

Document de cadrage

Projet de Visualisation Interactive de Données

Jules Bonhotal
Sherine Baya Kochkar
Yanis Aumasson
Gernido Hanampatra

Introduction

Ce projet vise à créer une visualisation interactive portant sur la consommation énergétique des centres de calcul. L'objectif principal est de représenter et d'explorer les différentes variables liées à la consommation énergétique et de proposer une visualisation adaptable à toutes les échelles de temps. Cela permettra d'identifier des opportunités d'optimisation, de réduire les coûts énergétiques et de limiter l'impact environnemental.

Problème abordé et besoin identifié

Les centres de calcul consomment une quantité importante d'énergie, avec des impacts environnementaux et économiques significatifs. Ce projet vise à répondre aux questions suivantes :

- Comment visualiser et analyser les variations temporelles de la consommation énergétique de manière interactive ?
- Comment permettre une visualisation intuitive des périodes de fortes et faibles consommations énergétiques, sur des échelles journalières, mensuelles et annuelles ?
- Comment mettre en valeur les activités inhabituelles pour différentes échelles de temps ?

Problématiques spécifiques identifiées :

1. Mettre en valeur les corrélations entre les variables (charge de travail, température ambiante, etc.) et la consommation énergétique.
2. Comparer l'efficacité énergétique entre différents centres de calculs ou infrastructures.
3. Visualiser les périodes de pic ou de creux énergétique pour aider à la planification et à l'optimisation.
4. Prévoir la consommation énergétique en fonction des tendances passées.

Public cible et tâches principales

La visualisation interactive s'adresse principalement aux publics suivants :

— **Gestionnaires de centres de calcul :**

Responsables de l'efficacité opérationnelle des data centers, ils ont besoin d'outils pour :

- Identifier rapidement les périodes de fortes et faibles consommations énergétiques afin de planifier des actions correctives.
- Évaluer l'impact des charges de travail sur la consommation énergétique.
- Optimiser la répartition des ressources pour réduire les coûts et l'empreinte carbone.

— **Chercheurs en informatique et ingénierie environnementale :**

Travaillant sur des solutions durables pour réduire l'impact environnemental des infrastructures numériques, ces chercheurs pourront :

- Explorer les corrélations entre la consommation énergétique et des facteurs externes (par exemple, température, humidité).
- Développer de nouveaux modèles de prédiction basés sur des visualisations des données existantes.
- Comparer différents scénarios d'optimisation énergétique.

— **Décideurs politiques et responsables de conformité environnementale :**

Chargés de définir des politiques de durabilité et de conformité, ils pourront utiliser cette visualisation pour :

- Suivre la conformité des centres de calcul aux normes environnementales en matière de consommation énergétique.
- Justifier des initiatives législatives pour réduire l'empreinte écologique des infrastructures numériques.
- Identifier les centres de calcul nécessitant des investissements prioritaires pour des améliorations énergétiques.

— **Opérateurs de maintenance :**

Directement impliqués dans la gestion quotidienne des équipements, ils utiliseront la visualisation pour :

- Détecter les anomalies dans la consommation énergétique en temps réel.
- Anticiper les besoins de maintenance liés aux périodes de stress énergétique élevé.
- Faciliter la prise de décision en cas de défaillances ou de surchauffes.

Trois tâches principales que permettra la visualisation :

1. Identifier les variables ayant le plus d'impact sur la consommation énergétique.
2. Comparer la consommation énergétique à travers différentes échelles

temporelles (heure, jour, mois, année).

3. Visualiser les variations saisonnières ou liées à des événements spécifiques, comme des pics d'utilisation.
4. Fournir une interface intuitive permettant de distinguer les périodes de fortes et faibles consommations énergétiques.

Sources de données choisies

- **Source principale :** *Data Server Energy Consumption Dataset.*
Intérêt principal : Données détaillées sur la consommation énergétique à différentes échelles temporelles d'un data center en Équateur.
Limites potentielles : Nécessite un prétraitement important pour harmoniser les formats, combler les données manquantes, et se ramener à un jeu de données exploitable avec D3.
Source du Dataset
- **Source Secondaire 2 :** *Top 50 Data Center Markets by Power Consumption.*
Table de données contenant la consommation en megawatts de chaque groupe de data centers par pays.
Intérêt principal : Ces données permettraient de faire une mise en abyme de la consommation de notre data center à l'échelle mondiale.
Limites potentielles : Contient uniquement la consommation des data centers agrégée par région, qui plus est seulement le top 50. Ce sera donc difficile de faire une comparaison pertinente avec un seul data server.
Source du Dataset
- **Source Secondaire 3 :** *Cost of Electricity by country*
Intérêt principal : Estimer le coût associé à la consommation énergétique du data center.
La première source ne contient que des données de 2022 à 2024 pour tous les pays. La seconde celle de 2017 à 2023 avec les moyennes des prix pour chaque marché (résidentiel, entreprise, vente en gros, industriel), mais uniquement pour l'Équateur. À priori on utilisera la seconde source.
World Population Review ou Global ClimateScope
- **Source Secondaire 4 :** *Données météorologiques locales et historiques.*
Intérêt principal : Analyser l'impact des températures et des saisons sur la consommation énergétique.
Limites potentielles : Peut être difficile d'obtenir des données précises pour certains centres.
Source NOAA Climate Data Online
- **Source Secondaire 5 :** *PUE (Power Usage Effectiveness) et autres*

indicateurs de performance énergétique des data centers.

