## Questions d'ingénierie :

# entretien avec Bian Gold de Pure Storage sur l'analyse Big Data pour Apache Spark

Q&R

La technologie Apache® Spark™ est devenue un incontournable pour les équipes de développement qui veulent utiliser un moteur de données in-memory ultra rapide pour l'analyse big data. Spark est une plate-forme open-source flexible, qui permet aux développeurs de créer des applications en Java, Scala, Python ou R. Avec Spark, les équipes de développement ont la possibilité d'accélérer les applications d'analyse en fonction des ordres de grandeur.

La croissance rapide de Spark ne s'est pas déroulée sans quelques défis à relever. Pour répondre aux besoins en termes de capacité et de performances d'applications riches en données, la plupart des organisations ont opté pour de vastes déploiements de systèmes de fichiers HDFS Hadoop, composés de racks de disques rotatifs. Mais les choses vont changer.

Pure Storage, pionnier des baies flash basées sur des blocs, a développé une technologie appelée FlashBlade, conçue spécialement pour les environnements de fichiers et d'objets. Avec FlashBlade, les équipes informatiques disposent maintenant d'une solution de stockage partagé facile à gérer, qui fournit les performances et la capacité nécessaires pour autoriser les déploiements Spark sur site.

Pour mieux comprendre les problématiques de stockage inhérentes à Spark et la manière dont FlashBlade y répond, Brian Gold de Pure Storage s'est entretenu avec Al Perlman, routard du journalisme technologique chez TechTarget, pour une discussion approfondie sur les tendances et les opportunités qu'offre Spark. Brian Gold est directeur R&D chez Pure Storage et l'un des fondateurs de l'équipe de développement FlashBlade. Ci-dessous les temps forts de l'échange.

**TechTarget :** Brian, quelle est votre vision d'Apache Spark et qu'est-ce qui explique l'engouement suscité ?

**Brian Gold :** Pour moi, Spark est une interprétation moderne d'un framework de traitement des données à hautes performances. Nous le voyons comme un moteur de base de données conçu à l'envers. Tout a commencé au sein de l'AMPLab de l'université de Berkeley où l'équipe est partie de fonctions primitives très puissantes et d'abstractions d'assez bas niveau pour permettre une manipulation de données évoluant facilement au niveau des cœurs de processeurs et des ordinateurs qui composent un centre de données.

PURESTORAGE®
Hello Possible™

TechTarget | Custom Media

À ce stade, quelques-uns des points les plus problématiques des implémentations antérieures d'Hadoop MapReduce ont alors été corrigés. Plus récemment, Spark a été enrichi d'abstractions de plus haut niveau, s'approchant de plus en plus d'une plate-forme analytique déclarative, telle qu'on pourrait la trouver dans un système de base de données complet.

Dans la dernière version 2.0, Spark incorpore des optimisations SQL impressionnantes qui étendent ses facultés d'application dans le domaine de l'entreposage de données traditionnel. En outre, les développeurs Spark ont continué à faire avancer la technologie de traitement central des données, proposant des performances de pointe en termes d'algorithmes (traitement de flux impressionnants et comparaison de tri).

Je dirais donc qu'en ajoutant ces moteurs SQL généraux hautement optimisés et certaines fonctionnalités de ses extensions de traitement de flux, Spark va désormais bien plus loin que ses fondements initiaux. Conçu à l'origine comme un système plus efficace pour l'analyse de lots et les algorithmes itératifs, il s'agit aujourd'hui d'un entrepôt de données complet, qui intègre des données structurées en provenance d'un magasin transactionnel, avec des flux de données en temps réel et qui, à partir d'une analyse à grande échelle, fournit des informations assez conséquentes.

Spark va désormais bien plus loin que ses fondements initiaux. Conçu à l'origine comme un système plus efficace pour l'analyse de lots et les algorithmes itératifs, il s'agit aujourd'hui d'un entrepôt de données complet, qui intègre des données structurées en provenance d'un magasin transactionnel, avec des flux de données en temps réel et qui, à partir d'une analyse à grande échelle, fournit des informations assez conséquentes.

**TT :** Comment les clients peuvent-ils bénéficier de ces analyses à grande échelle et pourquoi la solution Spark constitue-t-elle un bon choix pour eux ?

