



DATA2040

DATA 2040

Étude prospective sur le futur
des infrastructures numériques
et datacenters en 2040



RAPPORT PUBLIC - MARS 2026

UNE ÉTUDE CEA EN PARTENARIAT AVEC LE GERPA



PARTENAIRES DE L'ÉTUDE :



REMERCIEMENTS

LES PARTENAIRES DE L'ÉTUDE DATA 2040 :

APL : Christophe Weiss, Georges Ouffoué, Thomas Martin

NEXITY : François Desgardins, Gaspard Hantz

HYPERION : Nicolas Boulinguez, Laurent Pingault,

ECLAIRION : Charles Huot et Pierre Leca

SCHNEIDER ELECTRIC : Juan Ignacio Rubio, Jiyue SHEN, Janine Clamo Cramer, Roland Bohrer

EDF : Fabrice Decellas, Stéphane Dupré Latour, Isabelle Moussaoui

RTE : Maité Jaureguy-Naudin, Vincent Lefieux

ENEDIS : Bastien Gauthier, Gaëlle MBock, Jean-Marc Do Livramento, Martin Caron

TDF : Vi Olivet, Bertrand Deschard

ORANGE : Arnaud Braud, Gaël Fromentoux

MICHELIN : André Devillars

THALES : Thierry Betmont, Grégory Marmin

CREDIT AGRICOLE : Aldrick Zappellini, Julien Tournier, Frédéric Orsoni

SNCF : Pierre Fouillade

Le Ministère des Armées : Benjamin Chauvet, Laurent Vogel

CEA : Jacques-Charles Lafoucrière, Cédric Auliac, Eric Mercier

LES EXPERTS AYANT ENRICHIS L'ÉTUDE DATA 2040 :

SCHNEIDER : Rémi Paccou, Sébastien Cruz-Mermy

HPE : Philippe Rase, Ray Beausoleil

EVIDEN : Girault Gauvain, Eric Eppe, Laurent Cargemel

BIOMEMORY : Erfane Arwani

EXAION : Gilles Deleuze

GENVIA : Patrice Tochon

EQUINIX : Marc Penfeunteun, Régis Castagné

DATA4 : Jérôme Totel, Louis Ouvry, Olivier De Nomazy, Alexandre Delaval

OVH Cloud : Leo-Paul Souart

SCALEWAY : Emilien Maudet

GOOGLE : Maud Texier

ERICSSON : Benoit Parneix

OPCORE : Nicolas Fontes

NOKIA : Rany Pitch, Jean-Christophe Antoine

NALCO : Jean-Michel Cottin

SUPERNOVA INVEST : Carole Cazassus, Leo Fernandes, Damien Bretegnier

INRIA : Selma Souihel, Jacques Sainte-Marie

DGE Num : Emma Le Boulcault, Vincent Coudrain, Adrien Laroche

ADEME : Bruno Laffite,

CEA : Fabien Baligand, Nicolas Bihel, François-Marie Bréon, Christophe Calvin, Jean-Philippe Bourgoïn, Patrick Cappe de Baillon, Marc Duranton, Denis Dutoit, Jean-François Fourmigué, Christine Hennebert, Nicolas Lamaison, Benjamin Lucas-Leclin, Pascal Maugis, Blandine Merienne, Jens Merten, Jean-Noël Patillon, Sara Tucci Pergiovanni, Guillaume Ravel

L'ÉQUIPE D'ANIMATION ET FACILITATEURS DE L'ÉTUDE DATA 2040 :

CEA : Bénédicte Almozini, Jean-Baptiste Bel, Philippe Caillol, Tony Maindron, Alexandre Paléologue, Amélie Robuchon, Stéphane Ségard, Timothée Silvestre, Mélusine Suel, Sandra Tochon

GERPA : Régine Monti-Tessier, Stéphane Giron, Jean-Dominique Seval, Vincent Bonneau

Transition Data Lab : Rémi Grimaud

MISE EN FORME DU RAPPORT :

Marie Brochet (CEA)



AVIS AUX LECTEURS

La **prospective n'est pas une prédiction de l'avenir**. Elle n'est **pas non plus une prévision qui serait le prolongement de tendances actuelles** sous forme de « roadmapping ». La prospective est un exercice de réflexion qui s'appuie sur des tendances, signaux faibles et ruptures envisageables pour **décrire les futurs possibles**. Elle vise à **faciliter la prise de décision dans un monde incertain** et l'intégration des horizons de long-terme dans la stratégie. Les travaux présentés dans ce document sont issus d'une production collective qui **ne reflète pas nécessairement toutes les opinions des participants ayant contribué et n'engagent ni les organismes dont ils sont issus, ni le CEA**.

Enfin, les chiffres et estimations qui jalonnent cette étude, qu'il s'agisse de consommation énergétique, d'eau ou de volumes de données, doivent être interprétés comme des **ordres de grandeur relatifs et non pas des valeurs absolues**. Leur unique fonction est d'illustrer des dynamiques et de donner corps à nos scénarios.

PRÉFACE

Face à la dépendance croissante de nos sociétés aux services numériques, les **centres de données constituent aujourd'hui l'épine dorsale de l'économie numérique**. Interconnectés par des réseaux toujours plus performants, ils soutiennent des secteurs tels que la santé, la défense et la sécurité civile, l'administration, la recherche, la finance ou l'industrie.

Au cœur de cette économie numérique, la valeur que représentent les données produites par ces secteurs clés est tout autant stratégique que monétaire. Depuis 2010, le trafic Internet mondial a été multiplié par vingt-cinq, et **la récente vague de l'intelligence artificielle générative** bouleverse encore davantage ce paysage. **L'émergence de « gigadatecenter » hyperscale**, atteignant parfois plusieurs gigawatts de puissance, transforme les équilibres techniques, énergétiques et géopolitiques à l'échelle mondiale.

Dans ce contexte de compétition accrue entre grandes puissances – principalement les États-Unis et la Chine –, la course au gigantisme numérique soulève des **questions essentielles : soutenabilité énergétique et hydrique, dépendance aux matières critiques, infrastructures télécoms, vulnérabilités cyber, ou encore acceptabilité sociétale**. Ces enjeux dépassent largement la seule dimension technologique. Ils appellent à une réflexion collective sur les modèles d'infrastructures et les choix d'indépendance qui façonneront l'Europe numérique de demain.

C'est dans cet esprit qu'a été lancée l'étude prospective **DATA 2040**, portée par le CEA – Direction de la Recherche Technologique (DRT), au sein de sa Direction Innovation, dont la raison d'être est d'apporter **une vision à long terme sur les mutations technologiques et leurs impacts**. Dans la continuité d'autres travaux prospectifs comme celui mené sur le quantique (Quantum

2042), cette démarche illustre la volonté du CEA de conjuguer innovations et stratégies industrielles sur l'ensemble de la chaîne de la valeur.

La prospective, par nature, ne cherche pas à prédire l'avenir mais à **éclairer les futurs possibles**. Elle repose sur une **approche systémique et multipartenariale**, croisant les regards d'experts, d'industriels et d'acteurs publics. À partir de jeux d'hypothèses construits et validés collectivement, quatre scénarios contrastés ont été construits pour explorer les trajectoires possibles du numérique et des datacenters à l'horizon 2040. Ces scénarios constituent **autant d'outils d'aide à la décision pour anticiper les risques, identifier les leviers d'action et préparer des stratégies face à l'incertitude**.

Au-delà du présent rapport d'étude, **DATA 2040** a vocation à **fédérer un écosystème d'entreprises et d'acteurs institutionnels** autour de ces enjeux, dans une logique conservée de **partenariat collaboratif (HUB)** pour réaliser des études techniques et des expérimentations qui permettent de lever les verrous technologiques.

Le futur ne sera certainement pas la simple matérialisation de l'un des scénarios, mais un **mélange dynamique de plusieurs trajectoires**. L'ambition de **DATA 2040** est d'aider à mieux comprendre ces possibles, pour bâtir dès aujourd'hui un **numérique soutenable, résilient, « souverain » et au service de la société**.

Nicolas BEDOUIN

Directeur Délégué à la Stratégie Partenariale
CEA / Direction de la Recherche Technologique

EXECUTIVE SUMMARY

Le rapport DATA 2040 explore les futurs possibles des infrastructures numériques à l'horizon 2040, à l'heure où l'explosion de l'IA générative et les besoins en HPC redéfinissent les règles du jeu.

Face à des défis énergétiques et de ressources, cette étude prospective articule son analyse autour de **quatre scénarios contrastés : Croissance accélérée, Instabilité et cycles technologiques, Nouvelle donne numérique européenne, et Résilience et sobriété**. Ces trajectoires interrogent les incertitudes majeures de demain : centralisation (Cloud) ou décentralisation (Edge) ? Polarisation géopolitique ou multipolarisation ? Continuité technologique ou rupture brutale ?

Les datacenters ne sont plus de simples infrastructures digitales, ils sont devenus des **actifs géopolitiques stratégiques**. On assiste à l'émergence d'un **nouveau paradigme « Compute is a new Oil »** en complément de « Data is gold » où s'opère une **convergence entre l'Énergie et le Digital**. Cette course à la puissance peut entraîner une **pression sur les réseaux électriques et les ressources en eau**, créant des zones de tension (cf. saturation des réseaux en Irlande). Le **rôle du Datacenter devra évoluer** pour ne plus seulement être un consommateur mais devenir un **acteur de flexibilité du réseau électrique** (effacement, stockage, production) et d'un continuum Digital-Énergie.

Cette **révolution digitale est irréversible**, mais sa trajectoire reste à définir. Pour l'Europe et la France, l'enjeu est double. Premièrement, elle doit **s'affirmer dans la compétition mondiale**. Si le retard infrastructurel sur les USA est marqué, **l'Europe dispose néanmoins de leviers stratégiques** pour renforcer son autonomie via **l'innovation technologique et la normalisation sur l'intégralité de la chaîne de valeur**. Deuxièmement, il faudra passer d'une logique de l'efficacité à la résilience car l'optimisation des systèmes ne suffit plus. Face aux chocs potentiels (conflits, crises climatiques), **l'approche doit devenir systémique**. Il ne s'agit plus seulement de subir la croissance, mais de coordonner le développement d'infrastructures dynamiques, capables de s'adapter aux perturbations.

DATA 2040 pose la **question fondamentale de la gouvernance de cette croissance** : laisser faire les lois du marché de manière subie et désordonnée, ou orchestrer une stratégie industrielle et territoriale cohérente ? L'avenir des infrastructures numériques se jouera sur leur capacité à **conjuguer puissance de calcul et sobriété, performance et résilience**.

MESSAGES CLÉS DES PARTENAIRES



TÉLÉCOM

- **Les réseaux de communication évoluent vers des architectures hybrides**, intégrant de manière croissante des solutions terrestres et satellitaires afin d'améliorer couverture, continuité de service et résilience.

- **La virtualisation des réseaux (SDN : SDN – Software-Defined Networking et NFV : Network Functions Virtualization) s'impose comme une tendance structurante**, permettant une automatisation accrue et une meilleure flexibilité opérationnelle, tout en intensifiant la concurrence des acteurs « over-the-top » sur les services à valeur ajoutée.

- **La France bénéficie aujourd'hui d'une latence moyenne faible (≈7 ms sur la plupart des métropoles en France)** et d'une bande passante suffisante pour les usages courants, mais un cycle de renouvellement majeur des infrastructures télécoms sera nécessaire à l'horizon 2033.

- **La collecte massive de données industrielles issues de l'IoT ne se traduit pas par des volumes élevés de données**, car elle repose majoritairement sur de petits paquets d'informations à faible poids unitaire.

HPC, IA, EDGE

- **Certaines applications industrielles critiques exigent que les données restent sur le site de production**. D'autres pourraient être hébergées dans des datacenters, mais **nécessitent alors une latence inférieure à 5 ms**, ce qui implique de disposer de

datacenters de proximité, situés à moins de 300 km.

- **Le modèle économique de l'IA générative dans l'industrie reste incertain**, tant que les cas d'usage, les niveaux de performance attendus et les taux d'erreur acceptables ne sont pas clairement stabilisés.

- **Un nouveau paradigme de calcul distribué émerge**, fondé sur la migration dynamique d'agents d'IA vers les lieux où se trouvent les données, afin de mutualiser et d'allouer les ressources de calcul de manière plus efficiente (« fluid computing »).

- **Le déploiement à grande échelle des infrastructures EDGE (IoT, véhicules autonomes) pose un défi majeur de financement**, les modèles économiques associés n'étant pas encore clairement établis.

HARDWARE & STOCKAGE

- **Les technologies de stockage basées sur l'ADN pourraient, à long terme, réduire drastiquement les coûts** et offrir des durées de conservation extrêmement longues pour les données dites « froides ».

- **Les architectures matérielles deviennent de plus en plus hétérogènes**, combinant CPU, GPU, TPU, QPU, ainsi que de nouvelles hiérarchies mémoire (CXL), afin de dépasser les limites actuelles liées au « memory wall » de l'IA.

- **L'inférence devrait constituer le principal moteur de croissance du marché de l'IA** avec l'apparition d'architectures de calcul spécialisées (ex. Cerebras).

ÉNERGIE & REFROIDISSEMENT

- **En France, la consommation d'eau de la filière datacenter reste faible par rapport à d'autres secteurs industriels**, et n'est pas critique dès lors que des technologies de refroidissement sobres en eau sont privilégiées.

- **L'accès à l'énergie et les délais de raccordement constituent le principal facteur limitant du développement des datacenters**, la problématique française étant moins la production que la capacité et la rapidité de raccordement et les longueurs des procédures administratives.

- **Les coûts d'investissement et d'énergie liés au HPC deviennent un enjeu stratégique y compris pour les grands groupes industriels**, freinant la capacité à suivre la croissance rapide des besoins de calcul.

- **Un amortissement progressif de la consommation énergétique des datacenters est envisagé à horizon 10 ans**, avec des taux de croissance plus modérés que ceux observés au cours des trois dernières années.

- **Les datacenters pourraient être incités à devenir plus flexibles sur le plan énergétique** (effacement, modulation de charge), bien que cette logique soit en tension avec leur modèle historique de redondance et de disponibilité maximale.

MODÈLES DE DATACENTERS

- **La demande reste très orientée vers les grands datacenters et les hyperscalers** principalement financés par les fonds d'investissement et les grandes entreprises technologiques, tandis

que le marché des datacenters EDGE de petite et moyenne taille demeure de niche actuellement.

- **Il y a encore une part d'incertitude pour les 5 ans à venir sur le marché des datacenters concernant l'IA qui reste encore à la marge en France actuellement** ; comme dans les années 2000, dans le secteur des Datacenters, le risque de bulle existe et une décélération peut advenir du fait du non remplissage des datacenters actuels.

- **Le recours croissant aux jumeaux numériques entraînera des besoins importants en puissance de calcul**, dont une part significative sera probablement externalisée vers le cloud.

- **Les acteurs du minage de cryptomonnaies reconvertissent progressivement leurs infrastructures vers le HPC pour l'IA**, tout en offrant des capacités d'effacement et d'arbitrage énergétique lors des périodes de surproduction aux USA (ex. ERCOT au Texas).

SOUVERAINETÉ & RÉSILIENCE

- **La notion d'autonomie stratégique est plus pertinente que celle de souveraineté technologique**, l'indépendance totale étant irréaliste ; dans ce cadre, le ministère des Armées entretient des relations avec les GAFAM, dont la nature varie en fonction des usages et des enjeux de souveraineté, tout en développant des solutions propres avec des fournisseurs français et européens. Les acteurs du domaine de la défense et de la sécurité sont particulièrement attentifs à la recherche d'un juste équilibre

- **Les technologies duales, à usage civil et militaire**, constituent des leviers majeurs de **résilience industrielle, technologique et stratégique**.

- **Le futur reposera sur des infrastructures énergétiques suffisantes, modulaires et résilientes**, couplées à des **réseaux de communication redondants et autonomes**, intégrant notamment des solutions satellitaires.

SOMMAIRE

REMERCIEMENTS	3
AVIS AUX LECTEURS	4
PRÉFACE	5
EXECUTIVE SUMMARY	6
MESSAGES CLÉS DES PARTENAIRES	7
INTRODUCTION	10
ACRONYMES ET DÉFINITIONS	11
MÉTHODE PROSPECTIVE.....	13
VARIABLES CLÉS	14
GÉOPOLITIQUE	14
DATACENTERS	15
ÉNERGIE.....	17
EAU ET REFROIDISSEMENT	21
IA, HPC ET EDG.....	24
HARDWARE ET MÉMOIRES	27
RÉSEAUX & TÉLÉCOMMUNICATIONS	28
MINAGE ET BLOCKCHAIN.....	29

HYPOTHÈSES QUANTITATIVES & SCÉNARIOS.....	32
SCÉNARIO « CROISSANCE ACCÉLÉRÉE »	34
SCÉNARIO « INSTABILITÉ ET CYCLES TECHNOLOGIQUES».....	38
SCÉNARIO « NOUVELLE DONNE NUMÉRIQUE EUROPÉENNE»	42
SCÉNARIO « RÉSILIENCE ET SOBRIÉTÉ».....	46
CONCLUSIONS.....	51
RECOMMANDATIONS	56
ANNEXE	58

INTRODUCTION



L'infrastructure de la donnée numérique est à la croisée des chemins. L'émergence fulgurante de l'intelligence artificielle générative, couplée à une prise de conscience planétaire des enjeux énergétiques, climatiques et souverains, place le datacenter **au cœur d'une tension stratégique** sans précédent. Il n'est plus un simple «entrepôt» numérique passif ; il est devenu l'usine de production de valeur, **l'infrastructure critique** dont dépendront la compétitivité économique et l'autonomie numérique stratégique des nations.

PÉRIMÈTRE ET AMBITION DE L'ÉTUDE

C'est dans ce contexte de transformation radicale que nous avons initié l'étude prospective «DATA 2040». Se projeter à un horizon de quinze ans, dans un domaine où les cycles technologiques et les bouleversements géopolitiques se comptent en mois, peut sembler un pari. C'est pourtant nécessaire pour éviter de « subir » l'avenir.

La méthode prospective permet **d'éclairer les choix stratégiques d'aujourd'hui** en identifiant les leviers de résilience, les zones de vulnérabilité et les espaces d'innovation à long terme. Elle ne fournit pas une «roadmap» technologique prédictive, ni ne fige une trajectoire unique. Mais elle **explore les futurs possibles en fonction de différentes hypothèses d'évolution** (géopolitiques, économiques, technologiques, environnementales et sociétales).

L'ambition de **DATA 2040** est de **prendre de la hauteur**, d'éviter les visions monolithiques qui réduisent le futur à une trajectoire unique et d'ouvrir le champ des possibles. Les futurs des datacenters et de la chaîne du numérique dans son ensemble ne sont ni certains, ni linéaires : ils dépendent d'interactions complexes entre technologies, politiques énergétiques, contraintes environnementales, modèles économiques, usages sociétaux.

Reste à la charge de **chaque partenaire de s'approprier les résultats**, confronter les hypothèses aux réalités de son métier et identifier les scénarios ou tendances qui méritent d'être approfondis dans **ses propres plans d'action**.

Le périmètre géographique de notre analyse est mondial, bien qu'avec un tropisme occidental naturel, l'objectif étant d'avoir une **vision globale**. Car nous n'avons pas eu l'ambition de fournir une étude exhaustive sur l'avenir du datacenter dans chaque pays du monde. Le contexte de chacun est unique : tensions sur les ressources, mix énergétique et stabilité de ses réseaux électriques, cadres réglementaires et acceptabilité sociétale ou capacités de financement de l'innovation façonnent autant d'avenirs distincts. Nous n'avons pas non plus cherché à bâtir des modèles quantitatifs de consommation (énergie, eau, carbone). Notre but est **d'identifier les dynamiques d'innovation et les forces de tension structurantes** qui sont à l'œuvre, afin d'imaginer les grandes typologies de datacenters et les modèles d'infrastructures

qui pourraient en résulter. Quand cela s'est avéré pertinent et possible, un **focus particulier sur le cas français** a été réalisé.

DÉMARCHE COLLABORATIVE & PLURIDISCIPLINAIRE

La démarche **DATA 2040** repose sur la **co-construction** : elle a mobilisé **pendant douze mois** un collectif d'acteurs issus de secteurs variés – télécommunications, énergie, défense, banque, transport, numérique, bureaux d'études, opérateurs de datacenters, institutions publiques, etc.

Cette **diversité de points de vue** a permis d'enrichir les analyses et de limiter les biais sectoriels. Au-delà de la production intellectuelle, ce travail a également permis de renforcer les liens entre partenaires dans un cadre de confiance, sécurisé par un accord **de confidentialité**, propice à l'échange ouvert et au partage d'informations sensibles.

Le **CEA** a coordonné cette démarche en s'appuyant sur l'expertise du **cabinet GERPA**, reconnu pour ses travaux dans le champ de la prospective stratégique. Elle s'appuie sur la méthodologie de la « prospective à la française » enseignée au Conservatoire national des arts et métiers, initiée par Gaston Berger, orientée vers la décision et l'action pour l'innovation et la transformation.

ACRONYMES ET DÉFINITIONS

ACV

Analyse du Cycle de Vie
Évaluation environnementale des bâtiments

AI

Artificial Intelligence (Intelligence Artificielle)
On distingue 2 phases : phase d'apprentissage pour développer les modèles d'IA et la phase d'inférence qui est celle d'exécution.

AIE

Association Internationale de l'Énergie

B2B

Business-to-Business
Transactions entre entreprises.

B2C

Business-to-Consumer
Transactions entre entreprises et consommateurs.

BESS

Battery Energy Storage System
Utilisation de batteries pour alimenter les datacenter pendant des coupures de courant ou comme source de flexibilité

CAPEX

Capital Expenditure (Dépenses d'investissement)
Coût moyen ≈ 10 M€/MW pour un datacenter classique.

CXL

Compute Express Link
Protocole d'interconnexion haute performance pour relier processeurs et mémoire.

CPU

Central Processing Unit (Processeur central)
Souvent comparé aux GPU pour les charges AI.

DC

Data Center (Centre de données)

DCIM

Data Center Infrastructure Management (Gestion des infrastructures de centres de données)

DLC

Direct Liquid Cooling (Refroidissement liquide direct)
70 % du refroidissement du site ECLAIRION, capacité à délivrer de l'eau à ~50 °C.

EB

ExaByte 1012 octets

ENR

Énergies Renouvelables

EOP

European Operations (Opérations européennes)

EU

European Union (Union européenne)
Contexte des politiques de souveraineté et du European Chips Act.

ENISA

European Union Agency for Cybersecurity (Agence

européenne pour la cybersécurité)
Publication « Securing Machine Learning Algorithms » (2024).

GAFAM

Google, Apple, Facebook (Meta), Amazon, Microsoft
Les cinq grands acteurs technologiques américains.

GPU

Graphics Processing Unit (Unité de traitement graphique)
Principale unité de calcul pour les charges AI. Prévisions de « GPU-years » (ex. 300 000 GPU-years d'entraînement en 2025).

HPC

High-Performance Computing (Calcul haute performance)

IA

Intelligence Artificielle (voir AI)

IAM

Identity and Access Management (Gestion des identités et des accès)

IEA

International Energy Agency (Agence internationale de l'énergie)
Cité dans le rapport « Energy-AI synergy » (2025).

IOT

Internet des Objets (réseau d'objets connectés)

IXP

Internet eXchange Point

Infrastructure physique permettant aux fournisseurs d'accès et de contenus d'interconnecter leurs réseaux localement pour échanger du trafic de manière plus rapide, directe et économique

LLM

Large Language Model (Modèle de langage à grande échelle)

MW

Megawatt (Unité de puissance, 10^6 W)

Puissance installée des sites français (ex. Thales 30 MW) et coûts d'infrastructure (≈ 10 M€/MW).

NFV

Network Functions Virtualization

NFV consiste à virtualiser les fonctions réseau traditionnellement assurées par des équipements matériels dédiés

OCP

Open Compute Project (Initiative open-source pour la standardisation du matériel datacenter)

Référence à la spécification Open-Rack Base Spec v3.

OS

Operating System

Système d'exploitation des ordinateurs

OTT

Over-The-Top (Services de streaming contournant les opérateurs traditionnels)

PUE

Power Usage Effectiveness (Ratio de l'énergie totale consommée par le datacenter sur l'énergie utilisée par les équipements IT)

Cibles 2025 des hyperscalers (ex. Microsoft 1.05, Google 1.07, moyenne ≈ 1.10).

OPEX

Operational expenditure

les OPEX désignent les dépenses opérationnelles d'une entreprise, à savoir ce qu'elle dépense au quotidien pour faire fonctionner son activité

ORAN / VRAN

Open Radio Network / Virtual Radio Network

oRAN : Architecture de réseau mobile qui désagrège le matériel et le logiciel en utilisant des interfaces ouvertes et standardisées pour permettre l'interopérabilité entre différents fournisseurs. / vRAN : permet aux opérateurs de télécommunications d'exécuter leurs fonctions de bande de base en tant que logiciel

RISC-V

Reduced Instruction Set Computer – V (Architecture de jeu d'instructions ouverte)

Mentionnée dans le projet « RISC-V European Processor Initiative » pour la souveraineté des CPU/GPU.

SD-POWER

Software-Defined Power (Gestion logicielle dynamique de la puissance électrique)

Identifié comme « game-changer » pour le edge computing.

SLM

Small Language Model (Modèle de langage plus

petit)

SMR

Small Modular Reactor (Réacteur nucléaire modulaire de petite taille)

Étudié comme solution de secours de 300 MW (≈ 1 MM\$).

SDN

Software Define Network

SDN désigne une architecture de réseau pilotée par logiciel

TPU

Tensor Processing Unit (Unité de traitement tensoriel, Google)

UCIX

Universal Chiplet Interconnect Express (Standard d'interconnexion pour puces modulaires)

XR

eXtended Reality

Réalité augmentée, dont l'usage va être développé par les casques ou lunettes

ZERO TRUST

Modèle de sécurité informatique basé sur la méfiance par défaut

MÉTHODE PROSPECTIVE

Avec l'aide du cabinet de conseil GERPA¹ (Groupe d'Études Ressources Prospective en Action), nous avons adopté la méthode des scénarios par **l'analyse morphologique** pour construire nos scénarios.

Classiquement, une première étape a permis de décrire le **système prospectif** avec ses **principaux déterminants**. Rythmés par des interventions d'experts, visites de sites et ateliers de travail, des **cahiers prospectifs** ont permis de décrire quelles étaient les **tendances lourdes, émergentes, et aussi les signaux faibles et ruptures possibles** pour

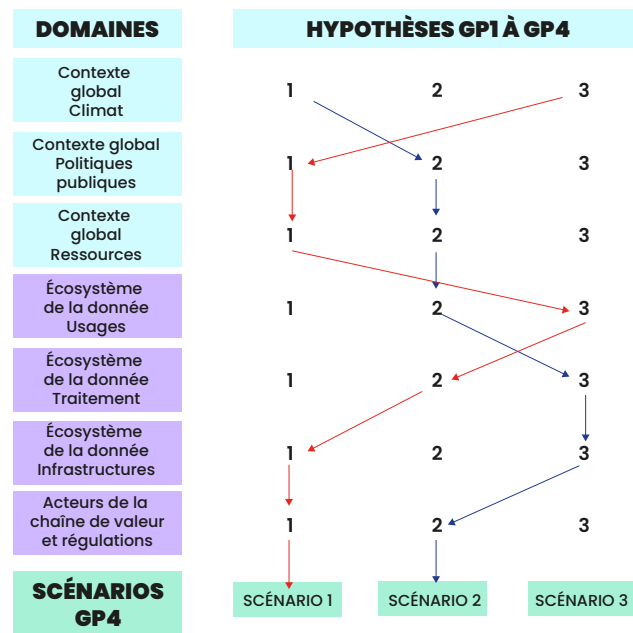


Figure 1 : Sélection des hypothèses dans la méthode des scénarios morphologiques

chaque déterminant.

