

Projet LotaCloud : Analyse de l'empreinte carbone

Le contexte

Présentation de LotaCloud :

Lota.cloud est une plateforme de gestion des dépenses Cloud dédiée à l'optimisation financière des services Cloud. Cette solution permet de Contrôlez toutes les dépenses MultiCloud dans un seul environnement grâce à leur plateforme unique dédiée au FinOps

Le projet :

La société LotaCloud nous a fourni des documents au format CSV. Ces fichiers contiennent des données d'utilisation de diverses ressources cloud d'un seul client. Notre mission sera de calculer l'empreinte carbone de certaines ressources spécifiques. Nous avons décidé de nous concentrer sur un seul fournisseur Cloud (Microsoft Azure) et sur deux ressources (Machines Virtuelles et le Stockage).

En résumé, à partir d'un ticket de caisse sur la consommation des ressources d'un client Azure, nous devons calculer une estimation de l'empreinte carbone de ses ressource utilisées dans le Cloud.

Les technologies utilisées seront :

Python, SQL et PowerBI.

L'essentiel à connaitre pour aborder ce projet

1. Qu'est ce qui caractérise le Cloud ?

- Le Cloud se caractérise par l'accès à des ressources informatiques (serveurs, stockage, bases de données, logiciels) via Internet, permettant une scalabilité, une flexibilité et une gestion à distance.

2. À quoi et à qui sert-il ?

- Le Cloud sert aux entreprises, aux développeurs et aux particuliers pour héberger des applications, stocker des données, analyser des données, et bien plus, sans avoir à gérer une infrastructure physique.

3. Quelles sont les ressources qui le composent ?

- Les ressources du Cloud comprennent les serveurs, les systèmes de stockage, les bases de données, les réseaux, les logiciels, et les services de sécurité.

4. Quel est le lien entre ces ressources Cloud et la production de gaz à effet de serre

- L'exploitation des centres de données qui hébergent les ressources Cloud consomme beaucoup d'électricité, souvent produite à partir de sources non renouvelables, ce qui génère des émissions de gaz à effet de serre.

5. Pour certaines de ces ressources, quels sont les catalyseurs qui influencent le taux d'émissions de CO2 ?

- Les catalyseurs incluent le type de matériel utilisé, l'efficacité énergétique des centres de données, le mix énergétique local (pourcentage d'énergie renouvelable vs fossile), et la gestion de la charge de travail.

6. Pourquoi la position géographique des ressources peut avoir un impact sur le CO2 qui est émit pour les faire fonctionner ?

- La position géographique impacte le mix énergétique utilisé (énergies renouvelables vs fossiles), les besoins en refroidissement (climats plus chauds nécessitent plus d'énergie pour refroidir les serveurs), et l'efficacité du réseau électrique local.

Prendre en compte le mixe énergétique pour le calcul des KPIs

7. Quels sont les autres éléments qui peuvent influencer les émissions de CO2 liées au Cloud ? ==> autre que l'énergie par exemple

- Les éléments incluent l'efficacité des logiciels utilisés, la virtualisation et la consolidation des ressources, l'optimisation des algorithmes et la gestion efficace des charges de travail.

8. Enfin, quels sont selon vous, les moyens de réduire l'empreinte carbone d'une application hébergée sur le Cloud ?

- Les moyens incluent :
 - Choisir des fournisseurs Cloud utilisant des énergies renouvelables.
 - Optimiser le code et les algorithmes pour réduire la consommation de ressources.
 - Utiliser des services de mise à l'échelle automatique pour éviter le surprovisionnement.
 - Choisir des régions Cloud géographiquement efficaces.
 - Migrer vers des architectures sans serveur (serverless) pour une utilisation plus efficace des ressources.

II Pour calculer l'empreinte Carbone d'un processeur

La métrique utilisée pour le calcul de l'empreinte carbone d'un processeur ou d'une machine virtuelle est généralement exprimée en kgCO₂e (**kilogrammes d'équivalent dioxyde de carbone**). Voici quelques KPI (Key Performance Indicators) couramment utilisés :

1. Consommation énergétique (kWh) :

- La quantité d'énergie consommée par le processeur ou la machine virtuelle est mesurée en kilowattheures (kWh). Cette consommation est ensuite convertie en émissions de CO₂ en fonction du mix énergétique de la région où se trouve le centre de données.

2. Émissions de CO₂ par kWh (kgCO₂e/kWh) :

- **Ce facteur de conversion dépend du mix énergétique local** (proportion d'énergie renouvelable vs fossile). Il permet de convertir la consommation énergétique en émissions de CO₂.

