AUDIT: Erreur sur le Cube OLAP

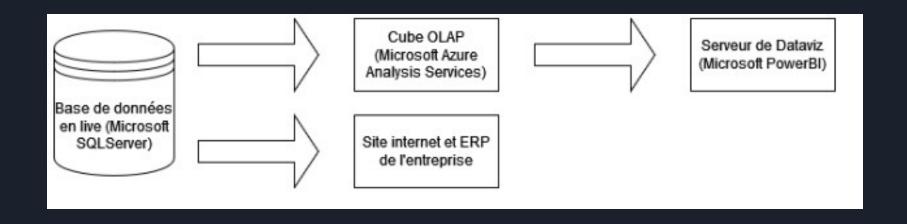


Présentation de la mission

La société SuperSmartMarket a trouvé des incohérences dans les données qu'elle utilise. Il faut:

- Vérifier les données
- Corriger les incohérences
- Faire un rapport

Analyse de la base de données Architecture de l'entreprise



Analyse de la base de données Dictionnaire des données du cube OLAP actuel

DICTIONNAIRE DES DONNÉES - Ventes

CODE	SIGNIFICATION	TYPE	LONGUEUR	NATURE	REGLE DE GESTION	REGLE DE CALCUL
id_vente	ID dans la base de données [PK]	Varchar	35	Elémentaire	Ne doit pas être nul	
id_client	ID du client [FK]	Varchar	17	Elémentaire	Ne doit pas être nul	
id_employe	ID de l'employé [FK]	Varchar	32	Elémentaire	Ne doit pas être nul	
ean	Code produit suivant la norme internationale [FK]	Varchar	13	Elémentaire	Ne doit pas être nul	
id_date	ID de la date d'achat [FK]	Integer	NC	Elémentaire	Ne doit pas être nul	
id_ticket	ID du ticket de caisse	Varchar	6	Elémentaire	Ne doit pas être nul	

DICTIONNAIRE DES DONNÉES - Produits

CODE	SIGNIFICATION	TYPE	LONGUEUR	NATURE	REGLE DE GESTION	REGLE DE CALCUL
ean	Code produit suivant la norme internationale [PK]	Varchar	13	Elémentaire	Ne doit pas être nul	
categorie	Catégorie du produit	Varchar	128	Elémentaire	Ne doit pas être nul	
rayon	Rayon du produit	Varchar	128	Elémentaire	Ne doit pas être nul	
libelle	Libellé du produit	Varchar	128	Elémentaire	Ne doit pas être nul	
prix	Prix du produit	Float	NC.	Flémentaire	Ne doit pas être nul	

DICTIONNAIRE DES DONNÉES - Clients

CODE	SIGNIFICATION	TYPE	LONGUEUR	NATURE	REGLE DE GESTION	REGLE DE CALCUL
id_client	ID du clent [PK]	Varchar	17	Elémentaire	Ne doit pas être nul	
date_inscription	Date d'inscription	Varchar	10	Elémentaire	Ne doit pas être nul	

DICTIONNAIRE DES DONNÉES - Calendrier

CODE	SIGNIFICATION	TYPE	LONGUEUR	NATURE	REGLE DE GESTION	REGLE DE CALCUL
id_date	ID de la date [PK]	Integer	NC	Elémentaire	Ne doit pas être nul	
annee	Année	Integer	NC	Elémentaire	Ne doit pas être nul	
mois	Mois de l'année	Integer	NC	Elémentaire	Ne doit pas être nul	
jour	Jour du mois	Integer	NC	Elémentaire	Ne doit pas être nul	
mois_nom	Nom du mois	Varchar	9	Elémentaire	Ne doit pas être nul	
jour_semaine	Jour de la semaine	Integer	NC	Elémentaire	Ne doit pas être nul	
annee_mois	Mois de l'année	Integer	NC	Elémentaire	Ne doit pas être nul	
trimestre	Trimestre de l'année	Varchar	2	Elémentaire	Ne doit pas être nul	

