

## Task 1

The human resource section of the company bigdata2022 is conducting a survey about employee background. In particular, they want to collect data different information for each employee:

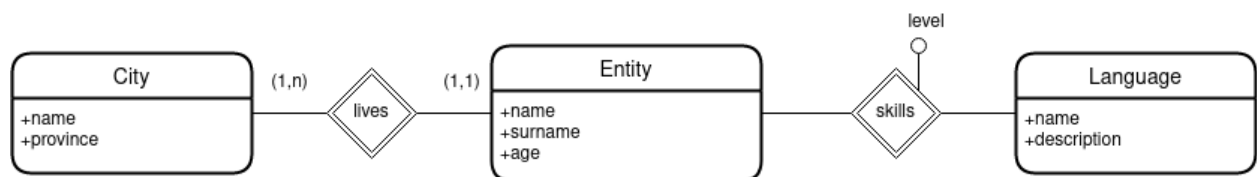
- name
- surname
- age
- living place - city (name and province)
- programming skills - list of programming languages (name and description) with confidence level for each one

Design a java application that the company can distribute in order to collect data of employee of different company sites

Note that data has to be stored as json objects

## Task 2

Data analysts want to reuse previous analysis performed on the following schema



Transform collected data into desired schema and execute the following query:

1. Find the language with the highest average level
2. Find the province with the minimum number of employee
3. For each language find the employee(s) with the highest level

## Task 3

Join the previous defined schema into a unique table named skills stored in a csv format

Design a mapreduce application that read the folder in which the *skills*' table data is stored and computes, for each city, the count of employees with at least 2 two language with a confidence level greater than 5

## Submission Instruction

Submit what you have done into a git repository with the following name  
*"name-surname\_01-test"*

In the file section of the general file there is an excel file named *"01-test\_git"* on which you are expected to add the link to your repository

Each folder is expected to contains:

- A project contain the java app of task 1
- A very simple report describing the design of the warehouse of task 2  
It is expected to write table creation statements and performed queries together with few works that motivates each statements
- A project containing the mapreduce job of task 3

## Useful material

```
<dependency>
  <groupId>com.google.code.gson</groupId>
  <artifactId>gson</artifactId>
  <version>2.10</version>
</dependency>
```