

# Cahier des Charges

## -Projet Distributech-

<b>1. Introduction.....</b>	<b>2</b>
<b>2. Motivations.....</b>	<b>2</b>
2.1. Le client.....	2
2.2. Le problème.....	2
2.3. L'existant.....	2
2.4. Le besoin non satisfait.....	2
2.5. Les objectifs.....	2
<b>3. Contexte.....</b>	<b>2</b>
3.1. Marché.....	2
<b>4. Profil des utilisateurs finaux.....</b>	<b>2</b>
4.1. En utilisation.....	2
4.2. En exploitation.....	3
4.3. En maintenance.....	3
<b>5. Documentation.....</b>	<b>3</b>
5.1. Terminologie métier.....	3
<b>6. Aspects contractuels.....</b>	<b>3</b>
6.1. Délais et protocole de livraison.....	3
<b>7. Aspect juridique et commerciaux.....</b>	<b>4</b>
7.1. Données personnelles et RGPD.....	4
<b>8. Choix technologiques.....</b>	<b>5</b>
<b>9. Fonctions à réaliser.....</b>	<b>5</b>
9.1. Ce que le système doit faire.....	5
<b>10. Création de la base de données SQL centralisée.....</b>	<b>6</b>
<b>11. Processus ETL.....</b>	<b>6</b>
<b>12. Environnement technique.....</b>	<b>7</b>
12.1. Infrastructure matérielle.....	7
12.2. Environnement logiciel.....	7
12.3. Bonnes pratiques.....	7
<b>13. Contraintes du système.....</b>	<b>7</b>
13.1. Contraintes matérielles.....	7
13.2. Contraintes logicielles.....	8
13.3. Contraintes fonctionnelles.....	8
13.4. Contraintes d'ergonomie.....	8

# **1. Introduction**

Le projet se nomme DISTRIBUTECH. Il consiste à concevoir et à mettre en place une base de données relationnelle ainsi qu'un processus ETL permettant d'intégrer chaque semaine des données issues de fichiers CSV et d'une base SQLite locale. Le but de ce projet est de centraliser et suivre l'historique des commandes des différents revendeurs ainsi que l'état des stocks régionaux.

## **2. Motivations**

### **2.1. Le client**

L'entreprise Distributech est un grossiste en équipements électroniques travaillant avec un réseau de revendeurs régionaux répartis sur l'ensemble de la France. Ces partenaires passent des commandes régulières pour réapprovisionner leurs stocks.

### **2.2. Le problème**

Distributech souhaite améliorer la gestion de son flux logistique et centraliser les informations commerciales envoyées par ses revendeurs et commerciaux.

### **2.3. L'existant**

Actuellement, les commandes des revendeurs sont envoyées via un fichier CSV. Les stocks, quant à eux, sont enregistrés localement dans une base de données au niveau de DISTRIBUTECH et mis à jour par les commerciaux.

### **2.4. Le besoin non satisfait**

L'information étant centralisée au niveau de Distributech, il y a de l'opacité sur les flux entre chaque revendeur et l'entreprise, alors qu'ils partagent tous le même stock. De plus, il y a une absence de suivi temporel des stocks.

### **2.5. Les objectifs**

L'objectif principal est de rendre transparent l'état des stocks pour tous les acteurs en créant une base de données SQL centralisée, développer une pipeline ETL et générer un fichier CSV hebdomadaire de l'état des stocks par produit.

## **3. Contexte**

### **3.1. Marché**

L'entreprise Distributech travaille en B2B (ou Business-to-Business). Ses transactions commerciales sont réalisées avec d'autres entreprises, les volumes d'achat sont donc très importants. Un meilleur suivi logistique et commercial est donc primordial pour éviter toutes ruptures de stocks et rester concurrentiel.

## **4. Profil des utilisateurs finaux**

### **4.1. En utilisation**

Les commerciaux et les revendeurs seront les premiers utilisateurs concernés. Ils saisissent ou transmettent les données sources utilisées dans le processus ETL. Aucune connaissance technique n'est attendue en base de données.

## 4.2. En exploitation

Les responsables logistiques exploiteront les données à des fins de gestion logistique et commerciale. Ils auront besoin d'avoir accès aux fichiers CSV générés par le système, les données doivent donc être à jour et compréhensibles sans traitement.