Dans un deuxième temps, des **hypothèses construites et validées par les experts** du domaine ont été formulées, nous permettant de construire des scénarios globaux par analyse morphologique (Figure 1).

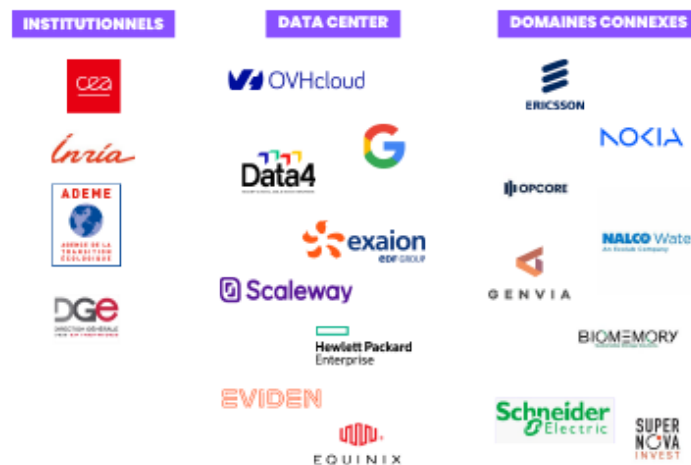


Figure 2 : Partenaires de l'étude DATA 2040

Nous avons confronté à cette étape les différents scénarios au point de vue d'un ensemble d'experts extérieurs au groupe de travail pour les enrichir et en identifier les limites. La constitution de ce groupe, appelé groupe « miroir » comptait les organisations de la Figure 2).

Ce travail **d'enrichissement** nous a permis aussi de repérer les **variables clés** et **approfondir les hypothèses**, dont **certaines ont pu être quantifiées**.

Nos scénarios ont pu être affinés et segmentés à l'aune de ces différentes hypothèses pour arriver à une **synthèse de l'étude et des recommandations**.

Un **rapport confidentiel** a été livré aux différents partenaires qui ont contribué à l'étude en plus de ce rapport public. Tout au long de cette étude, un processus itératif a permis d'affiner les différentes étapes pour réinjecter les apports des partenaires extérieurs ou des dernières informations de la veille et de l'actualité (Figure 3).

MÉTHODE PROSPECTIVE : UN PROCESSUS ITÉRATIF



Figure 3 : Processus global de la méthode prospective appliquée à DATA 2040

¹ <https://www.ressources-prospective.com/gerpa/>

VARIABLES CLÉS

GÉOPOLITIQUE

DESCRIPTION & ENJEUX

En 2025, le système international évolue vers un **ordre multipolaire fragmenté**, dominé par la **fin du leadership stabilisateur américain**, le retour du **protectionnisme et l'affaiblissement durable des institutions multilatérales**². Les États-Unis privilégient une logique nationaliste et coercitive, alimentant une guerre commerciale et normative qui **accélère la recomposition des alliances**. Dans ce contexte, la **Chine s'affirme comme une puissance structurante**³, promouvant un ordre fondé sur la souveraineté étatique, la projection économique et la maîtrise technologique (semi-conducteurs, données, métaux critiques, IA), tandis que **l'Inde adopte une stratégie de multi-alignement pragmatique**, cherchant à **tirer parti de la multipolarité sans rompre avec l'Occident**. L'Union Européenne, sous pression économique et stratégique, tente de préserver sa cohésion tout en renforçant son autonomie⁴. La **transition écologique** et les politiques technologiques **se fragmentent fortement** entre régions, faisant de **l'énergie, des infrastructures numériques et de l'IA** des champs centraux de **rivalités géopolitiques**.

MEGA TRENDS

• **Rattrapage technologique et découplage stratégique** : La Chine, l'Inde et plusieurs pays du Sud global accélèrent leur rattrapage technologique (IA, semi-conducteurs, énergie), entraînant une **régionalisation des chaînes de valeur**, une montée de **standards concurrents** et un usage croissant de la technologie comme levier de souveraineté.

• **Déclin du multilatéralisme et alliances de circonstance**⁵:

2 <https://www.chathamhouse.org/2025/03/competing-visions-international-order>

3 <https://www.geopolitics.fr/2025/07/16/brics-sud-global-et-defi-a-l-occident-vers-un-nouvel-equilibre-mondial/>

4 <https://www.bruegel.org/report/geopolitical-shifts-and-their-economic-impacts-europe-short-term-risks-medium-term-scenarios>

5 https://www.lemonde.fr/idees/article/2025/09/03/face-a-ce-sud-global-domine-par-la-chine-qui-n-existe-pas-l-occident-sous-la-ferule-de-donald-trump-existe-de-moins-en-moins_6638760_3232.html

Les grandes institutions internationales s'affaiblissent au profit d'**alliances opportunistes, bilatérales ou régionales**, fondées sur des intérêts économiques, énergétiques et sécuritaires immédiats plutôt que sur des règles communes stables.

- **Géopolitisation des ressources critiques** : L'accès à l'énergie, au calcul, aux données et aux matériaux critiques devient un facteur central de puissance, plaçant les infrastructures numériques et énergétiques au cœur des stratégies géopolitiques et industrielles.

SIGNAUX FAIBLES

- **Rattrapage technologique chinois plus rapide qu'anticipé** : Malgré les sanctions, la Chine progresse rapidement en semi-conducteurs et en IA via l'ingénierie inverse, l'optimisation logicielle et l'investissements massifs, réduisant l'écart avec les États-Unis, notamment sur l'inférence et les usages industriels et grand public (industrie 4.0, mobilité, robotique, e-commerce, etc.)

- **Fragmentation normative** et divergence des transitions : Les modèles de transition écologique, numérique et industrielle divergent fortement entre États-Unis, Europe, Chine et pays émergents, générant une fragmentation des normes (climat, IA, données) et des trajectoires de développement contrastées.

- **Montée des pays du Golfe dans le secteur des hubs IA et des datacenters** : Les États du Golfe investissent massivement dans l'IA, le cloud et les datacenters pour transformer leur rente énergétique en puissance numérique régionale et pour attirer hyperscalers, talents et capitaux.

- **Multi-alignement des puissances intermédiaires** : Des pays comme l'Inde, L'Asie du Sud-Est ou le Brésil tirent parti du découplage sino-américain pour capter investissements et technologies, sans alignement exclusif sur un bloc.

RUPTURES

- **Attaque des pays baltes par la Russie** sans soutien militaire direct des États-Unis : une **action militaire ou hybride russe** (incursion, déstabilisation armée, occupation partielle) vise un ou plusieurs pays baltes, sans engagement militaire direct des États-Unis en réponse malgré les accords de l'OTAN.

- **Avènement d'un « splinternet »** et dissociation technologique durable des blocs : la fragmentation d'Internet et des écosystèmes numériques devient structurelle, avec des blocs technologiques largement incompatibles (standards, clouds, IA, données, cybersécurité).

6 <https://www.ndc.nato.int/fr/what-if-12-dragon-king-scenarios-for-2028/>

- **Choc systémique sur les chaînes de valeur des semi-conducteurs : Blocus autour de Taiwan**, prise de contrôle partielle ou restriction d'exportations de TSMC⁶.

- Crises internes majeures dans les grandes puissances : **Implosion politique ou sociale aux États-Unis ou recomposition politique de l'Union européenne**.

DATACENTERS

TYPLOGIES ET MODÈLES

Il n'y a **pas vraiment de définition officielle** de ce qu'est un datacenter. D'une salle informatique dédiée ou un bâtiment construit à dessein, il existe **différents modes de classement** des datacenters. On peut les catégoriser selon leur puissance électrique, leur surface ou bien leur offre de service. Dans cette étude nous avons choisi de les classer selon **trois grandes typologies : hyperscaler, colocation et Edge** (Figure 4).

	SINGLE TENANT	COLOCATION DATACENTER			SINGLE TENANT
TYPE	PRIVATE / ENTREPRISE <i>Including server space</i>	REGIONAL	NATIONAL	INTERNATIONAL	HYPERSCALE
CUSTOMERS	SME Enterprise Public Semi-public	SME Public Semi-public	SME Enterprise Cloud Public Semi-public	SME Enterprise Cloud SaaS	Cloud SaaS
SPACE Impact	>10 m ² <i>Small</i>	>200 m ² <i>Small</i>	>200 m ² <i>Small</i>	>5000 m ² <i>Medium</i>	5 ha <i>Medium/Large</i>
ENERGY Impact	0,01 – 10 MW <i>Small</i>	0,5 – 10 MW <i>Small</i>	1 – 10 MW <i>Medium</i>	> 5 MW <i>Medium</i>	50 MW <i>Large</i>

Figure 4 : Différentes typologies de datacenter- Source : European Data center overview Oct. 2024

Ce que l'on désigne par hyperscalers sont des datacenters qui peuvent aller jusqu'à des puissances électriques de plusieurs Gigawatts et pour le compte unique de grands acteurs du digital (Google, AWS...).

La catégorie des datacenters en colocation se caractérise par des dimensions et consommations inférieures à celles des hyperscalers et ont la caractéristique d'être partagés par plusieurs utilisateurs.

La dernière catégorie est ce que l'on appelle les datacenters d'entreprise (ou parfois « On-Premise »), ils sont les plus petits et possédés souvent par des entreprises pour leur besoin spécifiques.

Les **États-Unis** dominent très largement le marché avec plus de **45 % du total mondial**⁷ mais la région Asie-Pacifique (APAC) affiche le **taux de croissance le plus élevé au monde**. L'Union Européenne est le second marché derrière les USA avec plus de 2 000 datacenters publics (et +500 au UK) mais il est largement dominé par les acteurs américains. En Europe, ce marché est concentré sur les villes FLAP-D (Francfort, Londres, Amsterdam, Paris, Dublin). Les hyperscalers consolident leur domination par une **forte verticalisation, intégrant infrastructures, plateformes logicielles et services** finaux. Cette maîtrise de bout en bout réduit leurs coûts et crée un verrouillage des clients, rendant les migrations complexes et coûteuses (« *vendor lock-in* »).

MEGATRENDS & ÉVOLUTIONS

On assiste depuis de nombreuses années à la **migration de données des utilisateurs vers les services du cloud** qui permettent de simplifier et de diminuer les coûts de gestion. Les « clouders » ne vendent pas seulement de l'infrastructure, mais une **solution métier complète** répondant à des besoins spécifiques et sont en train de devenir le **système d'exploitation central** de l'économie

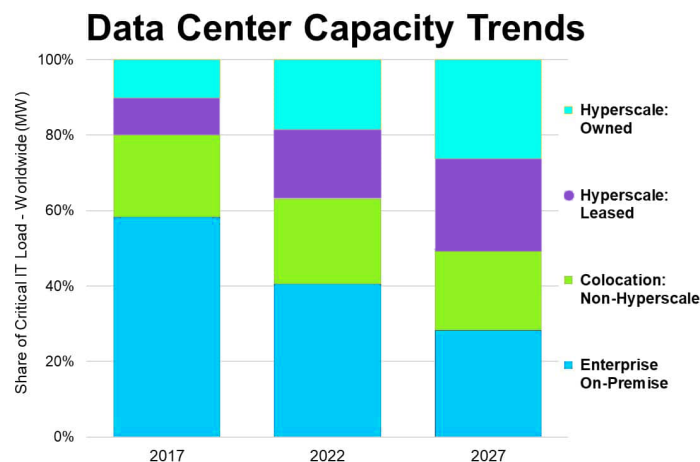


Figure 5 : Évolution des différentes typologies de datacenter - Source : Synergy Research Group

7 <https://www.cargoson.com/fr/blog/nombre-de-centres-de-donnees-par-pays>

8 <https://analysesetdonnees.rte-france.com/bilan-electrique-2024/synthese>

9 <https://systeme.developpez.com/actu/363400/La-hausse-de-prix-extreme-proposee-par-Broadcom-augmenterait-les-couts-de-VMware-de-1050-pourcent-et-Broadcom-empecherait-certains-fournisseurs-de-nous-vendre-des-produits-selon-AT-T/>

numérique sur lesquels les autres acteurs (FAI, éditeurs de logiciels) font tourner leurs applications. Même si, en nombre, les hyperscalers ne constituent pas la majorité des datacenters, ce sont eux **qui tirent le marché actuellement avec le développement de l'IA générative**. En termes de charge de travail, cette dynamique continuerait à croître selon certaines études (Figure 5).

En Europe, ce **sont les pays nordiques qui affichent les plus fortes croissances**, suivis par les pays de l'Europe du Sud (Figure 6).

La France dispose d'atouts majeurs : un mix électrique largement décarboné et excédentaire (10GW en 2024⁸), un réseau performant et une position géographique stratégique en Europe. En contrepartie, la complexité réglementaire et la longueur des procédures constituent un relatif frein comparé à certains pays (pays nordiques).

SIGNAUX FAIBLES

- **Le retour du « on-premise »** : Devant des prix de licences qui augmentent et des difficultés à maîtriser les coûts de gestion, on assiste à un mouvement des entreprises vers des solutions « on Premise ». Le mouvement a été amplifié par **le rachat de VMWARE par BROADCOM**⁹ en novembre 2023 qui a augmenté ses tarifs et modifié ses licences d'accès.

- **Utilisation de friches industrielles** : Reconversion de sites existants pour limiter l'artificialisation, bénéficier d'infrastructures électriques déjà présentes et

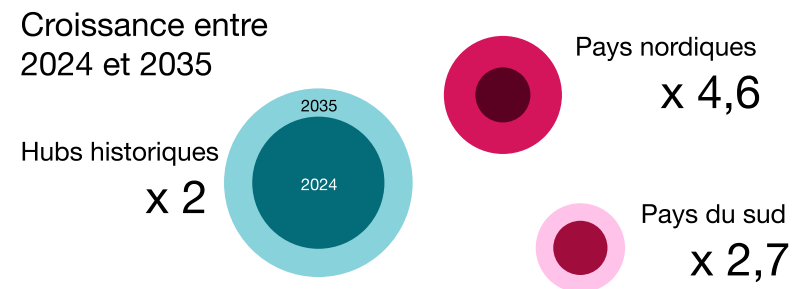


Figure 6 : Tendence d'évolution des datacenter en Europe - Source : Transition Data Lab - Le futur du cloud Européen à l'épreuve de l'énergie et du climat

accélérer les autorisations, tout en améliorant l'acceptabilité locale¹⁰.

• **Datacenters modulaires** : Déploiement par modules préfabriqués pour réduire le CAPEX, gagner en flexibilité et adapter rapidement les capacités à la demande, notamment pour l'edge et les usages incertains.

RUPTURES

• **Datacenters en milieux extrêmes** : Implantation dans des environnements contraints : sous-marins, spatiaux ou théâtres d'opérations militaires. De plus en plus de projets naissent pour développer des datacenters en conditions durcies pour profiter de conditions naturelles favorables ou pour rapprocher le calcul des zones d'opération. Certains satellites Starlink testent des capacités d'IA embarquée et de traitement local, inaugurant le concept de "**space edge computing**" : une distribution du calcul entre l'orbite et la Terre qui réduit la latence et prépare l'**émergence d'un Internet spatial distribué**¹¹.

• **Micro-datacenters intégrés** : La miniaturisation et la forte densification du matériel pourraient permettre l'intégration de capacités de calcul et de stockage directement dans des bâtiments tertiaires ou résidentiels, au plus près des usages, avec des logiques d'ultra-edge¹².

ÉNERGIE

DESCRIPTION & ENJEUX

Il est à noter que la **consommation électrique mondiale des datacenters en 2024, estimée à 415 TWh** par l'IEA¹³ (1,5% de la consommation mondiale), **est une estimation** et non une mesure exacte. L'absence de « reporting » obligatoire global rend les mesures directes impossibles, d'où le recours à des projections et modélisations. Par contre, il est clair que si, durant une dizaine d'années, la consommation des datacenters a stagné autour de 220 TWh de 2010 à 2020¹⁴, elle connaît une forte augmentation depuis cette date. C'est principalement dû

¹⁰ <https://www.entreprises.gouv.fr/la-direction-generale-des-entreprises/actualites/laureats-sites-cles-main-france-2030>

¹¹ <https://www.actia.com/ceos2030-une-solution-francaise-de-space-edge-computing/>

¹² <https://www.iledefrance.fr/toutes-les-actualites/qarnot-chauffe-les-batiments-avec-des-serveurs-informatiques>

¹³ <https://iea.blob.core.windows.net/assets/de9dea13-b07d-42c5-a398-d1b3ae17d866/EnergyandAI.pdf>

¹⁴ <https://informationssystemsbiology.blogspot.com/2025/01/2024-update-on-electricity-and-co2.html>

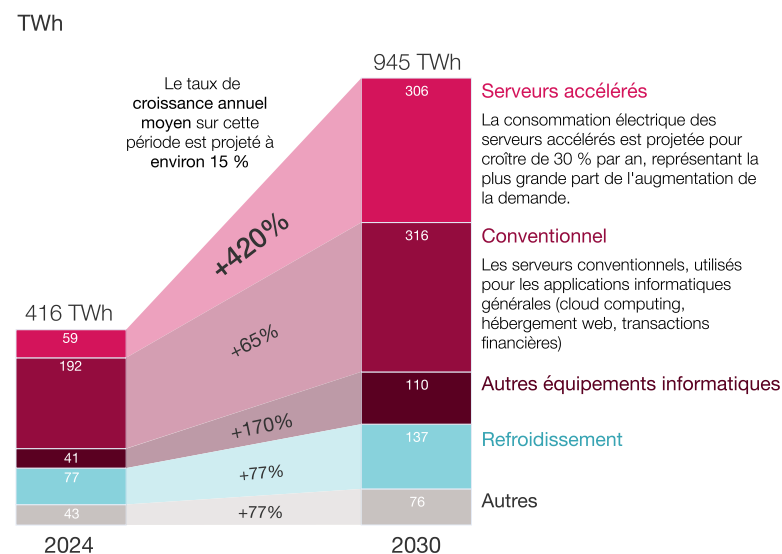
¹⁵ <https://ccaf.io/cbnsi/cbeci/comparisons>

à la **croissance des hyperscaler et de l'IA générative qui ont fait doubler les consommations** (Figure 7).

Il faut noter au passage la croissance de la **consommation liée au minage du bitcoin qui est passée de 70 TWh en 2020 à 191 TWh en 2025**¹⁵ (presque triplée).

Une croissance boostée par l'IA

Répartition et Évolution de la consommation mondiale d'électricité des centres de données, 2024/2030



Graphique : Transition Data Lab - Rémi Grimaud

Source : International Energy Agency, Energy and IA, 2025, Base Case Scénario

Figure 7 : Estimation de la croissance de consommation des datacenters liée à l'IA

Ces consommations de datacenters sont importantes à l'échelle mondiale mais à **relativiser également par rapport à d'autres secteurs énergétiques comme la chimie ou la métallurgie** qui représentent à eux deux près de 30% de la

consommation mondiale d'électricité¹⁶. Au niveau du mix énergétique, dans les années 2020, les hyperscalers prenaient des engagements d'une décarbonation totale des datacenters pour les prochaines années. Cependant ces dernières années, la compétition qu'ils se livrent **ne place plus les énergies renouvelables et décarbonées au centre de leur priorité. L'accès à l'énergie devient primordial**, bien avant la connexion à la fibre ou le prix du foncier, également en Europe¹⁷.

MEGATRENDS

S'il est déjà difficile d'estimer les consommations passées des datacenters dans le monde, il est encore plus difficile et incertain de se pencher sur les consommations futures. Il existe aujourd'hui une variété d'études et de modèles différents qui estiment des consommations des datacenters qui divergent d'un facteur 10 (Figure 8).

Ces estimations se basent sur **des modèles de croissances théoriques qui ne prennent pas en compte les aléas économiques ou de finitudes des ressources**. Dans le groupe de travail, nous croyons peu en un rythme de croissance exponentiel des consommations énergétiques des datacenters jusqu'en 2030.

Il y aura vraisemblablement un **ralentissement dans les 10 ans à venir**, grâce aux **progrès d'efficacité dans le hardware/software** qui permettront vraisemblablement de lisser cette croissance apparemment « débridée ». Nous assistons d'ailleurs à une **diminution des PUE qui tendent de plus en plus vers la valeur de 1** notamment chez les hyperscalers (Figure 9).

D'autres facteurs rentreront en jeu comme les **cycles économiques avec un ralentissement des investissements ou bien de la disponibilité de l'énergie disponible**. C'est déjà le cas de certains pays européens (Irlande, Hollande) qui concentrent historiquement les datacenters qui ont fait des **moratoires sur la construction de nouvelles infrastructures**. Des tensions apparaissent de plus en plus sur les réseaux électriques qui limitent cette progression, notamment dans la ville de **Dublin où 80% de l'électricité est utilisé pour les datacenters** (Figure 10).

Il convient aussi de relativiser les croissances électriques futures des datacenters par rapport à d'autres secteurs **comme la climatisation ou le chauffage des bâtiments qui ont des perspectives bien plus fortes** (Figure 11)

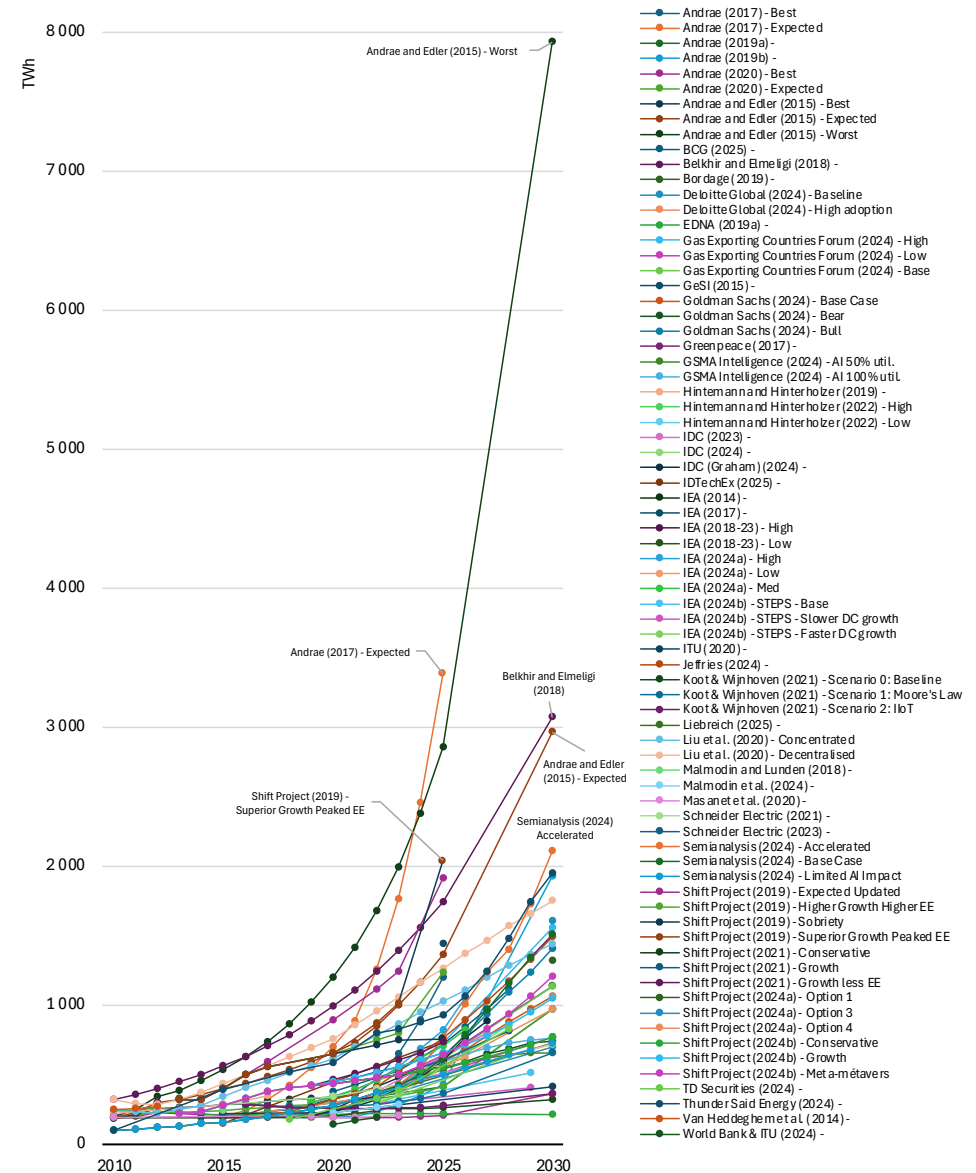


Figure 8 : Différents modèles de prévisions des consommations des datacenter au niveau mondial

16 <https://static1.squarespace.com/static/5877e86f9de4bb8bce72105c/t/64a6546bb2be6431f9a496bd/1688622201921/Global+Industrial+Electricity+Use+report-6.29.2023.pdf>

17 <https://www.investing.com/news/stock-market-news/is-renewables-role-in-powerhungry-data-centers-overstated-4224619>

TYPE D'ACTEUR	PUE 2008	PUE 2024	GAIN
HYPERSCALE (Google/Meta)	1,22-1,30	1,06-1,12	-15% à -20%
Cloud (AWS, Azure)	1,25-1,30	1,12-1,17	-10%
Colocation	1,8-2,0	1,4-1,6	-20%
Moyenne mondiale	1,65	1,55	-6%

Figure 9 : Évolution des Power Usage Effectiveness (PUE) des datacenter - Source : Tableau généré par Gemini 3.Pro

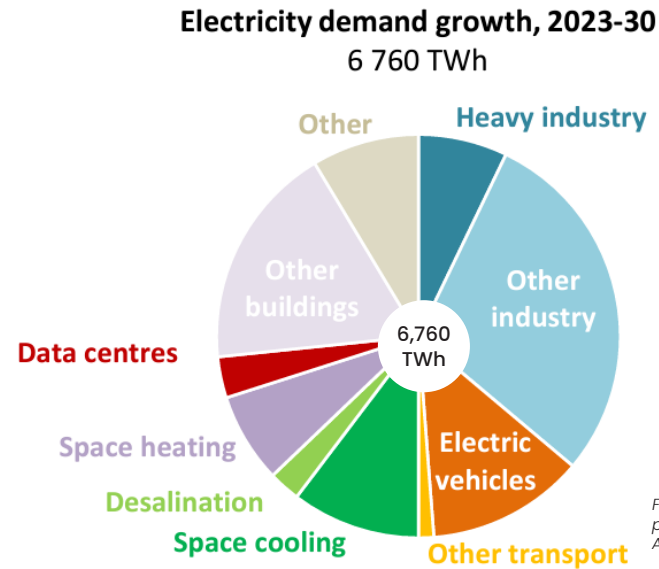


Figure 11 : Estimation des consommations des principaux secteurs en 2030- Source « Energy and AI » IEA Avril 2025

La saturation électrique de FLA-D

Contraintes de raccordement électrique dans les marchés FLAP-D

Londres

Les **raccordements peuvent prendre jusqu'à 13 ans**, poussant les développeurs à chercher des alternatives hors du Royaume-Uni.

