

# APPRENTISSAGE AUTOMATIQUE ET DEEP LEARNING D'ENTREPRISE AVEC STOCKAGE INTELLIGENT

## RÉSUMÉ

Alimentées par les données, les progrès de l'infrastructure et l'omniprésence des outils d'apprentissage automatique et de Deep Learning, les solutions d'intelligence artificielle (IA) sont rapidement en train de s'imposer dans le datacenter d'entreprise. L'IA convertit les données en renseignements pour un large éventail de secteurs professionnels tels que ceux de l'automobile, de la santé, des sciences de la vie, des finances, des technologies, de la vente au détail et bien d'autres. Les données constituent désormais un avantage concurrentiel dans des secteurs tels que les assurances (où l'IA prédictive supprime les risques de la garantie), la finance (où le Deep Learning en temps réel reconnaît les fraudes dès qu'elles se produisent) voire la gestion du datacenter (où les schémas sont analysés pour prédire les défaillances et les problèmes d'évolutivité).

L'intelligence artificielle et, plus particulièrement, le Deep Learning impliquent de nouvelles exigences quant à la façon de fournir les données aux moteurs de calcul qui les consomment. Les nouvelles réalités liées au déploiement de l'intelligence artificielle dans le datacenter font évoluer les exigences en matière de densité, de débit, d'accès simultané et même d'architecture de données scale-out. Afin de permettre à l'IA de tenir ses promesses pour l'entreprise, le département IT doit penser différemment lorsqu'il s'agit d'associer le stockage et le calcul.

Ce document explique comment le Deep Learning et l'intelligence artificielle dans l'entreprise apportent de nouveaux flux de travail et défis au niveau de l'architecture de datacenter. Il indique également comment des solutions peuvent être créées à partir d'architectures d'infrastructure spécialement conçues pour rapprocher le calcul et le stockage scale-out.

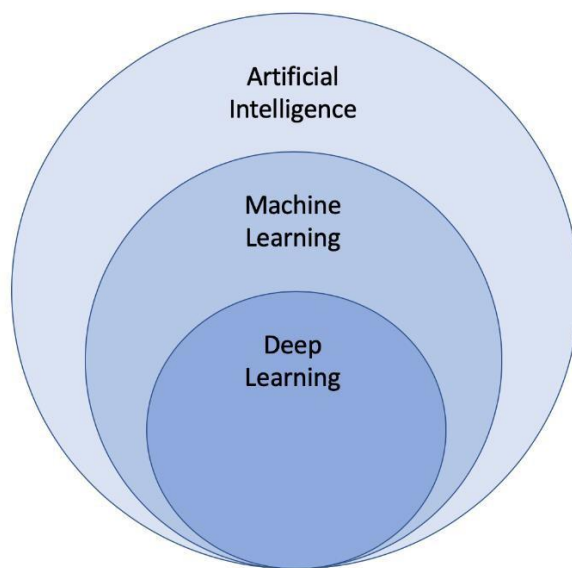
Le Deep Learning nécessite d'introduire de grandes quantités de données dans le processeur sans faire attendre ce dernier. En associant les ressources de calcul à la technologie de stockage appropriée, telle que la gamme Dell EMC Isilon, il est possible d'insérer des données dans le pipeline d'apprentissage automatique à la vitesse du processeur. Les systèmes correctement équilibrés accélèrent l'innovation et offrent la flexibilité et l'agilité nécessaires aux départements IT et aux scientifiques des données qui les utilisent.

## LE DEEP LEARNING EST EN TRAIN DE CHANGER L'ENTREPRISE

Il semblerait désormais que tout le monde ait vu la célèbre couverture du magazine The Economist mettant en scène des plates-formes pétrolières et portant le titre suivant : « The World's Most Valuable

Ressource » (La plus précieuse des ressources au monde)<sup>1</sup>. Les progrès réalisés dans les domaines de l'apprentissage automatique et du Deep Learning confèrent, en effet, une nouvelle importance aux données de chaque organisation : *les données sont devenues un facteur de différenciation concurrentielle*.

### FIGURE 1 : LA RELATION ENTRE IA, APPRENTISSAGE AUTOMATIQUE ET DEEP LEARNING



Source : Moor Insights & Strategy

L'intelligence artificielle décrit une classe générale de technologies dans lesquelles les ordinateurs prennent des décisions ou fournissent des renseignements généralement associés à l'intelligence humaine. L'IA correspond, par exemple, au moteur de recommandations que vous pouvez trouver sur le site Web en ligne d'un commerçant, et qui émet des recommandations sur les produits en fonction de votre historique d'achats ainsi que des articles que vous êtes en train de chercher.

<sup>1</sup> The Economist, The World's Most Valuable Resource is No Longer Oil, But Data. 6 mai 2017

---

L'apprentissage automatique est un type d'intelligence artificielle dans lequel des algorithmes analysent des données, apprennent de ces données et appliquent ces informations dans des environnements réels pour prendre des décisions. Un système d'apprentissage automatique servant à détecter les courriers électroniques indésirables, par exemple, peut être formé sur des échantillons provenant des millions d'e-mails déplacés vers le dossier de spam par des utilisateurs chaque jour.

