

# Le Big Data est un atout stratégique. Votre protection des données est-elle à la hauteur ?



## Les nouvelles solutions de stockage sont adaptées aux infrastructures d'aujourd'hui

Alors qu'il ne représentait voici quelques années qu'une poignée de projets expérimentaux, le Big Data a aujourd'hui acquis un statut stratégique au sein des entreprises numériques, et son importance ne fait que s'accroître. D'après IDC, d'ici 2020, les entreprises capables d'analyser toutes les données pertinentes et d'en tirer des informations utiles engrangeront 430 millions de dollars de plus que leurs homologues moins portées sur l'analytique<sup>1</sup>. Un temps limitée à des opérations occasionnelles, l'analytique Big Data est aujourd'hui effectuée quotidiennement dans de nombreuses entreprises, notamment Amazon, Walmart et UPS.

### Enjeux de sauvegarde du Big Data

L'ascension du Big Data jusqu'au sommet de l'échelle de l'information impose de le traiter avec soin. Auparavant considérée comme accessoire, la sauvegarde du Big Data est désormais primordiale et doit s'effectuer au bon moment et en toute fiabilité. Les responsables IT prennent la mesure de l'enjeu. Dans un récent sondage ESG sur les cinq grandes priorités de modernisation du datacenter en 2018, 31 % des sondés indiquent que l'amélioration de la sauvegarde et récupération des données est une priorité élevée. Les ressources de sauvegarde et récupération vont être de plus en plus consacrées au Big Data. IDC prévoit que le marché du Big Data représentera 210 milliards de dollars d'ici à 2020, contre 150,8 milliards en 2017<sup>2</sup>.

Pour sauvegarder les données avec efficacité, de nombreuses entreprises doivent procéder à des changements opérationnels importants. Les architectures existantes, qui étaient suffisantes pour les charges de travail antérieures, doivent être renouvelées avec des architectures modernes afin de protéger les environnements exécutant les charges de travail Big Data telles que Hadoop. Voici quelques exemples.

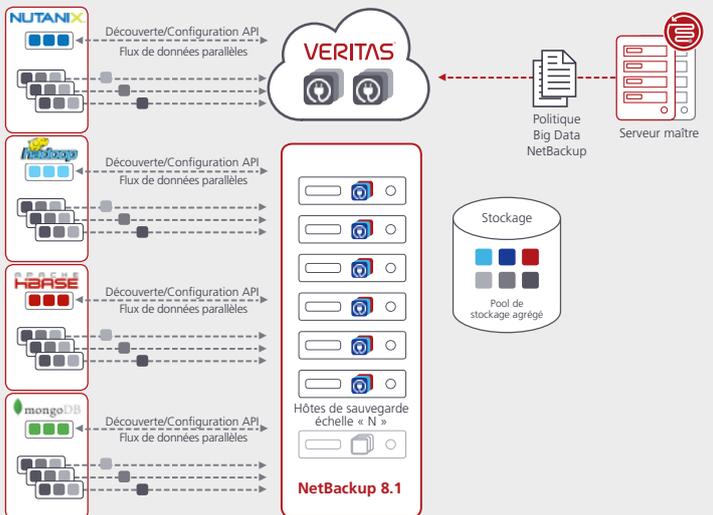
- ▶ **De nombreuses organisations ont implémenté des architectures de sauvegarde client-serveur**, qui fonctionnent bien avec les charges de travail conventionnelles, mais n'ont jamais été conçues pour traiter les grandes quantités de données du Big Data ou analyser en continu des données non structurées. Ces dernières, et tous les points de données dans le temps correspondants, sont tout aussi essentiels que les vastes quantités de données transactionnelles contenues dans les systèmes ERP ou CRM traditionnels.
- ▶ **La sauvegarde de plusieurs copies des mêmes ensembles de données** dilapide les ressources de stockage. Une architecture de sauvegarde efficace pour les environnements Big Data doit éliminer les ensembles de données répliqués de manière superflue afin de préserver les ressources de stockage, d'accélérer le traitement des données et de réduire les coûts.
- ▶ **L'intégrité des données est importante pour les entreprises** parce qu'elles prennent des décisions en s'appuyant sur des analyses de données stratégiques. La corruption des données à la suite d'une erreur humaine entraîne généralement la réplification d'informations incorrectes dans les environnements Big Data Hadoop.
- ▶ **Dans les entreprises qui exigent des processus analytiques journaliers**, les accords de niveau de service (SLA) imposent que les tâches de sauvegarde s'effectuent rapidement. Et les règles de conformité obligent généralement à prévoir une récupération sans délai des données.
- ▶ **Les architectures de sauvegarde dépassées** freinent le trafic Big Data. Par exemple, les blocs de données répliqués qui transitent de plusieurs points de données vers le NameNode Hadoop créent un goulet d'étranglement qui ralentit les flux de données Big Data.

