

Rapport Hidoop v.0

(Map/Reduce)

Architecture

Daemon_dataNode, serveur RMI tournant sur chaque machine « DataNode »

Job, le client RMI est appelé par MyMapReduce qui tourne sur la machine client

SlaveMap/SlaveEnvoyerVers/SlaveRecevoir permettent tout simplement la parallélisation des appels, ils sont utilisés par Job.

Job appelle la méthode Map sur tous les datanodes, appelle la méthode recevoir sur le datanode reducer puis appelle la méthode EnvoyerVers sur les datanodes non reducer (afin d'envoyer les resultats des maps vers le reducer), enfin il appelle reduce sur le reducer.

Callback permet de savoir quand on a fini d'exécuter une methode remote mais il n'est pas utilisé car on utilise un `.join()` pour le remplacer.

KV est un type contenant une valeur et une clef.

KVFormat est une classe qui permet de lire et d'écrire un fichier de type KV.

LigneFormat est une classe qui permet de lire et d'écrire un fichier de type « ligne » (c'est à dire, un fichier « normal » contenant des phrases).

Project contient le PATH du projet.

Points délicats

Dans le reduce de MyMapReduce on a été forcé de faire une boucle for au dessus du `while ((kv = reader.read()) != null)` car kv prenait la valeur null alors que le fichier n'avait pas été lu entièrement.

L'architecture a aussi posé quelques problèmes car elle est assez complexe et imbriquée (toutes les classes s'appellent les unes les autres).

Opérations essentielles

Cette partie du projet consiste principalement en de la communication entre machines et en de la lecture/écriture de fichiers. Tout doit se faire en parallèle donc il y a beaucoup de Threads (les Slaves).

Comment utiliser le système Map/Reduce

Modifier le PATH dans project afin qu'il corresponde à l'emplacement du projet.

Lancer Daemon_dataNode avec l'argument 6060 sur toutes les machines dataNodes.

Modifier liste_addr dans la classe Job pour mettre les hostnames des machines dataNodes.

Modifier addr_reduce dans la classe Job pour mettre le hostname de la machine qui servira de reducer (cette dernière doit avoir lancé un Daemon_dataNode !).

Lancer MyMapReduce avec l'argument « nom_fichier » sur une machine quelconque.

Le reducer aura créé un fichier « nom_fichier »-rec-rec qui est le résultat du Map/Reduce.