

Répartition des tâches Hidoop

1er rendu : 29 Janvier

1. Déploiement centralisé via SSH

Responsable : Wilfried

Description

Permettre le déploiement des démons sur tous les noeuds depuis le master.

2. Mesure de l'overhead (surcôt)

Responsable : Charles

Description

Elimination de la parallélisation (threads) et comparaison avec l'algorithme itératif. Ce surcôt devra être mesuré au moins pour le `WordCount`

2ème rendu : 14 Février

3. Etude de l'influence de la taille des blocs

Responsable : Charles

Description

Tout est dans le titre. Faire une étude des performances de la version *Map Reduce de Wordcount* pour des tailles de bloc différentes.

4. Ecriture des applications de Benchmark en Hidoop

Responsables : Rachid, Wilfried, Obeida

Description

Il s'agira de regarder à quoi ressemble de code *Hadoop* d'une application et s'en inspirer pour écrire le code *Hidoop*.

| Application | Responsable |
|-------------------------------|-------------|
| Algorithme X de Knuth | Rachid |
| Page Ranking | Wilfried |
| Décimales de Pi (Monte Carlo) | Obeida |

5. Etudes des scalabilité pour les autres application.

Responsables : Rachid, Wilfried, Obeida

Description

- On conserve la répartition précédente. Chacun étudiera la scalabilité sur son application.
- Pour une taille de fichier et de bloc on augmente le nombre de noeuds
- Pour une taille de fichier et un nombre de noeuds, on augmente la taille des blocs
- Chaque responsable devra modifier le script `data-gen` pour qu'il permette générer des données appropriées pour une application. Par exemple, `data-gen wordcount mb 128` devra générer un fichier d'entrée de 128Mb pour wordcount.
- Pour chaque application il faudra donc se renseigner sur le type de données qu'elle attend en entrée.

3ème rendu : 18 Mars

6. Amélioraton Hdfs : Intégration du facteur de réplication dans Hdfs

Responsables : Rachid, Obeida

Description

Un fragment devra posséder plusieurs réplicas réparties sur différents noeuds (autant que le facteur de réplication). L'interface d'Hdfs doit rester la même. (Transparence du point de vu de utilisateur)

7. Amélioration Hidoop : Introduction de shuffles

Responsables : Wilfried, Charles

Description

- Des shuffles qui s'effectuent sur chaque noeud => introduction d'un nouveau format `KVlist`
- Un shuffle traite tous les résultats de map se trouvant sur le noeud.
- On garde un reducer sur le master.

4ème rendu (final & Oral) : 03 Avril

8. Nouvelle étude de scalabilité

Responsables : Tous

Description

- On reprend les applications du 2ème Rendu et on étudie à nouveau la scalabilité de la plateforme (améliorée) sur ces applications. On compare les résultats avec ceux obtenus lors du 2ème Rendu.
- Chacun s'occupe de son application.

9. Rédaction du rapport

Responsables : Tous

| |
|--|
| <div>Remarques</div> <div><ul style="list-style-type: none">• L'étude de la scalabilité sur une application se fera<div><div>- pour des tailles de données allant de 64Mb à 10Gb</div><div>- pour un nombre de noeuds allant de 1 à 5</div></div>• Pour toutes les études, prévoir des résultats et les comparer avec les résultats effectivement obtenus.</div> |
|--|

