

# IT542 Homework 1

## Due: July 15<sup>th</sup>, 2019

---

### Data Processing using Apache Spark (RDDs) and Python

1. Using **Capitals.txt** dataset;  
Find two closest capital cities and the distance between them.
2. Using **Capitals.txt** dataset;  
Find two capital cities furthest away from each other and the distance between them.
3. Using **DepremVerileti.txt** dataset;  
Please find the **list of foreshocks and aftershocks (within 20 km and in 24 hours)** for top ten earthquakes between years 1990 and 2019.
4. Using **SayisalLoto.txt** dataset,  
  
Please find top most commonly drawn **triple numbers**.  
  
(9,38,42) → 2 times
5. Using **DolarVerileri.txt** dataset,  
Please find top 5 greatest daily decrease (percentage

Please create one python file for each question as following.

YourName-YourLastname-Question1.ipynb  
YourName-YourLastname-Question2.ipynb  
YourName-YourLastname-Question3.ipynb  
YourName-YourLastname-Question4.ipynb  
YourName-YourLastname-Question5.ipynb

(Please do not use Turkish characters in your filename)

You can find all datasets in resources folder at course web page.