

Exercice

Soit le tableau suivant une partie d'un jeu de donnée qui contient des informations relatives aux patients qui subissent des traitements anti dépression dans une clinique. Ces données sont décrites par trois variables ; une variable binaire appelée *Anti* qui exprime est ce que le patient a pris ou non des antidépresseurs, une variable catégorique appelée *Longueur* qui représente la durée de dépression et quatre variables numériques qui présentent l'inventaire de dépression de Beck (IDB), mesurés avant la dépression et après 2, 4 et 6 mois. On souhaite mettre en place un programme Hadoop qui permet de calculer pour chaque période de dépression la moyenne des inventaires mesurés après 2 mois du traitement.

Observation	Anti	Longueur	IDB_Avant	IDB_2	IDB_4	IDB_6
1	Oui	<6m	21	18	11	12
2	Oui	>6m	12	12	12	10
3	Non	>6m	10	22	17	9
4	Oui	>6m	3	7	9	3
5	Oui	<6m	12	17	13	13
6	Non	<6m	11	10	10	11
7	Non	>6m	8	10	9	12
8	Non	<6m	7	5	2	5
9	Non	<6m	11	7	9	14
10	Oui	>6m	12	10	17	9

1. Donner les types des clés valeurs utilisés pour développer ce programme.
2. Écrire les classes java MoyenneDriver, MoyenneMapper et MoyenneReducer qui implémentent respectivement la fonction main(), map() et reduce().
3. Donner les commande Hadoop qui permettent de :
 - ✓ démarrer les services
 - ✓ importer les données dans HDFS
 - ✓ exécuter le programme