

<u>Moez Skanji</u> <u>Master II SEP: Big Data</u>	<u>Test de connaissance</u> <u>BIG DATA</u>	<u>Université de Reims</u> <u>Champagne Ardenne</u> <u>FÉVRIER-2020</u>
--	--	---

Exercice: Système de gestion d'emprunt dans une librairie

Soit le modèle relationnel suivant :

```
author(aid, aname)
book(bid, title, category)
student(sid, sname, dept)
write(aid, bid)
borrow(sid, bid, checkout-time, return-time, book_summary)
```

book représente une table d'auteurs. Chaque ligne contient le nom et l'identifiant d'un auteur **book** représente une table de livres. Chaque ligne est un livre décrit par son identifiant, son titre et sa catégorie (roman, science-fiction, musique, etc.).

student représente une table d'étudiants. Chaque ligne est un étudiant décrit par son identifiant, son nom et son département (informatique, mécanique...).

write représente l'association entre les auteurs et les livres. Une ligne de cette table signifie que l'auteur a écrit le livre **bid**.

borrow représente les informations de prêt de livre. Une ligne de cette table signifie que l'étudiant a emprunté le livre **bid** , à la date **checkout-time** et l'a retourné à la date **return-time**

Questions:

à partir d'un spark session réaliser ces requêtes en DSL et en SQL

0-Créer les données.

1-Trouver les titres de tous les livres que l'étudiant sid='S15' a emprunté.

2-Trouver les titres de tous les livres qui n'ont jamais été empruntés par un étudiant.

3-Trouver tous les étudiants qui ont emprunté le livre bid='002'

4-Trouver les titres de tous les livres empruntés par des étudiants en Mécanique (département Mécanique) (ERRATUM !!)

5-Trouver les étudiants qui n'ont jamais emprunté de livre.

6- Déterminer l'auteur qui a écrit le plus de livres.

7- Déterminer les personnes qui n'ont pas encore rendu les livres.

8- Créer une nouvelle colonne dans la table borrow qui prend la valeur 1, si la durée d'emprunt est supérieur à 3 mois, sinon 0. (utiliser la fonction spark to_date) par la suite faire un export des données en CSV. dans un répertoire nommé contention (Attention pas de path en dur!)

9-déterminer les livres qui n'ont jamais été empruntés.

10-Commit et Push sur un repo Git.

Tous les attributs sont des chaînes de caractères.

Author

aid	name
07890	Jean Paul Sartre
05678	Pierre de Ronsard

book

bid	title	category
0001	L'existentialisme est un humanisme	Philosophie
0002	Huis clos. Suivi de Les Mouches	Philosophie
0003	Mignonne allons voir si la rose	Poeme
0004	Les Amours	Poème

Student

sid	sname	dept
S15	toto	Math
S16	popo	Eco
S17	fofo	Mécanique

write

aid	bid
07890	0001
07890	0002
05678	0003
05678	0003

borrow

sid	bid	checkout-time	return-time
S15	0003	02-01-2020	01-02-2020
S15	0002	13-06-2020	null
S15	0001	13-06-2020	13-10-2020
S16	0002	24-01-2020	24-01-2020
S17	0001	12-04-2020	01-07-2020