

L'INFRASTRUCTURE
DE **CALCUL**
NÉCESSAIRE POUR
L'ANALYTIQUE
PRÉDICTIVE

Les organisations modernes tirent parti des avantages de la transformation de l'analytique prédictive pour progresser sur le plan opérationnel. L'analytique prédictive vous permet de vous familiariser avec les connaissances et les informations qui peuvent faire la différence entre les profits et les pertes. Elle utilise la modélisation, l'apprentissage automatique et l'intelligence artificielle pour analyser les données actuelles et effectuer des prévisions à l'avenir. Tous les secteurs peuvent tirer parti de l'analytique prédictive. Les personnes qui ignorent ce potentiel passent à côté d'une incroyable opportunité d'augmenter ou de créer de meilleures expériences.

La progression récente de l'analytique prédictive et de l'apprentissage automatique basés sur l'IA offre aux organisations la possibilité de tirer pleinement parti des données qui étaient auparavant difficiles à déchiffrer. Grâce à la puissance de l'infrastructure de l'IA, vous pouvez tirer parti de l'analytique prédictive/de la vague d'apprentissage automatique et bénéficier d'avantages commerciaux, ce qui vous permettra non seulement de préserver une place concurrentielle, mais aussi de rester à la pointe du progrès. Les fonctionnalités de serveur automatisées et le serveur approprié pour la tâche en question vous permettront d'aller là où vous devez aller. Cet eBook se concentre sur la façon dont l'automatisation des serveurs, avec les serveurs appropriés pour la charge applicative, permet aux entreprises de faire des progrès considérables dans le monde de l'apprentissage automatique et de l'analytique prédictive.



Un nouveau rapport d'étude d'ESG, [trois technologies de calcul de transformation vérifiées pour accélérer l'IA et les principaux atouts](#), les organisations segmentées en fonction de leur niveau d'avancement en matière d'intelligence artificielle. Les trois critères utilisés pour l'analyse comparative étaient l'automatisation des serveurs, les accélérateurs et l'infrastructure convergée/hyperconvergée. Les fonctions de serveur automatisées ont permis d'intégrer les organisations dans la catégorie la plus avancée (phase 3). Tandis que seule la moitié (55 %) des entreprises en phase 1 développe, déploie et optimise les modèles d'IA en production pour l'analytique prédictive, 74 % des entreprises de phase 3 poursuivent ces objectifs.¹

Pourquoi est-ce important ? Cet eBook ne peut pas présenter tous les cas d'utilisation de l'analytique prédictive et de l'apprentissage automatique. Voici juste quelques exemples :

- Créer des moteurs de recommandations pour permettre aux clients d'acheter vos produits de manière appropriée.
- Les usines et chaînes logistiques peuvent utiliser l'analytique pour prévoir quand les défaillances se produiront, évitant ainsi les temps d'arrêt.
- Prédire le prix optimal de la vente d'un produit sur la base de tendances des données passées telles que la circulation piétonne et les prévisions météorologiques.

Ces scénarios n'effleurent même pas la surface de ce que vous pouvez faire avec l'analyse prédictive. En automatisant vos serveurs, vous libérez du temps pour les employés afin de vous concentrer sur la mise en place d'initiatives d'IA plutôt que sur l'infrastructure babysitting. Votre personnel peut ensuite consolider vos informations dans des lacs de données. Il utilisera des applications telles que Hadoop et Spark pour résoudre les problèmes et récupérer des prévisions exploitables à partir de cette immense quantité de données collectées depuis des années. Il exécutera tout cela sur les serveurs à l'aide d'excellentes performances en matière de traitement, de stockage et de mémoire. Lorsque vous faites de l'analytique prédictive et de l'infrastructure moderne une priorité, vous devenez davantage une organisation de phase 3. Imaginez que vous puissiez tirer parti de ces avantages.

Organisations en phase 3 :²

sont **2,6 fois** plus susceptibles que celles de la phase 1 de se démarquer de la concurrence en matière d'analytique et de business intelligence.

sont **2 fois** plus susceptibles de voir un délai de valorisation plus court.

améliorent la vitesse de prise de décision avec l'IA de près de **2 fois** plus que les organisations en phase 1.

améliorent la précision de la décision avec l'IA de près de **2,6 fois** plus que les organisations en phase 1.

voir les coûts moyens réduits **2,7 x** grâce à l'automatisation des processus métier et/ou des opérations par rapport aux organisations de l'étape 1.