Part de la demande d'électricité des centres de données à Londres en 2024

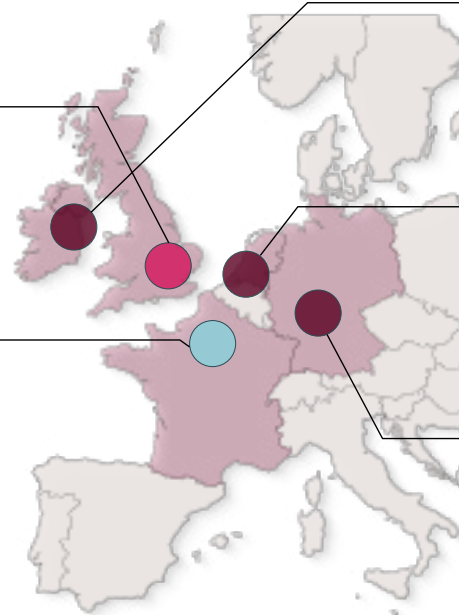
36%

Paris

La France reste la **seule place FLAP-D** où le **réseau électrique** demeure **relativement non contraint**, permettant la poursuite des investissements.

Une procédure de **raccordement accéléré** ("fast track") a été adoptée en mai 2025, pour permettre aux grands sites de se connecter d'ici **2028-2029**, consolidant la France comme hub de croissance européen.

Graphique : Transition Data Lab - Rémi Grimaud
Source : EMBER, Greenpeace



Dublin

L'opérateur EirGrid a imposé un **moratoire sur les nouveaux raccordements jusqu'en 2028, voire 2030**

Part de la demande d'électricité des centres de données à Dublin en 2024

80%

Amsterdam

Amsterdam est soumise à un **moratoire** pouvant durer jusqu'en **2035** et les Pays-Bas ont **interdit la construction** de nouveaux **hyperscalers** hors zones désignées,

Part de la demande d'électricité des centres de données à Amsterdam en 2024

33%

Francfort

Les autorités ont dû **bloquer toute nouvelle implantation** jusqu'à au moins **2030**, les projets devant attendre 7 à 10 ans, voire jusqu'à 13 ans pour un raccordement au réseau.

Part de la demande d'électricité des centres de données à Francfort en 2024

41%

Figure 10 : Part de l'électricité consommée par les datacenter dans les principales villes Européennes - Source : Transition Data Lab

SIGNAUX FAIBLES :

• **Tensions sur réseaux électriques** : l'accroissement des consommations des datacenters a déjà des impacts sur certaines villes comme Dublin et pourrait aussi augmenter les tensions sur les réseaux des pays nordiques, qui ont la dynamique d'implantation la plus élevée en Europe (Figure 12).

• **Apparition de datacenters « Off-grid »** : face aux contraintes sur la disponibilité électrique dans certaines régions du monde, des projets de solutions énergétiques « off-grid » voient le jour (ex : ECL à Houston Texas)¹⁸. La plupart du temps, ces solutions sont hybrides et permettent une connexion au réseau électrique centralisée.

• **Flexibilité énergétique des datacenters** : Face à l'explosion des besoins énergétiques, la **modulation de la charge de travail des datacenters** est une solution intéressante pour la stabilité des réseaux électriques, soumis à des perturbations liées aux énergies renouvelables. Différentes possibilités de flexibilité sont envisagées comme le **déplacement temporel de certaines charges numériques (batches, calculs différables) ou l'usage de batteries de stockage (BESS)**.

RUPTURES

• **SMR et fusion nucléaire** : De nombreux investissements ont été annoncés de la part des hyperscalers pour alimenter les datacenters avec des petits réacteurs nucléaires SMR (Small Modular Reactor). Des annonces optimistes sont faites pour les années à venir mais de **nombreux experts s'accordent à dire que la technologie ne sera mature qu'à partir de 2035**, et ce dans les pays qui auront la législation adéquate. De même un engouement pour la **fusion nucléaire** est palpable mais la mise sur le marché de telles solutions **n'est pas envisagée avant 2050**.

• **La transition vers des architectures de puissance en courant continu (DC) dans les datacenters** crée un lien direct avec **l'intégration des sources d'énergie renouvelable (ENR)** et des batteries, qui produisent naturellement du DC. En évitant les conversions successives AC<->DC des chaînes d'alimentation traditionnelles, des bus DC haute tension – potentiellement couplés à des **Solid-State Transformers (SST)** – permettent de

réduire significativement les pertes énergétiques, de diminuer la taille des modules et de **réduire la quantité de matériaux nécessaires**, notamment le cuivre, utilisé traditionnellement pour les transformateurs et câblages lourds. Cette évolution est rendue possible par les progrès récents des semi-conducteurs de puissance à large bande interdite, notamment le **carbure de silicium (SiC)**, qui autorise des commutations à plus haute fréquence, une

La montée des data centers accentue les tensions d'adéquation en Europe

Carte choroplèthe du LOLE (h/an) à l'horizon 2035 et cercles proportionnels représentant la consommation énergétique des data centers.

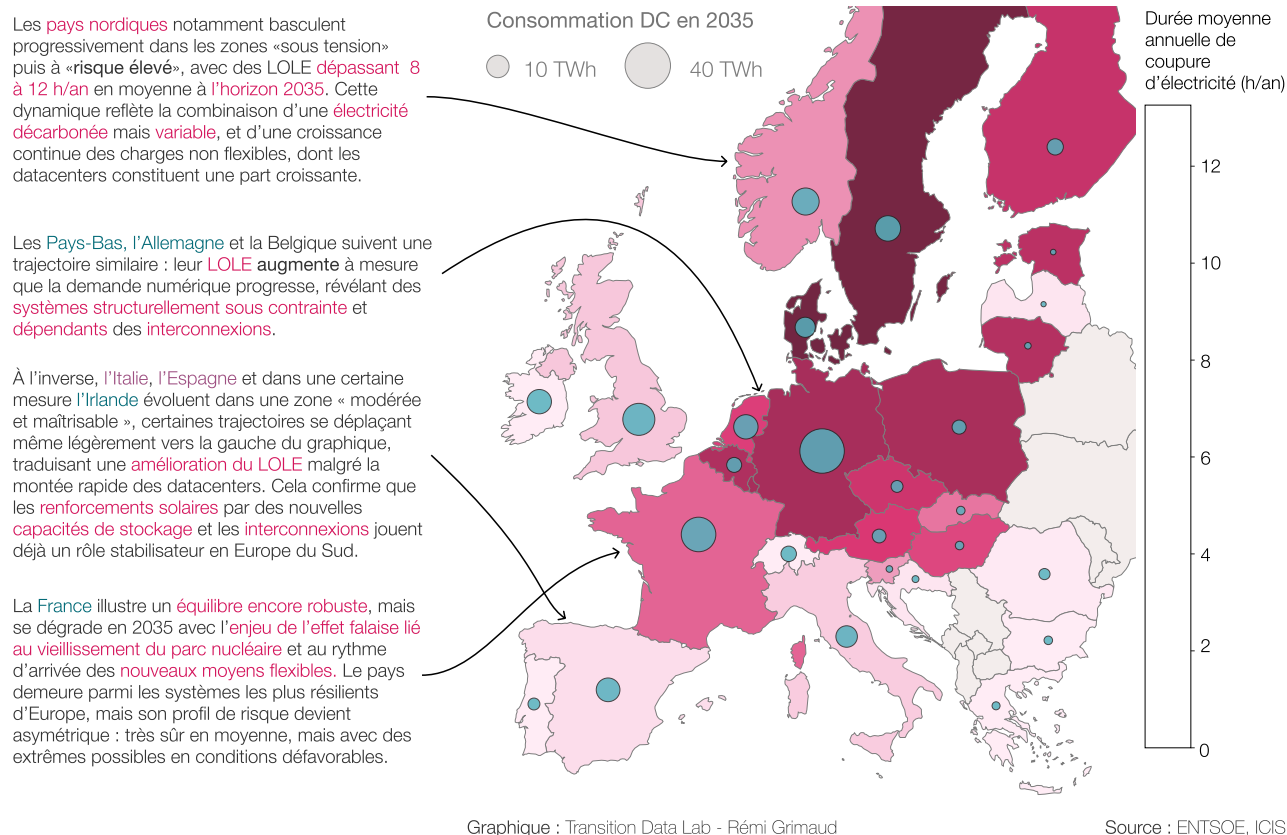


Figure 12 : Prévisions des tensions sur les réseaux électriques en Europe liée à la croissance des datacenter - Source : Transition Data Lab

¹⁸ <https://dcmag.fr/ecl-construit-le-premier-datacenter-dia-off-grid-de-1-gw/>

densité de puissance accrue et des rendements supérieurs¹⁹. Le SiC permet ainsi de concevoir des étages de conversion plus efficaces, plus compacts et mieux adaptés aux architectures DC haute tension (800 VDC et au-delà). Des projets concrets, comme le projet **DC Power en partenariat avec le CEA**²⁰, démontrent l'intégration opérationnelle de ces technologies pour alimenter directement des datacenters en DC depuis des sources renouvelables et des batteries, confirmant le potentiel de cette approche pour réduire les pertes et améliorer la résilience énergétique. La **convergence entre ENR, stockage batterie, distribution DC et électronique de puissance à base de SiC** constitue ainsi une **rupture dans la conception des architectures électriques des centres de calcul**, améliorant simultanément efficacité énergétique, résilience et capacité d'intégration aux micro-réseaux bas carbone.

EAU ET REFROIDISSEMENT

DESCRIPTION & ENJEUX

CHIFFRES CLÉS :

• Les **datacenters génèrent une chaleur substantielle** à partir de leurs équipements informatiques, **nécessitant des systèmes de refroidissement robustes** pour maintenir des températures de fonctionnement optimales et prévenir les dommages matériels. **L'eau est largement utilisée dans diverses technologies de refroidissement**, comme les tours de refroidissement par évaporation et les groupes froids. De nombreux datacenters **dépendent de l'eau potable en raison de sa propreté et de sa facilité d'utilisation**. Dans ce rapport nous parlerons uniquement de l'eau consommée et non de l'eau prélevée par les datacenters. Elle correspond à la part prélevée qui n'est pas restituée au milieu aquatique au même moment et au même endroit que le prélèvement. **Le refroidissement par évaporation, bien qu'économe en énergie, est gourmand en eau. Actuellement, la plupart des centres de données utilisent le refroidissement par air** et 22 % d'entre eux utilisent le refroidissement liquide, majoritairement dans les installations de grande taille et celles prenant en charge l'IA, qui nécessitent une puissance de calcul plus importante²¹. Il faut savoir que **le refroidissement représente généralement entre 30 et 40% des**

coûts d'exploitation²². **L'enjeu de l'eau pour les datacenters reste relativement modeste en volume à l'échelle nationale en France**, mais il est croissant et certains territoires en France peuvent subir des contraintes hydriques (ex : Marseille). De toute façon, l'implantation de datacenters en France fait l'objet de consultations avec les élus et les représentants de la société civile. Des solutions existent et sont déployées dans les cas où l'enjeu de la ressource en eau est primordiale pour les autres usages (particuliers, industries, etc.). On peut estimer la consommation d'eau (marge haute) en se basant sur les consommations d'électricité des datacenters. Sur une base de **10 TWh/an** comme **ordre de grandeur haut** pour la consommation électrique totale des infrastructures de calcul en France²³, si l'on applique à ce parc un **WUE direct moyen compris entre 0,3 et 0,5 L/kWh**²⁴, on arrive à une estimation de **3 à 5 millions de m³/an**. À l'échelle de la ressource nationale, ces chiffres restent **modestes comparés aux retraits agricoles et d'autres industriels**.

TECHNOLOGIES DE REFROIDISSEMENT :

• L'efficacité de ces systèmes est principalement évaluée à l'aide de deux indicateurs complémentaires : le **PUE**, qui **mesure l'efficacité énergétique globale du site**, et le **WUE** (Water Usage Effectiveness), qui **quantifie la consommation d'eau rapportée à l'énergie informatique délivrée**. Si l'amélioration du PUE est historiquement au cœur des stratégies des opérateurs, la maîtrise du WUE devient un enjeu critique. Un paradoxe structurel apparaît ainsi : **les technologies offrant les meilleurs rendements énergétiques sont souvent celles qui consomment le plus d'eau**, tandis que les **solutions sobres en eau présentent généralement une efficacité énergétique plus faible**. En exploitant l'évaporation de l'eau pour refroidir l'air, les systèmes adiabatiques comme les tours de refroidissement permettent d'atteindre d'excellents niveaux de PUE, mais au prix d'une consommation d'eau élevée. À l'inverse, les systèmes de climatisation reposant sur **des groupes froids affichent un WUE quasi nul** lorsqu'ils fonctionnent sans tours de refroidissement, mais avec un **PUE dégradé**, en raison de la forte consommation électrique des compresseurs. L'ajout de tours de refroidissement utilisées ponctuellement en France lors d'épisodes de canicules améliore l'efficacité énergétique, tout en augmentant significativement l'empreinte hydrique. Pour dépasser cette opposition, les architectures de refroidissement évoluent vers des **approches hybrides**, capables d'**adapter dynamiquement les technologies mobilisées en fonction**

¹⁹ <https://nfp-energie.ch/fr/projects/1008/>

²⁰ <https://www.dcpower.tech/about/project/>

²¹ <https://www.google.com/search?q=https://datacenter.uptimeinstitute.com/rs/711-RIA-145/images/2024.Cooling.Survey.Report.pdf>

²² <https://thenetworkinstallers.com/blog/data-center-operating-costs>

²³ <https://polenergie.org/ressources/actualites/datacenters-un-defi-energetique-qui-sintensifie/>

²⁴ Chiffres moyens WUE dans l'étude de l'ADEME de 2026 : <https://bibliothèque.ademe.fr/energies/8910-prospective-d-evolution-des-consommations-des-data-centers-a-court-moyen-et-long-terme-de-2024-a-2060.html>

des conditions climatiques et des charges informatiques. Le refroidissement moderne repose ainsi sur une double logique : capter la chaleur au plus près des composants et l'évacuer efficacement vers l'environnement. À l'intérieur des serveurs, le **refroidissement liquide direct**, via des plaques froides appliquées sur les composants les plus chauds, permet de capter la majorité de la chaleur dissipée, tout en conservant une part résiduelle de **refroidissement par air pour les composants secondaires**. Les **solutions par immersion**, où les serveurs sont entièrement plongés dans des fluides diélectriques, offrent une capacité de dissipation thermique encore supérieure. A noter que cela soit en DLC ou en immersion, l'utilisation de fluides diphasiques exploitant les changements d'état du fluide améliorent encore l'efficacité du refroidissement. Ces technologies sont particulièrement adaptées aux charges intensives de l'IA, mais leur performance environnementale dépend étroitement des modalités d'évacuation de la chaleur à l'extérieur du site. C'est à ce stade que se concentre l'essentiel de la consommation d'eau. Les architectures les plus sobres privilégient un fonctionnement en mode « sec », utilisant des aéroréfrigérants et l'air ambiant pour dissiper la chaleur, avec une consommation d'eau nulle et une bonne efficacité énergétique lorsque les températures extérieures le permettent. Le **free cooling** constitue un autre pilier des stratégies de refroidissement durable. Qu'il repose sur l'injection directe d'air extérieur, sur des échanges indirects via des fluides caloporteurs ou sur l'utilisation de sources naturelles d'eau froide, il permet de limiter, voire d'éliminer, le recours à la réfrigération mécanique. L'**exploitation de rivières, de nappes phréatiques, de lacs** ou de l'eau de mer offre des performances particulièrement élevées, avec des **PUE très bas et un WUE quasi nul**, sous réserve du respect de contraintes réglementaires strictes liées à la protection des milieux aquatiques. Déjà mise en œuvre dans **plusieurs datacenters européens et nordiques**, cette approche apparaît comme un levier structurant pour la durabilité à long terme du secteur, en particulier dans les régions tempérées à froides. Il faut noter qu'avec le changement climatique, ce type de refroidissement ne peut pas être le seul mis en place en France car les périodes où il ne fonctionnerait pas correctement auront tendance à s'allonger.

- Un des **enjeux cruciaux est** de réduire la consommation des systèmes informatiques pour réduire le besoin en refroidissement. La **valorisation des chaleurs fatales liées au refroidissement est également un enjeu depuis de nombreuses années. Une problématique majeure est la température sortie qui est trop faible pour être directement valorisée sans pompes à chaleur (maximum 55°C), que cela soit pour des process industriels ou bien pour des réseaux de chaleur (>90°C)**. Ce surcoût de Capex est un frein majeur à l'heure actuelle.

25 <https://www.iea.org/reports/energy-and-ai>

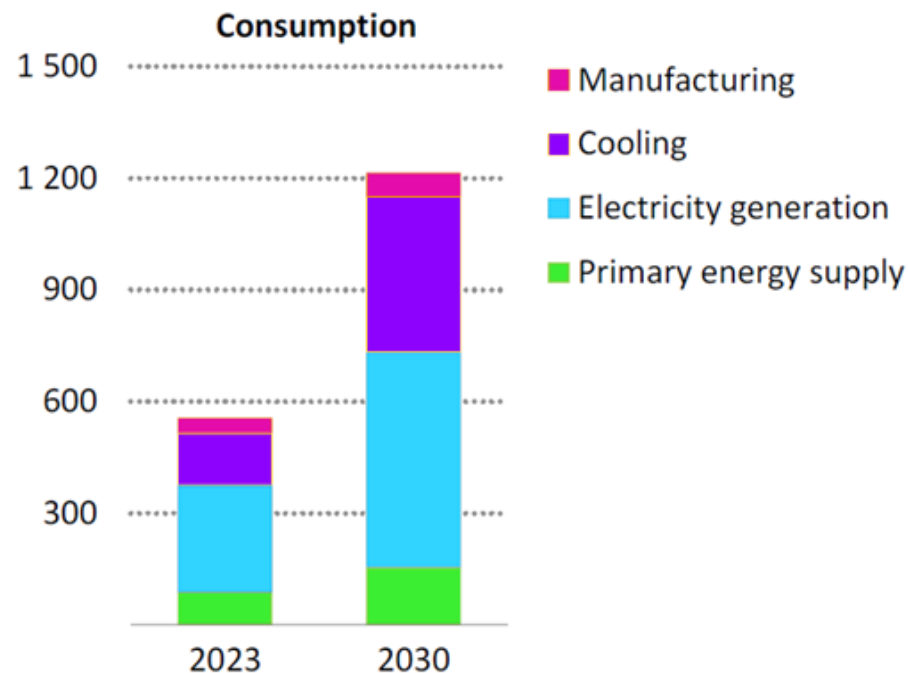


Figure 13 : Estimation des consommations d'eau des datacenters dans le monde – Source « Energy and AI » IEA Avril 2025

MEGATRENDS

- Basé sur une estimation de l'IEA, la consommation d'eau des datacenters pour le refroidissement (« cooling » dans la Figure 13) dans le monde pourra passer de **110 millions de m³ à 230 millions de m³ en 2030**²⁵.)

- **L'essor de l'Intelligence Artificielle (IA), avec ses processeurs à très haute densité thermique (GPU, TPU), rend les solutions de refroidissement traditionnelles par air insuffisantes et pousse à l'adoption massive du refroidissement liquide, voire diphasique (immersion et DLC). Autre intérêt de ces techniques, la possibilité d'un niveau de température de la chaleur évacuée plus élevé pour augmenter les possibilités de valorisation de la chaleur fatale.**

- Face à des réactions de la société civile de plus en plus fortes, les acteurs des datacenters travaillent sur des solutions pour la **diminution de la consommation d'eau**.

• La **valorisation de la chaleur fatale des datacenters** est un enjeu clé de la filière. Trouver des solutions pour valoriser de grandes quantités de chaleur basse température et surtout **pendant les périodes d'été est un challenge technologique majeur** de ces prochaines années.

SIGNAUX FAIBLES

• **Vers des WUE ambitieux en Europe** : la pression réglementaire et industrielle s'intensifie autour de la **sobriété hydrique des datacenters**, avec l'émergence d'objectifs explicites de **WUE (Water Usage Effectiveness)**. Le *Climate Neutral Data Centre Pact* engage ses signataires à viser un **WUE \leq 0,4 L/kWh IT en zones de stress hydrique**, préfigurant une possible normalisation ou obligation réglementaire à moyen terme²⁶.

• **Eaux grises et traitement sur site** : AWS s'est engagé à devenir « **water positive** » d'ici 2030, en restituant plus d'eau aux bassins versants qu'il n'en consomme. Cette stratégie repose notamment sur l'usage croissant d'**eaux usées traitées (« reclaimed water »)**, le **traitement et recyclage de l'eau sur site**, et l'optimisation des systèmes de refroidissement pour réduire les prélèvements d'eau potable²⁷.

• **Datacenters et stress hydrique** : en Chine, certaines régions peuvent connaître de fortes contraintes locales²⁸. Dans des scénarios de forte densité énergétique, la pression sur la ressource en eau pourrait inciter à privilégier des **architectures «water minimal»**, à boucle fermée, ou valorisant l'eau non potable et les sources naturelles froides. D'après l'étude de Transition Data Lab, avec le changement climatique, certains modèles prévoient

Europe 2030–2080 : les **datacenters** basculent vers des zones de **stress hydrique élevé**
Exposition des datacenters européens au stress hydrique

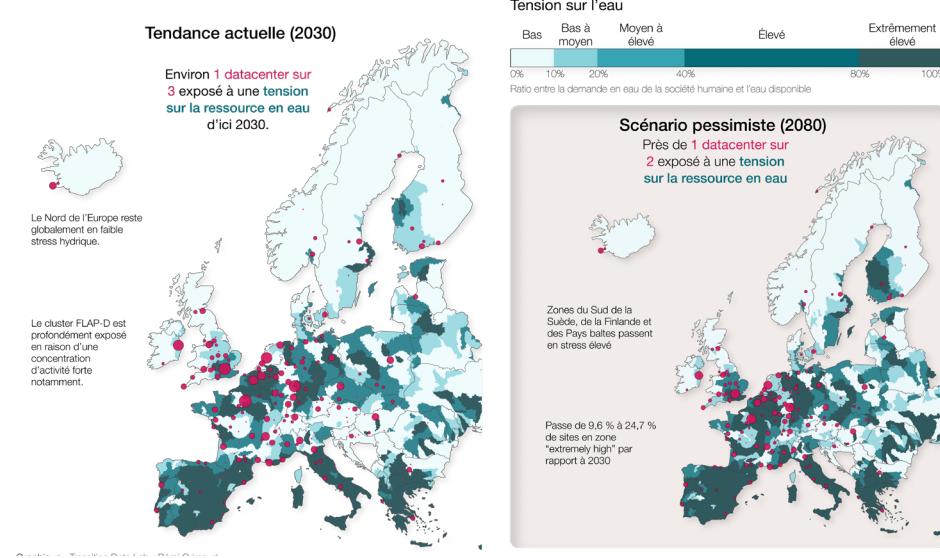


Figure 14 : Modèles d'évolution du stress hydrique en Europe en 2030 et 2080 – Source : Transition Data Lab

que des **stress hydriques pourront aussi toucher les zones industrialisées de l'Europe du Nord** (Figure 14).

RUPTURES

• **Innovations pour la valorisation de la chaleur fatale des datacenters** : des projets européens explorent comment rendre la chaleur fatale des datacenters **utile pour le chauffage urbain, serre²⁹ ou industriel**. Le projet **COOL DH³⁰** a montré que des quartiers entiers peuvent être chauffés grâce à de la **chaleur basse température recyclée**, tandis que **CoolHeatDC³¹** combine immersion et pompes à

chaleur pour augmenter la température récupérable. Le projet **MODERATOR³²** quant à lui utilise des **matériaux à changement de phase** pour stocker et restituer efficacement cette chaleur. Ces initiatives illustrent le potentiel de la chaleur des datacenters comme **ressource énergétique valorisable**.

• **Avec l'essor de l'IA, on envisage des racks de plusieurs centaines de kW**, rendant nécessaires des solutions avancées comme le **refroidissement liquide ou diphasique, en direct par des plaques froides au contact des composants, ou par immersion**, pour maintenir fiabilité et performance.

• Des programmes de recherche explorent

²⁶ <https://www.climateneutraldatacentre.net/>

²⁷ <https://www.datacenterdynamics.com/en/news/amazon-to-expand-number-of-data-centers-using-recycled-water-to-120/>

²⁸ <https://cwrrr.org/wp-content/uploads/2024/04/CWR-2024-China-ICT-running-dry-The-rise-of-AI-climate-risks-amplify-existing-water-risks-faced-by-thirsty-data-centres.pdf>

²⁹ https://france-chaleur-urbaine.beta.gouv.fr/documentation/datacenter_equinix.pdf

³⁰ <https://cordis.europa.eu/article/id/442938-warming-neighbourhoods-with-recycled-heat/fr>

³¹ <https://www.ait.ac.at/en/research-topics/heating-and-cooling-networks-for-cities-and-industry/projekte/coolheatdc>

³² <https://cordis.europa.eu/project/id/101136156>

activement le **refroidissement à eau chaude (« hot-water cooling »)** et des architectures thermiques conçues pour produire de l'eau en sortie à **60–70 °C**, rendant la chaleur fatale des datacenters plus **facilement valorisable** dans des réseaux de chaleur ou des usages industriels.

- Des voies de recherche sont explorées également pour **augmenter la température d'entrée permettant de** limiter les consommations d'eau au moment des canicules. **Il faut aussi travailler sur des températures de fonctionnement du hardware supérieures** afin de pouvoir fonctionner en « free cooling » à air direct ou indirect.

- **Datacenter hub et écologie industrielle** : dans son offre "**Datacenter et Territoires**", le groupe EDF met en avant la possibilité de **valoriser la chaleur fatale via des réseaux de chaleur** et d'optimiser les échanges d'énergie et de matière avec **d'autres acteurs industriels d'un même territoire**³³.

IA, HPC ET EDG

DESCRIPTION & ENJEUX

IA (INTELLIGENCE ARTIFICIELLE):

- Les datacenters dédiés à l'IA générative ne représentent que **quelques pourcents mais tirent le marché des datacenters actuellement**³⁴.

- **Deux trajectoires d'IA**, deux modèles économiques : d'un côté, nous assistons à l'avènement des «**Frontier Models**», des intelligences artificielles colossales développées par les «Hyperscalers». L'ambition est de créer des **plateformes globales et centralisées**. En parallèle, une **stratégie émerge, favorisant des modèles plus frugaux, spécialisés et surtout souverains**. Depuis 2024, une nouvelle génération de modèles d'IA dits Tiny Reasoning Models (TRM) remet en cause le paradigme du « toujours plus gros ». A noter que les hyperscalers développent aussi ces modèles spécialisés.

