

Simulation d'algorithmes d'équilibrage de charge dans un environnement distribué

Identifications des besoins

Kevin Barreau

Guillaume Marques

Corentin Salingue

30 janvier 2015

Résumé

Ce document dégage une première identification des besoins. Il s'agit d'un document support pour l'élaboration du cahier des charges

Sommaire

| | | |
|----------|---|----------|
| 1 | Définition du projet | 3 |
| 1.1 | Contexte | 3 |
| 1.2 | Finalité | 3 |
| 2 | Besoins fonctionnels | 5 |
| 2.1 | Environnement distribué | 5 |
| 2.2 | Gestion d'un réseau | 5 |
| 2.2.1 | Gestion des noeuds | 5 |
| 2.2.2 | Gestion des objets | 5 |
| 2.2.3 | Popularité d'un objet | 5 |
| 2.2.4 | Topologie du réseau | 5 |
| 2.2.5 | Sauvegarde d'un réseau | 6 |
| 2.2.6 | Importation d'un réseau | 6 |
| 2.3 | Test des algorithmes | 6 |
| 2.4 | Simulation de requêtes | 6 |
| 2.4.1 | Gestion des requêtes | 6 |
| 2.4.2 | Envoi d'une requête | 6 |
| 2.4.3 | Création d'un jeu de requête | 6 |
| 2.4.4 | Sauvegarde d'un jeu de requête | 6 |
| 2.4.5 | Importation d'un jeu de requête | 6 |
| 2.5 | Visualisation des données | 7 |
| 3 | Besoins non fonctionnels | 8 |
| 3.1 | Cassandra | 8 |
| 4 | Livrables | 9 |

1 Définition du projet

1.1 Contexte

L'expansion, au cours des deux dernières décennies, des réseaux et notamment d'Internet a engendré une importante création de données, massives par leur nombre et leur taille. Stocker ces informations sur un seul point de stockage (ordinateur par exemple) n'est bien sûr plus envisageable, que ce soit pour des raisons techniques ou pour des raisons de sécurité (pannes potentielles par exemple). Pour cela des systèmes de stockages dits distribués sont utilisés en pratique afin des les répartir sur différentes unités de stockages.

Définition Un environnement distribué est constitué de plusieurs machines (ordinateurs), appelées *noeuds*, sur lesquelles sont stockées des données.

Définition Une donnée est une suite binaire de 0 et de 1 dont le contenu n'est pas important pour l'application.

Le client souhaite répartir toutes ces données de manière équitable entre les noeuds. De plus, il souhaite que ces données soient accessibles afin de pouvoir les requêter et récupérer de l'information.

Définition Une requête est un message envoyé à une machine (ou plusieurs machines) afin de récupérer ou de modifier de l'information sur des données. Nous noterons que la nature de l'information est inutile pour le bon fonctionnement de l'application.

Pour répartir toutes ces données, notre client a développé de nouveaux algorithmes d'équilibrage de charge et de réplication qu'il souhaite tester dans un environnement distribué.

Définition Une *charge* est associée à un noeud et désigne le nombre de requêtes que le noeud doit traiter. La réplication d'une donnée consiste à faire des copies de cette donnée sur d'autres noeuds.

1.2 Finalité

Nous devons développer une solution logicielle permettant de tester ces nouveaux algorithmes d'équilibrage de charge et de réplication dans un environnement distribué.

Définition Un réseau est un ensemble de noeuds qui sont reliés entre eux (en général par Internet) et qui communiquent ensemble afin de traiter toutes les requêtes reçues.

Définition La topologie d'un réseau représente l'architecture physique ou logicielle des liens entre les noeuds. Elle comporte aussi des informations sur la hiérarchie des noeuds, le placement spatial et les divers équipements reliant les noeuds.

Définition L'état d'un réseau est l'ensemble des informations caractérisant un réseau (topologie par exemple) ainsi que des informations sur les noeuds (comme leur charge actuelle).

Cette solution doit permettre le paramétrage d'un réseau, c'est-à-dire le nombre de noeuds souhaité et la topologie du réseau. Le client pourra simuler différents jeux de données et jeux de requêtes sur ce réseau. Il pourra tester ses algorithmes implémentés. Enfin, il pourra visualiser la topologie et l'état du réseau à tout moment.