## 4.3. En maintenance

L'équipe informatique et le support technique assureront la maintenance et le bon fonctionnement du pipeline ETL et de la base de données centralisée. Le code fourni devra donc être bien documenté afin de faciliter la compréhension des différentes étapes du processus ETL.

# 5. Documentation

## 5.1. Terminologie métier

*Grossiste* : marchand en gros, intermédiaire entre le détaillant et le producteur ou le fabricant.

*Revendeurs* : intermédiaire commercial qui lie un distributeur et des consommateurs.

*Commandes* : document par lequel un client demande une marchandise dans un délai déterminé.

*Stock* : quantité de marchandises en réserve chez le grossiste.

# 6. Aspects contractuels

## 6.1. Délais et protocole de livraison

Semaine du 14 au 18 juillet : MCD/MLD et création de la base de données SQL

Semaine du 21 au 25 juillet : Scripts ETL et rédaction du cahier des charges

Semaine du 28 juillet au 1 août : Finalisation des scripts ETL, tests d'intégration des données dans la base SQL et génération du fichier CSV + création du diaporama pour présentation finale

(Vacances des équipes du 4 au 17 août)

Semaine du 25 août : présentation du projet et des livrables attendus

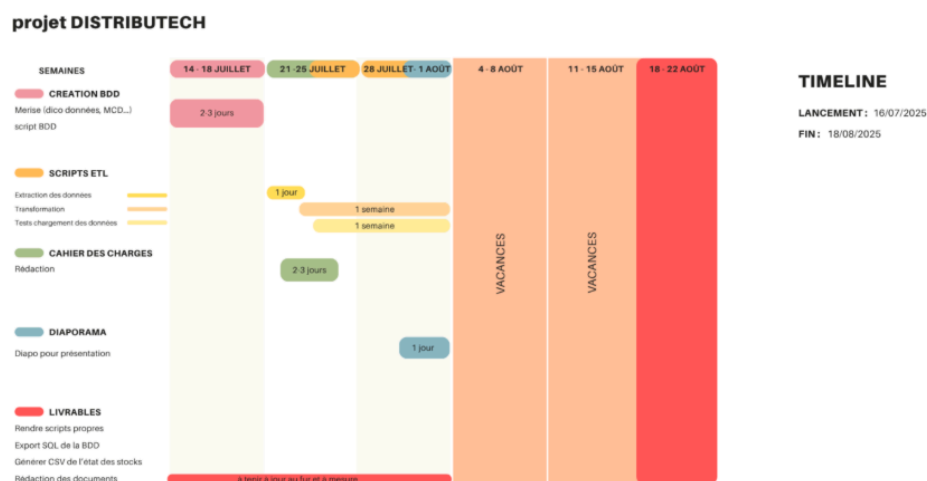


Figure 1 : Retroplanning prévu pour le projet Distributech

## **7. Aspect juridique et commerciaux**

- Les données clients ne doivent pas être partagées en dehors du cadre de ce projet.
- Le code développé est la propriété du client (l'entreprise Distributech).
- Pas de publication externe sans autorisation explicite.

### **7.1. Données personnelles et RGPD**

#### **Nature des données traitées :**

- Le système traite principalement des données professionnelles, liées aux produits, aux stocks, aux revendeurs (noms, identifiants uniques internes, localisation géographique).
- Aucune donnée personnelle sensible n'est prévue à ce stade.

#### **Sécurisation des accès :**

- Les fichiers de données sont stockés dans un répertoire sécurisé avec des droits restreints en écriture.
- L'accès à la base de données SQL centralisée doit être protégé par authentification.

#### **Conformité aux RGPD :**

- Les données sont traitées uniquement à des fins de gestion logistique et commerciale propre à l'entreprise Distributech.
- Les historiques des commandes sont conservés sur une période définie suivant le cadre légal en vigueur.
- Dans le cas où des données nominatives seraient présentes, les droits d'accès, de rectification et de suppression doivent être garantis sur demande de la personne concernée.

# Spécifications techniques

## 8. Choix technologiques

Dans le cadre de ce projet, le langage Python a été choisi comme langage de développement. Ses packages pandas et sqlite3 ont été retenus pour manipuler les données sources, issues des fichiers CSV et de la base de données locale SQLite. Concernant la base de données centralisée, il s'agit d'une base de données SQL relationnelles structurée, créée à partir d'un script SQL.

