



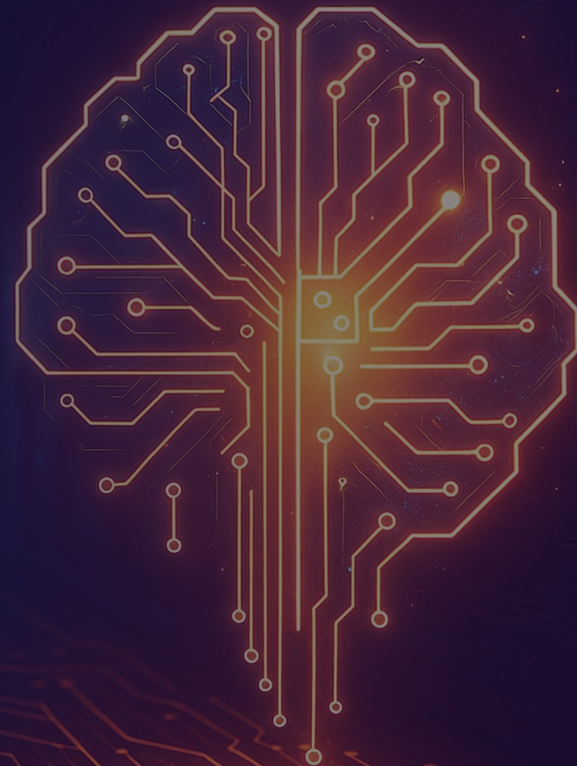
Datacenters IT & IA

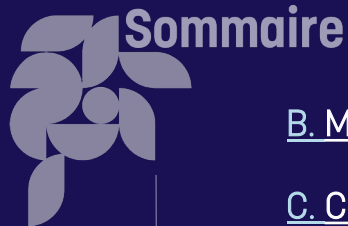
Restitution intermédiaire

27/11/2025



Une étude menée par MDC,
en partenariat avec l'ADEME,
soutenue par EDF





1 - Rappel de l'étude 2024 et présentation de l'étude 2025

- Méthodologie

2- Projections

3- Axes d'amélioration de la performance

A. Infrastructure technique et datacenters

- DLC (Direct to chip)
- Freecooling
- Immersion
- Régulation cascade groupes froid
- Monitoring

B. Monitoring serveurs

C. Composants serveurs – IT Hardware

- Analyse CPU
- Analyse GPU
- Conclusion intermédiaire : CPU et GPU

D. Efficacité énergétique des serveurs – IT infrastructure

4 - Conclusions générales et étapes suivantes de l'étude

5 - Glossaire



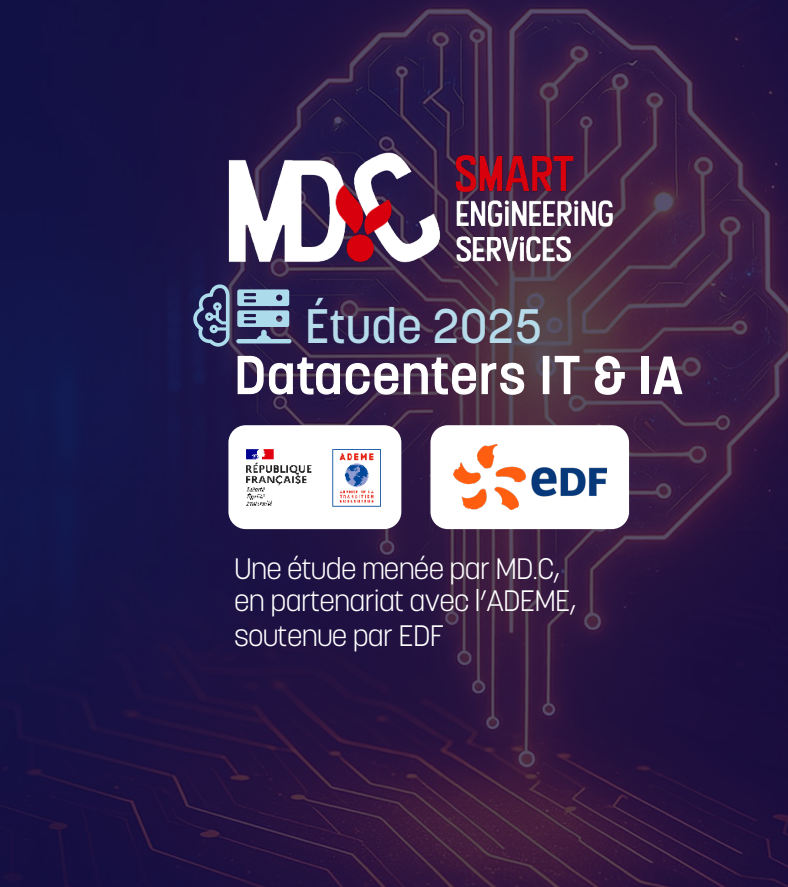
Une étude menée par MDC, en partenariat avec l'ADEME, soutenue par EDF

1

Rappel de l'étude 2024 et présentation de l'étude 2025



[Retour au sommaire](#)



MDC **SMART
ENGINEERING
SERVICES**

 **Étude 2025**
Datcenters IT & IA

 **RÉPUBLIQUE
FRANÇAISE**
Exercice
2024-2025

 **ADEME**
Agence de l'Environnement
et de la Transition

 **edf**

Une étude menée par MDC,
en partenariat avec l'ADEME,
soutenue par EDF

● **FOCUS 2024** : Réalisation d'études « Efficacité énergétique dans les datacenters »

MD.C a été missionné en 2023 par l'ADEME et l'ATEE pour la réalisation d'une étude globale sur l'efficacité énergétique afin de contribuer à établir un **état des lieux précis de la consommation énergétique des datacenters** et in fine **renforcer le dispositif des CEE**, avec une **évolution/création de fiches**, donc des **primes liées** à l'efficacité énergétique.

Cette étude a été réalisée notamment via des entretiens avec les acteurs pertinents et porte :

- Sur le **cadre réglementaire** du dispositif des CEE et associés
- Sur une **étude technique du parc existant** des Datacenters
- Sur les **technologies et solutions innovantes** de la filière afin d'accompagner financièrement leur mise en place via des **dispositifs de primes**

Pour télécharger
l'extrait de l'étude
cliquez ici



**L'EFFICACITÉ
ÉNERGÉTIQUE
DANS LES
DATACENTERS**

MDC
SMART ENGINEERING SERVICES

Restitution de l'étude
Mardi 15 octobre 2024
Espace Hamelin
17 Rue de l'Amiral Hamelin, 75116 Paris
de 9h30 à 12h30 (accueil et café - Pot & échanges)

Après le succès de l'étude 2024 sur l'efficacité énergétique des datacenters,

MD.C lance une nouvelle phase d'analyse

Datacenters IT & IA

Anticiper l'impact énergétique à horizon 2035

Un projet
soutenu par
l'ADEME

Un enjeu stratégique
pour la filière
numérique en France

● Pourquoi cette étude est essentielle ?

1

Projets IA à horizon 2035 : cartographier et anticiper

Identifier les projets datacenter liés à l'IA, analyser leur calendrier, et évaluer les besoins techniques associés : foncier, puissance électrique, refroidissement avancé (DLC, immersion).

2

Modernisation du parc : repenser les équipements critiques

Analyser les besoins de renouvellement des infrastructures existantes (refroidissement, ASI, groupes électrogènes) pour gagner en performance et en résilience..

3

Sobriété énergétique : mesurer les gains et les opportunités

Actualiser les économies d'énergie mobilisables à l'échelle nationale et identifier les nouvelles solutions technologiques au service de l'efficience.

4

Performance IT : créer les bons indicateurs

Étudier la consommation réelle des serveurs, en particulier IA, et proposer des indicateurs pertinents pour mesurer l'efficacité énergétique de la partie IT.

● Objectifs & Finalité de l'étude

- ▲ Fournir des repères fiables pour piloter la transition énergétique de la filière datacenter
- ▲ Créer des indicateurs de performance IT exploitables pour les serveurs IA et non IA
- ▲ Outiller les opérateurs et les institutions pour prendre de meilleures décisions d'investissement
- ▲ Soutenir un développement énergétiquement soutenable de l'IA en France
- ▲ Proposer des préconisations concrètes pour développer des nouvelles aides publiques et orienter les politiques futures



Étude 2025 Datacenters IT & IA



Une étude menée par MD.C,
en partenariat avec l'ADEME,
soutenue par EDF

Méthodologie



[Retour au sommaire](#)

● Rappel du contexte intermédiaire de l'étude

La présente restitution correspond à un rapport intermédiaire, pensé comme un point d'étape destiné à partager l'état actuel de l'étude et les axes de réflexion en cours de MD.C.

L'objectif est donc de créer un échange durant et à la suite de cette restitution entre MD.C et les acteurs concernés, visant à renforcer les données sources de l'étude et la réflexion idoine afin d'affiner les projections de déploiement et les estimations d'économies d'énergie.

L'ensemble des données et éléments contenus dans ce rapport sont ouverts à la discussion, et surtout à des révisions ultérieures, avec pour objectif une consolidation en vue de la restitution finale.

Calendrier

Mai – Octobre 2025

Collecte et analyse

Novembre 2025

Restitution intermédiaire

Mai 2026

Restitution finale

Aucune projection ou estimation contenue dans cette présentation ne saurait être entendue ou comprise comme constituant une conclusion ou une préconisation définitive de MD.C. L'ensemble des éléments avancés sont sujets à révision éventuelle pour la restitution finale.

● 1. Méthodologie

L'ensemble des données présentées dans le cadre de la présente restitution sont issues des données récupérées dans le cadre de l'étude, ou de sources (benchmark, études scientifiques et techniques) listées dans le présent rapport intermédiaire.

Les opérateurs audités relèvent quatre catégories :

- Opérateurs de datacenters (infrastructure)
- Acteurs IT (fabricants de serveurs, acteurs de l'IA/HPC, spécialistes IT)
- Fournisseurs (équipements d'infrastructure : production de froid ; distribution électrique ; monitoring)
- Installateurs (principaux acteurs de la construction et de l'installation spécialisés dans le datacenter)

Les entretiens de collecte de données réalisés avec les acteurs professionnels sont strictement confidentiels et réalisés selon une méthode standardisée



RÉPUBLIQUE
FRANÇAISE

*Liberté
Égalité
Fraternité*



- **Le pool d'opérateurs ayant accepté participé à l'étude à date du rapport intermédiaire**

Nombre d'opérateurs de datacenters
(opérant au moins un site à date)

22

soit

Nombre de sites existants et
en cours de construction

63

Nombre d'opérateurs IT
(fabricants de serveurs, acteurs
de l'IA/HPC, spécialistes IT) :

6

Nombre de fournisseurs
et installateurs

8

● Méthodologie relative aux projections de consommation

Etant donné :

- La nature intermédiaire de la restitution (et donc l'absence de données complètes permettant de réaliser les projections) ;
- La nature évolutive très rapide des déploiements en France de datacenters et des typologies de datacenters concernés

➔ Les calculs liés aux projections globales et les estimations d'économies d'énergie ont été systématiquement réalisés sur la base d'hypothèses hautes et basses avec pour sources :

- Les précédentes projections réalisées par MDC en 2024
- Les données publiques de RTE sur les raccordements de datacenters
- Les premières données collectées dans le cadre de l'étude 2025

Hypothèse basse d'évolution du parc IT & Datacenter :

données issues de la précédente étude, mises en relation avec les premières données collectées et réactualisées en 2025

Hypothèse haute d'évolution du parc IT & Datacenter :

données issues des demandes de raccordement RTE, mises en relation avec les premières données collectées et réactualisées en 2025

● Méthodologie relative aux projections de déploiement IA

Etant donné :

- La nature intermédiaire de la restitution (et donc l'absence de données complètes permettant de réaliser les projections)
- Les évolutions en cours des projets de déploiement IA/haute densité et des technologies déployées dans le cadre de ces évolutions

→ Les calculs liés aux projections de la part de l'IA en France sur la totalité du parc IT ont été réalisées en prenant en compte une hypothèse basse et une hypothèse haute, basées sur :

- Des études internationales sur le déploiement de l'IA et la répartition des usages, extrapolés à la France
- Les premières données collectées dans le cadre de l'étude 2025

Hypothèse basse et haute d'évolution du parc IT dédié à l'IA :

données issues d'études internationales, mises en relation avec les premières données collectées en 2025

Hypothèse de répartition des usages de l'IA (entraînement, inférence, etc...) :

données issues de rapports internationaux et de benchmarks, mises en relation avec les premières données collectées en 2025

● 2. Méthodologie (sources)

Les sources pour les données chiffrées relatives à l'évolution du parc français et aux performances des technologies citées sont donc exclusivement issues :

Des données fournies par les acteurs ayant participé à l'étude

Des données issues de rapports publics, de données open source, et de rapports/études techniques et scientifiques

Principales sources hors collecte directe des données (entretiens) :

- Spec
- Epoch AI
- 4E TCP (étude)
- ML Communs
- Chip data Set
- Harvard University (étude)
- Fsas technology
- Stratégie nationale bas-carbone (SNBC)
- RTE
- Statistique publique de l'énergie, des transports, du logement et de l'environnement
- Goldman Sachs Research Forecast (2025)
- Groupe d'étude géopolitique (2025)
- ...

2

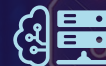
Projections



[Retour au sommaire](#)



SMART
ENGINEERING
SERVICES



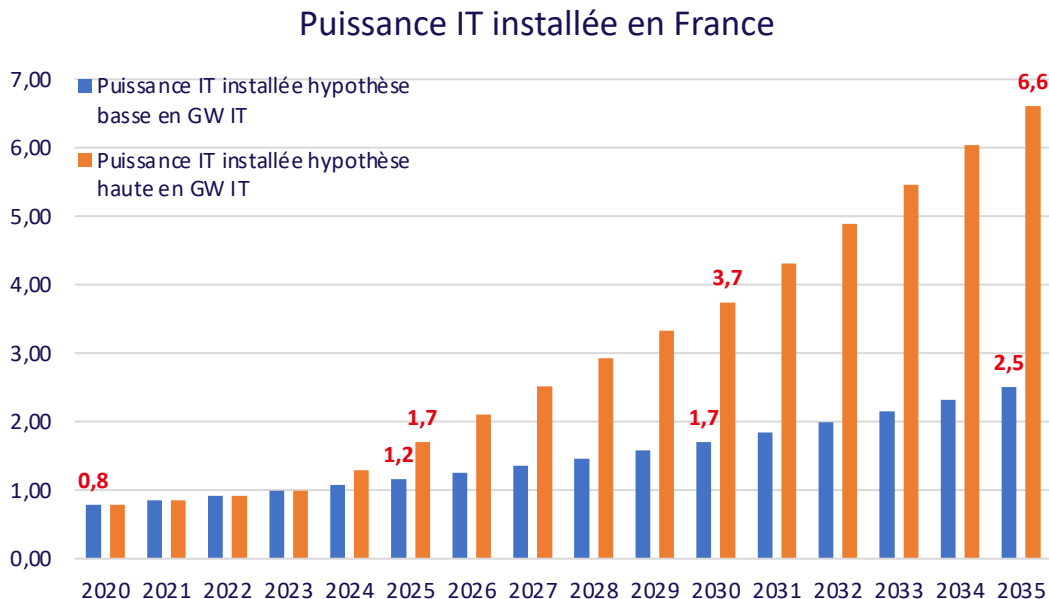
Étude 2025

Datcenters IT & IA



Une étude menée par MDC,
en partenariat avec l'ADEME,
soutenue par EDF

● Evolution de la puissance IT installée de la filière Datacenter



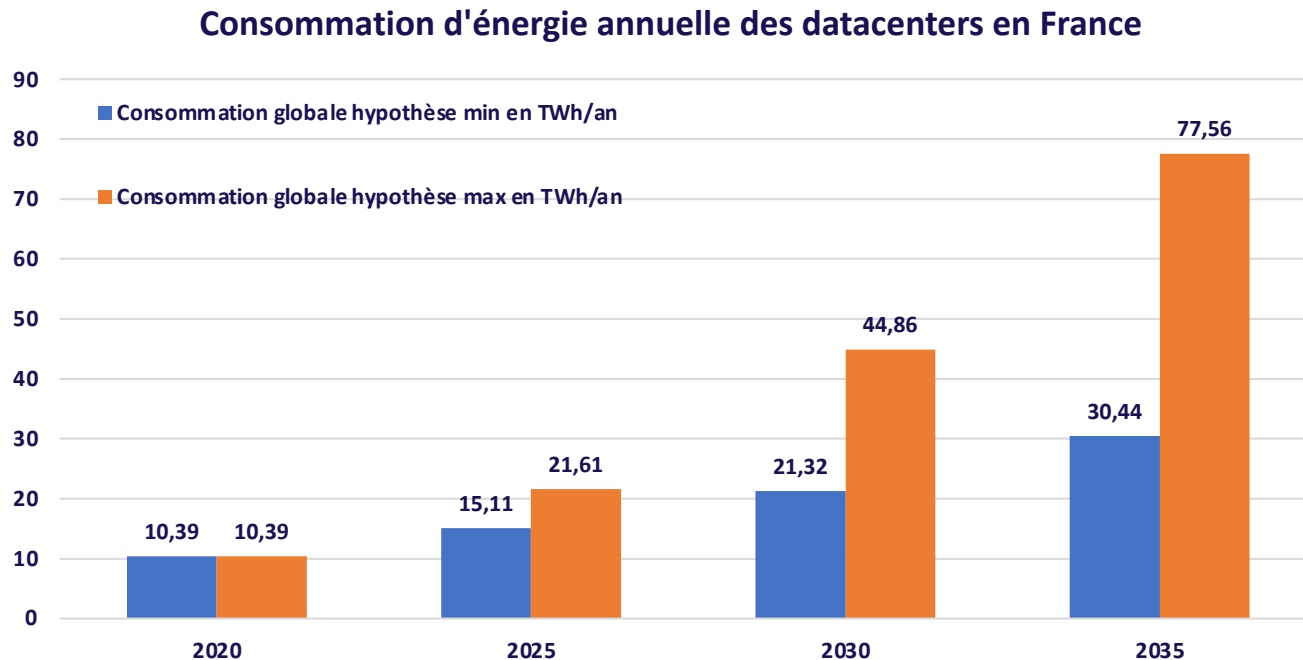
La puissance IT installée devrait être :