Phases de maturité de l'IA par ESG

Phase 1

(**42 %** des organisations de l'étude) : faible niveau d'automatisation, utilisation très limitée des accélérateurs et/ou peu ou pas d'infrastructure convergée/hyperconvergée pour l'IA.

Phase 2

(**33 %** des organisations) : niveaux d'automatisation modérés, certaines utilisations des accélérateurs et/ou une infrastructure convergée/hyperconvergée pour l'IA.

Phase 3

(**24 %** des organisations) : niveaux d'automatisation élevés, large utilisation des accélérateurs et/ou utilisation élevée d'une infrastructure convergée/hyperconvergée pour l'IA.

Applications pour vous aider à être les meilleurs avec une analytique prédictive : Spark en cours d'exécution sur Hadoop

Hadoop est une plate-forme Open Source qui fournit la base de vos activités d'analytique prédictive et d'apprentissage automatique. Lors de l'exécution d'une analytique prédictive, votre première problématique (après avoir pris connaissance des problèmes métiers que vous souhaitez résoudre ou des opportunités de marché que vous souhaitez exploiter) consiste à comprendre comment rassembler toutes vos données pertinentes. Rassembler toutes les données pertinentes vous permet de regrouper vos ressources de façon à ce que les prévisions générées soient aussi précises que possible. Hadoop présente l'avantage d'être un système de fichiers hautement distribué qui vous permet de stocker n'importe quel type de données au format natif sans transformation, ce qui en fait la solution idéale non seulement pour les données structurées, mais également pour les données non structurées.

Voici quelques exemples d'utilisation de Hadoop :



Les détaillants servent mieux leur clientèle en l'utilisant pour analyser des données structurées et non structurées afin de mieux adapter leurs offres.



Hadoop fournit des données aux applications qui permettent aux sociétés de services financiers de prévenir la fraude.



Les entreprises logistiques utilisent l'analytique alimentée par Hadoop pour évaluer si la maintenance préventive est requise sur leur équipement.

Il est important de garder à l'esprit que Hadoop doit être associé à une autre application pour exécuter une analytique prédictive et un apprentissage automatique réels. Il s'agit souvent d'une application Spark. Spark est un projet Open Source avec des fonctionnalités améliorées. Hadoop comporte des fonctions de mappage et de réduction. La fonction de mappage prend un jeu de données et le convertit en un autre jeu de données. Ce nouveau jeu décompose les éléments en paires clé/valeur. La fonction Reduce combine ensuite ces paires clé/valeur dans un ensemble plus petit de paires clé/valeur. Dans Spark, vous pouvez également joindre, regrouper, filtrer et ajouter ou supprimer des données. Vous bénéficiez ainsi d'opportunités supplémentaires pour manipuler vos données. Spark offre une analytique ultrarapide et dispose d'une bibliothèque d'apprentissage automatique intégrée. Le moteur de traitement en mémoire de Spark garantit des résultats prévisibles en temps réel. Il est important de garder à l'esprit que Spark et Hadoop travaillent en partenariat pour offrir ces avantages à votre organisation.

Fonctions importantes de la mise en œuvre de Spark :



Interrogation des données en temps réel



Une vaste communauté de support



Traitement rapide des flux de données à faible latence ; exécute les charges applicatives 100 fois plus vite que Hadoop seul ³



Bibliothèques complètes

Des serveurs automatisés et complets sont indispensables pour exécuter l'analytique prédictive

Pour tirer le meilleur parti de l'appariement Hadoop et Spark, vous aurez besoin des éléments suivants :

1. Des serveurs automatisés qui libèrent du temps et des ressources pour les formations d'apprentissage automatique et fournissent des informations plus rapidement à vos clients.
2. Des processeurs cœur puissants et complets pour exploiter toutes ces données et exécuter ces algorithmes.
3. Un stockage important pour s'assurer que Hadoop s'exécute sans limites.
4. Mémoire serveur sérieuse pour permettre au traitement en mémoire de Spark de fonctionner de manière optimale.

