

Chapitre 1

Cahier des charges du système décisionnel DataCo Supply Chain

1.1 Contexte et problématique

L'entreprise **DataCo** est un acteur international du e-commerce spécialisé dans la vente de multiples catégories de produits à travers plusieurs marchés géographiques. Elle dispose d'un volume important de données transactionnelles issues de sa chaîne logistique : commandes, clients, produits, livraisons, paiements et performance financière.

Actuellement, ces données sont stockées sous forme de fichiers plats, principalement le fichier *DataCoSupplyChainDataset.csv*, et exploitées de manière limitée pour le pilotage de l'activité. L'absence d'entrepôt de données structuré et d'outils décisionnels adaptés rend difficile l'analyse fine des ventes, de la rentabilité et de la performance logistique, ainsi que l'identification des axes d'amélioration opérationnelle.

1.2 Objectifs du projet

L'objectif général du projet est de concevoir et de mettre en place un système décisionnel complet permettant à l'entreprise DataCo de mieux analyser et piloter sa chaîne logistique.

Plus précisément, le projet vise à :

- Construire un entrepôt de données (Data Warehouse) reposant sur un modèle en étoile, alimenté à partir du fichier *DataCoSupplyChainDataset.csv*.
- Mettre en place un processus ETL (Extract–Transform–Load) pour extraire, nettoyer, transformer et charger les données sources vers le Data Warehouse.
- Fournir des tableaux de bord décisionnels (Power BI) permettant le suivi des principaux indicateurs de performance (KPIs) : ventes, bénéfices, retards de livraison, performance logistique, segmentation clients, etc.

- Offrir aux décideurs une vision consolidée et multidimensionnelle de l'activité (temps, client, produit, localisation, mode d'expédition).

1.3 Périmètre fonctionnel

Le périmètre fonctionnel du projet couvre les axes d'analyse suivants :

- **Ventes et rentabilité** : analyse du chiffre d'affaires, des quantités vendues et du bénéfice par commande, produit, catégorie, client, pays, segment et période (année, mois).
- **Clients et marchés** : analyse des segments de clientèle, des pays et villes les plus performants, identification des meilleurs marchés et des clients à forte valeur.
- **Produits** : analyse des produits et catégories les plus vendus, identification des produits les plus rentables ou les moins performants.
- **Logistique et livraison** : analyse des délais de livraison réels et planifiés, mesure des retards, suivi du statut de livraison et de la performance des modes d'expédition.
- **Analyse temporelle** : analyses par année, trimestre, mois, jour, permettant de suivre les tendances et la saisonnalité des ventes et des retards.

Les fichiers pris en compte sont :

- **DataCoSupplyChainDataset.csv** : fichier principal contenant les transactions (commandes, clients, produits, livraisons, ventes, profits).
- **DescriptionDataCoSupplyChain.csv** : dictionnaire de données utilisé pour comprendre le sens des champs et documenter les transformations.

Le fichier *tokenized_access_logs.csv* (logs techniques) est hors périmètre pour la première version du projet, mais pourra être exploité ultérieurement pour des analyses avancées.

1.4 Périmètre technique

Le périmètre technique comprend :

- La mise en place d'un schéma de Data Warehouse en étoile, composé d'une table de faits centrale et de plusieurs tables de dimensions (temps, client, produit, localisation, expédition).
- Le développement d'un processus ETL permettant :
 - l'extraction des données à partir des fichiers CSV ;
 - le nettoyage et la transformation des données (gestion des valeurs manquantes, formats de dates, types, calcul des indicateurs dérivés, etc.) ;
 - le chargement des données dans les tables de faits et de dimensions.

- L'utilisation d'un SGBD relationnel pour héberger le Data Warehouse (par exemple Oracle, PostgreSQL ou équivalent).
- La création de rapports et tableaux de bord avec Power BI connectés au Data Warehouse ou à un modèle analytique dérivé.