Le Deep Learning, qui est l'un des domaines les plus intéressants et les plus actifs de l'intelligence artificielle, est un sous-ensemble de l'apprentissage automatique. Le Deep Learning utilise des algorithmes appelés *réseaux de neurones* pour affiner continuellement ses prévisions à partir de données, à mesure qu'il les trouve. Le Deep Learning est au cœur des véhicules autonomes, de l'analyse du sentiment qui reconnaît les humeurs humaines, ainsi que de la plupart des autres techniques d'IA qui traitent des données concrètes pour prendre des décisions ou faire des recommandations dynamiques. Ce document porte plus particulièrement sur le Deep Learning, mais des similitudes architecturales peuvent apparaître lors de la mise en œuvre de tout type de système d'IA gourmand en données.

Les exemples d'utilisation de l'IA sont nombreux et variés, qu'il s'agisse de chatbots et de systèmes de réponse vocale basés sur l'IA, de la prédiction du comportement du client ou encore de la gestion optimisée de la chaîne logistique. L'ère de l'entreprise intelligente est arrivée et nombreuses sont celles qui se demandent comment exploiter les technologies d'apprentissage automatique pour mieux prendre l'avantage sur leurs concurrents dans leur secteur d'activité.

Selon une récente étude<sup>2</sup> menée auprès de plus de 1 300 professionnels de l'IT, plus de 60 % des personnes interrogées travaillent pour des organisations qui prévoient de consacrer au moins 5 % de leur budget IT à l'intelligence artificielle. Un cinquième de ces personnes interrogées travaillent pour des organisations qui envisagent de consacrer plus de 20 % de leur budget IT à l'IA, ce qui est impressionnant.

L'intelligence artificielle faisant appel à des techniques de Deep Learning agit sur toutes les entreprises, souvent de manière inattendue. Voici quelques exemples illustrant l'impact de l'apprentissage automatique et de sa ramification plus spécialisée qu'est le Deep Learning sur l'entreprise moderne :

- Le **secteur des médias et du divertissement** utilise l'apprentissage automatique pour ajouter de l'intelligence à un certain nombre de tâches. L'analyse du sentiment est utilisée pour classer la réaction de l'audience à la

---

<sup>2</sup> O'Reilly Media, *The State of Machine Learning Adoption in the Enterprise*, 2019.

programmation des films et des émissions de télévision. Le secteur des médias et du divertissement s'appuie également sur la reconnaissance d'images formée par des algorithmes de Deep Learning pour automatiser la génération de métadonnées sur des quantités innombrables de contenus vidéo.

- De nombreux aspects des opérations de **fabrication** modernes de l'ensemble des secteurs d'activité s'appuient sur l'intelligence artificielle et l'apprentissage automatique. Les systèmes de reconnaissance d'images analysent les produits sur les lignes de fabrication afin d'identifier d'éventuels défauts. Les systèmes d'apprentissage automatique participent également aux analyses prédictives des pannes en scrutant les capteurs d'une usine afin de reconnaître et d'identifier des schémas susceptibles de provoquer des défaillances s'ils ne sont pas résolus. L'apprentissage automatique est également utilisé pour prendre des décisions de chaîne logistique permettant d'optimiser une opération à flux tendus grâce à une logistique et à un approvisionnement intelligents.
  - Les secteurs **de l'automobile et des transports** utilisent le Deep Learning pour modifier notre conception des voitures. Les techniques de Deep Learning alimentent la course vers l'objectif ultime que constitue la livraison de véhicules autonomes.
-

Cette nouvelle voie est pavée d'applications concrètes, telles que les systèmes de régulateur de vitesse adaptatif et intelligent, le pilotage semi-automatique, l'analyse prédictive des pannes et même la surveillance du conducteur, pour permettre à ce dernier de savoir ce qu'il se passe dans le véhicule. Aucune de ces innovations n'aurait été possible sans les progrès permanents de l'apprentissage automatique et du Deep Learning.

Pourtant, bien que le déploiement des technologies d'IA dans l'entreprise ait un réel impact, il n'en reste pas moins nouveau pour la plupart des organisations. Il est donc important de simplifier le processus et d'examiner les éléments courants avant de lancer un projet d'IA. Presque toutes les solutions de Deep Learning, qu'il s'agisse de prendre en charge la détection d'images, la classification d'images, la segmentation, le traitement du langage naturel et/ou l'analytique prédictive, utilisent un ensemble commun de technologies de base. Ces techniques sont déployées sur des plateformes qui, nativement, prennent en charge et sont réglées pour des packages logiciels courants, tels que TensorFlow, PyTorch et Caffe2, omniprésents dans la mise en œuvre de ces exemples d'utilisation.

