.apache.org/>
- e. Spark
 - Implement an application using Spark APIs **(10 points)**
 - Resource: <https://spark.apache.org/>
- f. Spark streaming/Apache Flink
 - Implement a streaming application either using Spark streaming or Apache Flink **(10 points)**
 - Resource: <https://spark.apache.org/>
 - Resource: <https://flink.apache.org/>
- g. Zookeeper
 - Implement three distributed co-ordination protocols using Apache Zookeeper APIs **(10 points)**
 - Resource: <https://zookeeper.apache.org/>

Submission instructions:

Please upload all the application code in a public GitHub repository. Use separate folders per system to store the application code. Also, include a [README.txt](#) (per directory) to describe the implemented application. Github resource: <https://github.com/>

Please email the web-link of the github repository, with your name and university registration number to the following email address: [pramod \(dot\) bhatotia \(at\) ed \(dot\) ac \(dot\) uk](mailto:pramod(dot) bhatotia(at) ed(dot) ac(dot) uk)

Deadline: Please send your submission before 20th August, 2017.

Questions: For any questions or queries, please contact the ECI organizers:

eci2017@dc.uba.ar