# TI2736-B: Assignment 4 Big Data Processing

## Wing Nguyen, 4287118

**Public Repository:** https://github.com/codesalad/hahadoop
*The files are also included as a .zip file (wnguyen-4287118-code.zip)*

1. **Script:** assignment4/data_assignment4/query1.pig
   **Results:** assignment4/data_assignment4/q1_results

2. **Script:** assignment4/data_assignment4/query2.pig
   **Results:** assignment4/data_assignment4/q2_results

3. **Script:** assignment4/data_assignment4/query3.pig
   **Results:** assignment4/data_assignment4/q3_results

4. **Script:** assignment4/data_assignment4/query4.pig
   **Results:** assignment4/data_assignment4/q4_results

5. **Script:** assignment4/data_assignment4/query5.pig
   **Results:** assignment4/data_assignment4/q5_results

6. **Script:** assignment4/data_assignment4/query6.pig
   **Results:** assignment4/data_assignment4/q6_results

7. **Script:** assignment4/data_assignment4/query7.pig
   **Hadoop**: assignment4/data_assignment4/query7_hadoop.java
   **Results:** assignment4/data_assignment4/q7_results

   The Hadoop job took 15.71 seconds whereas the Pig Latin script took 24.16 seconds. A pig script has to be translated into a Hadoop job, so the extra ~10 seconds is for translating the pig script. The Hadoop job doesn't need to be translated and thus it is slightly faster.

8. **Script:** assignment4/data_assignment4/query8.pig
   **Jar:** assignment4/data_assignment4/udf_percentage.jar
   **Results:** assignment4/data_assignment4/q8_results

Usage scripts:
pig -x local query1.pig
>> creates output q1_results/