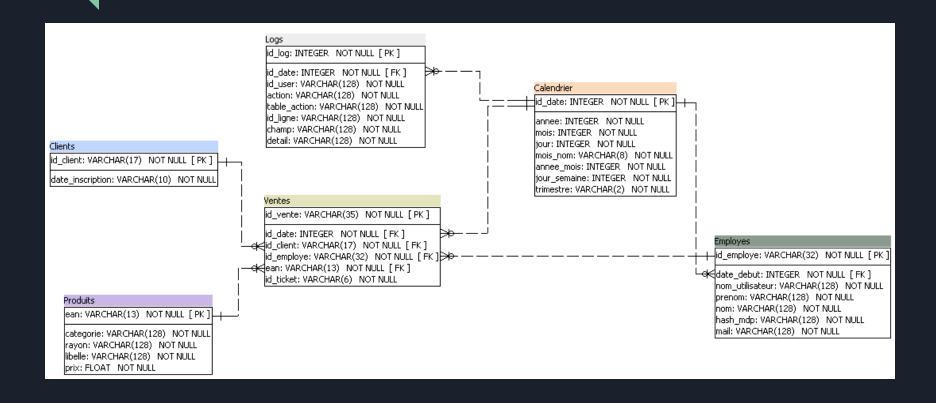
DICTIONNAIRE DES DONNÉES - Employés

CODE	SIGNIFICATION	TYPE	LONGUEUR	NATURE	REGLE DE GESTION	REGLE DE CALCUL
id_employe	ID de l'employé [PK]	Varchar	32	Elémentaire	Ne doit pas être nul	
nom_utilisateur	Nom d'utilisateur	Varchar	128	Elémentaire	Ne doit pas être nul	
prenom	Prénom de l'employé	Varchar	128	Elémentaire	Ne doit pas être nul	
nom	Nom de l'employé	Varchar	128	Elémentaire	Ne doit pas être nul	
date_debut	Date d'embauche [FK]	Integer	NC	Elémentaire	Ne doit pas être nul	
hash_mdp	Hash du mot de passe de l'employé	Varchar	128	Elémentaire	Ne doit pas être nul	
mail	Mail de l'employé	Varchar	128	Elémentaire	Ne doit pas être nul	

DICTIONNAIRE DES DONNÉES - Logs

CODE	SIGNIFICATION	TYPE	LONGUEUR	NATURE	REGLE DE GESTION	REGLE DE CALCUL
id_log	ID de la date de l'action [PK]	Integer	NC	Elémentaire	Ne doit pas être nul	AUTOINCREMENT
id_user	ID de l'utilisateur	Varchar	128	Elémentaire	Ne doit pas être nul	
id_date	ID de la date de l'action [FK]	Integer	NC	Elémentaire	Ne doit pas être nul	
action	Action effectuée	Varchar	128	Elémentaire	Ne doit pas être nul	
table_action	Table impactée par l'action	Varchar	128	Elémentaire	Ne doit pas être nul	
id_ligne	ID de la ligne impactée	Varchar	128	Elémentaire	Ne doit pas être nul	
champ	Nom du champ impacté	Varchar	128	Elémentaire	Ne doit pas être nul	
detail	Valeur du champ impacté	Varchar	128	Elémentaire	Ne doit pas être nul	

Analyse de la base de données Schéma relationnel



Présentation du prototype de BDD Explication de la base de données

- Prototypage basé sur le jeu de données de la feuille Excel fournie
- Une table de faits : « Ventes »
- 4 tables de dimension : « Produits», « Clients », « Calendrier »,
 «Employes »
- 1 table de logs: « Logs»

