

De la fraîcheur pour l'été : raisons pour lesquelles le refroidissement liquide est le prochain impératif de l'IA

Les processeurs avancés repoussent les limites thermiques dans les datacenters. Pour suivre ce rythme, les entreprises ont recours au refroidissement liquide, qui équilibre la performance, les coûts et la durabilité.

L'IA est la force motrice de certaines des charges de travail les plus exigeantes de l'entreprise. Il s'agit de diagnostiquer les maladies en temps réel, d'optimiser la logistique dans les chaînes d'approvisionnement mondiales et de stimuler la prochaine vague d'innovation produit. Mais derrière chaque information et chaque automatisation se cache un serveur fonctionnant à plein régime et générant de la chaleur.

Les unités centrales et les GPU haute performance d'aujourd'hui consomment plus d'énergie et génèrent plus de chaleur que les systèmes de refroidissement traditionnels n'étaient supposés en gérer, de par leur conception. Les ventilateurs et les dissipateurs thermiques ne parviennent pas à suivre, et les datacenters en paient le prix en termes de restriction de la performance, de gaspillage d'énergie, de niveaux sonores dangereux et de coûts d'infrastructure à la hausse.

Autrefois réservé aux supercalculateurs et aux environnements spécialisés, le refroidissement liquide est désormais adopté plus largement, offrant une réponse pratique aux exigences thermiques de l'IA. Voici un aperçu de la façon dont les systèmes de refroidissement liquide direct aident à gérer plus efficacement la chaleur, à réduire le gaspillage d'énergie et à maintenir la viabilité de l'infrastructure d'IA à grande échelle.

La crise de chaleur de l'IA

L'essor des grands modèles de langage et de l'IA agentique a fait basculer les demandes de calcul dans un territoire inexploré. Désormais, la consommation électrique des unités centrales haut de gamme dépasse régulièrement les 500 watts, et certains GPU frisent la barre des 1 000 watts. Dans le même temps, les tolérances thermiques diminuent. Alors qu'autrefois, les puces fonctionnaient en toute sécurité à 90, voire 100° C, la température maximale de certains des siliciums actuels les plus avancés ne dépasse pas 60°¹.

Le refroidissement par air traditionnel n'est tout simplement pas conçu pour cette nouvelle réalité thermique. Gérer la chaleur des processeurs modernes exige des dissipateurs thermiques massifs, des ventilateurs puissants et davantage d'espace physique, ce qui augmente les coûts et réduit la densité des serveurs. Cela se traduit par un cycle inefficace qui nuit aux performances et à l'évolutivité.

Une montée en flèche de la consommation d'énergie en résulte également. La consommation électrique des datacenters américains est passée de 76 TWh en 2018 à 176 TWh en 2023, et pourrait atteindre 580 TWh d'ici 2028.² La chaleur est devenue le goulet d'étranglement caché de l'IA et les enjeux augmentent rapidement.

Mode de fonctionnement du refroidissement liquide

Cette méthode consiste à faire circuler du liquide réfrigérant sur les composants les plus chauds d'un serveur (à savoir les unités centrales et les GPU), en vue d'absorber et d'évacuer la chaleur. Ce liquide traverse les plaques froides fixées aux composants clés, évacuant la chaleur et la transportant vers un collecteur externe, où elle est acheminée jusqu'à l'échangeur de chaleur permettant de l'éliminer. On obtient ainsi un transfert thermique plus rapide et plus efficace que celui que peut fournir le refroidissement par air.

Hewlett Packard Enterprise propose trois principales options de refroidissement liquide. Les systèmes en boucle fermée sont entièrement contenus dans le châssis du serveur, ce qui est idéal pour les mises à niveau incrémentielles de l'infrastructure existante. L'option alternative est le refroidissement liquide-air, qui emploie l'eau de l'installation pour refroidir un système de refroidissement par air connecté aux racks. Ce système transmet l'air frais aux composants informatiques les plus chauds avec plus de précision, et transfère la chaleur de l'air chaud vers l'eau d'installation. Quant au refroidissement liquide direct, il opère au niveau du rack, distribuant le liquide de refroidissement à plusieurs nœuds, et assurant une gestion thermique supérieure pour les charges de travail à haute densité et performance.

Il existe également une approche combinée qui utilise le refroidissement liquide pour les composants les plus gourmands en énergie, et l'air pour le reste. Cette conception équilibrée garantit un refroidissement efficace de chaque composant, sans surcharge du système.

Comparés au refroidissement par air, les systèmes de refroidissement liquide offrent une meilleure performance, permettent une densité de rack supérieure et limitent grandement la consommation en énergie. Les coûts d'exploitation tout comme les émissions de carbone s'en trouvent ainsi réduits.

¹ "HPE ProLiant Gen11 Servers with Direct Liquid Cooling," HPE, 2024.

