**Hadoop cluster implementation:**

**Après l’installation et la configuration de ‘hadoop cluster’ sur trois VM , premier est ‘namenode’ et les deux autres comme ‘datanode’ de ces deux références.**

**1-installation:**

<https://www.youtube.com/watch?v=cSTQs83GtnI>

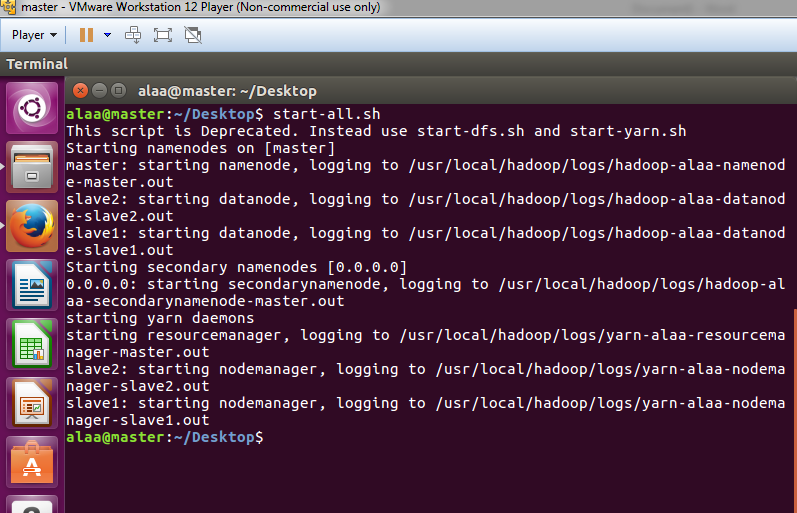
**2-configuration:**

<https://www.youtube.com/watch?v=MzdyM3N5SlE>

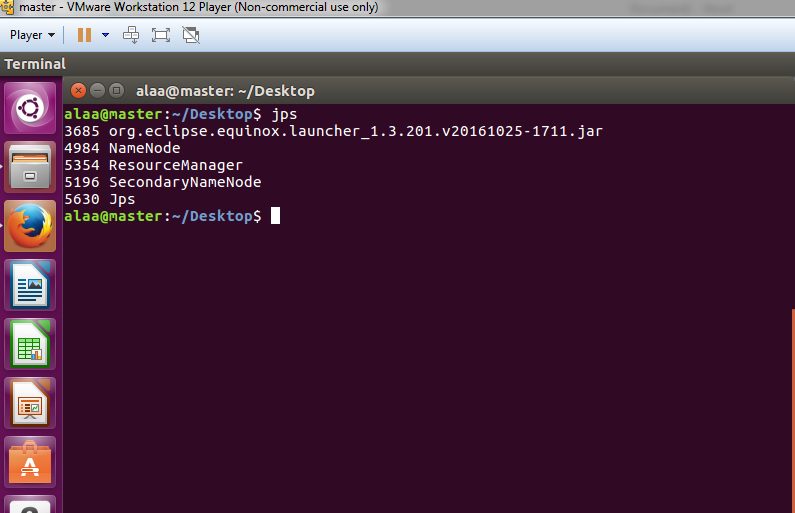
on peut demarrer notre node par le commande

