

TP1 – Introduction à Hadoop

Initialisation d'un cluster

2024-2025

M2 Informatique

UE Big Data

Responsable :

Jacob Michalski

1 Contexte

L'ensemble des TP se déroule dans la salle réseau, depuis laquelle vous avez accès aux serveurs dédiés disposant de Proxmox, un hyperviseur permettant la gestion d'environnements virtuels. Nous allons utiliser Hadoop 3.4.0 et Java 11 pour réaliser nos TP.

2 Configuration

Pour commencer, nous allons récupérer le template permettant de créer des VM pré-configurées. Pour cela, nous allons nous connecter à l'espace de stockage distant. Depuis Proxmox, sélectionnez Datacenter -> Storage -> NFS et ajoutez un nouveau support.

Les informations sont les suivantes :

ID : pedago-pve
server : nas-wks01.ad.univ-avignon.fr
export : /pedago-pve
content : VZDump backup file.

Une fois le template récupéré, nous pouvons créer des VM avec la bonne configuration initiale.

Créez une machine virtuelle et connectez vous dessus. Assurez vous que les variables d'environnement JAVA_HOME et HADOOP_PATH sont bien définies, sinon définissez les. JAVA_HOME doit pointer vers le dossier où est installé Java, par exemple `/usr/lib/jvm/java-11-openjdk-amd64` et apparaître dans le PATH (`export PATH=$PATH:$JAVA_HOME/bin`). HADOOP_PATH doit pointer vers le dossier où est installé Hadoop, dans notre cas `~/hadoop`

Vous pouvez tester que tout est en ordre en exécutant la commande

```
$HADOOP_PATH/bin/hadoop version
```

Vérifiez que vous pouvez vous connecter à localhost en SSH sans mot de passe. Si ce n'est pas le cas, ajoutez une clé SSH sans mot de passe.

```
ssh-keygen -t rsa  
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys  
chmod 0600 ~/.ssh/authorized_keys
```

Suivez les instruction du site suivant pour configurer le cluster :
<https://cdiese.fr/configurer-un-cluster-hadoop-multi-nud/>