

# Exploration de nouvelles technologies

---

## Projet 2

---

Durée : 3 semaines

Équipe : max 3 personnes

### Objectif

Récupérer des données disponibles sur Internet, puis les stocker dans une base de données **MongoDB Atlas**.

Ce projet vise à vous familiariser avec les technologies suivantes :

- **Webscraping** (extraction de données depuis un site web)
  - **Fichiers CSV** (format d'échange de données tabulaires)
  - **MongoDB** (base de données NoSQL dans le cloud)
  - **GitHub** (gestion de versions et collaboration)
- 

### Étapes du projet

1. Choisir un site web présentant des données pertinentes à extraire et informer votre enseignant du choix du site et des données (au plus tard à la 2e semaine).
2. Créer un dépôt **GitHub** pour héberger votre projet.
  - Inviter votre enseignant au dépôt.
  - Chaque membre de l'équipe doit contribuer (commits visibles).
3. Sélectionner un outil de webscraping.
4. Extraire un minimum de **2 500 éléments** (items) à l'aide de l'outil choisi (voir *Spécifications des données*).
5. Exporter les données dans un fichier **CSV**.
6. Créer une base de données dans **MongoDB Atlas** ([cloud.mongodb.com](https://cloud.mongodb.com)).
7. Importer le fichier CSV dans une collection de votre base de données.

8. Produire 3 **requêtes de sélection** avec la commande [find](#).

9. Rédiger un court **rapport** documentant vos démarches.

---

## Spécifications des données

Les données collectées doivent :

- contenir des champs permettant de **trier ou regrouper** les éléments (ex. : *catégorie*, *sous-catégorie*, etc.);
  - inclure des textes en **français avec accents**;
  - contenir des **nombres entiers et décimaux**;
  - (optionnel) si des images sont présentes, ne pas stocker les fichiers, mais uniquement les **URLs** pointant vers ces images.
- 

## Remise

Un fichier **ZIP** à déposer dans **Léa**, contenant :

1. **Un rapport (PDF)** comprenant :

- La description du site et des données choisies (incluant un URL cliquable).
- La présentation de l'outil de webscraping utilisé et de son fonctionnement.
- La méthode d'importation des données dans MongoDB Atlas.
- La description du schéma des documents dans votre collection (champs et types).
- Les 3 requêtes de sélection avec explication.
- Un lien vers votre base MongoDB Atlas (voir Notes).
- Un lien vers votre dépôt **GitHub**.
- Une conclusion avec vos impressions sur les technologies utilisées.
- Un tableau indiquant la contribution de chaque coéquipier (tâches réalisées et estimation du pourcentage de participation).

2. **Le fichier CSV** ayant servi à l'importation.

---

## Notes

1. Le port **27017** est bloqué par le cégep. Vous ne pouvez donc pas vous connecter à MongoDB Atlas via des outils comme *Compass* ou *mongoimport.exe* depuis le réseau du cégep. L'accès doit se faire uniquement via l'**interface web**.
  2. Pour permettre à l'enseignant d'accéder à votre base de données, vous devez l'inviter avec l'adresse : [nicolas.payre@cegepsherbrooke.qc.ca](mailto:nicolas.payre@cegepsherbrooke.qc.ca), en lui attribuant les droits suivants :
    - *Project Data Access Read Only*
    - *Project Read Only*
  3. Tous les membres de l'équipe doivent avoir effectué des **commits** dans le dépôt GitHub pour démontrer leur participation.
- 

## Évaluation et critères de correction (100 %)

### 1. Qualité du scraping et des données collectées (25 %)

- Respect du minimum de 2 500 items.
- Données cohérentes, complètes et exploitables.
- Respect des spécifications (accents, nombres, attributs communs, etc.).

### 2. Fichier CSV et importation dans MongoDB (20 %)

- Structure claire et bien formatée du fichier CSV.
- Importation réussie et correcte dans MongoDB Atlas.

### 3. Requêtes MongoDB (15 %)

- Trois requêtes pertinentes et fonctionnelles.
- Capacité à filtrer, trier ou regrouper les données.

### 4. Rapport (20 %)

- Clarté et précision des explications.
- Documentation complète (site, outil, importation, schéma, requêtes).
- Conclusion pertinente et réflexive.

### 5. GitHub et collaboration (20 %)

- Dépôt bien structuré et accessible (enseignant invité).
- Lien valide remis dans le rapport.
- Contributions visibles de tous les membres (commits).