La plupart des professionnels IT traditionnels n'ont pas les compétences requises pour concevoir et déployer efficacement des solutions d'IA pour des exemples d'utilisation très différents. L'apprentissage automatique et le Deep Learning sont des technologies synonymes de nouveaux défis, qui obligent à réfléchir aux données de façon novatrice.

L'étude susmentionnée montre que le manque de compréhension quant à la manière de déployer le Deep Learning, associé à une infrastructure qui n'est pas prête pour ces charges applicatives, entrave considérablement son adoption.

Il est donc essentiel, pour tout département IT concurrentiel, de combler ces lacunes et d'acquérir les compétences requises pour déployer le Deep Learning, secondé par des architectures d'analytique évolutives et flexibles.

## CRÉATION D'UNE ARCHITECTURE POUR LE DEEP LEARNING DANS LE DATACENTER

Pour saisir l'impact multidimensionnel du Deep Learning sur l'architecture de stockage, il faut comprendre un flux de travail d'apprentissage type. Chaque phase du pipeline d'apprentissage implique des exigences différentes vis-à-vis de l'infrastructure sous-jacente. Celles-ci sont illustrées dans la figure 2.

FIGURE 2 : PIPELINE DE MACHINE LEARNING/DEEP LEARNING TYPE

	<u>INGEST</u>	<u>DATA PREP</u>	<u>REFINE</u>	<u>TRAIN</u>	<u>DEPLOY</u>	<u>RETENTION</u>
	IOT, Logs, Sensors, Users, Etc	CPU-intensive Servers	GPU-enabled Server & Workstations	High Performance GPU-based Servers	CPU or Inference Accelerated Edge, Client, or Server	Long-term Storage
<b>Access Pattern</b>	Sequential	Sequential or Random	Random	Random	Random	Sequential
<b>Access Type</b>	Write	Read & Write	Read	Read	Read	Write
<b>Concurrency</b>	Variable	Low	Moderate	High	Low	Low
<b>Performance</b>	High	High	High	High	Moderate	Low
<b>Storage</b>	Block, File, or Object	Block or File	File or Object	File or Object	Block or Memory	Block, File, or Object
<b>Scale</b>	MB-GB	MB-TB	TB-PB	TB-PB	KB-MB	TB-PB

Source : Moor Insights & Strategy

Ces étapes se résument comme suit :

- **Acquisition des données** : les données proviennent d'une source externe (ou de plusieurs sources), telles que des périphériques, des fichiers journaux, des flux vocaux ou vidéo, ou des systèmes de gestion de la relation client. Les données arrivent et sont stockées. La solution de stockage doit simplement être aussi performante que les données entrantes l'exigent.
- **Préparation des données** : les données sont nettoyées et transformées pour la formation. Cette étape essentielle permet de s'assurer que les données sont cohérentes, que les anomalies sont identifiées et que le jeu de données est optimisé pour les algorithmes de formation. Certains types d'apprentissage automatique, tels que l'apprentissage supervisé, exigent que les données soient étiquetées au cours de cette phase.
- **Découverte et visualisation des données** : les experts en science des données travaillent avec les données pour optimiser les algorithmes et les paramètres de formation. Il s'agit d'un processus très itératif, mais qui ne nécessite que de modestes volumes de stockage et de calcul.

- **Formation et développement de modèles** : l'essentiel du travail a lieu à ce stade. Les données nettoyées sont introduites dans un cluster de processeurs graphiques, ou dans d'autres moteurs de calcul hautes performances, dans lesquels elles sont souvent itérées pendant de longues périodes. La formation nécessite un stockage haut débit, optimisé pour les opérations de lecture aléatoires hautement simultanées.
- **Modèle de déploiement ou inférence de production** : les modèles générés au cours de la phase de formation sont déployés par rapport aux données du monde réel. Les caractéristiques de cette phase dépendent fortement du type de Deep Learning déployé. La reconnaissance d'images, par exemple, peut se produire sur un périphérique client tel qu'une webcam intelligente, nécessitant peu d'interactions avec les systèmes de stockage, tandis que des applications plus avancées peuvent s'exécuter au sein d'un datacenter d'entreprise.
- **Conservation de données** : les données utilisées dans la formation du modèle ou dans l'inférence répétitive pour le Deep Learning sont conservées à des fins d'archivage ou de réutilisation. Cette étape est essentielle. L'archivage des données garantit que les modèles peuvent être recréés et que les données peuvent être extraites en vue d'une utilisation ultérieure.