**start-all.sh**

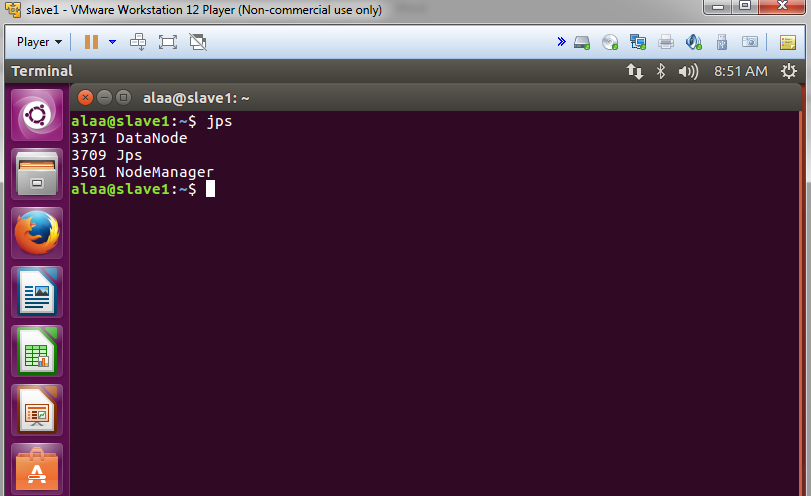
sur le master qui le namenode avec l’ip 192.168.0.1

****

Après on peut voir les processus qui sont en cours d’execution sur chaque machine comme suit:



C’est dans le master.



C’est dans le slave1 et slave2 qui jouent le travail de datanode1,datanode2 et ont les ip 192.168.0.2

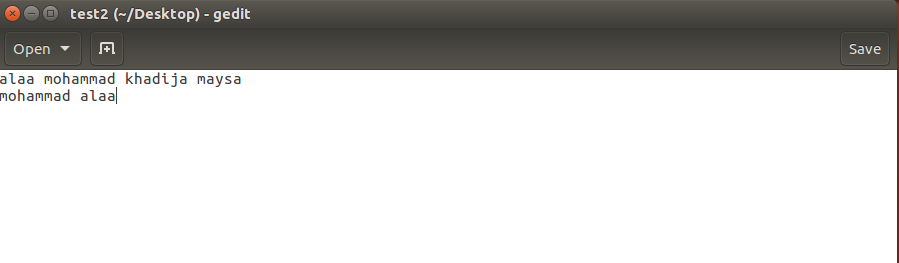
192.168.0.3

Ensuite on peut voir notre cluster.

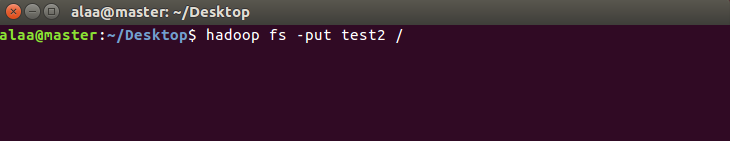


Qui est avec 2 live nodes et 0 dead nodes.

Après ça on va mettre un fichier dans le cluster pour faire un mapreduce qui travaille sur ce fichier.



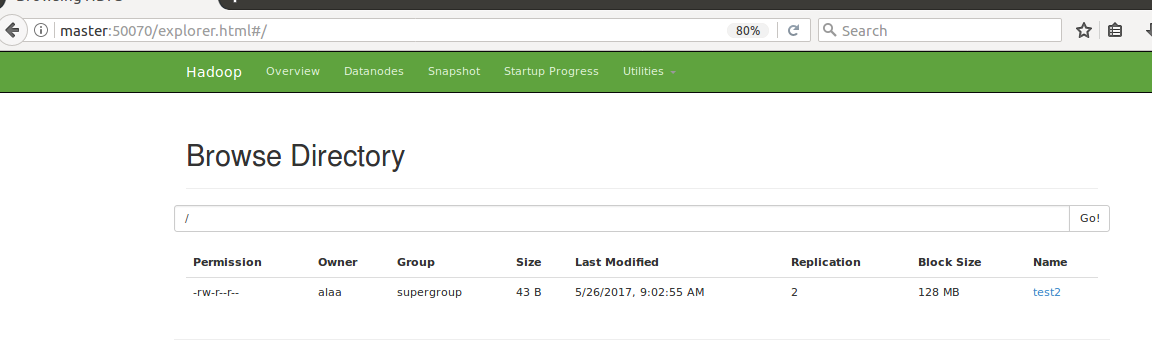
Par exemple en debut on a ce simple fichier on va le mettre dans le cluster puis faire un word count map reduce travail.



