

## ***Choix des technologies pour le stockage : File System Local***

Nous avons choisi d'utiliser un file system local pour le stockage des données dans notre Data Lake, car c'est l'approche la plus adaptée à notre contexte de projet académique et à nos besoins actuels. Voici les raisons principales de ce choix :

### **1. Simplicité et rapidité de mise en œuvre :**

Contrairement à des solutions comme HDFS ou Amazon S3, un file system local ne nécessite pas de configuration supplémentaire. Nous avons pu directement utiliser notre machine locale sans installer ni configurer des outils complexes, ce qui a permis de consacrer notre temps à l'implémentation des flux de données.

### **2. Adapté à notre volume de données :**

Étant donné que notre projet traite un volume limité de données (transactions clients, logs web, médias sociaux, flux publicitaires), un stockage distribué comme HDFS ou une infrastructure cloud comme Amazon S3 serait disproportionné. Un file system local offre des performances suffisantes pour nos besoins actuels.

### **3. Aucun coût et indépendance :**

Contrairement à Amazon S3, qui implique des coûts liés au cloud, ou à HDFS, qui demande des ressources matérielles supplémentaires, un file system local est gratuit et exploite les capacités de stockage de notre machine personnelle. Cela garantit un environnement de développement simple et sans dépendance à Internet.

### **4. Focus sur l'essentiel :**

Le choix d'un file system local nous permet de nous concentrer sur la conception logique et les traitements de données dans le Data Lake, plutôt que sur des problèmes d'infrastructure. Cela est particulièrement important dans un contexte pédagogique où la priorité est l'apprentissage.

Comparaison avec HDFS et Amazon S3 :

Critère	File System Local	HDFS	Amazon S3
Facilité d'utilisation	Très simple	Nécessite Hadoop	Configurations AWS
Coût	Gratuit	Gratuit	Coût du stockage
Adapté au projet	Oui	Idéal pour solutions à grande échelle	Idéal pour solutions à TRÈS grande échelle
Dépendances	Aucune	Besoin de Java, Hadoop	?

Pour ce projet, nous avons privilégié la simplicité et l'efficacité. En utilisant un file system local, nous avons pu rapidement mettre en place un Data Lake fonctionnel adapté à notre contexte.