De **790 MWIT** en 2020

Comprise entre
1 710 et 3 740 MW IT
en 2030

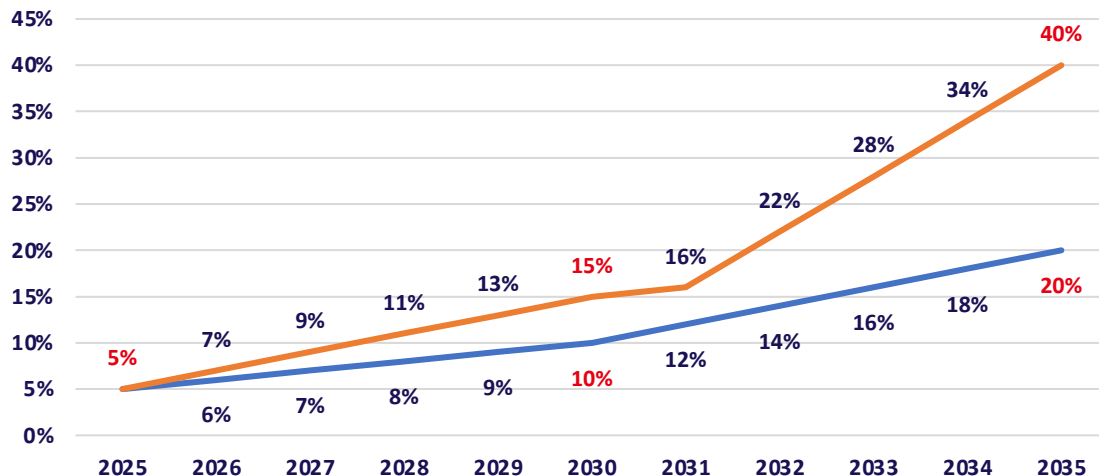
Comprise entre
2 510 et 6 600 MW IT
en 2035

● Evolution de la consommation totale dédiée aux datacenters en France (TWh/an)



● Evolution de la part de l'IA dans le parc IT/Datacenter français

Evolution de la part de l'IA dans le parc datacenter



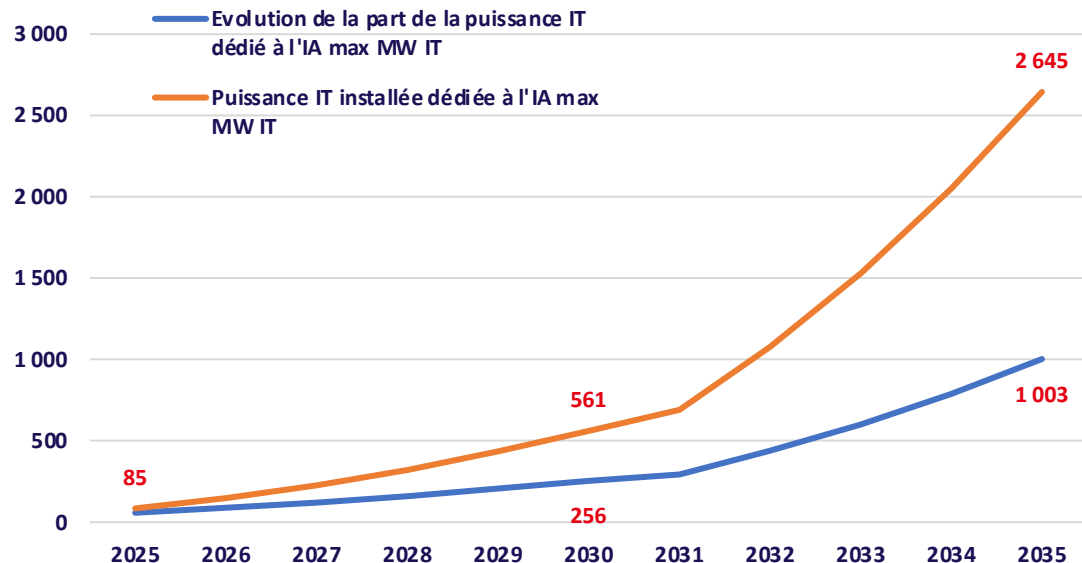
D'après les données collectées en entretien, corrélées à plusieurs projections internationales extrapolées à la France, la part de la puissance IT dédiée à l'IA dans les datacenters pourrait être comprise entre :

10 et 15 % en 2030



20 et 40 % en 2035

● Evolution de la part de l'IA dans le parc IT/Datacenter français



En suivant les hypothèses citées précédemment (évolution de la puissance IT installée en France et évolution de la part de l'IA dans les datacenters), la puissance IT dédiée à l'IA en France serait comprise entre :

256 et 561 MW IT en 2030



1 et 2,6 GW IT en 2035

3

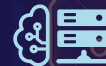
Axes d'amélioration de la performance



[Retour au sommaire](#)

MDC

**SMART
ENGINEERING
SERVICES**



Étude 2025

Datcenters IT & IA



Une étude menée par MDC,
en partenariat avec l'ADEME,
soutenue par EDF



 Étude 2025
Datacenters IT & IA



Une étude menée par MD.C,
en partenariat avec l'ADEME,
soutenue par EDF

3.A

Infrastructure technique et datacenters



[Retour au sommaire](#)

● Evolution des projections relatives aux économies d'énergie (équipements techniques : distribution et production du froid, monitoring infra, distribution électrique)

Lors de l'étude 2024 'Efficacité énergétique dans les datacenters', réalisée avec le concours de l'ATEE, de l'ADEME et de France Datacenter : 12 technologies avaient été relevées comme pertinentes pour la réalisation d'économies d'énergie pour la partie infrastructure technique des datacenters, avec des préconisations opérationnelles de création de fiches CEE,

Plusieurs évolutions sont à retenir :

- **Avancées significatives sur la création de certaines fiches** dans le cadre des groupes de travail pilotés par l'ATEE et en bonne intelligence avec l'ADEME
- **Réactualisation dans le cadre de l'étude 2025 des gisements d'économies d'énergie** associés aux technologies.

Focus sur cinq technologies - fiches potentielles :

DLC (Direct-to-chip)

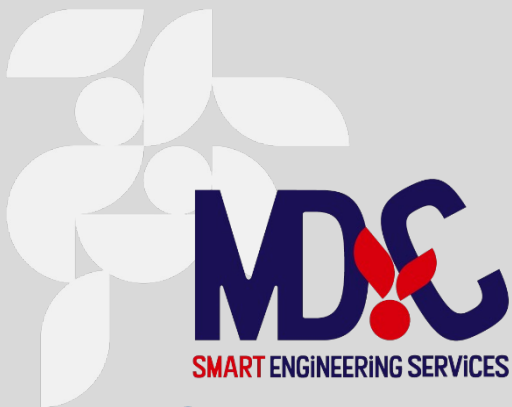
Freecooling -
Freechilling

Immersion

Monitoring

Régulation de la
cascade des chillers





Étude 2025

Datacenters IT & IA



Une étude menée par MD.C,
en partenariat avec l'ADEME,
soutenue par EDF

DLC (Direct to chip)

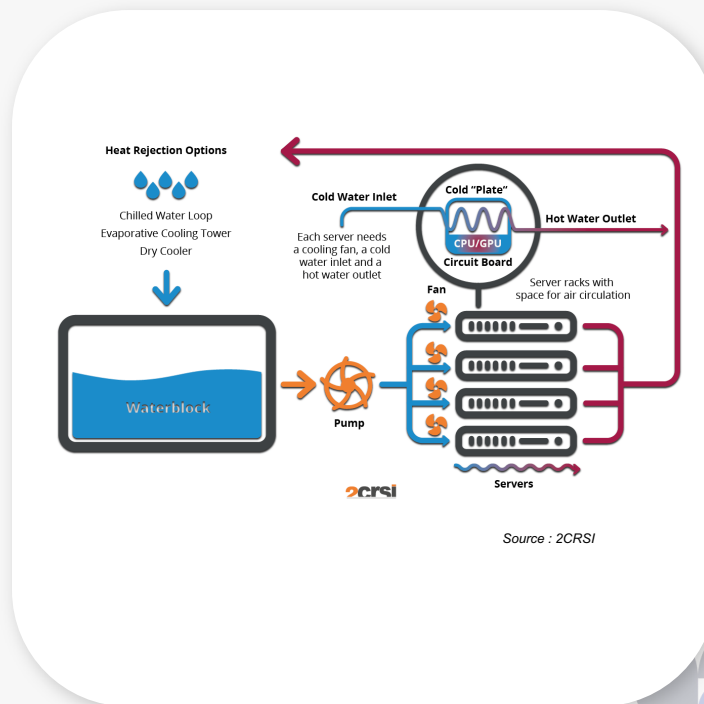


[Retour au sommaire](#)

● Présentation de la technologie

Le Direct Liquid Cooling « Direct to chip » revient à concentrer le refroidissement uniquement des parties les plus sensibles des serveurs (GPU et/ou processeur) au travers d'échangeurs placés directement sur ces derniers. En se rapprochant des sources chaudes et en réduisant le volume de refroidissement, les boucles d'eau glacée peuvent monter en température et donc **augmenter les performances** du système de refroidissement.

PUE Moyen de **1,15**





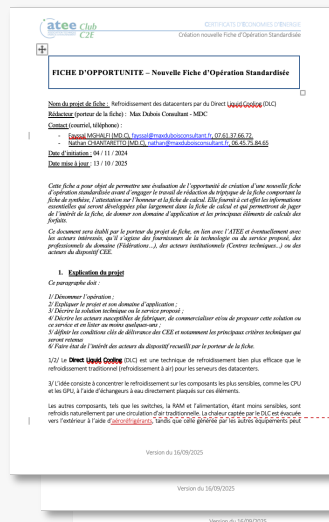
● Point d'étape sur la technologie et réactualisation des projections

Dans le cadre d'une étude spécifique liée au projet de création de fiches et de l'étude 2025 :

- Les estimations de **gisements d'économies d'énergie** ont été réactualisés
- Les **projections** liées à la pénétration du parc par la technologie ont été **revues à la hausse**
- Une **fiche d'opportunité** (étape préalable à la création d'une fiche CEE) a été créée et validée par l'ATEE et l'ADEME

Next steps (prévisionnels) :

- Lancement du GT officiel : **Q1 2026**
- Complétion de la fiche par le GT pour soumission à la DGEC : **Q2 2026**
- Publication de la fiche CEE standardisée



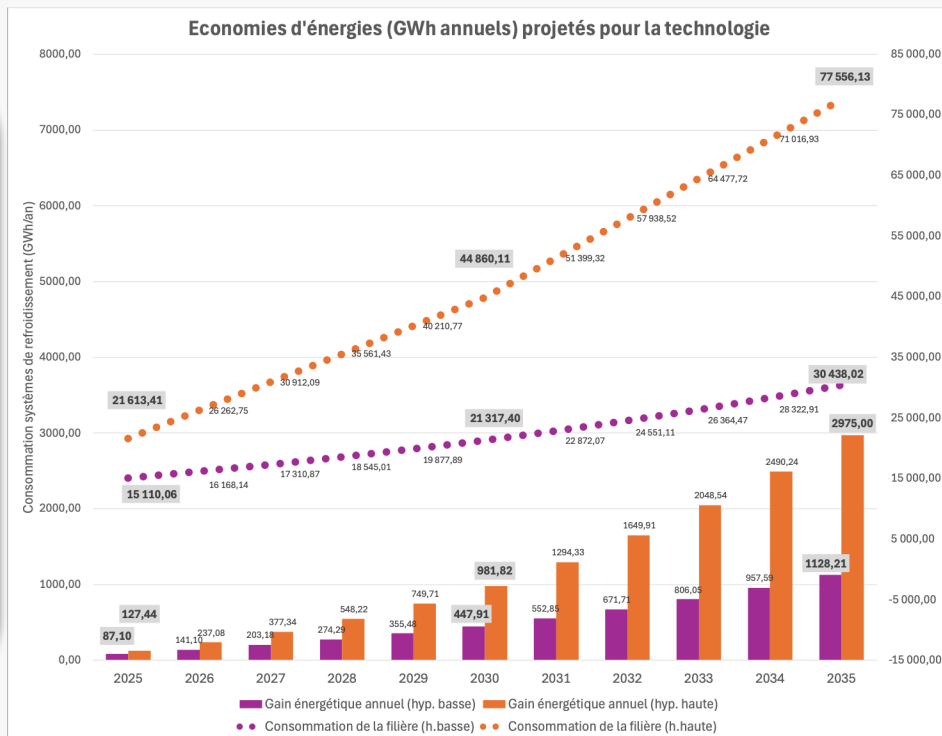
Gain énergétique moyen	1499,86
Puissance IT (MW IT)	1,00
Durée de vie (an)	14
Coefficient d'actualisation (MWh cumac/MWh)	10,563
Gisement estimé (MWh cumac)	15 843,21
Prime potentielle (€)	110 902,50 €

Attention : Les calculs et estimations réalisés par MDC à date sont des estimations basées sur les chiffres fournis par les acteurs dans le cadre de l'étude. Ces estimations peuvent être modifiées par la prise en compte de données complémentaires. La création de fiches CEE nécessite leur validation par l'ATEE, l'ADEME et la DGEC. La présente étude ne constitue en aucun cas une garantie de délivrance de prime ou de création de fiche pour les technologies citées.

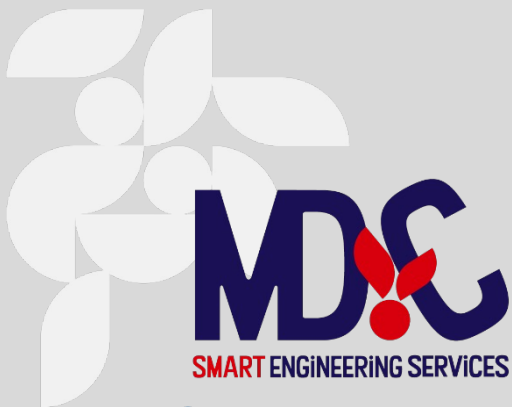
● Economies d'énergie via le DLC (Direct to chip) et projections

En prenant en compte le renforcement clair de l'utilisation prévisionnelle du DLC dans le parc datacenter et les nouvelles prévisions de croissance IT :

- **30 % de la puissance IT** installée totale équipée d'ici 2035
- Un gisement compris entre **11,9 et 31 TWh cumac** d'ici 2035
- Des économies d'énergie annuelles comprises entre **1,1 et 2,9 TWh annuels** en 2035
- Environ **80 et 200 Millions d'€ de primes CEE** potentielles d'ici 2035
- Entre **0,75 et 2 GW de puissance IT** équipée d'ici 2035



Attention : Les calculs et estimations réalisés par MD.C à date sont des estimations basées sur les chiffres fournis par les acteurs dans le cadre de l'étude. Ces estimations peuvent être modifiées par la prise en compte de données complémentaires. La création de fiches CEE nécessite leur validation par l'ATEE, l'ADEME et la DGEC. La présente étude ne constitue en aucun cas une garantie de délivrance de prime ou de création de fiche pour les technologies citées.



Étude 2025

Datcenters IT & IA



Une étude menée par MD.C,
en partenariat avec l'ADEME,
soutenue par EDF

Free cooling



[Retour au sommaire](#)

● Présentation de la technologie

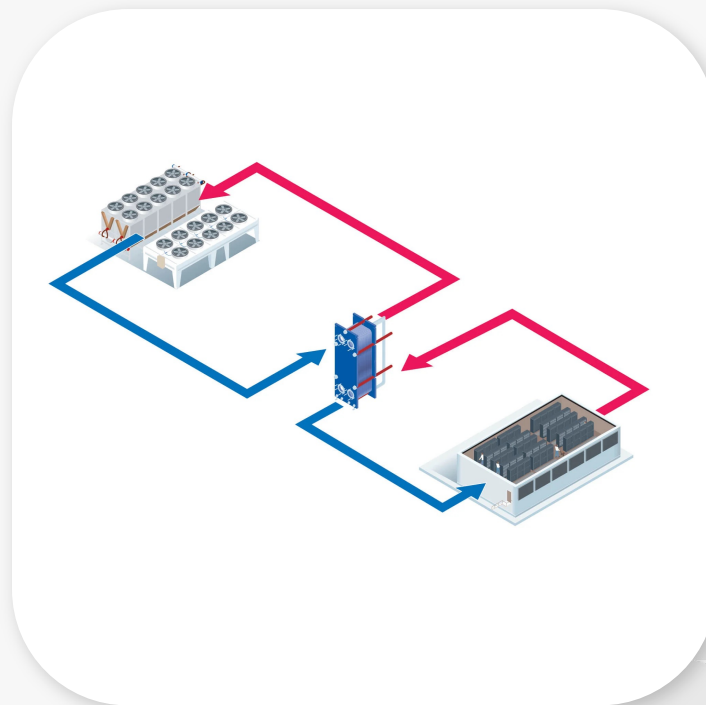
Le **Freecooling efficient** permet de diminuer la part de compression mécanique dans la production de froid dans un Datacenter, et donc, **augmenter les performances** du système de refroidissement.