Ces étapes sont secondées par un ensemble de principes de conception qui doivent être pris en compte lors de la mise en œuvre d'une infrastructure pour prendre en charge le Deep Learning :

- **Performances et évolutivité** : les performances ne peuvent pas se dégrader avec la mise à l'échelle. Chaque composant, qu'il s'agisse du calcul, du stockage ou de la gestion de réseau, doit évoluer de manière linéaire et indépendante pour que le système puisse s'adapter de manière transparente à la charge applicative afin d'éviter les goulots d'étranglement au niveau du calcul, des E/S et du réseau.
- **Flexibilité** : les systèmes IA sont conçus autour des données. La réalité inhérente à cette dynamique est que les logiciels, les techniques analytiques et les cas d'utilisation changeront inévitablement à mesure que l'écosystème IA évoluera, alors que les données d'une organisation resteront relativement constantes. Les systèmes doivent prendre en charge le stockage de données à long terme, tout en conservant la flexibilité nécessaire pour évoluer avec les besoins métier.
- **Gestion des données d'entreprise** : les données utilisées par le Deep Learning, en dépit de leur utilisation inhabituelle, sont des données d'entreprise et doivent être gérées comme telles. La sécurité, la protection des données, la conformité aux normes et d'autres problèmes classiques de gestion des



données s'appliquent également aux données de Deep Learning. Les solutions de stockage déployées dans ces environnements doivent s'intégrer correctement aux règles et procédures existantes de l'entreprise en matière de gestion des données.

Bien que les discussions sur l'apprentissage automatique et le Deep Learning s'intéressent naturellement au calcul, il est évident que ces solutions obligent à penser autrement au sujet des données. Pour le Deep Learning, il faut réfléchir autrement à la façon de gérer, d'analyser et de stocker les données.



## LES DONNÉES D'UN ENVIRONNEMENT DE DEEP LEARNING

Les caractéristiques des données d'un flux de travail de Deep Learning sont différentes de celles de la plupart des autres applications IT :

- **Les données sont principalement non structurées.** Elles comprennent des images, de l'audio, du texte libre voire des flux de données de séries chronologiques. L'architecture de stockage d'un environnement de Deep Learning doit être optimisée pour les données non structurées. Le stockage doit également prendre en charge plusieurs protocoles d'accès aux données tels que SMB, NFS, HDFS, S3 et HTTP pour offrir la plus grande flexibilité opérationnelle.
- **L'échelle des données** augmente considérablement avec les capteurs vidéo et de périphérie. Un contenu de résolution supérieur, notamment, génère des téraoctets de données à des fins d'analyse sur des périodes concises. La conservation de ces données à des fins d'analyse ou de reformation ultérieure peut nécessiter plusieurs pétaoctets de stockage. Pour que le Deep Learning permette d'extraire des renseignements fiables, il faut disposer d'un long historique de données à analyser. Les solutions de stockage de cet environnement doivent pouvoir évoluer (scale out) de manière simple et sans interruption.
- **L'utilisation des données varie considérablement**, car les besoins sont différents pour chaque phase du pipeline d'apprentissage. Le nettoyage ou l'attribution d'étiquettes aux données, par exemple, exige des performances très différentes de celles requises par les processus qui alimentent ces mêmes données dans un cluster à des fins de formation ou d'inférence en temps réel. Une extrémité de ce pipeline peut être satisfaite par du stockage local traditionnel, DAS ou milieu de gamme, alors que l'autre extrémité nécessite un débit et des fonctions d'entreprise capables de suivre le rythme des technologies de traitement modernes.
- **Les données arrivent de partout.** Les applications de Deep Learning présentent des sources de données très diverses. Les données servant à l'analyse ou à la génération de modèles peuvent venir de la périphérie, d'applications Cloud natives, de services vocaux et même d'applications d'agrégation de journaux de serveurs. Le stockage doit être conçu pour pouvoir ingérer des données provenant de diverses sources.
- **Le cycle de vie des modèles de données.** L'IA nécessite un ensemble cohérent d'outils de gestion couvrant la gamme de solutions de stockage d'archive hautes performances et profond pour que les données restent actives dans une architecture de stockage alignée sur le flux de travail d'IA global d'une entreprise. De même, transformer les données existantes en entrées pour

de nouvelles fonctions d'IA exige des outils de gestion des données permettant à un département IT de déployer de nouvelles solutions par rapport au stockage existant.

Ces caractéristiques générales se traduisent en éléments concrets à prendre en compte lors du choix d'une solution de gestion des données pour le Deep Learning. Il est important de souligner que ces données restent des données d'entreprise qui doivent être protégées contre les pannes matérielles et logicielles, sécurisées contre les violations et gérées de manière efficace.

Le type de Deep Learning qu'une organisation déploie a également un impact sur l'architecture de stockage qui prend en charge ces flux de travail. La reconnaissance d'images, par exemple, qui est très utilisée dans les secteurs des médias et du divertissement, de la fabrication et de l'automobile, repose sur l'application de réseaux de neurones convolutifs (CNN) et de réseaux de neurones profonds (DNN).

Le CNN est un type de réseau de neurones qui apprend à classer et à reconnaître des images par un certain nombre d'étapes très répétitives. Pour les CNN, les modèles d'accès aux données au cours de la formation et de la reconnaissance nécessitent une architecture de stockage réglée pour un très grand nombre d'accès en lecture par petits blocs à la baie de stockage sous-jacente.

