

Cahier des Charges

Distributech ETL

Sommaire

1. Introduction	2
1.1 Motivations	2
1.2 Contexte	3
2. Documentation	3
3. Profil des utilisateurs finaux.....	4
3.1 En utilisation	4
3.2 En exploitation	4
3.3 En maintenance	5
3.4 Environnement d'utilisation	5
4. Fonctions à réaliser	6
4.1 Ce que le système doit faire	6
4.2 Mesure de la réussite du projet	6
4.3 Ce que le système ne doit pas faire	6
5. Contrainte du système.....	7
5.1 Contraintes matérielles	7
5.2 Contraintes logicielles	7
5.3 Contraintes fonctionnelles	7
5.4 Contraintes d'ergonomie	7
5.5 Autres facteurs de qualité exigés par le client	7
6. Aspects contractuels.....	8
6.1 Délais et protocole de livraison	8
6.2 Aspect juridique et commerciaux	8

1. Introduction

Le projet Distributech-ETL vise à concevoir une base de données relationnelle **MySQL** centralisée alimentée par des données nettoyées par ETL, intégrant les données de revendeurs, régions, produits, commandes et mouvements de stock, ainsi que suivre l'évolution des niveaux de stock par produit dans le temps.

1.1 Motivations

1.1.1 Le client

Distributech est un grossiste en équipements électroniques qui collabore avec un réseau de revendeurs régionaux. Ces partenaires transmettent régulièrement des commandes afin de réapprovisionner leurs stocks locaux.



1.1.2 Le problème

Les données de revendeurs, régions, produits, commandes et mouvements de stock ne sont pas intégrés dans une base de données centralisée qui engendre des difficultés pour assurer un suivi logistique fiable et en temps réel.

1.1.3 L'existant

Actuellement, les données de commande sont transmises par fichiers CSV hebdomadaires, tandis que les informations de stock sont maintenues par les commerciaux dans des bases SQLite locales.

1.1.4 Le besoin non satisfait

Il y a des difficultés pour assurer un suivi logistique fiable et en temps réel. Il est donc nécessaire de centraliser et fiabiliser les données afin de :

- Assurer des données nettoyées, cohérentes et non redondantes,
- Suivre l'historique des commandes de manière pérenne,
- Conserver une photographie des stocks par région,
- Permettre un pilotage commercial et logistique à partir de données structurées et centralisées.

1.1.5 Les objectifs

Le projet vise à :

- Mettre en place un pipeline **ETL régulier** (fréquence initiale : hebdomadaire) permettant l'intégration de fichiers CSV (commandes) et de bases SQLite (stocks) nettoyées,
- Concevoir une base de données relationnelle **MySQL** centralisée intégrant les données de revendeurs, régions, produits, commandes et mouvements de stock,
- Suivre l'évolution des niveaux de stock par produit dans le temps,
- Faciliter le suivi commercial, la gestion logistique et l'analyse de performance des zones géographiques.

1.2 Contexte

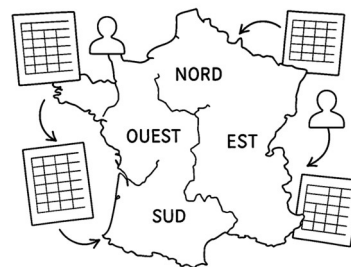
1.2.1 Précisions sur le client

Distributtech est un grossiste en équipements électroniques qui collabore avec des partenaires - un réseau de revendeurs régionaux.

Cet environnement de travail exige la communication rapide de la part des commerciaux vers les revendeurs qui passent leurs commandes. Les commerciaux ont besoin d'avoir une vue en temps réel sur les stock, mais les revendeurs passent commande via fichiers CSV. Ces fichiers ne sont pas encore automatiquement intégrés dans une base de données utilisée par les commerciaux. Il manque donc une solution centralisée qui permettra de récupérer des données de stock à jour.

1.2.2 Marché

La distribution d'équipements électroniques en B2B implique plusieurs enjeux, notamment l'optimisation de la logistique et du suivi des commandes, la réduction des erreurs ou des doublons dans les fichiers, une meilleure gestion des stocks pour éviter les ruptures ou les surplus, ainsi qu'une vision claire de l'activité régionale afin de faciliter la prise de décision.



La mise en place d'une base centralisée permettra de gagner du temps, d'améliorer la fiabilité des données et de mieux exploiter les informations commerciales.

2. Documentation

2.1 Terminologie métier

Vocabulaire spécifique au métier, dans le cadre du projet :

- **Commande** : demande d'un ou plusieurs produits passée par un revendeur.
- **Production** : réapprovisionnement des stocks centralisés par Distributtech.
- **Stock** : niveau de disponibilité des produits pour l'ensemble des revendeurs.
- **ETL** : outil d'extraction, transformation et chargement des données.

