

Compte-rendu de réunion : Conteneurisation de l'infrastructure microservices JobMarket

Date : 28 MAI 2025 **Participants** : Arnaud, Ilan, Sambou, Mickael **Sujet** : Comment organiser notre application avec des conteneurs Docker

1. Pourquoi utiliser Docker ?

Notre projet JobMarket utilise plusieurs outils différents (Python, bases de données, machine learning, etc.). Chaque outil a ses propres besoins d'installation et de configuration.

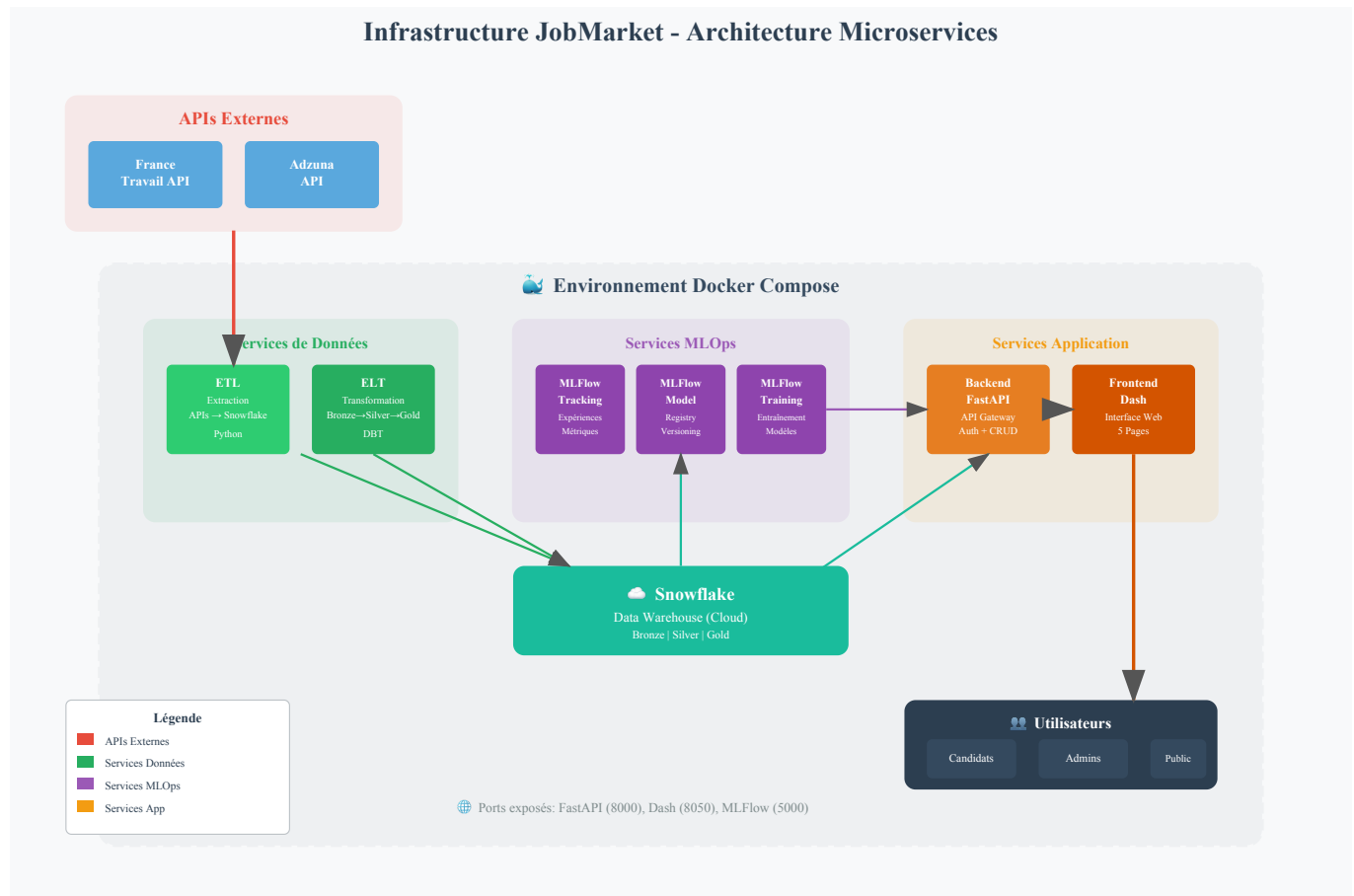
Avec Docker, on peut :