Plusieurs types de freecooling existent (air/eau mixte ; air direct ; air/air indirect).

Le principal type de freecooling utilisé est le **Freecooling air/eau mixte**

PUE avec Freecooling efficient entre

1,25 et **1,40**





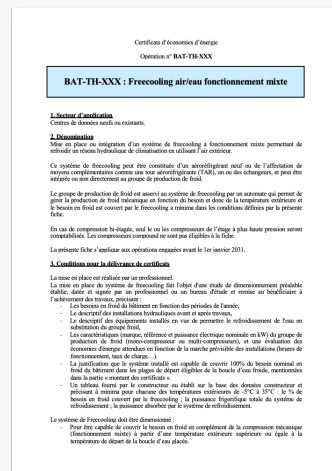
● Point d'étape sur la technologie et réactualisation des projections

Dans le cadre d'une étude spécifique liée au projet de créations de fiche et de l'étude 2025 :

- Les estimations de **gisements d'économies d'énergie** ont été réactualisés en vue de la création d'une fiche Freecooling air/eau mixte
- Une **fiche d'opportunité** finalisée et validée est en cours de publication par la DGEC pour le **Freecooling air/eau mixte**
- Une fiche d'opportunité est en cours de travail en GT pour le **Freecooling air direct**

Next steps (prévisionnels) :

- Publication de la fiche CEE standardisée FC air/eau mixte par la DGEC : Q1/Q2 2026



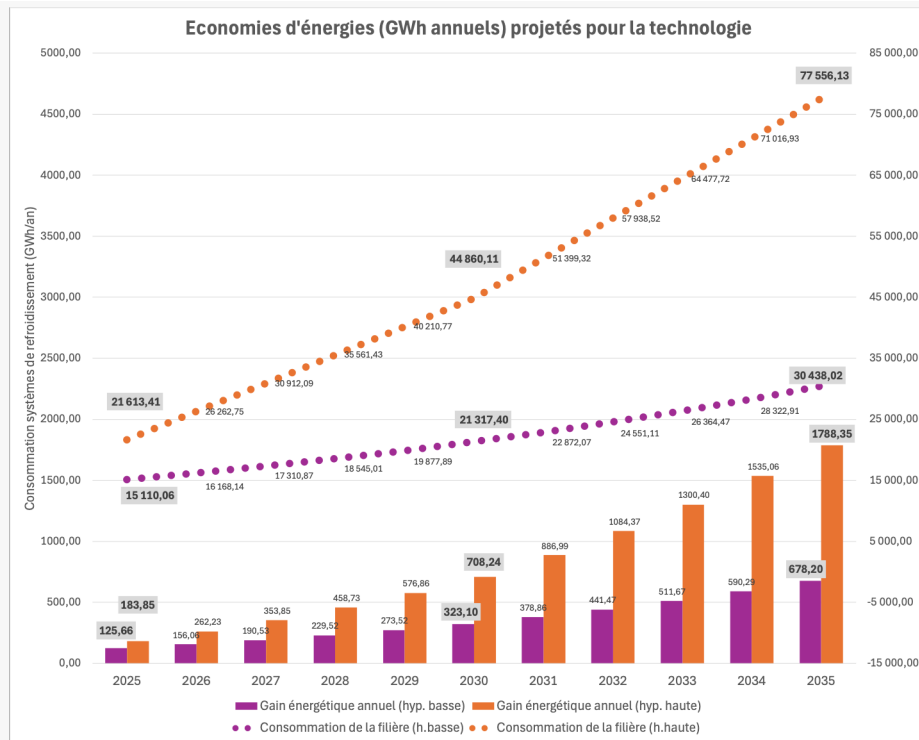
Gain énergétique moyen	540,97
Puissance IT (MW IT)	1,00
Durée de vie (an)	14
Coefficient d'actualisation (MWh cumac/MWh)	10,563
Gisement estimé (MWh cumac)	5 714,28
Prime potentielle (€)	39 999,96 €

Attention : Les calculs et estimations réalisés par MDC à date sont des estimations basées sur les chiffres fournis par les acteurs dans le cadre de l'étude. Ces estimations peuvent être modifiées par la prise en compte de données complémentaires. La création de fiches CEE nécessite leur validation par l'ATEE, l'ADEME et la DGEC. La présente étude ne constitue en aucun cas une garantie de délivrance de prime ou de création de fiche pour les technologies citées.

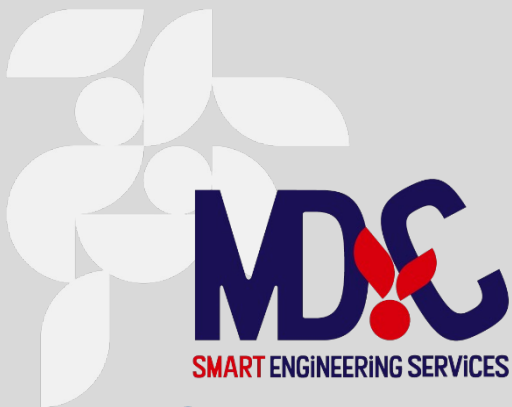
● Economies d'énergie via la technologie et projections

En prenant en compte les estimations réalisées en GT ATEE sur le potentiel de la fiche et les nouvelles prévisions de croissance IT :

- **50 % de la puissance IT** installée totale équipée d'ici 2035
- Un gisement compris entre **7,2 et 18,9 TWh cumac d'ici 2035**
- Des économies d'énergie annuelles comprises entre **0,68 et 1,79 TWh annuels** en 2035
- Environ **50 et 132 Millions d'€ de primes CEE** potentielles d'ici 2035
- Entre **1,25 et 3,3 GW de puissance IT** équipée d'ici 2035



Attention : Les calculs et estimations réalisés par MDC à date sont des estimations basées sur les chiffres fournis par les acteurs dans le cadre de l'étude. Ces estimations peuvent être modifiées par la prise en compte de données complémentaires. La création de fiches CEE nécessite leur validation par l'ATEE, l'ADEME et la DGEC. La présente étude ne constitue en aucun cas une garantie de délivrance de prime ou de création de fiche pour les technologies citées.



Étude 2025

Datcenters IT & IA



Une étude menée par MD.C,
en partenariat avec l'ADEME,
soutenue par EDF

Immersion



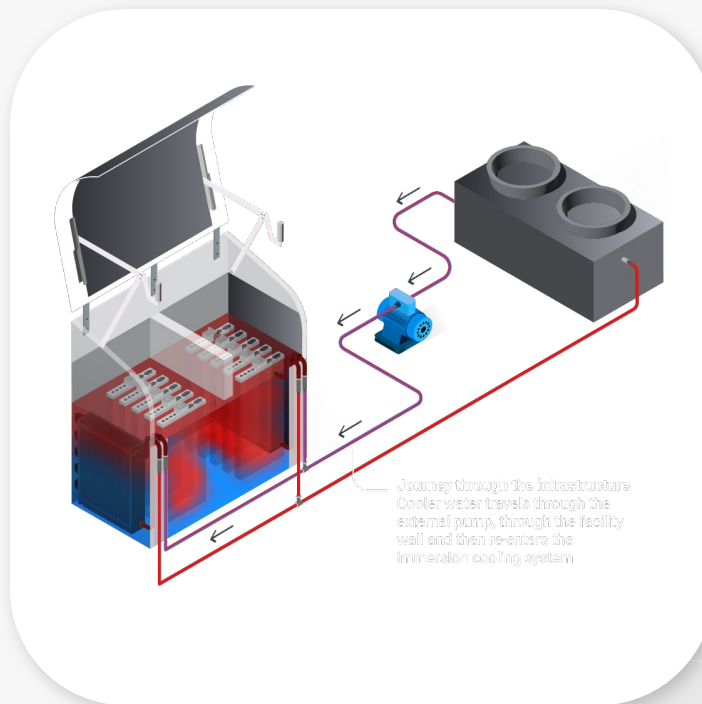
[Retour au sommaire](#)

● Présentation de la technologie

L'**immersion cooling** remplace l'air, utilisé traditionnellement comme **medium de refroidissement**, par un **fluide diélectrique** dans lequel les serveurs sont directement plongés. Ce fluide diélectrique est maintenu en température à travers des échangeurs et une boucle d'eau glacée. Les **températures des boucles d'eau sont plus élevées** car le **volume à maintenir en température est plus faible** et la **capacité calorifique** du fluide est beaucoup plus importante.

PUE avec Immersion entre

1,10 et **1,15**





● Point d'étape sur la technologie et réactualisation des projections

Dans le cadre d'une étude spécifique liée au projet de création de fiches et de l'étude 2025 :

- Les estimations de **gisements d'économies** d'énergie ont été réactualisés
- Les **projections** liées à la pénétration du parc par la technologie ont été **revues**
- Une **fiche d'opportunité** (étape préalable à la création d'une fiche CEE) a été **créée et validée** par **l'ATEE et l'ADEME** pour organisation du GT

Next steps (prévisionnels) :

- Lancement du GT officiel : **Q1/Q2 2026**
- Complétion de la fiche par le GT pour soumission à la DGEC : **Q3/Q4 2026**
- Publication de la fiche CEE standardisée

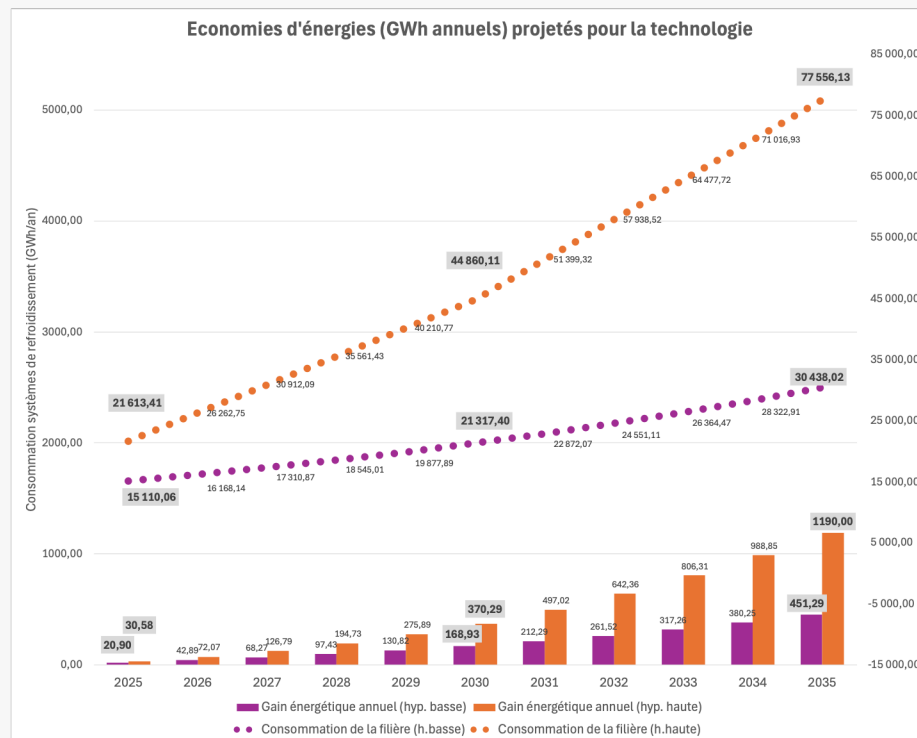
Gain énergétique moyen	1799,83
Puissance IT (MW IT)	1,00
Durée de vie (an)	14
Coefficient d'actualisation (MWh cumac/MWh)	10,563
Gisement estimé (MWh cumac)	19 011,86
Prime potentielle (€)	133 083,00 €

Attention : Les calculs et estimations réalisés par MD.C à date sont des estimations basées sur les chiffres fournis par les acteurs dans le cadre de l'étude. Ces estimations peuvent être modifiées par la prise en compte de données complémentaires. La création de fiches CEE nécessite leur validation par l'ATEE, l'ADEME et la DGEC. La présente étude ne constitue en aucun une garantie de délivrance de prime ou de création de fiche pour les technologies citées.

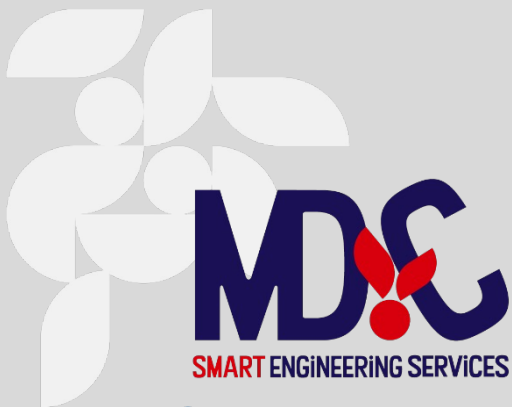
● Economies d'énergie via la technologie et projections

En prenant en compte les estimations réalisées/réactualisées et les nouvelles prévisions de croissance IT :

- **10 % de la puissance IT** installée totale équipée d'ici 2035
- Un gisement compris entre **0,75 et 2,1 TWh cumac d'ici 2035**
- Des économies d'énergie annuelles comprises entre **0,45 et 1,19 TWh annuels** en 2035
- Environ **33 et 88 Millions d'€ de primes CEE** potentielles d'ici 2035
- Entre **0,25 et 0,66 GW de puissance IT** équipée d'ici 2035



Attention : Les calculs et estimations réalisés par M.D.C à date sont des estimations basées sur les chiffres fournis par les acteurs dans le cadre de l'étude. Ces estimations peuvent être modifiées par la prise en compte de données complémentaires. La création de fiches CEE nécessite leur validation par l'ATEE, l'ADEME et la DGEC. La présente étude ne constitue en aucun cas une garantie de délivrance de prime ou de création de fiche pour les technologies citées.



Étude 2025

Datacenters IT & IA



Une étude menée par MD.C,
en partenariat avec l'ADEME,
soutenue par EDF

Régulation cascade Groupes de production de froid



[Retour au sommaire](#)

● Présentation de la technologie

La **régulation cascade** permet d'**optimiser l'usage du free-cooling et de la compression mécanique** en utilisant l'ensemble des condenseurs et surfaces d'échangeurs des groupes froids, y compris les **éléments en redondance**, grâce à l'asservissement de la boucle d'eau glacée.

