

Point d'avancement 2

1- Modifications envisagées dans le but d'améliorer les temps d'exécution :

- Au lieu de créer des fichiers intermédiaires qui contiennent les résultats des map, on pourrait envoyer directement les résultats à la machine qui s'occupera de faire l'opération reduce.
- Permettre de faire plusieurs reduce ce qui permettrait de paralléliser le traitement et donc d'améliorer la vitesse (Au lieu de sommer le nombre d'occurrences d'un mot dans les différentes parties d'un fichier. On pourrait faire la somme dans la moitié dans le premier reduce et l'autre moitié dans le deuxième reduce puis refaire un reduce sur les résultats des deux.
- Ne plus envoyer un seul KV mais envoyer un tableau de KV.

2- Plan de test :

- Critère à évaluer : durée d'exécution.
- Paramètres à varier :
 - Taille des fichiers
 - Nombre de machines
 - Nombre de reducers.
 - Taille du tableau de KV à envoyer
- Procédure de test :
 - Faire varier la taille des fichiers et le nombre de machines pour la version actuelle et garder les résultats (vitesse).
 - Pour les mêmes tailles et le même nombre de machines, faire varier le nombre de reducers et la taille du tableau pour optimiser la vitesse.