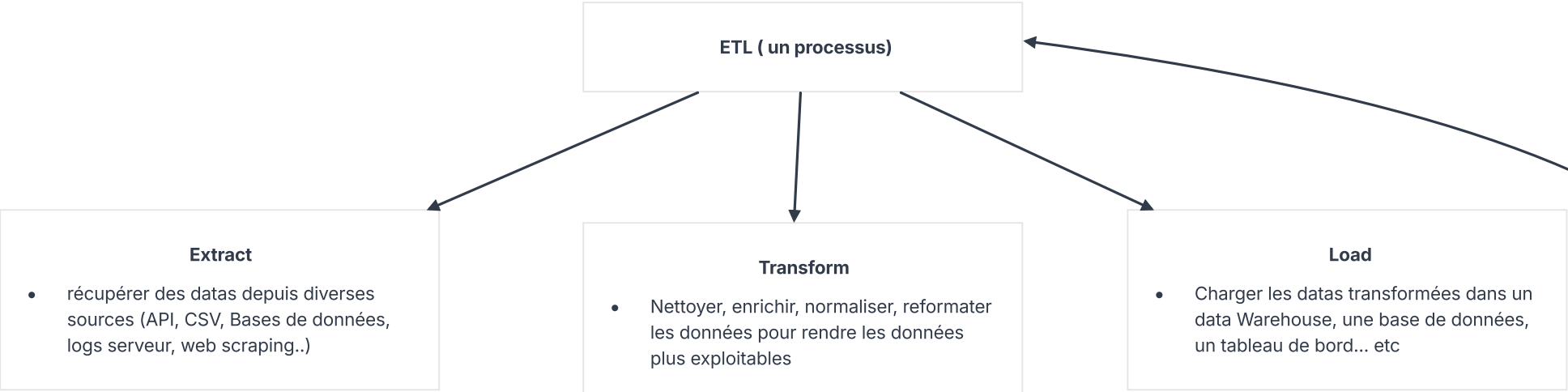


**NOTES :**

- **OpenRefine ne transforme pas automatiquement le projet complet.** Il sert juste à **définir les règles** et à tester comment les données devraient être nettoyées
- On exporte la décision qu'on veut et après on doit rédiger le script qu'on veut



**ETL X Laravel**

- il existe des packages ETL directement dans Laravel :
  - a. **jwhulette/pipes** : créer des pipelines ETL en PHP.
  - b. **redesign/etl** : migrations de données complexes
  - c. **laravelplus/etl-manifesto** : ETL déclaratif

**Transform interatif**

**OpenRefine (gratuit et open source)**

- Logiciel libre de nettoyage et de mise en forme des données, similaire dans son apparence à un tableur mais fonctionne comme une base de données.
- Permet de :
  - détecter les doublons
  - corriger les fautes
  - normaliser
  - transformer colonnes
  - visualiser les distributions

**Complémentaire**

**Apache Airflow (gratuit et open source)**

- Orchestrateur d'ETL : il planifie, orchestre, automatise, surveille des pipelines ETL
- Agit le long des 3 étapes ETL
- Permet de :
  - lancer un ETL automatiquement
  - gérer les dépendances entre étapes
  - relancer en cas d'échec
  - logger et surveiller
  - planifier
  - paralléliser

**Exemple :**

- lance pas LOAD si Transform a échoué
- imposer un ordre de traitement extract - transform-load
- airflow garde : logs d'exécution, durée des taches, messages d'erreur donc debug rapide, audit....

**Exemple :** dans une colonne pays

FR / france / France / FRA

- permet de repérer ces variations, les regrouper, et décider que tout sera mis à France

**Orchestration**