1.a.

Source 1 : Base de données relationnels (structuré)

Source 2 : Fichier plat (structuré)

Source 3 : Fichier semi-structurés (semi-structurés)

1.b.

Préparation :

* Nettoyage des données : Supprimer les espaces supplémentaires dans la source « category.csv »
* Gestion des valeurs manquantes : Il manque les villes dans la source « client.json », donc il faut remplir les valeurs manquantes via un API ou un fichier externe.
* Filtrage des données : Supprimer les données inutiles dans la source « client.json »

Standardisation :

* Transformer la source « client.json » en format relationnel

Calcul/Transformation :

* Mapping et jointures : Associer les différentes sources de données

1.c.

La différence principale entre le chargement incrémental et le chargement fusion c’est la façon de définir si une ligne de donnée entré est nouveau ou pas. Le chargement incrémental est basé sur l’identifiant ou la date de dernier chargement. Si l’identifiant ou la date de chargement est plus grande que le dernier chargement, on pense que cette ligne est nouvelle. Cependant, le chargement fusion est basé sur l’existence d’identifiant d’une ligne de donnée dans la base de données. Si l’identifiant n’existe pas dans la base de données, on pense que cette ligne est nouvelle.

1.d.

Une image contenant texte, diagramme, ligne, Police

Le contenu généré par l’IA peut être incorrect.

Figure 1 Processus d'ETL

1.e.

Une image contenant texte, capture d’écran, diagramme, Police

Le contenu généré par l’IA peut être incorrect.

Figure 2 Processus dans Talend

2.a.

Tableau 1 Dictionnaire des mesures

|  |  |  |  |
| --- | --- | --- | --- |
| Code | Description | Type | Formule d’extraction |
| CA | Le chiffre d’affaires en fonction d’un client, d’un produit et d’un jour | Réel | SUM(contenir.quantité \* Produit.prixUnitaire) |

2.b.

Une image contenant texte, Police, logiciel, nombre

Le contenu généré par l’IA peut être incorrect.

Figure 3 Représentation graphique du système décisionnel complet

2.c.