Automatiser vos serveurs

Qu'apporte l'automatisation des serveurs à l'analytique prédictive et à l'apprentissage automatique ? En fait, elle leur apporte beaucoup. Lorsque vous remplacez vos serveurs vieillissants par des serveurs modernes dotés de fonctions d'automatisation, vous pouvez compter sur les éléments suivants :⁴

43 %

Déploiements plus rapides

43 %

Systèmes plus fiables

38 %

mises à jour plus rapides des applications

37 %

moins de temps consacré aux tâches manuelles et routinières de gestion de l'infrastructure IT

Cela se traduit par un déploiement plus rapide des informations et plus de temps consacré par le personnel à la transmission de ces informations. Vous souhaitez que vos serveurs d'analytique soient opérationnels rapidement et qu'ils soient rétablis tout aussi rapidement s'ils tombent en panne. Vous souhaitez être en mesure de déployer vos clusters Hadoop avec efficacité et de mettre à jour Hadoop et Spark le plus rapidement possible en cas de besoin. Plus vous automatisez votre transition, moins vous consacrez de temps aux processus de maintenance manuelle susceptibles de ralentir votre transition vers la valeur de l'analytique.

Trouvez le serveur complet adapté à Hadoop et Spark

Ce dont vous avez besoin pour la tâche d'analytique prédictive à l'aide d'Hadoop et Spark, c'est d'un serveur qui a tout : des performances solides en matière de traitement, de stockage et de mémoire.

Les processeurs sont le cerveau derrière vos opérations. Ils analysent et interprètent toutes les instructions que d'autres matériels et logiciels leur donnent. Dans l'apprentissage automatique, le processeur est chargé de l'exécution de la logique de votre algorithme. Il s'agit d'un rôle important ; le nombre de cœurs est extrêmement important. Il est important de vous assurer que vous choisissiez une infrastructure capable de traiter ces charges applicatives gourmandes en traitement.

Hadoop est une application gourmande en stockage qui effectue un traitement par lots volumineux. Elle traite de vastes quantités de données de l'ordre des pétaoctets. Pour exécuter Hadoop, assurez-vous que vous provisionnez des serveurs qui ne risquent pas de manquer d'espace de stockage. Bien que Hadoop suive un modèle distribué, chaque serveur ou nœud doit disposer d'un espace de stockage suffisant pour s'exécuter sans problème sur les disques locaux. Il est important de comprendre qu'il existe plusieurs types de nœuds dans Hadoop. Cet eBook se concentre sur les offres pour le nœud de périphérie, qui joue le rôle de volume système pour le cluster, et les nœuds de traitement, qui génèrent le traitement réel de vos données.

Spark, en revanche, est gourmand en mémoire. Bien que la recommandation minimale de Spark soit de 8 Go de mémoire par machine, une plus grande quantité de mémoire est idéale. Certains experts recommandent au moins 128 ou 256 Go. Spark recommande d'allouer au maximum 75 % de la mémoire disponible pour Spark et de laisser le reste pour le système d'exploitation et le cache de la mémoire tampon.⁵ La mémoire est particulièrement essentielle pour cette application, car elle met en cache les données en mémoire sur plusieurs opérations parallèles.

En résumé, vous avez besoin de serveurs automatisés qui peuvent faire tout cela lorsqu'il s'agit d'explorer l'analytique prédictive et l'apprentissage automatique. Découvrez les serveurs Dell EMC conçus pour ces charges applicatives gourmandes en ressources.

