

[Articles](#) [Contact](#) [Notre Solution](#) [À propos de nous](#)[Français](#)

Prise en main de Apache Ozone with CDP 7.1

Découverte et prise en main de Apache Ozone with CDP 7.1

Temps de lecture difficulté	Date de publication	Auteur
10 min facile	Jeudi 10 Décembre 2020	Lucas BAKALIAN

Cet article a pour but de présenter Apache Ozone à travers une prise en main pour voir comment fonctionne ce nouveau système de fichier, et en quoi il diffère de Apache HDFS, à l'aide de d'une première copie de données.

Introduction

Afin de présenter Apache Ozone, il faut d'abord rappeler ce qu'est Apache HDFS et comment Apache Ozone a été pensé pour combler les manques et limites d'Apache HDFS. Rappelons aussi qu'Apache Ozone faisait parti au début d'Apache Hadoop (HDFS + YARN), puis est devenu un projet Apache à part entière. Comme chaque administrateur de cluster HDFS a pu l'apercevoir, HDFS fonctionne de manière optimale seulement avec de gros fichiers, puisque c'est pour ce type de fichier que HDFS a été développé. HDFS peut aussi supporter un nombre maximale d'environ 300 Millions de fichiers.

Apache Ozone a été développé pour pouvoir supporter encore plus de fichiers, sans impacter ses performances, que ce soit des petits fichiers ou de gros fichiers.

Dans l'air du temps et des technologies de stockage cloud, Apache Ozone essaie de faire le lien entre les cluster Big Data traditionnels hébergeant aApche HDFS et les système de fichier de type Cloud.

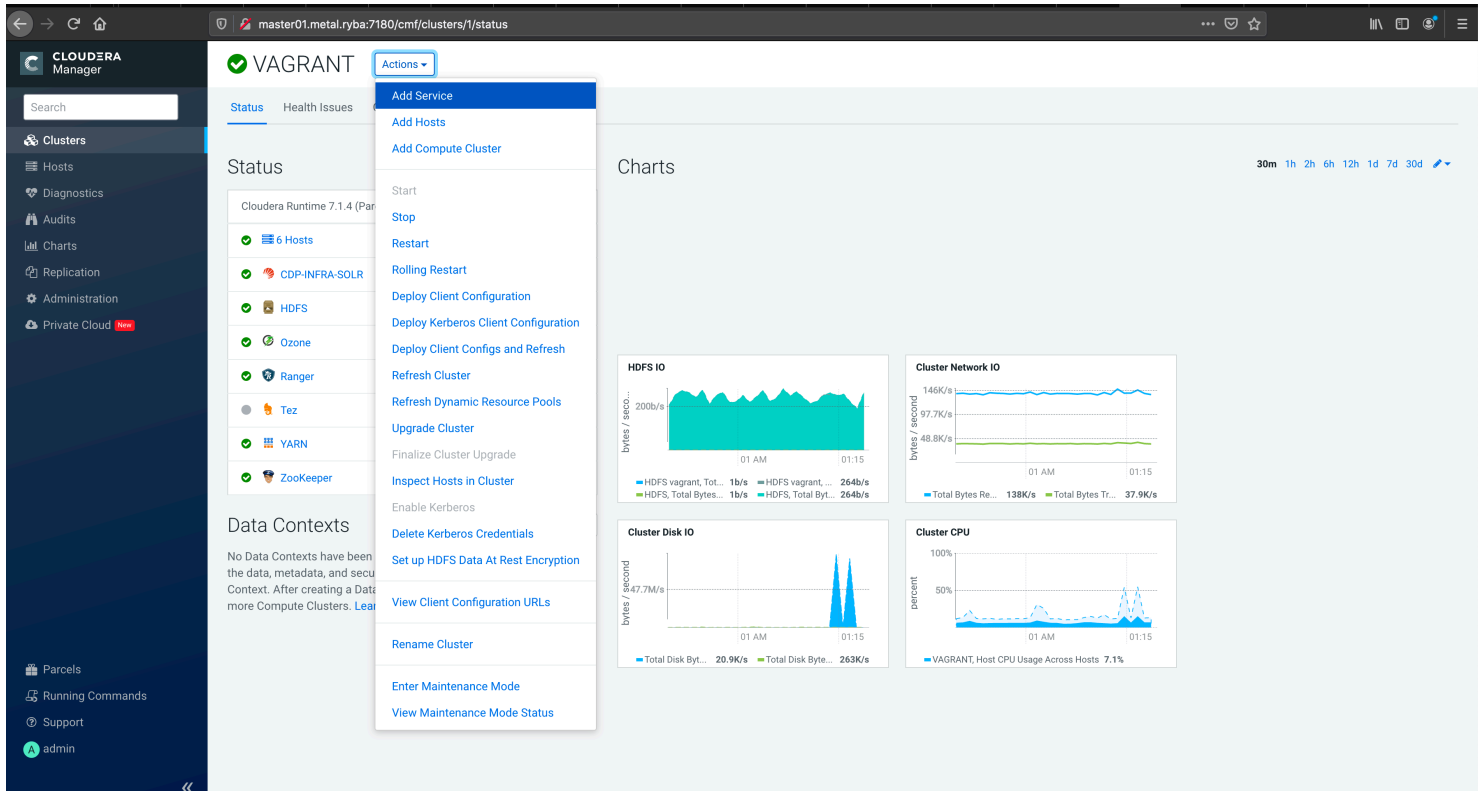
Afin d'écrire cet article, j'ai préparé un cluster virtuel constitué de:

