

Série de TD N°3
HDFS

Exercice 1

Dossier *patchesFolder* :

Fichier1 : *patches2018.txt*, taille= 513MB

Fichier2 : *patches2019.txt*, taille= 515MB

Nombre maximum d'instances Mapper : 5

Bloc HDFS = 512MB.

Fichier1 : on utilise 2 blocs

Fichier2 : on utilise 2 blocs

Le nombre d'instances Mapper utilisées lorsqu'on exécute une application MapReduce sur les deux fichiers du dossier *patchesFolder* doit supporter les quatres blocs des données en entrée. **On utilise donc 04 instances de Mapper.**

Exercice 2

Nombre maximum d'instances Mapper : 10

La fonction Mapper détermine les couples (date, temperature) pour chaque fichier

La fonction Reduce détermine la valeur maximale de la température *maxTemp* pour chaque date

Le résultat du programme MapReduce est le couple (2016/01/01, 1.5)

Exercice 3

Taille du fichier logs.txt : 5000 MB

Pour forcer Hadoop à exécuter 10 mappers en parallèle, on doit diviser le fichier en 10 blocks, la taille de chaque block sera : $5000 / 10 = 500$ MB, la réponse choisie est donc : d) Block size=512 MB

Exercice 4

Taille du fichier logs1.txt: 1036MB

Taille du fichier logs2.txt : 500MB

Facteur de réplication : 4

Taille du block : 512 MB

Nombre de blocks nécessaires pour stocker le fichier logs1.txt : $1036 / 512 = 2.023$ soient 3 blocks

Nombre de blocks après réplication : $3 * 4 = 12$ blocks

Nombre de blocks nécessaires pour stocker le fichier logs2.txt : $500 / 512 = 0.97$ soit 1 block

Nombre de blocks après réplication : $1 * 4 = 4$ blocks

Nombre total de blocks =16 blocks

La réponse d) est choisie.