MESURER LES EMISSIONS DE CARBONE DES TRAITEMENTS ML

PROBLÉMATIQUE SOCIÉTALE

- Augmentation de la taille & complexité des modèles
- Augmentation de l'utilisation de modèles de ML

Consumption	CO_2e (lbs)
Air travel, 1 passenger, NY↔SF	1984
Human life, avg, 1 year	11,023
American life, avg, 1 year	36,156
Car, avg incl. fuel, 1 lifetime	126,000
Training one model (GPU)	
NLP pipeline (parsing, SRL)	39
w/ tuning & experimentation	78,468
Transformer (big)	192
w/ neural architecture search	626,155

Table 1: Estimated CO_2 emissions from training common NLP models, compared to familiar consumption.[1]

- Besoin d'un outil de mesure pour plusieurs buts :
  - Encourager la sobriété numérique
  - Point de vue Y. Bengio : Data Scientists doivent mesurer pour pouvoir mettre en place des taxes pigouviennes pour forcer les corporations à réduire leur empreinte carbone

Energy and policy considerations in NLP, Strubell et al, 2019, University of Massachusetts

00:00:00:00

EMPREINTE CARBON DANS LES TRAITEMENTS ML

Conception

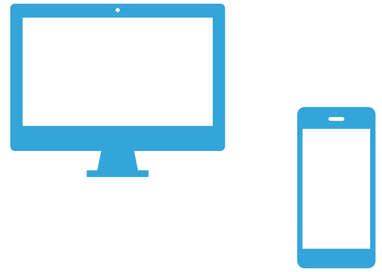
Utilisation des machines

Infrastructure de développement

Opérations

Infrastructure d’exploitation

Utilisation



Terminaux utilisateurs

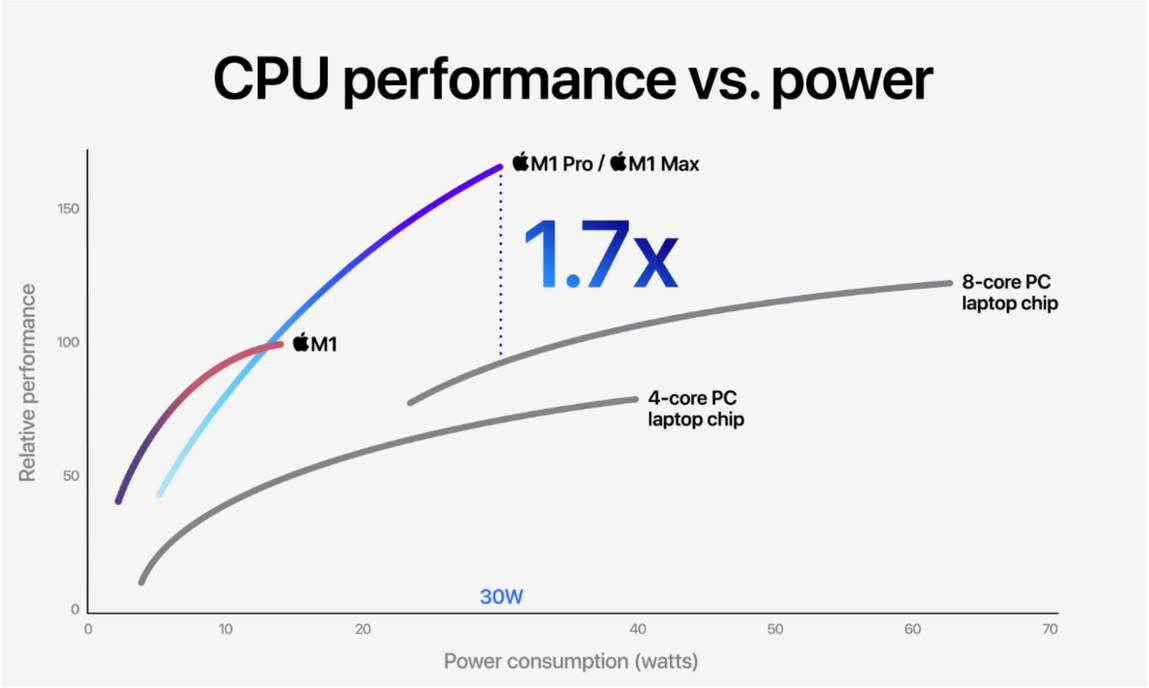
Réseaux

Serveurs mobilisés

MESURER LA CONSOMMATION D'UNE MACHINE PHYSIQUE

- Métrique de mesure de la consommation d'un CPU : taux d'utilisation * Thermal Design Power

"the power consumption (in watts) under the maximum theoretical load. Power consumption is less than TDP under lower loads. The TDP is the maximum power that one should be designing the system for."



