HADOOP :

Hadoop est un système distribué, tolérant aux pannes, pour le stockage de données et qui est hautement scalable. Cette capacité de monter en charge est le résultat d'un stockage en cluster à haute bande passante et répliqué, connu sous l'acronyme de HDFS (Hadoop Distributed File System) et d'un traitement distribué spécifique et robuste connu sous le nom de MapReduce.

https://big-data.developpez.com/tutoriels/cloud/hadoop-debuter/

HBASE :

Une base de données adaptée aux gros volumes

https://juvenal-chokogoue.developpez.com/tutoriels/apprendre-travailler-hbase/

https://soat.developpez.com/tutoriels/bigdata/datalakes-architecture-big-data/

<https://hbase.apache.org/>

HIVE :

Hive a un rôle particulier, son origine reste de générer du map reduce à partir de SQL. La notion de vue externe vous permet de « faire comme si » vous disposiez d'une base au-dessus du HDFS. Évidemment, compte tenu des volumes, il va vous falloir partitionner intelligemment votre structure de répertoires.

https://juvenal-chokogoue.developpez.com/tutoriels/sql-dans-hadoop/

https://soat.developpez.com/tutoriels/bigdata/datalakes-architecture-big-data/

https://hive.apache.org/