environnement distribué : le traitement tourne en parallèle sur plusieurs machines

Hadoop : stockage + traitement

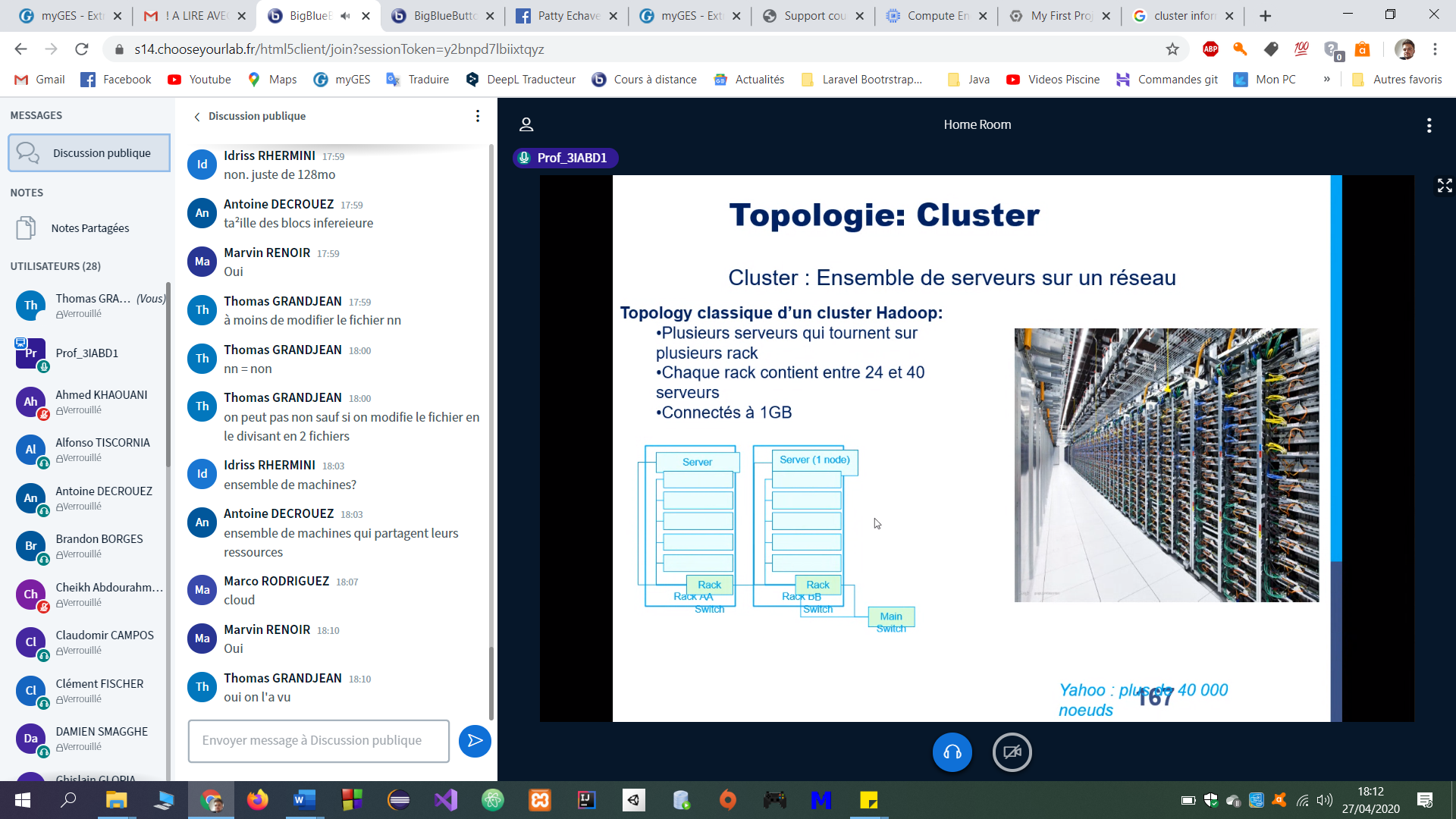
copie des données sur d’autres machines : quand on perd un noeud ce n’est pas un probleme car les données sont sur plusieurs machines

Hadoop : 3 composants : - HDFS (Stockage)

- Traitement (MapReduce)

- Gestion des ressources : Yarn

Yarn : qu’est ce que j’ai de disponible en ressources (CPU et GPU) et comment je les orchestre



1 rame : un serveur : stockage + ram + cpu

Name node : contient les informations sur tous les blocs (métadata) distribués en data nodes.