TD1 - Enjeux des données en entreprise

Contexte : MODA & CO

MODA & CO est une enseigne fictive de prêt-à-porter de taille intermédiaire (50 magasins, 300 employés) qui possède les systèmes suivants :

Canaux de vente

- **Site e-commerce** (Magento) : commandes, paniers, navigation, avis clients — données stockées dans une base MySQL, exports quotidiens en CSV
- **50 magasins physiques** équipés de caisses (logiciel propriétaire) : tickets de caisse, remises, retours — synchronisation une fois par nuit, parfois en échec
- **Application mobile** (iOS/Android) : historique d'achats, géolocalisation, notifications push, programme de fidélité — données envoyées à un service tiers (Firebase)

Outils de gestion

- **CRM** (Salesforce) : fiches clients, historique des interactions, segments marketing — alimenté manuellement par les équipes commerciales **et** automatiquement par le site web (parfois en doublon)
- **Logiciel de gestion des stocks (WMS)** : niveaux de stock par entrepôt et par magasin, mis à jour toutes les 4 heures, avec des identifiants produits différents de ceux du site e-commerce (SKU non harmonisés)
- **ERP** pour la comptabilité et la facturation, mis à jour en fin de mois — c'est la seule source « officielle » validée par la finance

Marketing et relation client

- **Plateforme publicitaire** (Google Ads, Meta Ads) : coûts, clics, conversions — données disponibles via API, mais avec un délai d'attribution de 7 jours
- **Service client** (Zendesk) : tickets, réclamations, retours produits — texte libre, non structuré, difficile à croiser avec les ventes
- **Fichiers Excel** transmis par les équipes commerciales terrain : remises négociées localement, opérations promotionnelles ponctuelles, envoyés par email de façon irrégulière, avec des formats différents selon les régions

Un contexte volontairement problématique

- Le **nom du client** est orthographié différemment selon les systèmes (CRM, site web, magasin)
- Les **identifiants produits (SKU)** ne sont pas les mêmes entre le WMS, le site e-commerce et l'ERP
- Certaines données du CRM sont **en anglais**, d'autres en français (expansion internationale récente)
- Personne ne sait avec certitude **qui est responsable de la qualité des données clients**

Demande de la direction



1. Quel est le chiffre d'affaires réel ?
2. Quels clients sont les plus rentables ?



3. Quels produits risquent d'être en rupture ?
4. Quelle campagne marketing fonctionne ?
5. Peut-on proposer des recommandations personnalisées ?

Consigne

En sous-groupe (2-2-3), répondez aux séries de questions ci-dessous. Prenez le temps de discuter avant de formuler une réponse écrite courte pour chaque point.

Partie 1 — Repérage

1. Combien de systèmes différents stockent des données sur un même client ?
2. Combien de systèmes différents stockent des données sur un même produit ?
3. Si vous deviez dessiner un schéma avec toutes les flèches entre ces systèmes, à quoi ressemblerait-il ?

Partie 2 — Confrontation à une question métier

Reprenez la question : **« Quels clients sont les plus rentables ? »**

1. Dans quels systèmes se trouvent les informations nécessaires pour répondre ?
2. Comment sait-on que « Jean Dupont » dans le CRM est la même personne que « J. DUPONT » sur le site web et « Dupont Jean » en magasin ?
3. Si un client achète en magasin ET en ligne, ses achats sont-ils comptés une fois ou deux fois ?
4. Qui, dans l'entreprise, a la responsabilité de dire « ce chiffre est le bon » ?
5. Si deux services (marketing et finance) calculent chacun un chiffre d'affaires différent, lequel croyez-vous ?

Partie 3 — Confrontation au temps

1. L'ERP est mis à jour une fois par mois, l'app en temps réel, le WMS toutes les 4 heures : peut-on croiser ces données directement ?
2. Si un produit tombe en rupture de stock maintenant, dans combien de temps le saura-t-on vraiment ?
3. Une alerte marketing basée sur des données vieilles de 7 jours (délai d'attribution publicitaire) a-t-elle encore de la valeur ?
4. Qu'est-ce qui est plus important pour la direction : avoir une donnée exacte mais tardive, ou une donnée approximative mais immédiate ?

Partie 4 — Confrontation à la confiance

1. Un commercial envoie un fichier Excel avec une remise négociée à l'oral, jamais validée par personne : peut-on l'intégrer aux analyses ?
2. Un ticket de service client contient du texte libre, en français ou en anglais selon les cas : comment

l'exploiter automatiquement ?
3. Si l'entreprise achète une IA de recommandation produit, sur quelles données va-t-elle s'appuyer ? Sont-elles prêtes pour ça ?
4. Si vous deviez donner une note de confiance sur 10 à chacun des systèmes cités, laquelle donneriez-vous et pourquoi ?

Synthèse

En grand groupe, répondez à la question suivante :

> **Si toutes ces données existent déjà quelque part dans l'entreprise, pourquoi est-il si difficile de répondre aux 5 questions de la direction ?**

Essayez d'associer chacune des difficultés identifiées à un mot-clé parmi la liste suivante (sans les définir pour l'instant, juste par intuition) :

- Silo de données
- Référentiel unique / identité client
- Fraîcheur de la donnée
- Qualité de la donnée
- Gouvernance des données
- Intégration de données

TD2 - Les briques d'une architecture de données

From:
http://slamwiki2.kobject.net/ - **SlamWiki 2.1**

Permanent link:
http://slamwiki2.kobject.net/eadl/bloc5/fm1/td1

Last update: **2026/08/30 18:43**