BLM4821-Big Data Processing and Analysis Course

Term Project

Spring 2018

In this project, you will design and implement your own big data processing application that implements following requirements.

1. **Details & Requirements**
   - You will install a multimode hadoop cluster on your computer (ex: two-node cluster). You may use virtual machines or you may also use Amazon AWS to deploy your hadoop cluster.
   - You will write a java class that reads/write the dataset into the hdfs file system.
   - You will also write java classes that implement 5 different descriptive statistics function by using map reduce programming model. (Ex: average, min-max, range, standard deviation, etc.).
   - You will do a performance analysis of the functions that you implement.
   - Your application will have a Graphical User Interface from which you select the dataset to be analyzed and the different types of descriptive statistics function. You will also use your GUI to start hadoop jobs.
2. **First thing to do : Submit a proposal paper**
   - Write a 1 page proposal for your own application. The proposal should include;
     - Which dataset you will be using? Describe the dataset and provide a link to the Web site where you find it.
     - Give the list of descriptive statistics functions that you will implement on the dataset.
     - Give the list of students in your project group (You can form a group up to 2 people).
     - Development plan with reasonable milestones.
   - **Due Date: 18/04 (Hard copy of the proposal will be submitted in class.)**

3. **Second thing to do : Implement your system and write a 2-page long report about your application**
   - Your report should include
     - General information about your application
       A. Motivation, Goal, Implementation Environment (APIs,...), and so on

- Use case scenario of your application
- Technical challenges and how you solved them
- Explanation of your implementation
- Performance Evaluation
- Experience and Discussion

- How to submit
  - A. Please submit the files in .zip format with the filename corresponding to your student id.(e.g TP_StudentID1_StudentID2.zip) via email to course assistant Ahmet Elbir (aelbir@gmail.com) Email Title : [TermProject] Your Student ID, Your Name
  - Zip file should include your all source codes & report
  - **Due Date: Sunday, 22/05, 11:59PM**

4. Late Assignment Policy

- A. Term Projects are due at 11:59pm on the date specified. A grading penalty will be applied to late assignments. (10% penalty up to the first 24 hours, 20% for 24 to 48 hours, with no credit received after that)