**BG :** Si vous prenez chacun de nos clients, vous constaterez que les données et les informations qu'ils obtiennent à partir de ces données occupent une place centrale dans leur activité. Le volume de données que vous pouvez collecter dans une entreprise s'est développé à une vitesse telle, que les systèmes existants pour stocker et traiter ces données n'arrivent plus à tenir le rythme.

Ce n'est pas nouveau. Le sujet central est exactement le même qu'il y a sept à dix ans de cela, lorsqu'a eu lieu la première déferlante big data. Mais de plus en plus, on voit que les entreprises parviennent à exploiter ces informations pour obtenir un avantage conséquent sur leurs concurrents, un avantage qui leur permet d'offrir plus à leurs clients.

Nous constatons que nos clients utilisent Spark pour deux types de tâches qui sont liées entre elles.  D'abord, il faut dire que les données big data sont en générales assez "vilaines" : déstructurées, mal formatées, contenant des erreurs qu'il faut corriger. Spark a les qualités requises pour permettre à nos clients d'effectuer une des tâches les plus importantes du secteur, quoique souvent oubliée : nettoyer et transformer les données.

Une fois que les clients disposent d'un ensemble de données mieux organisé, le second cas d'utilisation le plus répandu peut en grande partie être affilié à l'intelligence artificielle ou à l'apprentissage automatique.  Cela peut prendre la forme de requêtes analytiques relativement traditionnelles qui exécutent des rapports dans un entrepôt de données. Mais de plus en plus, on voit nos clients détecter des anomalies dans leur données, construire des modèles de prédiction avancés, relier des événements entre des ensembles de données disparates et des analyses profondes similaires.

Je fais partie des ingénieurs qui ont contribué à concevoir le produit chez Pure. En tant que tel, je suis particulièrement intéressé par cette évolution parce que Pure Storage, tout comme n'importe lequel de nos clients, détient des ensembles de données de l'ordre du pétaoctet, qui exigent des outils sophistiqués pour générer une visibilité plus large.

**TT :** À quel genre de visibilité plus large faites-vous allusion ? Quel impact sur l'utilisation des outils d'analyse par une entreprise moderne pour générer des opportunités commerciales ?

**BG :** Je vais vous décrire comment nous utilisons les données de grande échelle chez Pure et vous expliquer à quoi ressemble notre infrastructure d'analyse. Dans la mesure où le client est d'accord, chacun de nos système peut renvoyer des données télémétriques internes à Pure. Nous en collectons d'ailleurs beaucoup : sur la santé du stockage flash, sur les performances de mise en réseau, la gestion des métadonnées internes, le comportement du protocole front-end, les statistiques sur la réduction des données, l'utilisation de l'espace etc.

En plus des données sur le terrain, nous intégrons également des journaux de test en provenance de nos lignes de fabrication pour contrôler la santé du matériel, à partir du moment où l'on soude les puces sur la carte. Dès lors que nous développons des logiciels et exécutons des tests, nous voulons aussi comparer les performances de nos tests d'ingénierie internes. Étudier comment ils se comportent en fonction de la manière dont les clients interagissent avec nos produits.

À force, tous les déploiements de nos baies de stockage dans le monde entier ont fini par générer un ensemble de données massif, à savoir plusieurs pétaoctets de données télémétriques. Nous utilisons ces informations pour comprendre comment nos produits fonctionnent dans le monde réel, comment l'expérience de nos clients pourrait être encore améliorée, quels problèmes nous pourrions résoudre de façon proactive.

En interne, nous avons une blague à ce sujet : nous avons dû créer notre deuxième produit, FlashBlade, rien que pour analyser toutes les données télémétriques que nous collectons avec notre premier produit, FlashArray. C'est un processus important pour toute l'entreprise : le support technique, les ingénieurs, nos équipes en charge de la commercialisation.

Grâce à toutes ces analyses de données en interne, les clients sont extrêmement contents des produits dont ils bénéficient. Nous en avons la preuve grâce à des éléments tels que notre score d'avis positifs (Net Promoter Score) de 83 certifié par Satmetrix. Nous avons travaillé dur pour obtenir ce résultat et nous sommes très fiers de la satisfaction de nos clients vis-à-vis des produits et des services de notre entreprise.