MESURER LA CONSOMMATION D'UNE MACHINE PHYSIQUE

- Mesure lue, dans les OS équipés de processeurs compatibles, dans le fichier **RAPL**
- Utilisation de l'Intel Power Gadget pour les processeurs sans interface **RAPL**
- Consommation de la mémoire également grâce au TDP & dans le fichier **RAPL**

```
cpu0: MSR_IA32_TEMPERATURE_TARGET: 0x005a0a00 (90 C)
cpu24: MSR_IA32_TEMPERATURE_TARGET: 0x005a0a00 (90 C)
cpu0: MSR_IA32_PACKAGE_THERM_STATUS: 0x88300000 (42 C)
cpu0: MSR_IA32_PACKAGE_THERM_INTERRUPT: 0x00000000 (90 C, 90 C)
cpu24: MSR_IA32_PACKAGE_THERM_STATUS: 0x88370000 (35 C)
cpu24: MSR_IA32_PACKAGE_THERM_INTERRUPT: 0x00000000 (90 C, 90 C)
cpu17: MSR_PKGC3_IRTL: 0x00000000 (NOTvalid, 0 ns)
cpu17: MSR_PKGC6_IRTL: 0x00000000 (NOTvalid, 0 ns)
cpu17: MSR_PKGC7_IRTL: 0x00000000 (NOTvalid, 0 ns)
Package Core    CPU     Avg_MHz Busy%   Bzy_MHz TSC_MHz IRQ     SMI     CPU%c1  CPU%c6  CoreTmp PkgTmp  Pkg%pc2 Pkg%pc6 PkgWatt RAMWatt PKG_%   RAM_%
-       -       -       0       0.00    2164    3000    257     0       100.00  0.00    42      42      0.00    0.00    117.21  73.34   0.00    0.00
0       0       0       1       0.05    1783    3001    26      0       99.95   0.00    42      42      0.00    0.00    58.97   33.62   0.00    0.00
1       0       24      0       0.00    1200    3000    2       0       100.00  0.00    33      37      0.00    0.00    58.23   39.72   0.00    0.00
```

Estimating AWS EC2 instances power consumption, Benjamin DAVY (Teads)

