

Lab Assignments for ITM 511

➤ Overview

Students will have 3 lab assignments throughout the semester. Total points for lab assignments are 10.

➤ Feedback strategy

The feedbacks will be given by the online submission and evaluation system (<http://eclass.seoultech.ac.kr>)

Lab Assignment #1 – Big data manipulation

➤ Assignment Overview

- **Released date:** 3/27
- **Due date:** 4/2 23:59 PM
- **Assigned points:** 3
- **Where to submit:** to e-class (<http://eclass.seoultech.ac.kr>)
- **Submission:** a word file including the followings: 1) random.py, 2) the captured result for each step, and 3) the explanation of each step
- Late submission is not allowed
- Description

Practice – Inserting Data from a File

1. Make a simple python script to generate a given numbers of strings (one per line)
 - `python random_strings.py 100`
2. Make three files using the python script to generate random strings with 10, 100, and 1000 numbers, respectively (called, `ten_strings.txt`, `hundred_strings.txt`, and `thousand_strings.txt`)
3. Upload the generated files into HDFS
4. Create a table in Impala to store the generated data
5. Load a file, called `thousand_strings.txt`, into the table and check the total inserted number of data
6. Append a file, called `hundred_strings.txt`, into the table and check the total inserted number of data
7. Overwrite a file, called `ten_strings.txt` into the table and check the total inserted number of data