05/02/2024

NEKRUZ Kayumars

3Il

Rapport Power BI

Sous titre

Table des matières

[Introduction 2](#_Toc157977679)

[1. Contexte et Objectifs 2](#_Toc157977680)

[2. Identification et Sélection des Sources de Données 2](#_Toc157977681)

[3. Nettoyage et Préparation des Données 2](#_Toc157977682)

[4. Architecture du Datawarehouse 3](#_Toc157977683)

[1. Architecture 1 : Data Lake avec approche Serverless 3](#_Toc157977684)

[2. Architecture 2 : Data Warehouse centrée sur le Cloud 4](#_Toc157977685)

[5. Développement des Tableaux de Bord 5](#_Toc157977686)

[6. Analyses et Insights 5](#_Toc157977687)

[7. Conclusion 5](#_Toc157977688)

[Annexes 5](#_Toc157977689)

# Introduction

* Brève présentation du contexte et des objectifs des missions.
* Importance de l'analyse BI pour atteindre les objectifs de l'UE.
* Aperçu de la structure du rapport.

# 1. Contexte et Objectifs

Mission 1

* Présentation détaillée de la première mission : identifier des villes européennes à fort potentiel touristique sous-exploité et à faible impact écologique.

Mission 2

* Présentation de la deuxième mission : identifier des destinations attractives pour les jeunes européens, avec un focus sur le budget et la qualité de l'offre touristique.

# 2. Identification et Sélection des Sources de Données

* Description des sources de données utilisées (e.g., all hotels.csv, Berlin.json, Cities\_scores.csv, etc.).
* Justification de la sélection de ces sources par rapport aux objectifs des missions.
* Méthodologie de collecte des données.

# 3. Nettoyage et Préparation des Données

* Détails des étapes de nettoyage et de correction des données.
* Exemples de corrections apportées.
* Préparation des données pour l'analyse et l'intégration dans Power BI.

# 4. Architecture du Datawarehouse

## Architecture 1 : Data Lake avec approche Serverless

Une image contenant texte, diagramme, Plan, Dessin technique

Description générée automatiquement

* **Collecte et Préparation des Données** : Services serverless (Azure Functions, AWS Lambda) pour l'ingestion des fichiers (JSON, CSV, HTML) dans un Data Lake (Azure Data Lake Storage, AWS S3). Transformation des données via Apache Spark pour assurer la qualité et la cohérence.
* **Stockage et Organisation des Données** : Les données brutes sont stockées dans le Data Lake, organisées par zone fonctionnelle. Services de métadonnées (AWS Glue Data Catalog, Azure Data Catalog) pour cataloguer les données, facilitant leur découverte.
* **Modélisation des Données** : Création d'un modèle logique des données avec Azure Databricks ou AWS Glue pour transformer les données brutes en données structurées, prêtes pour l'analyse.
* **Analyse et Visualisation des Données** : Connexion de Power BI au Data Lake pour développer des rapports et des tableaux de bord interactifs, avec des fonctionnalités de traitement en direct pour des visualisations à jour.

**Avantages** :

* **Flexibilité et Scalabilité** : Les services serverless et le Data Lake offrent une grande flexibilité et une scalabilité automatique.
* **Efficacité de la Transformation** : Apache Spark garantit une transformation efficace et de qualité.
* **Découverte Facilitée** : Les services de métadonnées permettent une découverte et une organisation efficaces des données.
* **Visualisations Actualisées** : Power BI assure des analyses et visualisations en temps réel avec des données toujours actualisées.

**Inconvénients** :

* **Complexité de Configuration** : La mise en place et la gestion d'une architecture serverless et d'un Data Lake peuvent être complexes.
* **Dépendance aux Services Cloud** : Forte dépendance aux fournisseurs de services cloud pour la gestion de l'infrastructure.
* **Coûts Variables** : Les coûts peuvent varier en fonction de l'utilisation des services serverless et du stockage des données.

Cette transformation de l'architecture serverless en Data Lake en un modèle simplifié met en avant les étapes clés de la collecte à la visualisation des données, soulignant les avantages de l'approche serverless tout en reconnaissant les défis potentiels liés à la complexité et aux coûts.

## Architecture 2 : Data Warehouse centrée sur le Cloud

Une image contenant diagramme, Plan, croquis, Dessin technique

Description générée automatiquement

* **Collecte et Préparation des Données :**Utilisation d'outils ETL pour extraire, transformer, et charger les données de diverses sources vers l'entrepôt de données.
* **Stockage et Organisation des Données**: Les données sont stockées et organisées dans un Data Warehouse comme BigQuery, Redshift, ou Azure Synapse Analytics, utilisant un modèle en étoile ou en flocon pour structurer les tables de faits et de dimensions.
* **Modélisation des Données :** Mise en place d'un modèle en étoile ou en flocon pour simplifier les requêtes et améliorer les performances.
* **Analyse et Visualisation des Données**: Utilisation d'outils BI pour l'analyse et la création de tableaux de bord interactifs et de rapports à partir des données organisées.

**Avantages :**

* **Efficacité et Qualité** : Automatisation avec les outils ETL améliore la qualité des données et réduit les erreurs.
* **Performance** : Le stockage dédié et les modèles de données optimisent les performances des requêtes.
* **Insights Rapides** : Les outils BI permettent des analyses rapides et la création facile de visualisations.

**Inconvénients :**

* **Complexité et Coûts** : La configuration initiale et la maintenance peuvent être complexes et coûteuses, surtout avec l'augmentation des volumes de données.
* **Compétences Spécialisées** : La gestion optimale du Data Warehouse nécessite des compétences spécialisées, notamment pour l'optimisation des requêtes et la maintenance de l'infrastructure.

L'architecture est conçue pour maximiser l'efficacité de la collecte, du stockage, de l'organisation, et de l'analyse des données. Le modèle en étoile ou en flocon centralise les données pour des requêtes simplifiées, réduisant ainsi la complexité et améliorant la vitesse d'exécution. Cela facilite des analyses en temps réel et permet une prise de décision basée sur des données actualisées et bien structurées. La sélection des technologies est guidée par la recherche d'un équilibre entre performance, coût, et flexibilité, pour répondre aux besoins spécifiques et évoluer avec l'organisation.

# 5. Développement des Tableaux de Bord

* Description de la conception des tableaux de bord pour chaque mission.
* Inclusion des captures d'écran des tableaux de bord.
* Détail des visualisations choisies et leur pertinence par rapport aux objectifs d'analyse.

# 6. Analyses et Insights

* Présentation des principales découvertes pour chaque mission.
* Utilisation de calculs statistiques (DAX) : explication et captures d'écran des formules utilisées.
* Interprétation des données et recommandations basées sur les insights obtenus.

# 7. Conclusion

* Résumé des résultats clés et de leur impact potentiel sur les objectifs de l'UE.
* Discussion sur les limites de l'étude et les opportunités d'amélioration.
* Perspectives futures pour des analyses plus approfondies.

# Annexes

* Code DAX supplémentaire, si applicable.
* Toute autre documentation supportant le projet.