- Mettre chaque outil dans sa propre "boîte" (conteneur)
- Éviter les conflits entre les outils
- Installer le projet facilement sur n'importe quel ordinateur
- Travailler tous avec le même environnement

2. Les outils qu'on utilise

- **Snowflake** : Notre base de données principale (dans le cloud)
- **Python + DBT** : Pour nettoyer et préparer nos données
- **MLFlow** : Pour gérer notre intelligence artificielle
- **FastAPI** : Pour créer notre API (le lien entre le frontend et nos données)
- **Dash** : Pour créer notre site web

Schéma de l'architecture



3. Comment on a organisé nos conteneurs

3.1 Les conteneurs pour les données

Conteneur ETL (Extraction des données)

- **Ce qu'il fait :** Il va chercher les offres d'emploi sur les sites France Travail et Adzuna
- **Comment ça marche :**
 - Un script Python se connecte aux APIs de ces sites
 - Il récupère les nouvelles offres d'emploi
 - Il les envoie dans notre base Snowflake
- **Quand ça tourne :** Tous les jours automatiquement

Conteneur ELT (Transformation des données)

- **Ce qu'il fait :** Il nettoie et organise les données brutes qu'on a récupérées
- **Comment ça marche :**
 - DBT prend les données "sales" de Snowflake
 - Il les transforme en données propres et utilisables
 - Il suit la méthode Bronze → Silver → Gold (de plus en plus propre)
- **Avantage :** Si on fait une erreur, on peut revenir en arrière facilement

3.2 Les conteneurs pour l'intelligence artificielle

On a séparé MLFlow en 3 conteneurs différents pour mieux s'organiser :

Conteneur MLFlow Tracking

- **Ce qu'il fait** : Il garde en mémoire tous nos tests de machine learning
- **Contenu** : Les résultats, les paramètres, les performances de chaque test
- **Interface** : On peut voir tout ça sur une page web

Conteneur MLFlow Model

- **Ce qu'il fait** : Il stocke nos modèles d'IA qui marchent bien
- **Contenu** : Les différentes versions de nos modèles (test, production)
- **Utilité** : Notre API peut utiliser ces modèles pour faire des recommandations

Conteneur MLFlow Training

- **Ce qu'il fait** : Il entraîne nos modèles d'intelligence artificielle
- **Quand ça tourne** : Quand on veut créer un nouveau modèle ou l'améliorer
- **Ressources** : Il utilise beaucoup de calcul, donc on lui donne plus de puissance

mlflow2.22.0ExperimentsModelsPrompts

matching_offres_candidats_grid_search>

test_pipelineModel registered

OverviewModel metricsSystem metricsTracesArtifacts

DescriptionNo description

Details

Created at	06/02/2025, 09:37:31 PM
Created by	mluser
Experiment ID	3
Status	Finished
Run ID	f7b37af12dac445a80a9507087fab170
Duration	26.2s
Datasets used	—
Tags	Add tags
Source	train_model.py
Logged models	sklearn
Registered models	jobmarket v2 +1
Registered prompts	—

Parameters (8)

Parameter	Value
model_name	paraphrase-multilingual-MiniLM-L12-v2
batch_size	32
n_neighbors	3
threshold	0.75
max_length	512
device	cpu
num_workers	4
random_state	42

Metrics (15)

Metric	Value
System metrics (8)	
system/cpu_utilization_percentage	14.3
system/system_memory_usage_m...	2765
system/system_memory_usage_pe...	33.6
system/disk_usage_percentage	42.8
system/disk_usage_megabytes	25433.4
system/disk_available_megabytes	34020.9
system/network_receive_megabytes	0.036496999999997115
system/network_transmit_megabyt...	0.0126469999999999409
Model metrics (7)	
similarity_score	0.3156909096986056
mean_similarity	0.5047434568405151
diversity_score	0.004586935043334961

mlflow2.22.0ExperimentsModelsPrompts

Registered Models> MatchingPipeline>

Version 19Promote model

Registered At: 06/05/2025, 04:37:42 PMLast Modified: 06/05/2025, 04:37:42 PMSource Run: test_pipeline

