

Test

Python et Data Engineering:

1. Data pipeline

Fichier **main.py**

2. Traitement ad-hoc

Extraire depuis le JSON produit par la data pipeline le nom du journal qui mentionne le plus de médicaments différents ?

3. Pour aller plus loin

Quels sont les éléments à considérer pour faire évoluer votre code afin qu'il puisse gérer de grosses volumétries de données (fichiers de plusieurs To ou millions de fichiers par exemple) ?

Pourriez-vous décrire les modifications qu'il faudrait apporter, s'il y en a, pour prendre en considération de telles volumétries ?

SQL :

1. Première partie du test

2. Seconde partie du test