

16 JUIN 2022

@ Grand Mogador City Casablanca  
1<sup>ère</sup> édition en Afrique



## L'INTELLIGENCE ARTIFICIELLE DANS LES DATA CENTERS



Animé par M.EL HITMI Mohamed Amine, Fondateur et CEO du **Bureau** d'étude et d'expertise **INGENIOS**.

Expert Data center et doctorant en **IA**

Certifié Expert ATD , ATD, et AOS à **Uptime Institute**

# SOMMAIRE :

Pourquoi l'IA dans les DC ?

L'impact de l'IA dans l'infra de la conception des DC.

L'IA au service d'un meilleur PUE.

L'IA pour une maintenance efficiente et prédictive.

L'IA importe la robotisation au service des gestions de proximités des nos DC !!!



# Pourquoi l'IA dans les Data centers ?

Loin de s'essouffler, ou être un effet de mode, l'intelligence artificielle se démocratise, et joue un rôle déterminant sur les centres de données.



# Pourquoi l'IA dans les Data centers ?

Les data centers sont devenus de plus en plus complexes. C'est une combinaison du cloud, de l'intelligence artificielle (IA) et d'autres technologies avancées exigeant plus d'espace, de puissance et de connectivité.

Les spécialistes de l'IA se sont généralement occupés d'avoir suffisamment de performances de serveur pour accomplir leur tâche, mais ce n'est que le début de la chaîne de ressources. Aujourd'hui, les entreprises commencent à se pencher sur le défi plus large de l'infrastructure pour étendre l'utilisation de l'IA.

En technologie, nous sommes habitués à des cycles **de trois à cinq ans**. Aussi complexe que soit la technologie, une grande partie de celle-ci est éphémère. Ce n'est pas le cas avec l'infrastructure **physique**. Lorsqu'une nouvelle installation est prévue, on s'attend généralement à ce qu'elle ait un cycle de vie d'au **moins vingt à vingt-cinq ans**. Cette différence crée des défis à la fois budgétaires et opérationnels.



# Pourquoi l'IA dans les Data centers ?

Les nouvelles exigences du calcul haute performance (**HPC**), de l'**IA** et d'autres technologies créent donc des problèmes de centres de données qui vont bien au-delà des logiciels.

Le défi des installations n'est pas seulement d'optimiser l'espace rack lorsque de nouveaux serveurs plus puissants sont ajoutés.

C'est au-delà des considérations impliquées dans les nouvelles demandes de puissance. Par exemple, les nouvelles demandes de puissance créent plus de chaleur, ce qui signifie de nouvelles exigences en matière de **CVC (chauffage, ventilation, climatisation)**. Le bâtiment a-t-il suffisamment **d'espace** sur le toit et la toiture peut-elle supporter un poids supplémentaire ? Ce sont des questions que les programmeurs ne considèrent généralement pas.



# Pourquoi l'IA dans les Data centers ?

Ainsi , **l'IA s'imposera dans les centres de données** , en s'appuyant sur les nouveaux systèmes développés pour **l'apprentissage en profondeur** sont eux-mêmes une **expérience d'apprentissage**.

Un système **d'analyse de puissance** qui importe des **données similaires** provenant de plusieurs centres de données contribuera à réduire les risques et à mieux contrôler les coûts des entreprises qui passent à **des solutions basées sur l'IA**.

Ainsi comprendre la consommation d'énergie, afin de mieux prolonger la durée de vie des centres de données existants et d'améliorer à la fois l'expansion de ceux-ci et le développement de nouveaux centres, est essentiel pour aider l'IA à être rentable.



# L'impact de l'IA dans l'infra et la conception des DC

*L'IA touche tous les aspects du Data center, des équipements IT serveurs, réseaux, infrastructures, environnements, efficiences énergétiques, gestions de proximité, jusqu'aux maintenances préventifs et prédictifs.*

# L'impact de l'IA dans l'infra dans la conception des DC.

Le **machine ou deep learning** peut être exécuté sur une seule machine, mais grâce à l'incroyable débit de données, il est généralement exécuté sur plusieurs machines interconnectées pour assurer une communication continue pendant les phases de formation et de traitement des données, avec une faible latence et sans aucune interruption de service.