MESURE DE CONSOMMATION ÉLECTRIQUE SUR GPU

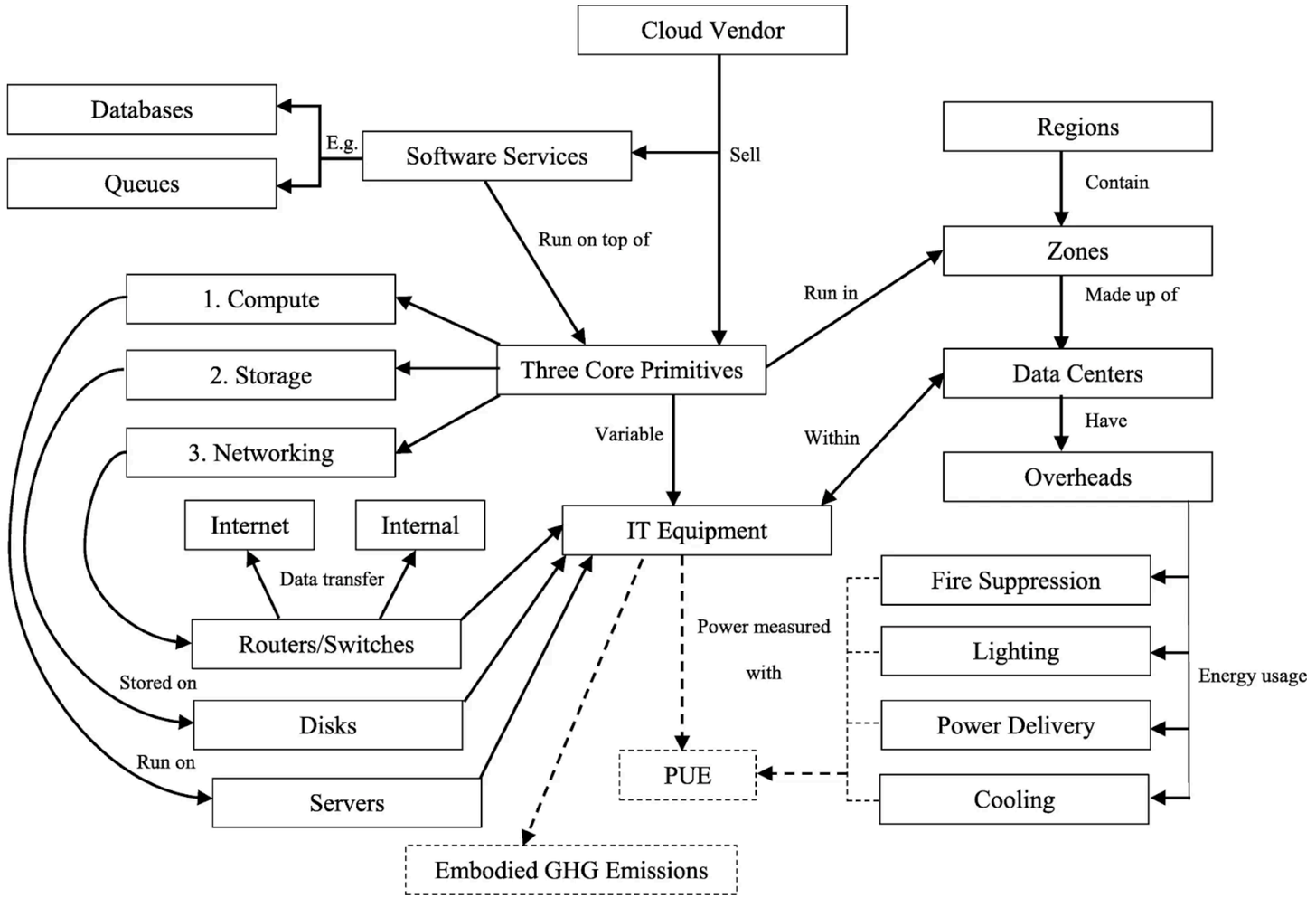
- L'interface SMI propose nativement une valeur de consommation lue directement par le driver NVIDIA

```
Every 2.0s: nvidia-smi

Fri Jun 29 11:25:43 2018
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 390.48                 Driver Version: 390.48                    |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|===============================+======================+======================|
|   0  Tesla M40           On   | 00000000:03:00.0 Off |                    0 |
|  0%   57C    P0   217W / 250W |  10806MiB / 11448MiB |     99%      Default |
+-------------------------------+----------------------+----------------------+
|   1  Tesla M40           On   | 00000000:04:00.0 Off |                    0 |
|  0%   58C    P0   200W / 250W |   9992MiB / 11448MiB |     99%      Default |
+-------------------------------+----------------------+----------------------+
|   2  Tesla M40           On   | 00000000:84:00.0 Off |                    0 |
|  0%   56C    P0   222W / 250W |   9955MiB / 11448MiB |     99%      Default |
+-------------------------------+----------------------+----------------------+
|   3  Tesla M40           On   | 00000000:85:00.0 Off |                    0 |
|  0%   57C    P0   203W / 250W |   9960MiB / 11448MiB |     98%      Default |
+-------------------------------+----------------------+----------------------+

+-----------------------------------------------------------------------------+
| Processes:                                                       GPU Memory |
|  GPU       PID   Type   Process name                             Usage      |
|=============================================================================|
|    0     35346      C   python2                                      110MiB |
|    0     55323      C   python                                     10673MiB |
|    1     55323      C   python                                      9970MiB |
```

