

Business Intelligence

Projet

Introduction

Ce document décrit les consignes du projet qui devra être réalisé pour l'examen oral de janvier 2023. 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 15/11/2022 au plus tard via le formulaire ci-joint (1 seule réponse par groupe) :

<https://forms.office.com/Pages/DesignPageV2.aspx?origin=NeoPortalPage&subpage=design&id=SAQmcanxKUe2MEiC5NxY9QdHgHKgqjNMmW6SBAold6xUQjROQ0tLM0RaOUw3VUE1SlRVvpJWDhXNy4u&topview=Preview>

Pour l'examen : votre projet doit tourner sur votre machine. La date de remise de votre projet sera La date de remise de votre projet est fixée au 24/12/2022 23:59:59. A cette date, vous devrez avoir rendu via un devoir Moodle qui sera créé pour l'occasion l'ensemble de votre projet (code source ETL, scripts de création et d'initialisation de DB, documentation...).

Description du domaine d'application

Vous travaillez pour une SSII active dans le domaine de la BI et êtes envoyé en mission chez BlablaCo, groupe commercial gérant deux marques de produits alimentaires: Foodies997 et Zen1984.

Ces marques enregistrent toutes les commandes de produits par le biais de ses employés, commerciaux déployés géographiquement. En plus de ses commerciaux, chaque enseigne dispose d'un webshop proposant le même catalogue de produits que ceux proposés par les commerciaux. Le groupe souhaite pouvoir tirer de l'information de toutes les données. 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 Foodies997. Celles de Zen1984 suivront dans un futur proche. Votre système doit être conçu de manière à absorber facilement l'intégration de Zen1984 dans le futur.

Le groupe souhaiterait entre autres 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.

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

- Quel est la société de livraison qui a acheminé le plus grand nombre de colis sur une période donnée ?
- De quel pays et de quelle ville viennent les clients qui représentent le plus grand total d'achat (en termes de montant)
- Quel commercial vend le plus de produits d'une catégorie X
- Le top 5 des produits les plus vendus en ligne
- Le top 3 des semaines de l'année durant lesquelles les ventes sont les plus importantes
- Pour une période donnée, vend-t-on plus en ligne ou par les commerciaux ?
- Pour une catégorie de produit X, vend-t-on plus en ligne ou par les commerciaux ?
- Les montants vendus répartis par catégorie de produit
- Les clients disposant d'une grande superficie de vente achètent-ils plus de produits que les autres ?

Le métier souhaiterait que le multilinguisme soit supporté. Vous devrez offrir une consultation des données en Anglais et en Français.

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.

Délivrables

Le dossier que vous remettrez sur Moodle sera un fichier ZIP et 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 ?
- La partie Reporting

Dans Src :

- Le script qui permettra de créer le Data Warehouse, sous SQL Server
- 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. Utilisez les technologies vues aux cours.

Modélisation

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

Gestion des erreurs

Votre ETL devra résister aux erreurs. Le jour de l'examen, lorsque nous testerons ensemble votre ETL, il est possible que je modifie les données de la BD source afin de générer des incohérences. Votre package ne devra pas s'arrêter. Adoptez une stratégie de gestion des erreurs qui vous semble pertinente.

Qualité des données

Pensez à intégrer à votre ETL des écrans de qualité pour détecter/signaler/corriger les anomalies. Le jour de l'examen, il est possible que je modifie les données de la BD source pour y intégrer des informations qui seraient de nature à fournir des données incohérentes. Votre DW ne doit pas en souffrir.

Ressources

- 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