La régulation de la cascade sur les groupes froids est limitée aux sites disposant d'au moins quatre groupes froids, et l'**optimisation des performances est fortement relative** à la performance de départ de la cascade

Optimisation Freecooling entre

30% et **70%**



[illegible]

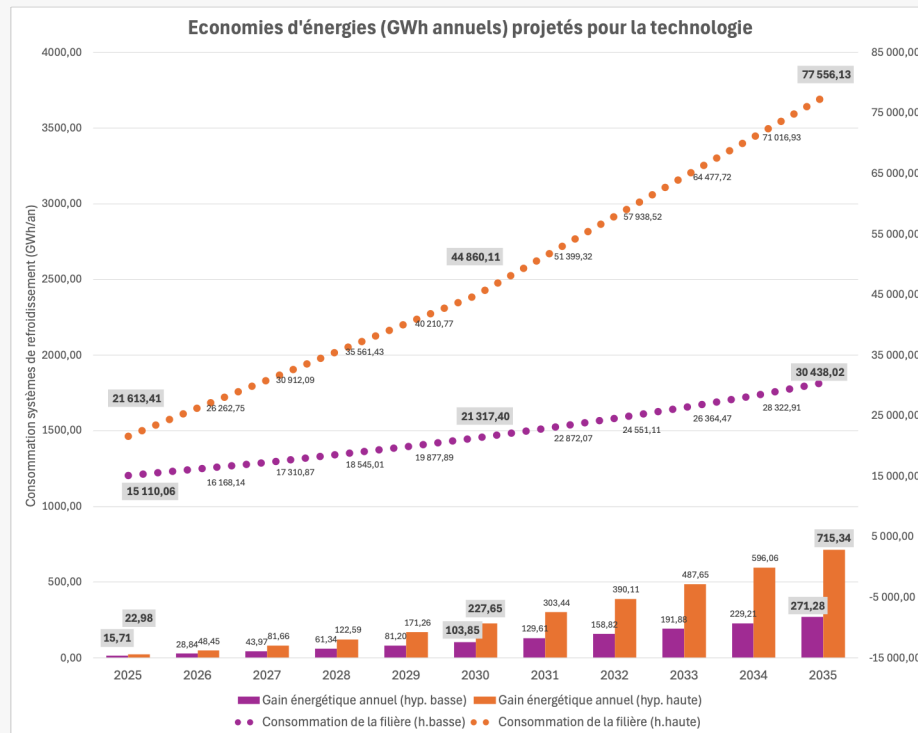
Gain énergétique moyen	270,48
Puissance IT (MW IT)	1,00
Durée de vie (an)	14
Coefficient d'actualisation (MWh cumac/MWh)	10,563
Gisement estimé (MWh cumac)	2 857,1
Prime potentielle (€)	19 999,98

38

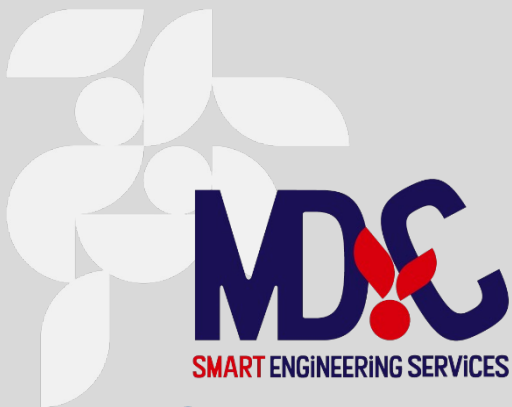
● Economies d'énergie via la technologie et projections

En prenant en compte les estimations réalisées/réactualisées et les nouvelles prévisions de croissance IT :

- **40 % de la puissance IT** installée totale équipée d'ici 2035
- Un gisement compris entre **2,9 et 7,6 TWh cumac** d'ici 2035
- Des économies d'énergie annuelles comprises **entre 0,27 et 0,7 TWh annuels** en 2035
- Environ **20 et 53 Millions d'€ de primes CEE** potentielles d'ici 2035
- Entre **1 et 2,6 GW de puissance IT** équipée d'ici 2035



Attention : Les calculs et estimations réalisés par MD.C à date sont des estimations basées sur les chiffres fournis par les acteurs dans le cadre de l'étude. Ces estimations peuvent être modifiées par la prise en compte de données complémentaires. La création de fiches CEE nécessite leur validation par l'ATEE, l'ADEME et la DGE. La présente étude ne constitue en aucun cas une garantie de délivrance de prime ou de création de fiche pour les technologies citées.



Étude 2025

Datcenters IT & IA



Une étude menée par MD.C,
en partenariat avec l'ADEME,
soutenue par EDF

Monitoring



[Retour au sommaire](#)

● Présentation de la technologie

Le **monitoring** permet de prendre du recul sur la **performance du bâtiment et des équipements** afin d'identifier le plus rapidement possible les **potentielles dérives de consommation** et d'agir au plus vite et de façon ciblée, avec certaines solutions permettant une **gestion automatisée des consommations**.

Gain énergétique entre

5% et 10%





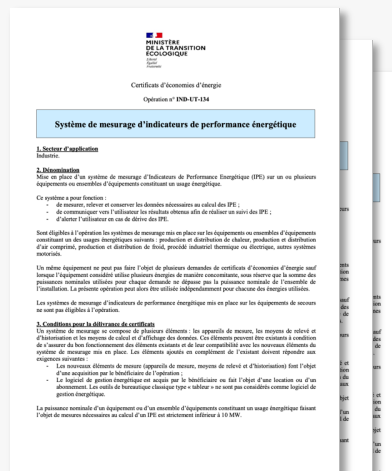
● Point d'étape sur la technologie et réactualisation des projections

Dans le cadre d'une étude spécifique liée au projet de création de fiches et de l'étude 2025, ainsi que de projet d'élargissement de la fiche Industrie IND UT 134 :

- Les estimations de **gisements d'économies d'énergie** ont été réactualisées
- Les projections** liées à la pénétration du parc par la technologie ont été **revues**

Next steps (prévisionnels) :

- Création/réactualisation de la fiche d'opportunité
- Lancement du GT officiel : **à déterminer**
- Complétion de la fiche par le GT pour soumission à la DGEC : **à déterminer**
- Publication de la fiche CEE standardisée



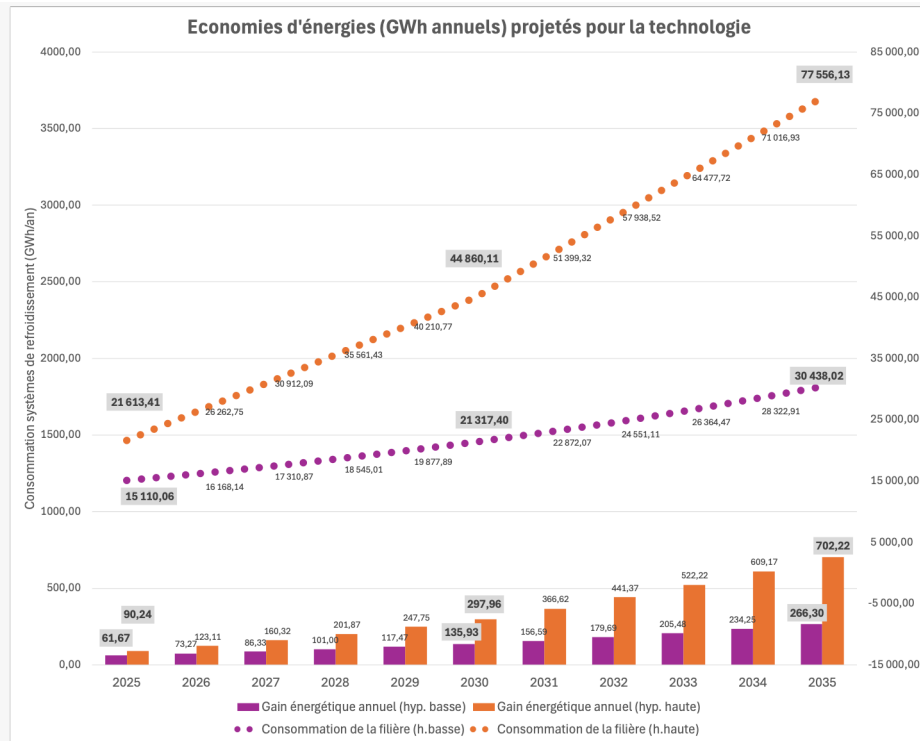
Gain énergétique moyen	132,76
Puissance IT (MW IT)	1,00
Durée de vie (an)	6
Coefficient d'actualisation (MWh cumac/MWh)	5,450
Gisement estimé (MWh cumac)	723,54
Prime potentielle (€)	5 064,79 €

Attention : Les calculs et estimations réalisés par M.D.C à date sont des estimations basées sur les chiffres fournis par les acteurs dans le cadre de l'étude. Ces estimations peuvent être modifiées par la prise en compte de données complémentaires. La création de fiches CEE nécessite leur validation par l'ATEE, l'ADEME et la DGEC. La présente étude ne constitue en aucun cas une garantie de délivrance de prime ou de création de fiche pour les technologies citées.

● Economies d'énergie via la technologie et projections

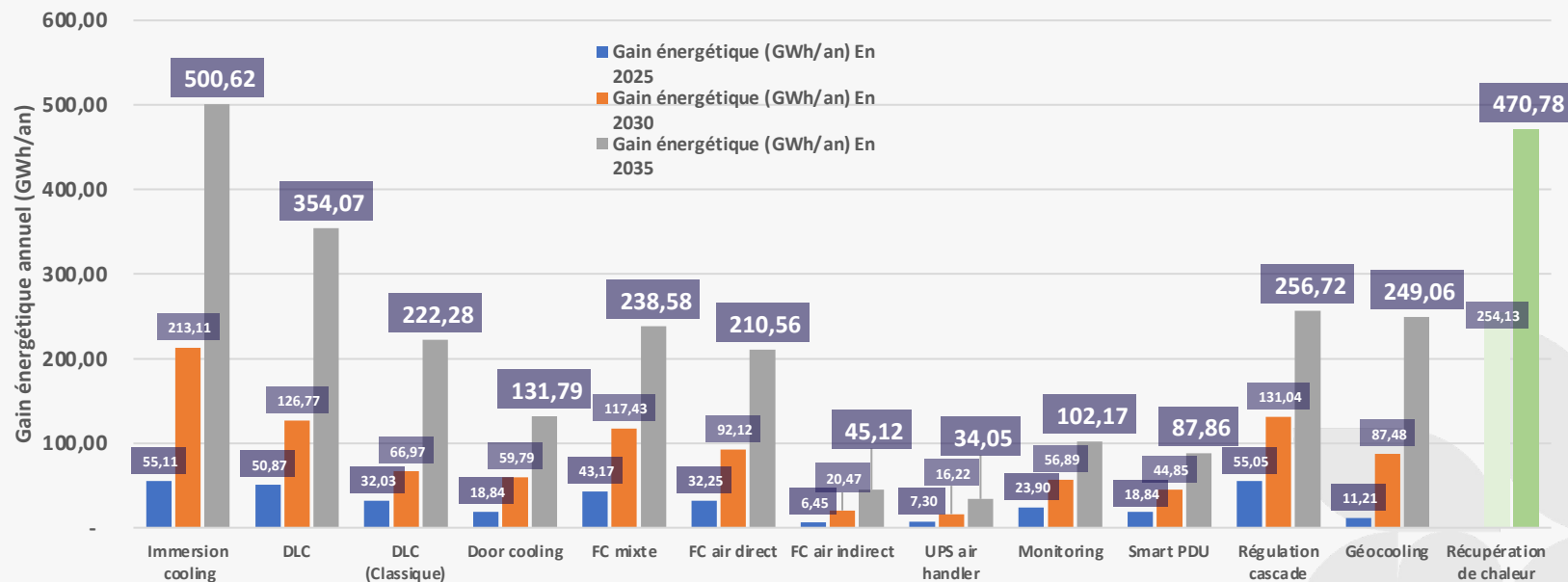
En prenant en compte les estimations réalisées/réactualisées et les nouvelles prévisions de croissance IT :

- **80 % de la puissance IT** installée totale équipée d'ici 2035
- Un gisement compris entre **2,8 et 7,5 TWh cumac** d'ici 2035
- Des économies d'énergie annuelles comprises entre **0,27 et 0,7 TWh annuels** en 2035
- Environ **19 et 53 Millions d'€ de primes CEE** potentielles d'ici 2035
- Entre **2 et 5,3 GW de puissance IT** équipée d'ici 2035



Attention : Les calculs et estimations réalisés par M.D.C à date sont des estimations basées sur les chiffres fournis par les acteurs dans le cadre de l'étude. Ces estimations peuvent être modifiées par la prise en compte de données complémentaires. La création de fiches CEE nécessite leur validation par l'ATEE, l'ADEME et la DGEC. La présente étude ne constitue en aucun cas une garantie de délivrance de prime ou de création de fiche pour les technologies citées.

● Rappel du périmètre global des économies d'énergie détectées en 2024 (étude précédente)*



* Graphique Issue du rapport de l'étude « efficacité énergétique dans les datacenters » - octobre 2024



 Étude 2025
Datacenters IT & IA



Une étude menée par MD.C,
en partenariat avec l'ADEME,
soutenue par EDF

3.B

Monitoring serveurs



[Retour au sommaire](#)

● Présentation de la technologie

Un axe majeur de **réduction des consommations** énergétiques liées à l'IT, détecté, est l'**optimisation de l'utilisation des serveurs** (taux de charge des serveurs ; mode Idle), avec l'emploi de logiciels de **monitoring/automates** permettant de détecter et optimiser la gestion des serveurs en sous-usage et donc une **gestion automatisée des consommations**.

**Consommation estimée des serveurs,
en mode Idle* :**

30% à 40%



● Economies d'énergie via la réduction des consommations des serveurs idle

1

Constat

Une part importante des serveurs installés est très peu, voire pas sollicitée.

Même à faible charge, ils conservent une consommation énergétique non négligeable (mode Idle)

2

Objectifs

Identifier les serveurs pas ou sous-utilisés le plus rapidement possible.

Optimiser l'usage des infrastructures en répondant aux besoins

3

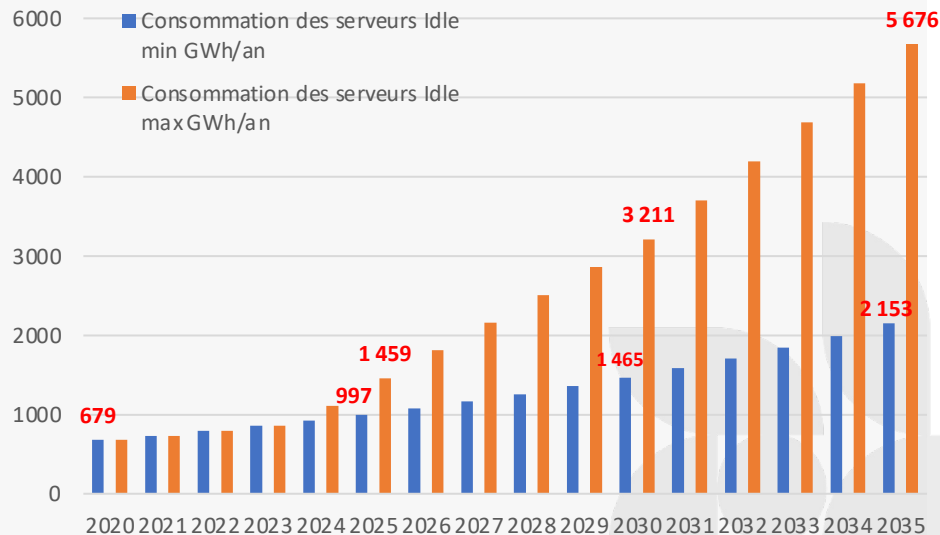
Démarche

Collaboration avec les parties prenantes (hébergeurs, datacenters privés, fournisseurs de solutions) pour analyser le parc et proposer des pistes d'amélioration

● Evolution de la consommation des serveurs Idle dans le parc français

- Les données intermédiaires remontées des fournisseurs et des opérateurs de sites indiquent une fourchette **de 10 à 35 %** des serveurs exploités en France fonctionnant en **mode Idle** pendant **au moins la moitié de l'année**, représentant une consommation électrique annuelle entre **1,5 et 3,2 TWh/an en 2030**.
- **Un système de management de l'énergie** poussé permettrait d'analyser et de mieux comprendre les périodes de travail de ces serveurs et donc de **minimiser leur consommation**.
- Cette analyse peut **varier et devra être revue** pour les serveurs **haute densité, dont l'activité est généralement plus ponctuelle**, à l'exception des usages liés à l'inférence IA.

Estimation de la consommation nationale des serveurs Idle



● Evolution du gain énergétique pour la mise en place de cette technologie

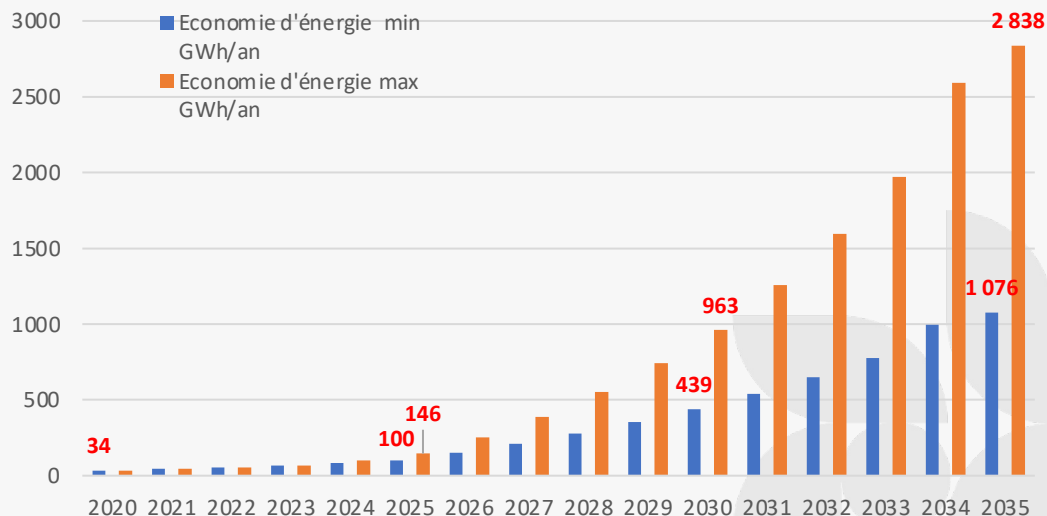
Via un taux de pénétration de la technologie allant de 10% en 2025 jusqu'à 50% en 2035, nous pouvons attendre un gain énergétique :

**Compris entre
439 et 963 GWh/an
en 2030**



**Compris entre
1 076 et 2 838 GWh/an
en 2035**

Estimation des économies d'énergie potentielles via la réduction des serveurs Idle





● Possibilité de création d'une Fiche d'opportunité et d'un GT CEE

Dans le cadre de l'étude en cours, suffisamment d'éléments ont été relevés à date pour la création d'une Fiche d'opportunité et (sous réserve d'un accord ATEE et ADEME) d'un groupe de travail :

- Des estimations de **gisements d'économies d'énergie** unitaires et nationales ont été réalisées
- Des premières données de **projection du potentiel national de la technologie** sont disponibles.

