

RAPPORT D'AUDIT

Mathilde LE SOLLIEC

2024-09-11

Table des matières

1. Résumé Exécutif	3
2. Introduction	4
3. Description de l'Entité Auditée	5
Présentation de l'organisation	5
Systèmes audités	5
4. Constats de l'Audit	6
Constat 1 – Décalage de l'insertion des données dans le cube OLAP	6
Constat 2 – Table de logs limitée	6
Évaluation des risques	6
5. Analyse des Résultats	7
6. Recommandations et plan d'action	8
Etape 1	8
1. Améliorer la table de log	8
2. Configurer l'intégrité des données :	8
3. Vérifier les contraintes d'intégrités	8
Etape 2	9
1. Mise en place d'un suivi des flux / dashboard	9
2. Mise en place de triggers	9
7. Annexes	10
Documents et preuves	10
Glossaire	11
Références	11

1. Résumé Exécutif

Objectifs de l'audit

L'audit a été mené pour identifier les causes d'un décalage de chiffre d'affaires constaté le 14-08-2024 dans le cube OLAP.

L'objectif est pour proposer des mesures visant à garantir la sécurité et la qualité des données de l'entreprise.

Conclusions principales

- 1 379 ventes ont été insérées dans Azure Analysis Services avec un jour de retard, générant un écart de 9 057,29 €.
- La table de logs actuelle ne capture pas toutes les informations nécessaires pour tracer et corriger les anomalies.
- Le monitoring des flux est insuffisant et les contraintes d'intégrité ne sont pas exploitées pour déclencher des alertes.

Recommandations principales

- Améliorer la table de logs avec les champs Statut, heures exactes du job, et valeurs Avant/Après.
- Utiliser des transactions atomiques sur les lots.
- Déployer un monitoring global sur les flux
- Mettre en place des Triggers pour anomalies critiques.

Portée de l'audit

L'audit évalue la cohérence des données remontées entre source SQL Server et cube OLAP, ainsi que la qualité des logs.

2. Introduction

Contexte

L'audit a été initié suite à l'identification d'un écart de chiffre d'affaires sur la date du 14-08-2024, où 1 379 ventes ont été insérées dans le cube OLAP un jour plus tard, générant un manque de 9 057,29 €.

L'objectif est de comprendre les causes de ce décalage et de proposer des mesures pour sécuriser le flux ETL et la remontée de données dans le cube OLAP.

Objectifs de l'audit

- Vérifier la fiabilité et la complétude des données remontées dans le cube OLAP.
- Identifier les causes des écarts de chiffre d'affaires.
- Proposer des recommandations pour sécuriser le flux ETL, améliorer le logging et le monitoring.

Périmètre de l'audit

- Base de données en live (SQLServer)
- Cube OLAP dans Azure Analysis Services (table Ventes, Produits, Clients, Employés)
- Table de logs existante et processus de monitoring associé.

Méthodologie

- Analyse de l'architecture de données
- Création d'un prototype de base de données en local (PostgreSQL) pour tester les données du cube OLAP.
- Analyse des logs.

3. Description de l'Entité Auditée

Présentation de l'organisation

Le service est le pôle Business Intelligence mobilise les informations sur les ventes, produits, clients et employés pour suivre les performances de l'entreprise.

Systèmes audités

- **SQL Server** : base relationnelle pour stocker les données brutes
 - **Cube OLAP (Azure Analysis Services)** : alimenté par la base de données SQLServer, le cube OLAP permet de préparer et agréger les données qui seront utilisées pour l'analyse.
 - **Microsoft PowerBI** : visualise les données sous forme de tableaux de bord.
-

4. Constats de l'Audit

Constat 1 – Décalage de l'insertion des données dans le cube OLAP

- **Description** : 1 379 ventes réalisés le 14-08-2024 sont insérées dans le cube OLAP le 15-08-2024.
- **Impact** : Écart de 9 057,29 € sur le chiffre d'affaires observé le 14-08-2024.
- **Preuve** : Analyse des logs

Constat 2 – Table de logs limitée

- **Description** : La table de logs ne capture pas toutes les informations nécessaires pour diagnostiquer les anomalies (ex. statut, heures du jobs).
- **Impact** : Difficile de retracer les causes des rejets et retards de chargement.
- **Preuve** : Analyse de la structure existante de la table de logs

Évaluation des risques

- **Décalage de données** : risque élevé pour la fiabilité du reporting et la prise de décision.
- **Logs incomplets** : risque moyen, rend le diagnostic plus long et augmente la probabilité de récurrence d'erreurs.