- Les datacenters deviennent l'infrastructure stratégique d'un **internet d'agents intelligents fondé sur la spécialisation coopérative**. La véritable

rupture ne réside pas uniquement dans l'autonomie des agents, mais dans leur capacité à se spécialiser (raisonnement, planification, exécution, vérification, négociation, perception, etc.) et à **coopérer dynamiquement pour atteindre des objectifs complexes** qu'aucun modèle monolithique ne pourrait traiter seul. Cette spécialisation distribuée implique des mécanismes d'interaction structurés via de **nouveaux protocoles (MCP, A2A)**, qui organisent les échanges, la délégation de tâches et la coordination sécurisée entre agents. **L'orchestration multi-agents** devient alors une conséquence nécessaire de cette architecture coopérative : **allocation dynamique** des rôles, gestion des dépendances, arbitrage des ressources, supervision et traçabilité. Ces protocoles et mécanismes structurent un nouvel espace stratégique, marqué par des **risques de verrouillage économique** (contrôle des standards d'interopérabilité, dépendance aux plateformes dominantes) et de fragmentation géopolitique. Plusieurs trajectoires sont possibles : concentration dans des hyper-campus IA fortement intégrés, maillage souverain distribué favorisant l'interopérabilité régionale, ou montée en efficacité logicielle à l'Edge via des essaims d'agents spécialisés.

HPC (HIGH PERFORMANCE COMPUTING) :

- **Au-delà de l'exascale : vers un nouveau paradigme du calcul haute performance**. Les lois de l'échelle qui ont soutenu l'informatique depuis cinquante ans, comme la **loi de Moore** et le « **Dennard scaling** »³⁵, montrent leurs limites et n'apportent plus les gains attendus en performance et en énergie. Pour continuer à progresser, le HPC entre dans une ère d'**hétérogénéité et de co-conception**. Les architectures de demain combinent processeurs CPU, accélérateurs GPU, puces spécialisées (ASIC, IPU, DPU), co-processeurs quantiques (QPU : Quantum Processing Unit) et unités neuromorphiques, chacune optimisée pour un type de tâche précis. Cette diversification des composants permet de maintenir une **croissance de performance effective, non pas en augmentant la fréquence, mais en réduisant le coût énergétique du calcul utile**.

- **D'une architecture «centralisée» à «distribuée»** : l'avenir est une **architecture hybride** où les grands HPC nationaux gèrent les défis de pointe, tandis que les architectures distribuées offrent la flexibilité et la scalabilité du cloud. Le HPC évolue d'un bâtiment de calcul statique vers une **plateforme de services hybrides, fédérée et intelligente**. Le HPC, historiquement financé par

³³ <https://cner-france.com/content/uploads/2025/05/DatacenterTerritoires-SMCI-2024-V-def.pdf>

³⁴ « Selon les inventaires industriels, il existe environ 11 800–12 000 datacenters physiques dans le monde aujourd'hui (<https://www.datacentermap.com/>). Parmi eux, environ 1 100+ sont des installations hyperscale (<https://www.srgresearch.com/articles/the-worlds-total-data-center-capacity-is-shifting-rapidly-to-hyperscale-operators>) – les principaux sites de pointe exploités par les grands cloud. Les datacenters explicitement optimisés pour les charges de travail IA sont minoritaires et peuvent être estimés à 300–400. Cette distribution permet d'illustrer que l'infrastructure IA représente une petite fraction du parc total, tout en tirant une part croissante de la croissance et des investissements dans le secteur. »

³⁵ Principe physique (1974) selon lequel la consommation d'énergie d'un transistor diminue proportionnellement à sa taille, permettant d'augmenter la puissance de calcul sans chauffer davantage. Sa fin (vers 2005) marque l'apparition du « mur thermique » : la miniaturisation n'offre plus d'économie d'énergie automatique, obligeant les datacenters à adopter des systèmes de refroidissement toujours plus complexes (comme le refroidissement liquide) pour gérer la chaleur accumulée.

subventions publiques, **adopte petit à petit le modèle «as-a-Service»** (HPCaaS) pour survivre et attirer les industriels. D'un point de vue **matériel et poussé par l'IA**, le HPC a adopté le GPU d'abord pour le parallélisme massif et la densité FLOPS/rack, puis pour sa performance énergétique.

EDGE :

• La « **promesse de la périphérie** » : L'Edge Computing consiste à **stocker et traiter l'information au plus près de sa source** afin de répondre aux exigences de réactivité des usages numériques modernes. Cet ancrage local permet de **traiter la donnée à proximité immédiate de l'utilisateur, ce qui réduit drastiquement la latence** pour les applications critiques tout en décongestionnant les réseaux de transport de données (Figure 15). L'enjeu est aussi résilience hors réseau et le contrôle local des données sensibles. Toutefois, la notion d'Edge demeure une frontière mouvante dont la **définition varie selon la perspective de chaque acteur industriel**. À titre d'exemple, un fabricant de capteurs situera le « Far Edge » au cœur même de la puce, alors qu'un opérateur télécom fixera cette limite à l'antenne 5G et qu'un fournisseur de Cloud l'étendra jusqu'aux datacenters régionaux. Cette infrastructure décentralisée représente un enjeu économique majeur puisque, selon la Fondation Linux, le marché mondial de l'Edge devrait atteindre 800 milliards de dollars dès 2028³⁶.

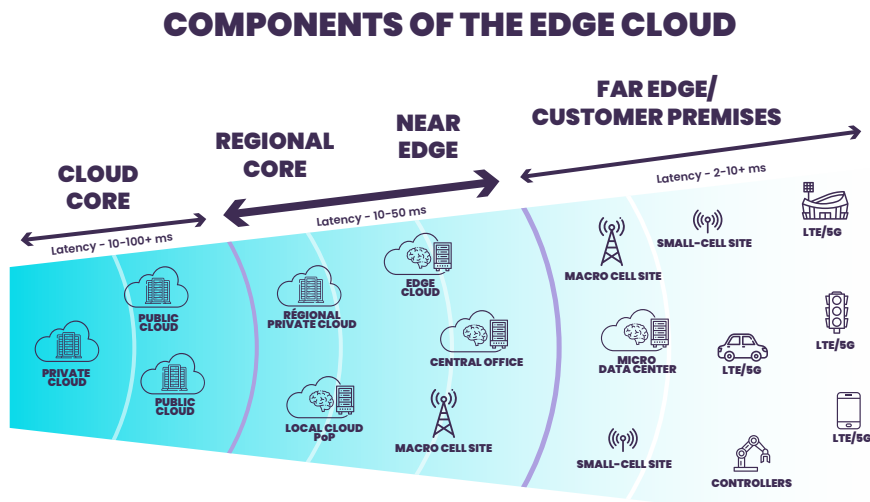


Figure 15: Architecture simplifiée du continuum Edge to Cloud

36 A. Joshipura, directeur général des réseaux au sein de la Fondation Linux (2019). Discours lors de l'Open Networking Summit

37 <https://www.hipeac.net/vision/2025/introducing-hipeacs-vision-for-the-future-the-next-computing-paradigm/#/>

• On distingue **plusieurs philosophies d'infrastructures EDGE suivant son interaction avec le Cloud** (Figure 16) :

♦ **Vision « Infrastructure » (Cloud-to-Edge)** : pour les géants du Cloud et des télécoms, l'Edge est une extension de leurs infrastructures (datacenters, Distributed and Central Units, etc.). On déploie des micro-datacenters proches des utilisateurs – Metro Edge en zones métropolitaines, Last Mile Edge en zones semi-urbaines ou rurales – pour compléter les grands datacenters.

♦ **Vision « Embarquée » (Cloud-to-Chip)** : pour les fabricants de puces et d'objets connectés (Apple, Qualcomm, Tesla), **l'Edge est directement sur l'appareil** : smartphones, voitures, casques AR, robots... L'IA s'exécute localement sur la puce, ce qui réduit la latence, protège les données, assure la résilience et économise de l'énergie. Il n'y a **pas besoin d'intermédiaires** entre l'appareil utilisateur et le cloud.

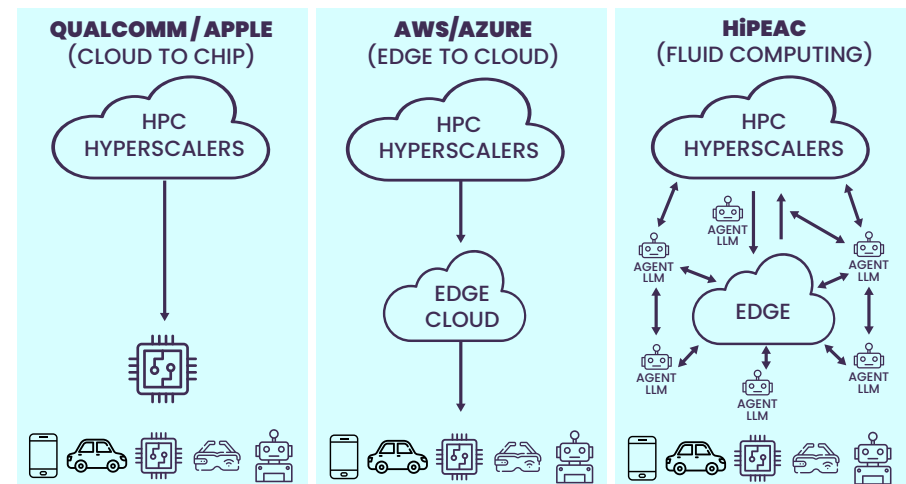


Figure 16 : Les différentes approches du continuum Edge to Cloud variant selon les acteurs

• le « **Fluid computing** » : dans ce modèle, les données peuvent non seulement migrer au sein de l'infrastructure, mais le code peut également s'y déplacer et s'exécuter là où des ressources de calcul ou de mémoire sont disponibles. C'est un **nouveau paradigme de cloud distribué qui est évoqué notamment dans le programme Européen Hipeac en tant que NCP: New compute Paradigm**³⁷. Cela engendre de **nombreux défis dont l'orchestration et l'allocation dynamique de ressources** avec des questions d'ouverture de standard et de modèles (libre versus propriétaire).

- **Freins au déploiement du Edge** : au niveau des EDGE Datacenter, la **dynamique de marché est bien moindre que celle des hyperscalers**. Au niveau de l'IA dans le edge, certains de nos partenaires sont dubitatifs sur le déploiement dans l'industrie (pas de « use-case » évident pour le moment). D'autre part, **la décentralisation complexifie la gestion** : gérer un parc de millions de nœuds edge est complexe, surtout à cause de la diversité des puces et systèmes et du manque de standardisation.

- **Modèle économique** : la lutte pour les standards (**ouverts vs. propriétaires**) est avant tout une lutte pour définir **qui captera la valeur de l'IA et de la data à la périphérie**.

MEGATRENDS

- Pendant de nombreuses années, l'intelligence artificielle dans le cloud s'était imposée comme un service parmi d'autres, sur une infrastructure de datacenter relativement « classique ». Avec l'essor des modèles de fondation et de l'IA générative, cette logique bascule : **l'IA ne vient plus après l'infrastructure, elle la redéfinit** (haute densité énergétique des **rack**³⁸, refroidissement liquide, infrastructures pouvant supporter de plus fortes charges au sol...). Cette transformation fonctionne dans les deux sens : **l'IA peut être utilisée pour garantir la continuité de service** (optimisation énergétique, orchestration intelligente, détection de défaillances), tandis que la **continuité de service devient une exigence essentielle pour l'IA**, qui repose sur des infrastructures fiables et performantes pour supporter des charges massives et des traitements critiques en temps réel.

- Émergence des Neoclouder ou « GPUaaS » : Avec cette convergence du HPC et des datacenters vers

38 <https://www.lemondeinformatique.fr/actualites/lire-la-consommation-energetique-des-racks-en-forte-hausse-dans-les-datacenters-97044.html>

39 <https://hellofuture.orange.com/fr/intelligence-artificielle-comment-les-neocloudeurs-transforment-le-marche-du-cloud/>

40 <https://www.forbes.fr/technologie/waicf-2026-a-la-recherche-d'une-troisieme-voie-europeenne-pour-une-ia-souveraine-et-frugale/>

41 <https://linuxfr.org/news/saga-openclaw-clawdbot-moltbot-enjeux-techniques-juridiques-et-ethiques-d-un-assistant-ia-open-source>

42 <https://en.fnnews.com/news/202510271256540125>

l'IA, on assiste à l'émergence de **nouvelles structures de calcul distribuées : NEOCLOUDERS³⁹ ou GPUaaS** (au service des « Clouder » mais aussi de NVIDIA ou de plus petits acteurs comme Mistral AI, etc.)

SIGNAUX FAIBLES

- **Émergence de modèles d'IA frugaux et souverains** : les architectures **TRM** ou **MoE** ou les modèles entraînés sur des jeux de données maîtrisés montrent qu'il est **possible d'obtenir des performances compétitives avec moins de calcul**, ouvrant une voie crédible à une IA européenne plus sobre et souveraine⁴⁰.

- **Promesse des agents intelligents et de l'IA agentique** : l'essor d'agents capables de planifier, coopérer et agir de manière semi-autonome annonce un changement de paradigme, où la valeur se déplace du modèle isolé vers des systèmes d'agents orchestrés au sein des infrastructures numériques. Le dernier né, en date qui permet de s'exécuter dans le système d'exploitation des PC, est OpenClaw⁴¹.

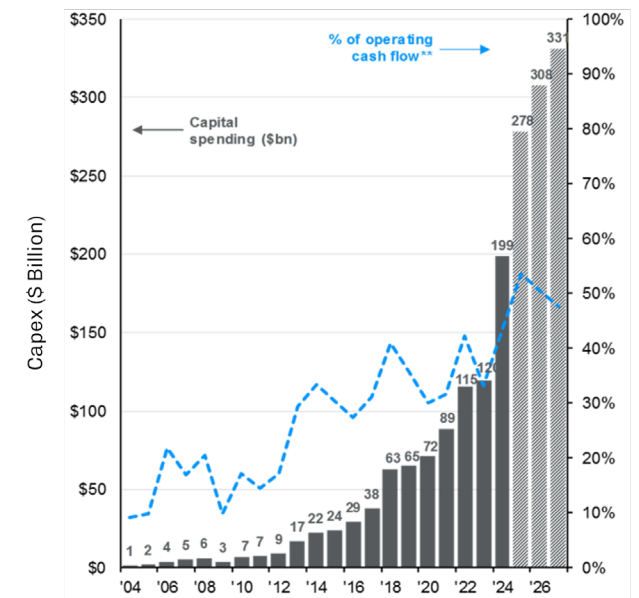
- **Hybridation datacenters, HPC et calcul quantique** : **l'intégration progressive du HPC et des premiers accélérateurs quantiques dans les datacenters impose des architectures hétérogènes**, combinant différents types de calcul, de réseaux et de chaînes logicielles au sein d'une même infrastructure.

RUPTURES

- **Bulle IA ?** Des niveaux d'investissements records sont faits par les BigTech dans l'IA, qui représentent des parts importantes de leur cashflow **sans pour autant**

dégager à l'heure actuelle de marge rentables, font penser à certains analystes économiques que les conditions d'une bulle sont réunies. Certains experts comme Yann Le Cun pensent que l'IA générative va bientôt atteindre la courbe de désillusions et que de nouveaux modèles d'IA comme **les « world models » sont l'avenir**⁴² (Figure 17)

AI Hyperscaler Capital Expenditures (Microsoft, Amazon, Meta, Alphabet)



Source: JP Morgan Asset Management. Data as of June 30, 2025.

Figure 17 : Part de l' Operating Cash Flow investi dans par les hyperscaler dans l'IA - Source : J.P.Morgan Asset Management

- Une rupture pourrait émerger si les architectures de continuité numérique s'appuyaient sur **des standards et protocoles ouverts, interopérables et**

agnostiques d'infrastructure. Un tel basculement limiterait la captation de valeur par quelques Big Tech et ouvrirait une fenêtre stratégique pour l'Europe, en favorisant un **Edge ouvert et souverain**, aligné avec ses contraintes énergétiques et réglementaires.

HARDWARE ET MÉMOIRES

DESCRIPTION & ENJEUX

- Les datacenters sont au cœur d'une accélération matérielle et architecturale portée par l'essor de l'IA, plus particulièrement de l'IA générative (GenAI) et des modèles de type « large language models ». Cette dynamique se traduit concrètement par la croissance du marché des semi-conducteurs sur ce segment devenant le **deuxième plus grand marché derrière les smartphones** selon Gartner⁴³. Les ventes de puces pour datacenters ont quasi doublé entre 2023 et 2024 atteignant 112 milliards de dollars en 2024.

- Les semi-conducteurs sont **devenus un enjeu géostratégique et de souveraineté majeur**, avec une production très concentrée autour de TSMC et d'ASML, alimentant une compétition accrue entre grandes puissances. Les États-Unis cherchent à **relocaliser et contrôler la filière** via subventions et restrictions, tandis que la **Chine accélère la construction d'un écosystème autonome**, misant sur des architectures massives pour compenser un retard technologique relatif.

- Les **GPU** (Graphics Processing Units) sont devenus la **pièce angulaire de l'intelligence artificielle moderne** grâce à leur calcul massivement parallèle – un rôle amplifié par NVIDIA et son écosystème

CUDA

- **MEMORY WALL** : Le **mouvement des données** est devenu le principal goulet d'étranglement dans les infrastructures pour l'apprentissage l'IA. Déplacer un bit à travers une machine consomme souvent plus d'énergie que de le traiter (Figure 18).

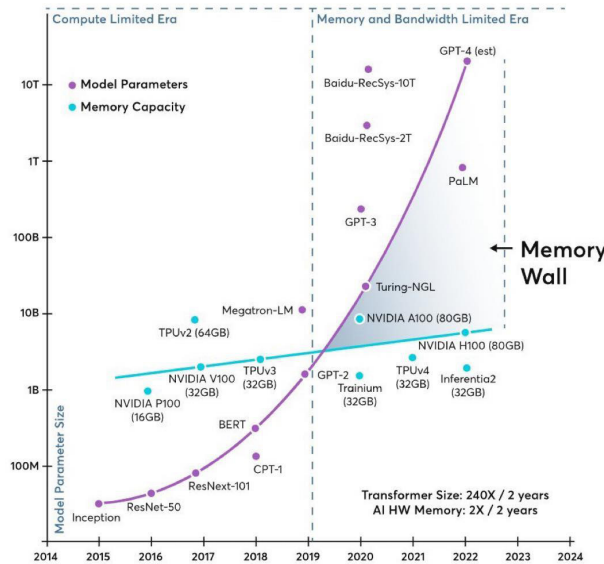


Figure 18 : Concept de memory wall - Source Ayar Labs

MEGATRENDS

- **Nouvelles architectures mémoires** : Les architectures actuelles, fondées sur une hiérarchie complexe (SRAM, HBM, DRAM, NVMe), **atteignent leurs limites face à la croissance des modèles**, imposant une recomposition de la mémoire « chaude ». D'où **l'émergence d'architectures «memory-centric»**, mais aussi de la mémoire empilée (HBM 3D⁴⁴) et

des interconnexions photoniques, qui **rapprochent la donnée du calcul et réduisent drastiquement la latence**.

- On observe une **hybridation systémique des architectures de calcul**, où CPU, GPU, TPU, NPU, QPU et puces neuromorphiques sont désormais **orchestrés au sein de systèmes hétérogènes**, chacun étant mobilisé selon la nature des charges (IA, calcul intensif, temps réel, optimisation).

SIGNAUX FAIBLES

- Cette dynamique s'étend au **sous-système mémoire**, avec l'émergence de **mémoires unifiées** et hiérarchisées (HBM, **CXL**, mémoire partagée CPU-accélateurs) permettant de réduire les latences, les copies de données et la fragmentation des espaces mémoire. Ensemble, ces évolutions dessinent un modèle de **co-design calcul-mémoire-logiciel**, clé pour soutenir l'IA à grande échelle tout en maîtrisant les coûts énergétiques et la complexité des systèmes.

- Le marché du matériel d'IA se distingue de plus en plus entre entraînement et inférence. Bien que l'entraînement reste très intensif et dominé par quelques acteurs avec des GPU/TPU de haute performance, **le segment de l'inférence connaît une croissance annuelle rapide, une adoption généralisée sur le cloud, le edge et les dispositifs embarqués, et une multiplication des architectures spécialisées (ASICs, NPU, FPGA)**. Les projections de croissance du marché des puces d'inférence jusqu'en 2030 (~ 90 Mds \$) et l'entrée de nombreux fournisseurs sur ce segment montrent que **le matériel d'inférence pourrait, en valeur et en volume, dépasser progressivement celui dédié à l'entraînement**⁴⁵.

43 <https://www.gartner.com/en/newsroom/press-releases/2025-02-03-gartner-says-worldwide-semiconductor-revenue-grew-18-percent-in-2024>

44 <https://www.yolegroup.com/strategy-insights/advanced-packaging-fueling-the-next-era-of-semiconductor-innovation/>

45 <https://www.deloitte.com/us/en/insights/industry/technology/technology-media-and-telecom-predictions/2026/compute-power-ai.html>

RUPTURES

• **Vers des mémoires à « coût marginal zéro »** : les progrès dans le stockage de données sur ADN⁴⁶ pourraient faire émerger un modèle où le stockage pur devient presque gratuit, avec des coûts projetés autour de **1\$/To/an**. Cela amène à des modèles **“Zero Storage Billing”** avec une facturation à l’écriture, à la restauration et à la transformation, plutôt qu’à la capacité.

• **Hardware dédié à l’inférence rapide et basse consommation** : l’essor de puces spécialisées pour l’inférence, comme **Cerebras Wafer-Scale Engine**, Graphcore IPU ou Habana Labs, permettent d’exécuter des modèles d’IA massifs avec une latence minimale et une consommation énergétique fortement réduite⁴⁷.

• **Open Hardware et souveraineté numérique** : le recours à des architectures open-source comme **RISC-V** ou des initiatives telles que l’**Open Compute Project (OCP)** offrent des opportunités pour construire des infrastructures numériques souveraines et modulaires⁴⁸. En Europe, des initiatives comme Open Source EDA Europe sont des leviers intéressants à suivre⁴⁹. L’adoption de **RISC-V** et de l’**Open Compute** offre à l’Europe une voie vers la souveraineté technologique⁵⁰, mais se heurte à des barrières opérationnelles majeures. Les entreprises manquent de ressources internes pour assurer la maintenance et l’évolution de ces systèmes ouverts, dont la maîtrise exige une expertise rare et coûteuse. De plus, des freins juridiques subsistent, notamment le risque de revendications de brevets sur certaines implémentations physiques, ainsi que des inquiétudes sécuritaires liées à l’exposition des plans de conception face au piratage.

RÉSEAUX & TÉLÉCOMMUNICATIONS

La frontière entre télécoms et datacenters n’existe plus vraiment : on est face à un continuum où les **opérateurs réseaux peuvent devenir des clients, voire des sous-traitants d’infrastructures cloud**, pendant que les **hyperscalers capturent la valeur sur le transport, le stockage et désormais l’IA**.

46 <https://www.cnrsinnovation.com/actualite/biomemory-le-stockage-des-donnees-sur-adn-pour-les-faire-durer-presque-eternellement/>

47 <https://cambrian-ai.com/big-ai-inference-has-become-a-big-deal-and-a-bigger-business/>

48 <https://www.opencompute.org/projects/open-chiplet-economy>

49 <https://fossi-foundation.org/resources/eu-roadmap>

50 <https://dl.acm.org/doi/10.1145/3731599.3767383>

51 https://sandvine.com/hubfs/Sandvine_Redesign_2019/Downloads/2024/GIPR/GIPR%202024.pdf

DESCRIPTION & ENJEUX

• **Le streaming et la vidéo dominant de 60 à 70 % le trafic IP mondial** dans le downstream fixe et mobile, confirmée par les rapports Global Internet Phenomena⁵¹.

• Le reste (gaming, réseaux sociaux, cloud, visio, OTT, e-commerce, apps mobiles, télétravail, puis IA générative grand public, etc.) vient densifier la demande, mais sans changer la logique : les réseaux télécoms sont devenus le système sanguin d’un écosystème dominé par quelques grandes plateformes et par l’essor des usages B2C intensifs (Figure 19).

Le streaming vidéo représente 61 % du trafic Internet

Répartition du trafic Internet descendant mondial en 2019 (%)



Source : Sandvine | The Global Internet Phenomena Report



statista

Figure 19 : Répartition du trafic internet descendant mondial selon les usages - Source : Sandvine, The global Internet Phenomena Report

• Les réseaux existants semblent capables d'absorber l'évolution actuelle du trafic Internet, mais la **croissance exponentielle des usages B2C** (XR, métavers, streaming 8K) et industriels IoT pourra soulever des **problématiques de capacité et de latence**.

• Les **données industrielles restent minoritaires en volume** comparées aux flux B2C (par ex. capteurs IoT envoyant quelques kilo-octets par événement), mais leur **fréquence élevée et leur criticité** augmentent les exigences de résilience et de **qualité de service**.

• L'enjeu stratégique est : « **Qui capte la valeur ?** » – les opérateurs réseau risquent de rester cantonnés au **rôle de « fournisseur de tuyau » à faibles marges si les hyperscalers intègrent le transport et la couche d'IA**.

MEGATRENDS

• **Désengagement immobilier des opérateurs historiques** : Verizon vend des datacenters à Equinix, SFR vers Ultraedge⁵², Iliad et Altice cèdent des sites au profit de fonds d'infrastructure, illustrant le **déplacement de valeur aux acteurs du cloud** et les fonds spécialisés.

• **Virtualisation et cloud-native** : les technologies NFV, vRAN, Open RAN/O-RAN, désagrégation du core 5G, microservices et orchestration Kubernetes **déplacent les fonctions traditionnellement embarquées dans du hardware propriétaire vers des serveurs standards, modifiant potentiellement la captation de valeur**⁵³.

52 <https://www.ultraedge.com/en/press/press-release/parteneriat-lancementultraedge>

53 <https://www.lemagit.fr/actualites/366614156/la-revolution-silencieuse-des-telecoms-le-virage-cloud-native>

54 https://www.bfmtv.com/economie/space-x-devient-un-peu-plus-un-operateur-mobile-l-entreprise-de-musk-rachete-pour-17-milliards-de-dollars-de-frequences-pour-utiliser-starlink-depuis-son-smartphone_AD-202509080576.html

55 <https://www.journaldugeek.com/2025/11/28/starlink-vs-amazon-leo-la-guerre-est-declaree-pour-linternet-par-satellite-a-l-gb-s/>

56 <https://www-file.huawei.com/admin/asset/v1/pro/view/8c64c0710ee04bee8e85385be5d944ad.pdf>

57 <https://digital-strategy.ec.europa.eu/en/policies/digital-networks-act>

58 <https://www.banquedesterritoires.fr/resilience-des-reseaux-telecoms-il-faut-orchestrer-la-reponse-des-pouvoirs-publics>

SIGNAUX FAIBLES

Space-to-cloud : En 2025, SpaceX, via sa filiale **Starlink, a racheté pour 17 milliards \$ les fréquences d'EchoStar**, afin de déployer une nouvelle génération de satellites capables de communiquer directement avec les smartphones⁵⁴. Aujourd'hui, c'est **AWS** qui lance son offre Leo (anciennement Project Kuiper), un service de connectivité par satellites en orbite basse conçu pour intégrer nativement les terminaux mobiles et les objets connectés au cloud. Contrairement aux réseaux satellites traditionnels, l'offre Leo d'Amazon ne se contente pas de fournir un accès internet : elle crée un lien privé et sécurisé entre les actifs mobiles sur le terrain et l'infrastructure AWS, permettant un traitement des données en temps réel directement à la périphérie du réseau (Edge) grâce à une nouvelle génération d'antennes compactes⁵⁵. Sans oublier la Chine qui a inauguré mi-2025 son projet de Constellation informatique avec le lancement de 12 premiers satellites, renforcé par l'annonce de la China Aerospace Science and Technology Corporation (CASC) qui intègre dans son nouveau plan quinquennal de développement (janvier 2026), l'envoi en orbite de plusieurs datacenters

RUPTURES

• **Déluge de données** : L'explosion des usages B2C (XR, métavers, streaming immersif) pourrait saturer les réseaux et les serveurs centraux, surtout avec les données ascendantes, **renforçant le développement du Edge Computing** pour traiter les flux localement et réduire la latence⁵⁶.