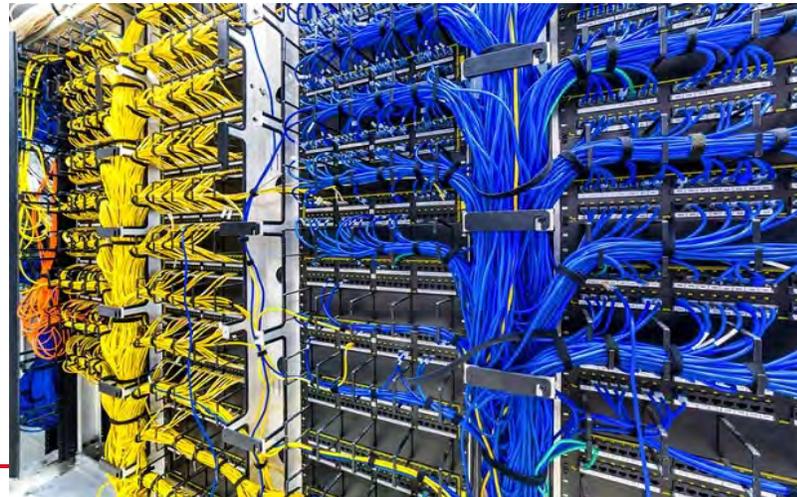
tant qu'utilisateur , notre désir d'obtenir toujours plus de données entraîne une **croissance exponentielle** de la quantité de **bande passante** nécessaire pour satisfaire nos souhaits. Cette bande passante doit être répartie à l'intérieur de plusieurs installations, mais aussi entre elles, avec des architectures complexes où les réseaux **spine-and-leaf** ne suffisent plus : nous parlons de réseaux **super-spine** et **super-leaf** pour permettre à tous les calculs algorithmiques complexes de circuler entre différents dispositifs et, finalement, retourner vers nos récepteurs.



# L'impact de l'IA dans l'infra dans la conception des DC.

**La fibre optique** est devenue le support de transmission numéro un de nos centres de données, grâce à ses capacités de haute vitesse et de très haute densité par rapport aux câbles cuivrés.

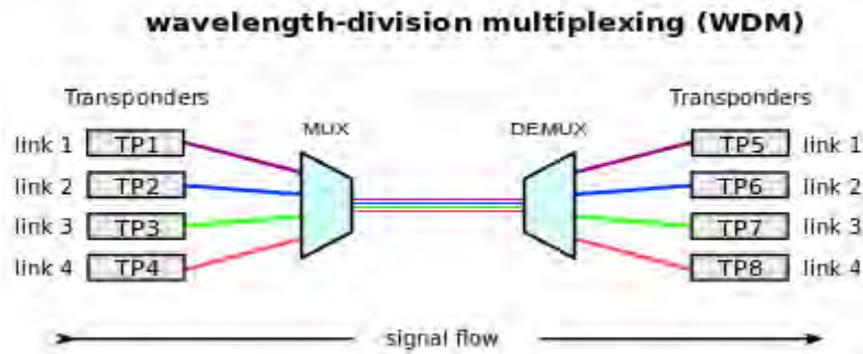
Les réseaux **traditionnels à trois niveaux** utilisaient la commutation par noyau, agrégat et périphérie, pour connecter les différents serveurs au sein du centre de données, où le trafic inter-serveur circulait dans une direction **nord-sud** à travers les dispositifs actifs pour communiquer entre eux. Cependant, aujourd'hui, grâce aux exigences de calcul et à l'interdépendance que **l'IA et le machine Learning** proposent, un plus grand nombre de ces réseaux sont installés via un réseau à deux niveaux, où les serveurs communiquent entre eux dans une direction **est-ouest** en raison de la très faible latence requise par les réseaux de production et de formation.



# L'impact de l'IA dans l'infra dans la conception des DC.

Depuis **l'approbation 40G et 100G (IEEE) en 2010**, ainsi que les besoins de la technologie déployant **l'IA à grande échelle** conçoivent dans leurs réseaux pour aujourd'hui et demain : l'optique parallèle monomode. (*Wavelength Division Multiplexing - WDM*) non interopérables approuvées ou brevetées, soit des accords multi-sources et des techniques d'optique parallèle **utilisant 8** (4 pour émettre et 4 pour recevoir) **ou 20 fibres** (10 pour émettre et 10 pour recevoir).