Next steps (prévisionnels) :

- Création d'une fiche d'opportunité : **2026**
- Lancement du GT officiel : **sous réserve validation ATEE ADEME**
- Complétion de la fiche par le GT pour soumission à la DGEC :
- Publication de la fiche CEE standardisée

MINISTÈRE
DE LA TRANSITION
ÉCOLOGIQUE

Certificats d'économies d'énergie

*Création potentielle
d'une Fopp CEE
dédiée sous réserve
de manifestation
d'intérêt de la filière*

Gain énergétique moyen	516,84
Puissance IT (MW IT)	1,00
Durée de vie (an)	5
Coefficient d'actualisation (MWh cumac/MWh)	4,630
Gisement estimé (MWh cumac)	2 392,92
Prime potentielle (€)	16 750,41 €

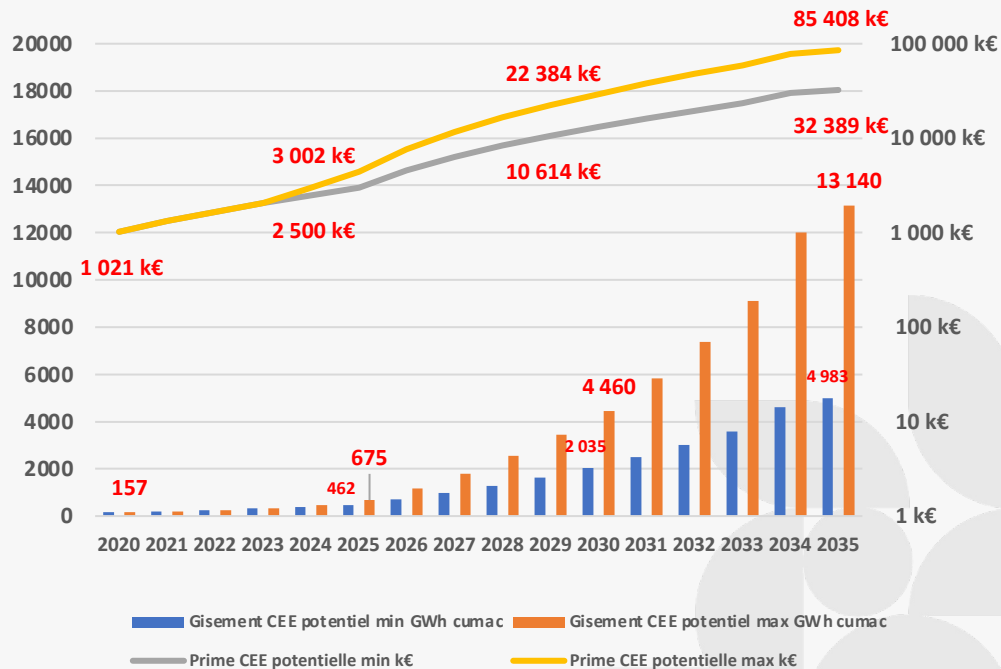
Attention : Les calculs et estimations réalisés par MD.C à date sont des estimations basées sur les chiffres fournis par les acteurs dans le cadre de l'étude. Ces estimations peuvent être modifiées par la prise en compte de données complémentaires. La création de fiches CEE nécessite leur validation par l'ATEE, l'ADEME et la DGEC. La présente étude ne constitue en aucun cas une garantie de délivrance de prime ou de création de fiche pour les technologies citées.

● Economies d'énergie via la technologie et projections

En prenant en compte les estimations réalisées/réactualisées et les nouvelles prévisions de croissance IT :

- **50 % de la puissance IT** installée totale équipée d'ici 2035 (hypothèse)
- Un gisement compris entre **5 et 13 TWh cumac d'ici 2035**
- Des économies d'énergie annuelles comprises entre **1 et 2,8 TWh annuels** en 2035
- Entre **32 et 85 Millions d'€ de primes CEE** potentielles d'ici 2035
- Entre **1,25 et 3,3 GW de puissance IT** équipée d'ici 2035

Gisements et primes CEE potentiels





 Étude 2025
Datacenters IT & IA



Une étude menée par MD.C,
en partenariat avec l'ADEME,
soutenue par EDF

3.C

Composant serveurs – IT Hardware



[Retour au sommaire](#)

● Estimation des performances énergétiques liées aux composants des serveurs – optimisation Hardware

Méthodologie

Pour chaque famille d'usage :

1. **Identification** du ou des **composant(s)** au cœur de la structure en termes de consommation.
2. A partir de son niveau de performance opérationnel et de son **Thermal Design Power (TDP en W)** : déduction d'un **niveau de performance énergétique**.

Forces

- Données disponibles dans le temps.
- Possibilité de voir l'évolution de la performance dans le temps

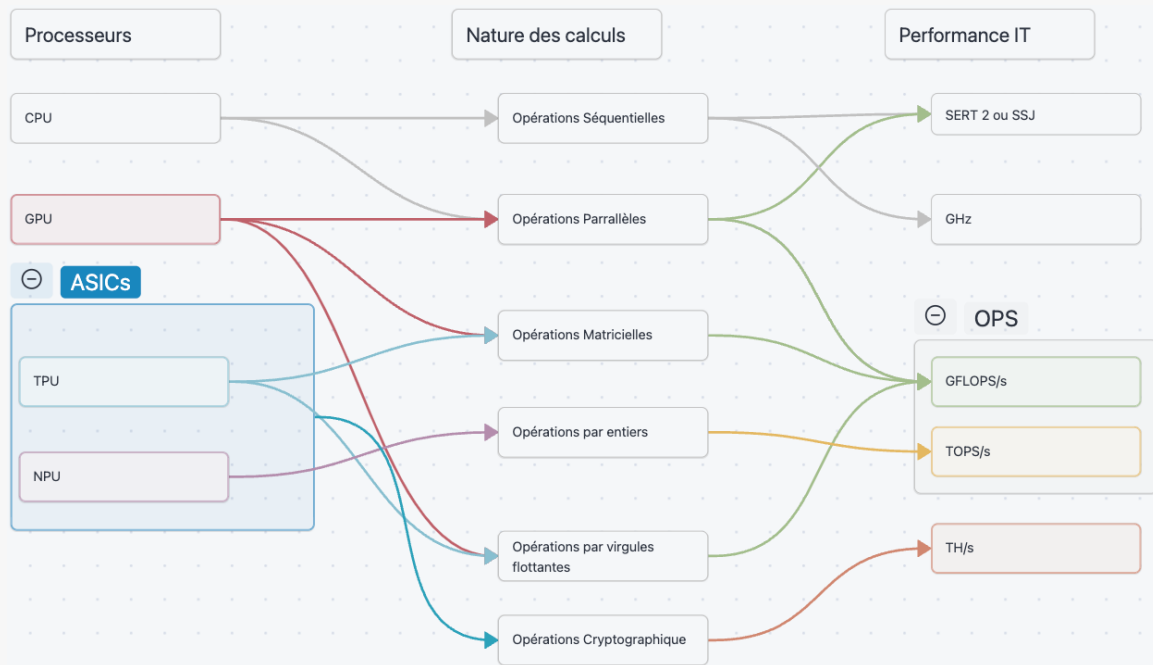
Faiblesses

- Les composants ne fonctionnent pas à leur TDP en permanence (dépendance d'un taux de charge)
- Les serveurs étant un ensemble de composants, se concentrer sur un composant ne permet pas de comprendre comment le système (Hardware + Software) fonctionne globalement

● Les composants d'un serveur

Un serveur est composé de plusieurs éléments : bloc d'alimentation, composant de refroidissement, des entrées et des sorties, de la mémoire vive, de la mémoire secondaire et des processeurs.

Ces derniers concentrent la plus grande part de l'énergie consommée d'un serveur.



● Les usages des différentes typologies de serveurs :

CPU
(Central processing unit)



Est plus performant dans le traitement séquentiel des données et le pilotage des actions générales du serveur. Utilisé dans l'ensemble des serveurs « classiques » et IA. Est associé à des accélérateurs (GPU, TPU, NPU ...) pour l'IA, le HPC et de manière générale les applications hautes densité/besoins en calculs élevés.

GPU
(Graphics processing unit)



Initialement utilisé pour la production graphique (jeux vidéos, CGI, 3D), permet d'effectuer un grand nombre de calculs en parallèle. Est aujourd'hui utilisé pour les applications IA/HPC, dont les modèles nécessitent d'énormes volumes d'opérations simultanées.

ASICs
(Application-specific integrated circuit, TPU ; NPU)



Les ASICs sont des accélérateurs conçus pour une tâche très spécifique.
TPU : Tensor Processing Unit -> Par exemple pour l'entraînement de certains modèles IA
NPU : Neural Processing Unit -> Par exemple pour les calculs de réseaux neuronaux en local (vision, audio, IA embarquée)

● Méthodologie globale relative aux estimations et projections de consommation - IT Hardware

Périmètre retenu

Focalisation sur les CPU et GPU, représentant les processeurs les plus courants et sur lesquels le plus de données disponibles ont pu être analysées.

Sources

L'ensemble des données présentées sont basées sur :

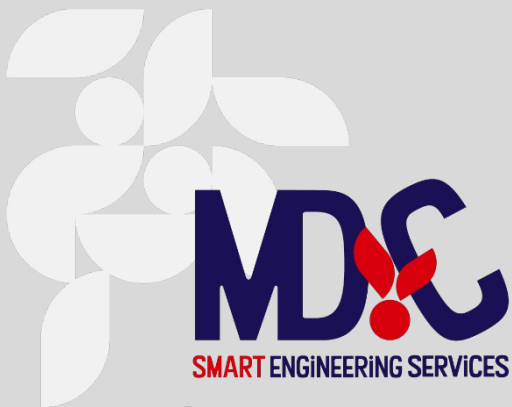
- Des méthodes de calcul des performances reconnues internationalement
- Des benchmarks publics ou des données de consommation en open source, fournies par les fabricants et utilisateurs de serveurs

Approche

Comparaison des données disponibles et des indicateurs retenus pour obtenir un historique de performance fiable

Projections et économies d'énergie

- Réalisation de projections sur la base de la tendance observée sur l'historique des données
- Simulation d'économies d'énergie sur des quantités de données/calculs/opérations produites rapportées aux projections de la croissance IT nationale



Étude 2025

Datcenters IT & IA



Une étude menée par MD.C,
en partenariat avec l'ADEME,
soutenue par EDF

Analyse - CPU



[Retour au sommaire](#)

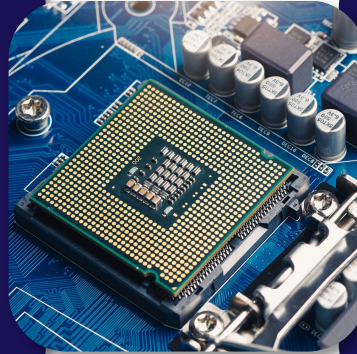
● Evaluation de la performance des CPU

Méthode

Approche sur deux niveaux : composants et benchmarks.

Niveau 1

- Récolte de données auprès des benchmarks/données open source Chip data set, SPEC, 4E TCP
- Mise en commun des sources pour évaluer la vitesse d'horloge moyen en GHz
- Mise en commun du thermal design power moyen par année
- Etablissement d'un indicateur de performance énergétique via le ratio entre vitesse d'horloge et consommation (W).



Niveau 2

- Récolte de donnée historique sur deux benchmarks SPECpower_ssj2008 et SERT2. Moyennisation des indicateurs de performances par année.

Les deux niveaux d'examen permettent une approche cohérente en évaluant à la fois l'évolution du matériel et l'évolution de sa performance

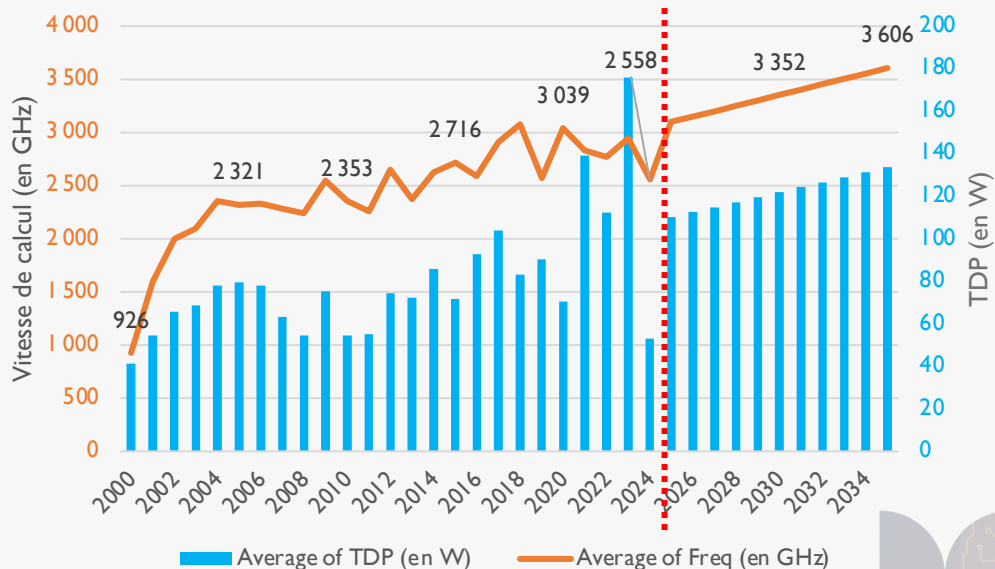
● Evolution de la vitesse d'horloge des CPU

La vitesse d'horloge est la fréquence à laquelle un processeur exécute ses cycles de calcul, c'est-à-dire le nombre d'opérations qu'il peut effectuer par seconde (en Hertz).

On observe un **gain de 30%** en vitesse d'horloge (GHz) entre 2010 et 2020.

Sur la base d'une **projection** considérant une **tendance constante** on peut estimer un **gain de 10%** entre 2020 et 2030

Evolution des vitesses d'horloge moyennes des CPU



Aucune projection ou estimation contenue dans cette présentation ne saurait être entendue ou comprise comme constituant une conclusion ou une préconisation définitive de M.D.C. L'ensemble des éléments avancés sont sujets à révision éventuelle pour la restitution finale.

Source : Chip data set

● Calcul SERT2

Le **benchmark SERT** (méthodologie **SERT 2**) est un outil normatif utilisé pour certifier des serveurs aux Etats-Unis et au sein de l'Union Européenne.

- Il évalue le CPU, la mémoire vive et la mémoire secondaire. Le système est soumis à différents niveaux de charge et doit effectuer des charges de travail.
- Les valeurs individuelles de chaque composant sont ensuite regroupées dans une moyenne géométrique pondérée.

Méthode de calcul de performance :

1. Récolte de données historiques sur le benchmark SERT2.
2. Moyennisation des indicateurs de performance par année.

$$SERT2 = \exp (0.65 * \ln(EffCPU) + 0.3 * \ln(EffMemory) + 0.05 * \ln(EffStorage))$$

Source : 4E TCP

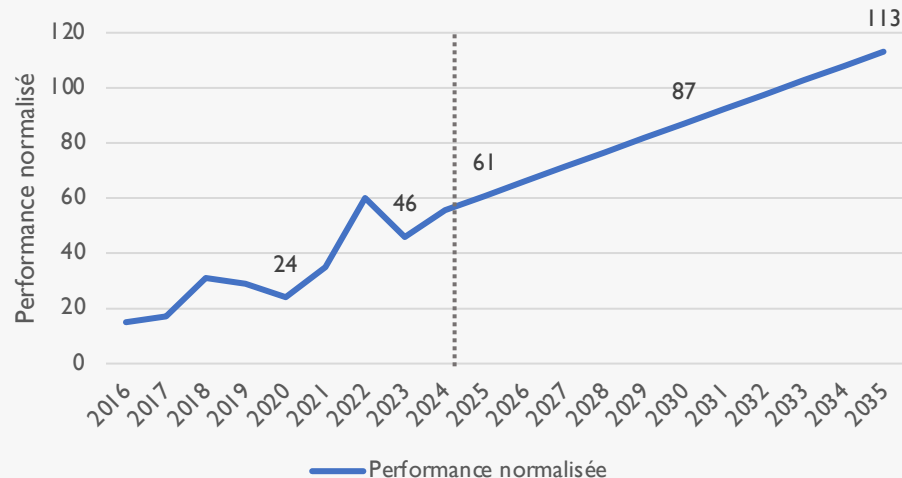
● Evolution de la performance des serveurs (SERT2)

SERT2 est un outil normatif, il teste le serveur entier (CPU, mémoire vive et mémoire secondaire) :

$$EE = \frac{Perf\ norm}{W}$$

Gain de performance
91,6% entre 2020 et
2023

Evolution efficience des serveurs



Aucune projection ou estimation contenue dans cette présentation ne saurait être entendue ou comprise comme constituant une conclusion ou une préconisation définitive de M.D.C. L'ensemble des éléments avancés sont sujets à révision éventuelle pour la restitution finale.

Source : 4E TCP

● Calcul SPECpower_ssj2008

Le benchmark **SPECpower_ssj2008** est un outil de test de performance sur une application Java. Il soumet au système une charge de travail consistant à réaliser des opérations standardisées.

Le système doit accomplir des Server-Side Java operations à divers taux de charge allant de 0% (Active Idle) à 100% par saut de 10%.

Méthode de calcul de performance :

1. Récolte de données historiques sur le benchmark **SPECpower_ssj2008**.
2. Moyennisation des indicateurs de performance par année.