Résultat des requêtes SQL Le chiffre d'affaires total pour le 14 Août à partir des logs

```
"ld dates"
          id date as le 14
         "Calendrier" \
                              annee=2024 /
                                                mois=8
                                                             jour=14
"Transactions date valeur le 14" AS (
          o.id date
         "Logs" o, "Id dates" d
                                  o.table insert='Ventes' A
           o.action='INSERT' Al
                                                              o.champs='Date' AND CAS
                                                                                           (o.detail AS INT)=d.le 14
              o.id date
"Chiffres affaires"
                          E '1899-12-30' + t.id_date) AS date, ROUND(CAST(SUM(p.prix) AS
                                                                                            NUMERIC), 2)
         "Logs" o, "Produits" p, "Transactions_date_valeur_le_14" t
           o.action='INSERT' AND
                                  o.table insert='Ventes' AND
                                                               o.champs='EAN' AND o.detail=p.ean A
                                                                                                      VD o.id date=t.id date
              date
              date
                               "Date d'enregistrement", ca AS "Chiffre d'affaires enregistré pour le 14 août 2024"
              (date as text)
       "Chiffres affaires"
        'Total'.
                    (ca)
       "Chiffres affaires"
```

Résultat des requêtes SQL Le chiffre d'affaires total pour le 14 Août à partir des logs

Date d'enregistrement	Chiffre d'affaires enregistré pour le 14 août 2024
2024-08-14	275186.59
2024-08-15	9057.29
Total	284243.88

Résultat des requêtes SQL Le chiffre d'affaires par client pour le top 10 des clients

```
SELECT v.id_client, ROUND(SUM(p.prix), 2) AS ca_par_client
FROM calendrier c, ventes v, produits p
WHERE c.id_date=v.id_date
    AND v.ean=p.ean
    AND c.annee=2024
    AND c.jour=14
    AND c.mois=8
GROUP BY v.id_client
```

ORDER BY ca par client DESC LIMIT 10

Client	Chiffre d'affaires (euros)
CUST-JNSOZSFORR88	846.86
CUST-GM6VBAYAB8SF	666.86
CUST-L2ST2JHI7K9O	644.18
CUST-WU7ZKQJE4L17	608.93
CUST-9WM83101QDTI	582.03
CUST-ZMAOVX8XYGJY	576.39
CUST-3K66CV0OHH7Q	571.44
CUST-CG23SXJDRNYR	531.09
CUST-D8IOFHVUFX3Y	477.35
CUST-IHN1HQRI7PYJ	463.73

Résultat des requêtes SQL Le chiffre d'affaires encaissé par employé

```
WITH ca par employe AS (
      SELECT v.id employe AS id employe, ROUND(SUM(p.prix), 2) AS
ca
      FROM calendrier c, ventes v, produits p
      WHERE c.id date=v.id date
         AND v.ean=p.ean
         AND c.annee=2024
         AND c.jour=14
         AND c.mois=8
      GROUP BY v.id employe
      ORDER BY ca DESC
SELECT CONCAT(SUBSTR(e.prenom,1,1),'. ',e.nom), cpe.ca
FROM employe e, ca par employe cpe
WHERE e.id employe=cpe.id employe
```

Résultat des requêtes SQL Le chiffre d'affaires encaissé par employé

Employé	CA (€)	Employé	CA (€)	Employé	CA (€)	Employé	CA (€)	Employé	CA (€)	Employé	CA (€)	Employé	CA (€)
A. Boulet	7818.82	P. Rodier	6094.62	A. Donnet	5513.15	G. Granet	5187.01	C. Vérany	4613.51	J. Leloup	4223.81	J. Vasseur	2798.76
E. Jacquier	7736.16	S. Deslys	5989.84	S. Jacquet	5490.3	J. Auch	5146.52	P. Beaumont	4597.93	L. Rochefort	4150.37	R. Bechard	2726.18
C. Delisle	6995.14	C. Gachet	5936.36	S. Pélissier	5490.26	A. Besson	5128.45	L. Boissonade	4577.33	D. Courtet	3902.65		
P. Manoury	6616.46	V. Baume	5909.73	J. Courbet	5450.27	A. Coquelin	5013.37	N. Picard	4512.28	A. Vannier	3852.21		
T. Arsenault	6483.84	B. Cazal	5827.75	E. Grosjean	5363.58	L. Maret	4941.48	B. Boutroux	4426.2	S. Grosjean	3851.0		
A. Dufresne	6361.22	I. Escoffier	5734.83	V. Genet	5345.16	V. Jacquier	4775.66	J. Beauvau	4410.01	N. Frère	3728.54		
A. Dutertre	6133.34	É. Blanchard	5705.28	Y. Grinda	5315.51	R. Rémy	4658.91	C. Morin	4395.59	E. Lussier	3558.79		
P. Ange	6112.37	U. Chevalier	5567.78	A. Marchal	5220.68	M. Poincaré	4638.77	S. Valluy	4333.48	D. Cazenave	3170.83		
A. Lièvremo nt	6111.18	S. Giraud	5557.84	M. Rochefort	5187.41	G. Bonhomme	4621.61	A. Bruneau	4232.0	É. Bissonnette	3001.75		