5. Analyse des Résultats

Analyse des écarts

- Les données ne sont pas chargées au moment attendu dans le cube OLAP.
- Les pratiques actuelles ne garantissent pas que toutes les anomalies soient loguées et prises en charge.

Hypothèse :

Plusieurs raisons pourraient expliquer le décalage d'intégration de certaines données dans le cube :

- Ressources insuffisantes : un manque de mémoire de la base SQLserver pour ingérer un gros volume de données peut avoir pour conséquences que certaines insertions échouent ou prennent beaucoup de temps et provoque un décalage dans le cube
- Taille excessive des données : Si un lot est très volumineux, il peut dépasser les limites du serveur ou provoquer un timeout.

La table LOG ne donne pas suffisamment d'information et de suivi pour s'assurer que ce type d'erreur ne se reproduise pas.

Plusieurs recommandations peuvent être évoquées pour résoudre cela et avoir une meilleure prise en charge des erreurs.

6. Recommandations et plan d'action

Recommandations

Etape 1

Les éléments suivants peuvent être introduit rapidement dans le process de l'entreprise :

1. Améliorer la table de log

- **Objectif** : avoir une trace des erreurs de chargement des lignes qui ne sont pas chargées vers le cube OLAP – centralisés les anomalies
- **Comment** : ajouter une colonne statut (rejetée, réussie, bloquée), et les détails sur les jobs dates et heures.
- **Outil** : SQLServer (table de log)

2. Configurer les transaction et l'isolation

- **Objectif** : préserver l'intégrité des données : si une étape échoue, tout le bloc est annulé
- **Comment** : paramétrage des lots pour que les transactions soient atomiques
- **Outil** : SQL Server – gestion des transactions.

3. Vérifier les contraintes d'intégrités

- **Objectif** : Empêchent qu'une donnée incohérente entre dans la base ou passe le flux
- **Comment** : paramétrage SQL : PK, FK, CHECK, NOT NULL, etc.
- **Outil** : SQL Server – contraintes intégrité

4. Sécuriser la gestion des droits d'administration

- **Objectif** : limiter les risques liés aux modifications non contrôlées dans SQL Server et le cube OLAP.
- **Comment** : Gérer l'accès des droit avec l'authentification , limiter les permissions attribuée, définir des niveaux (administrateur –lecteur – writer – admin...) , tracer les actions via un audit log
- **Outil** : SQL Server Management Studio

Etape 2

Dans un second temps, il est proposé d'intégrer un meilleur suivi des flux de données :

1. Mise en place d'un outil de monitoring des flux ETL

- **Objectif** : visualiser les echec de job, le temps de traitement des jobs et des écarts entre source et cube
- **Comment** : connecter les logs et métriques ADF pour créer un suivi automatisé.
- **Outil** : Azure Data Factory (ADF)

2. Mise en place de triggers

- **Objectif** : alerter dès qu'une erreur survient
- **Comment** : associer un trigger (ex : envoi d'un email à l'équipe) si les contraintes définies ne sont pas respectées
- **Outil** : SQL Server – Triggers + SQL Server Agent Jobs pour notifications.

Plan d'action

Étape	Délai	Action	Objectif	Outil
1	2 sem.	Améliorer Table de log	Centraliser les erreurs	SQL Server
		Transactions atomique	Evite les incohérence	SQL Server
		Contraintes d'intégrité	Bloquer données incohérentes	SQL Server
		Gestion des droits	Gérer les accès et suivre les modifications	SQL Server Management Studio
2	1 mois	Monitoring ETL	Suivre échecs et écarts	Azure Data Factory
		Triggers	Alerter sur erreurs	SQL Server + Agent Jobs