1.5 Description des données et besoins d'analyse

Les besoins d'analyse s'appuient sur les principaux groupes de données présents dans *DataCoSupplyChainDataset.csv*.

1.5.1 Données de commandes et ventes

- **Exemples de champs :** *Order Id, Order Date, Order Status, Order Item Quantity, Sales per customer, Benefit per order.*
- **Besoins :**
 - analyser le chiffre d'affaires par période, pays, produit, catégorie, segment client ;
 - mesurer la rentabilité (bénéfice) par commande, produit, catégorie, marché ;
 - suivre le volume de commandes et de lignes de commande.

1.5.2 Données clients

- **Exemples de champs :** *Customer Id, Customer City, Customer Country, Customer Segment.*
- **Besoins :**
 - analyser les ventes par pays, ville et segment de clientèle ;
 - identifier les clients et segments les plus rentables ;
 - cibler les marchés à fort potentiel.

1.5.3 Données produits

- **Exemples de champs :** *Product Name, Category Name, Product Price, Product Card Id.*
- **Besoins :**
 - identifier les produits et catégories les plus vendus ;
 - comparer la performance des catégories ;
 - analyser la rentabilité par produit et par catégorie.

1.5.4 Données logistiques et livraison

- **Exemples de champs :** *Days for shipping (real), Days for shipment (scheduled), Shipping Mode, Delivery Status, Late_delivery_risk.*

- **Besoins :**
 - mesurer les retards de livraison (écart entre délai réel et délai planifié) ;
 - suivre la performance logistique par mode d'expédition, pays, région ;
 - analyser la qualité de service via le statut de livraison et l'indicateur de risque de retard.

1.5.5 Données temporelles

- **Exemples de champs :** date de commande, année, mois, jour.
- **Besoins :**
 - analyser les tendances de ventes et de retards dans le temps ;
 - comparer les performances entre années, trimestres et mois ;
 - détecter les pics d'activité et la saisonnalité.

1.6 Contraintes et exigences non fonctionnelles

1.6.1 Qualité des données

- Les données doivent être nettoyées (gestion des doublons, des valeurs manquantes et des incohérences).
- Les formats de dates et de montants doivent être homogénéisés pour faciliter les agrégations.
- Les règles de calcul (par exemple, calcul du délai de livraison réel, du bénéfice, etc.) doivent être clairement définies et documentées.

1.6.2 Performance et volumétrie

- Le Data Warehouse doit être capable de supporter un volume de plusieurs centaines de milliers de transactions.
- Les requêtes d'analyse courantes (par période, produit, pays, segment client) doivent être exécutées dans des temps acceptables pour un usage décisionnel.

1.6.3 Évolutivité

- L'architecture doit permettre l'ajout futur de nouvelles sources de données (par exemple des logs d'accès ou d'autres systèmes opérationnels).
- Le modèle en étoile doit être extensible (ajout de nouvelles dimensions ou de nouveaux indicateurs dans la table de faits).

1.7 Livrables attendus

Les principaux livrables attendus dans le cadre de ce projet sont :

- Un cahier des charges fonctionnel et technique détaillé.
- Un schéma de Data Warehouse en étoile (modèle conceptuel et logique).
- Les scripts et procédures ETL pour le chargement des données.
- La base de données décisionnelle (tables de faits et de dimensions).
- Les rapports et tableaux de bord Power BI pour l'analyse des KPIs.
- Une documentation technique et utilisateur décrivant l'architecture, les règles de transformation et l'utilisation des tableaux de bord.

1.8 Critères de succès

Le projet sera considéré comme réussi si :

- le Data Warehouse permet d'interroger les ventes, les profits et la performance logistique selon plusieurs axes d'analyse (temps, client, produit, localisation, mode d'expédition) ;
- les tableaux de bord fournissent une vision claire et synthétique des principaux indicateurs de performance ;
- les utilisateurs métiers peuvent exploiter facilement les rapports pour la prise de décision ;
- les règles de transformation et les résultats obtenus sont cohérents avec les données sources.