



# BIG DATA, EXIGENCES HORS NORMES

## LE STOCKAGE SCALE-OUT RÉPOND AUX BESOINS DES ENTREPRISES EN CE QUI CONCERNE LE BIG DATA

Le concept de Big Data n'a rien de nouveau et les outils et technologies permettant d'en tirer parti sont maintenant accessibles à toutes les entreprises, quelle que soit leur taille. Des secteurs comme les médias, le divertissement et la recherche et le développement manipulent depuis longtemps de gros volumes de données dans des formats non structurés, qui changent en temps quasi réel. Cependant, l'exploitation de ces données en vue d'en extraire des informations a un coût prohibitif et exige généralement des technologies sur mesure onéreuses. Désormais, grâce aux progrès des techniques de stockage et d'analyse prédictive, toutes les entreprises peuvent tirer le meilleur parti du Big Data pour identifier les tendances permettant de rendre leur modèle plus agile, novateur et compétitif.

Plusieurs facteurs décisionnels importants expliquent l'intérêt croissant pour le Big Data :

- **La volonté de mieux comprendre les clients** : les entreprises veulent identifier les clients les plus rentables et trouver des solutions pour réduire le taux d'attrition et améliorer les niveaux de revenu par client.
- **La quête de l'efficacité opérationnelle** : les entreprises souhaitent bénéficier d'une vue d'ensemble de leurs processus métiers, qu'il s'agisse du contrôle qualité, de la fabrication ou du service clients, afin d'en améliorer l'efficacité et la fluidité.
- **Les impératifs de gestion des risques** : concrètement, les entreprises souhaitent améliorer la sécurité informatique et réduire les fraudes, en détectant par exemple les menaces réseau avancées et les anomalies dans les transactions.
- **L'identification de nouvelles possibilités d'innovation** : la découverte de nouveaux produits, de nouveaux services et de nouvelles opportunités commerciales est un facteur déterminant.

**InfoWorld**  
Custom Solutions Group

**COMPUTERWORLD**  
Custom Solutions Group

**EMC<sup>2</sup>**

« Ces objectifs n'ont rien de nouveau, mais les avancées en matière de stockage et d'analyse prédictive permettent désormais aux entreprises de valoriser leurs données plus rapidement, plus facilement et de manière plus économique. Il est même possible d'inclure des sources de données externes, pour une analyse encore plus fine. » explique Generosa Litton, Directrice marketing du Big Data chez EMC.

Pour concrétiser ces objectifs métiers, les entreprises doivent repenser leur mode de gestion des données. Et parce que le Big Data est par définition volumineux, non structuré et en constante évolution, il implique une nouvelle approche en rupture avec les techniques utilisées pour les données traditionnelles. Pour tirer parti des opportunités offertes par le Big Data, les entreprises doivent se tourner vers des solutions technologiques intégrant les composants suivants :

- une infrastructure de stockage scale-out polyvalente, efficace et facile à gérer permettant d'obtenir rapidement et facilement des résultats ;
- une plate-forme d'analyse prédictive unifiée et agile pour les données structurées et non structurées, avec une couche de productivité autorisant la collaboration ;
- des fonctions permettant de devenir plus prédictif, en déclenchant des actions concrètes à partir des tendances identifiées et en favorisant la réactivité.

Ces composants essentiels permettent de mettre en place des infrastructures à la mesure des promesses du Big Data.

« Dans notre cas, le Big Data est un atout pour les parties prenantes de l'entreprise qui sont chargées d'effectuer une analyse prédictive critique à partir de Datasets volumineux. » constate Paul English, directeur informatique de 3TIER, une entreprise spécialisée dans l'analyse des risques liés aux énergies renouvelables.<sup>1</sup> « Nous avons mis en place une plate-forme capable de prendre en charge les besoins informatiques de notre entreprise, ainsi que les calculs hautes performances indispensables pour valoriser des Datasets volumineux et en faire bénéficier nos clients. En résumé, nous extrayons de la valeur et la monétisons. »

« Dans notre cas, le Big Data est un atout pour les parties prenantes de l'entreprise qui sont chargées d'effectuer une analyse critique à partir de Datasets volumineux. »

Paul English  
Directeur informatique de 3TIER



### ► Le Big Data transforme les énergies renouvelables

Voir comment 3TIER s'appuie sur EMC Isilon pour tirer parti du Big Data. Avec Isilon, 3TIER aide ses clients à trouver le juste équilibre entre risques et opportunités afin d'exploiter au mieux les projets d'énergie propre.

<http://youtu.be/iqSNXZL8GFK>

### Le Big Data fait la différence

Le Big Data apporte de nombreux avantages, mais impose aussi des contraintes à sa mesure. En effet, il présente plusieurs aspects clés qui le différencient des données d'entreprise traditionnelles :

- **Il est volumineux** : les entreprises génèrent et collectent de gros volumes de données ordinaires, mais le Big Data est tout simplement à une autre échelle.
- **Il est principalement non structuré** : le Big Data inclut des fichiers journaux Internet, des images numérisées, des vidéos de surveillance, des commentaires sur un site Web, des informations biométriques et d'autres types d'informations numériques. Ces données sont difficiles à ranger dans les lignes et les colonnes d'une base de données ; pourtant, elles représentent près de 80 % de la croissance des données aujourd'hui, selon les experts.
- **Il change constamment** : le Big Data évolue souvent en temps réel ou quasi réel. C'est le cas par exemple des commentaires des clients sur un site Web. Ces données doivent être collectées sur de longues périodes pour qu'il soit possible d'identifier des schémas et des tendances.