Analyse des logs

- Nombre de logs: 207399
- Nombre de logs par action:
 - > DELETE: 2
 - > INSERT: 206905
 - > UPDATE: 582
- Explication des logs:
 - Table Clients: 20 INSERT (ajout nouveaux clients)
 - Table Employés: 2 DELETE, 7 UPDATE (modification mots de passe)
 - Table Produits: 575 UPDATE (modification prix)
 - Table Ventes: 206885 INSERT (ajout information ventes: date, article, client, ticket)

Compréhension des problèmes sur la base des logs

- Problème d'exploitation des données :
 - Certaines ventes du jour J ont été intégrées J+1.
 - Les logs ne sont pas facilement exploitables
- Problème d'attributions des rôles : L'ID user qui opère sur les produits, les ventes, les employés est toujours le même.
- Problème de cohérence de données : Des prix ont un format date dans l'extraction Excel. Les codes EAN des produits ont une taille qui varie (13 ou 12 chiffres).
- Problème d'historique : Des employés ont été supprimés.

Recommandations pour l'entreprise

- Monitoring des données: utilisation d'un outil comme Grafana pour monitorer le chiffre d'affaires.
- Toutes les tables : Spécifier des clés étrangères pour la COHÉRENCE des données.
- Table « Produits » :
 - o S'assurer que les codes EAN des produits soient à la même norme (EAN 13 en Europe).
- Table « Employés » :
 - Ne pas pouvoir supprimer les employés pour la COHÉRENCE des données et garder l'historique des ventes.
 - o Ajouter une colonne indiquant le niveau de privilèges de l'employé pour la SÉCURITÉ.
 - o Ajouter une colonne informant si l'employé est toujours dans l'entreprise.
 - o Ajouter une colonne donnant la date de fin de contrat.
- Les logs doivent être divisés : 1 table logs associée à chaque table Ventes, Produits, Employés, Calendrier et Clients.
- Reconstruction des tables à partir des logs.

Prototype Nouveau dictionnaire des données

DICTIONNAIRE DES DONNÉES - Ventes

CODE	SIGNIFICATION	TYPE	LONGUEUR	NATURE	REGLE DE GESTION	REGLE DE CALCUL
id_vente	ID dans la base de données [PK]	Varchar	35	Elémentaire	Ne doit pas être nul	
id_client	ID du client [FK]	Varchar	17	Elémentaire	Ne doit pas être nul	
id_employe	ID de l'employé [FK]	Varchar	32	Elémentaire	Ne doit pas être nul	
ean	Code produit suivant la norme internationale [FK]	Varchar	13	Elémentaire	Ne doit pas être nul	
id_date	ID de la date d'achat [FK]	Integer	NC	Elémentaire	Ne doit pas être nul	
id_ticket	ID du ticket de caisse	Varchar	6	Elémentaire	Ne doit pas être nul	

DICTIONNAIRE DES DONNÉES - Produits

CODE	SIGNIFICATION	TYPE	LONGUEUR	NATURE	REGLE DE GESTION	REGLE DE CALCUL
ean	Code produit suivant la norme internationale [PK]	Varchar	13	Elémentaire	Ne doit pas être nul, Doit faire 13 caractères exactement	
categorie	Catégorie du produit	Varchar	128	Elémentaire	Ne doit pas être nul	
rayon	Rayon du produit	Varchar	128	Elémentaire	Ne doit pas être nul	
libelle	Libellé du produit	Varchar	128	Elémentaire	Ne doit pas être nul	
prix	Prix du produit	Float	NC	Elémentaire	Ne doit pas être nul	