Aliases: AddStage (deprecated): None

New model registry UI

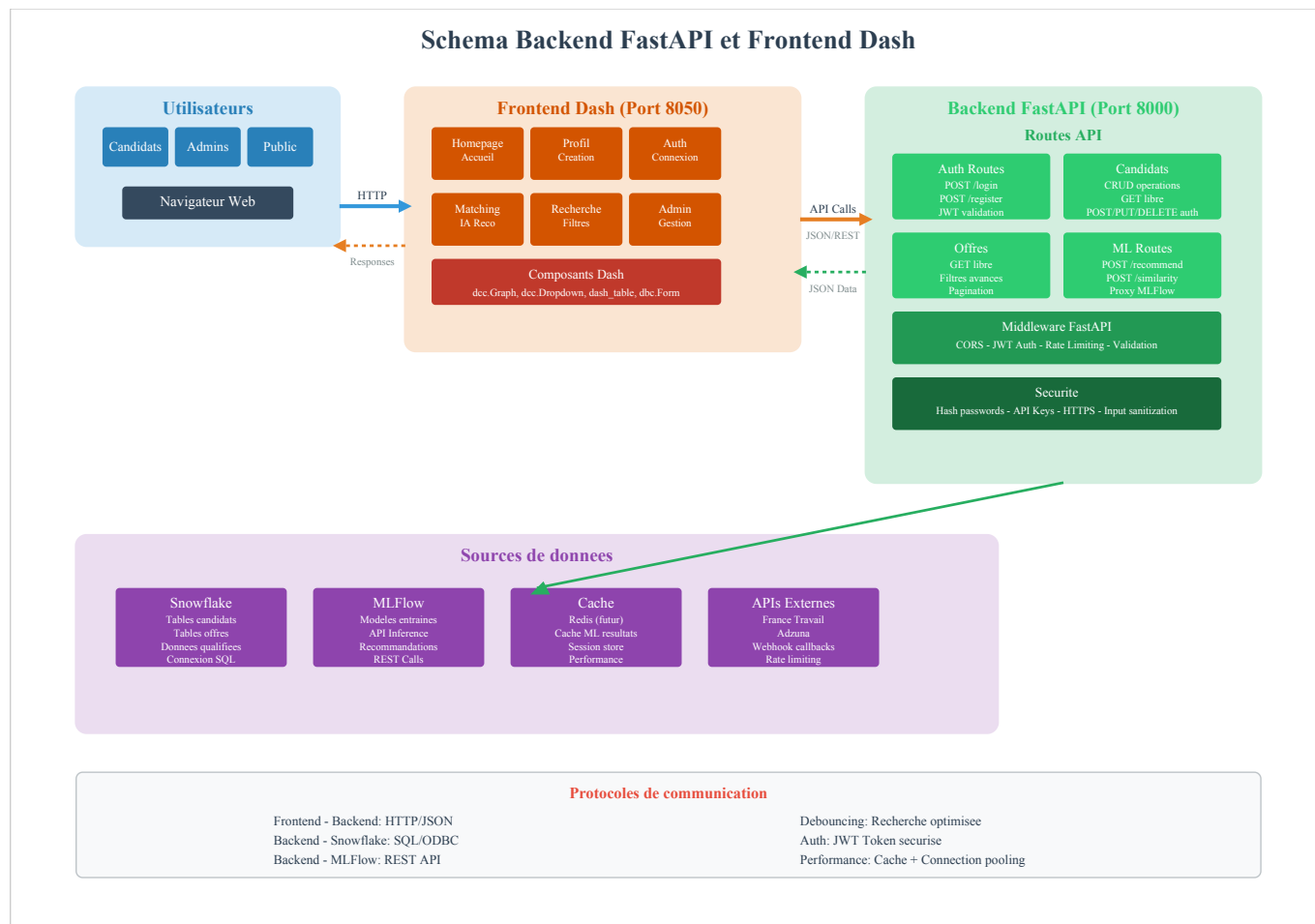
> Description Edit

> Tags

> Schema

Name	Type
Inputs (1)	
TEXT (required)	string
Outputs (2)	
match_index (required)	double
similarity_score (required)	double

3.3 Les conteneurs pour l'application



Conteneur Backend (FastAPI)