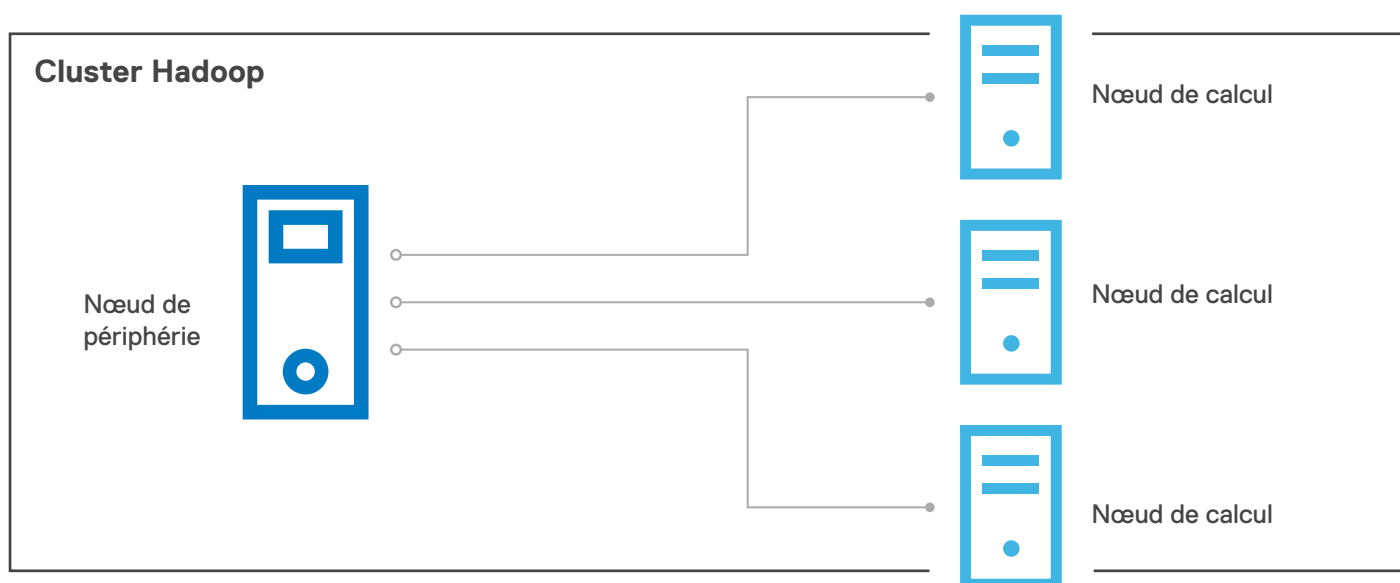
Spark recommande de provisionner au moins 8-16 cœurs par machine afin de faire en sorte que toutes les données soient indispensables à l'analytique prédictive.⁵

⁵ | L'infrastructure de calcul nécessaire pour l'analytique prédictive

Serveurs PowerEdge pour vos besoins en analytique prédictive

Tout d'abord, toutes les recommandations de produits ci-dessous mentionnent des options d'automatisation extrêmement convaincantes. OpenManage, la première plate-forme de gestion de serveurs Dell EMC propose des API RESTful. Les API vous permettent de scripter de nombreux aspects du déploiement, de la maintenance et du provisionnement de votre serveur. Vous pouvez écrire un script, le définir et l'oublier. Le portefeuille OpenManage prend également en charge le cadre Ansible, qui propose des modules permettant à l'IT d'être automatisée dans un scénario hétérogène. Ces modules sont fournis sous la forme de scripts pré-écrits pour une automatisation ajustée.⁷

Pour créer un cluster Hadoop, vous aurez besoin de plusieurs types de nœuds. Le schéma ci-dessous se concentre sur un seul nœud de périphérie et plusieurs nœuds de traitement. Le nœud de périphérie gère l'accès au cluster Hadoop et agit comme un point d'entrée entre votre cluster Hadoop et le réseau externe. Le nœud de périphérie ne prend pas en charge les tâches de calcul intensif. Cette tâche incombe aux nœuds de traitement. Les nœuds d'opérateur prennent en charge le gros travail de traitement des données pour vous et nécessitent une grande quantité de mémoire et de stockage, ainsi qu'une grande puissance de traitement. Pour obtenir des informations détaillées sur Hadoop, rendez-vous [ici](#).



Nœud de périphérie

PowerEdge R640

Bien que le nœud de périphérie ne soit pas recruté pour une grande quantité de données, il doit toujours disposer de puissants processeurs, ainsi que d'un stockage et d'une mémoire importants pour garantir le bon fonctionnement du cluster. Le Dell EMC PowerEdge R640 est tout juste adapté à cette tâche ; ni trop ni pas assez pour ce rôle. Le R640 vous permet de faire évoluer les ressources de calcul avec jusqu'à deux processeurs évolutifs de 2^e génération Intel® Xeon® et jusqu'à 56 cœurs. Il est livré avec un maximum de 24 barrettes DIMM, dont 12 peuvent être NVDIMM ou DCPMM, et jusqu'à 7,68 to de mémoire.⁸ Le PowerEdge R640 vous permet de créer un pool de cache NVMe et d'utiliser des disques de 2,5" ou 3,5" pour le stockage de données. Créez une configuration de nœud de périphérie optimale avec des processeurs, de la mémoire et du stockage évolutifs.



