

# Exercises with Coordination and Management

These are a set of exercises to practice with some aspects of the Spark Core, focusing on data coordination and management. Comment the code explaining what is going on. You should deliver the solutions using [this template](#). For this exercises you will need:

- [pagecounts](#)
- [bible.txt](#)

Write Spark code, in python, to perform the following tasks:

1. Add the phone prefix to the phone numbers of the provided dataset (in the template) using as reference the [ISO-3166 country codes](#). Use a Broadcast Variable.
2. Count the number of lines that contain at least a 'to' in the file and the number of lines in the bible.txt file. Use Accumulators.
3. Read the pagecounts dataset, then (use Repartition):
  - a. Write the RDD containing the pagecounts dataset
  - b. Write the RDD but with only 2 partitions
  - c. Write means write it to your file system
4. Check the differences in computation time when using cache method on an RDD. Read pagecounts files and count lines with and without using cache method show the time differences