

Aide à la migration du contenu
d'une base de données -
Création d'un système
d'extraction de tweets
Tuteur : Philippe Roulet

Plan

- Contexte du stage
- Projet de migration de données
- Projet d'extraction de tweets
- Etat actuel des projets et évolutions

Contexte du stage

- L'entreprise Isover
 - Filiale du groupe Saint-Gobain
 - Leader mondial de l'isolation
- Le département Business Intelligence
 - Développement et maintenance d'applications et solutions
 - Fonction de support aux utilisateurs

Projet de migration de données

- Contexte et origine
 - Données des ventes d'Isonat dans un format et un serveur différent
 - Nécessaires à la maison mère Isover
 - Problèmes de correspondances de certaines données
 - Lancement d'un projet de migration

Projet de migration de données

- But du projet
 - Permettre la migration des données de vente
 - Mise en place d'une correspondance des données différentes
 - Création d'une table contenant les ventes Isonat au sein des serveurs Isover

Projet de migration de données

- Technologies utilisées
 - SQL Serveur Management Studio
 - Manipulation des données des serveurs
 - Création de tables de correspondances
 - Rédaction d'une procédure stockée
 - SAP Business Object
 - Définition des données nécessaires
 - Création d'une requête d'extraction

Projet de migration de données

- Développement de la procédure
 - Modification de la requête fournie
 - Obtention des correspondances
 - Définition et création des tables de stockage
 - Rédaction du corps de la procédure
 - Encapsulation des zones critiques
- Étapes en parallèle
 - Tests sur données locales
 - Optimisation du code

Projet de migration de données

- Mise en pause du projet
 - Nombreuses correspondances nécessaires
 - Fichiers incomplets temporaires fournis pour la création de la procédure
 - Actuellement en attente des fichiers finaux
- Lancement du second projet pendant la période d'attente

Projet d'extraction de tweets

- Contexte et origine
 - Idée soulevée par le département marketing
 - Connaitre le contenu des tweets liés aux activités du groupe
 - En obtenir des informations utiles
- Lancement du projet par la BI
 - Anticipation de la demande
 - Projet secondaire aux activités BI

Projet d'extraction de tweets

- But du projet
 - Extraction de tweets selon plusieurs critères
 - Sélectionner quoi extraire par hashtag ou mot-clé
 - Utiliser les données des tweets obtenus pour obtenir différents graphiques et informations

Projet d'extraction de tweets

- Technologies d'origine
 - Machine virtuelle Ubuntu
 - Python
 - Programme d'extraction utilisant l'API Twitter
 - IDE PyCharm de JetBrains
 - MongoDB
 - Base de données NoSQL
 - Outil graphique MongoBooster

Projet d'extraction de tweets

- Problèmes de ces technologies
 - Limitation par l'API Twitter du nombre de tweets extraits par jour
 - Difficulté de manipulation des données stockées dans la base MongoDB
- Changement de technologie
 - GNU R
 - Langage de statistique et science des données
 - Librairies twitterR et ROAuth
 - Export et opérations statistiques simplifiées

Projet d'extraction de tweets

- Développement
 - Obtention des clés et tokens de connexion
 - Rédaction d'un script de connexion à Twitter
 - Tests d'extraction de tweets à petite échelle
 - Rédaction d'un script d'extraction selon hashtag, mot-clé et dates limites
 - Tests d'exportation au format csv des données extraites

Etat actuel des projets et évolutions

- Procédure SQL
 - En attente des dernières correspondances
 - En attente d'un passage en production
- Scripts d'extraction R
 - Actuellement en cours de développement
 - Travail en cours sur les opérations d'analyse et sur l'export dans un fichier utilisable



Conclusions