MESURER LA CONSOMMATION ÉLECTRIQUE CHEZ UN CSP



Mesure d'efficacité énergétique

$$\mathrm{PUE} = \frac{\text{Total Facility Energy}}{\text{IT Equipment Energy}} = 1 + \frac{\text{Non IT Facility Energy}}{\text{IT Equipment Energy}}$$

Mesure d'efficacité de refroidissement

$$\mathrm{WUE} = \frac{\text{Annual Water Usage}}{\text{IT Equipment Energy}}$$

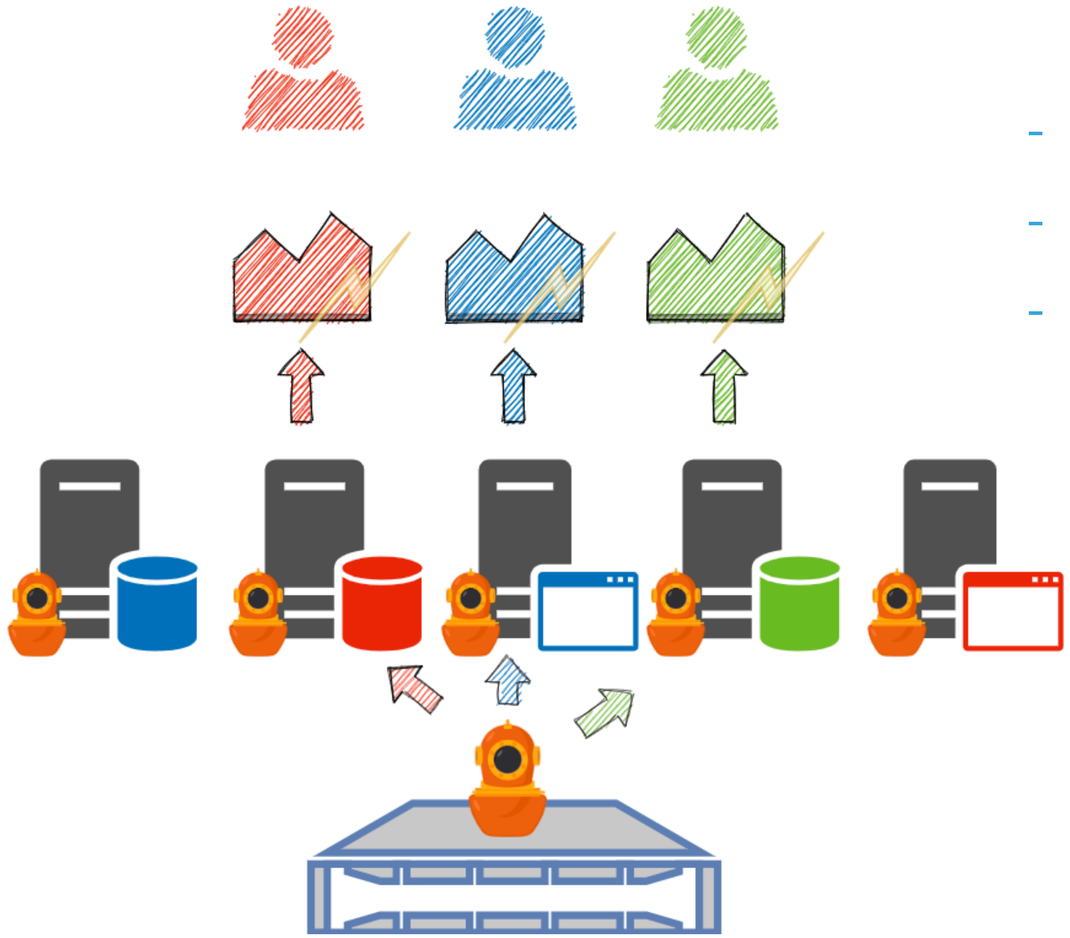
Assessing the suitability of the Greenhouse Gas Protocol for calculation of emissions from public cloud computing workloads, David Mytton

