

Atelier 1 - Conception de cluster(s) Hadoop/HDFS

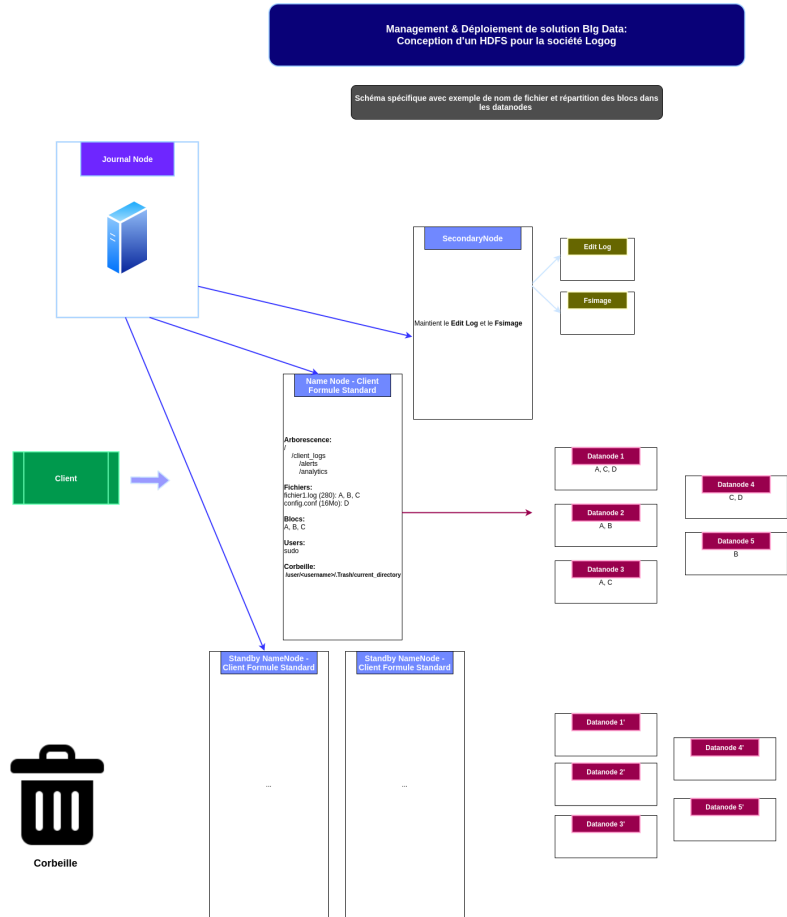
Acteurs

Nom
Jérémie LAERA
Arthur TOULOU MOND

Proposition de schémas & Justifications en notes

Nous proposons en premier lieu une schématisation spécifique pour un cluster Hadoop donnée, suivi de notes justifiant de nos choix architecturaux.

Enfin nous proposons la schématisation de 3 clusters en fonction de la formule choisie:



Notes

Ratio de réplication choisi: 3

Nombre de NameNodes impact de préférence afin de calculer la perte éventuelle (n - 1)/2.

Nous proposons dans ce cas 3 NameNodes, afin d'avoir au moins 1 NameNode opérationnel en cas de perte.

Nous proposons 3 illustrations de clustering en fonction de l'offre pour nos clients: un cluster Express, un cluster Standard et cluster Premium.

Base de taille des blocs: 64 Mo

Fichier log des clients, volumineux, qui justifient l'allocation de 3 blocs.

Fichier conf plus léger nécessitant 1 bloc.

Le FsImage contient les informations sur le namespace du HDFS depuis la création du NameNode principal.

Edit Log contient toutes les MAJs sur le namespace du FS en pointant sur la plus récente FsImage.

Le secondaryNode permet d'agir en tant que Helper pour le NameNode et maintient le FsImage ainsi que le Edit Log.

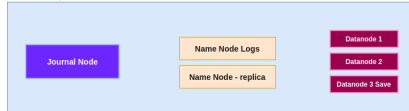
Il y a la prise en compte des logs et analyses pour les clients mais également les données concernant l'infrastructure de Logog

Le système de corbeille permet de référencer les anciens blocs éventuellement supprimés mais toujours répertoriés dans les NameNodes.

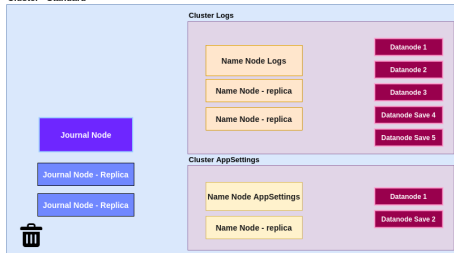
Par défaut, les fichiers supprimés seront stockés dans un répertoire de type: user/username/Trash/current_directory. Dans la formule, nous laissons et imposons le système de corbeille désactivé par défaut. Il faudra payer une formule standard afin d'y avoir accès.

Schémas génériques appliqués aux 3 formules pour les clients

Cluster - Express



Cluster - Standard



Cluster - Premium