Pour citer un exemple concret, dans le cadre d'une analyse comparative réalisée par Dell EMC et NVIDIA, un système de stockage Dell EMC Isilon F800 a été associé à des serveurs NVIDIA DGX-1 équipés de plusieurs processeurs graphiques NVIDIA Tesla V100. Chaque processeur graphique a exécuté plus de 5 000 threads parallèles, ce qui équivaut à une moyenne de 703 lectures de fichiers simultanées par processeur graphique<sup>3</sup>. Il est essentiel que le système de stockage associé à un système de Deep Learning soit capable de fournir des données à grande échelle et de prendre en charge une simultanéité extrême, afin d'éviter que les éléments de traitement n'aient à attendre les données.

Il ne s'agit là que d'un exemple. Les autres systèmes de Deep Learning présenteront d'autres exigences. Les systèmes intelligents qui fournissent une reconnaissance de schémas en temps réel pour la détection des fraudes financières, par exemple,

---

<sup>3</sup> Livre blanc : Dell EMC Isilon and NVIDIA DGX-1 Servers for Deep Learning, <https://www.dell EMC.com/fr-fr/collaterals/unauth/white-papers/products/storage/Dell EMC Isilon and NVIDIA DGX 1 servers for deep learning.pdf>

peuvent nécessiter un stockage en mode bloc à très hautes performances. Les applications impliquant de telles contraintes gagneront peut-être à faire appel à des baies de stockage en mode bloc à faible latence et haut débit, telles que celles de la gamme Dell EMC PowerMax.

Les tailles de bloc, les modèles d'E/S de fichiers et l'échelle sont également à prendre en compte. Ce qu'il faut retenir, c'est que la distribution de données pour l'apprentissage automatique et le Deep Learning est très différente des autres charges applicatives d'entreprise. La gestion des données pour le Deep Learning nécessite le déploiement de solutions conçues pour fournir des performances d'accès hautement simultanés et multidimensionnelles à grande échelle avec une hiérarchisation sur un seul espace de nommage et une gestion simplifiée grâce à un ensemble cohérent d'outils.

---

## DELL EMC : FOURNISSEUR DE STOCKAGE POUR LE DEEP LEARNING

La puissance de l'IA ne peut être exploitée que si les données sont distribuées de manière efficace et performante. Plusieurs facteurs doivent donc être pris en compte lors de la conception de solutions de stockage pour les applications d'apprentissage automatique et de Deep Learning, d'autant que les différentes phases du pipeline d'apprentissage présentent différentes exigences en matière de performances, d'évolutivité et d'accès simultané.

Dans le même temps, il est logique de déployer des architectures de stockage qui procèdent à la hiérarchisation et évoluent de manière transparente pour répondre aux exigences de toutes les phases d'une charge applicative de Deep Learning.

La gamme Dell EMC Isilon constitue une base solide à partir de laquelle vous pouvez proposer des fonctions de stockage pour prendre en charge l'ensemble du cycle de vie du Deep Learning d'entreprise. Cela permet de suivre le flux de travail, depuis la formation, l'apprentissage et le déploiement jusqu'aux besoins ultimes d'archivage à long terme.

### *DELL EMC ISILON ONEFS*

La puissance de tout système de stockage réside dans son logiciel de système d'exploitation sous-jacent. Le système d'exploitation Dell EMC Isilon OneFS permet aux solutions de stockage NAS scale-out Dell EMC Isilon de fonctionner de manière intelligente.

Les puissantes fonctionnalités de OneFS optimisent et simplifient le stockage des données au cœur de chaque flux de travail d'intelligence artificielle. Le logiciel offre une hiérarchisation transparente, tout en fournissant un seul espace de nommage, en gérant le positionnement des données, en optimisant et en réglant les performances de chaque baie d'après les schémas de trafic détectés, et en assurant la mise à l'échelle linéaire et sans interruption du stockage. Le système d'exploitation Dell EMC Isilon OneFS offre chacune de ces fonctionnalités.

La simplicité de gestion du stockage permet aux experts en science des données de concentrer leurs efforts sur la gestion du processus d'apprentissage automatique, sans se soucier des détails de l'infrastructure de stockage sous-jacente. Cette simplicité permet également aux administrateurs IT de déployer la bonne combinaison de solutions de stockage flexibles et efficaces pour répondre aux besoins de l'apprentissage automatique et du Deep Learning.

- **Data Lake consolidé** : consolide toutes les données du flux de travail analytique dans un seul et même emplacement pour simplifier les pipelines d'analytique des données.
- **Prise en charge multiprotocole** : permet à l'analytique d'accéder aux données afin de prendre en charge une méthodologie de type « stockage unique, utilisation multiple » et d'améliorer l'agilité.