MESURE DANS UN ENVIRONNEMENT VIRTUALISÉ GRÂCE A SCAPHANDRE

End-users

VMs

Hyperviseur



- Données exportées grâce à un agent
- Volume partagé entre toutes les VMS
- Fichier des données lues depuis le RAPL exportées grâce à un Prometheus branché sur l'hyperviseur

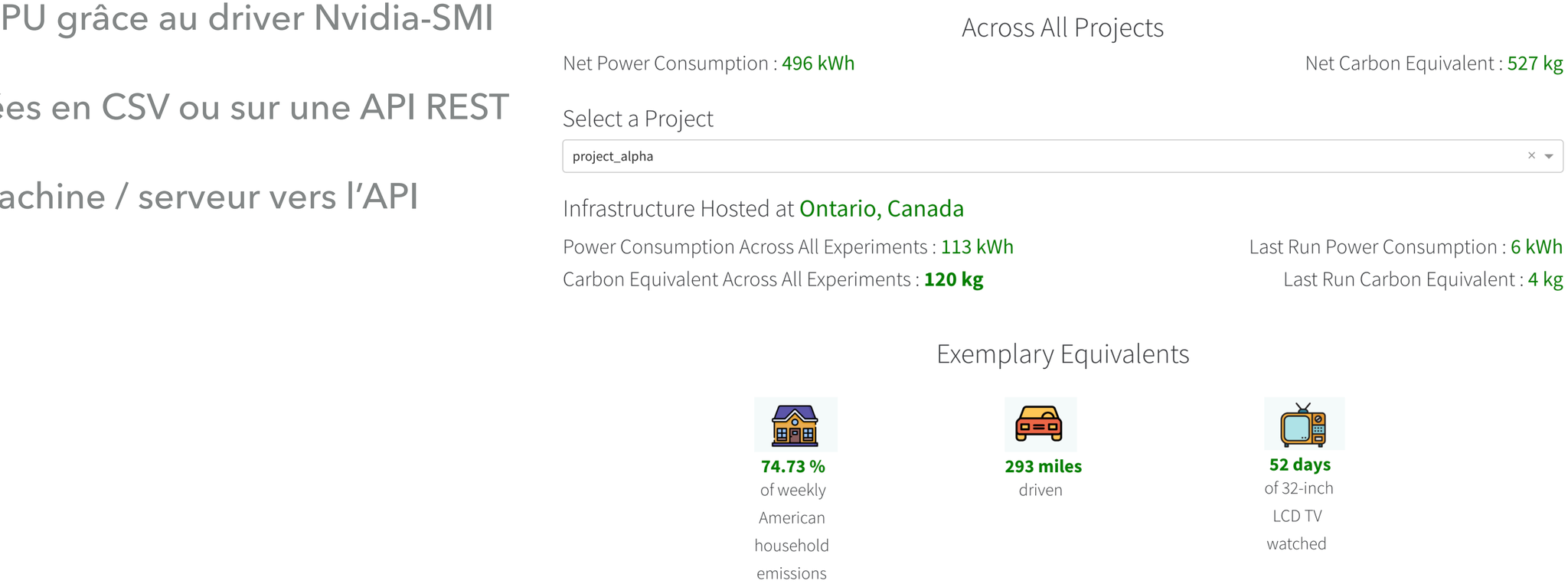
Propagate power consumption metrics from hypervisor to virtual machines (Qemu/KVM), Hubble.io

CODE CARBON

- Outil à destination des data scientists & développeurs pythons
- Lit le fichier RAPL ou ceux émis par l'Intel Power Gadget
- Fonctionne sur GPU grâce au driver Nvidia-SMI
- Données exportées en CSV ou sur une API REST
- Monitoring de machine / serveur vers l'API

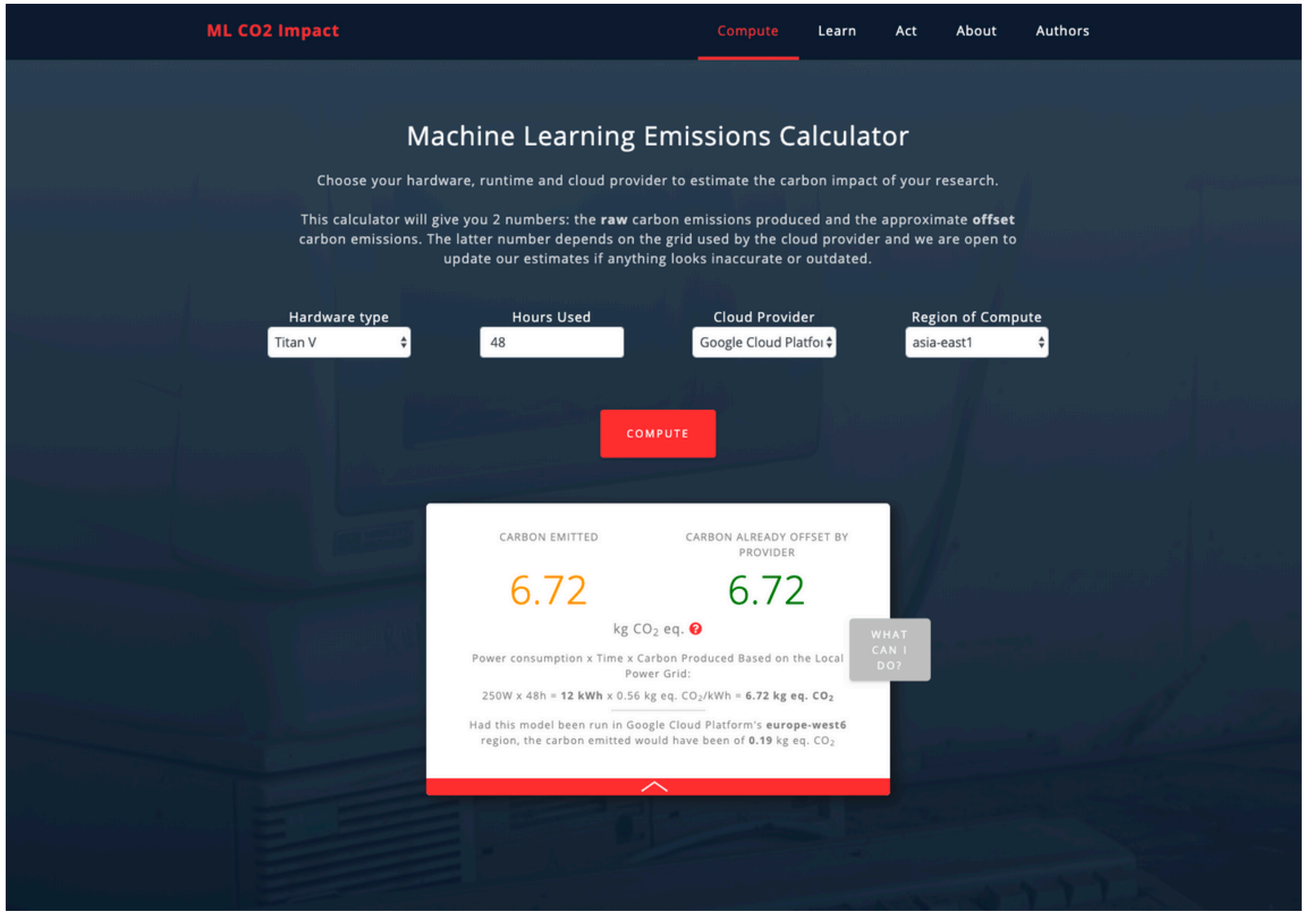
```python
from codecarbon import EmissionsTracker

tracker = EmissionsTracker()
tracker.start()
# GPU Intensive code goes here
tracker.stop()
```

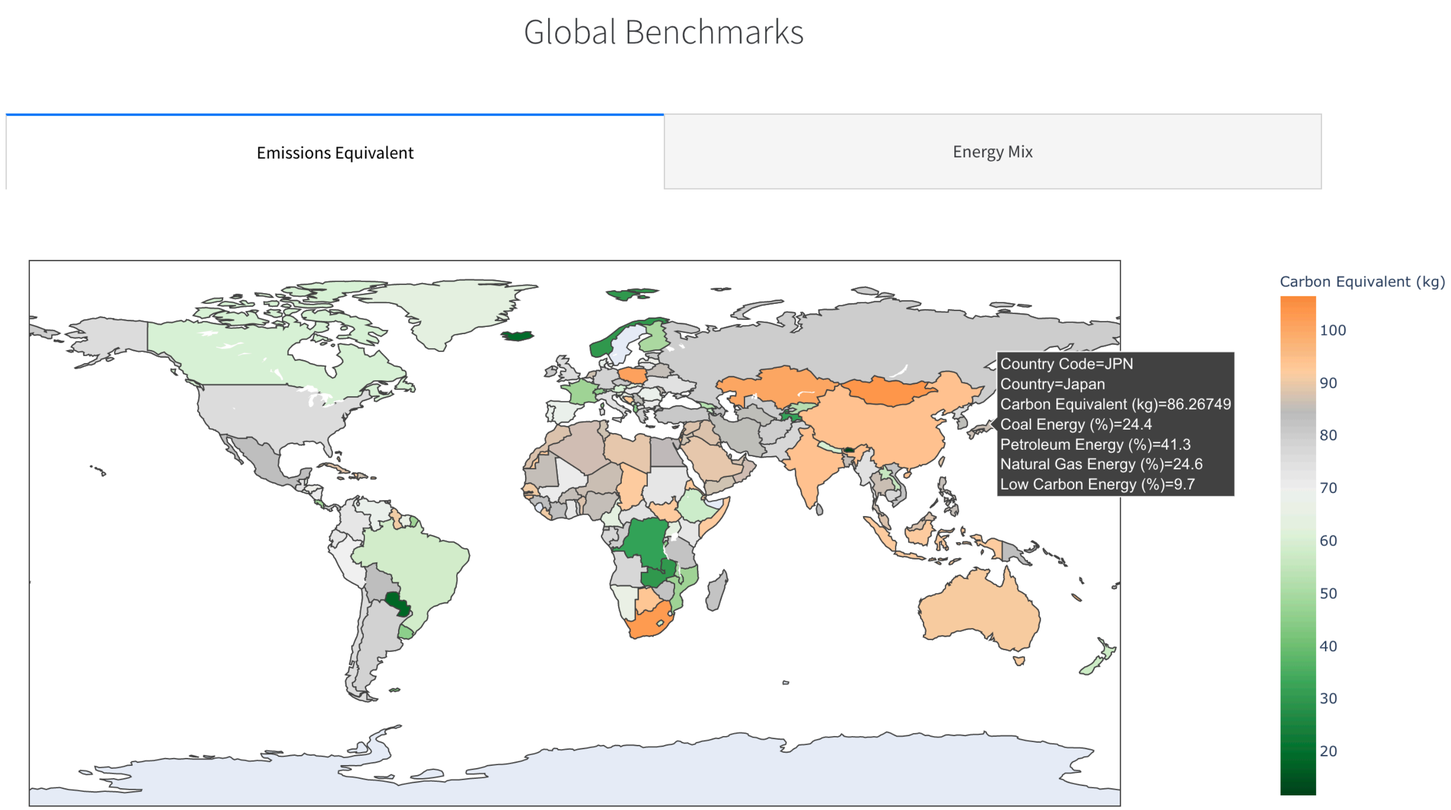


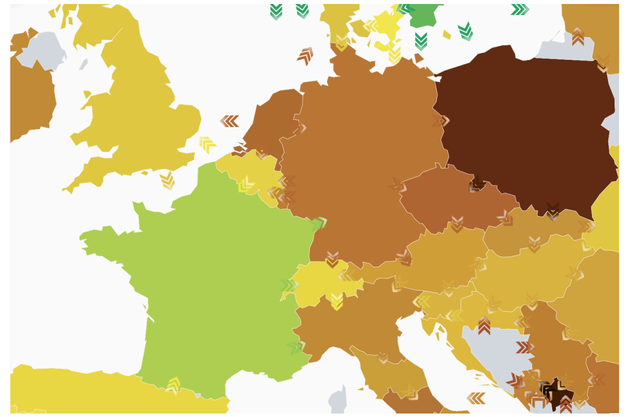
MLCO2 EMISSIONS CALCULATOR

- Choix du CSP & des data centers
- Comparaison du hardware
- Partage des émissions estimées



CODE CARBON : ESTIMATION DE L'EQUIVALENT CARBON





Basé sur l'API d'[ElectricityMap]

Croise les données de consommation avec celles de mix énergétique de la localisation du traitement

Cartographie les données de PUE des gros cloud providers US

CODE CARBON USERS

divyshah / **text-categorization** ♡ like 1

Text Classification | PyTorch | Transformers | divyshah/autonlp-data-text-ca

Model card | Files and versions

Model Trained Using AutoNLP

- Problem type: Multi-class Classification
- Model ID: 38039618
- CO2 Emissions (in grams): 214.76275947482927

Status	Name	Tags	Server end time	File name	Duration	user_name	steps
✓	f2466290d		11/11/20 06:43 PM	Jupyter interactive	00:00:48	dsblank	1875

Charts | Panels | Code | Hyper Parameters | Metrics | Graph Definition | Output | System Metrics | Installed Packages | Notes | Graphics | Audio