- **Ce qu'il fait :** C'est le cerveau de notre application
- **Fonctions principales :**
 - Gérer les comptes utilisateurs (création, connexion)
 - Récupérer les données de la base
 - Utiliser l'IA pour faire des recommandations
 - Vérifier que les utilisateurs ont le droit de faire certaines actions
- **Sécurité :**
 - Seules les personnes connectées peuvent modifier des données
 - Tout le monde peut consulter les offres d'emploi

Modèles			Accès au version du modèle	^
Offres d'emploi				^
GET	/api/offres	Récupérer la liste des offres d'emploi		▼
POST	/api/offres	Créer une nouvelle offre d'emploi		▼
GET	/api/offres/{id}	Récupérer une offre d'emploi par son ID		▼
PUT	/api/offres/{id}	Mettre à jour une offre d'emploi		▼
PATCH	/api/offres/{id}	Mettre à jour partiellement une offre d'emploi		▼
DELETE	/api/offres/{id}	Supprimer une offre d'emploi		▼
POST	/api/offres/search	Recherche avancée d'offres d'emploi		▼
Modèles			Accès au version du modèle	^
Offres d'emploi				^
GET	/api/offres	Récupérer la liste des offres d'emploi		▼
POST	/api/offres	Créer une nouvelle offre d'emploi		▼
GET	/api/offres/{id}	Récupérer une offre d'emploi par son ID		▼
PUT	/api/offres/{id}	Mettre à jour une offre d'emploi		▼
PATCH	/api/offres/{id}	Mettre à jour partiellement une offre d'emploi		▼
DELETE	/api/offres/{id}	Supprimer une offre d'emploi		▼
POST	/api/offres/search	Recherche avancée d'offres d'emploi		▼

Conteneur Frontend (Dash)

- **Ce qu'il fait** : C'est le site web que voient les utilisateurs
- **Pages créées** :
 - Page d'accueil avec présentation
 - Page pour créer un profil candidat
 - Page de connexion
 - Page de recommandations personnalisées d'offres
 - Page de recherche libre avec filtres
- **Communication** : Il parle avec le backend pour récupérer les données

👤 Rechercher une offre

Recherche d'offre :

devops

Ville :

France

Type de contrat :

×

CDI

×

Séniorité :

×

Confirmé

×

Domaine :

Sélectionnez un domaine

Rechercher

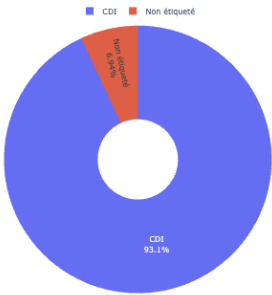
72 offre(s) trouvée(s)

Titre	Entreprise	Ville	Type de contrat	Domaine	Séniorité	Compétences
filter data...						
Développement MISTRAL – Ingénieur validation confirmé H/F (H/F)	SNCF D5	Saint-Denis	CDI	Personnel sédentaire du transport ferroviaire et réseau filo guidé	Confirmé	
INGENIEUR DEVOPS CONFIRME (H/F)	ADSEARCH	Anet	CDI	Conception, recherche, études et développement	Confirmé	DevOps
TECH LEADER (H/F)	TRIANGLE 67	Aix-en-Provence	CDI	Conception, recherche, études et développement	Confirmé	
CDI – Chef de Projet Innovation IT (H/F)	MERITIS	Paris			Confirmé	Chef
Chef de projet it transformation digitale (h/f) (CDI)	MERITIS	Paris			Confirmé	Chef
CDI – Chef de Projet Amélioration Continue & Qualité H/F	MERITIS	Paris		Banque	Confirmé	Chef
filter data...						
Développement MISTRAL – Ingénieur validation confirmé H/F (H/F)	SNCF D5	Saint-Denis	CDI	Personnel sédentaire du transport ferroviaire et réseau filo guidé	Confirmé	
INGENIEUR DEVOPS CONFIRME (H/F)	ADSEARCH	Anet	CDI	Conception, recherche, études et développement	Confirmé	DevOps
TECH LEADER (H/F)	TRIANGLE 67	Aix-en-Provence	CDI	Conception, recherche, études et développement	Confirmé	
CDI – Chef de Projet Innovation IT (H/F)	MERITIS	Paris			Confirmé	Chef
Chef de projet it transformation digitale (h/f) (CDI)	MERITIS	Paris			Confirmé	Chef
CDI – Chef de Projet Amélioration Continue & Qualité H/F	MERITIS	Paris		Banque	Confirmé	Chef
Chef de Projet Angular PHP H/F (CDI)	MERITIS	Paris			Confirmé	Chef
Consultant Dataiku – Junior / Confirmé / Senior H/F (CDI)	NEXTEP HR 1	Paris	CDI	Développement territorial et emploi	Senior	
DevOps Confirmé F/H	ITRUST	Toulouse	CDI		Confirmé	DevOps
Ingénieur de développement logiciel JAVA (H/F)		Aix-en-Provence	CDI	Conception, recherche, études et développement	Confirmé	Java, Linux

