

**Ecole Normale Supérieure de l'Enseignement Technique**  
**Département Mathématique et informatique**

**Examen du 2<sup>ème</sup> semestre 2021/2022**

**La note :.....**

**Date : 08/06/2022**

**Module : BIG DATA**

**Durée : 2h**

**Nom & Prénom :.....**

**Filière :.....**

**Enoncé:**

IMOMAROC est une entreprise d'immobilier qui souhaite traiter ces données au moyen d'une application Spark d'une manière parallèle est distribuée. L'entreprise possède des données stockées dans une base de données relationnel et des fichiers Json. L'objectif est de traiter ces données en utilisant Spark SQL à travers les APIs DataFrame et Dataset pour extraire des informations utiles afin de prendre des décisions.

**I. Traitement de données stockées dans Mysql**

IMOMAROC possède une application web pour gérer ces projets immobiliers, les données sont stockées dans une base de données MYSQL nommée **DB\_IMOMAROC**, qui contient deux tables **PROJETS** et **TACHES** (Voir les figures 1 et 2).


|                          | # | Nom   | Type        | Interclassement   | Attributs | Null | Valeur par défaut | C |
|--------------------------|---|---|-------------|-------------------|-----------|------|-------------------|---|
| <input type="checkbox"/> | 1 | ID_PROJET  | int(11)     |                   |           | Non  | Aucun(e)          |   |
| <input type="checkbox"/> | 2 | TITRE   | varchar(50) | latin1_swedish_ci |           | Non  | Aucun(e)          |   |
| <input type="checkbox"/> | 3 | DESCRIPTION   | text        | latin1_swedish_ci |           | Non  | Aucun(e)          |   |
| <input type="checkbox"/> | 4 | LIEU  | varchar(50) | latin1_swedish_ci |           | Non  | Aucun(e)          |   |
| <input type="checkbox"/> | 5 | DATE_DEBUT  | date        |                   |           | Non  | Aucun(e)          |   |
| <input type="checkbox"/> | 6 | DATE_FIN  | int(11)     |                   |           | Non  | Aucun(e)          |   |

Figure 1: structure de la table PROJETS



|                          | # | Nom  | Type        | Interclassement   | Attributs | Null | Valeur par défaut | Commentaires | Extra          |
|--------------------------|---|--|-------------|-------------------|-----------|------|-------------------|--------------|----------------|
| <input type="checkbox"/> | 1 | ID_TACHE    | int(11)     |                   |           | Non  | Aucun(e)          |              | AUTO_INCREMENT |
| <input type="checkbox"/> | 2 | TITE   | varchar(50) | latin1_swedish_ci |           | Non  | Aucun(e)          |              |                |
| <input type="checkbox"/> | 3 | DATE_DEBUT   | date        |                   |           | Non  | Aucun(e)          |              |                |
| <input type="checkbox"/> | 4 | DATE_FIN   | date        |                   |           | Non  | Aucun(e)          |              |                |
| <input type="checkbox"/> | 5 | ID_PROJET  | int(11)     |                   |           | Non  | Aucun(e)          |              |                |

Figure 2: structure de la table TACHES

### Travail à faire :

Vous créez la base de données et les tables et vous répondez aux questions suivantes :

1. Afficher les projets en cours de réalisation.
2. Afficher pour chaque projet, le nombre de tâches dont la durée dépasse un mois. Le format d'affichage est le suivant :

ID\_PROJET | TITRE | NOMBRE

3. Afficher pour chaque projet les tâches en retard (avec la durée de retard).

## **II. Traitement de données stockées dans fichier Json.**

L'entreprise dispose d'un grand fichier Json qui contient les ventes d'appartements pour ces clients, le fichier contient pour chaque client les appartements achetés (Voir le fichier Json).

### **Travail à faire :**

1. Afficher les deux première villes ou l'entreprise a réalisé plus de ventes (en terme de revenues).
2. Afficher la liste des clients qui possèdent plus qu'un appartement à rabat.
3. Afficher le nombre d'appartements vendues à Casablanca.