Performance			Power	Performance to Power Ratio
Target Load	Actual Load	ssj_ops	Average Active Power (W)	
100%	99.8%	1,855,713	196	9,485
90%	90.0%	1,672,373	162	10,324
80%	79.8%	1,482,832	142	10,414
70%	70.1%	1,303,849	129	10,080
60%	60.1%	1,117,171	117	9,584
50%	50.0%	930,399	104	8,913
40%	40.1%	745,737	93.3	7,992
30%	29.9%	555,427	82.8	6,705
20%	20.1%	373,154	72.3	5,159
10%	10.1%	187,092	61.0	3,067
Active Idle		0	36.0	0
Σ ssj_ops / Σ power =				8,550

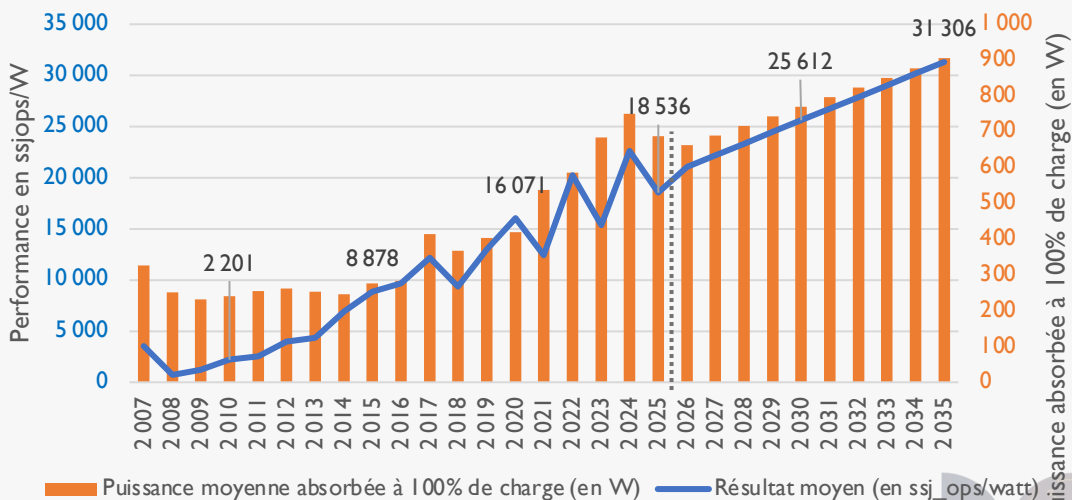
Source : Fsas
Technologies

● Evolution de la performance des serveurs (SPECpower_ssj2008)

SPECpower_ssj2008 est un benchmark qui évalue un **ratio** entre performance du serveur et sa consommation réelle

**Gain de performance
énergétique de 800%
entre 2010 et 2025**

Evolution de la performance serveur double-CPU selon SPECpower_ssj2008



Aucune projection ou estimation contenue dans cette présentation ne saurait être entendue ou comprise comme constituant une conclusion ou une préconisation définitive de M.D.C. L'ensemble des éléments avancés sont sujets à révision éventuelle pour la restitution finale.

Source : Spec.org

● Synthèse de l'évolution des performances des CPU

On observe donc une concordance de l'évolution de la performance des CPU selon les divers indicateurs.

Entre 2018 et 2023 :

Gain de de **48,38%** de
performance énergétique avec le
benchmark SERT2

Gain de **63,87%** avec le
benchmark SPECpower_ssj2008

Gain de 95,77% dans le TDP des CPU dans cet intervalle.

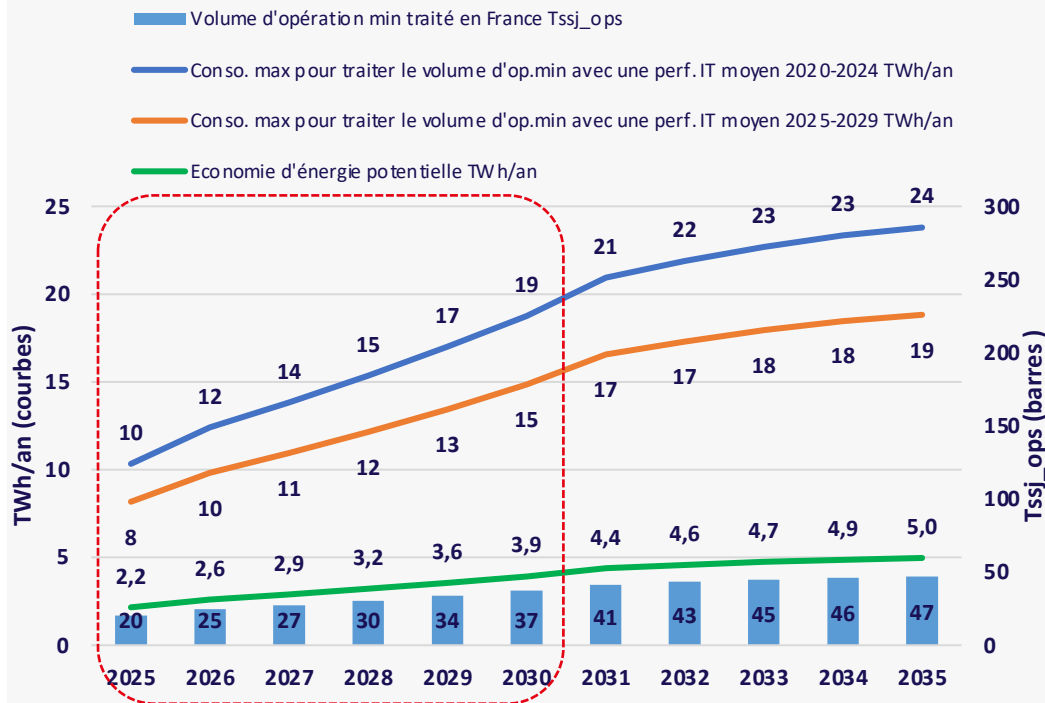
● Méthodologie de simulation des gains énergétiques via le déploiement de CPU plus performants indexés sur la croissance de la filière datacenter/IT

A partir la **performance des CPU en ssj_ops/watt** et de la **puissance IT installée** (hypothèses haute et basse), nous pouvons déterminer une **équivalence de quantité d'opérations en ssj_ops, réalisées dans le parc IT français.**

Les consommations énergétiques estimées plus loin représentent la quantité d'énergie qui aurait été consommée dans les datacenters en France pour traiter cette quantité d'opérations avec 2 niveaux de performance :

- Le niveau de performance moyen des CPU compris entre 2020 et 2024 inclus
- Le niveau de performance moyen des CPU compris entre 2025 et 2029 inclus

Consommation et gain énergétique pour le traitement du volume de données minimum - hypothèse basse croissance IT

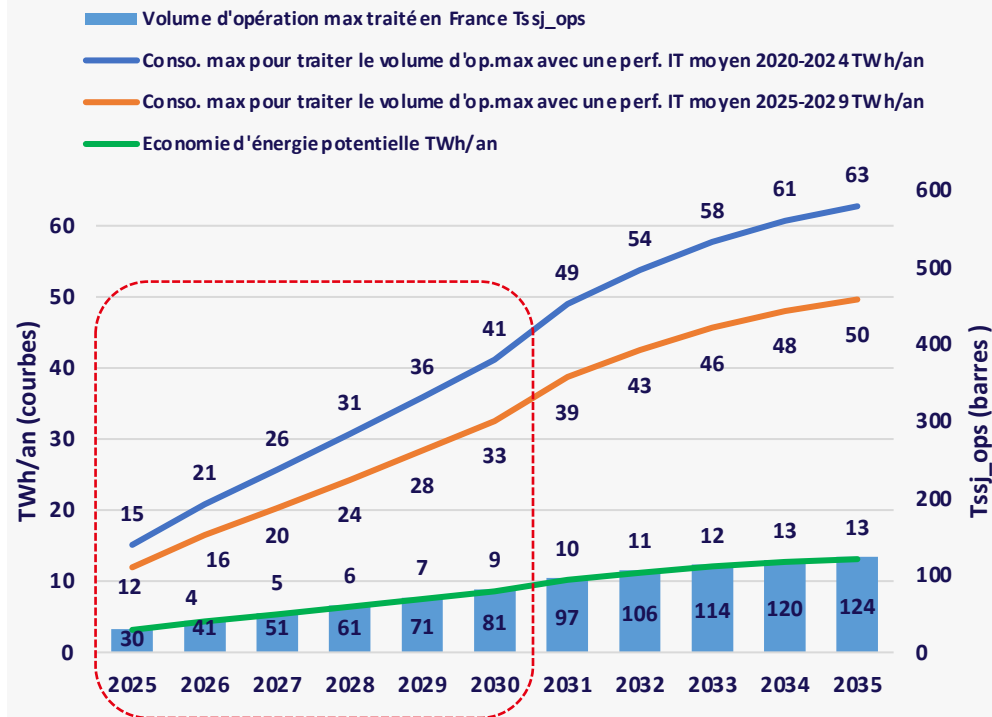


En appliquant l'approche détaillée dans la méthodologie, on peut estimer, de manière relative les éléments suivants :

- L'évolution des performances des CPU est significative et représente des économies d'énergie non négligeables au niveau national
- L'évolution générale des performances IT entre la période 2020-2025 et la période projetée 2025-2030 des CPU pourrait permettre un gain énergétique moyen (à volume d'opérations égal et indexé sur la croissance IT) de :

3,07 TWh/an

Consommation et gain énergétique pour le traitement du volume de données maximal - hypothèse haute croissance IT



En appliquant l'approche détaillée dans la méthodologie, on peut estimer, de manière relative les éléments suivants :

- L'évolution des performances des CPU est significative et représente des économies d'énergie non négligeables au niveau national
- L'évolution générale des performances IT entre la période 2020-2025 et la période projetée 2025-2030 des CPU pourrait permettre un gain énergétique moyen (à volume d'opérations égal et indexé sur la croissance IT) de :

5,67 TWh/an

● Conclusion de l'évolution des CPU et projection

Les performances des CPU augmentent de manière significative dans le temps, quel que soit l'indicateur ou le benchmark retenu.

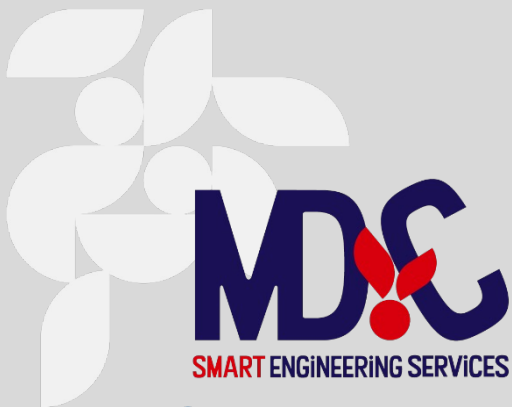
A date, et en indexant l'évolution des performances sur la croissance de la puissance IT déployée en France, on peut projeter une économie d'énergie relative comprise entre

15 et 25 TWh cumulés

à horizon 2030, en comparaison au niveau de performance atteint en 2025

Eléments à renforcer/compléter :

- L'évaluation de la vitesse d'horloge n'est pas suffisante pour évaluer la performance du CPU. C'est pourquoi nous allons creuser l'évolution du nombre de cœurs et l'accès à la mémoire pour tirer des conclusions sur la performance globale des CPU en lien avec les deux benchmarks.
- Par ailleurs, pour déterminer des préconisations opérationnelles d'actions d'efficacité énergétique, **une situation de référence doit être déterminée et consolidée**



Étude 2025

Datcenters IT & IA



Une étude menée par MD.C,
en partenariat avec l'ADEME,
soutenue par EDF

Analyse - GPU



[Retour au sommaire](#)

● Evaluation de la performance des CPU

Méthode

- Récolte de données auprès des benchmarks/données open source Chip data set, Epoch AI, 4E TCP
- Mise en commun des sources pour évaluer la vitesse de traitement en GFLOP/s
- Mise en relation avec le thermal design power (TDP) moyen par année
- Etablissement d'un indicateur de performance énergétique via le ratio entre GFLOP/s et consommation (W).



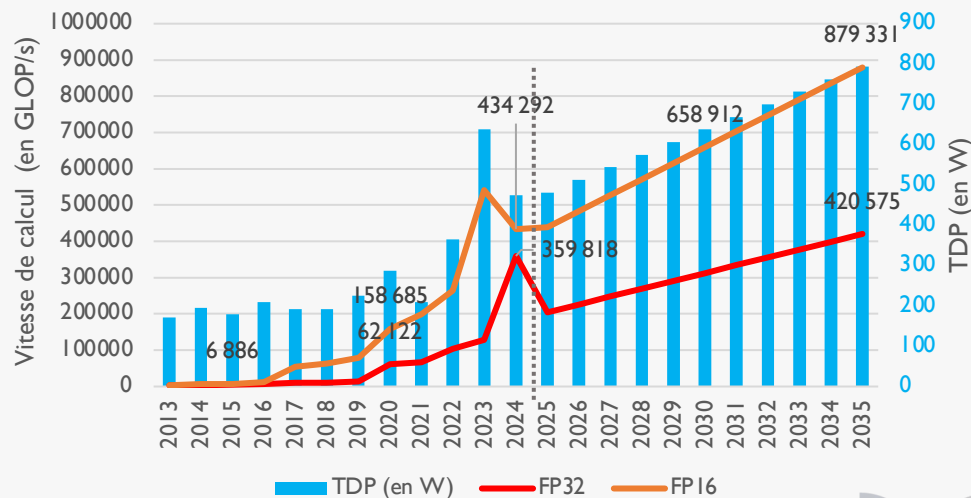
● Evolution de la vitesse de traitement des GPU et de leur TDP

En se basant sur la vitesse de calcul et le TDP, issus des fiches techniques des GPU :

En fonction de deux niveaux de précisions, FP32 et FP16

273% de gain entre 2020 et 2024 pour FP16, comparable en FP32

Evolution de la puissance de calcul moyenne des GPU dans le temps



Source : Epoch AI,
Chip data set & 4E TCP

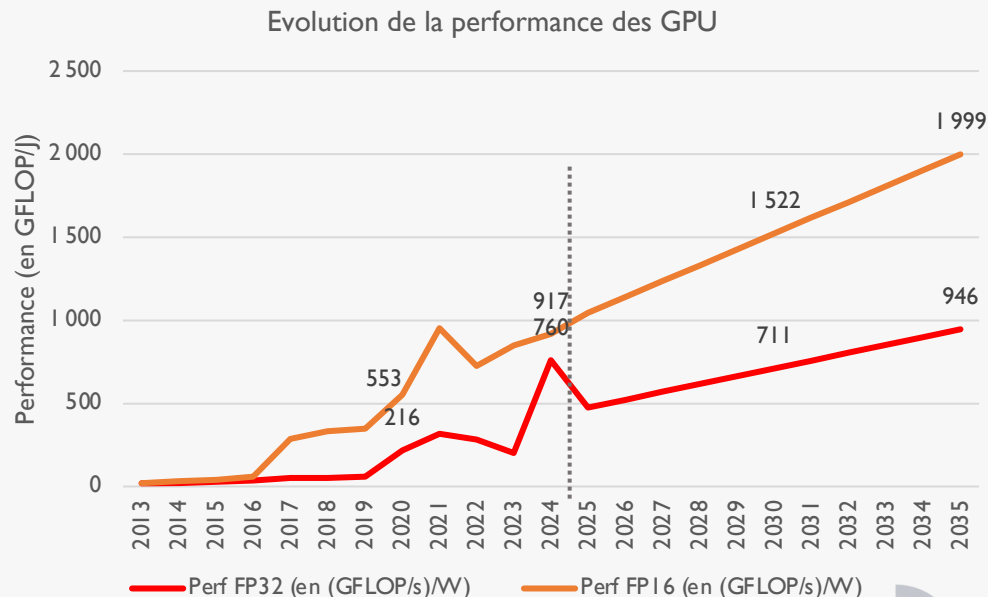
Aucune projection ou estimation contenue dans cette présentation ne saurait être entendue ou comprise comme constituant une conclusion ou une préconisation définitive de M.D.C. L'ensemble des éléments avancés sont sujets à révision éventuelle pour la restitution finale.

● Evolution de la performance énergétique des GPU

Performance énergétique
d'un GPU : GFLOP/s/W

Ratio entre puissance de
calcul et TDP

**165% de gain entre 2020
et 2024 en moyenne en
FP32 et FP16**



Source : Epoch AI,
Chip data set & 4E TCP

Aucune projection ou estimation contenue dans cette présentation ne saurait être entendue ou comprise comme constituant une conclusion ou une préconisation définitive de MD.C. L'ensemble des éléments avancés sont sujets à révision éventuelle pour la restitution finale.

● Synthèse de l'évolution des performances des GPU

On observe donc une concordance
de l'évolution de la performance des GPU.

Entre 2018 et 2023 :

Gain de de **400 %** de performance
énergétique en FP32

Gain de **254,65%** de performance
énergétique en FP16

Augmentation de 332,3% dans le TDP des GPU dans cet intervalle.