2 Besoins fonctionnels

L'environnement de simulation voulu est un système distribué constitué de n noeuds de stockage dans lequel on souhaite stocker m objets.

2.1 Environnement distribué

2.2 Gestion d'un réseau

Comme défini précédemment, un réseau est un ensemble de noeuds qui sont reliés entre eux (en général par Internet) et qui communiquent ensemble afin de traiter toutes les requêtes reçues.

2.2.1 Gestion des noeuds

Un noeud est une machine (ordinateur généralement) pouvant stocker des données et traiter des requêtes. Un noeud possède des données internes. Ces données sont propres au fonctionnement du noeud (comme la charge par exemple).

Le projet nécessite de :

- Créer un noeud (*Priorité* : 1, *criticité* : 1)

La création d'un noeud se fait au moment où le réseau n'existe pas encore. Le projet ne prend en compte ni la création dynamique de noeud (création et ajout d'un noeud après que le réseau soit créé), ni la suppression de noeud. Il est possible de séparer ce besoin en plusieurs sous-besoins :

- Créer un noeud dans l'environnement de simulation
- Initialiser les données internes d'un noeud
- Récupérer les données internes d'un noeud (*Priorité* : 1, *criticité* : 1)
- Mettre à jour les données internes d'un noeud (*Priorité* : 1, *criticité* : 1)

2.2.2 Gestion des objets

Un objet est...

Le client souhaite pouvoir :

- Créer un objet
- Supprimer un objet

2.2.3 Popularité d'un objet

Les algorithmes du client nécessitent de connaître la popularité d'un objet dans le réseau. La popularité d'un objet est fonction du nombre de requêtes sur cet objet. Plus ce nombre de requête est grand, plus l'objet est populaire.

2.2.4 Topologie du réseau

Un noeud maître est un noeud connu de tous les autres noeuds dans le réseau. Un noeud communique à un intervalle de temps régulier avec les autres noeuds du réseau. Tous les noeuds du réseau possèdent des informations sur les autres noeuds du réseau.

- Communiquer les données internes à un noeud maître (*Priorité* : 1, *criticité* : 1)
- Paramétrer l'intervalle de temps de communication (*Priorité* : 2, *criticité* : 3)
- Stocker les informations des autres noeuds (*Priorité* : 1, *criticité* : 2)

2.2.5 Sauvegarde d'un réseau

Lorem ipsum

2.2.6 Importation d'un réseau

Lorem ipsum

2.3 Test des algorithmes

La conformité des algorithmes implémentés est assuré par des jeux de tests qui suivent la démarche :

- Créer un jeu de test à la main, comprenant les paramètres de création et l'ensemble des requêtes à traiter
- Faire fonctionner l'algorithme à la main
- Stocker l'état final du fonctionnement de l'algorithme
- Faire valider le processus à la main par le client
- Lancer la simulation de l'algorithme avec le jeu de test créé à la main
- Vérifier les résultats constatés avec les résultats attendus

Si une différence est visible lors de la vérification, une vérification par le client peut être envisagée dans le cas de résultats *presque* similaires. La notion de similitude est laissée à l'appréciation de l'équipe en charge du projet, lors de la vérification.

2.4 Simulation de requêtes

Une requête est un message envoyé à une machine (ou plusieurs machines) afin de récupérer ou de modifier de l'information sur des données.

2.4.1 Gestion des requêtes

Le client souhaite pouvoir :

- Créer une requête
- Supprimer une requête

2.4.2 Envoi d'une requête

Lorem ipsum

2.4.3 Création d'un jeu de requête

Lorem ipsum

2.4.4 Sauvegarde d'un jeu de requête

Lorem ipsum

2.4.5 Importation d'un jeu de requête

Lorem ipsum

2.5 Visualisation des données

- Temps de réponse moyen sur les requêtes passées.
- Charge d'un noeud
- Popularité des objets

Note Bien définir ces items.

3 Besoins non fonctionnels

3.1 Cassandra

Cassandra est une base de données distribué.

4 Livrables