Il est désormais admis que pour tirer pleinement parti du Big Data, il est indispensable d'analyser l'ensemble des données numériques disponibles et de les soumettre à des itérations. De simples snapshots de données n'apportent que des éléments de réponse généralement insuffisants pour dresser un tableau complet ou résoudre un défi métier spécifique. D'autre part, la collecte et le stockage efficaces de ces données pour les soumettre à une analyse itérative ont un impact significatif sur les ressources de gestion informatique et de stockage d'une entreprise. Les professionnels du stockage se voient dans l'obligation de trouver des solutions de Big Data qui répondent aux attentes, sans pour autant dépasser des budgets déjà très serrés ou imposer des investissements importants en personnel spécialisé.

<sup>1</sup> [https://community.emc.com/community/events/live\\_events/theater\\_27](https://community.emc.com/community/events/live_events/theater_27)

Au vu de ces nouvelles exigences imposées par le Big Data et de l'importance d'un traitement correct des données, la plupart des entreprises estiment que la gestion de la croissance des données, le provisionnement du stockage et l'exécution d'analyses prédictives rapides, fiables et itératives constituent des priorités absolues, selon une étude récente menée par Enterprise Strategy Group (ESG).<sup>2</sup> Dans cette étude, réalisée en mars 2012, les 399 participants ont reconnu unanimement que le Big Data mettait leurs infrastructures informatiques actuelles à rude épreuve et qu'ils envisageaient de mettre en place des infrastructures évolutives dans leurs datacenters. C'est précisément cette surcharge qui dicte une nouvelle approche du stockage.

« Les entreprises ont pris conscience de la valeur de leurs données et la plupart cherchent un moyen d'en tirer le meilleur parti afin d'optimiser leurs résultats. Elles cherchent également à découvrir des éléments d'information leur permettant d'identifier de nouvelles opportunités commerciales, de nouveaux services aux clients et de meilleures méthodes de réduction des coûts. » explique Mike Noble, Directeur du marketing produits chez EMC. « Elles savent aussi qu'il leur faut protéger ces données, les conserver en lieu sûr et se conformer aux exigences réglementaires. »

« Les entreprises ont pris conscience de la valeur de leurs données et la plupart cherchent un moyen d'en tirer le meilleur parti afin d'optimiser leurs résultats. »

Mike Noble  
Directeur du marketing produits, EMC

## Repenser le stockage

Le Big Data implique d'accroître la capacité, l'évolutivité et d'assurer une accessibilité efficace sans augmenter les ressources nécessaires. Les architectures de stockage classiques ont été conçues pour évoluer en fonction du volume de données. Dans ce cas, l'évolution consiste à augmenter la capacité en ajoutant du matériel de stockage et en créant des silos, ce qui ne résout pas le problème de l'impact des données supplémentaires sur les performances.

« Dans les architectures de stockage traditionnelles, les systèmes reposant sur un contrôleur RAID entraînent la prolifération des systèmes de stockage et créent un environnement organisé en silos. » explique Eric Seidman, Directeur du marketing produits chez EMC. « Pourtant, les entreprises doivent être en mesure de consolider leurs ressources au sein d'une infrastructure de stockage



### ➔ EMC Isilon : le système NAS le plus rapide au monde pour le Big Data

[Le système NAS scale-out EMC Isilon](http://youtu.be/B1OT06ooGhc) a été certifié par SPECsfs comme le système NAS le plus rapide au monde sur la base de résultats de test validés démontrant des performances de plus de 1,1 million d'E/S par seconde NFS et de plus de 1,6 million d'E/S par seconde CIFS. Les performances exceptionnelles d'Isilon permettent aux entreprises de générer de la valeur à partir du Big Data tout en respectant des contraintes de budget et de temps. <http://youtu.be/B1OT06ooGhc>

unique, hautement évolutive. Elles ont également besoin de fonctions automatisées de gestion, de provisionnement et de hiérarchisation pour faire face à la croissance rapide du Big Data. »

Les architectures de stockage doivent être centrées sur le Big Data et inclure les caractéristiques suivantes :

- **Évolutivité** pour prendre en charge des datastores toujours plus volumineux, y compris la possibilité d'ajouter aisément des ressources de stockage supplémentaires au gré des besoins.
- **Hautes performances** pour assurer des temps de réponse et d'acquisition de données aussi courts que possible afin de ne pas ralentir le rythme des activités.
- **Efficacité maximale** afin de réduire les coûts de stockage et dérivés du datacenter.
- **Simplicité opérationnelle** pour rationaliser la gestion d'un environnement de données à grande échelle sans avoir à recruter des informaticiens.
- **Protection des données d'entreprise** afin d'assurer une haute disponibilité pour les utilisateurs internes et la continuité de l'activité en cas de sinistre.
- **Interopérabilité** pour intégrer des environnements très complexes et mettre en place une infrastructure agile qui prenne en charge une vaste gamme d'applications métiers et de plates-formes d'analyse prédictive, telles que Hadoop.