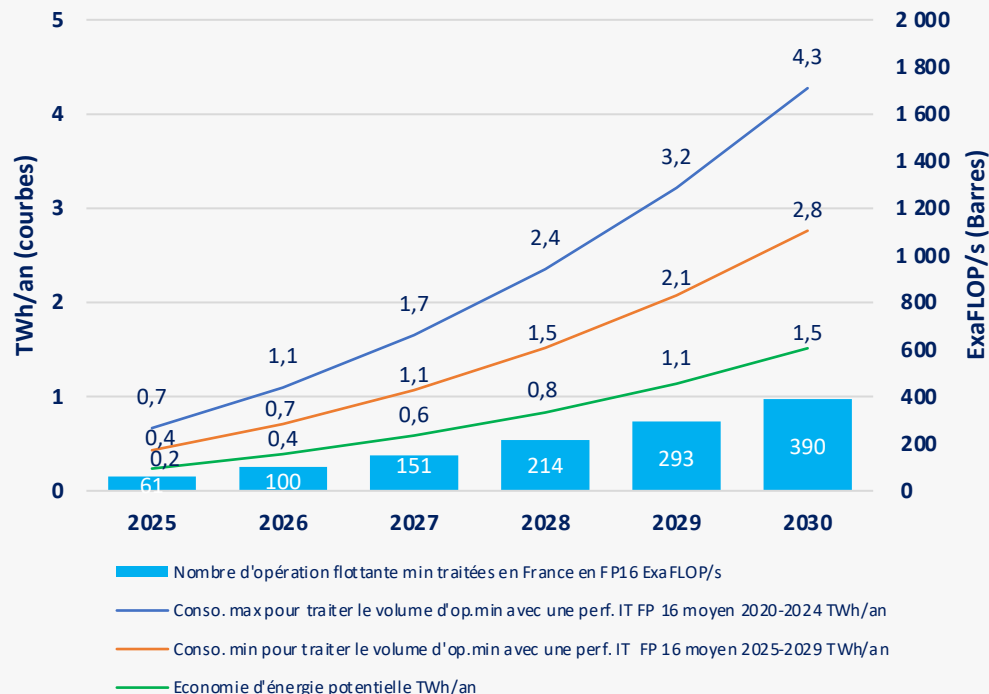
● Méthodologie de simulation des gains énergétiques via le déploiement de GPU plus performants indexés sur la croissance de la filière datacenter/IT

A partir la **performance des GPU en Flops/watt** et de la **puissance IT installée** (hypothèses haute et basse), nous pouvons déterminer une **équivalence de quantité d'opérations en Flops, réalisées dans le parc IT français**.

Les consommations énergétiques estimées représentent la quantité d'énergie qui aurait été consommée dans les datacenters en France pour traiter cette quantité d'opérations avec deux niveaux de performance :

- Le niveau de performance moyen des GPU compris entre 2020 et 2024 inclus
- Le niveau de performance moyen des GPU compris entre 2025 et 2029 inclus

Consommation et gain énergétique pour le traitement du volume de données minimum en FP16 - hypothèse basse de croissance de la filière

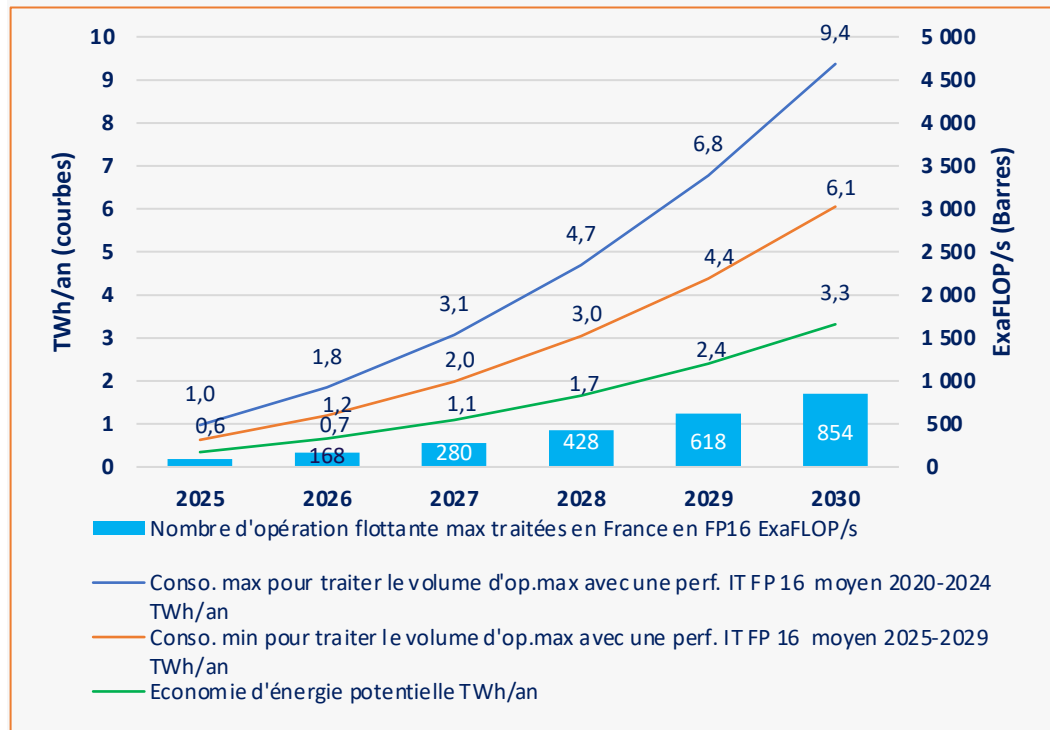


En appliquant l'approche détaillée dans la méthodologie, on peut estimer, de manière relative les éléments suivants :

- L'évolution générale des performances IT des GPU entre la période 2020-2025 et la période projetée 2025-2030 des GPU pourrait permettre un gain énergétique moyen (à volume d'opération égal et indexé sur la croissance IT) de :

0,67 TWh/an

Consommation et gain énergétique pour le traitement du volume de données minimum en FP16 - hypothèse haute de croissance de la filière

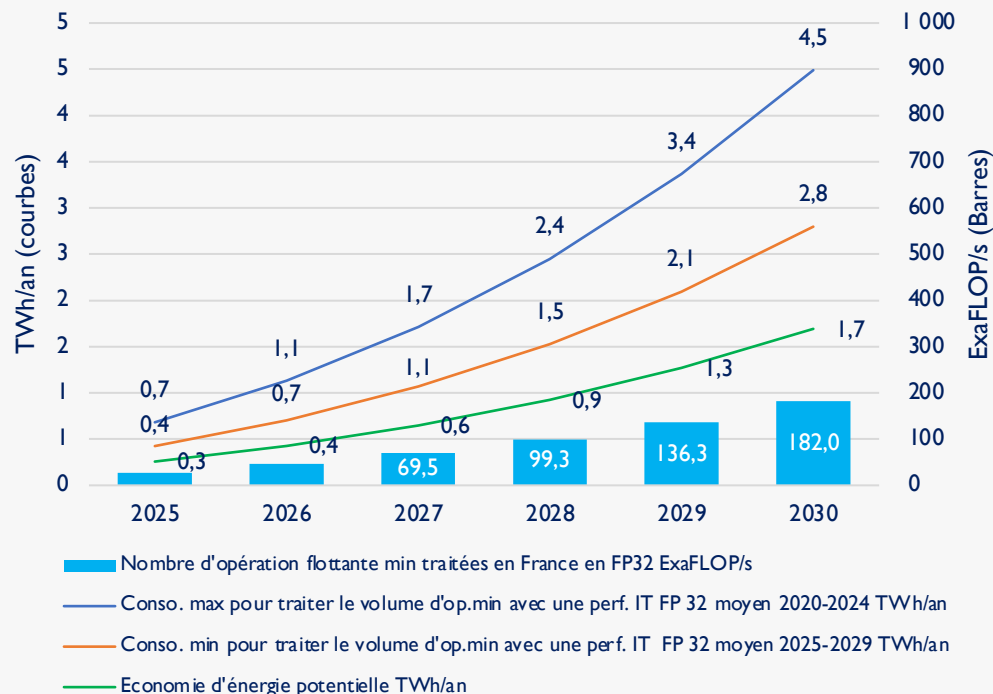


En appliquant l'approche détaillée dans la méthodologie, on peut estimer, de manière relative les éléments suivants :

- L'évolution générale des performances IT des GPU entre la période 2020-2025 et la période projetée 2025-2030 des GPU pourrait permettre un gain énergétique moyen (à volume d'opération égal et indexé sur la croissance IT) de :

1,6 TWh/an

Consommation et gain énergétique pour le traitement du volume de données minimum en FP32 - hypothèse basse de croissance de la filière

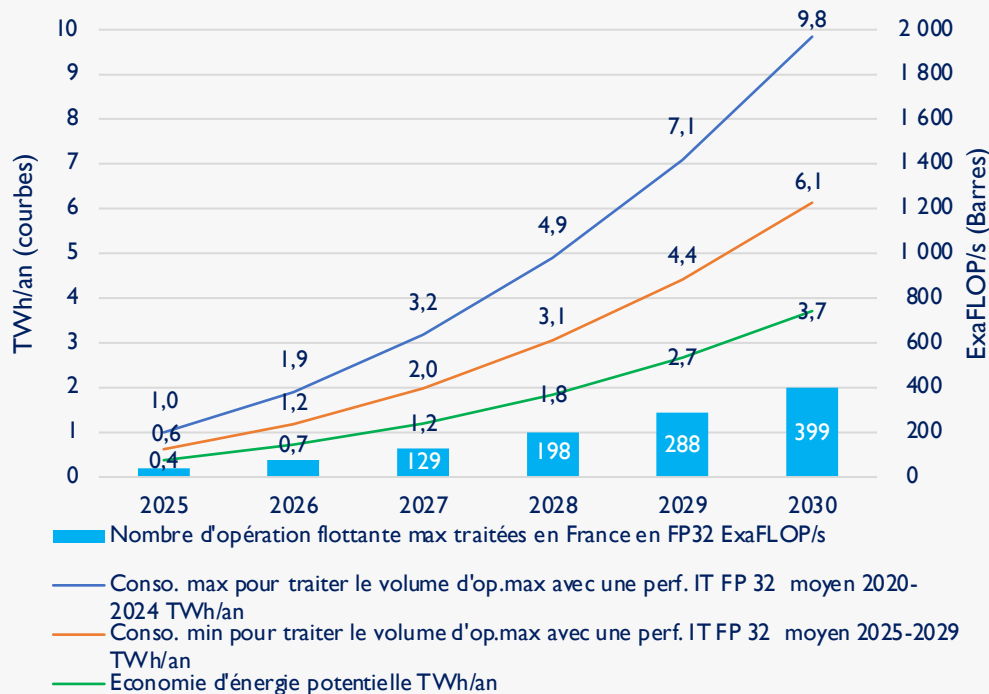


En appliquant l'approche détaillée dans la méthodologie, on peut estimer, de manière relative les éléments suivants :

- L'évolution générale des performances IT des GPU, entre la période 2020-2025 et la période projetée 2025-2030, pourrait permettre un gain énergétique moyen (à volume d'opérations égal et indexé sur la croissance IT) de :

0,87 TWh/an

Consommation et gain énergétique pour le traitement du volume de données minimum en FP32 - hypothèse haute de croissance de la filière



En appliquant l'approche détaillée dans la méthodologie, on peut estimer, de manière relative les éléments suivants :

- L'évolution générale des performances IT des GPU, entre la période 2020-2025 et la période projetée 2025-2030, pourrait permettre un gain énergétique moyen (à volume d'opérations égal et indexé sur la croissance IT) de :

1,75 TWh/an

● Conclusion de l'évolution des GPU et projection

La performance moyenne des GPU augmente de manière significative dans le temps

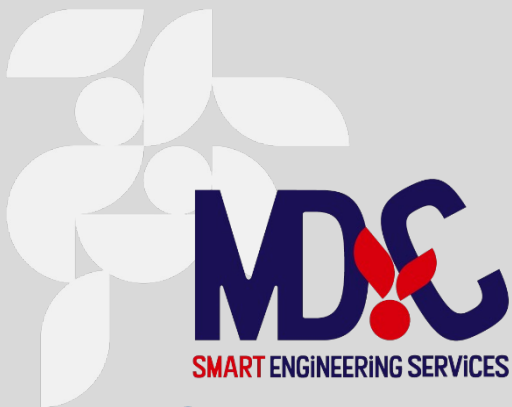
A date, et en indexant l'évolution des performances sur la croissance de la puissance IT déployée en France, on peut projeter une économie d'énergie relative comprise entre

3,35 et 8,6 TWh cumulés

à horizon 2030, en comparaison en niveau de performance atteint en 2025

Éléments à renforcer/compléter :

- La suite consistera à trouver des benchmarks permettant d'évaluer le GPU.
- Par ailleurs, pour déterminer des préconisations opérationnelles d'actions d'efficacité énergétique, **une situation de référence doit être déterminée et consolidée.**



Étude 2025

Datacenters IT & IA



Une étude menée par MD.C,
en partenariat avec l'ADEME,
soutenue par EDF

Conclusion intermédiaire : Analyse – GPU & CPU



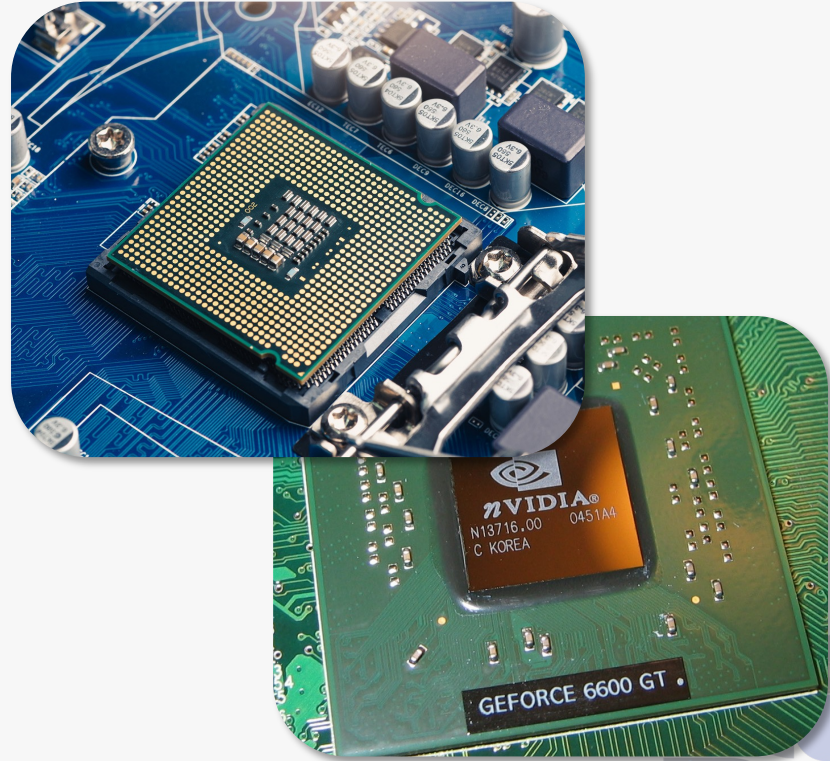
[Retour au sommaire](#)

● Conclusion intermédiaire - Analyse de performance GPU & CPU

Les premières données analysées dans le cadre de l'étude 2025 soulèvent :

- Une évolution rapide des performances énergétiques des principaux processeurs composant les serveurs ('standards', IA/HPC)
- Des indicateurs fiables de performance existants

L'aspect Hardware « pur » (performance « usine ») des équipements n'est qu'un des aspects de la performance IT, la mise en adéquation des équipements/usages doit également être considérée



MDC **SMART**
ENGINEERING
SERVICES

 Étude 2025
Datacenters IT & IA



Une étude menée par MD.C,
en partenariat avec l'ADEME,
soutenue par EDF

3.D

Effacité énergétique des serveurs – IT infrastructure



[Retour au sommaire](#)

● Estimation des performances énergétiques liées aux composants des serveurs – optimisation infrastructure IT

Méthodologie

Pour chaque famille d'usage :

1. **Identification** de l'indicateur ou des indicateurs permettant de relever et comparer les performances énergétiques.
2. A date, et en l'absence de données avec une antériorité significative, réalisation d'une première analyse sur la base des benchmarks existants - MLPerf

Forces

- Sources académiques fiables et reconnues
- Benchmark utilisé reconnu par des acteurs de l'industrie et académiques
- Existence d'indicateurs pour l'inférence et l'entraînement de l'IA

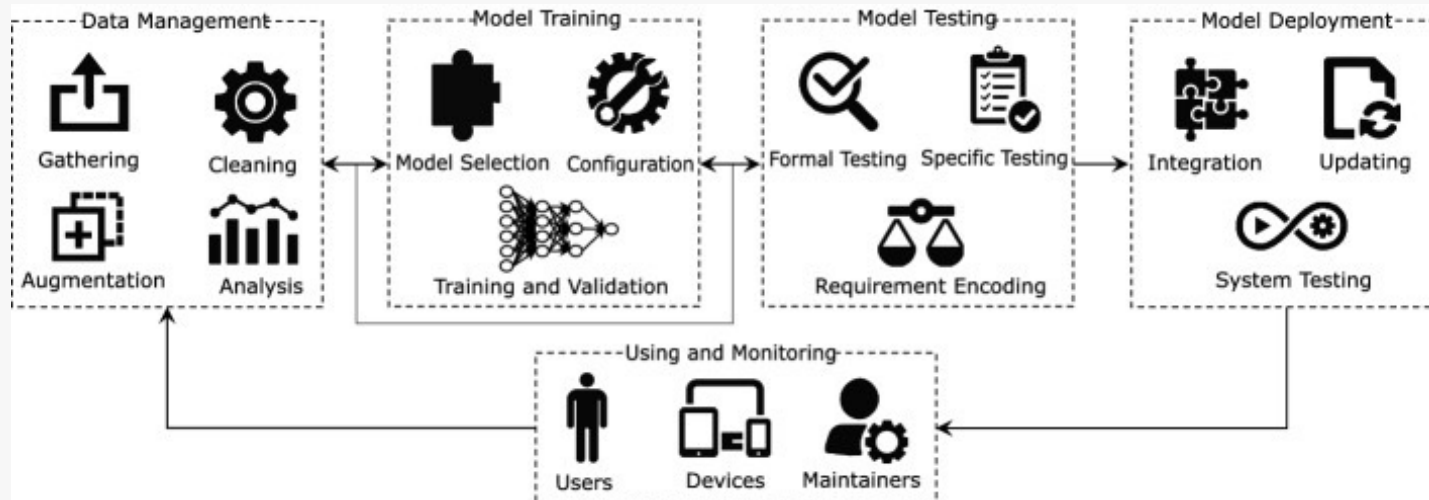
Faiblesses

- Absence de période significative avec des données relevées et comparables
- Difficulté de déterminer les facteurs pertinents du point de vue énergétique étant donné la nature multifactorielle des données analysées

● Les deux phases de l'IA

On peut séparer l'intelligence artificielle en deux segments:

1. Machine Learning : la phase d'entraînement du modèle à partir d'une base de données
2. Inférence : Modèle confronté à de nouvelles données et doit se servir de son modèle pour agir.