L'un des principaux avantages du déploiement de l'optique parallèle est la possibilité de prendre un port de commutation à grande vitesse (**40G par exemple**) et de le diviser en ports serveur (**4x10G**). La division des ports permet d'obtenir des économies d'échelle, car cette division à plus faible vitesse peut considérablement réduire le nombre de châssis ou d'unités de rack pour l'électronique de 3:1 (**et l'immobilier des centres de données coûte cher**) et **utilise moins de puissance**, ce qui nécessite moins de refroidissement et réduit encore la facture. Cela équivaut à une économie **de 30 %** sur une solution monomode.



# L'impact de l'IA dans l'infra et la conception des DC.

La carte technologique des principaux fournisseurs de commutateurs et d'émetteurs-récepteurs montre une voie de migration très claire pour les clients qui déploient des optiques parallèles. La majorité des entreprises de technologie ont suivi cette voie, de sorte que lorsque les optiques sont disponibles et qu'elles passent de **100G à 200G ou 400G**, leur infrastructure de fibre optique reste en place et ne nécessite aucune mise à niveau



# L'IA au service d'un meilleur PUE.

*L'IA met au régime les data centers pour les rendre moins énergivores*

# L'IA au service d'un meilleur PUE

## L'IA met au régime les datacenters pour les rendre moins énergivores

- Bardé de capteurs, un centre de données génère un grand nombre de data. Passant ces flux à la moulinette, le machine learning ajuste la capacité du système de refroidissement en fonction de la charge. L'efficacité énergétique des data centers est, de fait, un cheval de bataille de longue date des acteurs du numérique, GAFA en tête. Au-delà de l'enjeu d'image autour de réduction de leur empreinte carbone
- Il s'agit plus prosaïquement de réduire les coûts d'exploitation. Les concepteurs comme les exploitants de centres de données ont les yeux rivés sur le fameux PUE (Power Usage Effectiveness),
- Le principal indice d'efficacité énergétique. Il s'agit du rapport entre l'énergie totale utilisée par un centre de données et l'énergie consommée effectivement par ses équipements informatiques (serveurs, stockage, réseau). Il met notamment en évidence la part consacrée au système de refroidissement mis en place pour prévenir l'augmentation de la température des composants IT et éviter la surchauffe. Un PUE de 1,5 signifie que pour 1 Watt consommé par la charge IT, il faut 1,5 Watt pour faire fonctionner l'infrastructure.



# L'IA au service d'un meilleur PUE

- Pour tendre vers un PUE de 1, le chiffre théorique idéal, différentes techniques permettent d'améliorer le rendement de la chaîne d'alimentation électrique et d'optimiser la consommation des systèmes de refroidissement. Le **watercooling** exploite, par exemple, les propriétés caloriporteuses de l'eau. Un liquide réfrigéré vient évacuer la chaleur dégagée par les serveurs. La conception même du data center, son implantation dans des pays aux climats polaires ou le recours aux énergies renouvelables sont d'autres pistes déjà exploitées.





## Exemple d'une IA devenue autonome chez

Les acteurs du numérique se sont aussi tournés vers l'[intelligence artificielle](#). Si un centre de données traite des milliers d'informations par seconde, il génère lui-même un grand volume de données. De multiples capteurs remontent les températures intérieure et extérieure ou l'évolution de la charge de chaque serveur.

Des données qu'il est possible de croiser avec, entre autres, les prévisions météorologiques. Sorte de thermostat augmenté, **l'IA permet d'anticiper au plus juste, et en temps réel, l'énergie nécessaire au refroidissement du data center.**

[Google](#) est un pionnier dans le domaine. **Depuis 2014**, le géant du numérique fait appel à l'IA pour réduire la consommation énergétique de ses centres de données, sa filiale spécialisée en IA, DeepMind, avançait un gain allant jusqu'à **40%**, soit une réduction de **15%** de la surcharge globale du PUE. **Formé à partir des données d'historique** de production et alimenté en temps réel par les milliers de capteurs de température ou d'hydrométrie propres à chaque data center, en effet l'IA de Google, à base de réseaux neuronaux, recommande un scénario d'exploitation optimal en fonction des prédictions de température et de charge IT.