² 2024 United States Data Center Energy Usage Report, Lawrence Berkeley National Laboratory, U.S. Department of Energy, December 2024



Façon dont HPE déploie le refroidissement liquide à grande échelle

HPE a conçu ses solutions de refroidissement liquide pour répondre aux exigences d'échelle, de densité et d'efficacité de l'IA moderne. Les serveurs HPE ProLiant Compute Gen12 qui prennent en charge les configurations de refroidissement liquide se trouvent au cœur de cet effort. Les systèmes en boucle fermée constituent une solution de refroidissement autonome idéale pour les unités centrales de grande puissance dans les configurations de rack existantes. Le refroidissement liquide direct va plus loin, en distribuant le liquide de refroidissement sur l'ensemble d'un rack pour gérer la chaleur à grande échelle. Ce qui s'avère idéal pour les charges de travail denses et gourmandes en GPU.

Pour les environnements présentant des charges de travail mixtes, le refroidissement adaptatif en cascade de HPE offre une approche unifiée. Ce système breveté passe dynamiquement et en temps réel du refroidissement par air à celui liquide, et inversement, selon la charge thermique des composants connectés. Il en résulte un refroidissement bien plus efficace, avec une consommation d'énergie et une complexité d'infrastructure moindres.

Quelle que soit l'approche que vous choisissez, les gains en termes de coût et de durabilité sont substantiels, offrant un moyen de réduire la consommation d'énergie sans sacrifier les performances. Cela fait de cette démarche une option stratégique pour les entreprises désireuses de concilier les exigences opérationnelles et les objectifs environnementaux. Dans un déploiement de 10 000 serveurs, le refroidissement liquide direct de HPE rend possible une économie de plus de 2 millions de dollars par an en coûts de refroidissement globaux, tout en réduisant de 17 millions de livres les émissions de CO₂ annuelles.³

Ces systèmes sont déjà employés dans les supercalculateurs HPE Cray 100 % sans ventilateur et les plateformes HPE ProLiant XD, prenant en charge les déploiements haute densité où le refroidissement traditionnel s'avérerait sinon insuffisant. Tandis que les charges de travail d'IA continuent de s'étendre, le refroidissement liquide offre un moyen pratique de répondre aux exigences thermiques sans repenser l'ensemble du datacenter, car il ne nécessite pas de refonte complète. Les entreprises peuvent introduire graduellement le refroidissement avancé là où il est le plus nécessaire (les charges de travail à haute densité ou à forte chaleur, notamment). Cela en fait une étape commode pour atteindre les objectifs environnementaux, sociaux et de gouvernance.

³ [Liquid Cooling: A Cool Approach for AI](#), HPE Newsroom Blog, August 2024.

Concevoir à des fins de croissance : L'approche de HPE en matière d'IA évolutive

L'innovation de HPE dans le domaine du refroidissement liquide n'est qu'un volet d'une stratégie plus large visant à pérenniser l'infrastructure d'IA. Les solutions comme HPE Private Cloud AI combinent des systèmes HPE ProLiant Compute à refroidissement liquide au calcul accéléré NVIDIA, à un réseau intégré et à un logiciel prêt pour l'IA, le tout assemblé en une plateforme clé en main. Conçus pour l'IA à l'échelle de la production, ces systèmes offrent un délai de rentabilisation plus rapide tout en simplifiant le déploiement et la mise à l'échelle.

Comme il s'agit de solutions sur site, les entreprises conservent un contrôle total sur leurs données, ce qui constitue un avantage pour celles ayant des prérequis de souveraineté, de conformité ou de faible latence. Dans le même temps, des données économiques prévisibles et une efficacité de refroidissement intégrée facilitent la gestion des coûts opérationnels à mesure que les demandes d'IA augmentent.

HPE AI Services peut aider les entreprises à planifier et à mettre en œuvre une infrastructure prenant en charge aussi bien la performance que les exigences thermiques des charges de travail d'IA. Cela inclut la sélection de l'approche de refroidissement appropriée selon l'échelle de déploiement et le type de charge de travail. Ce genre d'alignement (entre l'informatique, le refroidissement et les opérations) peut réduire la complexité et faciliter une évolution efficace au fil du temps.

Garder une longueur d'avance sur la surchauffe

L'IA ne ralentit pas, et les exigences qu'elle impose à l'infrastructure non plus. À mesure que les puces chauffent et que les charges de travail gagnent en complexité, le refroidissement traditionnel finit par ne plus pouvoir suivre. Le refroidissement liquide change la façon dont les entreprises doivent aborder la planification de l'infrastructure. En gérant la chaleur plus efficacement, il prend en charge des déploiements plus denses, réduit la consommation d'énergie et contribue à étendre la durée de vie du matériel : autant d'aspects essentiels à l'évolution durable des charges de travail d'IA.

L'avenir de l'IA s'annonce brûlant, mais grâce au refroidissement liquide, votre infrastructure n'en souffrira pas nécessairement.

Pour en savoir plus

HPE.com/ai/insights

[Visiter HPE.com](#)

[Live Chat](#)

© Copyright 2025 Hewlett Packard Enterprise Development LP. Les informations figurant dans ce document sont susceptibles d'être modifiées sans préavis. Les seules garanties relatives aux produits et services Hewlett Packard Enterprise sont stipulées dans les déclarations de garantie expresses accompagnant ces produits et services. Aucune partie du présent document ne saurait être interprétée comme offrant une garantie supplémentaire. Hewlett Packard Enterprise décline toute responsabilité en cas d'erreurs ou d'omissions de nature technique ou rédactionnelle dans le présent document.

a50013353FRE

HEWLETT PACKARD ENTERPRISE

hpe.com