Source : [C. Yang & al](#)

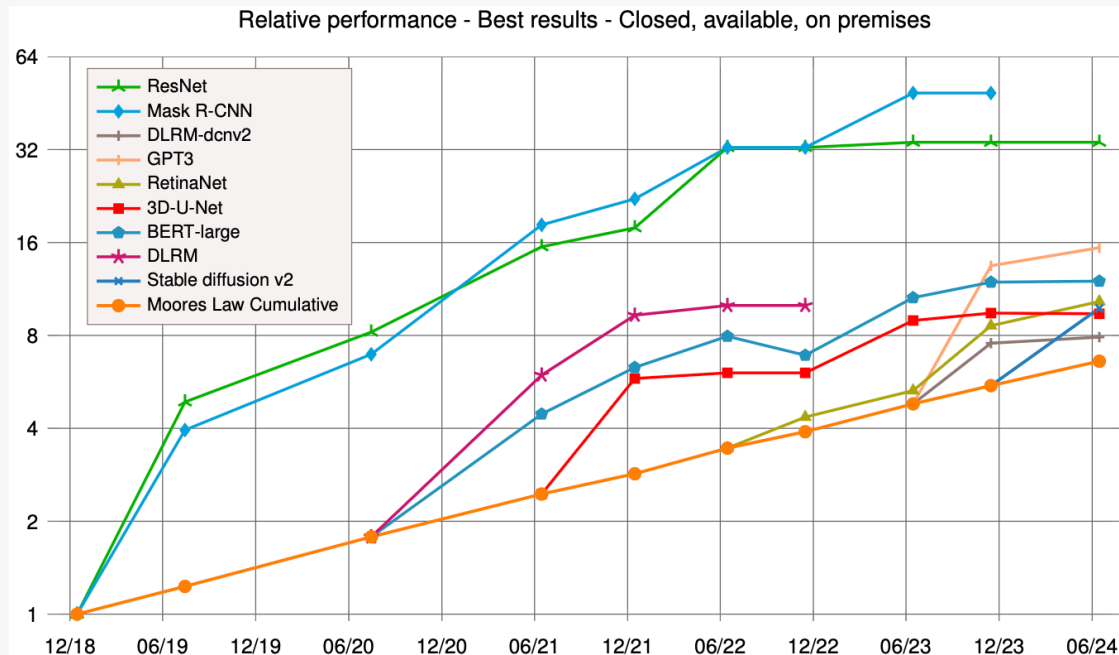
● Evolution de la performance du machine learning

Données issues de l'étude d'Harvard, basée sur MLPerf Training.

MLPerf Training est un benchmark basé sur des données OpenSource fournies par les fabricants de serveurs et les opérateurs IA

Performance : temps nécessaire à l'entraînement - base 1 ; 2018

Augmentation moyenne de 1700% de la performance entre 2018 et 2024



Source : [étude Harvard](#)

Aucune projection ou estimation contenue dans cette présentation ne saurait être entendue ou comprise comme constituant une conclusion ou une préconisation définitive de M.D.C. L'ensemble des éléments avancés sont sujets à révision éventuelle pour la restitution finale.

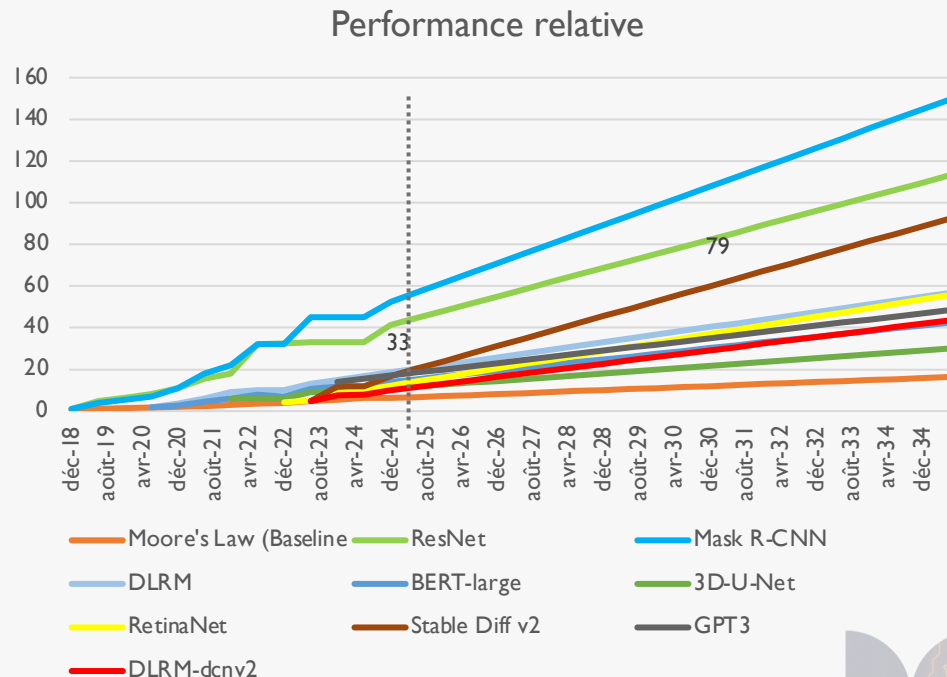
● Evolution et projections de la performance du machine learning - exemples

Gain entre juin 2019 et juin 2024 :

- **ResNet : + 660%**
- **Mask R-CNN : + 1125 %**

Si l'optimisation des systèmes continue avec la même tendance jusqu'en 2030, on peut projeter un gain additionnel :

- **ResNet : + 239,4 %;**
- **Mask R-CNN : + 228,9%.**



● Evolution de la performance de l'inférence

Données issues d'une extrapolation (étude Harvard) basée sur **MLPerf Inference**

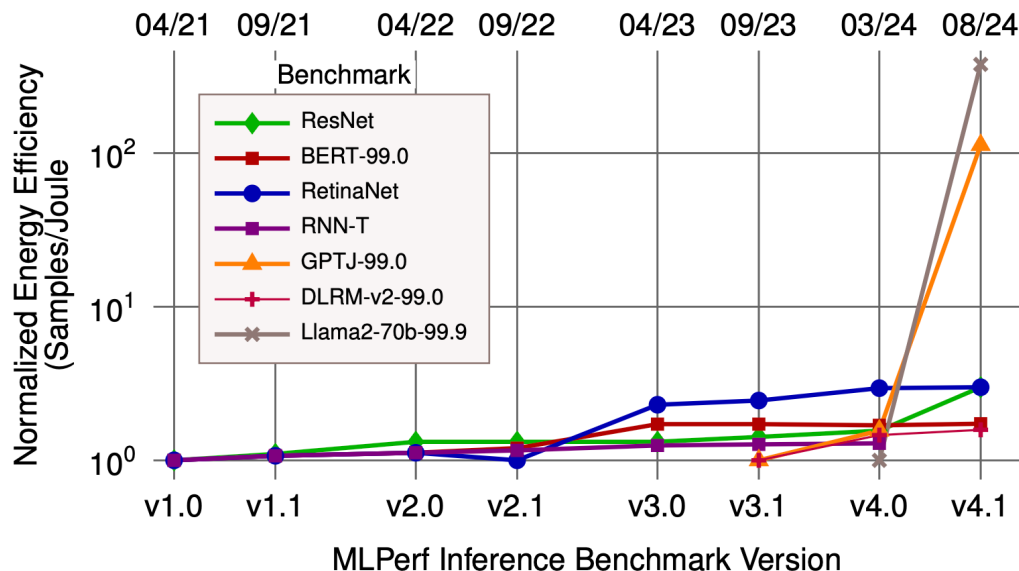
MLPerf Inference est un benchmark qui relève le coût énergétique nécessaire pour produire une inférence.

Performance :

- Samples/J,
- Queries/J,
- Token/J.

Gain de performance :

- **GPTJ-99.0 : + 73,33% en 5 mois;**
- **RetinaNet : + 290% en 28 mois.**



● Synthèse de l'évolution des performances des systèmes AI (I)

On observe donc une évolution positive de la performance des systèmes AI en machine learning et inférence, soulignée par les données disponibles à date et les études techniques disponibles, mais difficilement standardisables à ce stade :

Limites à ce stade :

Manque de recul historique sur les indicateurs et les benchmarks d'évaluation des systèmes IA

Cartographie non exhaustive des leviers d'amélioration de la performance énergétique des systèmes IA

Données relatives aux indicateurs de performance énergétique à compléter

● Synthèse de l'évolution des performances des systèmes AI (II)

Néanmoins, possibilité d'utiliser les données existantes pour créer et proposer une méthode de comparaison des performances entre infrastructures IT pour les deux usages étudiés (AI Inference ; AI Training) :

Indicateurs standardisés déjà utilisables et pris en compte pour la partie inférence :

Samples/J

Token/J

Queries/J

Indicateurs relevés à date pour la partie training :

Time to train rapporté aux conso
énergétiques (Wh)

4

Conclusion générale et étapes suivantes de l'étude



[Retour au sommaire](#)

MDC

**SMART
ENGINEERING
SERVICES**



Étude 2025

Datacenters IT & IA



Une étude menée par MDC,
en partenariat avec l'ADEME,
soutenue par EDF

● Compilation des perspectives d'optimisation 'DC, IT et IA' : En résumé

Infrastructure Technique Datacenter

- Des suites concrètes données à l'étude 2024 à travers, et en parallèle de l'étude 2025
- Une réactualisation en cours des projections et un focus sur les technologies à fort potentiel
- Une transition en cours de la filière vers le refroidissement liquide
- Des économies d'énergie potentielles de **2,8 TWh à 7,3 TWh** annuels d'ici 2035 pour les cinq technologies citées

Monitoring des serveurs

- Une première approche standardisable identifiée
- Une solution à maturité et déployable dans le parc
- Des économies d'énergie potentielles significatives

IT Hardware

- Des indicateurs identifiés et standardisables
- Des évolutions claires des principaux composants vers plus efficacité énergétique
- Une évolution en termes d'économies d'énergie mesurable dans le temps

IT Infrastructure

- Des données existantes
- Des indicateurs partiels relevés
- De premières tendances et méthodes de mesurage de l'efficacité énergétique déterminées

● Next steps – Reste à faire

Partie Infrastructure Technique Datacenter :

- Renforcement du pool de datacenters et fournisseurs audités afin de consolider les projections et les estimations de performance
- Détection et renforcement du travail sur les autres technologies efficaces (citées ou complémentaires)

Partie Infrastructure Technique Datacenter :

- Renforcement du pool de datacenters et fournisseurs audités afin de consolider les projections et les estimations de performance
- Détection et renforcement du travail sur les autres technologies efficaces (citées ou complémentaires)

Partie IT Infrastructure :

- Evaluation des systèmes standard et HPC
- Technologie de Dynamic Voltage and Frequency Management (DVFM) - mesure de l'impact
- Définition des familles d'usage IT pour adaptation cohérente des indicateurs
- Détermination d'une méthodologie d'analyse des performances permettant d'isoler et mettre en relation les facteurs (langage, type d'IA, infrastructure physique etc)
- Impact de la virtualisation

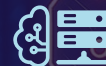


Call to Action

Acteurs et opérateurs IT & Infra DC



SMART
ENGINEERING
SERVICES



Étude 2025
Datcenters IT & IA



Une étude menée par MDC,
en partenariat avec l'ADEME,
soutenue par EDF

Appel à participation

CEE

MD.C se propose de continuer à travailler avec les acteurs de l'administration en charge du dispositif (ATEE, ADEME) et en partenariat avec EDF en tant qu'acteur obligé du dispositif CEE pour mettre en place les **groupes de travail visant à la création des fiches** relatives aux technologies présentées

Étude 2025

La présente restitution est une étape intermédiaire de l'étude, dont la finalisation est prévue pour mai 2026

Afin que l'étude puisse être basée sur des données représentatives et permettre de véritables propositions opérationnelles en accord avec la réalité terrain des acteurs concernés, **nous avons besoin de votre participation !**

5

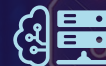
Glossaire



[Retour au sommaire](#)



SMART
ENGINEERING
SERVICES



Étude 2025
Datcenters IT & IA



Une étude menée par MDC,
en partenariat avec l'ADEME,
soutenue par EDF

● Définitions (1)

Monitoring énergétique

Aussi appelé pilotage énergétique, est un processus qui permet de récolter, regrouper, analyser et suivre l'ensemble des données et indicateurs de consommations énergétiques d'un site, ou d'un équipement.

High-Performance Computing (HPC)

Le calcul intensif se réfère à des systèmes informatiques avec une puissance de calcul extrêmement élevée qui sont capables de résoudre des problèmes extrêmement complexes et exigeants.

Machine Learning (ML)

Processus par lequel un algorithme évalue et améliore ses performances sans l'intervention d'un programmeur, en répétant son exécution sur des jeux de données jusqu'à obtenir, de manière régulière, des résultats pertinents.

Inférence AI (AI inference)

L'inférence IA est le processus qui consiste à appliquer ce que le modèle IA a appris grâce au ML pour prendre des décisions, faire des prédictions ou tirer des conclusions à partir des données.

● Définitions (4)

Processeur graphique (GPU)

Un processeur graphique, est un circuit électronique spécialisé conçu à l'origine pour accélérer la création d'images et de vidéos. Sa remarquable capacité à effectuer un grand nombre de calculs rapidement a conduit à son adoption dans divers domaines.

Application-specific integrated circuit (ASIC)

Un ASIC est un circuit intégré construit pour une fonction précise. Les NPU et les TPU en sont des exemples modernes orientés vers l'accélération du calcul d'IA, exploitant leur spécialisation matérielle pour offrir des performances très supérieures aux processeurs généralistes sur ces tâches.

Unité de traitement tensoriel (TPU)

Les TPU sont des accélérateurs d'IA conçus sur mesure et optimisés pour l'entraînement et l'inférence de modèles d'IA.

Unité de traitement neuronal (NPU)

Les NPU sont conçues pour accélérer les tâches et les charges de travail liées à l'IA, telles que le calcul des couches de réseaux neuronaux composées de mathématiques scalaires, vectorielles et tensorielles.

● Définitions (3)

Densité par rack

Quantité de puissance informatique (en kilowatts) ou de serveurs installés dans un seul rack (armoire) d'un centre de données.

Puissance thermique nominale (TDP)

La puissance thermique nominale (TDP) désigne la consommation électrique sous une charge théorique maximale.

Puissance de calcul (Compute power)

La puissance de calcul désigne la capacité de traitement dont disposent les équipements.

Fréquence d'horloge

La fréquence d'horloge mesure le nombre de cycles que votre processeur exécute par seconde. Il est mesuré en GHz (gigahertz).

● Définitions (2)

FLOP

Un flop est l'une des opérations arithmétiques élémentaires +, -, *, / effectuées sur des nombres à virgule flottante.

Nombre à virgule flottante

Les nombres à virgule flottante se structurent en trois segments binaires : le signe, l'exposant et la mantisse (ou fraction). Les différentes structures sont le FP64 (double précision), le FP32 (standard) et le FP16 (demi-précision).

Jeton (Token)

Les Tokens sont les unités fondamentales du texte que les modèles d'IA utilisent pour comprendre et traiter le langage. les Tokens peuvent être des mots ou des phrases.

Requête (Query)

Une requête est une demande unique envoyée au moteur d'inférence.



Merci
pour votre attention !

contact@maxduboisconsultant.fr

www.maxduboisconsultant.fr

