



Hadoop Mapreduce TD/TP

Exercice 1

L'objectif de cet exercice est d'exécuter le fameux exemple «Word Count » ou « Compteur d'occurrence de mots » afin de compter les occurrences de tous les mots dans un texte. En premier lieu via l'API Java, puis via Apache Hadoop (installé par l'étudiant).

Par exemple, pour le texte ci-après :

Valeur Vélocité Véracité Valeur Variété Volume

Le résultat sera:

| | |
|----------|---|
| Valeur | 2 |
| Vélocité | 1 |
| Véracité | 1 |
| Variété | 1 |
| Volume | 1 |

Exercice 2

Ecrire une job Mapreduce qui à partir d'un texte d'entrée, calcule le nombre des mots commençant par un caractère donné.

Par exemple, pour le texte ci-après :

HDFS Job Mapreduce JobTracker Namenode Hadoop

Le résultat sera:

H 2
J 2
M 1
N 1

Exercice 3

Ici, on cherche à trouver des anagrammes à partir d'un texte d'entrée.

Une anagramme est le résultat de la permutation des lettres d'un mot de manière à produire un autre mot qui a un sens.

Par exemple : « argent » ; « tanger » ; « gerant » sont des anagrammes.

Ecrire une job Mapreduce qui, à partir d'un texte d'entrée permet d'extraire tous les mots anagrammes.

Exercice 4

Un palindrome est un mot qui reste le même qu'on le lise de gauche à droite ou de droite à gauche (par exemple, PIERRE n'est pas un palindrome, alors que OTTO est un palindrome).

Ecrire une job Mapreduce qui, à partir d'un texte d'entrée permet d'extraire tous les palindromes.