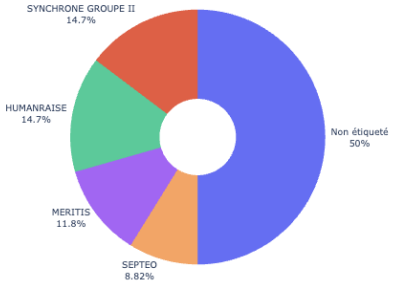
1 / 8

📊 Analyse des résultats

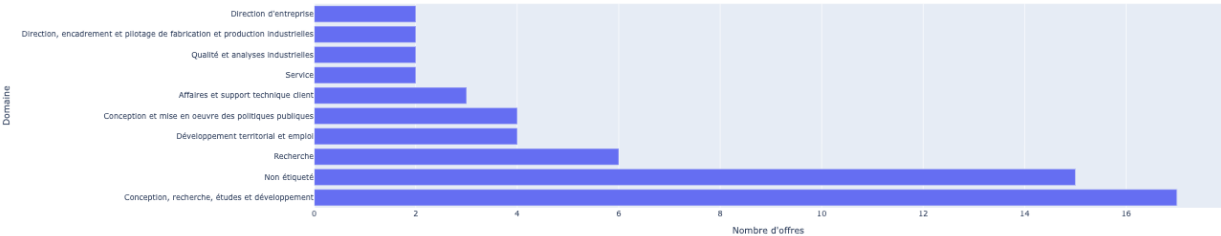
Répartition des offres étiquetées par type de contrat



