

Business Intelligence

Projet

Introduction

Ce document décrit les consignes du projet qui devra être réalisé pour l'examen oral de janvier 2019. Lors de cet examen, vous serez évalué sur :

- Le projet décrit dans ce document
- Des questions générales sur la démarche BI (voir supports de cours et séances de laboratoire).

Consignes

Ce projet est à réaliser par groupe de deux personnes. Vous devrez communiquer la constitution de votre groupe au professeur pour le 21 décembre à 23 :59 :59 au plus tard via le formulaire ci-joint : <https://goo.gl/forms/q825DVxOUiGDvE6D3>

Votre projet doit tourner sur les machines de laboratoire de l'école ou sur votre PC personnel. Attention : les versions des outils de développement étant différentes entre les stations de l'école et vos stations personnelles, assurez-vous de pouvoir montrer quelque chose qui tourne le jour de l'examen oral : venez avec votre matériel si nécessaire. Si vous ne pouvez pas montrer quelque chose qui tourne, vous partirez avec un gros handicap pour votre examen.

La date de remise de votre projet est fixée au premier jour d'examen du cours de BI (même si votre groupe ne présente pas ce jour-là). Les dates de la session n'étant pas encore connues, la date de remise vous sera communiquée ultérieurement. A cette date, vous devrez avoir commité et poussé sur un repository Git hébergé sur GitHub l'ensemble de votre projet (code source cube et ETL, scripts de création et d'initialisation de DB, documentation...). Ce repository Git devra être privé et rendu accessible à l'utilisateur GitHub « samiroquai ». Votre première tâche le jour de l'examen sera de réaliser un clone de votre repository sur une machine de laboratoire de l'école (ce clone sera utilisé pour archiver vos dossiers, de cette manière nous ne vous demandons pas de graver votre production sur CD).

L'horaire de passage de l'examen vous sera communiqué ultérieurement.

Description du domaine d'application

Vous êtes un consultant BI envoyé en mission chez un client, gros groupe industriel, qui possède plusieurs chaînes de supermarchés, dont les chaînes « Grand Souk » et « Alim 2000 ».

Ces chaînes de supermarché enregistrent tous les produits vendus grâce à des systèmes de caisses enregistreuses. Le groupe souhaite pouvoir tirer de l'information de toutes les données enregistrées par les caisses. Il souhaiterait disposer d'un outil simple à utiliser qui permettrait d'obtenir des réponses rapidement à des questions précises. Il souhaiterait pouvoir analyser les données sur plusieurs axes, faire des recoupements. Pour le moment,

le groupe s'intéresse particulièrement aux données produites par Grand Souk. Celles d'Alim 2000 suivront dans un futur proche. Votre système doit être conçu de manière à absorber facilement l'intégration d'Alim 2000 dans le futur.

Le groupe souhaiterait pouvoir analyser les ventes par critère temporel, par produit et catégorie et par les caractéristiques des clients qui ont acheté des produits.

Le système de caisses enregistreuses stocke ses données dans une base de données à laquelle vous avez accès en lecture seule. Cette base de données se trouve sur `vm-sql2.iesn.be\Stu3IG`. Une fois par an, vous importerez ces données en écrasant ce qui existe déjà (par souci de simplification).

Quelques exemples de questions auxquelles le client souhaite pouvoir répondre à l'aide de l'outil que vous lui proposerez :

- De quelle province viennent les clients qui représentent le plus grand total d'achat (en termes de montant)
- Quelle chaîne de supermarchés vend le plus de produits d'une catégorie X
- Le nom du mois de l'année au cours duquel les ventes sont les plus importantes (toujours en termes de montants)
- Le total de la surface de vente disponible dans chaque province du pays.
- Quelles sont les catégories de produits qui sont les plus achetées par les célibataires ?
- Les montants vendus ventilés par département.
- Les personnes nées en Janvier achètent-elle plus de galette des rois ?

Le multilinguisme doit être supporté. Vous devrez offrir une consultation des données dans les 4 langues déjà utilisées dans la DB opérationnelle.

Vous devez concevoir la base de données du Data Warehouse cible et développer le nécessaire pour l'alimenter à partir de la base de données opérationnelle qui vous est fournie. Le Cube ne vous est pas demandé et ne conditionnera pas votre réussite, il sera comptabilisé en bonus s'il est fourni.

Délivrables

Votre repository Git devra contenir deux répertoires :

- Docs (documentation de votre projet)
- Src (code source de votre projet)

Dans docs, un fichier word qui contient :

- Une page de garde (infos du groupe, infos de l'année, de l'établissement, du cours...)
- Une brève explication de la solution que vous proposez, des composants impliqués et des technologies que vous employez.
- Les diagrammes de vos différents modèles.
- La granularité de votre table de faits.
- La justification des choix :
 - De clés primaires du modèle Data Warehouse.

- De modélisation des dimensions et faits
- Les limites rencontrées (à quelles questions fournies par le métier ne pourrez-vous pas répondre, pour quelle raison).
- La stratégie d'exécution de l'ETL : que faire si l'ETL s'arrête en plein traitement ? Peut-on le relancer ? Comment les erreurs sont-elles traitées ?

Dans Src :

- Le script qui permettra de créer le Data Warehouse, sous SQL Server (vm-sql2.iesn.be\Stu3IG)
- Les sources du projet qui permettra d'alimenter le Data Warehouse à partir de la DB d'origine. Utilisez les technologies vues aux cours.
- Les sources du cube si vous visez le bonus. Utilisez les technologies vues aux cours.

Export du système de caisses

Les noms des champs devraient être suffisamment parlants. Les clients enregistrés proviennent uniquement de Belgique.

ETL

Gestion des erreurs

Si une erreur survient lors de l'exécution de votre ETL (ex : un lookup qui ne retourne pas de correspondance)

Modélisation

Vous devez viser la performance. Faites les choix de modélisation adéquats, tels que discutés durant les cours théoriques et laboratoires.

Ressources

- Fichier des codes postaux de Belgique :
<http://www.bpost.be/site/fr/envoyer/adressage/rechercher-un-code-postal>
- Documentation SSIS: <https://docs.microsoft.com/en-us/sql/integration-services/sql-server-integration-services>
- Supports de cours théoriques et laboratoires sur Moodle
- Livre de référence disponible à la bibliothèque de l'IESN :
https://www.amazon.com/Data-Warehouse-Toolkit-Definitive-Dimensional/dp/1118530802/ref=dp_ob_title_bk