Récompense

L'impulsion de la marque IT : **Leader du marché 2019, serveurs montés en rack** : « Les professionnels de l'IT ont à nouveau reconnu Dell EMC comme le leader du marché des serveurs montés en rack pour la 3^e fois en sept ans. »



Serveurs PowerEdge pour vos besoins en analytique prédictive

Nœuds d'opérateur

PowerEdge R740

Le PowerEdge R740 est une plate-forme adaptée à l'analytique, avec une mémoire hautement extensible (jusqu'à 7,68 To) qui permet d'exécuter Spark. Évoluez pour répondre aux exigences de capacité avec un maximum de 16 disques de 2,5" ou 8 disques de 3,5" et jusqu'à 128 To de stockage afin de garantir un fonctionnement optimal des clusters Hadoop. Le R740 comprend la gamme de processeurs évolutifs de 2^e génération Intel® Xeon® avec jusqu'à 56 cœurs. Il s'agit d'une source d'alimentation dont vous aurez besoin pour traiter de vastes quantités de données.



Récompense

IDEA – PRIX BRONZE 2018



PowerEdge R740xd

Outre les fonctions de PowerEdge R740, Dell EMC PowerEdge R740xd ajoute des options de capacité de stockage exceptionnelles, ce qui en fait la solution idéale pour les applications gourmandes en données qui nécessitent un stockage important tel que Hadoop. Vous pouvez choisir jusqu'à 24 disques NVMe ou un total de 32 disques de 2,5" ou 18 disques de 3,5" et jusqu'à 288 To de stockage. Le R740xd avec les processeurs évolutifs Intel® Xeon® de 2^e génération a conduit à 900 % de jeux de requêtes supplémentaires lors de tests récents par rapport à notre serveur de génération précédente, R720xd, et a fourni les données 99,8 % plus rapidement. Par rapport à notre version la plus récente de ce serveur, le R740xd avec des processeurs évolutifs Intel® Xeon® de 2^e génération abritait 400 % de jeux de requêtes en plus et fournissait les données 27 % plus rapidement.⁹



Review

ServeTheHome : [Review du serveur Dell EMC PowerEdge R740xd](#)



Options de nœud d'opérateur, comparées



-- Option 1 : R740



-- Option 2 : R740xd



Traitement

Jusqu'à 56 cœurs

Jusqu'à 56 cœurs



Stockage

128 To

288 To



Mémoire

7,68 To

7,68 To

Architectures Dell EMC Ready pour Hadoop

L'expertise et l'infrastructure font une différence lors de la création d'un environnement Hadoop.¹⁰ Vous avez la possibilité de réduire vos efforts et d'acheter un modèle validé pour la création de vos clusters. Les architectures Ready pour Hadoop sont conçues pour répondre aux exigences en matière d'analytique des données, réduire les coûts de développement et améliorer les performances. Bien qu'il y ait de nombreux cas d'organisations qui prennent du retard sur les délais et qui ont du mal à prendre des décisions pour mettre en place leurs plates-formes Hadoop, votre entreprise peut connaître une expérience différente.¹¹

Les architectures Dell EMC Ready pour Hadoop sont des solutions créées pour répondre à tous vos besoins en analytique de données. Dell EMC a commencé à créer des solutions Hadoop personnalisées en 2009 et possède l'expertise, les outils et les solutions nécessaires pour mener à bien un déploiement Hadoop. Cloudera Hadoop Design fournit les principaux éléments de Hadoop au sein d'une solution basée sur le logiciel Cloudera Enterprise et le matériel Dell EMC, avec des options de service pour vous aider. Ses avantages sont notamment :¹²

- Valorisation d'une solution déjà optimisée et validée
- Surmonter le manque d'expertise Hadoop en interne
- Écourter les délais du projet et réduire les problèmes de production

Pour en savoir plus sur les architectures Dell Ready pour Hadoop, [rendez-vous sur notre site](#) ou [contactez-nous](#) dès aujourd'hui.

Services de migration Hadoop

Vous êtes prêt à faire le grand saut et à migrer vers Hadoop, mais vous ne savez pas comment vous y prendre ? Les services de consulting Dell Technologies peuvent vous aider. Vous n'êtes pas seul si vous avez du mal à gérer les opportunités créées par l'analytique des données et les charges applicatives lourdes associées à de grandes quantités de données. Les services de Consulting Dell Technologies vous permettent de bénéficier d'une migration de données uniquement vers une plate-forme complète, en fonction des besoins de votre entreprise. Tout commence par une évaluation de l'état actuel, conduisant à une conception de solution pour un état futur. Nous établissons ensuite un pilote de la migration pour tester et valider un exemple d'environnement. Après un pilote réussi, la solution est entièrement mise en œuvre.¹³

Pour commencer, [obtenez des informations supplémentaires](#) sur les services de migration Hadoop ou [contactez](#) un expert en services Dell Technologies.