Répartition du top 5 des entreprises qui recrutent le plus



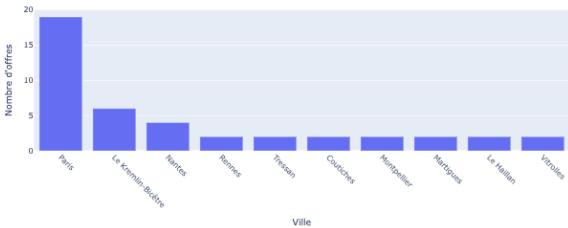
Top 10 des domaines les plus demandés



Répartition des offres étiquetées par niveau de séniorité

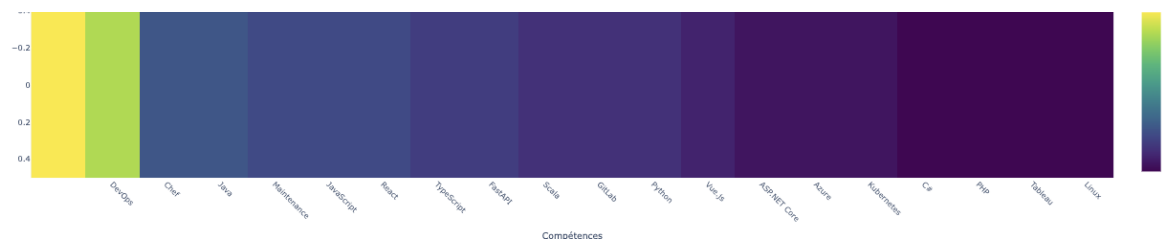


Top 10 des villes avec le plus d'offres étiquetées



Top 20 des compétences les plus demandées

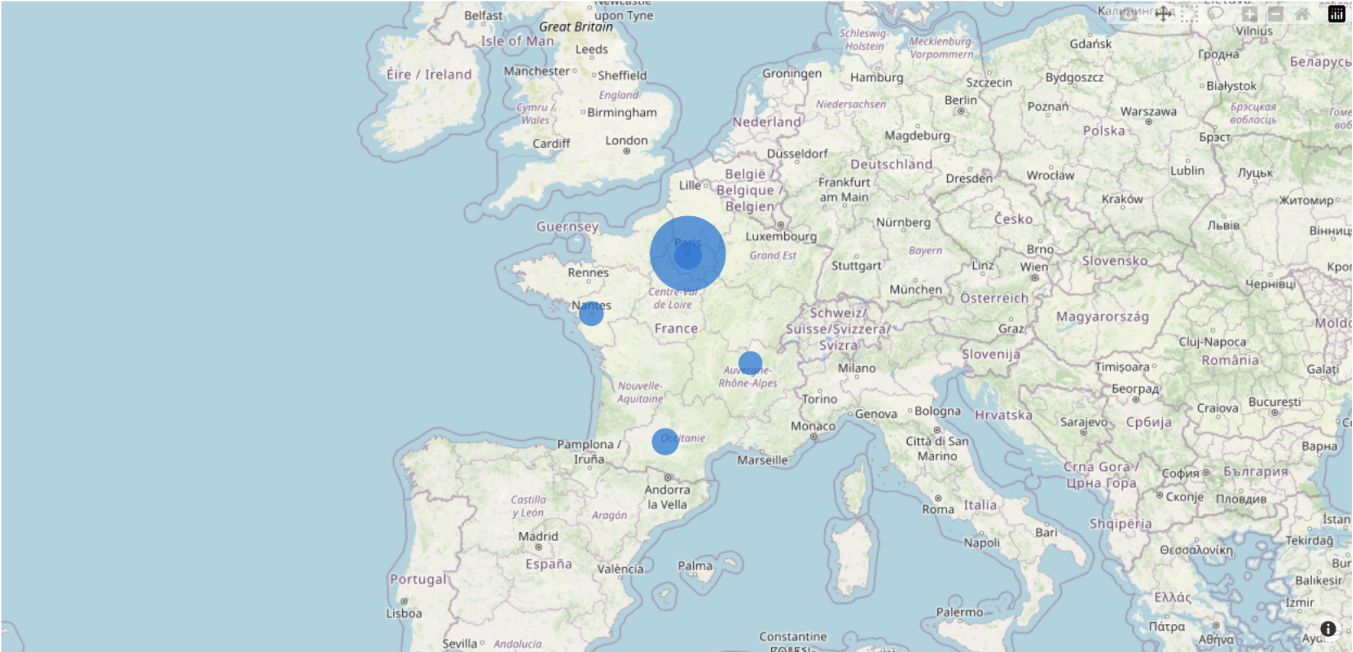




Job Market Dashboard

[Health Check](#) [Soft Skills](#) [Top Ville](#) [Search Offer](#)

Villes avec le plus d'offres d'emploi



Job Market Dashboard

[🔍 Health Check](#) [🧠 Soft Skills](#) [🏠 Top Ville](#) [🔍 Search Offer](#)

🧠 Soft Skills par Code ROME

Sélectionnez un code ROME :