- **Gouvernance des données d'entreprise** : protège les données grâce à des fonctions de résilience et de sécurité natives.
- **Hiérarchisation transparente** : hiérarchise le stockage entre les nœuds All-Flash, hybrides et d'archivage du même cluster afin de permettre une mise à l'échelle économique vers plusieurs pétaoctets et l'accès à des jeux de données plus volumineux.
- **Mise en cache intelligente** : offre la possibilité d'ajuster les caractéristiques de mise en cache du système de stockage de façon dynamique en fonction des charges applicatives qui consomment les données. La mise en cache Isilon OneFS cible les performances de lectures simultanées, qui sont essentielles dans les flux de travail de Deep Learning.
- **Évolutivité linéaire** : permet aux systèmes Isilon de garantir des performances prévisibles, tout en traitant les charges applicatives parallèles hautement simultanées des implémentations de Deep Learning.
- **Prise en charge des DevOps et du modèle as-a-service dès la première utilisation** : permet aux entreprises de mettre en place des environnements de développement, de test et de données de production ou de fournir plusieurs environnements de données de production en séparant clairement les clients via plusieurs zones d'accès au sein du même cluster Isilon.

Le logiciel gère l'expérience globale et l'intelligence inhérente à la gamme Dell EMC Isilon. La facilité de gestion et les solides caractéristiques de performances et d'évolutivité de la baie font d'Isilon une plate-forme attractive pour les charges applicatives de Deep Learning.

## *DELL EMC ISILON : UNE PLATE-FORME CONÇUE POUR L'APPRENTISSAGE AUTOMATIQUE ET LE DEEP LEARNING*

Le niveau supérieur de la gamme de stockage Dell EMC Isilon correspond au NAS scale-out All-Flash Isilon F800. Selon Dell<sup>4</sup>, le F800 offre des performances et une capacité approchant celles des meilleurs produits du secteur. Le F800 peut exécuter jusqu'à 250 000 E/S par seconde avec un débit global de 15 Go/seconde dans un seul châssis 4U et jusqu'à 15,75 millions d'E/S par seconde à 945 Go/seconde dans un cluster complet de 252 nœuds.

En matière de capacité, l'Isilon F800 propose initialement des dizaines de téraoctets de stockage et peut évoluer, sans interruption, jusqu'à plusieurs dizaines de pétaoctets dans un seul espace de nommage. Isilon offre jusqu'à 85 % d'efficacité de stockage. Il fournit également des technologies de déduplication et de compression qui peuvent

---

<sup>4</sup> Caractéristiques de Dell EMC Isilon F800 : <https://www.dell.com/fr-fr/collaterals/unauth/data-sheets/products/storage/h15963-ss-isilon-all-flash.pdf>

diviser les besoins en capacité de stockage des données par 3, augmentant ainsi la capacité effective de la solution.

Isilon F800 est capable d'alimenter correctement les nœuds de calcul Deep Learning. Dotées de 60 disques SSD hautes performances et de huit connexions Ethernet à 40 Gbit/s, ces machines sont conçues pour offrir des performances prévisibles aux niveaux élevés de simultanéité requis par le Deep Learning. Outre le simple fait de fournir des performances prévisibles, l'Isilon F800 peut être hiérarchisé avec des nœuds Isilon Hybrid et Isilon Archive pour fournir une évolutivité facile à gérer vers plusieurs pétaoctets.

Le meilleur exemple pour illustrer ces performances est celui des architectures de référence Dell EMC codéveloppées avec NVIDIA, alliant les fonctionnalités d'Isilon F800 et des serveurs accélérés par des processeurs graphiques NVIDIA Tesla V100, tels que les serveurs PowerEdge C4140, DSS 8440 et NVIDIA DGX-1. Les analyses comparatives de ces solutions ont révélé les performances du point de référence ResNet-50 avec jusqu'à 72 processeurs graphiques. Les performances exprimées en nombre d'images par seconde étaient linéaires de 8 à 72 processeurs graphiques, et ces derniers étaient utilisés à 97 %<sup>5</sup>.

Ces chiffres d'analyse montrent que, sur l'un des ordinateurs de Deep Learning les plus performants du marché actuellement, le processeur constitue le goulot d'étranglement, alors que le système Dell EMC Isilon F800 permet de l'alimenter constamment en données.

## *DELL EMC POWERMAX : STOCKAGE EN MODE BLOC HAUTES PERFORMANCES*

Dans le flux de travail IA et dans certains algorithmes d'apprentissage automatique et de Deep Learning, certaines étapes nécessitent un stockage en mode bloc à très faible latence pour les taux de réponse en temps réel lors de l'acquisition des données, de leur préparation et de l'inférence de production.

La gamme Dell EMC PowerMax de solutions de stockage en mode bloc, qui propose certaines des architectures de stockage les plus performantes actuellement disponibles, est bien conçue pour prendre en charge ces scénarios. PowerMax s'appuie sur la technologie NVMe de bout en bout, offrant des temps de latence inférieurs à 300 ms entre 1,7 et 10 millions d'E/S par seconde (pour PowerMax 2000 et PowerMax 8000, respectivement), avec jusqu'à 13 To par brique<sup>6</sup>.

