

Rapport de TP - Cassandra DB

loïc divad

Mai 2015

1 Introduction

...

2 Installation

Pour réaliser l'installation du logiciel avec la dernière version en date: *2.1.7*,

on procède d'un manière différente en indiquant toute les étapes.

```
lmdadm $ wget ftp://mirrors.ircam.fr/.../apache-cassandra-2.1.7-bin.tar.gz
```

```
lmdadm $ tar -zxvf apache-cassandra-2.1.7-bin.tar.gz
```

On redirige en suite les logs en éditant le fichier *logback.xml*.

...

```
<file> /home/lmdadm/log/cassandra/system.log </file>
```

...

Puis on édite le fichier *cassandra.yaml*. La liste suivante présente tous la paramètres personnalisés dans le cadre du TP:

- *cluster_name*: Cassandra Cluster
- *data_file_directories*: /home/lmdadm/data/cassandra
- *commitlog_directory*: /home/lmdadm/log/cassandra/commitlog
- *saved_caches_directory*: /home/lmdadm/tmp/cassandra/saved_caches

On lance ensuite la base de donnée. On note que l'option indiquée dans le TP -f signifie: *force foreground*. Conclusion, nous n'utiliserons sur tout pas cette option. Après l'avoir lancé en mode démon, on ouvre la console:

```
lmdadm $ ./bin/cassandra
```

```
lmdadm $ ./bin/cqlsh
```

2.1 Logiciel d'administration

Point sur les logiciels gui pour cassandra

3 Base de donnée

3.1 Description de l'espace de clef

Pour mieux comprendre la structure de donnée dans cassandra cette partie propose un petit point sur ce qu'est un espace de clef, une brève définition et une comparaison avec d'autre systemes connus. On propose également de décrire la commande de création proposé dans le TP:

```
cqlsh> CREATE KEYSPACE TP WITH REPLICATION = {
'class' : 'SimpleStrategy',
'replication_factor' : '1'
};
```

Definition(Keyspace): Il s'agit de la plus grande structure de données de cassandra. Elle est comparable au base (ou schéma) dans les systèmes classiques ou dans les systèmes document (mongodb, couchdb). Elle contient des familles de colonnes, qui elles, représentent la séparation entre les tables (mysql) ou les collections (mongodb). Chaque ligne insérée présente des informations pour certaines colonnes d'une famille de colonnes.

Propriétés:

- **Replication factor:** définit le nombre de noeud possédant une copie de la donnée.
- **Replica placement strategy:** il s'agit du mode de répartition des réplicats. Il en existe trois. simple strategy, old network topology strategy, network topology strategy
- **Column families:** Liste des familles des colonnes associées. Dans notre cas sur le keyspace TP il y a user, follows et message.

La commande de création du Keyspace TP signifie donc: crée l'espace de clefs TP dont les lignes ne sont présentes que sur un noeud. Si il doit y avoir de la réplication le *partitioner* redistribuera automatiquement sur le noeud suivant.

La commande DESCRIBE nous permet d'avoir la liste de keyspaces.

```
cqlsh> DESCRIBE KEYSPACES;
system_traces OpsCenter system tp
```

On peut ensuite accéder aux familles de colonnes sur un espace de clefs particulier.

```
cqlsh> use TP;
cqlsh:tp> DESCRIBE TABLES;
follows message user
```

On peut ensuite retrouver les colonnes d'une famille en particulier.

```
cqlsh:tp> DESCRIBE TABLE user;
CREATE TABLE tp.user (
  user_id bigint PRIMARY KEY,
  user_name text
)
```

3.2 le CQL, langage de requête Cassandra

remarque: On note que l'utilisation de double quotes (") dans la ligne de commande provoque l'erreur:

```
SyntaxException: <ErrorMessage code=2000 [Syntax error in CQL query] message="line
1:58 extraneous input hello world expecting ) (...) values (1,1 hello world[])...>
```