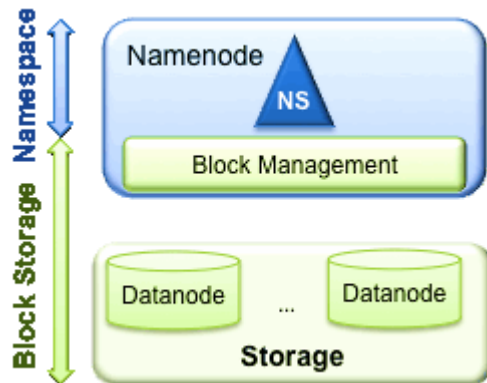


Atelier: HDFS

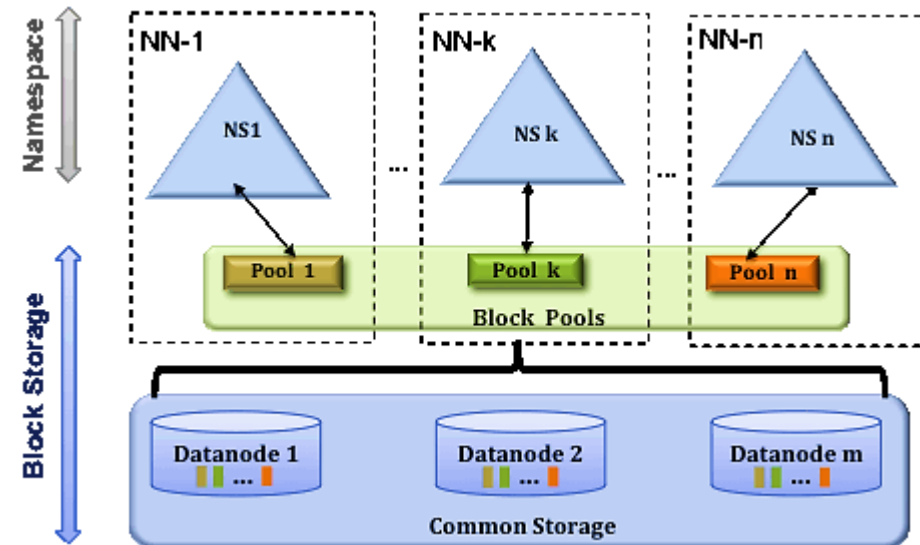
Pr. HIBBI Fatima-Zohra

f.hibbi@emsi.ma

Architecture de la fédération HDFS



Hadoop 1.0



Hadoop 2.0

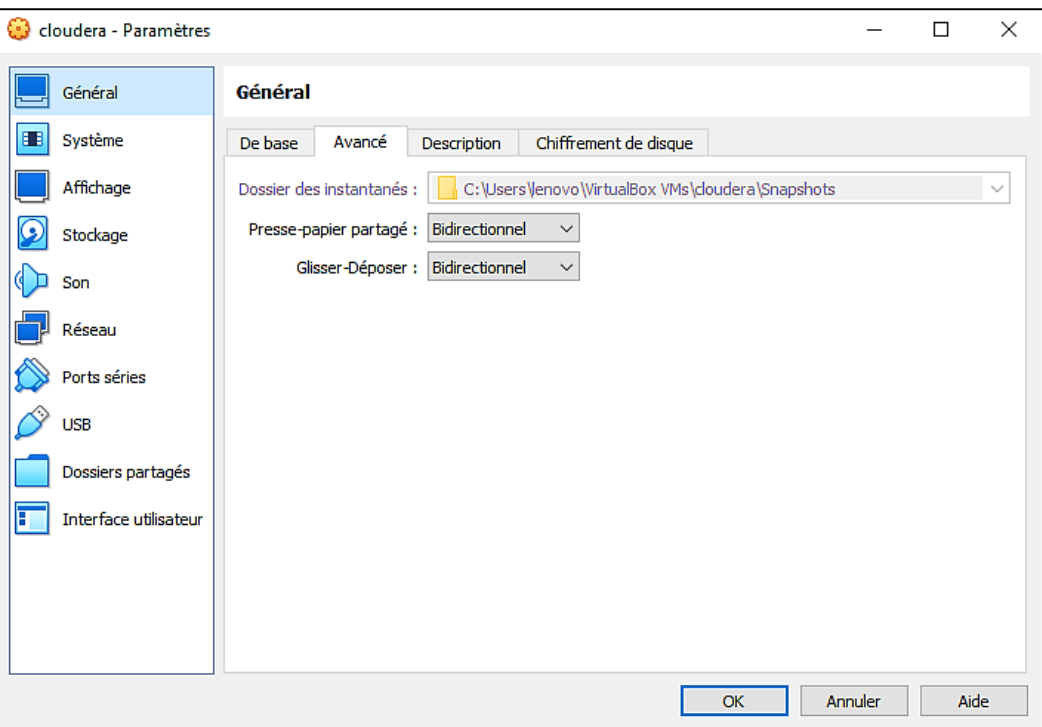
Hdfs fsck : states

AR: Actual Replication

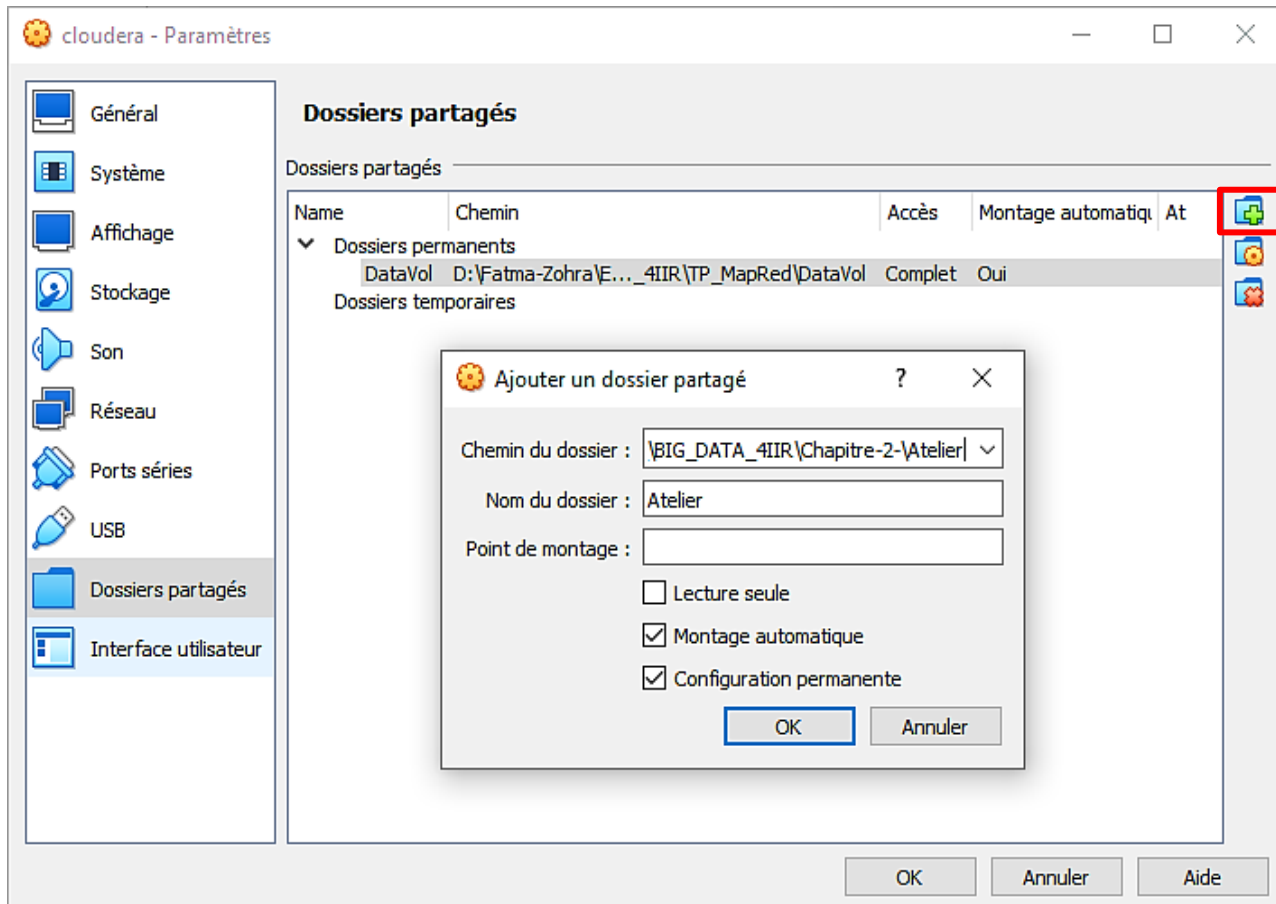
ER: Expected Replication

- § Over-replicated blocks: $AR > ER \rightarrow$ Suppression d'un block
- § Under-replicated blocks: $AR < ER \rightarrow$ Création d'un block
- § Mis-replicated blocks: Les blocs qui ne satisfont pas à la politique de placement des répliques de bloc
- § Corrupt blocks: Les blocs répliqué n'ayant pas les mêmes format (Hexadécimale)
- § Missing replicas: Ce sont des blocs sans répliques nulle part dans le cluster.

Configuration du dossier partagé



Configuration du dossier partagé



§ Pour créer un dossier taper la commande : `mkdir TPI`

§ Taper la commande :

`sudo mount -t vboxsf DataVol TPI`

§ DataVol est le nom du dossier partagé et le TP_I le nom du dossier créé

→ Vous allez remarquer qu'un raccourci vers le dossier partagé est ajouté sur le bureau de la virtuelle machine

Configuration du Hadoop et état de service

- § Dans un terminal, afficher le contenu du fichier : `cat /etc/hadoop/conf/hdfs-site.xml`
- § Quel est l'URL du processus serveur NameNode de HDFS ainsi que le port?
- § Quel est la valeur du facteur de réplication? `cat /etc/hadoop/conf/hdfs-site.xml`
- § Utiliser la commande `service nomservice status` pour vérifier l'état des services suivants:

`service hadoop-hdfs-namenode status`

`Service hadoop-hdfs-secondarynamenode status`

`Service hadoop-hdfs-datanode status`

- § Utiliser l'outil `jps` pour lister les processus Java en cours d'exécution: `$ sudo jps -l`