• **Réseaux télécom comme vecteurs de souveraineté et de résilience** :

◊ Les opérateurs historiques Européens pourraient jouer un rôle clé en garantissant **des infrastructures à faible latence et sécurisées pour les entreprises**, limitant la transmission de données sensibles vers des clouds étrangers⁵⁷.

◊ À terme, ils pourraient également participer au **doublage des infrastructures critiques** via des **réseaux autonomes et résilients**, notamment pour les communications d'urgence et la gestion de l'énergie⁵⁸.

MINAGE ET BLOCKCHAIN

DESCRIPTION & ENJEUX

Confrontés à l'incertitude du minage, et notamment à la pression sur les marges suite aux « halvings » successifs, les acteurs du minage du bitcoin les plus agiles ont opéré un pivot stratégique. Ils **«recyclent» leurs infrastructures en convertissant les processeurs dédiés au minage (ASIC) par des GPU pour répondre à la demande explosive du HPC (High-Performance Computing) et de l'IA, un marché plus stable et en croissance**. Ce pivot crée une **nouvelle classe d'acteurs hybrides : les Neo Clouder (Corweave, Hive, Riot, Marathon, etc.)**. Les technologies acquises dans le minage leur a donné l'expérience de la gestion d'infrastructures de calcul

à très haute densité, la maîtrise de l'orchestration de tâches complexes à grande échelle, et surtout, une expertise de pointe dans la gestion et l'optimisation des coûts énergétiques⁵⁹. En France, des sociétés comme **Policloud**⁶⁰ (associée au groupe américain Hivenet) promeuvent des **infrastructures de calcul modulaires**.

MEGATRENDS

• **Convergence du secteur du minage, de l'HPC et des datacenters** : l'infrastructure de minage (puissance électrique, refroidissement, surface) est réutilisée dans des offres de **colocation, AI hosting et services HPC**, reflétant une requalification économique de ces actifs⁶¹.

• **Cryptomonnaies, finance et datacenter** : les cryptomonnaies évoluent vers des **actifs financiers matures, de plus en plus intégrés dans les portefeuilles institutionnels** et les stratégies de trésorerie des entreprises⁶². Cette dynamique crée un nouveau marché où **la puissance de calcul devient une ressource financière**⁶³, et transforme le rôle des datacenters en plateformes multi-usages à la fois technologiques et économiques.

SIGNAUX FAIBLES

• GPUaaS et Neo Clouder « **off-grid** » : Le projet Stargate d'OpenAI repose sur une **autonomie énergétique partielle ou totale** par rapport au réseau électrique public – une logique issue du monde du minage de crypto où ces pratiques sont courantes. Ainsi ce ne sont pas les acteurs traditionnels qui ont été mandaté pour ce projet mais un **acteur issu du minage « Off-Grid »** qui opère sur des sites pétroliers, en exploitant le gaz torché comme source d'énergie (Figure 20).

• **Minage et arbitrage énergétique** : au Texas, **les mineurs de Bitcoin agissent comme une « batterie virtuelle »** pour le réseau électrique (ERCOT) : **ils absorbent les surplus d'énergie renouvelable** et coupent instantanément leurs machines du réseau lors des pics de demande pour éviter les pannes, un service qui est rémunéré⁶⁴. L'approche s'inscrit dans une logique de « **load balancing** » **énergétique** : **le minage, comme les datacenters de calcul, peut devenir un outil de flexibilité du réseau, à condition de rester modulable**. Le défi est d'en faire un usage **intelligent, flexible et couplé à la récupération de chaleur**, plutôt qu'un modèle spéculatif énergivore.



Figure 20 : C'est Crusoe Energy qui a été choisie pour faire le Gigadatecenter Stargate au Texas. Cette société de minage de bitcoin s'est fait connaître en exploitant le gaz torché sur des puits de pétrole pour alimenter ses calculs - Source : crusoe.ai

RUPTURES

• **Un cours du bitcoin en « flèche »** : Si le cours du bitcoin continue à croître jusqu'en 2040, le modèle économique tendra à renforcer les acteurs du minage et accélérer la convergence avec les datacenters. **Le modèle « Quantile »** de Sina Gopal prévoit une hausse progressive du prix, avec des jalons autour de 300 000 \$ d'ici 2029, 3 millions \$ vers 2039, et 10 millions \$ à l'horizon 2046⁶⁵.

59 <https://www.datacenters.com/news/bitcoin-miners-pivot-to-ai-data-centers-a-strategic-shift-in-2025>

60 https://tribuca.net/numerique_148287724-policloud-leve-75m-pour-batir-une-alternative-europeenne-aux-geants-du-cloud/

61 <https://www.datacenters.com/news/how-cryptocurrency-mining-farms-are-evolving-into-data-centers>

62 <https://www.weforum.org/stories/2026/01/digital-economy-inflection-point-what-to-expect-for-digital-assets-in-2026/>

63 <https://www.coindesk.com/coindesk-indices/2025/10/15/the-fortunes-of-tomorrow-will-be-built-on-compute-power>

64 <https://www.financierworldwide.com/the-evolving-energy-regulatory-landscape-for-cryptocurrency-mining-businesses-opportunities-and-risks>

65 <https://bitcoinmagazine.com/bitcoin-price-predictions>

• **Decentralized Physical Infrastructure (DEPIN) :**

Les DePIN sont des réseaux qui utilisent la **technologie blockchain**⁶⁶ et des **systèmes de récompense en tokens** pour développer des infrastructures physiques dans le monde réel. Avec l'exemple de la société Aethir, la puissance de calcul GPU devient une **ressource fluide, délocalisée** : elle n'est plus attachée à des datacenters propriétaires mais orchestrée à la demande sur un réseau mondial de GPU distribués. Le token **ATH** permet l'allocation dynamique, la rémunération et la gouvernance du calcul, faisant émerger un **marché temps réel de la puissance IA**. Cette rupture préfigure un **fluid computing** où le calcul circule vers les usages (edge, IA, rendu) et pourrait être complémentaire du modèle centralisé des hyperscalers⁶⁷.

• **Chip to Grid (C2G) Datacenter :** Ce concept développé par le NRL aux USA transforme les datacenters en **ressources énergétiques flexibles**, capables d'ajuster en temps réel la consommation de leurs puces en fonction des contraintes du réseau⁶⁸. Il repose sur une **intégration de la conception des puces, des serveurs et du réseau électrique** pour maximiser l'efficacité et la réactivité (Figure 21). Des acteurs comme **Emerald AI** appliquent cette logique à l'IA, orchestrant les workloads pour réduire les pics et optimiser l'efficacité, faisant des datacenters des **acteurs actifs du réseau électrique**⁶⁹. Il y a de plus en plus cette tendance de **fusion du calcul et de l'arbitrage énergétique** qui incite à penser que « **Compute is a New Power** ». Il faut néanmoins garder à l'esprit que les datacenters et le calcul ne constituent pas à l'heure actuelle le plus gros gisement de flexibilité et d'effacement, loin derrière les véhicules électriques où les consommations des bâtiments.

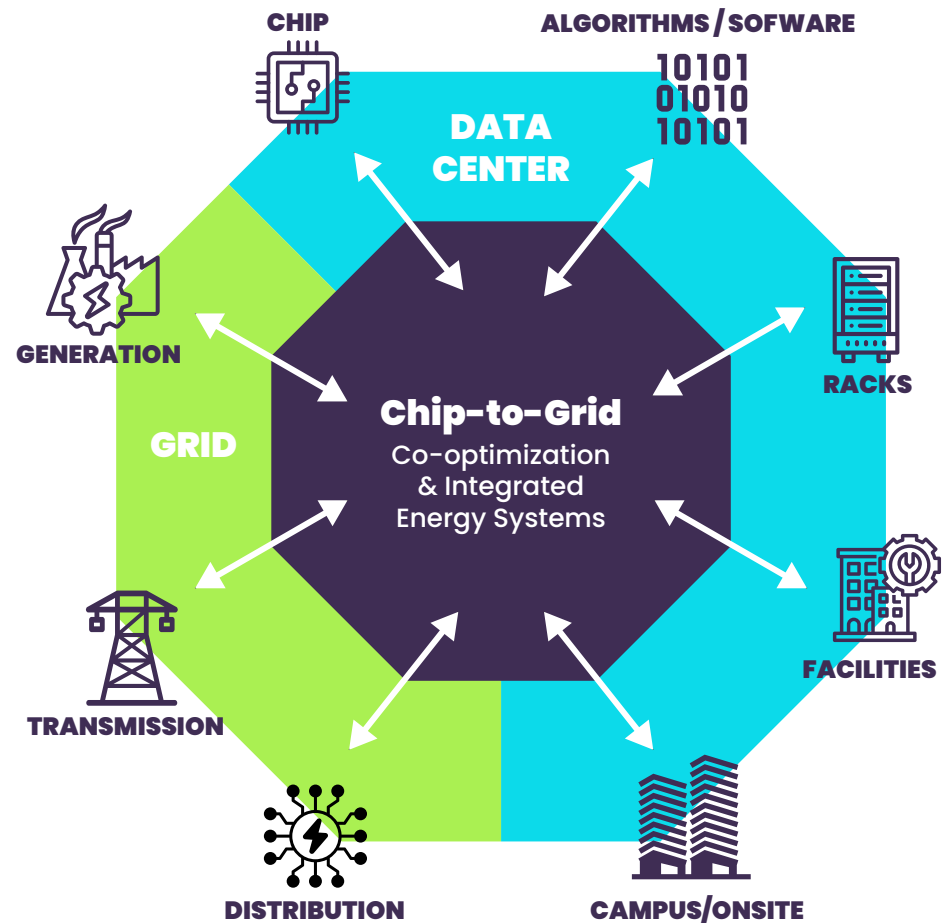


Figure 21 : Le concept de « chip-to-grid » datacenter permet aux réseaux électriques d'être en interaction proche avec les cœurs de calcul des datacenter pour un pilotage et « load balancing » instantané - Source : National Laboratories of the Rockies

66 <https://coinacademy.fr/academie/quest-ce-que-blockchain/>

67 <https://ecosystem.aethir.com/blog-posts/tokenized-gpu-compute-ath-as-liquid-ai-infrastructure>

68 <https://www.nrel.gov/computational-science/chip-to-grid-data-center-initiative>

69 <https://www.emeraldai.co/>

HYPOTHÈSES QUANTITATIVES & SCÉNARIOS

HYPOTHÈSES QUANTITATIVES

Nous avons pris des hypothèses à l'échelle mondiale d'ici 2040, sur la consommation d'énergie des datacenters, d'eau⁷⁰ et sur le trafic internet. Il est difficile de proposer des hypothèses sur des échelles de temps aussi grandes. Il s'agit plus de positionner les scénarios entre eux que de donner des valeurs absolues.

CONSOMMATION MONDIALE D'ÉNERGIE PAR LES DATACENTERS

Dans la littérature, on trouve des hypothèses très variées sur les consommations énergétiques dans le monde allant de quelques centaines de TWh/an à plusieurs dizaines de milliers (Figure 22).

Dans cette étude, les hypothèses de consommations énergétiques mondiales des DC sur lesquelles nous nous sommes basés dans nos scénarios sont issues de l'IAE : ENERGY AND AI, Février 2025⁷¹.

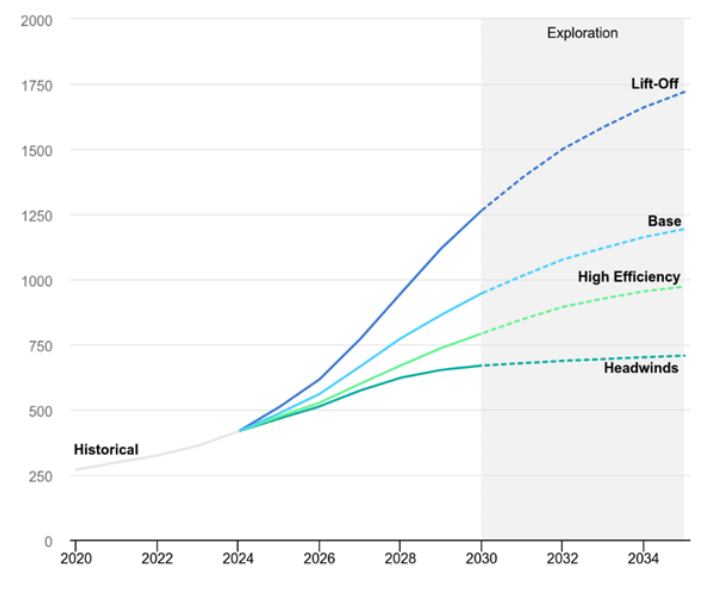


Figure 22: Hypothèses de consommation mondiale des datacenter jusqu'en 2035 - Source IEA, AI & Energy 2025

⁷⁰ Data Center Water Usage: A Comprehensive Guide - Dgtl Infra, <https://dgtlinfra.com/data-center-water-usage>

⁷¹ <https://www.iea.org/reports/energy-and-ai>

CONSOMMATION MONDIALE D'EAU

Pour projeter la consommation mondiale d'eau des datacenters⁷², la métrique Water Usage Effectiveness (WUE) est utilisé. Par conséquent, la consommation totale d'eau (L) est égale à la consommation totale d'énergie informatique (kWh) multipliée par le WUE (L/kWh).

Scénario « croissance accélérée » : Dans ce scénario de forte croissance, les hyperscalers atteignent des WUE très performants (0,19 L/kWh) grâce à des technologies de refroidissement avancées comme le Direct Liquid Cooling ou l'immersion. Les datacenters de taille moyenne atteignent également des WUE améliorés (0,30 L/kWh). Cependant, une part significative des anciens datacenters, qui constituent la majorité, ne sont pas modernisés et continuent de consommer beaucoup d'eau avec des systèmes de refroidissement moins efficaces.

Scénario « Instabilité et cycles technologiques » :

Ce scénario implique une croissance plus modérée avec moins d'hyperscalers et de datacenters à l'échelle mondiale. Il se caractérise par une forte hétérogénéité des WUE, bien qu'une amélioration globale soit observée. Les grands groupes continuent d'innover, mais l'adoption des meilleures pratiques est inégale.

Scénario « Nouvelle donne numérique européenne » :

Dans ce scénario, l'Europe impose des normes strictes de WUE pour les hyperscalers et tous les nouveaux datacenters (moyens ou petits), favorisant les systèmes innovants. Aux États-Unis, les normes sont moins contraignantes, mais un mouvement réglementaire existe et les hyperscalers y font des efforts significatifs. Cela conduit à une proportion

plus élevée de datacenters très efficaces.

Scénario « Résilience et Sobriété » : Ce scénario est caractérisé par des normes strictes de WUE partout dans le monde, où les faibles consommations d'eau sont privilégiées. Cela reflète une adoption généralisée des technologies et des pratiques les plus efficaces en matière de gestion de l'eau.

Les calculs ci-dessous modélisent la consommation totale d'eau (en milliards de litres/an) pour chaque hypothèse décrite par scénario, en **fonction de la répartition supposée des niveaux d'efficacité (WUE Moyen, Amélioré, Agressif) dans le parc de datacenters en pourcentage**⁷³.

Remarque : Nous parlons ici de données de **consommation d'eau directement pour le refroidissement des datacenters et non pas de prélèvements ou d'utilisation d'eau indirecte** (fabrication du courant qui alimente les datacenters, fabrication du hardware etc.)

TRAFIC DE DONNEES INTERNET

De même que pour l'énergie, il y a de grosses variations dans les hypothèses de trafic mondial internet (**de l'ordre du Yottabyte** dans le rapport Intelligent World 2030: Global Outlooks de Huawei). Il faut distinguer **trafic IP (réseaux Internet/transmission de données) et volume total des données générées ou stockées** : les chiffres ne sont pas directement comparables (à l'échelle mondiale, le **stockage et la gestion des données** atteignent déjà des **centaines de zettabytes**⁷⁴). Pour évaluer les hypothèses de trafic, nous nous sommes basés sur les données de l'UIT qui donne des volumes de trafic de l'ordre de 6174 EB /an⁷⁵ (environ 1 074 EB de trafic mobile et 5 100 EB de trafic fixe). Plusieurs études estiment que, à l'horizon 2030, le trafic fixe mondial pourrait plus que doubler, tandis que le trafic mobile pourrait être multiplié par un facteur proche de

SCÉNARIOS	TWh/an (Énergie totale)	% WUE Moyen (1,8 L/ kWh)	% WUE Amélioré (0,30 L/kWh)	% WUE Agressif (0,19 L/kWh)	Consommation d'eau totale (Mrds L/an)
CROISSANCE ACCÉLÉRÉE	1700	40 %	35 %	25 %	1 884
INSTABILITÉ ET CYCLES TECHNOLOGIQUES	1250	35 %	45 %	20 %	1 291
NOUVELLE DONNE NUMÉRIQUE EUROPÉENNE	970	15 %	45 %	40 %	690
SOBRIÉTÉ ET RÉSILIENCE	700	10 %	40 %	50 %	614

⁷² https://www.researchgate.net/publication/281719759_The_Water_Footprint_of_Data_Centers

⁷³ <https://librairie.ademe.fr/energies/8910-prospective-d-evolution-des-consommations-des-data-centers-a-court-moyen-et-long-terme-de-2024-a-2060.html>

⁷⁴ <https://gitnux.org/data-storage-industry-statistics/>

⁷⁵ Facts and Figures 2023 - ITU, <https://www.itu.int/itu-d/reports/statistics/2023/10/10/ff23-foreword/>

trois⁷⁶ ⁷⁷. Ces ordres de grandeur résultent d'extrapolations de tendances observées sur la période 2015-2023, intégrant la montée en puissance de la 5G, du streaming vidéo haute définition, du cloud et des usages numériques avancés. Nous estimons pour notre étude un trafic mondial de l'ordre de 13422 EB en 2030, base sur laquelle nous avons appliqué différents taux de croissance actualisés selon les scénarios en 2040.

SCÉNARIO CROISSANCE ACCÉLÉRÉE :

Compte tenu de la nature «débridée», un TCAC élevé soutenu de 35 % de 2030 à 2040 est supposé, reflétant potentiellement l'extrémité supérieure de la croissance de la DCI de l'IA ou la forte adoption de la XR de Nokia.

Calcul : $13\,422 \text{ EB/an} * (1 + 0,35) \wedge 10 \approx \mathbf{267\,000 \text{ EB/an}}$

SCÉNARIO INSTABILITÉ ET CYCLES TECHNOLOGIQUES :

Un TCAC de 25 % de 2030 à 2040 est supposé moins agressif que le scénario «accélérée» mais sans régulation ni optimisation du trafic.

Calcul : $13\,422 \text{ EB/an} * (1 + 0,25) \wedge 10 \approx \mathbf{125\,000 \text{ EB/an}}$

SCÉNARIO NOUVELLE DONNE NUMÉRIQUE EUROPÉENNE:

Un TCAC de 20 % de 2030 à 2040 est supposé, s'alignant sur le taux de croissance annuel projeté d'Ericsson pour le mobile d'ici 2030.4 Cela reflète un marché plus mature avec une optimisation des trafics.

Calcul : $13\,422 \text{ EB/an} * (1 + 0,20) \wedge 10 \approx \mathbf{83\,105 \text{ EB/an}}$

SCÉNARIO RÉSILIENCE ET SOBRIÉTÉ :

Un TCAC faible de 10 % de 2030 à 2040 est supposé, reflétant des contraintes significatives.

Calcul : $13\,422 \text{ EB/} * (1 + 0,10) \wedge 10 \approx \mathbf{34\,813 \text{ EB/an.}}$

SCÉNARIO « CROISSANCE ACCÉLÉRÉE »

PITCH

« Explosion des usages de IA générative et des agents, lunettes XR, véhicules autonomes et objets connectés – modèle énergivore plutôt centralisé avec une offre EDGE dominée par les big tech US – dans un monde peu régulé, piloté par la croissance économique. La décarbonation n'est pas le nerf de la guerre malgré des efforts réels et des progrès d'efficacité de certains acteurs. L'énergie est très convoitée mais des solutions existent toujours pour les besoins digitaux. L'Europe profite de cette croissance mondiale et de l'afflux d'investissements dans le secteur de l'IT mais elle est de plus en plus dépendante des Big Tech US dans un monde très polarisé, majoritairement américain, sous une influence chinoise grandissante »

Légende :

Un exemple typique de datacenter dans ce scénario est représenté par un campus d'hyperscalers au Texas, USA. Ces campus sont des hub énergétiques qui concentrent des hyperscalers qui mutualisent des sources de production d'énergie (SMR, Gas, photovoltaïque) dont une partie est off-grid. On est sur une explosion des usages comme la réalité augmentée, les jumeaux numériques, la robotique ou les véhicules autonome, ce qui fait exploser le trafic internet et notamment les flux ascendants. Cela nécessite un continuum edge to cloud avec différentes couches de traitement Edge. L'IA a continué sa croissance avec le déploiement d'agents intelligents distribués dans des infrastructures possédées en majorité par les Big Tech. Le réseau internet se développe dans toutes les parties du monde avec l'émergence du space cloud via satellite. Les serveurs IA ont de très fortes densités énergétiques (~1MW) et concentrent de nombreuses innovations : photonique, co-processeurs quantiques, datacenter robotisés, etc.

⁷⁶ Internet traffic trends & growth forecasts - <https://www.sofrecom.com/en/news-insights/internet-traffic-growth-trends-and-forecasts.html>

⁷⁷ Abiresearch : <https://www.abiresearch.com/press/increased-demand-for-bandwidth-intensive-services-cause-mobile-data-traffic-to-surge-threelfold-by-2030>

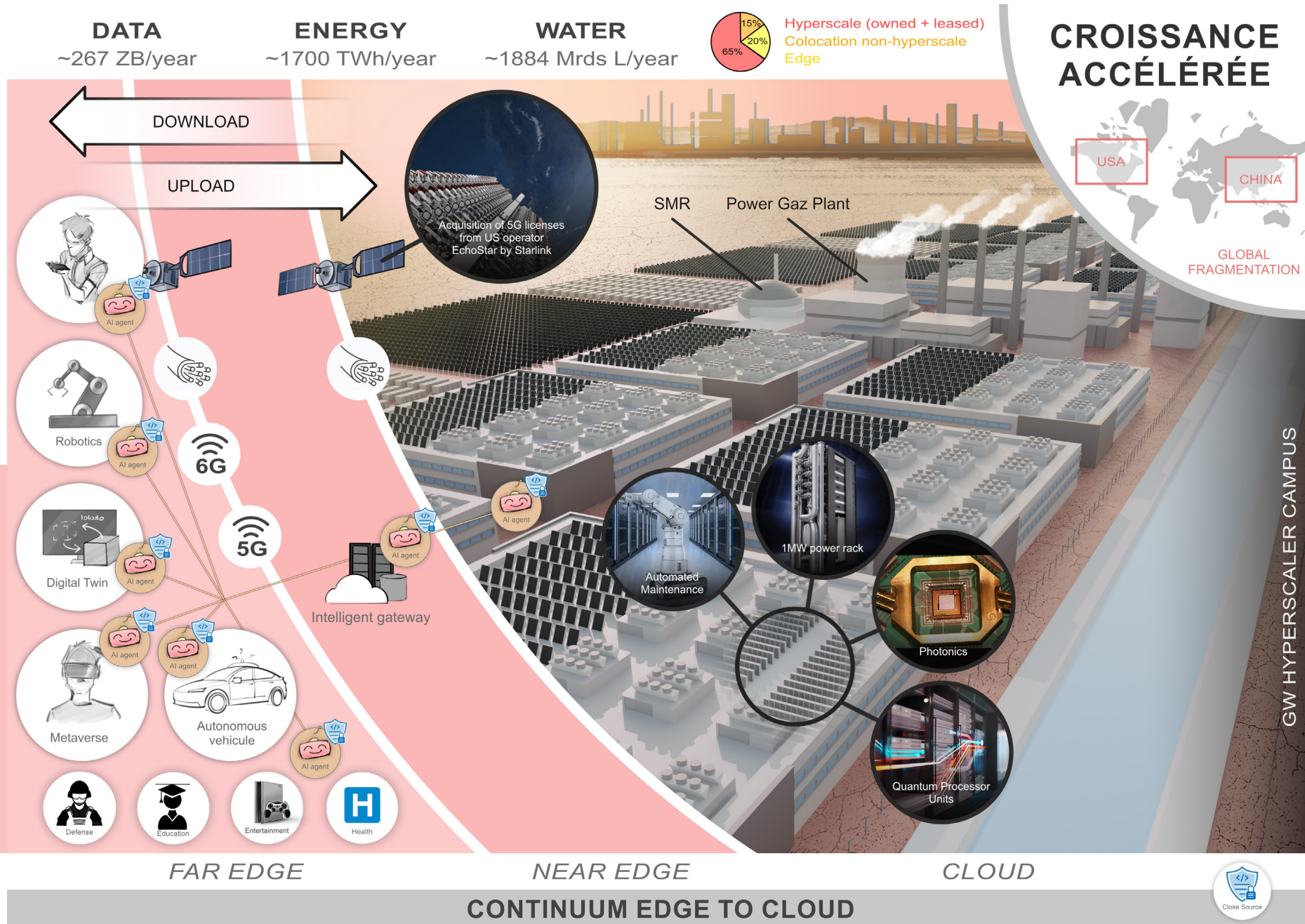


Figure 23. Scénario «Croissance accélérée» a course aux hyperscalerratories of the Rockies

GÉOPOLITIQUE : UNE POLARISATION USA-CHINE ET UNE EUROPE SOUS INFLUENCE

L'environnement géopolitique évolue vers une **bipolarisation durable entre les États-Unis et la Chine**, marquée par une **compétition stratégique intense mais contenue**. Les tensions récurrentes, notamment dans l'Indo-Pacifique, génèrent une **volatilité économique et commerciale accrue** sans rupture prolongée des échanges mondiaux.

Cette rivalité se traduit par une **concurrence renforcée sur les technologies critiques**, les infrastructures numériques et les chaînes de valeur industrielles, entraînant une **fragmentation partielle des écosystèmes** et une adaptation des acteurs globaux. Dans ce contexte, **l'Europe reste sécuritairement alignée sur les États-Unis** et **économiquement exposée aux deux pôles**, sans parvenir à stabiliser une autonomie stratégique complète.

DATA CENTER ET CLOUD : UNE DOMINATION DES BIG TECH ET DES HYPERSCALERS

La dynamique du cloud est marquée par une **domination structurelle des Big Tech et des hyperscalers**, appuyée sur des **architectures hautement centralisées et mondialisées**. Les infrastructures de datacenters deviennent **fortement hétérogènes et modulaires**, intégrant des workloads **hybrides IA-HPC-quantique** orchestrés de manière fluide au sein de **plateformes cloud unifiées**, ce qui marginalise progressivement les modèles **on-premise**. La montée en puissance de l'IA générative et du **GPU-as-a-Service** s'accompagne de **densités énergétiques très élevées**, de réseaux à très faible latence et du déploiement de **technologies avancées de connectivité et de photonique intégrée**.

