ISGA

INSTITUT SUPERIEUR D'INGENIERIE ET DES AFFAIRES
Rue Ibn Habousse Hivernage Marrakech
Tél : 05 24 43 48 59 / 44 / 63

Contrôle du 1er Semestre (S1C1)
Année universitaire 2021/2022

CLASSE	2CI ISI	DATE	30/03/2022
MATIERE	Big Data	DUREE	1h30
PROFESSEUR	Snineh Sidi Med	DOCUMENTS	Autorisés

EXERCICE 1 : APACHE HIVE

On suppose que :

- L'emplacement de l'enrepôt de données Hive dans un système de fichiers HDFS est : /user/isga/warehouse
- La source de données locale est un fichier des employés. Nous voulons créer des partitions et des buckets basés sur le sexe et l'année d'embauche (Figure 1). Le sepérateur des données est le point virgule(;)

D:\data\emp.txt

```
id      dateNais    Nom      Prenom                S AnneeEmb
-------------------------------------------------------------
10001;'1953-09-02';'Georgi';'Facello';           'M';'1986'
10002;'1964-06-02';'Bezalel';'Simmel';           'F';'1985'
10003;'1959-12-03';'Parto';'Bamford';            'M';'1986'
10004;'1954-05-01';'Chirstian';'Koblick';        'M';'1986'
10005;'1955-01-21';'Kyoichi';'Maliniak';         'M';'1989'
10006;'1953-04-20';'Anneke';'Preusig';           'F';'1989'
10007;'1957-05-23';'Tzvetan';'Zielinski';        'F';'1989'
10008;'1958-02-19';'Saniya';'Kalloufi';          'M';'1994'
```

Figure 1: Liste des employés

QUESTIONS

1- Ecrire les instructions qui permettent de créer la table interne (emp_part) avec une partition dynamique basée sur le sexe et l'année d'embauche.

2- Ecrire les instructions (hadoop et Hive) qui permettent d'alimenter la table interne (emp_part) à partir d'une source HDFS basée sur le fichier **D:\data\emp.txt**

3- Ecrire les instructions qui permettent de créer et d'alimenter la table externe (emp_annee) à partir de la source locale **D:\data\emp.txt**. Nous souhaitons avoir pour chaque partition dynamique, basée sur l'année d'embauche, deux buckets.

EXERCICE 2 : APACHE SQOOP

- Nous souhaitons importer des données de la BD MySql « isga » vers une destination du système de fichiers HDFS précisemment vers l'emplacement « /user/hive » . L'utilisateur de cette BD est « root » et son mot de passe est « 123 ».

1. Ecrire les commandes Sqoop qui permettent d'importer seulement les clients de la ville de « Marrakech ».
    - Structure de la table clients (id, raisonsociale, gsm, ville, ...)
2. Nous voulons que Sqoop utilise 6 tâches en parallèle pendant l'importation des données. Modifier les commandes de la question n°1 pour répondre à ce besoin.