7. Annexes

Documents et preuves

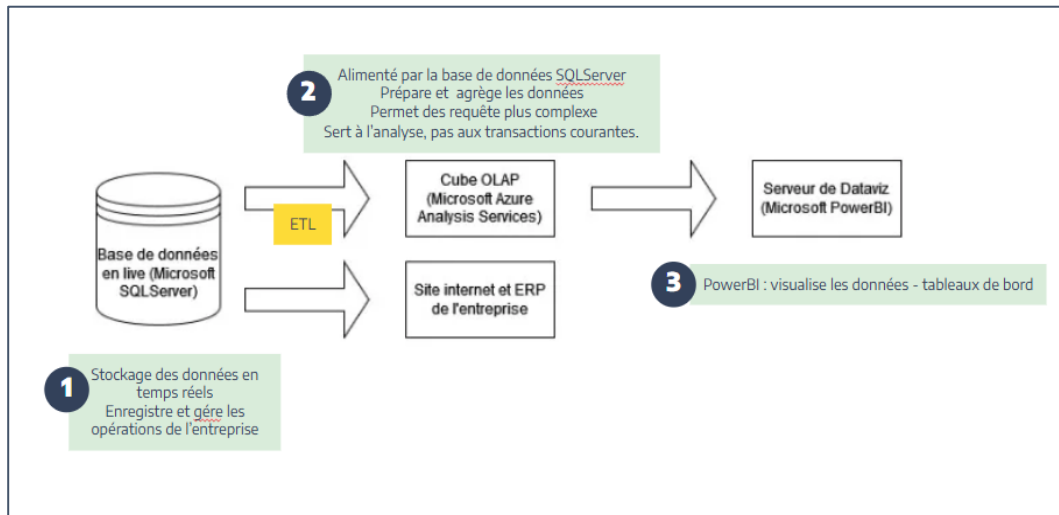


Figure 1 schéma de l'architecture de donnée de l'entreprise

logs
Columns (8)
logs_pk
id_user
date_action
type_action
table_insert
id_ligne
champs
detail

Figure 2 structure de la table logs actuelle

date	action	table_insert	champs	detail
15/08/2024	INSERT	Ventes	CUSTOMER_ID	CUST-B8W3YSU9HIUX
15/08/2024	INSERT	Ventes	id_employe	951e47dfa4c5298382d1b7d75
15/08/2024	INSERT	Ventes	EAN	6 386 228 298 857
15/08/2024	INSERT	Ventes	Date	14/08/2024

Figure 3 Extrait de la table de log actuelle : date d'insertion de la vente différente de la date de la vente

Glossaire

- Cube OLAP (Online Analytical Processing)
Base de données multidimensionnelle organisée en tables qui permet de traiter et d'analyser plusieurs dimensions de données beaucoup plus rapidement et efficacement qu'une base de données relationnelle.
- ETL (Extract, Transform, Load)
Processus qui récupère, transforme et charge les données depuis les sources vers le data warehouse.
- Transaction atomique
Ensemble d'opérations SQL qui réussissent toutes ou sont toutes annulées, pour garantir l'intégrité des données.
- Trigger
Commande SQL qui exécute automatiquement une action lorsqu'un événement se produit sur une table (ex. insertion, modification).
- Journal de Logs
Journal des transactions qui enregistre toutes les transactions et les modifications apportées par chaque transaction.

Références

- SQL Server Docs – transaction log
<https://learn.microsoft.com/fr-fr/sql/relational-databases/logs/the-transaction-log-sql-server?view=sql-server-ver17>
- SQL Server Docs – Triggers
<https://learn.microsoft.com/sql/t-sql/statements/create-trigger-transact-sql>
- SQL Server Docs – Autorisations utilisateur
<https://learn.microsoft.com/fr-fr/ssms/user-permissions?tabs=windows-10>
- SQL Server Docs – SQL Server Agent Jobs
<https://learn.microsoft.com/fr-fr/ssms/agent/sql-server-agent>
- SQL Server Docs – Integrity Constraints
<https://learn.microsoft.com/sql/relational-databases/tables/primary-and-foreign-key-constraints>
- Azure Data Factory – Monitoring and Alerts
<https://learn.microsoft.com/azure/data-factory/monitor-visually>

- Microsoft Learn – Surveiller les actualisations des flux de données
<https://learn.microsoft.com/fr-fr/power-query/dataflows/monitor-dataflow-refreshes>
- LeMagIT - Les bonnes pratiques pour bien protéger SQL Server
<https://www.lemagit.fr/conseil/Les-bonnes-pratiques-pour-bien-proteger-SQL-Server>