Faculté : NTIC, **Département :** IFA
Année Universitaire : 2023/2024
Module: TDSAI (Master 1 SDIA)

TP 01: Spark-RDD

Exercice 1

1. Créez un RDD contenant des nombres entiers inferieurs à 15.
2. Effectuer les opérations suivantes sur le RDD crée:

a. Multiplier les éléments du RDD par 2.
b. Afficher les éléments inferieurs à 5.
c. Transformer chaque élément du RDD crée en un seul RDD contenant des tuples où chaque tuple contient le nombre et le nombre suivant.
d. Trier les éléments du RDD crée dans un ordre décroissant.

Exercice 2

Ecrire un code Pyspark qui utilise les RDDs pour calculer la somme puis le nombre des éléments pairs et impairs de la liste suivante [12,10, 20, 7, 13, 19, 18, 23, 50].

Exercice 3

Ecrire un code Pyspark qui utilise les RDDs pour calculer le nombre d'occurrence de mots dans une collection de documents, ce problème est connu sous le nom de « Word count ».

Exercice 4

Nous disposons de l'historique de vente sur plusieurs années d'un magasin dans un fichier 'sales.csv', chaque enregistrement correspond à une vente de plusieurs articles. Les seules informations qui nous intéressent sont : la date, le nom de l'article, son prix ainsi que la quantité vendue.

Ecrire un programme Pyspark qui utilise les RDDs pour calculer le prix de vente total par produit.