## 9. Fonctions à réaliser

Deux fonctions sont à réaliser :

- Créer une base de données SQL relationnelle,
- Mettre en place un processus ETL hebdomadaire devant extraire, transformer et charger les données dans la base de données.

### 9.1. Ce que le système doit faire

Priorité	Fonction	Description	Critère de performance
1	Création d'une base de données SQL	La base de données doit être structurée autour des revendeurs, régions, produits, commandes et stocks	MCD et MLD documentés Export SQL valide
2	Extraction des données des fichiers CSV	Les informations des commandes sont récupérées pour être intégrées à la base SQL	Les fichiers sont traités sans erreur
3	Extraction des données de la base locale SQLite	Les informations concernant les revendeurs, les régions et les produits sont extraites et réparties dans des tables	Les tables sont extraites correctement et leurs relations sont bien conservées
4	Nettoyage des données	La cohérence des données est vérifiée, les doublons sont supprimés et les formats de date sont adaptés au besoin	Transformations des données sans erreurs
5	Chargement des données dans la base de données SQL	L'ensemble des données sont insérées dans la base de données centralisée	Données présentes et cohérentes dans la base SQL
6	Générer un fichier CSV de l'état des stocks	Le fichier CSV doit permettre d'avoir une photographie des stocks par produit	Fichier généré dans le format voulu

## 10. Création de la base de données SQL centralisée

Une modélisation structurée de la future base de données SQL a été réalisée. Un Modèle Conceptuel de Données (MCD) a servi de base pour définir les principales entités du système et les liens qui existent entre elles. Ce modèle (figure 2 ci-dessous) permet de s'assurer que toutes les informations nécessaires au fonctionnement du système sont bien prises en compte et correctement reliées entre elles.

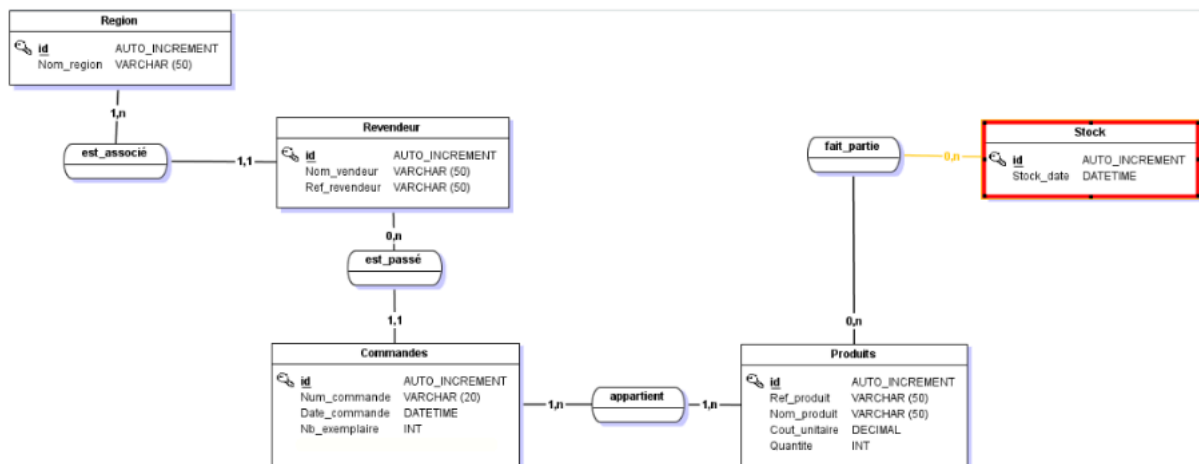


Figure 2 : MCD de la base de données SQL – vue conceptuelle

Ce MCD a ensuite été traduit en Modèle Logique de Données (MLD, figure 3 ci-dessous) afin de préciser la structure de la base de données relationnelle qui sera créée. Il permet également de définir les tables, les clés primaires et étrangères.