DICTIONNAIRE DES DONNÉES - Clients

CODE	SIGNIFICATION	TYPE	LONGUEUR	NATURE	REGLE DE GESTION	REGLE DE CALCUL
id_client	ID du clent [PK]	Varchar	17	Elémentaire	Ne doit pas être nul	
id_date_inscription	ID date [FK]	Integer	NC	Elémentaire	Ne doit pas être nul	

DICTIONNAIRE DES DONNÉES - Calendrier

CODE	SIGNIFICATION	TYPE	LONGUEUR	NATURE	REGLE DE GESTION	REGLE DE CALCUL
id_date	ID de la date [PK]	Integer	NC	Elémentaire	Ne doit pas être nul	
annee	Année	Integer	NC	Elémentaire	Ne doit pas être nul	
mois	Mois de l'année	Integer	NC	Elémentaire	Ne doit pas être nul	
jour	Jour du mois	Integer	NC	Elémentaire	Ne doit pas être nul	
mois_nom	Nom du mois	Varchar	9	Elémentaire	Ne doit pas être nul	
jour_semaine	Jour de la semaine	Integer	NC	Elémentaire	Ne doit pas être nul	
annee_mois	Mois de l'année	Integer	NC	Elémentaire	Ne doit pas être nul	
trimestre	Trimestre de l'année	Varchar	2	Elémentaire	Ne doit pas être nul	

DICTIONNAIRE DES DONNÉES - Employés

CODE	SIGNIFICATION	TYPE	LONGUEUR	NATURE	REGLE DE GESTION	REGLE DE CALCUL
id_employe	ID de l'employé [PK]	Varchar	32	Elémentaire	Ne doit pas être nul, ne peut pas être supprimé	
nom_utilisateur	Nom d'utilisateur	Varchar	128	Elémentaire	Ne doit pas être nul	
prenom	Prénom de l'employé	Varchar	128	Elémentaire	Ne doit pas être nul	
nom	Nom de l'employé	Varchar	128	Elémentaire	Ne doit pas être nul	
date_debut	Date d'embauche [FK]	Integer	NC	Elémentaire	Ne doit pas être nul	
date_fin	Date de fin de contrat [FK]	Integer	NC	Elémentaire		
hash_mdp	Hash du mot de passe de l'employé	Varchar	128	Elémentaire	Ne doit pas être nul	
mail	Mail de l'employé	Varchar	128	Elémentaire	Ne doit pas être nul	
privilege	Rôle de l'utilisateur: administrateur, utilisateur	Varchar	50	Elémentaire		
actif	L'employé est-il encore dans l'entreprise	Boolean	NC	Elémentaire		

DICTIONNAIRE DES DONNÉES - Logs Ventes

CODE	SIGNIFICATION	TYPE	LONGUEUR	NATURE	REGLE DE GESTION	REGLE DE CALCUL
id_log	ID du log [PK]	Integer	NC	Elémentaire	Ne doit pas être nul	AUTOINCREMENT
date	Timestamp du log	Timestamp	NC	Elémentaire	Ne doit pas être nul	
id_user	ID de l'utilisateur qui a effectué l'action	Varchar	128	Elémentaire	Ne doit pas être nul	
action	Action effectuée: INSERT, UPDATE, DELETE	Varchar	10	Elémentaire	Ne doit pas être nul	
id_vente	ID dans la base de données	Varchar	35	Elémentaire	Ne doit pas être nul	
id_client	ID du client	Varchar	17	Elémentaire	Ne doit pas être nul	
id_employe	ID de l'employé	Varchar	32	Elémentaire	Ne doit pas être nul	
ean	Code produit suivant la norme internationale	Varchar	13	Elémentaire	Ne doit pas être nul	
id_date	ID de la date d'achat	Integer	NC	Elémentaire	Ne doit pas être nul	
id_ticket	ID du ticket de caisse	Varchar	6	Elémentaire	Ne doit pas être nul	

