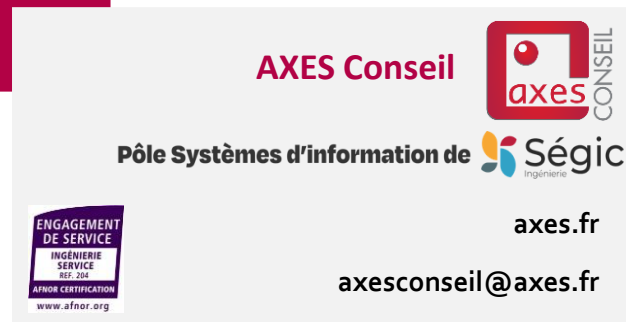


# Quid de l'après ETL « accessible » ?



## Manipuler des données SIG hétérogènes : un besoin devenu majeur

La généralisation des données numériques, spatialisées ou non, dans le fonctionnement des organisations, amène les administrateurs SIG à devoir travailler avec des sources de données hétérogènes en format et structure afin de produire de nouvelles données, réaliser des analyses, etc.

Il semble donc indispensable de disposer d'outils de manipulation de données SIG hétérogènes, parfois volumineuses. Cette évolution profonde est visible par la présence quasi généralisée dans les entités SIG d'outils de type « ETL » : **Extract, Transform, Load**.

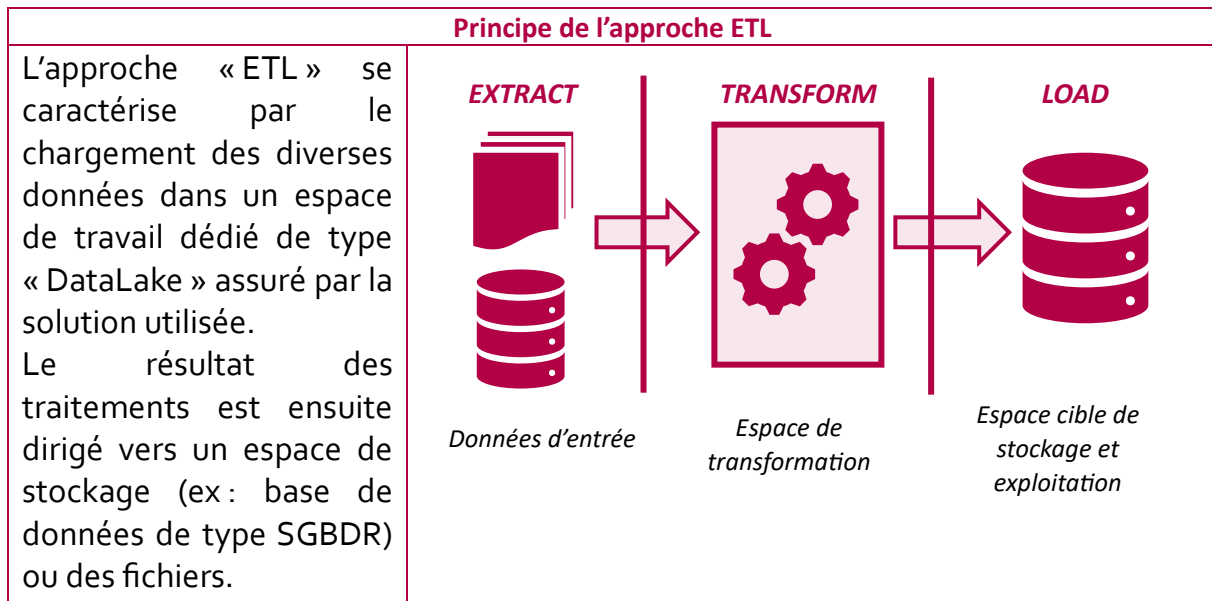
## Une récente évolution du marché des ETL spatiaux

Si dans les années 2010/2015 plusieurs solutions commerciales ou OpenSource co-existaient (GéoKettel, Pentaho, Talend, FME), le marché s'est progressivement concentré. Actuellement, la solution FME (éditeur SafeSoftWare) domine très largement le marché. Restent également présents des solutions comme Pentaho, Talend ou KNIME.

Cependant, depuis 2023 les conditions ont fortement changé :

- SafeSoftware a (très) fortement augmenté le tarif de ses licences. Pour le dire sobrement, de très nombreuses collectivités se sont donc interrogées sur le cout de l'investissement, etc.

- Les éditeurs respectifs de Pentaho ou Talend ont abandonné les versions OpenSource de leur solution respectives (Talend au 1<sup>er</sup> janvier 2025 ; Pentaho en 2024). Des cartouches spatiales n'ont pas été maintenues (Talend).



Fin 2024, la situation de l'offre de solutions ETL s'est un peu assombrie du point de vue des services comptables des collectivités, invitées à faire des économies. Or, le besoin de manipulation de données reste présent chez les administrateurs SIG. Existe-t-il des solutions tierces ?