# L'IA au service d'un meilleur PUE

## Une IA devenue autonome chez Google

Si le modèle se contentait d'émettre des recommandations aux gestionnaires de ses data centers, Google est passé un cran plus loin. Dans [une note publiée deux ans plus tard](#), DeepMind explique que le système d'IA contrôle désormais directement le refroidissement des centres de données, même si les opérateurs peuvent reprendre la main à tout moment. "La mise en œuvre des recommandations exigeait trop d'efforts et de supervision de la part de l'opérateur", argue DeepMind.

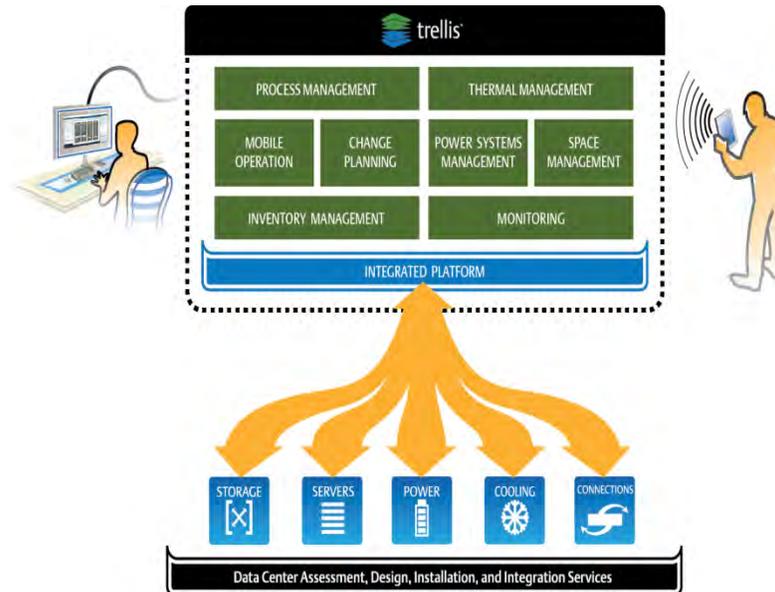


# L'IA pour une maintenance efficace et prédictive.

*L'Intelligence Artificielle en maintenance : prédire et prévenir*

# L'IA pour une maintenance efficiente et prédictive

- Anticiper une défaillance matérielle, prévoir une interruption de service, détecter un défaut sur une pièce en service offrent aux entreprises ou collectivités la capacité de réagir et d'éviter des incidents aux conséquences parfois dramatiques.
- Le calcul du temps moyen entre pannes (MTBF) est classique depuis de nombreuses années et est historiquement basé sur des modèles statistiques. L'apport de l'Intelligence Artificielle est d'offrir une plus grande finesse dans la détermination des pannes, une anticipation prenant en compte des facteurs « temps réels » voire de détecter des signaux faibles ou de découvrir des cas non répertoriés.
- Quelle maintenance ? Les processus de maintenance se classent en trois catégories :



# L'IA pour une maintenance efficiente et prédictive

- **Maintenance corrective** : celle-ci intervient alors que l'incident a eu lieu et que la panne a déjà produit ses effets. Le coût de la panne est subi par l'entreprise qui déclenche une opération de réparation et, éventuellement, de maintenance des matériels connectés ou ayant pu être impactés. Ce cas est le moins favorable pour l'entreprise.
- **Maintenance préventive** : ce schéma est le plus traditionnel. Il consiste à planifier des opérations de maintenance régulières sur les environnements en fonction du temps moyen entre pannes. Comme cette démarche est purement statistique, elle entraîne des actions parfois coûteuses sans qu'elles soient réellement justifiées. Pour éviter les incidents, il y a la probabilité d'en faire plus que nécessaire et donc d'engendrer un surcoût inutile.
- **Maintenance prédictive** : le modèle idéal consiste à s'appuyer sur des signaux faibles ou des éléments d'environnement qui permettent d'anticiper une panne et donc de déclencher de façon ciblée une action de maintenance. L'Intelligence Artificielle propose des outils efficaces pour répondre à cette problématique.



# L'IA pour une maintenance efficiente et prédictive

## ➤ Intelligence Artificielle Prédictive :

La mise en œuvre de processus de maintenance prédictive impacte les méthodes de fonctionnement et doit faire l'objet d'un changement des pratiques voire des matériels sur les chaînes de production.

