



Licence 3

INF351 : Business Intelligence

2024/2025

Valéry MONTHE

Projet

Mise en œuvre d'un Système d'Aide à la Décision

Objectif : Mettre en place un système d'aide à la décision.

Plus spécifiquement, il est question de collecter des données, les transformer, les stocker et d'en produire des informations nouvelles pour aider à la prise de décision.

Modalité : le travail est à réaliser individuellement pour une partie et par groupe de 4 étudiants pour l'autre partie.

Début : 03/01/2025

Durée : 21 jours (3 semaines)

Livraison :

Elle se fera progressivement avec des évaluations intermédiaires, comme décrit dans le calendrier de livraison présenté à la section IV.

I. Description

Une conférence scientifique est un évènement organisé par une société scientifique ou un groupe de chercheurs, au cours duquel des scientifiques dont les articles ont été acceptés pour publication, font des présentations de leurs travaux de recherches. La conférence RecSys(<https://recsys.acm.org/>) est l'une de ces conférences. Les organisateurs de RecSys souhaitent comprendre comment évolue leur conférence à l'effet de prendre les bonnes décisions pour sa gestion. Ils s'intéressent notamment aux articles publiés, leurs auteurs, les institutions et pays d'origine des auteurs, etc. Ils aimeraient par exemple avoir des informations sur le taux de publication d'articles par institution, ville, pays, continent, etc. Une telle information pourrait permettre par exemple de définir des politiques pour encourager les régions du monde où il y a peu d'auteurs pour la conférence.

Pour aider les organisateurs de RecSys, il vous est demandé de leur proposer un système d'aide à la décision. Ce système devra s'appuyer sur les données des publications d'articles et dégager des informations nouvelles et utiles aux organisateurs.

Ce projet qui est organisé en plusieurs tâches se déroulera en plusieurs étapes.

Dans le cadre du projet, vous avez donc pour mission de :

1. Analyser les besoins et définir les fonctionnalités du SID à mettre en place
2. Mettre en place une architecture de système décisionnel permettant de collecter les données des différentes sources (fichiers d'articles), de les stocker dans une base de données opérationnelle, puis d'alimenter une base de données décisionnelle (entrepôt de données) à partir de la base opérationnelle.
3. Identifier les indicateurs pertinents et concevoir des tableaux de bord pour la restitution des informations utiles aux organisateurs de la conférence.

4. Documenter votre travail.

Pour atteindre ces objectifs, le projet est découpé en trois (3) lots définis et organisés comme suit :

- a. **Lot 1** : Conception, implémentation et alimentation d'une base de données à partir des fichiers des articles. *(Travail individuel)*
- b. **Lot 2** : Conception, implémentation et alimentation d'un entrepôt de données à partir de la base opérationnelle. *(Travail de groupe)*
- c. **Lot 3** : Conception et implémentation des tableaux de bord pour la visualisation / restitution de l'information. *(Travail de groupe)*

II. Travail à faire

Partie 1 : Conception, implémentation et alimentation de la base de données opérationnelle (lot1)

Cette partie qui constitue le **lot 1** du projet et est structurée comme suit :

1. Analyse et conception
 - Nous l'avons réalisée ensemble lors de la séance de cours avant le contrôle continu
 - Un modèle conceptuel de données a été produit
 - Il vous revient donc de produire le modèle logique de données.
2. Implémentation
 - Choisir un SGBD, l'installer et implémenter la base de données.
3. Alimentation
 - Analyser et comprendre la structure (contenu) des différents fichiers d'articles
 - A l'aide du langage Python, mettre en place le processus ETL permettant d'extraire les données de ces fichiers, de les transformer (traiter et formater) et de les stocker dans la base de données.
4. Documentation
 - Produire un rapport pour cette partie (R1)

Partie 2 : Conception, implémentation et alimentation de l'entrepôt de données (lot2)

Cette partie qui constitue le **lot 2** du projet et est structurée comme suit :

1. Analyse

En tenant compte des besoins analysés et des données stockées dans la base de données, vous devez :

- Identifier et décrire les différents éléments qui vont constituer votre entrepôt de données : faits, mesures, dimensions et leurs niveaux de granularités appropriés, etc.
 - Définir en détail les faits, en produisant le dictionnaire des mesures.
 - Définir en détail les dimensions : dictionnaire des attributs.
2. Conception
 - Choix du modèle approprié pour l'entrepôt.
 - Définition des hiérarchies avec les cardinalités et définition du schéma conceptuel de l'entrepôt de données.
 3. Implémentation : approche ROLAP
 - Implémentez votre entrepôt dans un SGBD relationnel de votre choix, MySQL ou PostgreSQL.
 4. Alimentation
 - Faire le schéma du processus à mettre en place.
 - Utiliser l'outil Talent OS-DI pour réaliser le processus ETL et alimenter l'entrepôt de données à partir de la BD opérationnelle
 - Décrire les composants utilisés dans votre/vos job(s). NB : les composants et les jobs sont des notions que vous découvrirez dans l'outil Talent.
 5. Documentation
 - Produire un rapport pour cette partie (R2)

Partie 3 : Conception et implémentation de tableaux de bord et reporting(lot3)

Dans cette partie, vous devez proposer des rapports et tableaux de bord assez intuitifs, et fournissant des informations pertinentes pouvant aider les organisateurs de RecSys à prendre de bonnes décisions. Pour cela, il vous est demandé de :

1. Définir des indicateurs pertinents pour leurs besoins
2. Concevoir plusieurs types de tableaux de bord
3. Utiliser le langage Python et l'outil Power BI pour mettre en œuvre les tableaux de bord
4. Créer des comptes sur Power BI service et publier les rapports.
5. Documenter le travail réalisé : Rapport R3

III. Sources de données à utiliser

Les fichiers (sources de données) sont fournis, et sont à récupérer dans le drive.

La structure et le contenu de ces fichiers vous ont été expliqués en cours.

IV. Calendrier de livraison

- Projet à réaliser par groupe

Num	Livrables	Modalités	Equipes	Délais
1	Lot 1 - archive contenant : <ul style="list-style-type: none">- Rapport R1- Base de données : backup SQL- Code Python- Readme	Rendu via un google form	Individuel	J0+7
2	Lot 2	Suivi : <i>évaluation du niveau d'avancement</i>	Groupe de 4	J0+14
3	Lot 3	Suivi : <i>évaluation du niveau d'avancement</i>	Groupe de 4	J0+19
4	Rendu final : <ul style="list-style-type: none">- Rapport final : R1 + R2 + R3- Project final: projet Talend, code Python, projet Power BI	Evaluation finale: <i>Présentation + Démo</i>	Groupe de 4	J0+21