On met le fichier par la commande Hadoop fs –put.

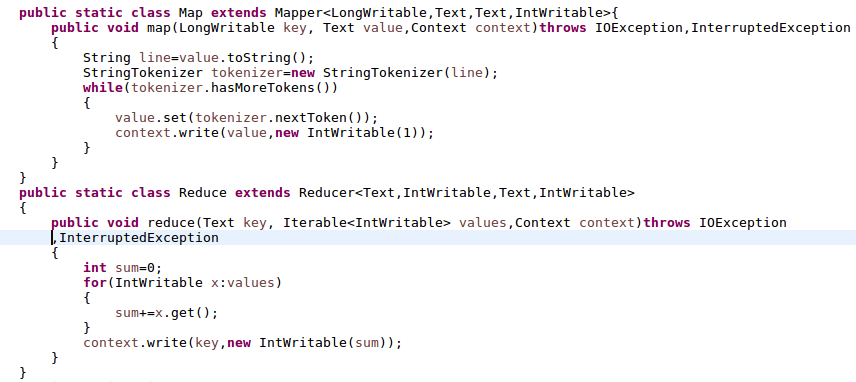
Le nom de fichier sur le desktop est test2.

on le met sous le root ‘/’.

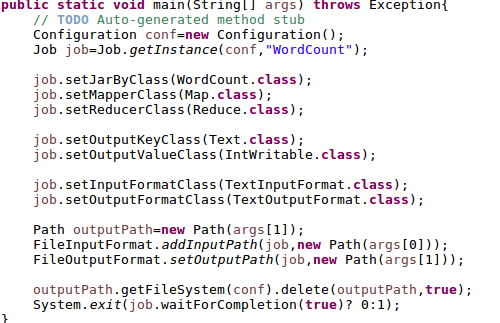


D’abord le fichier devient sur notre cluster avec 2 replications sur chaque datanode.

On veut maintenant écrire le code map reduce.



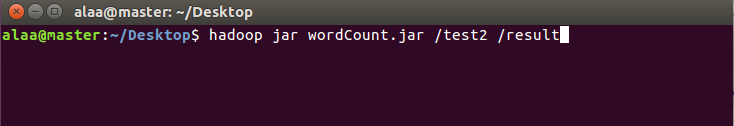
C’est le map reduce count .



C’est le main.

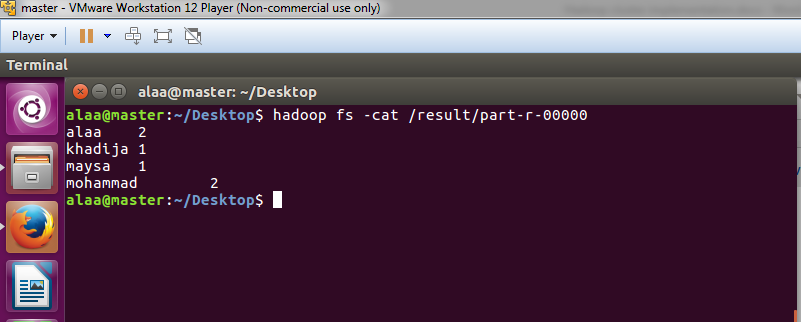
Ensuite on prend le jar de notre code.

Et on commence le mapreduce work sur notre fichier.



On veut le resultat dans un directory result.

Quand le travail se termine:



C’est le vrai resultat qu’on attend..

**on a testé ce travail sur un fichier de taille 130 mb qu’on a créé avec ce code..**

****

**Et on a vu le resultat du map reduce sur ce fichier après seulement 3 minutes.**

**Et on a fait un simple code pour faire le word count.. mais le resultat apparait après 35 minutes!**

**C’est le code..**

****

**Alors que map reduce travaille a donné le resultat plus rapide 10 fois qu’un simple code..**