

Contrôle de connaissances

Intégration de données et Big Data

Durée 1h30, documents autorisés

© Mourad Ouziri
mourad.ouziri@u-paris.fr

QCM 1 (6 pts) – Intégration de données

Répondre au QCM en ligne : <https://moodle.univ-paris8.fr/mod/quiz/view.php?id=312544>

QCM 2 (6 pts) – Big Data

Répondre au QCM en ligne : <https://moodle.univ-paris8.fr/mod/quiz/view.php?id=312586>

Exercice (8 pts) – Programmation Big Data avec Spark Core

Considérons la liste des salaires suivantes :

```
var salaires = List (3000, 200000, 2500, 4000, 5000, 500000, 2000, 300000, 150000, 3000)
```

Les salaires inférieurs à 10000 sont français (donc en Euro, dénoté avec le symbole EUR) et le reste des salaires sont japonais (donc en Yen, dénoté avec le symbole JPY). Nous considérons le taux de change suivant : 1 EUR = 100 JPY.

A partir de cette liste des salaires, réaliser les calculs suivants (avec affichage des résultats) avec l'API Spark Core en assurant un partitionnement en 4 :

1. Ajouter le symbole de devise correspondant aux salaires. Le résultat attendu est le RDD : [3000 EUR, 200000 JPY, ..., 150000 JPY, 3000 EUR]
2. Calculer le salaire moyen en Euro sur l'ensemble des salaires français et japonais.
3. Calculer, dans un seul RDD, le salaire moyen (en Euro) des salaires français et le salaire moyen (en JPY) des salaires japonais. Le résultat attendu est le RDD : [("FR", 3250), ("JP", 287500)]

Quelques opérations sur RDD pouvant être utiles : `map`, `filter`, `reduce`, `reduceByKey`, `sum`, `mean`, `count`

Rendu du travail

Le travail réalisé pour l'exercice est à rendre dans un seul fichier *scala* (format texte) portant vos noms et incluant le code Spark réalisé pour chaque question (indiquée dans un commentaire) avec la trace de son exécution dans *spark-shell* (résultat retourné ou message d'erreur).

Indiquer le nom des deux binômes tout au début de ce fichier.

Merci de déposer le fichier immédiatement en fin d'épreuve à : <https://moodle.univ-paris8.fr/mod/assign/view.php?id=312542>

Bonne réussite !