



Projet ANR DATAZERO (2015 – 2019)

Jean-Marc Pierson¹



Les centres de données nécessitent une gestion énergétique optimisée pour gérer la charge informatique résultant de la demande toujours croissante de services en ligne. En général, les centres de données alimentés en électricité produisent des émissions de CO₂ ou des déchets nucléaires. Le développement des énergies renouvelables ouvre la voie à une alimentation des centres de données en électricité locale, plus propre.

L'objectif du projet DATAZERO est d'étudier les solutions possibles pour concevoir et exploiter au mieux un centre de données, alimenté uniquement par des énergies renouvelables, en utilisant un stockage d'énergie à long et à court terme.

Il existe un besoin d'optimiser la charge informatique par rapport à la disponibilité énergétique, et inversement, d'optimiser la production d'énergie par rapport à la charge informatique entrante. Au lieu de résoudre un problème centralisé, nous avons proposé une approche innovante où la partie informatique et la partie électrique gèrent et optimisent leurs opérations indépendamment et négocient ensemble lorsque nécessaire. En plus de la conception de processus de négociation efficaces, plusieurs aspects ont été étudiés pour obtenir une vue globale du problème : le contrôle efficace de la puissance électrique, la répartition de cette puissance depuis les différentes sources d'énergie, la planification efficace des tâches sur les serveurs informatiques, sans oublier les problèmes sous-jacents de dimensionnement électrique et informatique.

1. Professeur à l'Université Toulouse 3 Paul Sabatier, porteur du projet.

Les partenaires du projet DATAZERO sont l'IRIT et le LAPLACE à Toulouse, le FEMTO-ST à Besançon et Belfort, et l'entreprise EATON.



Enjeux et problématique

Les centres de données nécessitent beaucoup d'énergie pour gérer la charge informatique résultant de la demande toujours croissante de services en ligne. Plusieurs études montrent que la demande énergétique de l'industrie informatique augmente à grande échelle et à une vitesse fulgurante, notamment en raison de l'industrie des *datacenters* [1, 2, 3].

Aujourd'hui, les centres de données s'appuient sur un ou plusieurs réseaux électriques pour fournir aux installations informatiques la puissance nécessaire. L'électricité provenant de ces réseaux peut cependant être plus ou moins propre, en termes d'émission de CO_2 ou de déchets nucléaires. Le développement des énergies renouvelables ouvre ainsi la voie à une alimentation plus propre des *datacenters*. Des initiatives innovantes font la transition des architectures traditionnelles aux centres de données alimentés, au moins en partie, par l'énergie renouvelable des panneaux solaires (Aiso, Phoenix IT, Intel, Sun, Google, HP), des éoliennes (Google, OWC, Green House Data, Baryonyx), de la géothermie et de l'hydroélectricité (Greencloud), de l'hydroélectricité (Yahoo, Facebook) ou des générateurs basés sur les piles à combustible (Panasonic). Pour montrer leur implication vers des solutions plus durables, les opérateurs de centres de données, comme Apple et Google, organisent leur production d'électricité verte localement pour ce qui est possible, et souvent ont recours à des fournisseurs d'électricité tiers et paient des certificats « verts » (RECS pour *Renewable Energy Certificate System*). Cela favorise indirectement la production d'énergie renouvelable et les opérateurs peuvent faire connaître leurs efforts dans ce sens. Il convient de mentionner que cette solution est principalement une astuce commerciale, car les électrons proviennent toujours des usines les plus proches (peut-être au charbon). Elle présente également plusieurs inconvénients par rapport à la production locale, tels que les pertes de transport et de distribution (près de 10 %).

Construire un *datacenter* uniquement alimenté en énergies renouvelables locales est cependant un défi. La production d'énergie renouvelable est intermittente et la charge informatique est variable, c'est-à-dire que les panneaux photovoltaïques sont inefficaces la nuit, les éoliennes ne produisent de l'électricité qu'avec le vent et les utilisateurs soumettent leur travail quand ils le souhaitent, et leurs périodes de pointe

ne coïncident pas nécessairement. Il est donc essentiel de stocker l'énergie pour les faire correspondre. Les batteries, généralement déployées dans les centres de données pour gérer les incidents à court terme sur la ligne électrique, ne sont cependant pas en mesure de couvrir les incidents qui durent plus de quelques minutes ou heures. Utiliser des électrolyseurs pour produire de l'hydrogène (H_2) à partir d'énergie non consommée, la stocker dans des réserves à hydrogène et utiliser des piles à combustible pour produire de l'électricité à nouveau en cas de besoin est donc complémentaire des batteries, des éoliennes et des panneaux photovoltaïques. En règle générale, les batteries sont utiles pour stocker l'énergie supplémentaire non consommée pendant la journée pour ensuite être utilisée pendant la nuit, tandis que le stockage d'hydrogène permet de combler l'écart de production d'énergie entre les différentes saisons.