A1102

ID	Soft Skill	Description
1	Être autonome	Exercer ses activités sans constante supervision, s'organiser en prenant des initiatives dans un cadre de
2	Sens de l'innovation/créativité	Raisonnement différencié, rechercher, imaginer, définir, mettre en œuvre des pratiques, des solutions nouve
3	Capacité d'adaptation	Faire face à la nouveauté et au changement, ajuster sa posture en fonction de l'environnement, de la situ
4	Curiosité intellectuelle	Être ouvert à la nouveauté, avoir le goût d'apprendre.
5	Etre persévérant	Faire preuve de ténacité, surmonter les obstacles en vue d'atteindre un objectif défini.
6	Faire preuve d'autorité	Faire accepter son autorité légitime en s'affirmant et en assumant ses responsabilités dans le respect d'i
7	Sens des responsabilités	S'engager, avec discernement, dans une action relevant de ses attributions et assumer ses choix.
8	Maîtrise de soi	Contrôler ses émotions, conserver son sang-froid en toute circonstance, en particulier dans les situation
9	Sens de l'organisation	Avoir une propension à agencer et structurer son environnement en vue d'optimiser le service rendu.
10	Orientation clients	Renvoie à la capacité de focaliser tous les efforts nécessaires pour comprendre les besoins des clients i
11	Être rigoureux	Prêter attention aux détails, avoir le souci de la vérification et du contrôle en vue de réduire les ince
12	Sens de la pédagogie	Exprimer avec clarté, expliquer et transmettre une information, une réglementation, une procédure ou un p
13	Être à l'écoute	Être réceptif et attentif aux messages et sollicitations de ses interlocuteurs et de son environnement.
14	Avoir l'esprit d'équipe	Coopérer avec autrui et contribuer, au sein d'une équipe ou d'un projet, à l'atteinte d'un objectif commu
15	Sens des relations humaines	S'intéresser à autrui et établir aisément avec ses interlocuteurs, une relation propice aux échanges et à
16	Sens des affaires et de la valeur de l'argent	Renvoie à la capacité d'empathie et d'analyse de l'environnement pour vendre et faire des affaires au mie
17	Faire preuve de discrétion	Faire preuve de retenue dans la divulgation d'informations et être capable d'en respecter la confi dentia
18	Faire preuve de diplomatie	Être pondéré, nuancé, faire preuve de tact et d'habileté dans sa relation avec autrui.
19	Réactivité	Répondre rapidement et de manière appropriée aux sollicitations, en particulier dans les situations d'urg
20	Sens de l'analyse	Examiner, décomposer et comparer les éléments constitutifs d'une situation, d'une procédure, d'un documen
21	Esprit de synthèse	Réunir, faire émerger et mettre en cohérence les informations essentielles concernant une situation, un d
22	Sens critique	S'interroger, porter un jugement constructif, avancer une contreproposition argumentée sur un fait, une s
23	Sens de l'initiative	Prendre spontanément, dans son champ de responsabilité, les dispositions susceptibles d'optimiser son act

4. Comment tout ça marche ensemble

1. **Les données arrivent** : Le conteneur ETL récupère les offres d'emploi
2. **Les données sont nettoyées** : Le conteneur ELT les transforme
3. **L'IA apprend** : MLFlow entraîne des modèles sur ces données
4. **L'utilisateur utilise le site** : Il va sur notre site Dash
5. **Le site récupère les infos** : Dash demande au backend FastAPI
6. **Le backend répond** : Il va chercher dans la base et utilise l'IA si besoin
7. **L'utilisateur voit le résultat** : Des recommandations personnalisées

5. Configuration technique

Fichier docker-compose.yml

Ce fichier dit à Docker comment organiser tous nos conteneurs :

- Quels conteneurs créer
- Comment ils communiquent entre eux

- Où stocker les données importantes
- Quels ports utiliser pour accéder aux services

Variables importantes

- Mots de passe pour Snowflake
- Clés d'accès aux APIs externes
- Paramètres de configuration de chaque service

Stockage des données

- Les données importantes sont sauvegardées même si on redémarre les conteneurs
- Les logs sont conservés pour pouvoir diagnostiquer les problèmes

6. Avantages de cette organisation

Pour le développement :

- Chacun peut installer le projet en une commande
- Pas de conflit entre les différents outils
- Facile de tester une nouvelle version

Pour la production :

- Si un service plante, les autres continuent de marcher
- On peut améliorer un service sans toucher aux autres
- Facile de voir quel service consomme quoi

Pour la maintenance :

- Chaque service a sa propre responsabilité
- Plus facile de trouver et corriger les bugs
- Documentation claire de ce que fait chaque partie

7. Ce qui reste à faire

Court terme

- Finir les tests pour vérifier que tout marche bien ensemble
- Créer un environnement de test avant la production
- Écrire la documentation pour les nouveaux développeurs

Moyen terme

- Ajouter de la surveillance pour savoir si tout va bien
- Optimiser les performances
- Renforcer la sécurité

Long terme

- Prévoir l'augmentation du nombre d'utilisateurs

- Ajouter de nouvelles fonctionnalités
- Améliorer l'intelligence artificielle

8. Points d'attention

Risques identifiés :

- Si les APIs externes (France Travail, Adzuna) ne marchent plus
- Coûts de Snowflake si on a beaucoup de données
- Plus difficile de déboguer quand tout est séparé

Solutions prévues :

- Surveiller la disponibilité des APIs
- Optimiser l'utilisation de Snowflake
- Bien organiser les logs pour faciliter le débogage

Ce qu'il faut retenir : Notre architecture avec Docker nous permet d'avoir une application robuste, facile à maintenir et qui peut grandir avec nos besoins. Chaque service a un rôle précis et peut évoluer indépendamment des autres.