1 DC FutureScape: Worldwide Big Data, Business Analytics and Cognitive Software Predictions, International Data Corporation, 2016.

2 Worldwide Semiannual Big Data and Analytics Spending Guide, IDC, 2017.

- **La sauvegarde des environnements Hadoop** a amené de nombreuses entreprises à créer des solutions de contournement en plusieurs étapes qui prévoient notamment l'ajout de clusters Hadoop secondaires et de dispositifs de stockage supplémentaires. La complexité de ces mesures crée un risque de perte de données à chaque étape et peut empêcher de respecter l'objectif de délai de récupération (RTO).

### Présentation de l'architecture



### La solution Veritas : NetBackup 8.1

Veritas NetBackup 8.1 est conçu pour répondre aux besoins de sauvegarde Big Data. Il comprend le framework Veritas NetBackup Parallel Streaming, qui est conçu pour les charges de sauvegarde et de clusters multinœud à grande échelle dans les environnements Hadoop. En sauvegardant le système de fichier HDFS (Hadoop Distributed File System) de manière native, le logiciel NetBackup 8.1 sans agent évite de recourir à des solutions de contournement complexes. Une architecture sans agent est un concept important pour les environnements modernes. Comme il n'y a pas d'agent sur les nœuds de cluster, vous n'avez pas à vous soucier de leur gestion. Vous gagnez donc du temps et réalisez des économies. Et vous n'avez pas non plus à vous préoccuper de mettre à niveau Hadoop.

Le fonctionnement est le suivant : un module externe téléchargeable pour NetBackup 8.1 recherche les métadonnées du cluster Hadoop, puis partitionne et répartit la charge de travail entre les hôtes de sauvegarde, en exploitant le pool de ressources constitué par les serveurs de médias ou clients NetBackup. Les ressources de tâches sont gérées de manière à maximiser les performances de sauvegarde et récupération sur les flux parallèles lorsque chaque nœud Hadoop envoie des données aux hôtes de sauvegarde. Les politiques Big Data sont totalement intégrées dans l'interface utilisateur de NetBackup. Grâce à l'utilisation de modules externes, vous n'avez pas à attendre la prochaine version de Hadoop.

La solution Veritas se différencie de l'architecture Big Data Hadoop, qui réplique trois copies des données de chaque nœud via le NameNode et envoie le tout au stockage via les hôtes de sauvegarde, ce qui induit un gaspillage du stockage et alourdit ses coûts. Au contraire, NetBackup 8.1 capture et stocke seulement les données uniques, ce qui réduit des deux tiers l'espace de stockage utilisé et permet d'économiser 67 % des coûts de stockage secondaire. En outre, le processus de sauvegarde est trois fois plus rapide.

Même si Hadoop se charge de répliquer les données, il ne fournit pas une protection des données à un point dans le temps, pour prévenir les erreurs humaines et la corruption des données. NetBackup 8.1 comble cette lacune en effectuant une sauvegarde à un point dans le temps pour préserver les données critiques tout en respectant les SLA et les obligations de conformité.

Comme la solution Veritas repose sur un module externe NetBackup sans agent (plutôt qu'un agent placé sur le cluster Hadoop), les performances analytiques du cluster ne sont pas altérées par les processus de sauvegarde et récupération. En outre, elle élimine le casse-tête lié à la gestion des agents, notamment l'installation, l'administration et la mise à niveau.

L'évolutivité est importante pour les environnements Big Data, et la solution Veritas peut être étoffée de manière à faire face à l'expansion des volumes de données Hadoop. Elle permet ainsi d'ajouter des nœuds NetBackup dès lors que des nœuds sont ajoutés au cluster Hadoop.

De plus, NetBackup 8.1 prend en charge plusieurs environnements au-delà de Hadoop, comme l'infrastructure hyperconvergente Nutanix, Apache HBASE et MongoDB. NetBackup prend en charge 40 environnements cloud grâce à des connecteurs cloud intégrés.

### Conclusion

L'analytique Big Data a quitté les laboratoires de test pour s'imposer dans l'entreprise et elle représente aujourd'hui un atout stratégique sans précédent au sein de nombreuses grandes organisations. Pour ne pas mettre en jeu leur prospérité commerciale, les entreprises doivent rapidement analyser de grandes quantités de données en temps réel.

L'infrastructure qui sous-tend la sauvegarde et récupération Big Data doit impérativement s'adapter à ce rôle stratégique. Le framework Veritas NetBackup Parallel Streaming lève les obstacles inhérents des environnements Big Data qui freinent la sauvegarde. Veritas NetBackup 8.1 accélère les performances tout en réduisant les risques et en consommant moins de ressources de stockage. Cette solution ouvre ainsi de nouveaux horizons pour les entreprises numériques qui cherchent à acquérir un avantage stratégique grâce à l'analytique Big Data.

Pour en savoir plus, cliquez [ici](#).