Assignment 4
STAT 3006, 2021/22

Due date: May 15, 2022 23:59, Sunday

Question 1

(40%)Please apply parallel computing techniques to the EM algorithm in Q4 of the Assignment 1, and contrast the execution time using parallel computing with that using the original EM code.

Question 2

(30%)There are three tables named "Student", "Book" and "Record" in the database "Library". Please write SQL queries to:

a) list the 'Book' table;

b) find the ids and the entry years of the students who have borrowed book from natural science

c) find the ids and majors of the students who have borrowed book 008 and occupied it for more than 30 days.

Hint:

(1) The code to access the data base is as follows:

```
library("RMySQL")

drv=dbDriver("MySQL")
con=dbConnect(drv,user="student", password="HappyStudy2022",
dbname="Library", port=3306,
host="rds-mysql-statclass.czyn7pdbk60s.us-west-2.rds.amazonaws.com")
```

(2) Please use the function `TIMESTAMPDIFF(day, BorrowingTime, ReturnTime)` to calculate the time difference.

Question 3

(30%)Please write programs to

a) find all the companies and ticker symbols listed on NASDAQ-100 from `https://www.slickcharts.com/nasdaq100`;

b) retrieve Market Cap, Price to Book Value, and Dividend Yield from Y-Charts (e.g. `https://ycharts.com/companies/FB` for facebook) for each of the NASDAQ-100 companies(note that the retrieve information may be not available for some companies) and output the table in your report;

c) list the top 3 companies with the highest Market Cap.

Hint: (1) When `htmlTreeParse` does not work directly with a given url, you may try the following command lines:

```
library(XML)
library(httr)
library(RCurl)
```

```
url_complist <- "https://www.slickcharts.com/nasdaq100"
doc_complist <- htmlTreeParse(rawToChar(GET(url_complist)$content),
useInternalNodes = TRUE)
```

(2) Please write code to systematically retrieve websites from Y-Charts using the company abbreviations you obtain in step (a).

Requirements:

-	in the paper report	in the R code file
Q1	the parallel computing algorithm the consumed time using parallel computing the consumed time using old EM algorithm	R code
Q2	SQL code answers to the three problems	R code
Q3	the informations in a) and b) the top 3 companies with the Highest Market Cap and their corresponding Y-Charts website screenshots	R code