

Traitement de données

ANNEXES

Tu trouveras en annexe deux fichiers :

- Une liste d'offres mobiles (id, label, seuil)

```
offer_id,label_tunnel,seuil_data  
1,Modulo 12 mois,1250000
```

- Une liste de consommations en data, voix et sms/mms (l'offer_id permet de faire le lien entre les deux listes)

```
anum,start,amount,carrier,emitter_location,receiver_location,item_unit,offer_id  
063XX23353,2020-09-01 00:10:26,6,data,france,france,ko,270
```

EXERCICES / QUESTIONS

01 - Regroupement par tel et par jour

En utilisant les données en annexe, l'objectif est de créer un script qui permet d'afficher un récapitulatif des consommations d'une ligne donnée (passée en argument de l'application).

Pour chaque ligne on affichera :

- La durée totale des consommation de type voix
- Le nombre de SMS/MMS envoyés
- Le volume total de data consommé (pour la conversion en gigaoctets on considérera qu'un giga est égal à 1 000 000 kilo octet)
- Un tableau récapitulatif des consommations data journalières (dans l'ordre de la plus ancienne à la plus récente).

On souhaite également pouvoir visualiser pour l'ensemble du parc les consommations moyennes des utilisateurs (voix, SMS/MMS et data).

Exemple :

```
› recap -tel "062XX00273"
```

```
Ligne :      062XX00273, 193 consommations.
Offre :      Le complet (353)
Voix :      12h25m41s
SMS/MMS :    69
Data :      5.504 Go / 50 Go

+-----+
| 2020-09-05 | 148.032 Mo |
| 2020-09-06 | 114.829 Mo |
| 2020-09-07 | 464.396 Mo |
| 2020-09-08 | 548.160 Mo |
| 2020-09-09 | 0.049 Mo |
| 2020-09-10 | 2181.637 Mo |
| 2020-09-12 | 541.726 Mo |
| 2020-09-13 | 538.267 Mo |
| 2020-09-14 | 130.848 Mo |
| 2020-09-15 | 256.998 Mo |
| 2020-09-16 | 339.509 Mo |
| 2020-09-17 | 229.910 Mo |
| 2020-09-18 | 10.028 Mo |
+-----+

-----
Consommation Voix Moyenne :    44mn
Consommation SMS/MMS Moyenne :   38
Consommation Data Moyenne :    1.10 Go
```

Si on ne passe pas de numéro de tel en entrée, on affichera le récapitulatif de chaque ligne puis la consommation moyenne sur l'ensemble du parc :

```
› recap
```

```
[...]
Ligne :    076XX11186, 11 consommations.
Offre :    Le complet (353)
Voix :     1m49s
SMS/MMS :  0
Data :     0.112 Go / 50.000

+-----+
| 2020-09-24 | 99.121 Mo |
| 2020-09-28 | 12.910 Mo |
+-----+

-----
Ligne :    067XX56643, 25 consommations.
Offre :    Le complet (353)
Voix :     8m51s
SMS/MMS :  5
Data :     1.943 Go / 50.000

+-----+
| 2020-09-22 | 4.185 Mo |
| 2020-09-23 | 157.952 Mo |
| 2020-09-24 | 371.189 Mo |
| 2020-09-25 | 842.373 Mo |
| 2020-09-28 | 253.439 Mo |
| 2020-09-29 | 314.265 Mo |
+-----+

-----
Ligne :    061XX52867, 14 consommations.
Offre :    Modulo 4 (119)
Voix :     14m15s
SMS/MMS :  0
Data :     0.302 Go / 8.000

+-----+
| 2020-09-21 | 74.934 Mo |
| 2020-09-22 | 35.416 Mo |
| 2020-09-23 | 36.182 Mo |
| 2020-09-24 | 26.908 Mo |
| 2020-09-26 | 128.156 Mo |
+-----+

-----

Consommation Voix Moyenne :    44mn
Consommation SMS/MMS Moyenne :   38
Consommation Data Moyenne :    1.10 Go
```

Quelques précisions :

- Les numéros de tel sont anonymisés mais pour cet exercice on peut les considérer comme des identifiants uniques.
- Pas besoin de vérifier la validité des données, on considère que ce qui se trouve dans les fichiers csv a été renseigné correctement.
- La performance de l'application est un critère important, les fichiers de consommations pourraient peser plusieurs dizaines de gigaoctets
- Il n'y a pas de pièges sur l'algorithme à mettre en place, ce qui nous intéresse c'est de pouvoir discuter des choix de structure de données, de comprendre quelle est la réflexion sur la performance et de voir comment le code est organisé, documenté, versionné, etc.

02 - Question : scaling ?

Maintenant si on imagine que l'on a 100 000 000 de consommations et qu'on a 1000 nouvelles consommations qui arrivent par seconde, quelle architecture privilégierais-tu sachant que l'on souhaite le résultat de l'exercice n°1 en quasi temps-réel ?

NB: pas de code nécessaire pour cette question mais un petit schéma d'architecture peut aider à expliquer les choix

FAQ

Quelles technos dois-je utiliser ?

Le choix des techno est laissé libre, nous avons surtout l'habitude d'utiliser du python et du go mais on recherche quelqu'un de pragmatique et curieux qui sait utiliser les bonnes technos au bon moment et pour les bons usages. Du coup le choix est libre et l'explication qui va avec nous intéresse beaucoup.

Quel est le rendu attendu ?

Pour le premier exercice, on s'attend tout simplement à voir un dépôt git bien rangé avec un bel historique qui nous permettra de mieux comprendre ta réflexion.

Pour le second, un petit doc en markdown ou un slide fera l'affaire.