## Exploitation des bibliothèques spécialisées comme OGR/GDAL et Python

A défaut d'offre « sur étagère », il reste possible d'exploiter des bibliothèques informatiques disponibles en Open Source : OGR, GDAL ou Python. Cette approche permet de manipuler les données mais requière des compétences en développement informatique. L'utilisation de Python implique un suivi régulier des diverses librairies exploitées pour assurer la stabilité du code. La maintenance du code peut donc rapidement être exigeante (et chronophage). Vis-à-vis d'une solution comme FME proposant des IHM graphiques (« drag & drop »), la marche peut sembler haute.

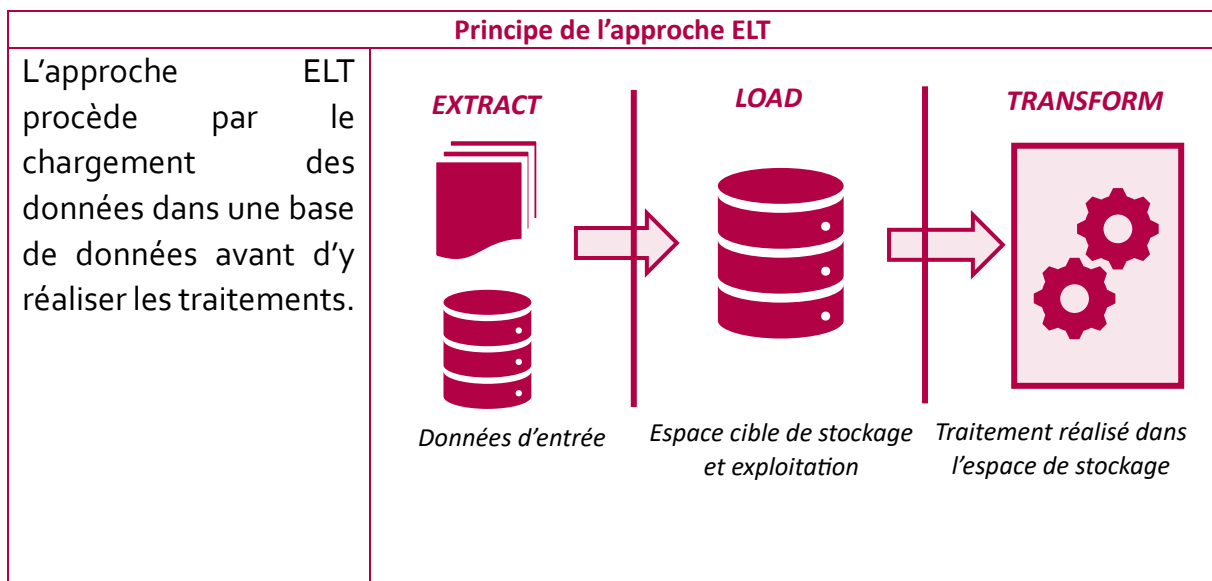
## Explorer l'approche « ELT ? » directement en base de données

Il existe une autre possibilité permettant de répondre à cette nécessité de travailler des données sources hétérogènes : il s'agit de l'approche dite « **ELT** » : **Extract Load, Transform**. Le principe est de charger dans une base de données (ex : PostgreSQL/PostGIS) toutes les données vectorielles ou texte requises. Une fois chargées, il devient possible de les exploiter via les outils de la base de données (SQL, PL SQL, etc.).

Le principe « ELT » peut répondre aux besoins de manipulation de diverses données SIG hétérogènes rencontrés par les administrateurs SIG de collectivités par exemple.

Contrairement aux outils d'ETL, disposant de nombreux connecteurs capables d'accéder à une quantité importante de formats, l'approche ELT se distingue par la nécessité de charger dans une base divers formats. C'est l'opération la plus sensible du processus. Car, s'il existe des outils d'intégration capables de traiter les formats les plus courants (ex: utilitaire OGR; QGIS), les formats plus « exotiques » ou propriétaires (ex: dwg) nécessiteront un effort plus important...sans garantie de succès.

Une fois l'effort de chargement des données en base (ex: PostGIS) achevé, l'administrateur SIG évolue alors directement dans son environnement courant de production avec des données techniquement homogènes et structurées. Il peut notamment utiliser le langage SQL pour mettre en place les opérations de traitement : choix d'attribut, renommage, re-typage, concaténation, fusion de tables, calculs, etc. Il est à noter que le langage SQL présente une très grande stabilité, propriété bien utile pour justifier le temps investi.



## En conclusion

L'évolution de l'offre de solutions ETL a un impact important pour les personnes amenées à manipuler des données notamment spatiales. A défaut de voir une offre abordable, pérenne et fonctionnellement satisfaisante, émerger rapidement, les solutions de recours ne sont pas nombreuses et ont en commun de devoir monter en compétence technique en développement informatique (SQL, python, etc.).

L'approche « ELT » permet de travailler au plus près de la donnée directement dans le stockage avec du code informatique stable (SQL).