



Module: Big Data Analytics Spécialité : Master MIDVI Année universitaire 2020/2021

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:

Le résultat sera:	Application of the second of t
Valeur	1
Vélocité	A STATE OF THE STA
Véracité	1 The second sec
Variété	1
Volume	E. IL Control of Children & Street

Exercice2

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:

Le résultat sera:	
10000000000000000000000000000000000000	是是是一种的基本是是是是一种的一种。 第一种的一种的一种,一种的一种一种一种一种一种一种一种一种一种一种一种一种一种一种
1000 全国的一个企业的企业,这种特殊的	。 (1)
Jan 2	THE RESERVE AND A STREET OF THE PARTY OF THE
MATTER	
2. 一种的一种 1. 10 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	
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.