- 3x serveur master and 3x serveur workers Centos7
- CDP Base en 7.1.4
- Apache HDFS configuré en HA
- Kerberos configuré avec un royaume FreeIPA

Architecture d'Apache Ozone et assignations des roles sur le cluster

Voyons comment installer Apache Ozone avec Cloudera CDP 7.1.4 et cloudera manager.

Depuis la Home Page de Cloudera Manager, cliquer sur le nom du cluster et ajouter un service.



Cliquer sur Ozone, puis une fois cliqué dessus vous arrivez sur la page des assignations de rôles. On tombe sur la liste des composants suivants:

- Storage Container View
- Ozone Manager
- Ozone Recon
- Ozone Datanode
- S3 Gateway
- Gateway

Il est intéressant d'utiliser l'assignation des composants, pour comprendre quel est le rôle de chaque composant et l'installer sur le bon serveur.

Depuis [la documentation d'Apache Ozone](#), on peut lire que Apache Ozone sépare la gestion des métadonnées concernant les volumes, de la gestion des métadonnées des fichiers (blocs physiques).

Listons

Ozone Manager: C'est le service qui contient les méta données du Namespace Ozone. Ozone Manager est responsable de la collection de volumes, buckets et opération sur les fichiers. En ce qui concerne la terminologie Apache Ozone, il existe trois différents types d'objets: volume, bucket et keys. Le volume Ozone est ce qui ressemble le plus au dossier racine d'Apache HDFS, à chaque volume peut être associé un ou plusieurs buckets. Et ce sont dans les buckets que les fichiers sont téléchargés. Un fichier Apache Ozone s'appelle aussi un bucket key ou key pour garder la forme anglo-saxonne. C'est cette répartition qui permet de rendre Apache Ozone plus scalable que Apache HDFS. On peut avoir plusieurs instances d'Apache Ozone. J'ai donc choisi d'installer un Ozone Manager par serveur master.

Ozone Storage Container Manager (SCM): Qu'est-ce que le SCM et pourquoi en avons-nous besoin ?

Dans Apache HDFS, toutes les informations des blocs constituant les fichiers sont stockées dans le Namenode HDFS, tandis que dans Apache Ozone, les blocs sont réunis de manière aléatoire dans ce qu'on appelle des containers:

- un container a une taille comprise entre 2Go et 15Go
- un container existe à l'intérieur d'un Datanode
- la réplication HDFS se fait au niveau des containers.

L'Ozone Manager donne les instructions au Storage Container Manager, et les requêtes clientes sont redirigées vers le Storage Container Manager.

La principale fonctionnalité du Storage Container Manager est de gérer le cycle de vie de la réplication des groupes de blocs organisés en container.

C'est cette gestion des blocs en container qui rend Apache Ozone beaucoup plus scalable que Apache HDFS. -> On peut donc assigner le rôle du SCM [Storage Container Manager] sur un rôle master.

NB: Pour l'instant la gestion de plusieurs SCM n'est pas gérée.

Ozone Datanode: Ce rôle est un peu plus subtil. Une Ozone Datanode peut s'exécuter à l'intérieur d'un datanode (HDFS) lui-même, ou s'exécuter en tant que