À l'horizon 2040, le cloud hyperscale — combinant infrastructures détenues et louées — concentre une **part majoritaire des capacités mondiales**, consolidant la position dominante des acteurs globaux du cloud et du calcul intensif.

USAGES B2C, B2B ET DÉFENSE

Dans le scénario de **croissance soutenue**, les usages numériques connaissent une **expansion rapide et largement acceptée**, portée par la **généralisation de services hyper-personnalisés** et fortement consommateurs de données.

En **B2C**, l'IA irrigue les usages du quotidien — assistants numériques, agents autonomes, mobilité autonome, environnements immersifs — fonctionnant en **temps réel** et mobilisant des volumes massifs de données issues de capteurs,

de vidéos haute définition et de jumeaux numériques individuels, avec une adhésion sociale élevée malgré des préoccupations résiduelles.

En **B2B**, les usages se structurent autour de **solutions cloud standardisées mais fortement segmentées**, permettant aux petites organisations de privilégier le coût et la productivité, tandis que les grandes entreprises recourent aux services avancés des Big Tech pour leurs fonctions support, tout en conservant un contrôle accru sur les données stratégiques.

Dans le domaine de la **défense**, les architectures reposent largement sur des **briques civiles issues des Big Tech**, adaptées, certifiées et sécurisées via des partenariats spécifiques, faisant de ces acteurs des **fournisseurs technologiques stratégiques de facto**.

HPC, IA ET EDGE : CONTINUUM EDGE TO CLOUD PILOTÉ PAR LES BIG TECH

Le scénario repose sur un **Continuum Edge-to-Cloud entièrement piloté par les Big Tech qui est une extension hiérarchique du cloud central**. Les architectures s'organisent en **niveaux complémentaires**, du **far edge** pour les pré-traitements locaux au **near edge** métropolitain pour les usages à **faible latence**, jusqu'aux **cœurs régionaux hyperscale** dédiés à la consolidation des données et à **l'entraînement des modèles massifs**.

L'IA et le HPC convergent dans des workloads hybrides, avec un **entraînement centralisé** et une **inférence largement distribuée à l'edge** pour garantir la réactivité. Ces capacités sont opérées au sein de **plateformes agentiques propriétaires**, tandis que les Big Tech s'appuient sur un **écosystème de partenaires locaux licenciés** pour déployer leurs technologies au plus près des usages, tout en conservant la maîtrise des couches critiques

MÉMOIRE ET STOCKAGE : LA COURSE À LA BANDE PASSANTE

La **mémoire et le stockage** sont dominés par une **course à la bande passante**, la **latence devenant critique**, tandis que le **coût et l'efficacité énergétique** passent au second plan face aux besoins des GPU, TPU et QPU. La **mémoire HBM empilée en 3D** s'impose pour l'IA haut de gamme, offrant des **bandes passantes extrêmes**, complétée par l'adoption massive de **CXL** pour créer de vastes **pools de mémoire partagée**, souvent propriétaires et générateurs de **verrouillage technologique**.

Les **interconnexions optiques co-packagées** et la **SRAM on-chip** renforcent encore la performance, au prix de **contraintes thermiques élevées** nécessitant

un **refroidissement liquide**. En parallèle, le **stockage froid** voit son **coût marginal chuter**, favorisant des **modèles de stockage quasi gratuit**, tandis qu'une partie du stockage à **haute bande passante** se rapproche du processeur pour accélérer l'accès aux modèles d'IA.

ÉNERGIE : L'ACCÈS À L'ÉNERGIE PRIORITAIRE

Dominé par une **priorité absolue à l'accès continu, massif et sécurisé à l'électricité**, condition sine qua non du développement des datacenters hyperscale. Lorsque les réseaux centralisés ne peuvent suivre le rythme de la demande, des **solutions off-grid** sont déployées dans les pays où la réglementation le permet. Les sources offrant la **densité de puissance et la fiabilité requises** — principalement le **nucléaire et le gaz** — s'imposent comme piliers du système, tandis que les datacenters se regroupent en **campus énergétiques** à proximité directe des capacités de production, notamment aux États-Unis et dans certains pays nordiques.

Cette dynamique s'accompagne d'une **accélération des SMR** et d'investissements soutenus dans la **fusion nucléaire**, encore non mature à l'horizon 2040. **Les énergies renouvelables**, majoritairement photovoltaïques et couplées à des **systèmes de stockage par batteries**, jouent un rôle complémentaire d'optimisation et d'écrêtement des pics, sans constituer le socle principal. Le **fuel** demeure enfin un **back-up stratégique**, garantissant l'uptime maximal dans une logique hyperscale où la continuité de service prévaut sur toute autre considération.

EAU & REFROIDISSEMENT : GÉNÉRALISATION DU LIQUID COOLING

Les hyperscalers de nouvelle génération atteignent des **performances hydriques très élevées** ($WUE \approx 0,19 \text{ L/kWh}$) grâce au **refroidissement liquide avancé** devenu la **norme pour les charges denses** (IA, HPC, quantique), seul capable de gérer des **densités extrêmes et des architectures photoniques intégrées**. Toutefois, une **dualité structurelle** persiste : une **majorité de datacenters existants** n'est pas modernisée et continue d'opérer avec des **systèmes à air et tours de refroidissement**, plus consommateurs d'eau, générant des **tensions locales sur la ressource** malgré une contribution moindre aux charges les plus intensives.

La **valorisation de la chaleur fatale** reste marginale et **opportuniste**, mise en œuvre uniquement lorsqu'un **modèle économique clair** existe, notamment au sein de **campus énergétiques**. Cette coexistence explique qu'à l'échelle sectorielle, le **WUE moyen** demeure pénalisé par l'héritage du parc existant,

malgré l'excellence des nouvelles installations.

TÉLÉCOM : UNE VALEUR DÉPLACÉE VERS LE CLOUD ET L'ÉMERGENCE DE RÉSEAUX 6G CLOUD-NATIVE

Les **télécoms** évoluent vers un modèle où la **valeur se déplace du matériel réseau vers le cloud et le logiciel**, sous l'effet de la généralisation de l'IA, de la XR et des besoins croissants en **uplink**, notamment dans les zones urbaines denses. Les hyperscalers privilégient **la fibre noire** et internalisent la gestion de leurs équipements, tandis que les opérateurs développent des **réseaux définis par logiciel (SDN)** pour orchestrer efficacement les charges IA. Les progrès de la **fibre optique** favorisent le rapatriement des données vers le cloud, où les **fonctions réseau virtualisées** (CU, DU, RU) sont hébergées sur des **serveurs standards**.

Cette dynamique accélère l'émergence de **réseaux 6G entièrement cloud-native**, portés par les acteurs du cloud et du logiciel, au détriment des fournisseurs historiques de hardware, malgré leurs tentatives d'adaptation. En parallèle, les **Big Tech déploient massivement des constellations de satellites**, captant des **parts de marché croissantes sur plusieurs continents** et renforçant leur contrôle de bout en bout sur la connectivité mondiale, au détriment des acteurs historiques du matériel télécom.

BLOCKCHAIN & CRYPTO : DU MINAGE RÉSIDUEL À L'OPTIMISATION ÉNERGÉTIQUE DES WORKLOADS

En **2040**, plus de **99,8 %** des bitcoins auront été extraits et le minage ne sera plus une activité autant spéculative que par le passé (Bitcoin post-quantique ou autres crypto).

Cette activité se concentre sur quelques gros acteurs attirés par une énergie peu chère et disponible. Les campus de datacenters qui sont de véritables campus énergétiques attirent également des **activités de minage qui viennent profiter des infrastructures (électricité et refroidissement)**. Ces grands acteurs industriels s'associent aux grands acteurs de datacenters pour participer à un **arbitrage énergétique** (trading d'énergie ou pilotage de charge).

SCÉNARIO « INSTABILITÉ ET CYCLES TECHNOLOGIQUES »

PITCH

« Phénomène de hype, crises économiques multiples et instabilités géopolitiques, avec des transitions énergétiques et écologiques mal maîtrisées. Malgré des volontés politiques affichées, la réalité de la décarbonation reste très difficile dans un monde où les ambitions écologiques ne sont que difficilement atteintes et où la compétition et les Retours sur Investissements (ROI) priment. L'Europe décroche et subit l'hégémonie des BigTech US dans un monde polarisé entre USA et Chine mais dans lequel viennent jouer de nouveaux acteurs comme l'Inde. »

Légende :

Dans ce scénario où de nombreuses crises économiques et tensions sur l'énergie ont lieu, les acteurs du digital privilégient des solutions éprouvées et limitent les prises de risques favorisant les faibles CAPEX. Les usages du digital sont importants (jeux, vidéos, etc.) mais on n'a pas le déluge de données comme dans le scénario de croissance accélérée. On voit ainsi des campus de datacenters qui se développent en suivant les différentes évolutions technologiques. On est en présence de datacenters hétérogènes où on cherche surtout à optimiser l'existant : co-existent sur un même site des datacenters modulaires en container (dont une partie servent au minage), des friches industrielles réutilisées en profitant des accès à l'énergie existant ou bien des datacenters utilisant des technologies avancées de refroidissement les plus économes. L'Intelligence artificielle continue d'avoir une place privilégiée mais se retrouve en majorité centralisée et dans les mains de quelques BIG tech. Les architectures Edge ne se sont développées que dans certaines niches. Il y a peu de strates intermédiaires entre les objets connectés et le cloud : on est dans un continuum cloud to chip.

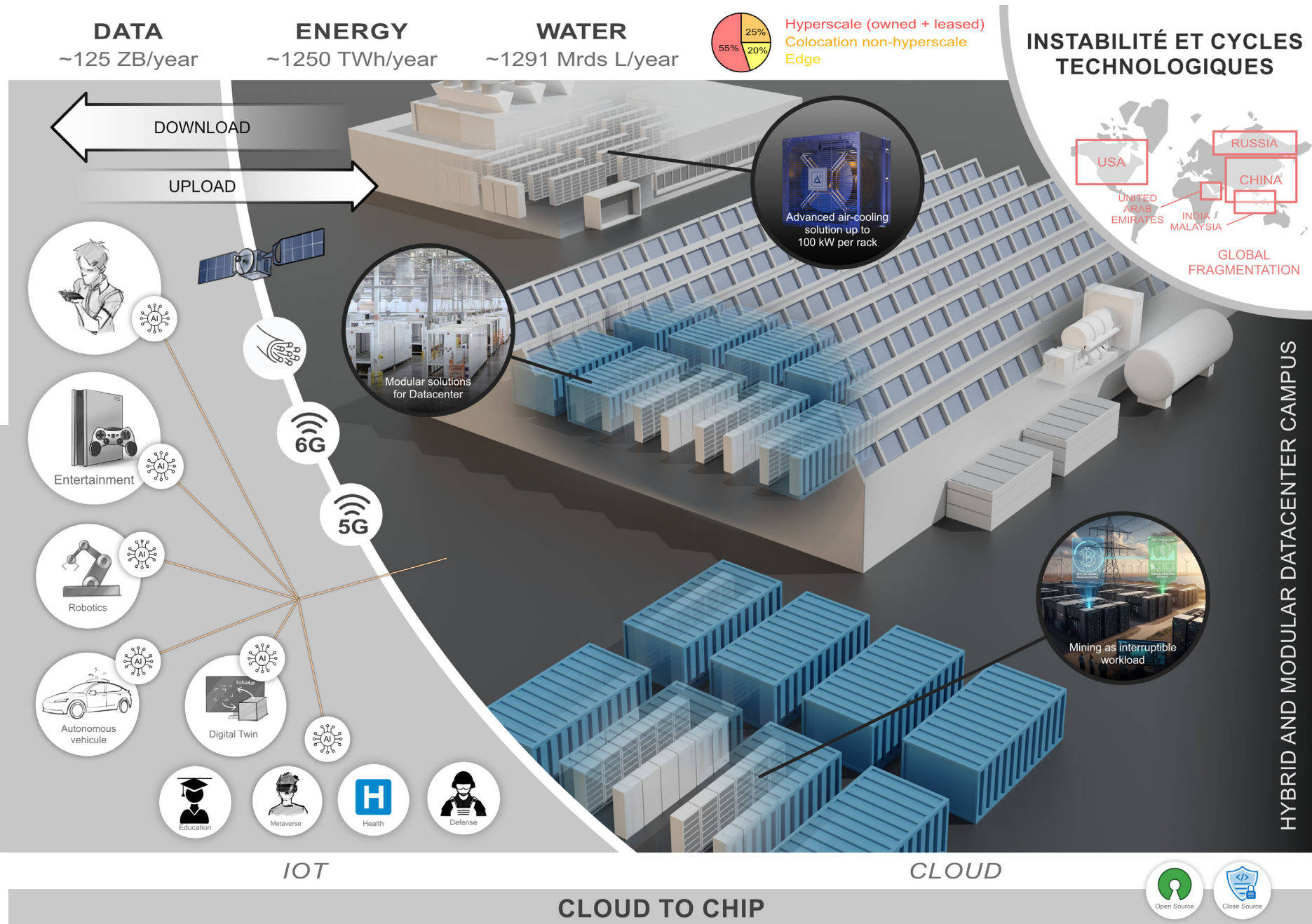


Figure 24 : Scénario «Instabilité et cycles technologiques»

HYPOTHÈSES GÉOPOLITIQUES : UN MONDE PLUS MORCELÉ AVEC UN DÉCROCHAGE EUROPÉEN

Dans le scénario d'Instabilité **et cycles technologiques**, le système international évolue vers une **multipolarité fragmentée et instable**, marquée par une **faible capacité de coordination entre puissances** et une succession de **chocs politiques, énergétiques et technologiques**. Les États-Unis connaissent des phases d'**instabilité intérieure** favorisant le **retour du protectionnisme** et des **politiques économiques erratiques**, alimentant des **cycles spéculatifs courts** dans l'IA et le numérique. Les crises énergétiques renforcent la **vulnérabilité des infrastructures critiques**, tandis que le centre de gravité mondial se déplace vers l'Asie du Sud et du Sud-Est autour d'un **axe Inde-Indonésie** structurant.

Dans ce contexte, **l'Europe connaît un décrochage politique, économique et technologique**, marqué par une **fragmentation interne** et une perte durable de compétitivité.

DATA CENTER ET CLOUD : CONSOLIDATION DU MODÈLE CENTRALISÉ APRÈS LA BULLE IA

Dans le scénario d'Instabilité **et cycles technologiques**, la dynamique du cloud se caractérise par un **renforcement du modèle centralisé**, au détriment des architectures distribuées et du micro-edge, qui restent marginales. Les **Big Tech privilégient des architectures matérielles et logicielles éprouvées**, plus robustes face à la volatilité des marchés, tandis que les **solutions de datacenters modulaires** se diffusent pour ajuster plus rapidement les capacités. Les acteurs de type **NeoCloud**, fortement exposés aux cycles spéculatifs du GPU et à des niveaux d'endettement élevés, disparaissent progressivement, leurs actifs étant **consolidés par les grands acteurs du cloud**.

À l'horizon 2040, le marché se stabilise autour d'un **cloud centralisé dominé par un nombre limité d'acteurs globaux**, combinant hyperscale et colocation, dans un contexte de **rééquilibrage post-hype** et de maintien d'une diversité d'architectures sans rupture majeure.

USAGES B2C, B2B ET DÉFENSE

Dans ce scénario, les usages numériques évoluent dans un **environnement fragmenté et instable**, marqué par des **inégalités croissantes d'accès, de qualité et de sécurité**. En **B2C**, le marché devient clairement **à plusieurs vitesses**, les niveaux de service et d'assistance numérique dépendant fortement des

revenus, des territoires et des infrastructures d'accès, transformant la fracture numérique en un phénomène à la fois économique et culturel. En **B2B**, les usages reposent sur des **solutions hétérogènes et peu standardisées**, allant d'offres premium des Big Tech pour les grands comptes à des solutions low-cost asiatiques, indiennes ou open source pour les PME et TPE, au prix de fortes tensions d'interopérabilité. Dans le domaine de la **défense**, la dépendance à un **écosystème numérique fragmenté** limite les ambitions d'autonomie technologique, favorise des systèmes semi-autonomes, modulaires et supervisés, et accentue les écarts capacitaires entre États.

HPC, IA ET EDGE : "CLOUD TO CHIP"

Approche « **Cloud to Chip** », parfois qualifiée de « **Silicon-locked Edge** », dans laquelle l'exécution des modèles d'IA est majoritairement **déportée sur les terminaux utilisateurs** (smartphones, objets connectés), sous le contrôle des **Big Tech** qui structurent l'ensemble de l'écosystème matériel et logiciel. L'**essoufflement de l'IA générative** entraîne une **concentration des investissements** sur quelques acteurs dominants, tandis que de nouveaux paradigmes (world models, approches neuro-symboliques) alimentent des cycles de hype successifs.

Les données continuent d'alimenter des **clouds propriétaires** pour la personnalisation et l'amélioration des modèles centraux. Le **Edge Computing intermédiaire** ne se généralise pas : faute de cas d'usage massifs et rentables, il reste **cantonné à des applications industrielles ciblées**, fragmentées et peu standardisées, dans un contexte où la **rentabilité du cloud centralisé** est à nouveau privilégiée.

MÉMOIRE & STOCKAGE : STAGNATION TECHNOLOGIQUE ET FRAGMENTATION DES STANDARDS

La **mémoire et le stockage** recherchent avant tout un **compromis coût-performance**. Les **technologies de mémoire chaude** évoluent lentement, freinées par des **contraintes physiques, thermiques et énergétiques** ; la **SDRAM standard (DDR5 puis DDR6)** demeure dominante, tandis que l'adoption de **CXL** reste **lente et fragmentée**, limitant l'émergence d'écosystèmes de mémoire partagée. Faute de standards unifiés, la **circularité matérielle** progresse peu et l'optimisation se fait principalement **au niveau logiciel** plutôt que par de nouvelles architectures. Côté **stockage froid**, les ruptures technologiques (ADN, etc.) peinent à se matérialiser commercialement, au profit de solutions **archivage sur bande** compétitives.

ÉNERGIE : ARBITRAGES ÉCONOMIQUES ET FIN DES GRANDS PROJETS NUCLÉAIRES

Le modèle énergétique est caractérisé par un **marché tendu et très volatil**, sous l'effet de **crises économiques successives**, ce qui **ralentit fortement les investissements de long terme**, notamment dans le **nouveau nucléaire** et les **réseaux énergétiques**, au profit d'une exploitation accrue des **infrastructures existantes et centralisées**. La **construction de SMR pour les datacenters est abandonnée**, en raison de **CAPEX élevés** et de **rentabilités insuffisamment démontrées**.

Les **groupes électrogènes de secours** restent dominants pour assurer la continuité d'alimentation, reposant sur le **gaz** et des carburants alternatifs. Les stratégies d'approvisionnement deviennent **opportunistes et différenciées selon les pays**, avec le développement de **mécanismes de flexibilité des datacenters**, tels que le « **Bring Your Own Power** », dans les zones sous tension réseau.

EAU & REFROIDISSEMENT : OPTIMISATION DE L'EXISTANT

Le **CAPEX devient le facteur déterminant**, conduisant à privilégier des **solutions éprouvées, peu coûteuses et compatibles avec les infrastructures existantes**. Le **free cooling à air** et le **refroidissement adiabatique** demeurent dominants, non par performance optimale mais par **avantage économique**, les acteurs cherchant avant tout à **optimiser l'existant** plutôt qu'à engager des refontes lourdes.

Cette logique se traduit par une **stagnation des performances moyennes**, avec un **PUE/WUE élevé** et une forte **hétérogénéité du parc**, majoritairement composé de systèmes vieillissants. Les **technologies de refroidissement liquide** (DLC, immersion) restent **marginales**, perçues comme coûteuses et risquées, et sont cantonnées à **quelques niches HPC** où la densité l'impose.

En l'absence de pression réglementaire forte, la **consommation d'eau reste mal maîtrisée**, les tours adiabatiques étant largement utilisées, ce qui maintient un **niveau élevé de WUE à l'échelle de l'industrie**.

TÉLÉCOM : VERS DE SIMPLES FOURNISSEURS D'INFRASTRUCTURES ?

Les **télécoms** connaissent une **explosion des volumes de trafic** sans capturer la **valeur ajoutée correspondante**, celle-ci étant majoritairement absorbée par

les **Big Tech** et les grandes plateformes **OTT** qui maîtrisent la couche applicative et l'IA. Les opérateurs tendent à se **replier vers un rôle d'infrastructures de connectivité**, avec des marges limitées, ou à intégrer les télécommunications comme **brique interne** au service d'acteurs disposant de leurs propres besoins critiques (IoT industriel, réseaux privés).

La valeur se déplace vers les plateformes capables de **court-circuiter les opérateurs traditionnels**, notamment via l'acquisition directe de spectre ou d'infrastructures. Dans ce contexte, les opérateurs télécom se contentent souvent de **louer leurs sites et leurs antennes** aux hyperscalers pour des offres d'edge managées, sans développer de **services différenciants**, accentuant leur dépendance stratégique.

BLOCKCHAIN & CRYPTO : LE « CRYPTO-REVIVAL »

Dans ce scénario, la **blockchain connaît un regain d'intérêt opportuniste**, porté par un **Bitcoin redevenu actif refuge** dont la valeur élevée maintient un **incitatif économique fort**, réservé aux **acteurs les plus capitalisés** et encadré par des **cadres réglementaires plus clairs**. Le **minage** s'intègre comme une **charge de travail flexible**, mobilisée pour **optimiser l'utilisation des infrastructures de calcul** lorsque la demande en IA ou en services cloud ralentit.

Dans les zones à **forte volatilité des prix de l'électricité**, il devient un **workload tampon**, activé lors des surplus ou des prix négatifs et combiné à des mécanismes de **demand response**. Certains opérateurs se positionnent ainsi comme de véritables **arbitres énergétiques**, capables de **basculer en temps réel entre calcul IA, services cloud et minage de Bitcoin** par **effet d'aubaine**, afin de **maximiser la valeur du MWh et le taux d'utilisation des actifs** des datacenters.

SCÉNARIO « NOUVELLE DONNE NUMÉRIQUE EUROPÉENNE »

PITCH :

« Suite à une politique américaine trop hégémonique s'inscrivant dans un conflit économique majeur avec la Chine et une Russie belliqueuse, l'Europe a pris le virage d'une relance technologique maîtrisée, politique industrielle souveraine avec un numérique piloté. Des fonds conséquents pour le financement des infrastructures numériques ont été débloqués en s'appuyant sur des champions européens. La dépendance aux Big Tech est toujours présente mais atténuée avec des règlements stricts auxquels ils doivent se plier : souveraineté, décarbonisation, etc. L'Europe arrive à pousser l'EDGE en s'appuyant sur un dynamisme économique et un réseau d'entreprises Européennes »

Légende :

Le scénario du « Nouvelle donne numérique Européenne » est caractérisé par une rationalisation des usages avec une consommation internet contenue. Cela n'exclue pas le développement de la robotique, jumeaux numériques ou des véhicules autonomes, mais dans un cadre régulé. Des règles strictes concernant la valorisation de la chaleur où les consommations d'eau sont appliquées aux datacenters dont l'implantation est privilégiée dans des logiques d'écologie industrielles. On favorise les synergies avec les acteurs du territoire pour mutualiser les ressources là où cela fait sens. Ainsi l'exemple d'un « eco-park » est illustré, avec des moyens de productions d'énergie renouvelables et de stockage comme les BESS. Des piles à combustibles sont utilisées comme source de flexibilité énergétique du fait d'un site de production d'hydrogène. La chaleur est également valorisée pour les besoins des acteurs du site : serres agricoles, scieries, stations d'épuration, etc. Le continuum Edge-to-cloud s'est développé avec des solutions et standards européen qui privilégient l'open-source et la multiplicité des solutions. Des agents intelligents permettent une coordination et une allocation dynamique des ressources (calcul, data et énergie) favorisant la souveraineté européenne.

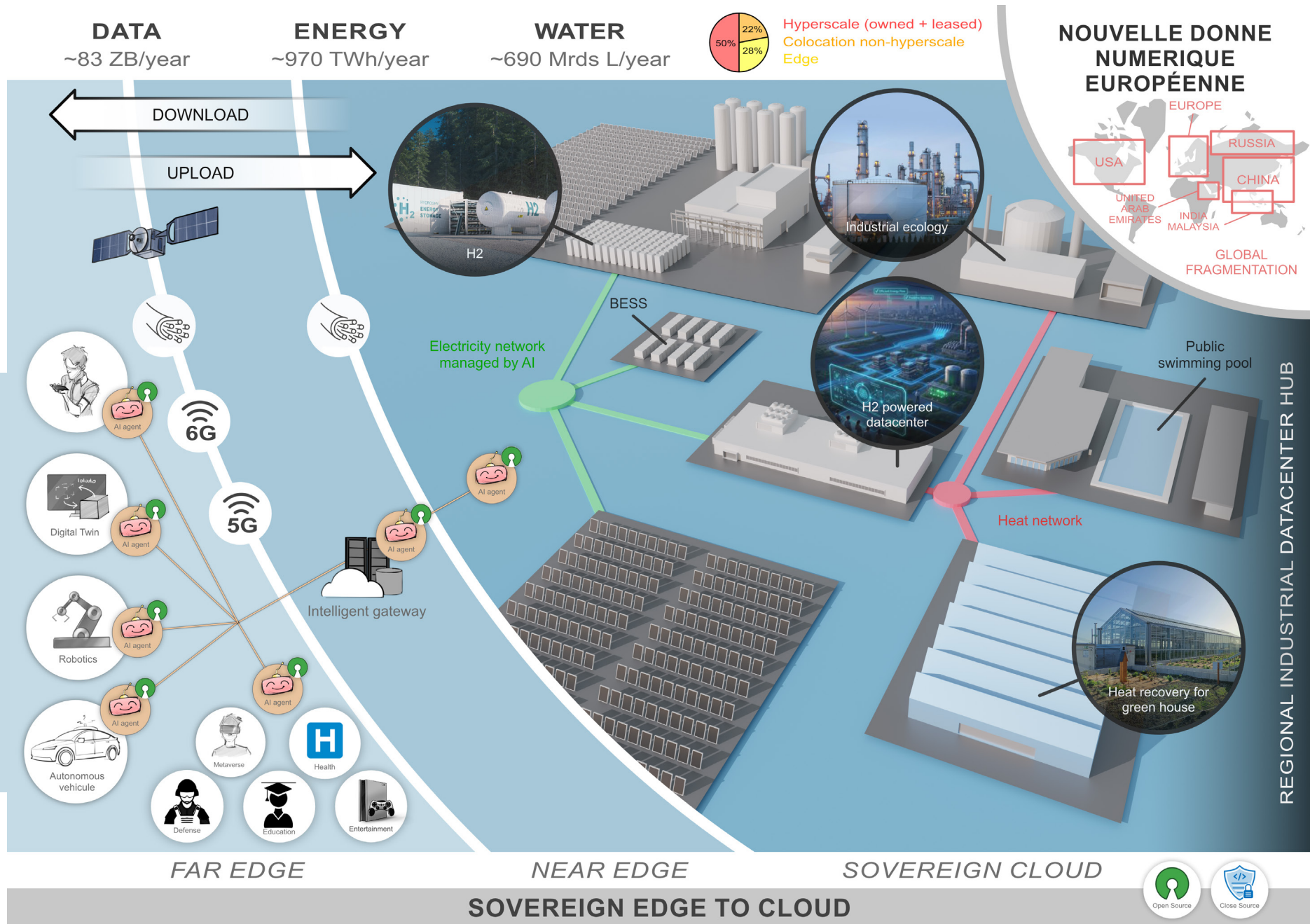


Figure 25: Scénario «Nouvelle donne numérique Européenne»

GÉOPOLITIQUE : L'ÉLECTROCHOC GÉOPOLITIQUE ET LE LANCEMENT D'UN «DIGITAL DEAL» EUROPÉEN

L'attaque des pays baltes par la Russie sans une riposte conséquente des USA crée un électrochoc majeur qui révèle **la vulnérabilité critique de la sécurité et des infrastructures numériques européennes** et contraint l'Europe à un choix stratégique explicite. Cette prise de conscience déclenche une **recomposition politique rapide** et le lancement d'un **plan massif de souveraineté stratégique**, combinant défense, énergie et numérique, qui permet une **reconvergence des États membres** et une capacité accrue d'action collective.

