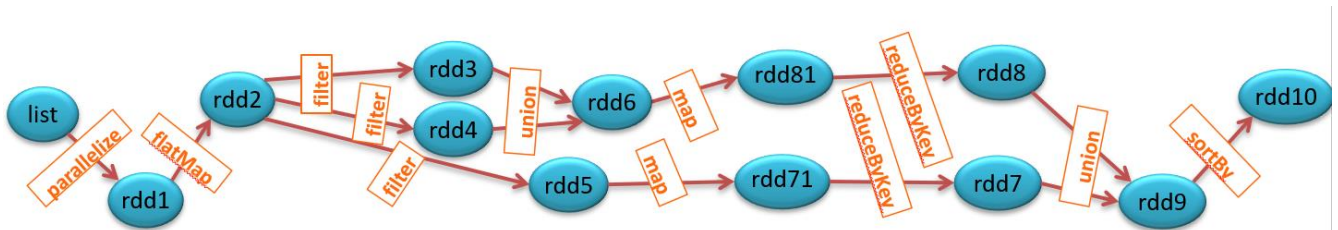


Exercices : Traitement parallèle et distribué avec Spark

Exercice 1 :

On considère le graphe suivant qui représente le lignée des RDDS :



Le premier RDD (RDD1) est créé en parallélisant une collection contenant des noms des étudiants. Ecrivez une application Spark qui permet de réaliser ce lignage.

Exercice 2 :

1. On souhaite développer une application Spark permettant, à partir d'un fichier texte (ventes.txt) en entrée, contenant les ventes d'une entreprise dans les différentes villes, de déterminer le total des ventes par ville. La structure du fichier ventes.txt est de la forme suivante :

date ville produit prix

Vous testez votre code en local avant de lancer le job sur le cluster.

2. Vous créez une deuxième application permettant de calculer le prix total des ventes des produits par ville pour une année donnée.