

# CS6502: Week 7 Lab Exercises

Dr. Andrew Ju

March 8, 2021

## Exercise 1

The lab in this week is **supervised**. The lecturer/lab assistant will be available between 15:00 – 16:00 for any questions you may have.

The main task for today's lab is to get started and get familiar with Spark.

In Sulis class page, there is a forum “Labs (week 7)” created. In the forum, there are two topics:

- *Discussion*, use this topic to post your questions and discuss with your fellow classmates.

There is no submission requirement for this week's lab.

Below is a list of steps you should follow for today's lab.

- Google Cloud clean up, please follow the link [here](#)
- Install Spark in your local machine (lots of installation guide online)
- Practice PySpark (based on lecture slide) via PySpark Shell
  - create new RDD
  - create TextFileRDD
  - Read though list of transformations/actions
  - try out a few transformations (map, flatMap, filter, etc)
  - try out a few actions (reduce, take, first, collect, etc)
  - try out `persist`, `collect`
- If there are more time left, please revisit the MapReduce examples from last week's lecture, and re-think how you can you implement them in PySpark (part of next week's lecture/lab)