## L'Intelligence Artificielle peut intervenir à deux niveaux :

**Détermination de la solution optimale** : le programme de maintenance résulte **d'un équilibre entre le taux de disponibilité des machines et équipements à contrôler et le niveau de coûts supportable**, l'objectif étant **d'obtenir un Taux de Rendement Synthétique\* (TRS) proche de 100%**. Cette recherche d'un optimum peut être complexe dans des environnements où plusieurs processus sont en œuvre et où la maintenance prédictive implique le déploiement de nouveaux outils de mesure. Dans ces cas, obtenir une solution optimale peut être facilité par des approches utilisant **de l'Intelligence Artificielle**.

**Anticipation des pannes** : l'usage de **nouveaux capteurs**, la prise en compte de **signaux faibles**, l'analyse d'écart aux valeurs nominales sont autant de facteurs que l'IA traite en temps réel pour déterminer qu'une opération de maintenance est nécessaire.



# L'IA importe la robotisation au service des gestions de proximités !

*Les Data Centers du Cloud seront gérés par des robots d'ici 2025*

# L'IA importe la robotisation au service des gestions de proximités !

La moitié des Data Centers Cloud exploiteront la robotique d'ici 2025. Et ces déploiements de l'IA pourraient **accroître l'efficacité opérationnelle des Data Centers** à hauteur de 30%.

Pour bien saisir l'impact de la robotique **IA** sur les Data Centers, il faut s'intéresser à leur fonctionnement. Le **workflow d'un Data Center** est composé de différentes actions comme la planification de capacité, l'ajustement des machines virtuelles et des environnements de conteneurs, ou encore la minimisation du gaspillage de ressources.

Toutes ces tâches sont essentielles, mais sont aussi fastidieuses et complexes. Or, **les robots se révèlent particulièrement efficaces** pour ce type de tâches nécessitant une exécution rigoureuse au quotidien.

Le déploiement de robots IA permet de **réduire le besoin d'une intervention humaine**, sans pour autant compromettre la sécurité, la précision ou l'efficacité de l'environnement.



<https://youtu.be/dL344MJPtEY>

# L'IA importe la robotisation au service des gestions de proximités !

Gartner dévoile l'impact de la robotique IA sur les Data Centers du Cloud.

À travers son étude, **Gartner va plus loin et identifie quatre domaines** dans lesquels les robots vont avoir un lourd impact sur l'automatisation des Data Centers au fil des **quatre prochaines années**.

Tout d'abord, les robots pourront prendre en charge **le déclassement et la destruction des disques durs** lors de la suppression de serveurs., ce travail peut être accompli plus efficacement par **l'IA**. Or, il s'agit d'un labeur fréquent dans le cadre des mises à jour hardware du Data Center.

L'intelligence artificielle excelle aussi pour **la surveillance à distance**. Le déploiement de **capteurs robots** permet de collecter en temps réel les données granulaires sur la température des racks des serveurs. D'autres données telles que les sons et les images peuvent aussi être capturées pour détecter rapidement d'éventuelles anomalies.

Les robots peuvent aussi **assurer la maintenance de sécurité** physique et numérique des Data Centers.

# L'IA importe la robotisation au service des gestions de proximités !

Le bénéfice de l'IA [et du Machine Learning](#) est aussi la gestion des processus de Data Centers. Les capacités de traitement naturel du langage permettent aux ingénieurs de fiabilité du site d'interagir et de communiquer avec la plateforme. De plus, cette dernière peut apprendre des situations passées pour améliorer l'efficacité future.

Selon Gartner, **les leaders IT peuvent exploiter l'automatisation des opérations et processus de Data Centers** pour différencier leurs services et les améliorer en termes de disponibilité. Ceci permet au fournisseur de services de tenir ses promesses en termes **d'accords de niveaux de service**.

**En conclusion**, comme de nombreuses industries, le secteur du Data Center va connaître **une profonde mutation grâce à l'IA**.

Les avantages seront nombreux, mais il convient de réfléchir au sort des opérateurs humains [qui seront remplacés par les robots...](#)





INGENIOS VOUS REMERCI POUR  
VOTRE ATTENTION

Questions / Réponses

Q & A

