

- ✓ **Sujet du projet :** Big Data.
- ✓ **Auteur :** Mehya MLAWEH & Sami FARHAT
- ✓ **Date :** Avril 2022

1. Introduction au projet :

Ce projet a pour objectif de comprendre le Big Data. Le Big Data est l'art de gérer de gros volumes de données, complexes et hétérogènes, pour la plupart non structurées, qui doivent circuler vite dans un système donné. Il présente **la théorie des 6 V**

(Volume/Variété/Vélocité/Véracité/variabilité /valeur).

Même si ce projet n'était pas notre premier choix, mais ce qui nous a attiré, il sert à :

- C'est notre domaine d'études tout d'abord.
- On veut plus connaître sur le big data.
- Big Data sert à réduire les couts/temps

2. Spécifications du projet :

a) Notions de base et contraintes du projet :

- La théorie des 3V :
 - **Volume :** Un système Big Data se caractérise d'abord par le volume de données en jeu.
Un système Big Data traite un volume de données largement supérieur à ce que traitent les bases de données traditionnelles.
 - **Vélocité :** Un système Big Data, c'est aussi un système dans lequel la donnée circule vite entre les outils, les bases, les applicatifs, les sources.
 - **Variété.** Les données sont en grand nombre et circulent vite dans le système. Mais ce n'est pas tout. Le Big Data se caractérise aussi par l'immense variété des données traitées. Les bases de données relationnelles ont affaire à des données structurées, bien définies, bien classées, bien normées.
 - **Valeur :** Ce V décrit la valeur qu'il est possible d'obtenir à partir des données et comment les mégadonnées obtiennent de meilleurs résultats à partir de données stockées.

- **Véracité** : La véracité permet de vérifier la qualité et l'origine de l'information.
 - **Variabilité** : définit la nécessité d'obtenir des données significatives en tenant compte de toutes les circonstances possibles.
- La protection des données.
 - La sécurité du stockage des données.

b) Fonctionnalités attendues du projet :

- Ajouter des données
- Filtrer les données
- Supprimer les données
- Analyser les données
- Prédiction.
- Classifier les données.
- ETL : Extraction de données/ Transformation des données/ Charger les données.