



Objectifs pédagogiques

- Récupération de données structurées et non-structurées
- Intégration de données dans une base de données

Démarche pédagogique (projet, ressources ...)

Ce module se décompose en trois parties introduisant des méthodes pour aller récupérer des données et les intégrer dans une base de données.

Les deux premières parties sont consacrées au scraping et à l'utilisation d'apis, la troisième à la structuration des données à l'aide d'ETL python.

Introduction : Plongée dans l'Univers de l'Extraction de Données

Imaginez un immense océan d'informations, un flot ininterrompu de données circulant sur internet. Chaque clic, chaque interaction génère des millions de données exploitables. Mais comment peut-on transformer ce flux en informations utiles et exploitables ? C'est là que l'extraction de données entre en jeu.

Autrefois, les analystes devaient manuellement collecter des informations, feuilleter des pages de rapports imprimés ou copier-coller des données depuis des sites web. Aujourd'hui, grâce aux avancées technologiques, il est possible d'automatiser cette collecte en utilisant le web scraping, les APIs et les pipelines ETL.

Ce cours va vous guider pas à pas dans la découverte et la mise en œuvre de ces techniques. À la fin, vous serez capable de construire vos propres pipelines d'extraction et de transformation des données, prêts à être utilisés dans des analyses, des dashboards ou même pour alimenter des modèles d'intelligence artificielle.



La Région
Auvergne-Rhône-Alpes

1. Pourquoi extraire des données ?

L'extraction de données est un outil puissant qui trouve des applications dans de nombreux domaines :

Cas d'usage concrets

- **Analyse de marché** : Imaginez un e-commerçant souhaitant surveiller ses concurrents. Grâce au scraping, il peut récupérer automatiquement les prix des produits concurrents et ajuster ses prix en conséquence.
- **Veille informationnelle** : Un journaliste peut extraire les dernières actualités de plusieurs sites pour générer un résumé automatisé des tendances.
- **Recherche scientifique** : Un chercheur en santé publique peut extraire des articles scientifiques et des statistiques depuis différentes sources académiques pour ses analyses.
- **Finance** : Les traders utilisent des APIs financières pour suivre les fluctuations du marché en temps réel et adapter leur stratégie d'investissement.
- **Automatisation des tâches** : Un professionnel du marketing peut récupérer les commentaires laissés sur les réseaux sociaux pour analyser les avis des clients sur un produit.

2. Comment extraire des données ?

L'extraction de données repose sur trois grandes méthodes :

1. **Le Web Scraping** : L'automatisation de la récupération des données directement depuis des pages web.
2. **Les APIs** : Des interfaces fournies par des services en ligne permettant de récupérer des données sous forme structurée.
3. **Les pipelines ETL** : Une méthode combinant extraction, transformation et chargement des données pour une exploitation en base de données ou fichiers structurés.

Compétences

- Scraping :
 - Récupération de données avec BeautifulSoup
- API
 - Utiliser une API avec des paramètres
- ETL
 - Introduction Extract Transform Load et Extract Load Transform
 - Opération ETL qui reprend API + Scraping + SQL

MODALITÉS

Durée

2 jours soit 14 heures au total.

Lancement le 22/12/2025 et clotûre le 23/12/2025.

Formateur(s)

David NADJAR

TRAME

		Planning		Jour	Sujet	Activités
22/12	Extraction de données	David	David	1	Scraping + API	
23/12	Extraction de données	David	David	2	API + ETL	