<sup>2</sup> Rapport ESG Research, juillet 2012 : « The Convergence of Big Data Processing and Integrated Infrastructure »

## L'approche EMC

EMC Isilon offre des solutions de stockage scale-out extrêmement flexibles, parfaitement adaptées aux besoins de stockage et d'analyse prédictive du Big Data. Pilotée par un système d'exploitation novateur, Isilon OneFS, la plate-forme Isilon de stockage scale-out rattaché au réseau (NAS) permet de mettre en place une infrastructure aussi efficace que fiable, capable d'évoluer facilement en termes de capacité et de performances. Elle simplifie et réduit la gestion du stockage à quelques heures par semaine et optimise les ressources en implémentant une stratégie de stockage automatisée et hiérarchisée.

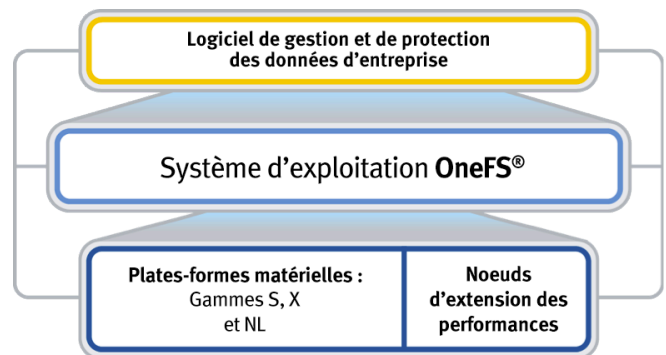
Elle optimise également le retour sur investissement, du fait de taux d'utilisation du stockage particulièrement élevés, et réduit aussi bien les dépenses d'investissement que les coûts opérationnels.

Vous trouverez ci-dessous quelques-unes des principales caractéristiques **de la plate-forme de stockage scale-out EMC Isilon** :

- ✓ **Évolutivité à grande échelle** : évolue facilement jusqu'à plus de 15 Po de stockage au sein d'un système de fichiers et d'un volume uniques simples à gérer.
- ✓ **Performances exceptionnelles** : assure un débit de plus de 100 Go et établit un record SPECSfs, avec 1,6 million d'E/S par seconde dans un environnement CIFS/SMB.
- ✓ **Efficacité optimale** : renforce l'efficacité avec des taux d'utilisation du stockage supérieurs à 80 % en empêchant la formation de « points sensibles » et en automatisant les fonctions de gestion et de hiérarchisation du stockage.
- ✓ **Sécurité et protection des données d'entreprise** : assure une sauvegarde et une protection efficaces des données, une reprise après sinistre fiable et la rétention des données de type WORM.
- ✓ **Simplicité de gestion** : permet de faire évoluer la capacité et les performances sans alourdir les dépenses d'exploitation (OPEX), y compris dans un système de fichiers unique, un volume unique et espace de nommage global.
- ✓ **Flexibilité opérationnelle** : inclut la prise en charge intégrée d'une vaste gamme de protocoles standard, tels que NFS, SMB, HTTP, FTP et iSCSI.
- ✓ **Prise en charge native de Hadoop** : assure la prise en charge native de Hadoop en intégrant le système de fichiers HDFS (Hadoop Distributed File System) dans le système d'exploitation EMC Isilon OneFS afin de faciliter le déploiement d'une solution complète de stockage.

## Conclusion

Face à la croissance du volume des données non structurées au sein des entreprises, la mise en place de nouvelles approches de gestion de ces données s'impose comme une nécessité. De même, un système de stockage efficace et évolutif est indispensable pour gérer cette croissance. L'étude d'ESG a fait ressortir trois grands avantages de l'implémentation d'un stockage scale-out : 1) évolutivité accrue, 2) performances accrues (en termes d'E/S comme de débit) et 3) disponibilité accrue des données. Le stockage scale-out, associé à des outils puissants d'analyse prédictive capables de mettre en lumière de précieuses informations dans ce foisonnement de données, constitue la configuration idéale pour exploiter au mieux le Big Data.



Reposant sur le système d'exploitation OneFS, les solutions de stockage scale-out EMC Isilon proposent aux utilisateurs une large gamme d'options pour satisfaire leurs besoins en matière de stockage du Big Data.

## Étapes suivantes

Pour en savoir plus sur les solutions Big Data d'EMC et découvrir comment elles peuvent contribuer à transformer votre entreprise, consultez les pages :

<http://france.emc.com/BigData>

[www.isilon.com](http://www.isilon.com)

[twitter.com/emcbigdata](https://twitter.com/emcbigdata)

**InfoWorld**  
Custom Solutions Group

**COMPUTERWORLD**  
Custom Solutions Group

**EMC<sup>2</sup>**