Progressivement, l'Union renforce son **autonomie technologique**, impose un cadre réglementaire structurant aux acteurs globaux et développe un **modèle numérique éthique, sécurisé et décarboné**. Dans un monde devenu **tripolaire**, les États-Unis conservent leur primauté militaire, la Chine reste une puissance économique majeure, et **l'Europe s'affirme comme une puissance normative globale**, exportant ses standards et ses règles auprès des économies émergentes.

DATA CENTER ET CLOUD : UNE ARCHITECTURE « REGION-CENTRIC »

La dynamique du cloud évolue vers un **modèle region-centric**, optimisé à l'échelle des **espaces souverains européens** et visant à **réduire les dépendances et les flux transcontinentaux**. Les **Big Tech américaines adaptent leurs offres** pour rester présentes sur le marché européen, en se conformant à des **exigences strictes de souveraineté, de sécurité et de conformité réglementaire**. Parallèlement, les **acteurs européens du cloud souverain** et les **data spaces européens** se structurent autour d'architectures interopérables, soutenant des **usages industriels, urbains et robotiques à faible latence**.

Cette dynamique conduit à une **coexistence durable entre cloud régional, infrastructures distribuées et maintien ciblé de solutions on-premise**, lorsque les contraintes de temps réel, de contrôle ou de résilience l'exigent.

USAGES B2C, B2B ET DÉFENSE

Les usages numériques évoluent vers un **modèle plus responsable, régulé et souverain**, sous l'impulsion d'un cadre européen renforcé sur la gouvernance des données, la cybersécurité et la soutenabilité environnementale.

En **B2C**, les consommateurs adoptent progressivement des **services labellisés européens**, fondés sur la transparence, la protection des données

et la décarbonation des usages, tandis que les grandes plateformes extra-européennes reculent sur le marché européen sans disparaître totalement.

En **B2B**, la numérisation repose sur un **écosystème interopérable de clouds et de data spaces européens**, favorisant des IA sectorielles, l'automatisation sur mesure et la co-construction de solutions, faisant de l'interopérabilité un **avantage compétitif stratégique**.

Dans le domaine de la **défense**, l'Europe consolide une **souveraineté technologique partielle mais structurée**, s'appuyant sur des infrastructures numériques communes, des programmes industriels coordonnés et une **IA de défense assistive, modulaire et conforme aux principes éthiques européens**, tout en maintenant des relations contractuelles maîtrisées avec les Big Tech étrangères.

HPC, IA ET EDGE : « SOVEREIGN EDGE TO CLOUD » :

A la différence du scénario de croissance accélérée, le continuum EDGE to Cloud est **interconnecté et piloté par des standards ouverts européens**, conçu comme une **alternative structurée au modèle centralisé des Big Tech**.

L'Europe s'appuie sur son **maillage télécom** pour déployer des **micro-datacenters et points de présence régionaux**, garantissant le **traitement local des données critiques** et l'interopérabilité via des cadres communs tels que **Gaia-X et Eurostack**. Les usages d'IA s'orientent vers des **architectures agentiques distribuées**, fondées sur des **protocoles ouverts** et des mécanismes d'orchestration multi-fournisseurs, limitant le verrouillage technologique.

Les **data spaces européens et l'apprentissage fédéré** s'imposent comme des briques clés, tandis que le **HPC et certaines capacités GPU** restent partiellement hors du cloud public pour des raisons de souveraineté et de temps réel. Cette dynamique s'appuie sur **l'émergence d'un écosystème matériel et logiciel européen**, associant cloud régional, open hardware et infrastructures publiques mutualisées, avec des gains d'efficacité énergétique progressivement intégrés aux architectures d'IA.

MÉMOIRE & STOCKAGE : «SOVEREIGN MEMORY» BASÉE SUR DES STANDARDS OUVERTS

La **mémoire et le stockage** s'inscrivent dans une **stratégie européenne de standards ouverts**, conciliant performance, durabilité et souveraineté. Pour les

données chaudes, le **CXL** s'impose comme socle d'architectures **désagrégées et interopérables**, permettant une **allocation dynamique de la mémoire**, une meilleure **efficacité énergétique** et une **circularité accrue** des infrastructures, soutenues par des projets européens de **production locale**.

Pour le **stockage froid**, des innovations européennes réduisent fortement le **coût marginal et l'empreinte énergétique**, avec le développement de **solutions de stockage passif** à très faible impact.

ÉNERGIE : FLEXIBILITÉ DES DATACENTERS ET VALORISATION ÉNERGÉTIQUE TERRITORIALE

Le modèle énergétique repose sur un **mix souverain et bas carbone**, combinant **nucléaire et énergies renouvelables**, complété par l'**hydrogène bas carbone** afin de réduire la dépendance aux importations fossiles et de sécuriser l'approvisionnement. Les datacenters sont pleinement intégrés aux **politiques énergétiques européennes**, avec un recours accru aux **PPA**, au **stockage** et à des mécanismes de **flexibilité valorisée économiquement** (BESS, pilotage de la charge, effacement).

Cette approche favorise le développement de **hubs territoriaux d'écologie industrielle**, dans lesquels les datacenters contribuent à la **valorisation énergétique locale**, notamment via la chaleur fatale. L'ensemble s'inscrit dans un **cadre européen structurant d'efficacité numérique et énergétique**, associant régulation, investissements coordonnés et maîtrise des prix de l'énergie à l'échelle de l'Union.

TÉLÉCOM : DES ACTEURS EUROPÉENS REDEVENUS CLÉS DE LA « SOUVERAINETÉ »

Les **opérateurs télécoms européens**, propriétaires des réseaux, s'affirment comme des **acteurs clés de la souveraineté numérique**, en élargissant leur rôle au-delà de la connectivité pour proposer des **services Edge de bout en bout**, fondés sur des **équipements et logiciels majoritairement européens**.

Ils jouent un rôle central dans l'**application des règles de résidence et de souveraineté des données**, en partenariat avec des **clouds locaux** et des initiatives paneuropéennes. Positionnés comme des **hébergeurs de confiance de l'Edge européen**, ils sont **incités et soutenus** pour investir dans un écosystème télécom-cloud cohérent à l'échelle européenne.

EAU ET REFROIDISSEMENT : DES NORMES POUR L'EFFICACITÉ ET LA VALORISATION DE CHALEUR

Le refroidissement est piloté par une **politique industrielle et réglementaire européenne exigeante**, plaçant l'**efficacité, la circularité et la souveraineté** au cœur de la conception des datacenters. Des **normes strictes sur l'énergie et l'eau** imposent des **PUE bas** et des **WUE fortement contraints**, faisant de la **valorisation de la chaleur fatale** un levier central d'intégration territoriale. Les techniques de **Direct Liquid Cooling** et de **refroidissement par immersion**, y compris **diphasique**, sont largement déployées, à la fois pour **refroidir efficacement les charges denses** et pour **faciliter la récupération et la valorisation de la chaleur**. Les systèmes consommateurs d'eau sont progressivement écartés, au profit de **boucles fermées sans prélèvement hydrique**, complétées par des exigences de **traitement et de réutilisation des eaux**. Le refroidissement devient ainsi une **brique structurante de l'écologie industrielle**, intégrée aux réseaux de chaleur et aux écosystèmes locaux.

CRYPTO & BLOCKCHAIN : REGISTRES DISTRIBUÉS AU SERVICE DE LA SOUVERAINETÉ

Les **technologies de registres distribués** sont intégrées à l'**Edge européen** comme **infrastructures de confiance**, au service de l'**interopérabilité**, de la **gestion des identités**, de la **traçabilité des données** et du **partage distribué des ressources de calcul**. La **blockchain** est ainsi mobilisée comme une **brique fonctionnelle** des écosystèmes numériques souverains, notamment pour l'orchestration de capacités Edge et la gouvernance des data spaces. Le **minage de cryptomonnaies** est **strictement encadré** et cantonné à des usages résiduels, n'étant autorisé que s'il respecte des **critères exigeants de décarbonation et d'utilité systémique**. Les mécanismes de **proof of stake** s'imposent, tandis que les approches **proof of work** sont progressivement marginalisées, alignant les usages blockchain avec les objectifs européens de **sobriété énergétique, de souveraineté et de conformité réglementaire**.

SCÉNARIO « RÉSILIENCE ET SOBRIÉTÉ »

PITCH :

« De graves crises climatiques et de conflits armés ont touché l'ensemble des pays du monde qui se fragmente. Les pays essaient de réguler les usages du numérique en développant des infrastructures résilientes, sobres, locales et durables. On assiste à un ralentissement des usages de la donnée, avec un recentrage sur les besoins essentiels (santé, environnement, éducation, etc.). Les innovations « right tech » se multiplient, soutenues par la puissance publique pour répondre aussi aux nouvelles contraintes de protection des données et de sobriété globale. Les Big Tech développent également des solutions plus frugales et efficaces. Les axes de recherches dans le monde sont dirigés pour une IA de confiance et frugale dans un monde multipolaire et fragmenté »

Légende :

Dans le scénario résilient et sobre, les hyperscalers sont limités au strict minimum et on privilégie les petits et micro-datacenters plus autonomes en énergie. Des règles strictes s'appliquent sur la consommation d'eau des datacenters où on privilégie les approches « zero water ». Les progrès dans la valorisation de la chaleur et d'un hardware dédié pouvant accepter des températures de fonctionnement plus élevées permettent d'intégrer des datacenters dans des immeubles d'habitation. Une partie de l'eau chaude sanitaire peut être ainsi produite pour les habitants et les constructions sont conçues dès le début pour pouvoir accueillir des datacenters. Les énergies renouvelables et la flexibilité énergétique sont priorisées au sein de micro-grid énergétiques dans lesquelles sont intégrés les datacenters. L'usage de la blockchain permet de partager des ressources de calcul ou de l'énergie avec les DEPIN (Decentralised Physical Infrastructures). La résilience des infrastructures est privilégiée avec des réseaux énergétiques et télécommunications doublés qui peuvent avoir des usages duaux en cas de conflits avec des systèmes capables de fonctionner en mode dégradé. Les usages de l'internet sont priorisés ainsi pour la défense, la santé ou l'éducation. L'architecture Edge s'appuie sur le développement d'une IA frugale, souveraine utilisant des briques open-source, avec un apprentissage décentralisé.

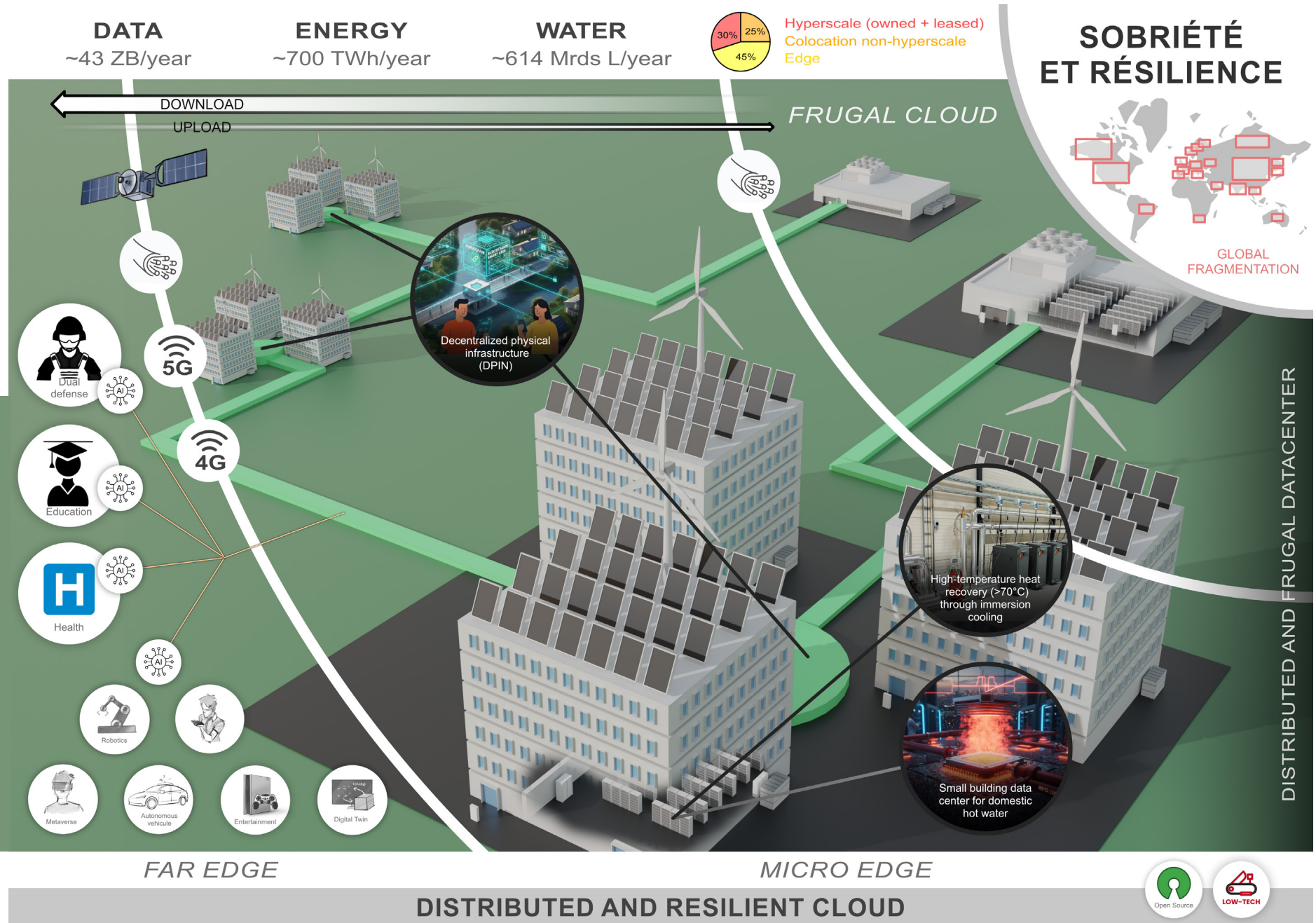


Figure 26 : Scénario «Résilience et sobriété»

GÉOPOLITIQUES : UN « ARCHIPEL DIGITAL » FRAGMENTÉ PAR LES CRISES CLIMATIQUES

Dans le scénario de **résilience**, l'environnement géopolitique est profondément reconfiguré par des **crises climatiques et sanitaires répétées**, entraînant une **démondialisation contrainte** et la fragmentation durable des échanges. La multiplication des chocs climatiques, des pandémies régionales et des ruptures d'approvisionnement conduit les États à **prioriser la sécurité alimentaire, énergétique et numérique**, réduisant fortement les interdépendances internationales. Les infrastructures numériques deviennent **directement exposées aux contraintes physiques et énergétiques**, remettant en cause la soutenabilité du modèle de croissance du numérique.

Dans ce contexte, un **nouveau paradigme de sobriété numérique** s'impose progressivement, sous l'impulsion européenne, à travers des normes contraignantes de sobriété, de relocalisation et d'usage prioritaire des énergies renouvelables. Le système international évolue vers un **« archipel » de souverainetés technologiques**, composé de blocs régionaux autonomes, inégalement dotés en ressources, où la **résilience prime sur la performance** et où l'instabilité géopolitique demeure élevée.

DATA CENTER DÉCARBONÉS, DISTRIBUÉS ET RÉSILIENTS

Dans le scénario de **résilience**, la dynamique du cloud s'organise autour d'un **modèle distribué et largement auto-contenu**, où les flux vers le cloud central sont **réduits aux fonctions essentielles de synchronisation et de résilience**. Le traitement des données est majoritairement **local**, fondé sur des **protocoles légers open source** et des volumes minimaux d'échange. Les infrastructures privilégient des **micro-datacenters territorialisés et des architectures edge**, intégrés aux écosystèmes locaux et optimisés pour la **sobriété énergétique**.

L'accent est mis sur **l'allongement de la durée de vie**, le **reconditionnement**, les **architectures électriques en courant continu** et une **approche cycle de vie globale**. Les hyperscalers restent présents mais **sous contraintes réglementaires strictes**, dans un paysage dominé par un **numérique frugal et résilient**.

USAGES B2C, B2B ET DÉFENSE

Les usages de la donnée évoluent vers un **numérique rationné et recentré sur l'essentiel**, considéré comme un **bien commun soumis à des contraintes fortes de ressources**.

En **B2C**, les usages sont hiérarchisés selon les besoins fondamentaux, la sobriété devenant un principe structurant sous l'effet de la régulation, de la tarification progressive et de la fin des usages illimités ; les citoyens privilégient des services à **faible impact environnemental**, des terminaux durables et réparables, et des **IA embarquées fonctionnant hors cloud**, tandis qu'une minorité accède à des services premium à coût élevé.

En **B2B**, les entreprises abandonnent la logique du Big Data au profit de stratégies de **Smart Data**, centrées sur la valeur, la traçabilité et l'efficacité, intégrant des solutions IT for Green, l'energy harvesting et des indicateurs d'impact environnemental des usages numériques, faisant de la sobriété un levier de performance et de compétitivité.

Dans le domaine de la **défense**, la priorité est donnée à des solutions **sobres, interopérables et autonomes**, reposant sur des architectures distribuées, des technologies open source, des IA locales et supervisées, et des infrastructures résilientes capables de fonctionner en mode dégradé, avec un recours limité aux solutions high-tech réservées aux segments critiques.

HPC, IA ET EDGE : IA FRUGALE ET EMBARQUÉE AU PLUS PRÈS DES BESOINS

L'architecture IA-Edge-HPC est orientée vers la **résilience locale et l'autonomie**, avec des systèmes capables de fonctionner en **mode dégradé** et une dépendance minimale au réseau global, utilisé comme **ultime recours**. Le Edge repose sur une **mosaïque de solutions ouvertes et frugales**, majoritairement **open source**, développées par des écosystèmes multiples, offrant une forte adaptabilité au prix de performances hétérogènes. Les entreprises privilégient des **solutions low-power, locales et interopérables**, y compris issues d'acteurs non occidentaux, tandis que la recherche se concentre sur une **IA frugale, explicable et embarquée**, dédiée à des usages essentiels (santé, énergie, maintenance). Les modèles sont **drastiquement réduits**, l'inférence limitée et planifiée, et **l'apprentissage fédéré** devient la norme afin de concilier utilité, sobriété énergétique et continuité de service.

MÉMOIRE ET STOCKAGE : LA SOBRIÉTÉ AVANT LA PERFORMANCE

La mémoire et le stockage sont guidés par une logique d'**efficacité énergétique**, visant le **maximum de calcul utile par watt** plutôt que la performance brute, avec des équipements conçus pour **durer et consommer peu**. Pour les **données chaudes**, l'innovation se déplace vers le **logiciel (« code over silicon »)** afin de

réduire les transferts coûteux en énergie, via compression, déduplication et IA conscientes de la hiérarchie mémoire ; des **mémoires basse consommation** accompagnant des **ASIC spécialisés** remplacent les architectures généralistes. Les **standards ouverts** sont privilégiés pour la **réparabilité et la résilience**, permettant l'interopérabilité et l'usage de matériel reconditionné. Pour les **données froides**, la priorité est à la **réduction des volumes**, à des solutions **très sobres et peu coûteuses**, et à des politiques actives de sélectivité et d'effacement, compensant la stagnation matérielle par des gains logiciels

ÉNERGIE : AUTONOMIE LOCALE ET SOBRIÉTÉ DES NŒUDS NUMÉRIQUES

Le modèle énergétique est fondé sur l'**autonomie locale des nœuds numériques** et la **sobriété des équipements**, avec un Edge conçu pour fonctionner avec des **apports énergétiques minimaux**, issus de **sources locales, décentralisées et décarbonées**. Le pilotage repose sur des **micro-réseaux**, l'**autoconsommation ENR** et une gestion fine de la demande, dans un contexte de **tensions électriques récurrentes** et de **black-out régionaux** liés aux canicules et aux crises hydriques.

Une **tarification dynamique du calcul** est mise en place, limitant l'accès aux ressources numériques en période de contrainte réseau, tandis qu'une **fiscalité dissuasive sur l'énergie fossile** accélère la sobriété des usages. Les datacenters sont fortement incités à la **flexibilité**, avec l'abandon progressif des groupes électrogènes au profit de **batteries locales robustes** – notamment **LFP** – et de productions renouvelables. Les **solutions hydrogène** sont écartées, jugées **trop coûteuses et centralisées** pour un modèle Edge **distribué et résilient**.

EAU ET REFROIDISSEMENT : VERS LE « ZERO WATER »

Le refroidissement privilégie des solutions **sobres et résilientes**, la **consommation d'eau devenant une contrainte réglementaire forte**. Le **refroidissement par immersion** est favorisé pour son **efficacité sans prélèvement hydrique**, sa robustesse en environnements contraints et sa capacité à **prolonger la durée de vie des équipements**, le **Direct Liquid Cooling** jouant un rôle complémentaire.

Le **free cooling** est limité par les **canicules récurrentes**. Des **seuils de WUE réglementaires** et des mécanismes fiscaux pénalisent les systèmes intensifs, favorisant des **performances hydriques élevées**. La **valorisation de la chaleur fatale est systématique** et conditionne toute implantation, dans une **approche territoriale et circulaire** fondée sur des **micro-datacenters intégrés aux**

bâtiments et aux usages locaux.

TÉLÉCOM : MUTUALISATION D'INFRASTRUCTURES DURABLES POUR PLUS DE RÉILIENCE

Les **opérateurs télécoms** font face à des **coûts croissants**, mais engagent une **mutation profonde** alignée avec la **transition écologique et la résilience des réseaux**. Ils privilégient la **mutualisation des infrastructures**, tant passives qu'actives, afin de réduire les investissements et l'empreinte environnementale.

Des **partenariats étroits avec les filières d'énergies renouvelables** se développent pour sécuriser l'approvisionnement énergétique des réseaux et limiter les émissions. En parallèle, les telcos s'orientent vers des **solutions open source et du matériel standardisé**, réduisant la dépendance aux équipements propriétaires. Enfin, la **résilience** devient un axe majeur : déploiement de **sites autonomes** (batteries locales, ENR, micro-réseaux) avec des capacités de **fonctionnement en mode dégradé** et priorisation des services critiques afin de **limiter la dépendance aux réseaux électriques centralisés**.

CRYPTO & BLOCKCHAIN : BLOCKCHAIN VERTE POUR UN CLOUD DÉCENTRALISÉ

les **technologies blockchain** s'inscrivent dans une logique de **sobriété et de résilience**, en s'appuyant sur des approches de **DEPIN (Decentralized Physical Infrastructure Networks)** pour mutualiser et orchestrer des **infrastructures physiques distribuées**. Le **minage de Bitcoin**, jugé **incompatible avec les contraintes énergétiques**, est largement écarté au profit de **mécanismes de consensus sobres** (proof of stake, proof of work allégé).

Les modèles de **cloud distribué décentralisé** reposent sur la **réutilisation de matériel reconditionné**, l'**activation locale de capacités de calcul, de stockage et de connectivité non utilisées**, et une **gouvernance par registres distribués**, permettant d'optimiser les ressources existantes, de réduire les besoins énergétiques et de renforcer l'**autonomie territoriale et la résilience systémique** des infrastructures numériques.

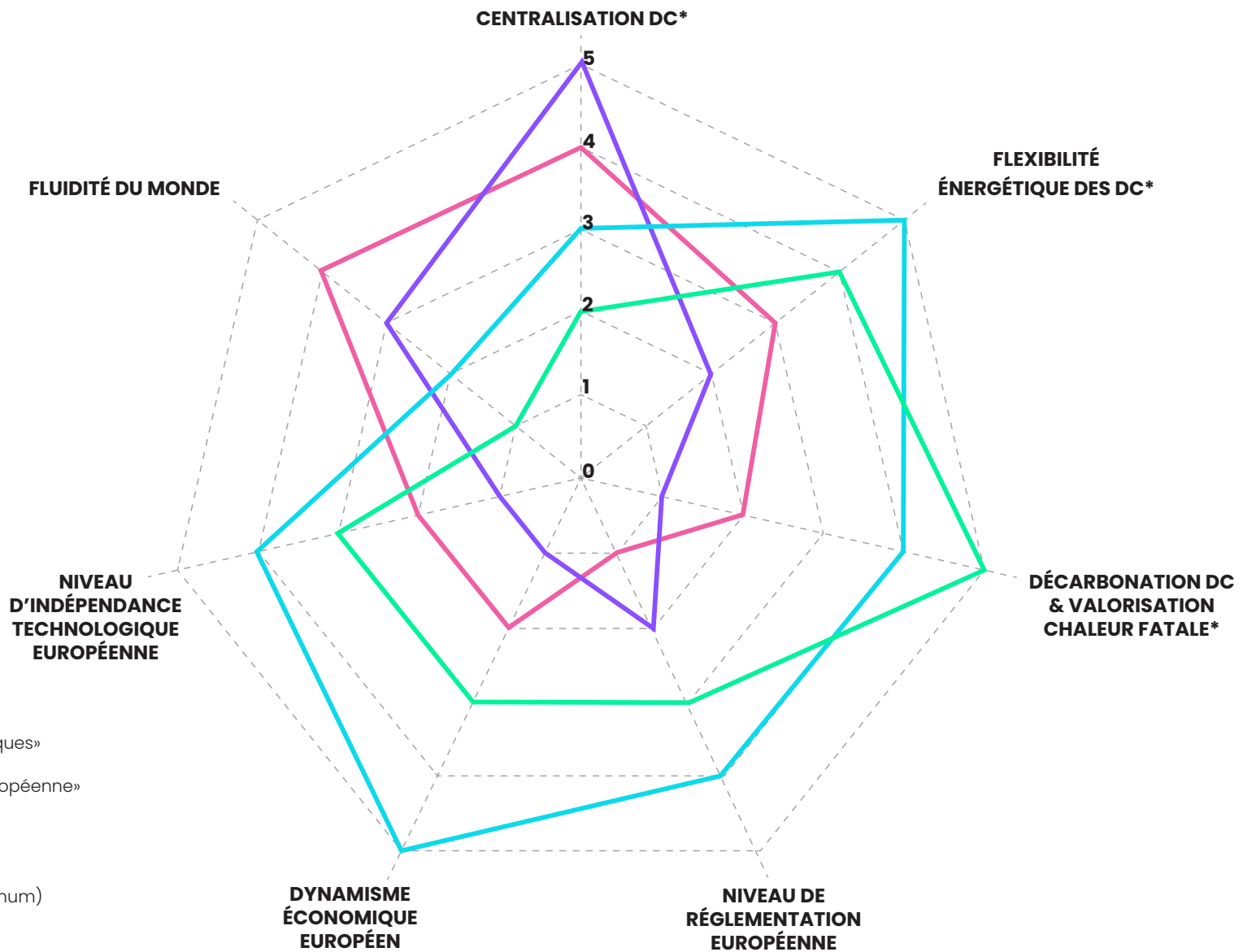


Figure 27 : Diagramme de comparaison multicritère des scénarios DATA 2040

CONCLUSIONS

QUATRE TRAJECTOIRES CONTRASTÉES MAIS INTERDÉPENDANTES

L'étude DATA 2040 met en évidence **quatre futurs contrastés**, allant d'une croissance accélérée portée par l'IA à un modèle de sobriété territoriale et résiliente. Entre ces pôles, se dessine progressivement une vision plus systémique où calcul, données et énergie s'ajustent dynamiquement aux besoins, aux contraintes physiques et aux tensions géopolitiques.