Intérêt principal : Fournir des mesures standards sur l'efficacité énergétique des centres de calcul pour des comparaisons globales.

Limites potentielles : Les données peuvent ne pas être disponibles pour certains centres, ou manquer de granularité.

Source Data Center Map

- **Source Secondaire 6 :** *Données ouvertes sur l'énergie nationale ou régionale.*

Intérêt principal : Comparer la consommation énergétique des data centers avec la consommation énergétique globale ou des données régionales.

Limites potentielles : La granularité peut ne pas être suffisante pour des comparaisons détaillées.

Source World Energy Data

- **Source Secondaire 7 :** *Open Compute Project (OCP) Data.*

Intérêt principal : Données ouvertes provenant d'une initiative internationale visant à améliorer l'efficacité des data centers.

Limites potentielles : Les données sont parfois limitées à des centres de calcul pilotes.

Source Open Compute Project

Travaux importants liés au projet

Voici une liste de travaux pertinents utilisant des datasets similaires ou ayant des objectifs proches :

- **Projet 1 :** *Optimisation énergétique des data centers grâce à des visualisations interactives*

Dataset utilisé : Données énergétiques historiques de centres de calcul.

Contribution : Ce projet propose des visualisations permettant d'identifier les périodes de fortes consommations énergétiques, d'analyser les inefficacités et de guider les gestionnaires dans leurs décisions d'optimisation.

Visualisation associée : Consommation d'un data center : tout savoir pour limiter les dépenses.

- **Projet 2 :** *Analyse des variations saisonnières de consommation énergétique dans les infrastructures IT*

Dataset utilisé : Logs énergétiques croisés avec des données météorologiques locales.

Contribution : Ce projet met en avant l'impact des variations climatiques (température, humidité) sur la consommation énergétique des data centers, avec des visualisations temporelles interactives.

- Visualisation associée : Comprendre la consommation énergétique des data centers.
- **Projet 3 : Crédit à un tableau de bord énergétique pour les centres de calcul intelligents**
- Dataset utilisé : Données énergétiques collectées en temps réel via des capteurs IoT.
- Contribution : Ce tableau de bord interactif permet un monitoring en temps réel de la consommation énergétique, mettant en évidence les pics, les anomalies et les inefficacités.
- Visualisation associée : Consommation d'énergie dans un data center.

Organisation du projet

Moyens de communication

Pour assurer une communication efficace et un suivi rigoureux du projet, nous utilisons les outils suivants :

- **Git (GitHub) :**
Utilisé pour :
 - La gestion de version du code source.
 - Le suivi des tâches via les issues.
 - Le partage de documentation à travers le wiki intégré.
 - La publication du projet en ligne via GitHub Pages.
- **Discord :**
Utilisé pour :
 - La communication rapide (textuelle et vocale) entre les membres.
 - L'organisation de réunions hebdomadaires pour synchroniser les avancées.
 - Le partage instantané de ressources et de documents.

Planning

Le projet est organisé en plusieurs phases claires pour garantir un avancement structuré :

- **Phase 1 : Collecte et nettoyage des données.**
Objectifs :
 - Récupérer les données provenant des sources identifiées.
 - Prétraiter les données (harmonisation, gestion des valeurs manquantes, nettoyage des anomalies).
 - Réaliser une analyse exploratoire pour identifier les tendances principales.
- **Phase 2 : Développement des premières visualisations interactives.**
Objectifs :

- Créer des prototypes basés sur les premières esquisses.
- Implémenter les visualisations de base à l'aide de D3.js ou d'autres outils.
- Tester les interactions principales (zoom, filtres, survol d'éléments).
- **Phase 3 : Améliorations et intégration des retours utilisateurs.**
Objectifs :
 - Affiner les visualisations en fonction des retours des pairs et du professeur.
 - Optimiser les performances des graphiques pour les datasets volumineux.
 - Ajouter des éléments interactifs supplémentaires pour enrichir l'expérience utilisateur.
- **Phase 4 : Présentation finale et documentation.**
Objectifs :
 - Préparer une démonstration claire et fluide pour la soutenance.
 - Finaliser la documentation technique et fonctionnelle.
 - Publier le projet sur GitHub Pages.

Rôles au sein de l'équipe

Chaque membre de l'équipe se concentre sur ses responsabilités :

Tâches	Gernido Hanampatra	Yanis Aumasson	Jules Bonhotal	Sherine Baya Kochkar
Prétraitement	X	X		
Développement		X	X	X
Design	X		X	
Coordination				X

TABLE 1 – Répartition des rôles au sein de l'équipe

Esquisses et premières pistes

Les premières esquisses papier ou numériques seront incluses ici pour définir les concepts visuels envisagés.

Exemples d'idées à explorer :

- **Timeline interactive** : Permet de visualiser les variations de consommation énergétique sur différentes échelles temporelles (heure, jour, mois, année) avec un zoom fluide.
- **Diagrammes en cercles concentriques** : Illustrer les variations saisonnières ou périodiques, facilitant l'identification des schémas récurrents.
- **Tableau de bord complet** : Combine plusieurs types de graphiques :
 - Graphiques temporels pour montrer les tendances.

- Cartes géographiques pour visualiser la consommation par région (si applicable).
- Histogrammes pour comparer les périodes de forte/faible consommation.



FIGURE 1 – Fiches de Brouillons de la séance du 3/12/2024