Des systèmes d'alimentation complexes doivent donc être mis en œuvre pour assurer le fonctionnement du centre de données. Concevoir et exploiter un *datacenter* utilisant des énergies renouvelables et un stockage à long terme n'est donc pas anodin. Outre le choix du type de stockage d'énergie à utiliser, il est nécessaire d'optimiser la charge informatique par rapport à la disponibilité énergétique, et inversement, d'optimiser la production d'énergie par rapport à la charge informatique entrante. Le dimensionnement du système en termes de composants informatiques et électriques est également un défi.

Approche et contribution

De tels problèmes d'optimisation peuvent être résolus de manière centralisée, lorsque l'ensemble du système est connu, du point de vue informatique et électrique. De telles solutions ont fait l'objet de nombreux travaux dans le passé, mais souffrent du manque de dynamisme lorsque quelque chose change sur l'un des côtés.

Nous avons proposé une approche innovante où la partie informatique et la partie électrique sont capables de gérer et d'optimiser leurs opérations de manière indépendante, mais aussi de négocier entre elles lorsque nécessaire. Par exemple, lorsque la charge informatique augmente et que davantage de puissance est nécessaire, ou lorsqu'un nuage couvre le soleil et que la production photovoltaïque chute, chaque partie peut commencer à négocier avec l'autre partie si elle ne peut pas résoudre le problème seule (cf. figure 1).

La principale contribution de notre approche est ainsi d'aborder les problématiques suivantes et de les regrouper dans un cadre commun :

- Privilégier une approche où chaque partie, électrique et informatique, optimise son propre problème. De cette façon, les deux problèmes d'optimisation peuvent être résolus en utilisant différentes méthodes, par exemple par heuristique ou programmation linéaire, et indépendamment, facilitant ainsi l'utilisation de solutions les mieux adaptées dans des cas particuliers.

- Proposer un processus de négociation donnant des conseils aux solveurs d'optimisation (gestion informatique et gestion électrique). Pendant la phase de négociation, les solutions pour les parties électriques et informatiques sont comparées et le meilleur compromis pour chaque côté est choisi.
- Inclure plusieurs sources d'énergie renouvelables et de stockage (panneaux solaires, éoliennes, supercondensateurs, piles à combustible H_2 , batteries), en modélisant précisément leur comportement avec des tests d'équipement et en incluant les effets du vieillissement dans une infrastructure de microréseaux complète.
- Intégrer la modélisation des profils d'application et des profils de charge du serveur, permettant l'exécution des applications à des performances dégradées, donnant plus de flexibilité lors de la planification des tâches, en plus des effets classiques sur l'équipement informatique et les tâches (allumer / éteindre les serveurs, changement de fréquence, migration).
- Proposer un intergiciel innovant facilitant la communication et la négociation entre le côté électrique et le côté informatique. L'innovation réside dans le fait que l'intergiciel gère à la fois le matériel réel, le matériel émulé, et la simulation, ou un mélange des trois (par exemple, une infrastructure électrique simulée et émulée, et des serveurs informatiques sous Openstack).

Résultats et exploitation

Les résultats individuels obtenus dans DATAZERO ont été publiés dans plusieurs

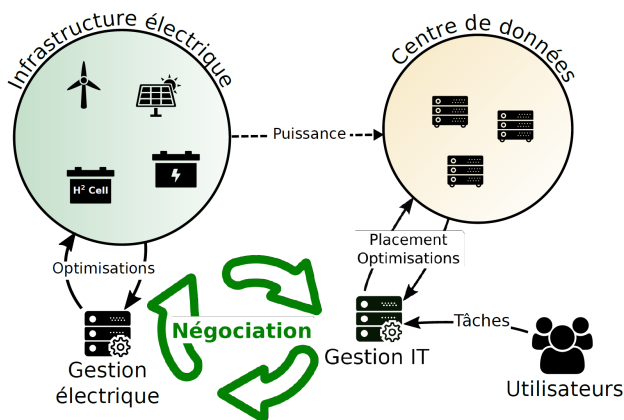


FIGURE 1. Une négociation est en cours entre la partie informatique traitant les tâches des utilisateurs et la gestion électrique traitant de l'intermittence des sources d'énergie. © Léo Grange.

conférences et revues de haut niveau tout au long du projet. Un document de recherche conjoint a été produit par le consortium et publié en *Open Access* dans IEEE Access [4].

