

# Big data: architectures and data analytics

## Spark - Exercises

## Exercise #49

- Input:
  - A csv file containing a list of profiles
    - Header: name,surname,age
    - Each line of the file contains one profile
    - name,surname,age
- Output:
  - A csv file containing one line for each profile. The original age attribute is substituted with an new attributed called rangeage of type String
    - rangeage = "[" + (age/10)\*10 + "-" + (age/10)\*10 + 1"]"

3

## Exercise #49

- Input:
  - name,surname,age
  - Paolo,Garza,42
  - Luca,Boccia,41
  - Maura,Bianchi,16
- Expected output:
  - name,surname,rangeage
  - Paolo,Garza,[40-49]
  - Luca,Boccia,[40-49]
  - Maura,Bianchi,[10-19]

4

## Exercise #50

- Input:
  - A csv file containing a list of profiles
    - Header: name,surname,age
    - Each line of the file contains one profile
    - name,surname,age
- Output:
  - A csv file containing one single column called "name\_surname" of type String
    - name\_surname = name+" "+surname

5

## Exercise #50

- Input:
  - name,surname,age
  - Paolo,Garza,42
  - Luca,Boccia,41
  - Maura,Bianchi,16
- Expected output:
  - name\_surname
  - Paolo Garza
  - Luca Boccia
  - Maura Bianchi

6