2.2 Bibliographie succincte

Documentation indispensable à la réalisation du projet, s'il y a lieu :

Fichiers CSV de commandes Fichiers transmis régulièrement par les revendeurs, contenant l'historique des commandes. Ils constituent la principale source de données transactionnelles à intégrer via le pipeline ETL.

Base SQLite consolidée Fichier local au format .sqlite contenant :

- Les données de référence (produits, revendeurs, régions),
- Les événements de stock (réceptions), utilisé comme source d'extraction complémentaire.

Documentation technique sur les formats attendus

- Spécifications du format des fichiers CSV d'entrée (structure, encodage, séparateur),
- Structure des tables SQLite à respecter pour assurer la compatibilité,
- Schéma relationnel cible MySQL (jointures, clés étrangères, types de données),
- Recommandations sur le traitement des erreurs et la journalisation.

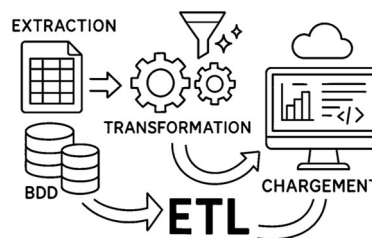
3. Profil des utilisateurs finaux

3.1 En utilisation

Profils et compétences des futurs utilisateurs du système

Les utilisateurs finaux du système sont principalement :

- Les **responsables commerciaux**,
- Et les **équipes commerciales terrain** (commerciaux régionaux, coordinateurs de zone).



Leur objectif est de :

- Accéder aux rapports consolidés de commandes et de stocks,
- Suivre les niveaux de disponibilité des produits par région,
- Identifier les tendances de réapprovisionnement ou les points de tension logistique.

Compétences attendues :

- Aisance avec les outils bureautiques (tableurs, lecteurs de fichiers CSV),
- Bonne compréhension des notions commerciales liées aux commandes et au stock,
- Capacité à interpréter un état de stock ou un historique de commandes.

3.2 En exploitation

Profils et compétences des responsables de l'exploitation du système

Le personnel chargé de l'exploitation technique du système est généralement un **technicien informatique** ou un **data engineer junior**, en interne ou via un prestataire.

Responsabilités clés :

- Exécution du pipeline ETL à la fréquence définie,
- Suivi des fichiers générés (logs, exports de stock),
- Mise à jour éventuelle des chemins de fichiers ou de la configuration du script.

Compétences attendues :

- Connaissance de base en Python (lancement et paramétrage d'un script),
- Savoir utiliser un outil de requête MySQL (console, DBeaver, etc.),
- Compréhension des formats tabulaires (CSV, SQL).

3.3 En maintenance

Profils et compétences du personnel en charge de la maintenance du système

La maintenance technique sera assurée par un :

- **Développeur Python** ou **data engineer**,
- Ou par un **prestataire tiers** si un contrat est établi.

Tâches à prendre en charge :

- Correction d'éventuels dysfonctionnements du pipeline,
- Adaptation à des évolutions de structure de données (ex : changement de format CSV),
- Refactorisation du code et documentation.

Engagements possibles :

- Délai d'intervention en cas de dysfonctionnement bloquant,
- Maintien à jour des dépendances et versions logicielles (Python, bibliothèques).

3.4 Environnement d'utilisation

Environnement matériel & technique :

- Ordinateur de bureau ou serveur compatible Linux ou Windows 10+,
- Python 3.11 ou version compatible installé,
- Serveur MySQL (local ou distant) disponible,
- Accès en lecture/écriture à :
 - Une base **SQLite (.sqlite)** fournie par les commerciaux,
 - Et aux fichiers **CSV** de commandes transmis par les revendeurs.

Environnement professionnel :

- Contexte logistique et commercial interne à l'entreprise,
- Usage par les services commerciaux pour exploitation métier,
- Collaborations ponctuelles avec le service technique ou informatique,
- Pas de contrainte d'exécution temps réel : les traitements sont manuels ou planifiés selon besoin métier (quotidien, hebdomadaire ou à la demande).

4. Fonctions à réaliser

4.1 Ce que le système doit faire

Le projet vise à :

- Mettre en place un pipeline **ETL régulier** (fréquence configurable) permettant l'intégration de fichiers CSV (commandes) et de bases SQLite (stocks),
- Concevoir une base de données relationnelle **MySQL** centralisée intégrant les données de revendeurs, régions, produits, commandes et mouvements de stock,
- Suivre l'évolution des niveaux de stock par produit dans le temps,
- Faciliter le suivi commercial, la gestion logistique et l'analyse de performance des zones géographiques.

4.2 Mesure de la réussite du projet

Indicateurs quantitatifs :

- Taux de réussite du pipeline ETL sans erreur critique (>95%),
- Taux de couverture des commandes intégrées dans la base SQL (>98% des commandes reçues),
- Taux de fiabilité des stocks reconstitués (>90%).

