

BPI France - plateforme Data

Procédure de chiffrement HDFS

—

Ce document a pour but de décrire les procédures permettant d'administrer et utiliser le HDFS chiffré.

Sommaire

[Introduction](#)

[Gérer les clés de chiffrement](#)

[Créer une clé](#)

[Lister les clés](#)

[Supprimer une clé](#)

[Manipuler les données chiffrées](#)

[Créer un dossier chiffré](#)

[Lister les dossiers chiffrés](#)

[Écrire des données](#)

[Lire des données](#)

Introduction

La fonctionnalité de chiffrement du HDFS permet de chiffrer certaines données afin qu'en cas de vol de données brutes ou d'intrusion dans le réseau, les données ne soient pas lisibles par l'attaquant.

En l'état actuel, l'administration et l'utilisation du chiffrement du HDFS n'est réalisable que via la ligne de commande `hdfs`. Ces lignes de commandes peuvent être lancées via un job Sqoop.

Un HDFS chiffré n'implique pas que la totalité des données sont chiffrées. L'utilisateur du HDFS a le choix de chiffrer ou non les dossiers qu'il crée.

Gérer les clés de chiffrement

Avant de pouvoir manipuler n'importe quelle donnée chiffrée, il est nécessaire de créer des clés de chiffrement qui serviront à chiffrer les dossiers souhaités.

Créer une clé

Avant de chiffrer un dossier, la clé de chiffrement correspondante doit être créée via la ligne de commande suivante :

```
hadoop key create le-nom-de-ma-cle -size 128
```

Attention, les noms de clés ne peuvent pas contenir de majuscule.

Il est également possible de définir une description sur la clé lors de sa création :

```
hadoop key create le-nom-de-ma-cle -size 128 -description "Ma description"
```

Lister les clés

A tout moment, il est possible de lister les noms des clés existantes via la ligne de commande suivante :

```
hadoop key list
```

S'il est nécessaire d'afficher également les métadonnées des clés, c'est possible avec une option supplémentaire :

```
hadoop key list -metadata
```

Supprimer une clé

Il est possible de supprimer une clé qui est devenu inutile grâce à la commande suivante :

```
hadoop key delete le-nom-de-ma-cle -f
```

Attention, supprimer une clé qui est encore utilisée provoquera la perte définitive des données correspondantes.

Manipuler les données chiffrées

Créer un dossier chiffré

Pour chiffrer un dossier, il est impératif que celui-ci soit vide. Par conséquent, il faut donc commencer par créer un nouveau dossier. Ceci peut se faire grâce à la ligne de commande suivante :

```
hdfs dfs -mkdir /monDossier
```

Ensuite, il est possible de chiffrer le dossier avec une clé précédemment créée en utilisant la ligne de commande suivante :

```
hdfs crypto -createZone -keyName le-nom-de-ma-cle -path /monDossier
```

A partir de cet instant, le dossier et toutes les données qu'il contient seront chiffrés.

Lister les dossiers chiffrés

Le nombre de dossiers chiffrés et de clés n'est pas limité. Il est possible de lister tous les dossiers chiffrés ainsi que les noms des clés de chiffrement correspondantes avec la ligne de commande suivante :

```
hdfs crypto -listZones
```

Le résultat ressemblera aux lignes suivantes :

```
/monDossier          le-nom-de-ma-cle  
/user/hdfs/encrypted kms-key1  
/unAutreDossier     kms-key2
```

La première colonne correspond aux dossiers chiffrés et la deuxième colonne indique les noms des clés liées aux dossiers de la même ligne.

Écrire des données

Après avoir créé des dossiers chiffrés, il est possible d'écrire des fichiers dans ces dossiers et chaque fichier à l'intérieur d'un dossier chiffré sera automatiquement chiffré lui aussi.

Dans l'état actuel, Hue n'est pas compatible avec la manipulation des données chiffrées. Par conséquent, il est nécessaire d'utiliser une solution de contournement pour envoyer un fichier dans un dossier chiffré du HDFS.

Tout d'abord, envoyez le fichier de données dans un dossier non chiffré via le navigateur de fichiers de Hue. Ensuite, il suffit de copier ce fichier vers un dossier chiffré en utilisant la ligne de commande suivante :

```
hdfs dfs -cp /monFichier /monDossier/monFichier
```

Le fichier est maintenant dans un dossier chiffré. Il est donc automatiquement chiffré également.

Ensuite, supprimer le fichier précédemment importé via Hue afin qu'il n'existe plus en clair dans le HDFS.

Lire des données

Les contenus chiffrés peuvent être parcourus et lus avec la ligne de commande `hdfs`. Celle-ci offre une grande partie des commandes native Linux habituelles. Ces lignes de commandes sont utilisables que les données soient chiffrées ou non.

En voici quelques exemples ci-dessous.

La commande suivante permet de lister le contenu d'un dossier :

```
hdfs dfs -ls /monDossier
```

La commande suivante permet lire le contenu d'un fichier :

```
hdfs dfs -cat /monDossier/monFichier.csv
```