



Cati Sicpa

Ce document est mise à disposition selon les termes de la  
[Licence Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/)

# Configuration d'un cluster Apache Solr sous CentOS 7

29 août 2018

## 1 Objectifs

Dans le précédent document nommé "**InstallSolr**", nous avons décrit la procédure d'installation de Solr dans une machine fonctionnant sous CentOS 7. Cette installation est suffisante si le volume des données à indexer pour la recherche est statique et reste dans la limite des caractéristiques physiques de cette machine. Dans la plupart des cas, la quantité des données à indexer pour la recherche augmente avec le temps. Ce qui accroît le risque de dépasser les caractéristiques physiques d'un seul serveur Solr et de dégrader ses performances. Dans une telle situation, il est recommandé de mettre en place un cluster constitué de plusieurs serveurs Solr qui partitionnera et stockera de façon distribuée ces données. Notons qu'un cluster de serveurs Solr est communément appelé SolrCloud.

SolrCloud est donc un système de stockage et de recherche distribués. Un des grands problèmes dans les systèmes distribués est celui de la coordination des processus et des services fournis par ce système. Apache Zookeeper est un projet mature qui propose une excellente solution à ces problèmes de coordination dans les systèmes distribués. Apache Solr a opté pour cette solution en confiant la coordination de ses services à Zookeeper. Ainsi, Apache Zookeeper est embarqué dans Solr. Lorsque Apache Solr est utilisé en mode SolrCloud, une seule instance Zookeeper est activée dans l'un des serveurs Solr constituant ce cluster pour assurer la coordination des services. Cette coordination inclut la réplication et la redistribution des données contenues dans d'autres serveurs Solr en cas de panne. Cependant, le fait qu'une seule instance Zookeeper orchestre tout le cluster Solr pose un problème, car si le serveur Solr contenant cette instance Zookeeper actif tombe en panne, tout le système devient défaillant.

SolrCloud possède ainsi un point unique de défaillance que l'on peut résoudre en confiant son orchestration à un ensemble externe constitué d'un nombre impair de serveurs Zookeeper distincts. SolrCloud couplé avec un cluster externe Zookeeper reste hautement disponible et ne présente plus de point unique de défaillance tant que la majorité des serveurs Zookeeper est en vie et se communique mutuellement. L'objectif de ce deuxième document consacré à Apache Solr est de mettre en place SolrCloud couplé à un ensemble externe Zookeeper. Nous disposons de 12 machines virtuelles CentOS 7 dans chacune desquelles Solr sera installé. Nous installerons notre ensemble Zookeeper dans 3 de ces machines.

## 2 Installation et démarrage d'un ensemble Zookeeper

Nous installons et configurons le cluster Zookeeper dans nos trois machines virtuelles nommées **node11**, **node12** et **node13**. Pour cela, nous avons suivi les instructions que nous avons fournies dans le document nommé "Install ZooKeeper Cluster 2" consacré à l'installation de ZooKeeper. Ensuite, nous démarrons et laissons actif chacun de ces trois serveurs Zookeeper avec la commande suivante :

```
# zkServer.sh start
```

## 3 Installation de Solr

Nous avons installé Apache Solr dans chacun de nos 12 machines virtuelles appelées **node2**, **node3**, ..., **node13** en suivant la procédure décrite dans notre document nommée "InstallSolr".

## 4 Connection SolrCloud-Zookeeper

On exécute la commande ci-dessous dans l'un des serveurs Solr :

```
# solr zk mkroot /solr -z 10.10.10.11:2181,10.10.10.12:2181,10.10.10.13:2181
```

Ensuite, on édite le fichier `solr.in.sh` dans chaque serveur Solr. Puis, on y modifie la valeur du paramètre `ZK_HOST`, on enregistre et on ferme ce fichier.

```
# cd /opt/solr-7.4.0/bin
```

```
# vi ./solr.in.sh
```

```
ZK_HOST="10.10.10.11:2181,10.10.10.12:2181,10.10.10.13:2181/solr"
```

Le but des actions ci-dessus est d'indiquer à chaque serveur Solr où se trouve l'ensemble Zookeeper à utiliser.

On charge les fichiers de configuration par défaut de Solr dans l'ensemble Zookeeper en exécutant les commandes ci-dessous dans l'un des serveurs Solr :

```
# solr zk cp file:/opt/solr-7.4.0/server/solr/solr.xml zk:/solr.xml
```

```
# zkcli.sh -cmd upconfig -zkhost 10.10.10.11:2181,10.10.10.12:2181,10.10.10.13:2181  
-confname my_first_config -confdir /opt/solr-7.4.0/server/solr/configsets/_default/conf
```

```
# zkcli.sh -cmd upconfig -zkhost 10.10.10.11:2181,10.10.10.12:2181,10.10.10.13:2181  
-confname my_second_config  
-confdir /opt/solr-7.4.0/server/solr/configsets/sample_techproducts_configs/conf
```

Précisons que les adresses IP indiquées dans les commandes ci-dessus sont celles des serveurs constituant l'ensemble Zookeeper.

## 5 Démarrage du cluster SolrCloud-Zookeeper

A ce stade, la configuration de base est terminée et notre cluster Solr est prêt à être mis en route. Puisque l'ensemble des serveurs Zookeeper est nécessairement resté actif durant cette configuration, Il ne reste plus qu'à démarrer successivement chacun des serveur Solr. On peut le faire en exécutant dans chacun d'eux la commande ci-dessous dans laquelle le nombre "X" est à remplacer avec celui qui figure dans le nom "nodeX" du serveur considéré et l'option "-force" est nécessaire si on est connecté sur le serveur en tant que utilisateur "root" :

```
# solr start -cloud -h 10.10.10.X -force
```

On peut voir le status de chaque serveur Solr avec la commande suivante :

```
# solr status
```

On peut arrêter un serveur en particulier avec la commande suivante exécutée dans ce serveur :

```
# solr stop
```

On peut arrêter simultanément tous les serveurs Solr du cluster avec la commande suivante dans l'un de ces serveurs :

```
# solr stop -all
```

## 6 Conclusion

Dans ce document, nous avons montré comment mettre en place un cluster Solr hautement disponible, résistant aux pannes et ne présentant pas de point unique de défaillance. Pour la suite, nous allons sécuriser ce cluster en configurant convenablement les paramètres d'authentification et d'autorisation. Ensuite, à travers quelques lignes de commandes, nous montrerons comment indexer et rechercher les données dans ce cluster Solr.

## Références