




# Lab 3: Earthquake

Dans ce lab on va essayer de se connecter à une source de données dynamique, ils s'agit d'un API public des mesures de tremblements de terre

## API Documentation - Earthquake Catalog

USGS Earthquake Hazards Program, responsible for monitoring, reporting, and researching earthquakes and earthquake hazards

 <https://earthquake.usgs.gov/fdsnws/event/1/#methods>

## Pipeline:

Nous avons dans notre pipeline 4 composants:

- Script python packagé dans un fichier exécutable `fetch_data.exe`
- Base de données Postgres avec le nom `landing_earthquake`
- transformation pentaho dans le fichier `earthquake_transformation.ktr`
- Rapport power BI `map_powerbi.pbix`

## Etapes:

1. Créez la base de donnée `landing_earthquake` dans Postgres
2. Lancez le programme `fetch_data.exe` et laissez le tourner sur Terminal

Le script python derrière ce programme va requêter chaque minute l'API Earthquake et capter les données des dernières 24h.

A chaque fois il va créer et/ou remplir une table appelée `raw_measures` dans la base de donnée de Postgres créée précédemment, vous n'avez pas besoin de créer cette table explicitement, le script s'en occupe.

Les données de la table sont remplacés et pas ajoutés à chaque itération du script

**Important:** Si vous choisissez un nom de base de données différent de celui en haut, le script risque de ne pas marcher correctement car par défaut ils cherche une base nommée `landing_earthquake` sur le localhost.

Vous pouvez modifier ces paramètres si vous regardez directement le code python dans le répertoire `earthquake` du repo github et vous installez l'environnement python sur votre machine

3. Créez la table `measures` dans la même base et utilisez la transformation de pentaho pdi: `earthquake_transformation.ktr` pour transformer la colonne time en timestamp (format date) et peupler la table nouvellement créée
4. A ce stade vous pouvez vous baser sur le rapport existant de power BI : `map_powerbi.pbix` ou créez un nouveau from scratch. L'idée ici est de charger la table `measures` et l'utiliser pour afficher dans une carte du les mesures en bulle centrée sur les coordonnées GPS et dont le diamètre dépend de la valeur de magnitude de tremblement de terre.