

---

## **SAE S4.C.01 - DEVELOPPEMENT AVEC UNE BASE DE DONNEES ET VISUALISATION**

---

Ce projet demande une maîtrise avancée des outils de gestion, analyse, filtrage et stockage de données.

**Comment se fait le travail ?** En binôme/trinôme.

**Composition du groupe :** libre.

### **DATES A RETENIR :**

- Vendredi 5 Avril, dépôt sur Moodle par un seul membre du groupe

### **DESCRIPTION DE LA SAE**

La problématique professionnelle est de créer une base de données à partir d'un ensemble de données complexe et incertain. Il faudra prendre en considération la qualité des données et les performances du modèle créé. Il sera également nécessaire de proposer des outils de visualisation des données. L'ensemble sera mis en œuvre au sein d'une équipe, dans une démarche de développement itérative ou incrémentale.

### **JEU DE DONNES : RESEAU DE CO-ACHAT DE PRODUITS SUR AMAZON**

#### **INFORMATIONS SUR LE JEU DE DONNEES :**

Le jeu de données présente un réseau de co-achat de produits collecté en effectuant un crawling sur le site d'Amazon. Il comprend des métadonnées de produits et des informations sur les critiques de 548 552 produits différents, tels que des livres, des CD de musique, des DVD et des cassettes VHS.

<https://snap.stanford.edu/data/amazon-meta.html>

Pour chaque produit, les informations suivantes sont disponibles :

- Titre
- Classement des ventes
- Liste de produits similaires (co-achetés avec le produit considéré)
- Catégorisation détaillée du produit
- Avis sur le produit : temps, client, notation, nombre de votes, nombre de personnes ayant trouvé l'avis utile

#### **STATISTIQUES DU JEU DE DONNEES :**

- Produits : 548 552

- Arêtes Produit-Produit : 1 788 725
- Avis : 7 781 990

Répartition des produits par groupe :

- Livres : 393 561
- DVD : 19 828
- CD de musique : 103 144
- Vidéos : 26 132

SOURCE : :

J. Leskovec, L. Adamic and B. Adamic. "The Dynamics of Viral Marketing." *ACM Transactions on the Web (ACM TWEB)*, 1(1), 2007.

REMARQUES :

Les données ont été collectées à l'été 2006, offrant ainsi une perspective historique de cette époque sur les habitudes d'achat et les critiques de produits sur la plateforme Amazon. Ce jeu de données offre une richesse d'informations pour explorer les dynamiques du marketing viral et les relations de co-achat entre les produits.

## TRAVAIL ATTENDU

INTRODUCTION :

Ce projet vise à exploiter le jeu de données Amazon Co-Achat pour créer une base de données relationnelle et développer une interface web permettant une consultation efficace des données stockées. Le processus inclura une analyse approfondie du dataset, l'extraction de données pertinentes, le nettoyage des informations et la modélisation en base de données relationnelle.

TACHES A REALISER :

1. Analyse du Dataset :  
Examiner le jeu de données Amazon Co-Achat pour comprendre sa structure et identifier les informations clés. Déterminer les entités, les relations, et les contraintes potentielles.
2. Modélisation de la Base de Données Relationnelle :  
Concevoir un schéma de base de données relationnelle en utilisant des entités, des relations, et des attributs pertinents. Appliquer les règles normales pour garantir l'intégrité des données.
3. Extraction et Nettoyage de Données :  
Extraire les données du data-set à l'aide d'une application développée dans le langage de votre choix et insérez-les dans le schéma réalisé. Effectuer, lorsque nécessaire, le nettoyage des données pour garantir la qualité et la cohérence des informations.
4. Modélisation de la Base de Données Multidimensionnelle :  
Concevoir un modèle multidimensionnel pour optimiser la représentation des données en vue d'une

analyse plus approfondie. Justifiez vos choix. Intégrez dans ce nouveau schéma les données précédemment extraites (les mêmes données doivent être incluses dans les deux schémas).

5. Requêtes :

Proposez au moins 20 requêtes (dont au moins 5 incluant des opérateurs OLAP), afin de démontrer l'efficacité de votre modèle multidimensionnel. Testez ces requêtes sur les deux modèles en fournissant les temps d'exécution respectifs. Justifiez vos réponses.

6. Développement de l'Interface Web :

Concevoir et créer une interface web permettant aux utilisateurs de consulter efficacement les données stockées dans la base de données multidimensionnelle. Intégrer des fonctionnalités de recherche, de filtre, et d'affichage des résultats.

## LIVRAISONS ATTENDUES :

- Les 2 scripts de création avec les commandes SQL pour créer vos bases de données.
- Un rapport détaillé décrivant
  - L'analyse du data-set, l'extraction et le nettoyage des données,
  - La modélisation de la base de données relationnelle
  - La modélisation de la base de données multidimensionnelle et les raisons de vos choix.
  - Les requêtes SQL et les temps d'exécution de chaque requête sur les 2 modèles. Motivez les réponses.
  - Documentation à propos de l'interface web (fonctionnalités proposées, technologies proposées, problèmes rencontrés).
- Le code source de l'interface web développée (partie facultative : hébergez quelque part votre site web pour le rendre opératif)

## MODALITES D'ÉVALUATION :

L'évaluation portera sur la qualité de l'analyse du dataset, la pertinence de la modélisation en base de données, la création réussie de la base de données, les requêtes et les explications données et la fonctionnalité de l'interface web.