

Accenture - Challenge Hackathome

Intro

Depuis 50 ans, la réduction de notre empreinte sur l'environnement est devenue une priorité majeure pour l'Humanité. Et le monde numérique ne fait pas exception : en 2019, on estimait que le domaine pesait pour 4% des émissions des gaz à effet de serre (GES), soit plus que le transport aérien !

Ces émissions ont deux sources :

- les appareils (serveurs, réseau, ordinateurs...) : production et fin de vie
- l'exploitation : principalement consommation, et entretien

Dans le même temps, l'utilisation exponentielle du numérique permet de réduire l'impact environnemental de nombreux services. Encore faut-il le faire intelligemment !

Objectif

A partir d'une liste de services à exécuter et des types de serveurs disponibles, votre objectif est de donner la liste des serveurs à acheter ainsi que la répartition des services sur ces serveurs, tels que l'impact environnemental de votre solution est minimal.

Description du problème

Serveurs

Dans l'archive de description du problème se trouve un fichier `servers_catalog.csv`. Vous y trouverez la liste des serveurs que vous pouvez utiliser pour votre solution. Les caractéristiques des serveurs sont les suivantes :

- `model` : nom du serveur
- `co2production` : émission de GES pour la production du serveur
- `co2usage` : émission de GES pour l'exploitation pendant 1 an du serveur
- `disk` : capacité de stockage
- `ram` : volume de RAM disponible
- `cores` : nombre de processeurs du serveur

Important : ce catalogue est unique, et est disponible pour toutes les fichiers d'entrée.

Services

Plusieurs listes de services sont disponibles, et vous devrez proposer une solution pour chacune.

La première ligne est un entier, donnant le nombre d'années pendant lequel tous les services de la liste devront être hébergés.

Sur les lignes suivantes se trouve la liste des services à héberger, avec leur caractéristiques. Dans l'ordre :

- Nom du service
- Volume de stockage nécessaire
- Volume de RAM nécessaire
- Nombre de processeurs nécessaire

Plusieurs services peuvent être hébergés sur un même serveur, mais un service ne peut pas être réparti sur plusieurs serveurs. Cela signifie que le serveur qui héberge le service doit avoir toutes les capacités nécessaires (disk, ram, cores) disponibles.

Format de solution

Pour chaque liste de service, vous devrez proposer une répartition de tous les services de la liste sur différents serveurs. Vous pouvez utiliser plusieurs fois le même type de serveur ; tous les services doivent être hébergés.

Votre soumission doit être un fichier au format CSV, ayant sur chaque ligne un serveur différent. Chaque ligne commence par le type de serveur, suivi par les noms de tous les services hébergés sur ce serveur, le tout séparés par des virgules. L'ordre des lignes ou des noms de services sur une ligne n'importe pas.

Un exemple est disponible plus bas à la section "Exemple".

Score

Score par jeu de données

Le score pour chaque liste de services est la somme des émissions en GES de chaque serveur de votre solution. Plus le score est faible, meilleure sera votre solution.

$$Score = \sum_{i \in \text{serveurs}} co2prod_i + N * co2usage_i$$

Les données `co2prod` et `co2usage` sont celles présentes dans le fichier `servers_catalog.csv`. La donnée `N` représente le nombre d'années pendant lequel les services doivent être hébergés (première ligne du fichier d'entrée).

Attention : pour que votre solution soit valide, il faut que tous les services soient hébergés, et que la somme des capacités requises par les services hébergés sur un serveur ne dépassent pas les capacités de ce serveur.

Score total

Sur chaque jeu de données, votre score sera obtenu en divisant le score de la meilleure soumission (toutes équipes confondues) par le score de la meilleure soumission de votre équipe. Votre score total sera la somme de ces scores sur tous les jeux de données, multiplié par 1000000.

Si vous n'avez soumis aucune solution valide pour un jeu de données, ce jeu de données ne contribuera pas à votre score total.

Attention : cela signifie que chaque dataset peut rapporter jusqu'à 1 000 000 de points. Travaillez donc sur tous les jeux de données, même les plus simples !

Exemple

Sur un jeu de données, vous obtenez un sous-score de 1600 points et toutes les autres équipes ont obtenu des scores moins bons (= plus élevés). Ce jeu de données vous rapportera $1\,000\,000 * (1600 / 1600) = 1$ million de points.

Plus tard dans la partie, une équipe concurrente soumet une solution meilleure que la vôtre, dont le sous-score est de 1000. Votre score sur ce jeu de données est recalculé et vaut désormais $1\,000\,000 * (1000 / 1600) = 625\,000$ points.

Exemple

Pour le fichier d'entrée suivant :

```
6
service_1,10,2,1
service_2,15000,4,2
service_3,500,30,1
```

Plusieurs solutions sont possibles. Une solution très simple mais très peu optimale est d'allouer un serveur de grande capacité à chaque service.

```
large1,service_1
large1,service_2
large1,service_3
```

Cette solution donne un score de $(25000 + 6 * 10000) * 3 = 255.000$. Une autre solution possible est la suivante, avec un score de 39.600 :

```
classic2,service_1,service_3
storage1,service_2
```