|  |  |
| --- | --- |
|  | **2015** |
|  | [Tapez le nom de la société]  Pascal AUREGAN |

|  |
| --- |
| **[Rapport projet]** |
| Test de chaine décisionnelle sur cas simple. Base opérationnelle en fichier et MySQL, ETL avec TALEND Data Integration, Reporting avec SAS |

# Introduction

Le but de se projet est de mettre en place une chaine décisionnelle. L'objectif de ce projet n'étant pas de pousser dans ses retranchements les outils de la chaine décisionnelle, il a été décidé de créer de toute pièce un jeu de données. Le thème de ce jeu de données est l'achat de BD comme dans le cours.

La base opérationnelle décrit les factures d'un ensemble librairies en France dont les clients sont connus.

La chaine décisionnelle doit être en mesure de répondre in fine à la question: analyser les ventes de BD en France en fonction des clients, du lieu et date d'achat.

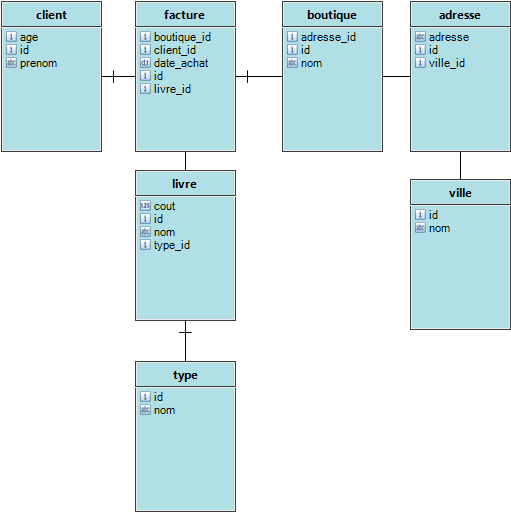
La couche opérationnelle est assurée par une base de données de fichier csv et une base de données PostgresSQL.

Pour la couche ETL, il a été choisi TALEND Data Integration.

La couche reporting étudiée lors de ce projet est assurée par SAS Studio Version Etudiante.

# Jeu de données de la base opérationnelle

## Modèle Relationnel



Le modèle relationnel est un modèle normalisé.

- Une facture a été établie pour un client donné dans une boutique donnée à une date donnée et en rapport avec un livre donné.

- Le livre appartient à un type donné (Roman, BD, etc.).

- La boutique a une adresse située dans une ville

Voici combien de lignes ont chaque table :

- adresse : 26

- boutique : 25

- client : 8819

- facture : 8819

- livre : 8

- type : 3

En annexe, le code R qui a servi à générer les données dont il a été estimé que la pertinence n'avait que peu d'importance pour l'exercice.

De plus, nous possédons un fichier de type csv contenant un dictionnaire de villes avec les départements associés ainsi que les régions.

Voici en exemple les quatre premières lignes de ce fichier villes\_def.csv

"ville","departement","region\_name"

"Beauvais","Oise","Picardie"

"Compiègne","Oise","Picardie"

"Clermont","Oise","Picardie"

# TALEND Open studio for Data integration 5.6

## Téléchargement / Installation

Le téléchargement se fait sur http://fr.talend.com/download/data-integration.

La documentation de l'outil est disponible au même endroit dans la partie Manuels d'utilisateurs et est facilement accessible.

L'outil est en fait basé sur eclipse ce qui le rend facilement appréhendable pour ceux qui connaissent le célèbre IDE mais vient aussi avec ses défauts.

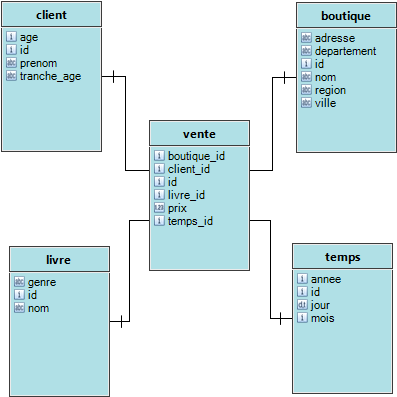
## Création du business model

### question a répondre

Analyse des ventes de bandes dessinées d'une enseigne de librairie possédant plusieurs boutiques dans plusieurs villes.

### business model choisi

Le modèle décisionnel retenu est le suivant :



#### Table de faits Vente:

##### indicateurs

prix (somme, moyenne)

#### TaBLE client

- id

- prenom

- age

- tranche\_age ("-12", "12-25", "35-65", "25-34", "+65")

\* age < tranche\_age

#### TABLE boutique

- id

- nom

- adresse

- ville

- region

\* adresse < ville < region

#### table livre

- id

- nom

- genre

\* nom < genre

#### table temps

- id

- jour

- mois

- annee

## Mise en place de l'integration