DICTIONNAIRE DES DONNÉES - Logs Produits

CODE	SIGNIFICATION	TYPE	LONGUEUR	NATURE	REGLE DE GESTION	REGLE DE CALCUL
id_log	ID du log [PK]	Integer	NC	Elémentaire	Ne doit pas être nul	AUTOINCREMENT
date	Timestamp du log	Timestamp	NC	Elémentaire	Ne doit pas être nul	
id_user	ID de l'utilisateur qui a effectué l'action	Varchar	128	Elémentaire	Ne doit pas être nul	
action	Action effectuée: INSERT, UPDATE, DELETE	Varchar	10	Elémentaire	Ne doit pas être nul	
ean	Code produit suivant la norme internationale	Varchar	13	Elémentaire	Ne doit pas être nul	
categorie	Catégorie du produit	Varchar	128	Elémentaire	Ne doit pas être nul	
rayon	Rayon du produit	Varchar	128	Elémentaire	Ne doit pas être nul	
libelle	Libellé du produit	Varchar	128	Elémentaire	Ne doit pas être nul	
prix	Prix du produit	Float	NC	Elémentaire	Ne doit pas être nul	

DICTIONNAIRE DES DONNÉES - Logs Clients

CODE	SIGNIFICATION	TYPE	LONGUEUR	NATURE	REGLE DE GESTION	REGLE DE CALCUL
id_log	ID du log [PK]	Integer	NC	Elémentaire	Ne doit pas être nul	AUTOINCREMENT
date	Timestamp du log	Timestamp	NC	Elémentaire	Ne doit pas être nul	
id_user	ID de l'utilisateur qui a effectué l'action	Varchar	128	Elémentaire	Ne doit pas être nul	
action	Action effectuée: INSERT, UPDATE, DELETE	Varchar	10	Elémentaire	Ne doit pas être nul	
id_client	ID du clent	Varchar	17	Elémentaire	Ne doit pas être nul	
id_date_inscription	ID date	Integer	NC	Elémentaire	Ne doit pas être nul	

DICTIONNAIRE DES DONNÉES - Logs Employés

CODE	SIGNIFICATION	TYPE	LONGUEUR	NATURE	REGLE DE GESTION	REGLE DE CALCUL
id_log	ID du log [PK]	Integer	NC	Elémentaire	Ne doit pas être nul	AUTOINCREMENT
date	Timestamp du log	Timestamp	NC	Elémentaire	Ne doit pas être nul	
id_user	ID de l'utilisateur qui a effectué l'action	Varchar	128	Elémentaire	Ne doit pas être nul	
action	Action effectuée: INSERT, UPDATE, DELETE	Varchar	10	Elémentaire	Ne doit pas être nul	
id_employe	ID de l'employé [FK]	Varchar	32	Elémentaire	Ne doit pas être nul	
nom_utilisateur	Nom d'utilisateur	Varchar	128	Elémentaire	Ne doit pas être nul	
prenom	Prénom de l'employé	Varchar	128	Elémentaire	Ne doit pas être nul	
nom	Nom de l'employé	Varchar	128	Elémentaire	Ne doit pas être nul	
date_debut	Date d'embauche [FK]	Integer	NC	Elémentaire	Ne doit pas être nul	
date_fin	Date de fin de contrat [FK]	Integer	NC	Elémentaire		
hash_mdp	Hash du mot de passe de l'employé	Varchar	128	Elémentaire	Ne doit pas être nul	
mail	Mail de l'employé	Varchar	128	Elémentaire	Ne doit pas être nul	
privilege	Rôle de l'utilisateur: administrateur, utilisateur	Varchar	50	Elémentaire		
actif	L'employé est-il encore dans l'entreprise	Boolean	NC	Elémentaire		

