

# Contrôle de connaissances

## Intégration de données et Big Data

Durée 1h30, documents autorisés

© Mourad Ouziri  
[mourad.ouziri@u-paris.fr](mailto: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 notes suivantes :

```
var notes = List (30, 16, 38, 10, 20, 28, 12, 18, 22, 32)
```

Ces notes sont exprimées sur 40 mais tous les résultats doivent être calculés et affichés sur 20.

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

1. Ajouter une appréciation à chaque note : « Bonne » si note  $\geq 13/20$ , « Insuffisante » si note  $< 10$  et « Satisfaisante » sinon. Le résultat attendu est le RDD : [15 Bonne, 8 Insuffisante, ..., 11 Satisfaisante, 16 Bonne]
2. Calculer le nombre de notes inférieures à la note moyenne. Le résultat attendu est : 6.
3. Calculer la moyenne des notes inférieures à 10/20 et la moyenne des notes supérieures ou égales à 10/20 dans un seul RDD. Le résultat attendu est le RDD suivant :  
[("inf\_10", 7), ("sup\_10", 14.16)]

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 !*