En termes d'impact scientifique, DATAZERO est le premier du genre à avoir étudié en profondeur un centre de données fonctionnant uniquement avec des sources renouvelables. Il montre la possibilité technique et scientifique d'exploiter un tel centre de données et ouvre la voie à l'industrie pour la mise en œuvre d'une solution pleinement opérationnelle. Nous avons montré qu'il est possible d'éviter les émissions de CO_2 liées au fonctionnement des centres de données en traçant un futur sociétal et environnemental différent pour la pérennité des infrastructures informatiques.

Sur les aspects industriels, l'exploitation des résultats en est à ses balbutiements : DATAZERO a développé un intergiciel qui a atteint le niveau de maturité technologique 3 ou 4. Des discussions sont en cours avec la SATT Toulouse Tech Transfer et avec le pôle de compétitivité DERBI afin de construire un logiciel plus mature.

Une autre exploitation des résultats est le grand nombre de conférences invitées dans des événements grand public et industriels où le projet DATAZERO a été présenté. En fait, nous avons eu un assez grand nombre d'articles et d'interviews dans des journaux, des magazines et des émissions de radio, ce qui a suscité l'intérêt du grand public pour le sujet de notre projet. Un certain nombre de voix politiques manifestaient de l'intérêt et encourageaient notre travail, atteignant *de facto* un public plus large que le public scientifique classique. On peut convenir que sur ce point, DATAZERO a contribué à l'intérêt mondial pour les émissions de CO_2 en cette période de réchauffement climatique.

Conclusions

Le projet DATAZERO est un succès. Au cours des quatre années du projet, les scientifiques de l'informatique et du génie électrique ont travaillé ensemble pour atteindre l'objectif d'évaluer la possibilité de construire et d'exploiter des centres de données fonctionnant uniquement avec des énergies renouvelables. Nous avons en fait prouvé que d'un point de vue technique et scientifique, c'est faisable !

Un grand nombre de publications scientifiques ont été produites dans le cadre du projet et les activités ont attiré plusieurs nouveaux venus sur le sujet chez les partenaires respectifs. À noter également qu'une HDR et cinq doctorats ont été menés en lien avec le sujet de DATAZERO (deux ont été financés par le projet, d'autres par voie interne), ainsi que dix mémoires de master et stages, démontrant également l'implication des partenaires DATAZERO dans le développement de la carrière des jeunes générations sur le thème important de la durabilité des centres de données.

Une suite de DATAZERO, nommée DATAZERO2, a été sélectionnée par l'Agence nationale de la recherche pour démarrer en 2020 et couvrira des perspectives identifiées dans DATAZERO, comme la robustesse des algorithmes d'optimisation et les liens entre l'optimisation à long terme et celle à court terme. Avec le même partenariat, il montre l'engagement des partenaires envers le concept de centres de données zéro émission ainsi que le soutien de l'Agence nationale de la recherche.

Remerciements

Ces travaux ont été en soutenus par l'Agence nationale de la recherche dans le cadre du projet DATAZERO (ANR-15-CE25-0012).

Références

- [1] Clicking clean : Who is winning the race to build a green internet ? Greenpeace, Amsterdam, Netherlands, Jan. 2017.
- [2] G. Cook. "how clean is your cloud?". <http://www.greenpeace.nl/Global/nederland/2012/publicaties/HowCleanisYourCloud.pdf>", April 2012.
- [3] G. Cook, T. Dowdall, D. Pomerantz, and Y. Wang. "clicking clean : How companies are creating the green internet: <http://media.oregonlive.com/silicon-forest/other/Clicking%20Clean%20Report.pdf>", April 2014.
- [4] J. M. Pierson, G. Baudic, S. Caux, B. Celik, G. Da Costa, L. Grange, M. Haddad, J. Lecuivre, J. M. Nicod, L. Philippe, V. Rehn-Sonigo, R. Roche, G. Rostirolla, A. Sayah, P. Stolf, M. Thi, and C. Varnier. Datazero : Datacenter with zero emission and robust management using renewable energy. *IEEE Access*, 7 :103209–103230, 2019.