---

<sup>5</sup> Livre blanc Dell EMC. Dell EMC Isilon and NVIDIA DGX-1 servers for deep learning. Novembre 2018. [https://www.dell.com/fr-fr/collaterals/unauth/white-papers/products/storage/Dell EMC Isilon and NVIDIA DGX\\_1\\_servers\\_for\\_deep\\_learning.pdf](https://www.dell.com/fr-fr/collaterals/unauth/white-papers/products/storage/Dell EMC Isilon and NVIDIA DGX_1_servers_for_deep_learning.pdf)

<sup>6</sup> Notice technique Dell EMC PowerMax : <https://france.emc.com/collateral/data-sheet/h16739-powermax-2000-8000-ss.pdf>



Dell a positionné PowerMax pour traiter les charges applicatives d'IA en temps réel les plus exigeantes actuellement déployées dans les entreprises.

## DELL EMC : DEEP LEARNING DE PILE COMPLÈTE

Dans les environnements de Deep Learning, le stockage et le calcul sont étroitement liés. Une infrastructure bien conçue pour le Deep Learning, avec toutes les complexités qu'il implique en matière de gestion des données, doit être synonyme d'équilibre, d'interopérabilité, de performances et de flexibilité. Bien qu'il puisse y avoir de nombreuses similitudes entre les implémentations, il n'existe pas de solution universelle. Chaque déploiement et chaque environnement diffèrent légèrement.

Il existe une multitude d'options pour le déploiement des charges applicatives d'apprentissage automatique et de Deep Learning. Les différentes phases nécessitent des accès aux données différents, mais également des solutions de calcul distinctes. Les professionnels de l'IA peuvent choisir d'exécuter des charges applicatives sur des serveurs vierges, dans des machines virtuelles ou encore dans des conteneurs de type Docker.

La société Dell EMC ne se contente pas de fournir des éléments individuels dans une infrastructure de Deep Learning, elle travaille à créer des solutions pouvant être déployées rapidement par les professionnels IT. Dell EMC simplifie les décisions architecturales et réduit les délais de déploiement en proposant des offres telles que les Ready Solutions et les architectures de référence, qui associent tous les éléments requis pour résoudre le problème. Dell EMC fournit des instructions de configuration afin d'aider les entreprises à dimensionner et mettre à l'échelle leurs solutions d'analytique de données et d'IA pour qu'elles répondent à leurs exigences de charges applicatives spécifiques.

Les Ready Solutions et les architectures de référence combinent des serveurs Dell PowerEdge de taille adaptée avec des commutateurs réseau Dell EMC, du stockage Isilon et une pile logicielle optimisée pour la solution. Les Ready Solutions sont des piles matérielles et logicielles validées, disponibles à la commande et optimisées pour accélérer les initiatives d'IA. Elles permettent de réduire le temps de conception d'une nouvelle solution de 6 à 12 mois. Les services de conseil, de support, de financement et de déploiement Dell EMC complètent parfaitement ces offres Dell EMC Ready Solutions. Ces services fonctionnent ensemble pour garantir un déploiement fluide de la solution.

Les architectures de référence sont des piles testées et validées, ciblant les clients et les partenaires de solutions Dell. Alors que les offres Ready Solutions peuvent être commandées directement auprès de Dell, les architectures de référence visent à aider les professionnels IT à créer leur propre solution de pointe, en s'appuyant sur des produits Dell Technologies éprouvés.



## TABLEAU 1 : EXEMPLES DE QUELQUES READY SOLUTIONS ET ARCHITECTURES DE RÉFÉRENCE DISPONIBLES

Type	Solution	Éléments clés	Principaux
Ready Solutions pour l'IA	Deep Learning with Intel	Isilon H600 PowerEdge R740xd PowerEdge C6420	Intel
	Deep Learning with NVIDIA	Isilon F800 PowerEdge R740xd PowerEdge C4140	NVIDIA
	Machine Learning with Hadoop	Isilon H500/H600 PowerEdge R640	Hortonworks
Architectures de référence pour l'IA	Dell EMC Isilon and NVIDIA DGX-1 for Deep Learning	Isilon F800 NVIDIA DGX-1	NVIDIA
	Dell EMC Isilon and PowerEdge C4140 for Deep Learning	Isilon F800 PowerEdge C4140	NVIDIA
	Dell EMC Isilon and DSS 8440 for Deep Learning	Isilon F800 DSS 8440	NVIDIA
	Dell EMC Isilon and PowerEdge R940 for Algorithmic Trading	Isilon F800 PowerEdge R940	Intel

Source : Moor Insights & Strategy

## CONCLUSION

Les données sont devenues les ressources les plus stratégiques et les plus différenciatrices de nombreuses organisations. Les techniques d'IA révolutionnent la façon dont les données sont interprétées et utilisées. Les entreprises investissent énormément dans l'acquisition de connaissances et le déploiement d'une infrastructure capable de prendre en charge cette réalité.

