HADOOP :

La version libre arrivée en 2008 et supporte :

HDFS/ Hadoop map reduce et HBASE.

Hadoop permet de traiter bcp de donné en peu de temps exemple avec 1300 nœuds yahoo a traité 500GB en 59s.

map reduce n’est pas utile pour des tâche analytique et itérativement

Il utilise l’algo map-reduce afin de réduire son temps de calcule algo dev par google

HBASE :

Système de base de donnée non relationnel (C'est-à-dire une BD no-SQL(Donc en gros une bd où les élément n’ont pas de type))

HIVE :

Permet principalement pour nous donc de faire des requêtes SQL au sein de HBASE

AVANTAGE :

L’avantage d’un datalake est surtout sa facilité de stockage de données ainsi que sa rapidité de traitement. En effet le fait de pouvoir stocker des données semi structurés voir pas du tout structuré et surtout sans type particulier permet de les stocker beaucoup plus facilement.

Un accès à plus de données. Effectivement l'avantage d’avoir la possibilité de stocker plus facilement les données nous permet d’en avoir et d’en gérer plus. Sur un but commercial.

Pour le projet : Plus de docs sur les outils…

INCONVÉNIENT :

Un désordre rapide possible étant donnés que les données n’ont pas de structures et donc une exploitation de ces données peut s'avérer rapidement compliqué.

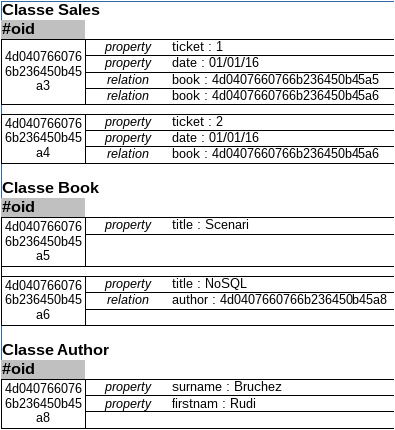
Ce qui fait leurs force (la rapidité) leurs fait aussi défaut effectivement l'absence de « sécurité » avec les règles peut poser des problèmes sur certaines base. Par exemple le fait de ne pas avoir de règles de transaction pour une base de paiement est impensable.

Un stockage de plus de données ce qui pose un soucis d’ordre physique peut amener une entreprise à délocaliser ses serveurs et donc exposé à un plus grande latence.

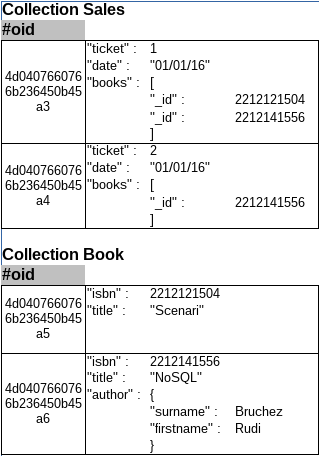
Exemple de no-sql:

JSON/XML donc soit en format graff/ colonne/document

graph:



Document:



Colone:

