

... et tenter de la réduire !

Estimer l'empreinte écologique de PostgreSQL



Estimer l'empreinte écologique de PostgreSQL

juin 2021

Dalibo SCOP

Estimer l'empreinte écologique de PostgreSQL

... et tenter de la réduire !

TITRE : Estimer l'empreinte écologique de PostgreSQL

SOUS-TITRE : ... et tenter de la réduire !

DATE: juin 2021

Estimer l'empreinte écologique de PostgreSQL

A PROPOS

- Je suis Damien Clochard
- DBA chez Dalibo / Président de l'association PostgreSQLFr
- Cette présentation est une synthèse d'un travail collectif
- [Démarche globale lancée en 2019]

Les personnes suivantes ont participé à la rédaction de cette étude :

Thibaut Walkowiak, Thibaud Madelaine, Christophe Courtois, Benoit Lobréau, Franck Boudehen, , Florent Jardin, Philippe Beaudouin

AU MENU

- **Définir** : De quoi parle t' on ?
 - **Mesurer** : Indicateurs et Outils de Mesures
 - **Réduire** : Bonnes pratiques
-

DEFINIR

EMPREINTE ÉCOLOGIQUE

Définition large et floue !

« Pression exercée sur la nature par une activité humaine »

Exemple :

- Artificialisation des sols
 - Consommation d'énergies non-renouvelables
 - Production de Gaz à Effet de Serre (GES)
-

EMPREINTE CARBONE

« Mesure des émissions de GES imputées à une activité humaine »

L'empreinte carbone est une facette parmi d'autres de l'empreinte écologique

MAIS....

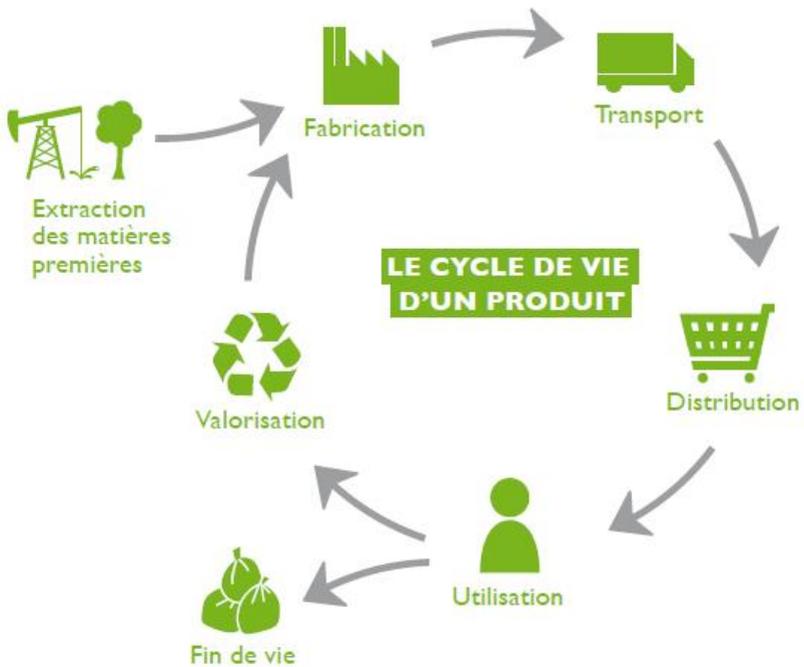
- Masque d'autres aspects (terres rares, articialisation, biodiversité, ...)
 - Le nucléaire est non-carbonné mais est-ce une "énergie durable" ?
 - L'économie de la "Compensation Carbone" sous le feu des critiques
-

Estimer l'empreinte écologique de PostgreSQL

INSTANCE POSTGRESQL

- Le logiciel PostgreSQL lui-même (moteur)
 - Les données
 - Les sauvegardes
 - Le trafic réseau
-

ANALYSE DU CYCLE DE VIE (ACV)



https://fr.wikipedia.org/wiki/Analyse_du_cycle_de_vie

ENERGIE GRISE

La quantité d'énergie consommée lors du cycle de vie d'un produit.

Egalement appelée « [énergie intrinsèque] »

énergie intrinsèque

MESURER

EQUIVALENT CARBONE

- Unité de mesure de l'empreinte carbone
 - "kilogrammes equivalent CO2" ou kgCO2eq
 - Mesure la production de différent gaz à effet de serre
 - C'est un outil abstrait pour la reflexion, pas une mesure réelle
 - Assez simple à utiliser
-

POWER USAGE EFFECTIVENESS (PUE)

Indicateur d'efficacité énergétique qui quantifie l'efficacité énergétique d'un datacenter.

- Conso Globale du centre divisée par la somme des conso de chaque équipement
 - En moyenne les datacenters francais ont un PUE de 2,5
 - pour 1 Watt consommé par l'informatique, il en faut 2,5 Watt à l'entrée du datacenter.
 - PUE d'un datacenter moderne : 1,5
 - PUE d'un datacenter sans climatisation : 1,10
-

ACV D'UN SERVEUR INFOMATIQUE

- DELL PowerEdge R640 - 1,2 TB - 32GB RAM - 2 CPU
- Conso électrique : 1760 kWh par an
- Facteur d'Emission (en France) : 0.053 kgCO2eq par kWh
- PUE du datacenter : 1.5
- **Fabrication** : 1300 kgCO2eq
- **Utilisation** : 140 kgCO2eq par an

Pour comparaison, voici le Bilan GES pour d'autres équipements informatiques:

https://www.bilans-ges.ademe.fr/documentation/UPLOAD_DOC_FR/index.htm?ordinateurs_et_equipement

La conception d'un ordinateur ou d'un téléphone, très polluante, a des répercussions sur l'environnement. Un ordinateur émet de 20 à 50 fois plus de CO2 lors de sa fabrication qu'au cours de son utilisation

<https://bfmbusiness.bfmtv.com/01-business-forum/les-ravages-insoupconnes-de-la-pollution-informatique-593131.html> <https://www.networkworld.com/article/2229029/computer-factories-eat-way-more-energy-than-running-the-devices-they-build.html>

L'extraction est beaucoup plus impactante <https://www.ibm.com/downloads/cas/GYR3MWQN> que le transport (https://fr.wikipedia.org/wiki/Impact_environmental_de_la_m%C3%A9tallurgie). <https://ecoinfo.cnrs.fr/wp-content/uploads/2019/09/Informatique-et-d%C3%A9veloppement-soutenable-LIMSI.pdf?#page=18>

AVC D'UNE INSTANCE POSTGRESQL

1 base applicative de 500Go

- Pendant 10 ans
- 2 instances de production en Hot Standby
- 3 instances de pré-prod / dev / qualif
- 1 instance développement / Qualif
- 1 semaine de sauvegarde PITR + 4 dumps
- remplacement du matériel tous les 4 ans

Pour comparaison, voici le Bilan GES pour d'autres équipements informatiques:

https://www.bilans-ges.ademe.fr/documentation/UPLOAD_DOC_FR/index.htm?ordinateurs_et_equipe

EMPREINTE SUR 10 ANS

- 1 instance = 4.6 TCO2eq sur 10 ans
- total pour l'appli = 32 TCO2eq sur 10 ans

En comparaison :

- un AR Paris-New York pour 1 passager = 1TCO2eq
- Dalibo en 2019 : 70 TCO2eq

AUTRE APPROCHE

- Définir un facteur volumétrique
- Exemple : "1 TB pendant 1an = 2000 kgCO₂eq"
- Trop simpliste
- Trop imprécis
- Pas de sources fiables

- Etude 1 : 320 kgCO₂e par an
- Etude IBM : 6319 kwh /an x 0,5 kgco₂eq par kwh = 3 TCo₂eq
- Etude DELL : 1,6TCo₂e par an
- https://i.dell.com/sites/csdocuments/CorpComm_Docs/en/carbon-footprint-poweredge-r640.pdf
- <https://www.goclimat.com/blog/the-carbon-footprint-of-servers/>
- <https://www.ibm.com/downloads/cas/GYR3MWQN>

Données active : 7kWh/GB ou 3kWh/GB hard disk, which requires about 0.000005 kWh per gigabyte to save your data,

1 TB pendant 1an = 2TeqCO₂

- <https://stanfordmag.org/contents/carbon-and-the-cloud>
 - <https://www.aceee.org/files/proceedings/2012/data/papers/0193-000409.pdf>
-

REDUIRE

GRANDS PRINCIPES

- Parier sur l'endurance : Allonger la durée de vie des serveurs
 - Viser la Sobriété : Sélectionner les fournisseurs
 - Eliminer le superflu : Réduire les volumes
 - S'équiper léger: Dimensionner correctement
 - Mutualiser : Regrouper les ressources
-

5 AXES PRINCIPAUX

- 1- Matériel
 - 2- OS
 - 3- PostgreSQL
 - 4- Applicatif / SQL
 - 5- Architecture
-

1- HÉBERGEMENT

- Quel est le PUE du datacenter ?
 - Quel est la source d'énergie du datacenter ?
 - DCIM : datacenter infra management
-

Estimer l'empreinte écologique de PostgreSQL

1- MATÉRIEL

- Tout se joue à l'achat !
- Intégrer l'empreinte écologique dans les critères de choix

Exemple :

Les disques SSD consomme 5X moins et produisent moins de chaleur

1- MATÉRIEL : DURABILITÉ

- Repenser « l'amortissement » !
 - Amortissement comptable sur 3 ans
 - "Pousser" la durée de vie à 5 à 6 ans, voire 8
-

2- OS : PRÉFÉREZ LINUX !

- Linux / Postgres : marche sur du matériel ancien
 - Pas obsolescence programmée dans les logiciels libres
-

2- OS : MUTUALISATION DU STOCKAGE / VIRTUALISATION

- Dimensionner les ressources "au plus juste"
 - La virtualisation complique le calcul du bilan GES
-

3- POSTGRESQL : OPTIMISER

- Optimiser la configuration !
 - La config par défaut est minuscule
 - Réécrire les requêtes lentes (EXPLAIN)
 - Identifier les requêtes courtes mais très nombreuses (powa / pgbadger)
-

3- POSTGRESQL : NETTOYER

- taille sur disque > taille réelle
 - Surveiller la fragmentation de la base ("bloat")
 - Surveiller l'autovacuum
-

4- APPLICATION : MINIMISATION

- Tordre le coup à la "Mentalité Big Data"
 - Supprimer les champs inutiles
 - Limiter la profondeur d'historique (pour chaque table)
 - La réduction des volumes est bonne pour l'optimisation des perfs
-

4- APPLICATION : MINIMISATION

- Bonus : C'est un principe RGPD !

Les données stockées doivent être adéquates, pertinentes et limitées à ce qui est nécessaire au regard des finalités pour lesquelles elles sont traitées.

4- APPLICATION : DONNÉES OBSOLÈTES, INACTIVES OU ARCHIVABLES

- Supprimer les index inutiles
- Partitionner
- Archiver
- Aggréger les données obsolètes
- Externaliser les données "moins actives" (fdw)

Attention: pour les index, un oubli peut couter cher en CPU

Piste à explorer : compression du FS : tradeoff entre CPU et Stockage

Estimer l'empreinte écologique de PostgreSQL

5- ARCHITECTURE : HAUTE-DISPONIBILITÉ

- cluster HA = facteur 2 à 5
 - -> A-ton besoin d'une haute dispo 99,99% ?
 - -> Quel est le bilan GES du PRA ?
-

5- ARCHITECTURE : MICROSERVICES - MACROCONSOMMATION

- PostgreSQL est conçu sur des principes forts de mutualisation
 - 1 instance = N bases + N Schemas + N Roles
 - Déployer une instance par microservice est un contresens
-

5- ARCHITECTURE : LE CLOUD DE TARTUFFE

« Cachez cette pollution que je ne saurais voir.... »

Le grand opérateur communiquent publiquement sur le Green IT

- Mutualisation
 - Location / Flexibilité
 - Expertise (datacenters) / Rationalisation
 - Corrélation entre volumes et tarif
-

5- ARCHITECTURE : LE CLOUD DE TARTUFFE

MAIS...

- Peu d'information sur la provenance du matériel, le retraitement, etc.
- Incitation à la "Sur-consommation" de services
- HA trop "facile" d'accès
- Logique de déresponsabilisation

Ce sont vos données, vous êtes responsables de leur empreinte, même si elles sont sur du matériel qui ne vous appartient pas.

CONCLUSIONS

PostgreSQL n'est pas le plus gros producteur de GES dans le SI

MAIS c'est un point central dans toute démarche d'éco-conception

- Collecter les données de base
 - Evaluer le bilan GES d'une instance PostgreSQL standard
 - Engager des actions de réduction
-

AGIR À TOUS LES NIVEAUX

- Pression sur les vendeurs de matériels (Transparence)
 - Pression sur les datacenters (PUE)
 - Pression sur les développeurs (Minimisation / Optimisation)
 - Réflexion sur les usages (Disponibilité / Archivage)
-

ATTENTION À L'EFFET REBOND !

- Les économies d'énergies génèrent un gain de productivité
- Le gain de productivité génère des bénéfices
- Les bénéfices sont réinjectés pour produire plus
- La hausse de production génère une surconsommation d'énergie

Les gestes "individuels" doivent s'inscrire dans une démarche globale