Témoignage client : Epsilon

Examinons l'analytique prédictive au travail. Epsilon est un leader du marketing en matière de fidélisation, messagerie électronique et d'autres programmes marketing pour des marques de classe mondiale. Elle utilise un grand nombre de points de données pour personnaliser les messages personnalisés de ses nombreux destinataires, ce qui garantit des résultats optimaux pour les clients. Comme l'explique Robert Walden, CIO d'Epsilon : « Toutes nos activités sont centrées sur les données, ainsi que notre capacité à proposer le bon message au destinataire approprié, au moment opportun. »

La vitesse est primordiale pour que l'entreprise poursuive sa trajectoire actuelle qui consiste à doubler sa croissance d'une année à l'autre. Cela nécessite une infrastructure et un personnel qui peuvent suivre le rythme. Epsilon automatise tous ses processus d'admission lors du déploiement de ses serveurs Dell EMC PowerEdge R740xd. Le R740xd offre la flexibilité, l'évolutivité et les performances nécessaires pour répondre aux besoins de ces entreprises exigeantes. Le fait d'exécuter des charges applicatives telles que Hadoop pour appliquer l'IA et l'apprentissage automatique à ses processus de personnalisation du courrier électronique permet à Epsilon de tenir ses promesses.

Robert Walden, CIO d'Epsilon, est le mieux placé pour le dire : « Nous sommes concentrés sur la réussite de nos clients. Tout ce que nous faisons au sein de notre infrastructure et de nos environnements applicatifs consiste à faciliter la réussite. Plus les applications, les solutions, les services, l'infrastructure et le matériel sont efficaces, plus nous y parvenons. »

[Lisez l'étude de cas Epsilon.](#)



[Regardez la vidéo de l'étude de cas Epsilon.](#)



Epsilon au travail



Robert Walden, CIO d'Epsilon

Conclusion

Grâce à des serveurs modernes et automatisés, dotés de processeurs haute puissance et d'options de mémoire et de stockage généreuses, votre entreprise sera bien placée pour tirer le meilleur parti des nombreuses opportunités générées par l'analytique prédictive. Dans le même temps, il est essentiel de maîtriser les analyses prédictives et les aspects de l'analytique prédictive pour préserver la pertinence de l'IT dans votre organisation. Associez-vous à Dell EMC pour des solutions qui ont du sens et qui vous permettent de vous lancer dans l'apprentissage machine et l'analyse prédictive.

Pour en savoir plus, [contactez un commercial Dell EMC](#) ou visitez DellEMC.com/fr-fr/Servers.

¹ <https://www.dell EMC.com/fr-fr/collaterals/unauth/analyst-reports/products/servers/esg-three-transformational-compute-technologies-verified-to-accelerate-ai-and-business-value-en.pdf>

² [ibid.](#)

³ <https://spark.apache.org/>

⁴ <https://www.dell EMC.com/resources/fr-fr/asset/analyst-reports/products/storage/forrester-delivering-outcomes-by-automating-compute-infrastructure.pdf>

⁵ <https://spark.apache.org/docs/latest/hardware-provisioning.html>

⁶ [ibid.](#)

⁷ <https://blog.dell EMC.com/fr-fr/ride-ansible-revolution-dell-emc-openmanage/>

⁸ <https://www.dell.com/fr-fr/work/shop/povw/poweredge-r640>

⁹ <https://www.principledtechnologies.com/Dell/PowerEdge-R740xd-analytics-comparison-0719.pdf>

¹⁰ <https://www.dell EMC.com/fr-fr/solutions/data-analytics/hadoop/index.htm>

¹¹ <https://blog.dell EMC.com/fr-fr/dell-emc-can-help-dive-successful-hadoop-projects/>

¹² https://www.dell EMC.com/fr-fr/collaterals/auth/white-papers/products/ready-solutions/Ready_Bundles_for_Hadoop_-_Solution_Overview_China.PDF

¹³ https://www.dell EMC.com/fr-fr/collaterals/unauth/offering-overview-documents/services/H16645_Big_Data_Migration_for_Hadoop_svo.pdf