3. Utilisation du processeur (CPU Utilisation %) :

- Le pourcentage d'utilisation du processeur aide à déterminer l'efficacité et la charge de travail. Une utilisation inefficace peut entraîner une consommation énergétique inutile, augmentant ainsi les émissions de CO₂.

4. **Puissance de conception thermique** (TDP, Thermal Design Power) :

- Le TDP est une mesure de la quantité maximale de chaleur dissipation par un processeur sous une charge de travail typique, souvent utilisé comme proxy pour estimer la consommation énergétique.

5. Temps d'utilisation (Operating Hours) :

- Le nombre total d'heures pendant lesquelles la machine virtuelle ou le processeur est en fonctionnement, influençant directement la consommation totale d'énergie et donc les émissions de CO₂.

Définition et exemple de calcul empreinte Carbone

$$\text{Empreinte carbone (kgCO}_2\text{e)} = \text{Consommation énergétique (kWh)} \times \text{Facteur d'émission (kgCO}_2\text{e/kWh)}$$

Définition :

- **Consommation énergétique (kWh)** peut être estimée à partir du TDP et de l'utilisation du processeur → Ce TDP peut-être trouver dans fiche technique du constructeur.
- **Facteur d'émission (kgCO₂e/kWh)** dépend du mix énergétique du centre de données.

Exemple de calcul :

Supposons qu'un processeur a un TDP de 150W (0.15 kW) et fonctionne à 50% d'utilisation pendant 24 heures. Le mix énergétique local a un facteur d'émission de 0.5 kgCO₂e/kWh.

1. Calcul de la consommation énergétique :

Consommation énergétique (kWh) = 0.15 kW x 24 heures x 0.5 (utilisation) = 1.8 kWh

2. Calcul de l'empreinte carbone :

Empreinte carbone (kgCO₂e) = 1.8 kWh x 0.5 kgCO₂e/kWh = 0.9 kgCO₂e

Ce calcul donne une estimation de l'empreinte carbone d'un processeur ou d'une machine virtuelle en fonction de la consommation énergétique et du mix énergétique local.

Le "e" dans "kgCO₂e" signifie "équivalent". L'abréviation "kgCO₂e" se lit "kilogrammes d'équivalent dioxyde de carbone". Cette métrique est utilisée pour exprimer l'impact climatique de divers gaz à effet de serre en termes d'équivalents de dioxyde de carbone (CO₂). Cela permet de comparer et de cumuler les effets des différents gaz à effet de serre sur le réchauffement climatique.

Sélection des colonnes dans le dataset de LotaCloud

Pour calculer l'empreinte carbone à partir de notre dataset provenant d'un client Azure, nous allons avoir besoin des colonnes qui fournissent des informations sur la consommation de ressources, l'emplacement géographique (car le mix énergétique varie selon la région), et les mesures spécifiques associées aux ressources utilisées. Voici les colonnes pertinentes à conserver dans un premier temps, une analyse plus fine nous permettra de valider ou pas notre choix:

1. **ResourceLocation** : Emplacement géographique de la ressource Azure, crucial pour déterminer le mix énergétique et les facteurs d'émission de CO2.
2. **Date** : Date de la consommation, utile pour suivre l'empreinte carbone sur une période donnée.
3. **MeterCategory** : Catégorie de mesure pour comprendre le type de service consommé (par exemple, calcul, stockage).
4. **MeterSubCategory** : Sous-catégorie de mesure, fournissant des détails supplémentaires sur le type de service.
5. **MeterName** : Nom de la mesure pour savoir précisément ce qui est mesuré (par exemple, heures de CPU, stockage en Go).
6. **UnitOfMeasure** : Unité de mesure utilisée pour la consommation (par exemple, heures, Go), nécessaire pour quantifier la consommation.
7. **Quantity** : Quantité de ressources consommées, fondamentale pour calculer l'empreinte carbone.
8. **ConsumedService** : Service consommé pour avoir une idée précise du service Azure utilisé.
9. **MeterRegion** : Région de la mesure, indiquant où la consommation a eu lieu, ce qui peut influencer les émissions en fonction de la source d'énergie.
10. **ResourceId** : Identifiant unique de la ressource Azure, pour une traçabilité précise des ressources.
11. **Tags** : Étiquettes associées à la ressource pour la classification et la gestion, utiles pour des analyses plus détaillées par projet ou département.
12. **EffectivePrice** : Prix unitaire effectif de la ressource consommée, bien que moins directement lié à l'empreinte carbone, cela peut être utile pour des analyses coût-carbone.

Ces colonnes devraient suffire pour calculer l'empreinte carbone des ressources consommées. Pour le calcul, il faudra associer la consommation (quantité) à un facteur d'émission spécifique à la région et au type de ressource, afin d'obtenir les émissions de CO2e correspondantes.