

## Homework 2

### Installation and Configuration:

Install Apache Spark on your computer. Encourage the use of virtual machines (Docker is an example).

### Data

Download the following data:

- airports.dat  
(<https://raw.githubusercontent.com/jpatokal/openflights/master/data/airports.dat>)
- routes.dat (<https://raw.githubusercontent.com/jpatokal/openflights/master/data/routes.dat>)

### Do the tasks with Spark:

- Create RDD from the files
- Count the number of airports
- Filter by country
- Group by country
- Count the number of airports in each country
- Join 2 RDD
- Count the number of flights arriving in each country