Bases NoSql

Hadoop

Machine Learning (IA), dont Deep Learning, recommendation, supervised Algo et no supervised Algo, reinforcing learning

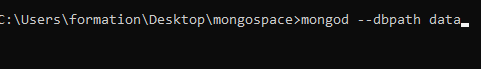
Moteur de recherche

ElasticSearch + Kibano + LogStach

CAP théorème : [clic](https://openclassrooms.com/fr/courses/4462426-maitrisez-les-bases-de-donnees-nosql/4462471-maitrisez-le-theoreme-de-cap)  
C : Consistence  
A : Availability  
P : Partition tolerance (distribution)

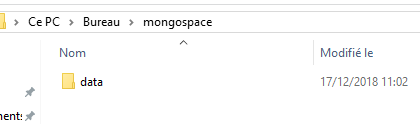
Il n'est pas possible d'avoir dans un système distribué les trois propriétés, deux au plus sont respectées.

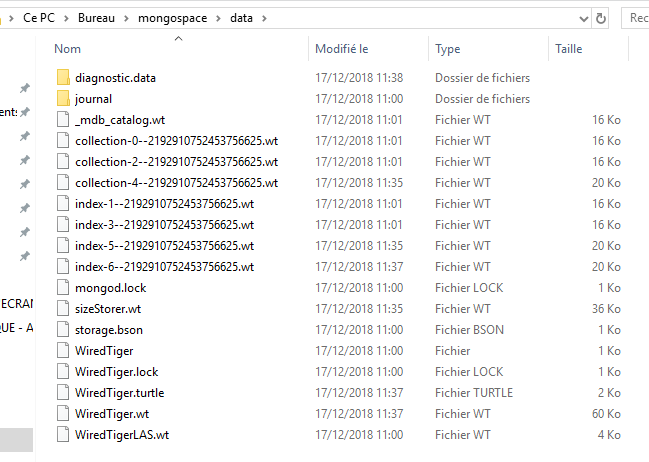
Le stockage est incomparablement moins cher désormais, les bases de données non relationnelles favorisent UN nœud de stockage et stockent en Json ou XML l'objet en intégralité.



* Va stocker dans le dossier data la future bdd.

On ajoute le path du bin dans les variables d'env avant évidemment, et on a créé le mongospace et le folder data dedans.



Le dossier s'est alors rempli :  


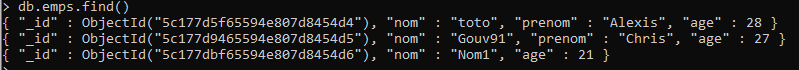
C:\Users\formation\Desktop\mongospace\data

On utilize ensuite db.emps.insert(monJson) :

On remplit la table emps directement !



On remplit quelques datas, un troisième sans prénom (schemaless !)

Fonction db.nomTable.find() montre les datas :  


GFS : Google File System  
HDFS  
Map Reduce pour récupérer les données (on ne peut SELECT \* )  
Namenode (node parent) qui donne des token pour accéder aux serveurs contenant la donnée,  
Les serveurs répliquent leurs données pour éviter les pannes.

Hive  
PIG  
Sgoop : exportation importation de données de HDFS vers une BDD transactionnelle.  
HBASE : une bdd noSql (comme Mongodb)  
Flume : connecteur Apache, permet de connecter les composants  
SPARK : repose directement sur HDFS, il peut faire des calculs sur la quantité de données que l'on veut, on peut l'utiliser sur Windows

Hive :facebook  
 Pig : Yahoo

Spark-shell : lancer spark  
pyspark : lancer spark en python

Quelques commandes (en spark-shell), dont inversion clef-val d'une map !  