Indicateurs qualitatifs :

- Satisfaction des équipes commerciales sur l'exploitation des rapports de stock,
- Capacité à reconstituer une photographie précise des stocks à une date donnée,
- Réduction des erreurs de saisie ou des incohérences entre revendeurs et catalogue.

Contraintes réglementaires et techniques

- **Cohérence des données** : clés primaires/étrangères, unicité des identifiants, relations conformes au modèle relationnel.
- **Séparation claire des sources** : CSV pour les commandes, SQLite pour les stocks et les entités statiques.
- **Traçabilité** : chaque événement de stock ou commande doit être daté et archivé pour reconstitution historique.
- **Sécurité des données et respect du RGPD** :
 - Les données à caractère personnel doivent être protégées selon les règles du RGPD.
 - Un système de pseudo-anonymisation ou de chiffrement peut être mis en place si nécessaire.
 - Les données doivent être stockées de manière sécurisée, avec journalisation minimale des traitements.



4.3 Ce que le système ne doit pas faire

Les éléments suivants ne sont pas couverts par cette phase :

- Interface utilisateur (front-end ou dashboards),
- Gestion des droits ou rôles utilisateurs,
- Intégration temps réel ou API,
- Paiement, facturation ou documents contractuels.

5. Contrainte du système

5.1 Contraintes matérielles

Configuration, machines de déploiement, compatibilité avec d'autres systèmes, etc. :

Ordinateur avec environnement Python installé. Capacité à gérer plusieurs fichiers CSV par semaine.

5.2 Contraintes logicielles

Système d'exploitation, environnement numérique d'exploitation, compatibilité avec des logiciels existants, etc. :

Utilisation de pandas, sqlite3, et SQL standard. Compatibilité avec Visual Studio Code pour le développement.

5.3 Contraintes fonctionnelles

Personnel, Ressources consommables, sécurité, etc. :

Ressources humaines : équipe projet composée d'étudiants développeurs. Sécurité des données sensible.

5.4 Contraintes d'ergonomie

En relation avec les profils utilisateurs décrits plus haut :

Fonctions accessibles via scripts, pas d'interface graphique. Utilisation prévue par des techniciens ou analystes.

5.5 Autres facteurs de qualité exigés par le client

Fiabilité, intégration, portabilité, maintenabilité, etc. :

- Fiabilité du processus ETL sur des fichiers corrompus
- Portabilité du projet sur plusieurs environnements locaux
- Maintenabilité du code via documentation technique claire

6. Aspects contractuels

6.1 Délais et protocole de livraison

Le projet est prévu pour une livraison finale au 18/08/2025, avec des étapes intermédiaires :

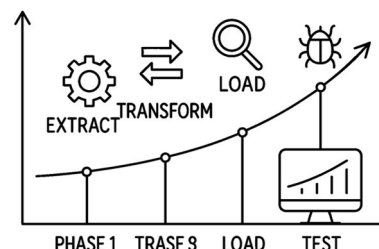
Phase 1 : Module d'extraction csv et SQLite (échéance : 25/07/2025)

Phase 2 : Création de la base mySQL (échéance : 28/07/2025)

Phase 3 : Module transform (échéance : 30/07/2025)

Phase 4 : Module load (échéance : 31/07/2025)

Phase 5 : Tests et intégration (échéance : 18/08/2025)



Les livrables seront :

- Un script Python complet pour le processus ETL
- Une base SQL exportée (fichier .sql)
- Un jeu de fichiers CSV générés automatiquement
- Une documentation technique détaillée (PDF)
- Les fichiers seront remis via Dépôt Git.

Aucune maintenance post-livraison n'est prévue / Une maintenance de 4 semaines sera assurée pour les corrections.

Une documentation d'utilisation sera fournie pour faciliter la prise en main du système.

6.2 Aspect juridique et commerciaux

Respect du RGPD

Le projet a été conçu dans le respect des principes du Règlement Général sur la Protection des Données (RGPD). Aucune donnée personnelle directement identifiable (nom, prénom, email, etc.) n'est traitée dans le système. Les identifiants utilisés pour les revendeurs sont anonymisés dès la phase d'extraction. Les noms des régions sont utilisés uniquement pour vérifier la cohérence des données, mais ne sont pas exploités dans les autres phases du traitement, ni conservés dans les journaux (logs).

Les données sont stockées dans une base MySQL sécurisée, avec accès restreint à l'équipe projet. Les fichiers sources (CSV, SQLite) sont archivés uniquement à des fins de traçabilité, dans un dossier sécurisé local.

Le projet ne fait l'objet d'aucun transfert hors de l'Union Européenne.

Propriété

Tous les éléments développés (code, base, documentation) seront la propriété exclusive de Distributech.

Le code n'est pas destiné à une réutilisation ou distribution externe.

Aucun dépôt de licence ou de copyright n'est prévu.