

Système d'intégration de données pour l'optimisation de la recherche d'emploi

Malek JAAFAR, Sarah LAKHZOUM, Malik TOUATI

April 17, 2025

Application

Le but de cette application est de faciliter la recherche d'emploi suivant des critères précis tels que le réseau de transport ou bien les événements de networking.

Sources de données

- Les données ont été récupérées via deux méthodes : des APIs ou du scraping.
- Les APIs d'Indeed et de France Travail ont permis de récupérer les données d'offres et d'entreprises au format `.json`, traitées avec la bibliothèque `json` de Python.
- Les données du réseau de transport ont été récupérées sur le site de la RATP au format `.csv`.
- Les événements de networking proviennent d'un scraping du site EventBrite (catégorie "Business", Paris), stockés d'abord en `.har` puis convertis en `.csv` et traités avec `pandas`.

Tables extraites

- **offres**(id, intitule, description, lieu_travail_libelle, entreprise_nom, type_contrat, experience_exige, salaire_libelle, secteur)
- **entreprise**(id, nom, adresse, codepostale, commune, section, lat, long)
- **networking**(id, nom, timezone, url, date_debut, categorie)
- **transports**(station, traffic, correspondance1, correspondance2, correspondance3, correspondance4, correspondance5, ville)

Vues SQL créées

- **offres_view** : enrichit les offres avec les URLs des entreprises.
- **offres_networking_view** : relie les offres aux événements du même secteur.
- **offres_transport_view** : associe les offres à des stations selon leur localisation.
- **recrutement_networking_view** : détecte les entreprises recrutant lors d'événements.
- **sector_networking_view** : analyse les secteurs qui recrutent et les événements associés.
- **recent_networking_offres_view** : affiche les offres liées à des événements récents (30 jours).

- **company_sector_city_view** : calcule le nombre de villes et secteurs pour chaque entreprise.
- **formation_events_view** : affiche les offres liées à des événements de formation.

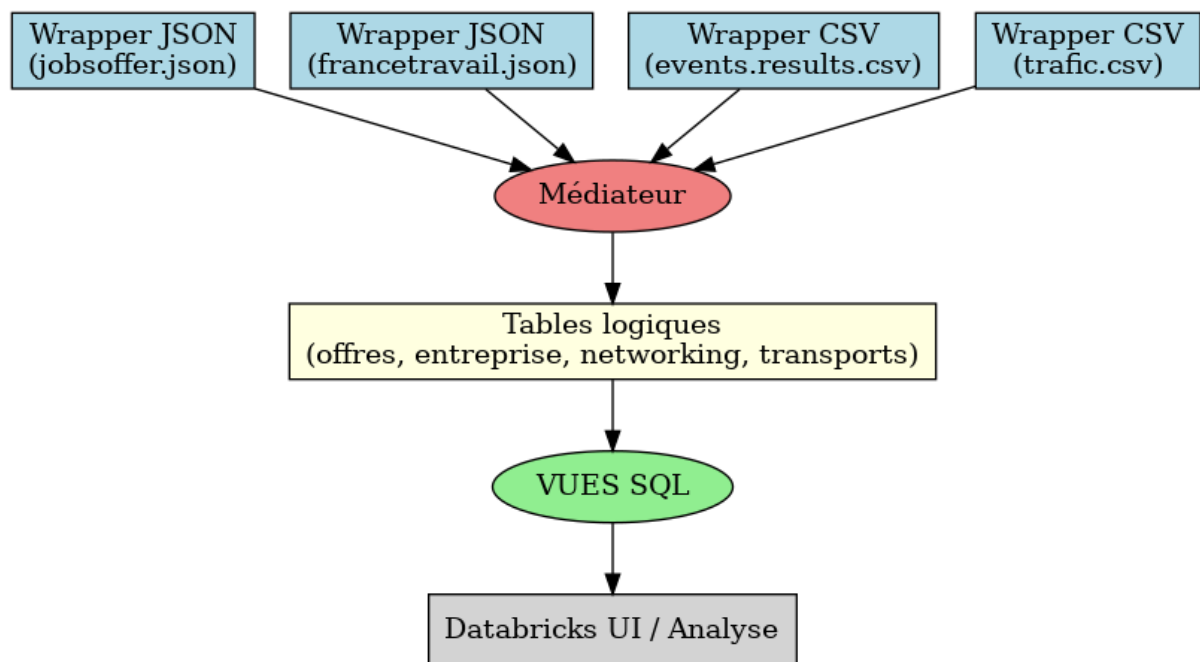
Triggers et mise à jour

Une table `trigger_log` enregistre les actions sur la table `offres`. Une fonction Python vérifie ce journal pour rafraîchir automatiquement les vues après chaque insertion, mise à jour ou suppression.

Scénarios d'utilisation

1. Recommander des événements pertinents aux chercheurs d'emploi selon leur secteur.
2. Évaluer l'accessibilité d'un poste via les transports en commun.
3. Identifier les entreprises qui recrutent dans plusieurs villes ou secteurs.
4. Repérer les secteurs en tension.
5. Trouver des événements de formation liés à certaines entreprises.

Schéma de l'application (Architecture)



APIs / Sites utilisés

- www.eventbrite.com
- data.ratp.fr
- Apify - Indeed Actor
- France Travail API - Offres d'emploi