



Master Spécialisé « Data Engineering » (MSDE)

Projet Module 4 :

BD SQL / NoSQL

Objectifs et données :

L'objectif de ce projet est de mettre en place une base de données relationnelle (MySQL) et une autre NoSQL (MongoDB) puis les interroger.

Le contenu des BD sera constitué des indicateurs du recensement général de la population et de l'habitat du Maroc de 2014 (RGPH 2014) disponibles sur le site web du HCP via le lien suivant :

<http://rgphentableaux.hcp.ma/Default1/>

Ces indicateurs sont répartis en 6 thèmes :

- Démographie
- Handicap
- Education et alphabétisation
- Langues locales utilisées
- Activité et emploi
- Conditions d'habitat

Chaque thème, présenté à différentes échelles géographiques (communal, provincial, régional et national), comporte un ensemble d'attributs répartis eux-mêmes par genre (masculin et féminin) et par catégorie territoriale (urbain et rural).

Dans ce projet, l'intégration des données dans la base de données se fera au niveau du découpage le plus fin (commune), les autres niveaux géographiques supérieurs (province, région et Maroc) seront calculés directement par agrégation en utilisant les fonctions adéquates.

Deadline de remise du projet : 7 février 2020

Livrables du projet :

Fichier zippé (.zip où .rar) portant votre (vos) noms et comportant les fichiers suivants :

- BD MySQL
- BD MongoDB

- Fichier de traitement et de requêtes (.ipynb ou tout autre format de fichier utilisé pour répondre aux questions de ce projet)

Remarques :

- Le projet peut être réalisé en monôme ou en binôme
- Les livrables sont à déposer sur la plateforme du Master dans l'espace du module 4 : **Dépôt du projet SQL/NoSQL**
- Le projet est à réaliser en utilisant les compétences acquises lors de ce module et des précédents modules.

Questions :

Partie 1 : Chargement des données et création de la BD relationnelle sous MySQL

1. Analyser les données d'indicateurs du RGPH2014 selon les 6 thèmes (au niveau des communes) puis définir le schéma global de la BD relationnelle qui va contenir ces données ;
2. Créer une BD MySQL
3. Charger les données dans cette BD

Partie 2 : Interrogation de la BD MySQL

En utilisant le langage SQL, répondez aux questions suivantes :

4. Quel est le total de la population municipale du Maroc ?
5. Quels sont les dix premières communes les plus peuplées du Maroc ?
6. Quel est le nombre de communes existant hors les régions de Rabat et Casablanca ayant une population municipale supérieur à 50 000 habitants, un taux de chômage inférieur à 15%, un Taux net d'activité féminin supérieur à 30% et un taux d'analphabétisme inférieur à 20% ?
7. Quelle est la province ayant le taux de prévalence d'handicap le plus faible ?
8. Quel est l'âge moyen au premier mariage par genre (masculin, féminin) et par secteur (urbain, rural) au niveau de chaque région ?
9. Dans quelles provinces se trouvent-elles les dix communes ayant le taux d'analphabétisme le plus élevé ?
10. Quel est le niveau d'études le plus prépondérant au niveau de chaque région ?

11. Quelles sont les provinces où la langue locale « Tachelhit » est plus utilisée que « Tamazight » dans le cas où le pourcentage cumulé de ces deux langues locales dépasse 50% ?
12. Quel est le nombre de provinces où le taux d'équipement des ménages en parabole est supérieur à celui en réfrigérateur ?
13. Générer les vues représentant les 6 thèmes d'indicateurs à l'échelle provinciale, régionale puis nationale.

Partie 3 : Migration des données de la BD MySQL vers la BD MongoDB

14. Migrer le contenu de la BD MySQL vers une BD MongoDB
15. Définir votre processus de migration :
 - Conception de cette BD MongoDB
 - Import de données

Partie 4 : Interrogation de la BD MongoDB

16. Répondez aux mêmes questions de la partie 2 en utilisant la BD MongoDB

Partie 5 (optionnelle) : Création d'une page web d'authentification

17. Ajouter dans les deux BD une structure pour stocker un utilisateur dans la BD avec son login et son mot de passe.
18. Créer une page web permettant d'authentifier cet utilisateur.