Dans le même temps, l'intelligence artificielle, qu'il s'agisse d'apprentissage automatique ou de Deep Learning, oblige les départements IT à réfléchir différemment aux données et à l'architecture de stockage, comparé à celles qui soutiennent des charges applicatives d'entreprise plus traditionnelles. Les attributs des données sont différents. La complexité de l'analytique est différente. Les besoins des consommateurs de ces données sont différents. La capacité à alimenter constamment ces nœuds de calcul accélérés est primordiale. Les solutions d'IA basées sur Dell EMC Isilon sont précisément conçues pour répondre à ces besoins.

Le déploiement de solutions de Deep Learning nécessite une soignée réflexion. Il implique également de collaborer avec des fournisseurs de technologies qui comprennent les exigences de ce nouveau monde et proposent toutes les solutions ciblées nécessaires pour faciliter la vie des professionnels IT entrés dans cette nouvelle ère.

Dell EMC est un excellent exemple de ce type de partenaire. Le Deep Learning donne la priorité aux données, et Dell EMC est l'un des leaders mondiaux de la gestion des données des datacenters, des Clouds privés et publics, et des réseaux de périphérie. L'ampleur de sa gamme de solutions d'IA confère à Dell EMC une position unique pour aider les clients à concevoir le meilleur environnement possible afin de répondre à leurs besoins. Dell EMC dispose d'une vaste gamme de solutions de stockage qui gèrent et protègent les données des clients, ainsi que de services et de solutions optimisés pour réussir à utiliser l'IA.

Pour en savoir plus, consultez le site Web de Dell EMC consacré à ce sujet : <https://www.dellemc.com/fr-fr/solutions/artificial-intelligence/index.htm>.

## INFORMATIONS IMPORTANTES SUR CE DOCUMENT

### *CONTRIBUTEUR*

Steve McDowell, analyste senior chez [Moor Insights & Strategy](#)

### *DIRECTEUR DE LA PUBLICATION*

Patrick Moorhead, fondateur, président et analyste principal chez [Moor Insights & Strategy](#)

### *DEMANDES DE RENSEIGNEMENTS*

[Contactez-nous](#) si vous souhaitez discuter de ce rapport. Moor Insights & Strategy vont répondre rapidement.

### *CITATIONS*

Cet article peut être cité par la presse accréditée et les analystes, mais doit être cité dans le contexte, en affichant le nom de l'auteur, son titre et « Moor Insights & Strategy ». Les organismes autres que la presse et les analystes doivent recevoir l'autorisation écrite préalable de Moor Insights & Strategy pour toute citation.

### *LICENCE*

Ce document, y compris les documents à l'appui, est la propriété de Moor Insights & Strategy. Cette publication ne peut être reproduite, distribuée ou partagée sous quelque forme que ce soit sans l'autorisation écrite préalable de Moor Insights & Strategy.

### *DIVULGATION D'INFORMATIONS*

Ce document a été commandé par Dell. Moor Insights & Strategy fournit des études, des analyses, des avis et des conseils à bon nombre des sociétés de haute technologie mentionnées dans le présent document. Aucun employé de l'entreprise ne détient des participations avec l'une des entreprises citées dans ce document.

### *CLAUSE DE NON-RESPONSABILITÉ*

Les informations présentées dans ce document sont fournies à titre informatif uniquement et peuvent contenir des inexactitudes techniques, des omissions et des erreurs typographiques. Moor Insights & Strategy décline toute garantie quant à l'exactitude, à l'exhaustivité ou à la pertinence de ces informations et n'aura aucune responsabilité quant aux erreurs, omissions ou insuffisances de ces informations. Ce document comprend les opinions de Moor Insights & Strategy et ne doit pas être interprété comme des déclarations de fait. Les opinions exprimées ici sont sujettes à modification sans préavis.

Moor Insights & Strategy fournit des prévisions et des déclarations prospectives en indicateurs directionnels et non en prédictions précises d'événements futurs. Bien que nos prévisions et déclarations prévisionnelles représentent notre jugement actuel sur ce que nous réserve l'avenir, elles sont soumises à des risques et incertitudes à cause desquelles les résultats réels pourraient sensiblement différer. Il vous est demandé de ne pas accorder d'importance indue à ces prévisions et déclarations prospectives, qui ne reflètent notre opinion qu'à partir de la date de publication de ce document. Veuillez garder à l'esprit que nous ne nous obligeons pas à réviser ou à dévoiler publiquement les résultats de toute révision de ces prévisions et déclarations prospectives à la lumière de nouvelles informations ou d'événements futurs.

©2019 Moor Insights & Strategy. Les noms de sociétés et de produits sont utilisés à des fins d'information uniquement et peuvent être des marques déposées de leurs propriétaires respectifs.