DICTIONNAIRE DES DONNÉES - Logs Calendrier CODE SIGNIFICATION TYPE NATURE REGLE DE GESTION REGLE DE CALCUL LONGUEUR id log ID du log [PK] Integer NC Elémentaire Ne doit pas être nul AUTOINCREMENT date Timestamp du log Timestamp NC Elémentaire Ne doit pas être nul id user ID de l'utilisateur qui a effectué l'action Ne doit pas être nul Varchar 128 Elémentaire Ne doit pas être nul action Action effectuée: INSERT, UPDATE ou DELETE Varchar 6 Elémentaire id date NC Ne doit pas être nul ID de la date Integer Elémentaire Année annee Integer NC Elémentaire Ne doit pas être nul Mois de l'année Ne doit pas être nul mois Integer NC Elémentaire jour Jour du mois Integer NC Elémentaire Ne doit pas être nul Jour de la semaine NC Ne doit pas être nul jour_semaine Integer Elémentaire Nom du mois mois nom Varchar 8 Elémentaire Ne doit pas être nul

2

Elémentaire

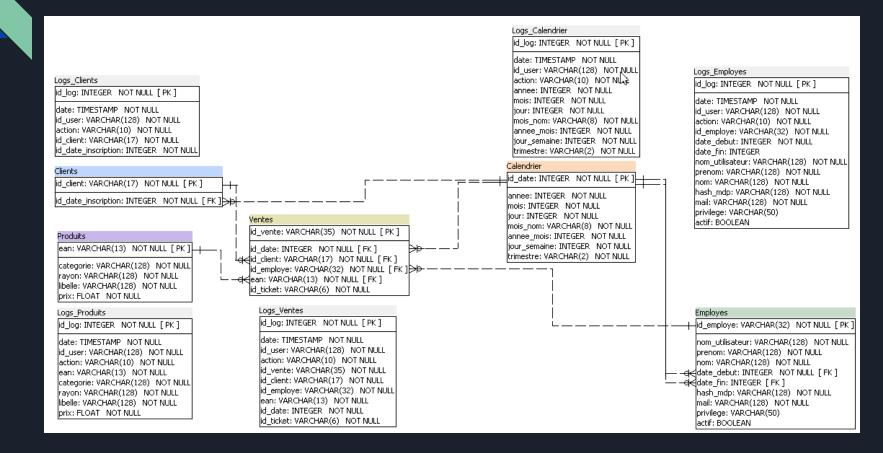
Ne doit pas être nul

Varchar

trimestre

Trimestre de l'année

Nouveau Schéma relationnel



Exemple de fonction trigger insérant un log après un INSERT client avec Microsoft SQL Server

```
CREATE TRIGGER trg_AfterInsert_Client
ON Clients
AFTER INSERT
AS
BEGIN
SET NOCOUNT ON;
INSERT INTO Logs_Clients
SELECT GETDATE(), USER_ID, 'INSERT', * FROM inserted;
END
```

Exemple de fonction trigger empêchant la suppression d'un employé avec Microsoft SQL Server

```
CREATE TRIGGER trg_BlockEmployesDeletion
ON Employes
AFTER DELETE
AS
BEGIN
RAISERROR('Suppression d'employé non autorisée.', 16, 1);
ROLLBACK TRANSACTION;
END
```

Contrôle du champ EAN dans la table produit

Exemple d'un rendu graphique généré par Grafana



Recommandations pour l'entreprise Avantages et inconvénients de la solution

Avantages:

- Meilleure cohérence des données.
- Meilleure sécurité.
- Monitoring régulier.

Inconvénients:

- Gestion des contraintes plus complexe.
- Mise en place de nouvelles procédures.

Recommandations pour l'entreprise Implications pour l'organisation actuelle

- Travailler sur la base de données en intégrant les contraintes.
- Mettre en place le monitoring avec Grafana.
- Mettre en place de nouvelles procédures de sécurité impliquant de définir les rôles et privilèges de chaque employé.
- Former les employés à ces nouvelles procédures.

Recommandations pour l'entreprise Planning de mise en place et coût éventuel

Tâches	Durée	Nombre de personnes y travaillant
Définition des rôles et privilèges de chaque employé	2 semaines	2
Refonte de la base de données pour intégrer les nouvelles colonnes et contraintes	2 semaines	2
Mise en place du monitoring avec Grafana	2 semaines	2
Tests	2 semaines	4
Formation des employés	3 semaines	2

Conclusion

En appliquant de nouvelles contraintes, en définissant de nouvelles règles de sécurité et en mettant en place une solution de monitoring, la résilience de la base de données sera améliorée.