Panels — Auto Refresh | Refresh | Edit Layout | Add Panel

CodeCarbon Footprint

CODE CARBON

Select an Experiment: f2466290de4e462b9cc9b32476e925 or Analyze Project

Showing results for experiment:

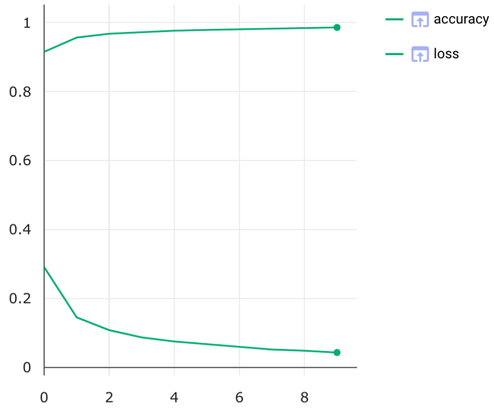
Infrastructure Hosted at United States

Power Consumption : 0.0003 kWh

Carbon Equivalent : **0.0001 kg**

Exemplary Equivalents

accuracy,loss VS epoch



BigScience | Mila | Buster.Ai

ROADMAP

- Sortie d’une V2 :
  - Amélioration de l’estimation (RAM, Windows, méthodes de fallback)
  - Intégration d’une API & dashboard communautaire
  - Support de nouvelles puces
- A venir :
  - Application web sécurisée et déployable
  - Amélioration du support des environments de calcul intensif (HPC, distribués, etc.)

RESOURCES

- Package documentation & repository
- Quantifying emissions in Machine Learning, V. Schmidt et al., 2019, MILA
- MLCo2 emissions calculator & manifest, MILA
- Code Carbon API