**CSC14118**

**Introduction to Big Data**

**Group 03**

**Lab03: Apache Spark with MongoDB**

|  |  |
| --- | --- |
| **Student ID** | **Full name** |
| 20127449 | Trần Quốc Bảo |
| 20127452 | Hồ Đăng Cao |
| 20127476 | Đỗ Đức Duy |

**LAB03 Version Date**

LAB03 v1.0 2023-08-

# Summary

|  |  |  |
| --- | --- | --- |
| **Section** | **Completed percentage** | **Issues** |
| 1a | 100% |  |
| 1b | 100% |  |
| 1c | 100% |  |
| 1d | 100% |  |
| 1e | 100% |  |
| 1f | 100% |  |
| 2a | 100% |  |
| 2b | 100% |  |
| 2c | 100% |  |
| 2d | 100% |  |

# Reflection

Does your journey to the deadline have any bugs?

The data in thuoc have many typo and inconsistent structures (have natural language, etc). This causes difficulties in gaining APIs for 2b and 2c questions.

We have to lowercase characters to compare, e.g. vitamin and Vitamin, but (vitamin) E also becomes (vitamin) e.

How have you overcome it?

What have you learned after this process?

If you cannot overcome the bugs, describe where the bottlenecks are in your work.

Contents

[Summary 2](#_Toc142728517)

[Reflection 2](#_Toc142728518)