	CROISSANCE ACCÉLÉRÉE	INSTABILITÉ ET CYCLES TECHNOLOGIQUES	NOUVELLE DONNE NUMÉRIQUE EUROPÉENNE	RÉSILIENCE & SOBRIÉTÉ
LOGIQUE DOMINANTE	Expansion illimitée des usages et IA-first	Multi-crisés et spéculation (bulles et corrections)	Régulation coopérative et autonomie stratégique	Frugalité choisie, sobriété fonctionnelle
ARCHITECTURE	Cloud-to-Edge mondial hypertrophié	Cloud centralisé avec à-coups	Cloud fédéré, dataspace, hybride maîtrisé	Edge local autonome, microgrids numériques
ÉNERGIE	Surconsommation gérée par la flex	Dépendance opportuniste aux prix	Décarbonation planifiée, compute fléché	Autonomie locale, couplage fort énergie-data
GOVERNANCE	Privée, Big Tech dominantes	Fragmentée, dominée par les marchés	Publique/partenaire, souveraineté ciblée	Locale, mutualisée, orientée résilience
RÉGIME D'INNOVATION	Technologique & spéculatif	Cyclique, guidé par les contraintes budgétaires et par les « buzz »	Systémique & normatif (Green Compute, dataspace)	Adaptatif, low-tech, techno-frugal

Ces trajectoires ne sont pas exclusives et dans le futur coexisteront des modèles de datacenters hyperscales organisés en campus géant, des petits datacenters plus frugaux et résilients, des hub régionaux de datacenters intégrés au sein d'un tissu d'entreprises ou finalement des campus de datacenters hétérogènes qui évoluent au fil des demandes en suivant l'évolution de technologies. Le monde de 2040 sera probablement **hybride**, mêlant des zones de surpuissance numérique (hyper-campus IA, hubs de calcul souverains, clouds globaux) et des poches de sobriété fonctionnelle, où le numérique sera dimensionné pour la résilience locale plutôt que pour la croissance illimitée.

VERS UNE INFRASTRUCTURE NUMÉRIQUE SYMBIOTIQUE ET RÉILIENTE

L'étude prospective **DATA 2040** met en lumière un point de bascule : l'infrastructure numérique va devenir une **infrastructure de plus en plus vitale**, au même titre que les réseaux d'eau, d'énergie ou de transports. Elle structure désormais la compétitivité industrielle, la souveraineté des nations et le fonctionnement quotidien de nos sociétés.

Nous entrons dans une période de transformation caractérisée par la convergence de tensions : demande énergétique, puissance de calcul notamment due à l'IA générative, ressources matérielles et naturelles dans un contexte géopolitique fragmenté. De ces analyses émergent cinq axes structurants pour l'avenir.

LE NOUVEAU PARADIGME : «COMPUTE IS A NEW OIL»

La donnée n'est plus la seule valeur ; la **capacité à la traiter devient un enjeu primordial ces dernières années**. Sous l'impulsion d'acteurs comme NVIDIA, la puissance de calcul s'impose comme un étalon de mesure de la puissance économique entre les états. Cette mutation redéfinit de plus en plus le lien entre datacenter et réseau électrique. Ainsi, le datacenter de demain ne sera plus un consommateur passif, mais un **acteur de la stabilité énergétique** :

- **stockage sur site** (batteries, hydrogène, chaleur, voire minage flexible) ;
- effacements et modulations pour accompagner les ENR ;
- **valorisation de la chaleur fatale** dans des réseaux urbains ou industriels.

Les **charges IA accentuent cette dynamique** : l'entraînement induit des plateaux de consommation élevés et prolongés, l'inférence des fluctuations rapides dépendant des usages. Les **infrastructures devront être** capables de moduler leur charge, à l'image des sociétés de minage de cryptoactifs pivotant vers l'IA (comme au Texas avec ERCOT) et **offrir une flexibilité capable d'équilibrer le réseau**.

L'ÉQUATION DES RESSOURCES ET L'ANCRAGE TERRITORIAL

La **course à la densité et à la puissance** nous confronte au **«heat wall»**. Si des solutions de refroidissement en boucle fermée ne consommant que très peu d'eau existent, la réalité économique impose souvent l'eau comme vecteur de refroidissement le moins cher, créant parfois une compétition locale pour cette ressource vitale, exacerbée par le changement climatique (cf. étude Rémi Grimaud). **L'innovation dans les technologies de refroidissement** (qui peuvent représenter jusqu'à 40 % de la consommation énergétique d'un site) et la **valorisation de la chaleur fatale** vont continuer à constituer des axes forts d'innovation dans les années futures. Cette équation impose désormais aux datacenters de **s'intégrer profondément aux territoires** et de participer à l'équilibre des écosystèmes locaux, habitants comme industriels. Il s'agit d'entrer dans une véritable **démarche d'écologie industrielle** où les déchets des uns deviennent les intrants des autres. Cette mutation oblige les acteurs de la filière à composer avec leur environnement et à **renforcer leur communication sur les enjeux du numérique**, en prenant en compte les **aspects sociétaux pour construire une acceptabilité** durable et éviter toute situation de blocage avec les populations.

UN NOUVEAU CYCLE CENTRALISATION / DÉCENTRALISATION

L'histoire du numérique est rythmée par des cycles successifs (Figure 28)

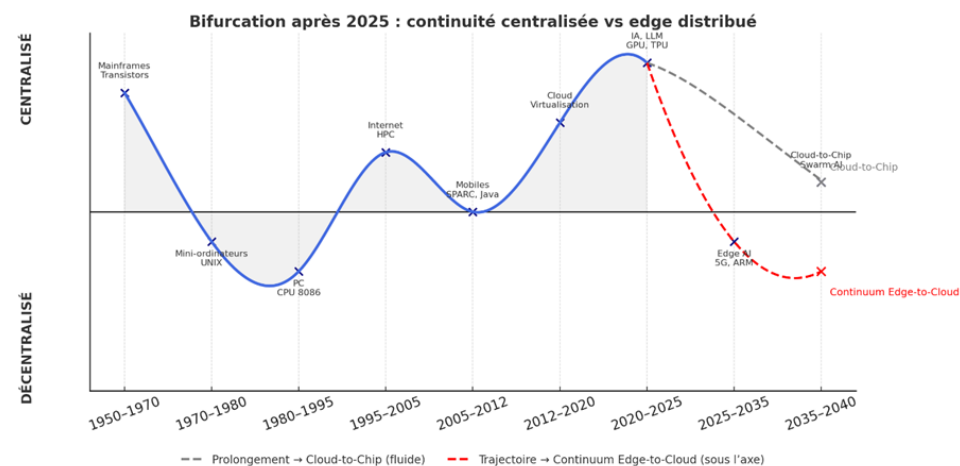


Figure 28 : Evolution des cycles de centralisation/décentralisation du numérique

La prochaine vague n'opposera pas centralisation et décentralisation, mais les combinera dans un **continuum Edge to Cloud**.

Ce continuum se fera, mais selon quelle architecture ? Ce sont les usages émergents – **robotique autonome, véhicules connectés, XR/AR, maintenance prédictive**, systèmes industriels – qui décideront de l'équilibre entre :

- un **edge distribué** proche des capteurs,
- un **cloud hyperscale** concentrant les modèles géants ou cloud souverains ou fugaux,
- des architectures MEC (Multi-access Edge Computing) ou Cloud-to-Chip, selon les besoins.

L'HYBRIDATION D'ARCHITECTURES HÉTÉROGÈNES

L'écosystème numérique bascule vers des systèmes de calcul massivement **hybrides** :

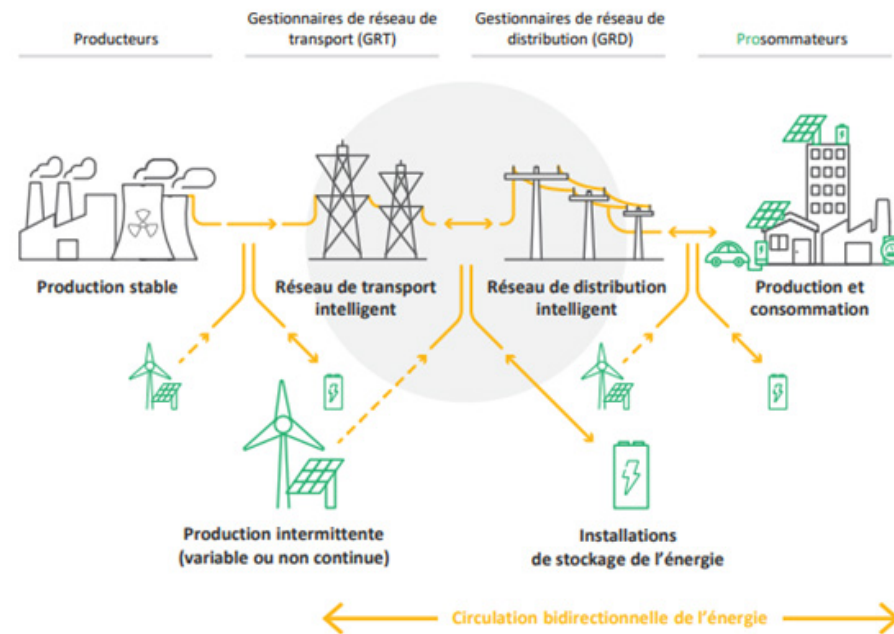
- Hybridation HPC/IA : coexistence et orchestration de processeurs hétérogènes (GPU, NPU, CPU, QPU, etc.)
- **Interconnexions photoniques** pour dépasser les limites électriques,
- **Utilisation de mémoires désagrégées** via CXL (Compute Express Link),
- Cloud hybride.

Cette hybridation touche aussi l'**organisation du stockage** de la donnée, les **réseaux télécom** (fibres, satellite et radio dans des architectures multi-couches, etc.) et les **réseaux électriques**, de plus en plus de systèmes de production/stockage variés (ENR, batteries, Stockage H2, etc.).

Ces transformations structurent un nouvel écosystème hétérogène distribué, où la performance ne dépend plus seulement de la data mais aussi du calcul, des réseaux télécom et de l'énergie. Cette convergence représente un challenge dans la gestion et l'orchestration de l'ensemble de cette **architecture systémique** digitale.

CONVERGENCE DIGITAL ET ÉNERGIE

La gestion centralisée du réseau se complexifie sous l'effet de sources d'énergie variées, parfois non pilotables, et de nouveaux acteurs. Nous passons ainsi d'une distribution unilatérale à un **échange d'énergie bidirectionnel** entre producteur et consommateur (figure 29).



Source : Cour des comptes européenne.

Figure 29 : Décentralisation des systèmes de production énergétiques – Source : rapport de la cour des comptes Européenne « Préparer le réseau électrique de l'UE à la neutralité carbone » 2025

Nous passons d'un système descendant à un «**Internet de l'énergie**» où la production est décentralisée et où les échanges se font de pair à pair (Figure 30 page suivante).

Cette convergence offre une nouvelle grille de lecture de nos scénarios :

- **Croissance accélérée & sobriété et résilience** : Tendent vers une production d'énergie distribuée sur des réseaux électriques fragmentés avec des sources hétérogènes (campus hyperscale off-grid ou micro-datacenters autonomes sur boucles locales ENR).

- **Instabilité et cycles technologiques & Nouvelle donne numérique Européenne** : Reposent davantage sur des réseaux stabilisés et pilotés de manière centralisées ou homogènes (Datacenter opportunistes utilisant la disponibilité des réseaux électriques ou hub régionaux qui participent à la stabilisation par leur flexibilité)

La capacité de l'Europe à maintenir et investir dans ses infrastructures énergétiques pour soutenir cette convergence du digital et de l'énergie sera clé.

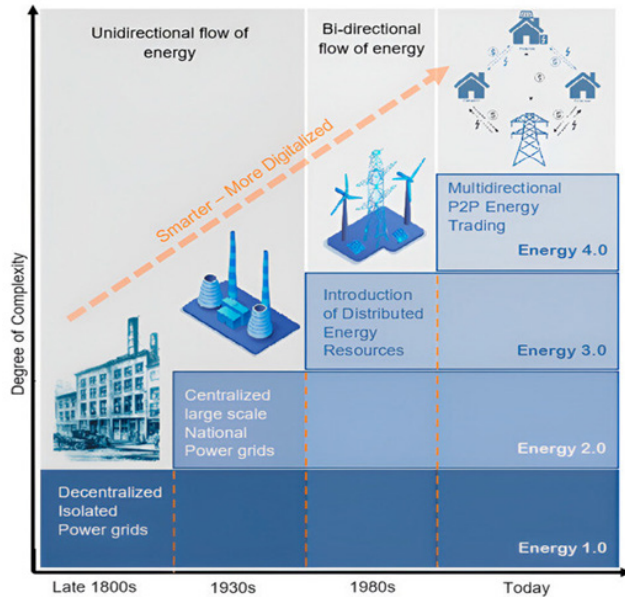


Figure 30 : Cycles de centralisation/décentralisation dans les réseaux électriques - Source : « Comparison between Blockchain P2P Energy Trading and Conventional Incentive Mechanisms for Distributed Energy Resources—A Rural Microgrid Use Case Study », A. Ghattas Aoun & Al.

Qu'elles soient centralisées ou distribuées, ces architectures qui constitueront le continuum digital/énergétique devront devenir **dynamiques** et **capables de s'adapter en continu aux contraintes énergétiques, de données, de réseaux ou de calcul**. L'horizon, tel que décrit par le concept de Next Compute Paradigm (HiPEAC), est une **infrastructure symbiotique évolutive, auto-adaptative en temps réels aux différentes contraintes**.

Ce concept de **Continuum Digital et Energétique** esquisse les contours de systèmes intelligents (agents ou autres) orchestrant en temps réel :

- L'allocation de calcul sur des sites distribués en fonction de leur charge ;
- La migration de données ou de programmes en fonction des vitesses ou de la saturation du réseau
- Le tout en fonction de la disponibilité voire du prix/décarbonation de l'énergie

Mais des freins subsistent : absence de standards,

fragmentation des écosystèmes, dépendances propriétaires, orchestration encore immature...

La question centrale demeure donc : qui captera la valeur ? Un modèle dominé par quelques plateformes propriétaires (NVIDIA, hyperscalers) ou un écosystème ouvert, fondé sur des standards partagés ?

LE DÉFI DE LA CROISSANCE : EFFICIENCE OU EFFET REBOND ?

L'IA est le moteur actuel des investissements, **mais aucune croissance exponentielle n'est physiquement durable à l'infini**. Si l'efficacité énergétique progresse, et que l'on arrive à stabiliser la consommation mondiale des datacenters, le secteur doit se prémunir contre le **paradoxe de Jevons**, où **chaque gain d'efficacité entraîne une explosion des usages**. De manière souhaitable,

la régulation de cette croissance ne pourra pas se faire uniquement par les lois du marché : elle devrait **nécessiter une coordination pour éviter la saturation des ressources**.

DE LA SOUVERAINETÉ À LA RÉSILIENCE SYSTÉMIQUE

La souveraineté technologique totale est, dans l'infrastructure actuelle, une illusion. La dépendance aux fondeurs asiatiques (TSMC, Samsung) ou aux écosystèmes logiciels et hardware américains est structurelle (Nvidia, VMWARE, Google, etc.). L'enjeu pour l'Europe n'est pas l'autarcie, mais l'**autonomie stratégique** : la capacité à choisir ses dépendances stratégiques et à sécuriser des maillons clés de la chaîne de valeur (à l'instar d'ASML, EuroHPC ou Mistral). Elle s'étend à la sphère matérielle, les datacenters devant devenir non seulement neutres en carbone, mais aussi **circulaires dans leur composition**.

DIGITAL VS ENERGIE

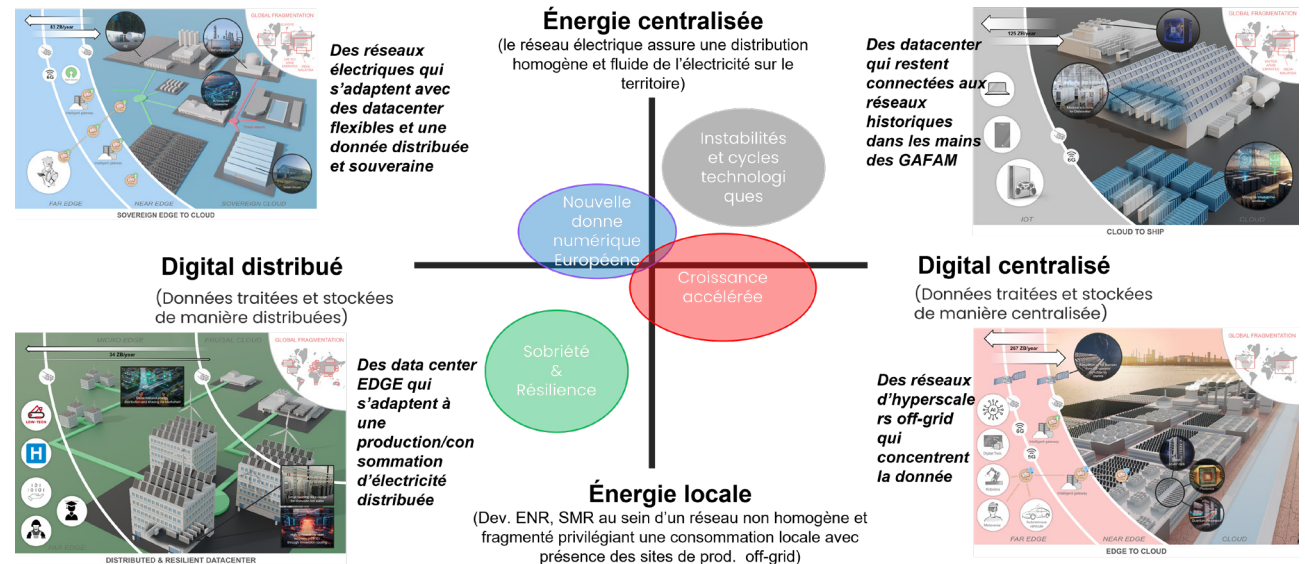


Figure 31 : Segmentation digital vs énergie des scénarios DATA 2040

Pour l'Europe, la véritable boussole pour 2040 doit être la **résilience systémique**. Face aux menaces hybrides, cybernétiques ou climatiques, nous **devons concevoir des infrastructures digitales «duales»** (civiles et défense), inspirées par exemple des modèles décentralisés (DePIN⁷⁸) ou de l'Open Source/Hardware. Il ne s'agit plus seulement de construire des datacenters, mais de bâtir **in fine** un **système digital résilient** sur lequel reposera la stabilité de nos sociétés futures.

CONCLUSION GÉNÉRALE

L'étude DATA 2040 met en lumière un **basculement de paradigme** : le numérique de 2040 ne sera pas seulement plus puissant, il sera — ou devra devenir — **« conscient » de ses limites physiques, matérielles et politiques**.

Les datacenters deviendront :

- plus denses,
- plus sobres,
- plus intégrés aux territoires et aux systèmes énergétiques,
- plus distribués,
- plus sécurisés.

Ils ne seront plus de simples "boîtes noires" de calcul, mais la **colonne vertébrale d'un « métabolisme planétaire »** où :

- le calcul sera intimement lié à l'énergie,
- l'architecture s'articulera avec le territoire,
- la souveraineté s'exprimera dans les choix de standards, de gouvernance et de répartition des risques,
- la résilience importera autant que l'efficacité.

Plusieurs chemins restent ouverts :

- Un futur **désordonné**, où l'infrastructure suivra les seuls signaux de marché et la recherche de profit à court terme, au risque de multiplier les vulnérabilités, les fractures et les impasses écologiques.
- Un futur **coordonné**, où la puissance de calcul sera pensée comme une ressource stratégique partagée, mise au service de la transition énergétique, de la connaissance scientifique et du bien commun.

L'Europe possède un mix décarboné, une culture réglementaire forte et des initiatives structurantes (CRMA, cloud de confiance, dataspace). La France, avec sa capacité excédentaire, pourrait accueillir jusqu'à 10 GW de datacenters. Mais la véritable question, pour les acteurs des datacenters, des réseaux, de l'IA et des politiques publiques, n'est donc plus :

- « Combien pouvons-nous croître ? »

mais bien :

- « Comment faire du continuum digital-énergie un système symbiotique, stable et soutenable, plutôt qu'un facteur de fragilité supplémentaire ? »

L'ambition de DATA 2040 est de contribuer à cette prise de conscience et d'offrir des repères pour **choisir le futur plutôt que le subir** : un futur où le numérique, loin d'être une fuite en avant, devient un levier de sobriété, de souveraineté maîtrisée et de résilience systémique.

Le futur n'est pas écrit, mais il peut être préparé.

78 Decentralized physical infrastructure network

RECOMMANDATIONS



Les recommandations de DATA 2040 convergent vers une même trajectoire : faire évoluer les datacenters d'infrastructures techniques optimisées localement vers des systèmes adaptatifs, intégrés et pilotés dynamiquement, capables de s'inscrire durablement dans les contraintes énergétiques, territoriales et sociétales de l'Europe. Il faut redire ce que nous avons souligné dans notre étude : c'est l'évolution des usages dont vont dépendre principalement les développements technologiques des datacenters. Déluge de données ou régulation des usages digitaux vont créer des infrastructures différentes.

DÉVELOPPER DES SYSTÈMES DYNAMIQUES D'ALLOCATION DES RESSOURCES, ORCHESTRÉS PAR L'IA ET FONDÉS SUR DES STANDARDS OUVERTS EUROPÉENS

Les scénarios de DATA 2040 montrent une alternance de phases de centralisation et de décentralisation des infrastructures numériques. Cette dynamique appelle moins une architecture figée qu'un **pilotage adaptatif des ressources** : calcul, mémoire, stockage et énergie devront être alloués de manière dynamique, en fonction des usages, des contraintes réseau et des signaux énergétiques.

Orientations recommandées :

- Développer des architectures spécifiquement optimisées pour l'inférence, appelée à devenir majoritaire.
- Soutenir des modèles d'IA frugaux et sobres, capables de contribuer à l'optimisation globale des infrastructures.
- Développer et promouvoir des standards européens ouverts pour l'orchestration dynamique des ressources (calcul, mémoire, stockage, énergie), s'appuyant sur l'IA pour piloter ce continuum numérique-énergétique.
- Faire co-évoluer software et hardware de manière coordonnées et pas de manières séparées.
- Intensifier l'autonomie stratégique et la « souveraineté » des ces briques, aussi bien au niveau logiciel que matériel.

DÉPASSER LE « HEAT WALL » PAR DES DATACENTERS THERMO-INTELLIGENTS ET INTÉGRÉS AUX ÉCOSYSTÈMES DE CHALEUR

Quel que soit le scénario envisagé, DATA 2040 met en évidence l'émergence d'un **mur thermique («heat wall»)** : l'augmentation continue des densités de calcul rend progressivement inopérantes les solutions de refroidissement conventionnelles. Le franchissement de ce seuil impose une **rupture technologique** dans la manière de gérer, valoriser et piloter la chaleur.

Orientations recommandées :

- Déployer des solutions de refroidissement liquide adaptées aux architectures haute densité afin de franchir durablement le « heat wall ».
- Élever le niveau de température de la chaleur fatale pour permettre son intégration effective dans des réseaux de chaleur urbains ou industriels.
- Adapter le design et les consommations d'eau aux zones à stress hydrique, et explorer des approches visant des datacenters à impact hydrique positif.
- Développer des outils avancés de modélisation et de jumeaux numériques afin d'identifier, territoire par territoire, les optima énergie-eau-performance.

INSCRIRE LES DATACENTERS DANS UNE PLANIFICATION ÉNERGÉTIQUE ET TERRITORIALE INTÉGRÉE, EN DÉVELOPPANT LEUR POTENTIEL DE FLEXIBILITÉ

Les datacenters ne doivent pas être pensés uniquement comme des charges électriques à raccorder, mais comme des infrastructures territorialisées, potentiellement contributrices à la flexibilité et à la stabilité du système énergétique, sous conditions économiques, techniques et réglementaires adaptées.

Orientations recommandées :

- Favoriser la co-localisation des datacenters avec des infrastructures énergétiques bas-carbone et des solutions de stockage.
- Orienter les choix technologiques de stockage d'énergie vers des solutions réduisant les dépendances critiques en matériaux, en cohérence avec les trajectoires industrielles européennes.
- Développer des maillages régionaux et des campus énergétiques permettant mutualisation et optimisation.

- Intégrer les datacenters aux stratégies de planification foncière, énergétique et hydrique.
- Étudier la participation volontaire des datacenters à des mécanismes de flexibilité (équilibre offre-demande, contraintes locales).
- Explorer des mécanismes de marché permettant d'articuler énergie, capacité de calcul et flexibilité à l'échelle nationale et européenne.

FAIRE DE LA RÉSILIENCE, DE LA CIRCULARITÉ ET DE L'ACCEPTABILITÉ SOCIALE DES FONDEMENTS DU NUMÉRIQUE SOUTENABLE

La montée en puissance des infrastructures numériques renforce les enjeux de résilience, d'autonomie stratégique et d'acceptabilité sociale. DATA 2040 montre que ces dimensions ne sont pas périphériques, mais conditionnent la viabilité de long terme des trajectoires technologiques.

Orientations recommandées :

- Renforcer la résilience des infrastructures critiques (hardware, énergie, télécom, données) par des approches systémiques plutôt que par la seule redondance.
- Allonger la durée de vie des équipements et structurer des filières de réemploi et de reconditionnement.
- Concevoir des datacenters circulaires, éco-conçus, sobres en ressources et intégrés à leur environnement.
- Développer des stratégies de « souveraineté » choisie (multi-cloud raisonné, open source, alternatives hardware & software européennes).
- Renforcer la cybersécurité, la surveillance et la sûreté des datacenter européen qui sont des cibles potentielles de notre infrastructure digitale.
- Investir dans les compétences humaines et la transparence afin de renforcer l'acceptabilité sociétale.

Concernant la partie climatique, les données sont extraites du rapport de Rémi Grimaud, sur le futur du cloud européen à l'épreuve de l'énergie et du climat.

Ce rapport fait en partenariat avec le Transition Data Lab est disponible sur le site :

<https://www.transition-data-lab.com/rapports/le-futur-du-cloud-europeen-a-lepreuve-de-lenergie-et-du-climat>





CONTACTS

Timothée Silvestre

Responsable de l'étude prospective

Timothee.silvestre@cea.fr

Sandra Tochon

Chef de projet

sandra.tochon@cea.fr

Philippe Caillol

Directeur Innovation de la Recherche Technologique du CEA

philippe.caillol@cea.fr