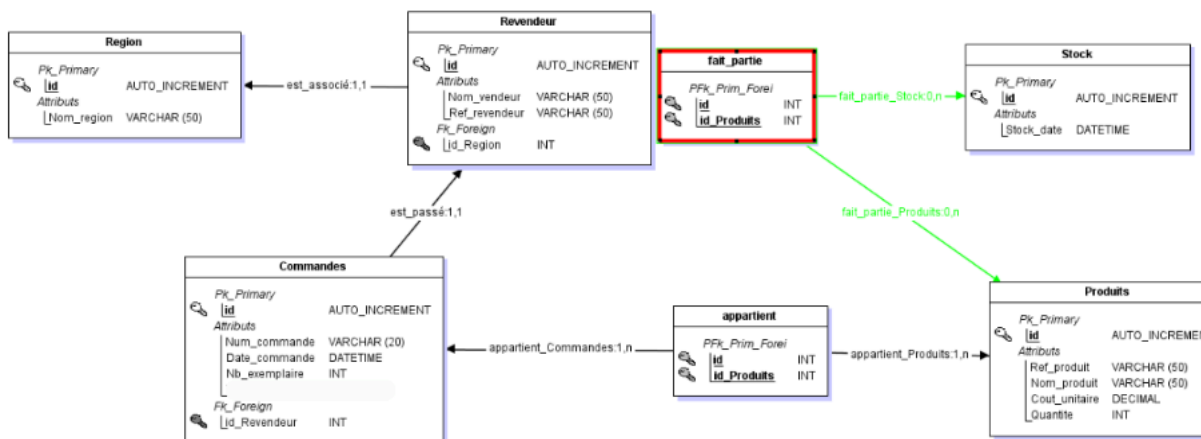


Figure 3 : MLD représentant les tables et leurs relations de la future base de données SQL

## 11. Processus ETL

Un processus ETL (Extract, Transform, Load) sera mis en place pour centraliser les données issues des différentes sources. Cela permettra de garantir l'intégrité des données avant leur insertion en base et de générer un fichier CSV de l'état des stocks par produit.

**Déroulement :**

1. Extraction des données depuis deux sources (Fichier CSV et base de données locale SQLite).
2. Transformation : suppression des doublons, harmonisation des formats de date, regroupement des données en sous forme de tableaux (ou dataframes).
3. Chargement : insertion des données transformées dans la base SQL centralisée et génération automatique d'un fichier CSV présentant l'état des stocks par produit.

**Fréquence et critère de performance :**

- Fréquence : hebdomadaire
- Aucun message d'erreur durant l'exécution du processus.
- Cohérence des données dans la base SQL.
- Génération du fichier CSV conforme au format attendu.

## 12. Environnement technique

### 12.1. Infrastructure matérielle

**Poste de travail** : ordinateur Windows, Linux ou macOS.

**Espace disque** : au moins 5 Go disponibles pour les fichiers CSV, exports générés et sauvegardes des bases SQLite et SQL.

**Connexion réseau** : accès stable à internet pour l'exécution du processus ETL et la synchronisation avec la base SQL centralisée.

### 12.2. Environnement logiciel

**Langage de développement** : Python 3.12

**Bibliothèques Python** :

- pandas : traitement, nettoyage et manipulation des données extraites.
- sqlite3 : interaction avec la base de données locale SQLite.
- mysql-connector : connexion à la base centralisée SQL

**Bases de données** :

Locale : SQLite pour les données sources.

Centralisée : MySQL structurée selon le MLD validé.

### 12.3. Bonnes pratiques

- Code Python lisible et commenté.
- Sauvegarde régulière de la base SQL centralisée.
- Gestion des accès selon les profils utilisateurs.

## 13. Contraintes du système

### 13.1. Contraintes matérielles

Le système ne nécessite pas d'infrastructure lourde. Un ordinateur classique avec un espace de stockage correct et une connexion réseau fiable est suffisant.

### **13.2. Contraintes logicielles**

Le système repose sur des logiciels open-source pour garantir la portabilité et l'accessibilité. Aucune dépendance à un logiciel propriétaire spécifique n'est prévue.

### **13.3. Contraintes fonctionnelles**

Le processus ETL doit être automatisé, fiable et exécutable chaque semaine. Les données ne doivent être accessibles qu'aux utilisateurs autorisés et la base de données SQL centralisée peut être restreinte en lecture/écriture selon les profils (lecture seule pour les responsables logistiques et accès complet pour le support technique).

### **13.4. Contraintes d'ergonomie**

- Pour les commerciaux et les revendeurs : aucune modification de leur outil actuel, le système s'adapte à l'existant.
- Pour les responsables logistiques : les fichiers CSV générés doivent être lisibles dans Excel, structurés avec des noms de colonnes clairs et utilisables directement.
- Pour l'équipe informatique et le support technique : le code du processus ETL doit être lisible et commenté.