

Spark 2020 2C 1

Enunciado

Tenemos un RDD con información de recetas:

(ID_Receta, Nombre, Categoría)

Y otro RDD con los ingredientes de cada receta:

(ID_Receta, Ingrediente, Cantidad_Kg)

Queremos obtener:

a) Listar todos los ingredientes que aparecen en alguna receta que usa "pollo" indicando en cuantas recetas el ingrediente y pollo aparecen juntos. El formato de salida es (ingrediente, cantidad de recetas en que aparece junto con pollo). Por ejemplo, la papa aparece en 10 recetas con pollo, por lo que tendríamos (papa, 10). (50 pts)

b) Queremos obtener todos los nombres de recetas Mediterráneas que no tengan ni papa ni pollo entre sus ingredientes.(50 pts)

Resolver los puntos usando la API de RDDs de PySpark.