**TT :** Nous comprenons les avantages que représentent toutes ces données si elles peuvent être utilisées intelligemment et stratégiquement. Mais nous savons aussi qu'elles génèrent une pression énorme sur l'infrastructure de stockage. Du point de vue du stockage, quels sont les problématiques inhérentes à Spark pour les utilisateurs et pour l'équipe informatique qui les assiste ?

C'est ce qu'on entend constamment de la part des clients qui utilisent Apache Spark et d'autres outils d'analyse de données modernes. Ils ont besoin de capacités de stockage importantes. Ils ont besoin d'un accès rapide à toutes leurs données. Et ils ont besoin de simplicité dans l'extension et la gestion de toutes ces données. FlashBlade agit efficacement dans ces trois domaines.

**BG :** En résumé, nous concevons FlashBlade pour permettre le fonctionnement de cette nouvelle génération de charges de travail analytiques, qui doivent être capables de s'adapter facilement à des évolutions de capacités et de performances. On ne doit pas noyer nos utilisateurs sous une complexité et une charge de gestion et d'administration superflues.

C'est ce qu'on entend constamment de la part des clients qui utilisent Apache Spark et d'autres outils d'analyse de données modernes. Ils ont besoin de capacités de stockage importantes. Ils ont besoin d'un accès rapide à toutes leurs données. Et ils ont besoin de simplicité dans l'extension et la gestion de toutes ces données. FlashBlade agit efficacement dans ces trois domaines et bon nombre de nos clients initiaux confirment ce que nous avons nous-mêmes observé. Les choses vont même plus vite que nous ne l'avions prévu.

**TT :** Qu'est ce que la solution FlashBlade et quelle a été la méthode de Pure pour la développer ?

**BG :** FlashBlade est notre système de stockage évolutif 100 % flash dédié au stockage de fichiers et d'objets. Pour remettre les choses dans leur contexte, Pure a commencé par un produit intitulé FlashArray, un périphérique de stockage de blocs 100 % flash. Se concentrant sur la réduction des données et sur une architecture conçue pour les supports solid-state, FlashArray fournit des performances 100 % flash à un coût inférieur ou égal à celui de baies de disques professionnelles. Depuis la création de l'entreprise en 2009, FlashArray a été confronté à quelques-uns des scénarios de stockage les plus stratégiques pour les entreprises.

À mesure du succès de FlashArray sur le marché, nous avons constaté qu'il existait un ensemble annexe de charges de travail et de cas d'utilisation qui exigeaient un protocole d'accès axé sur les fichiers. Pour ces charges de travail, de nombreuses machines doivent pouvoir accéder à un ensemble de données partagé. À ce stade, nous avions donc un nouvel objectif clair : un système de fichier distribué. Il nous a en effet semblé qu'un système de stockage 100 % flash et conçu à partir de zéro pourrait apporter une véritable transformation à nos clients. Mais nos produits existants n'étaient pas en mesure d'offrir cela et très franchement, aucun des produits de nos concurrents ne le pouvait non plus.

FlashBlade est un système bénéficiant d'une conception matériel/logiciel conjointe. Nous avons tout créé, depuis les premiers prototypes jusqu'au système de fichiers et d'objets modulaire à hautes performances entièrement fonctionnel.

**TT :** Qu'est-ce qui fait de FlashBlade un bon choix pour Spark ?

**BG :** Prenons le déploiement typique de Spark aujourd'hui. Spark est issu de l'écosystème Hadoop, donc par défaut on a généralement le système de fichiers Hadoop, HDFS. Du fait de son évolutivité, le système HDFS peut atteindre de très grandes capacités ; c'est pour cela qu'il a été conçu.

Mais il a également été créé à une époque où, pour obtenir des capacités étendues, il fallait des disques rotatifs. Partir d'une plate-forme de stockage composée de disques rotatifs entraîne souvent des problèmes de performances (ou un surprovisionnement massif d'axes de disques pour éviter de tels

problèmes) et constitue un cauchemar opérationnel compte tenu des taux de panne élevés des disques durs.

C'est pour cet environnement que le système HDFS a été créé et il présente de profondes lacunes en cas de passage à un environnement 100 % flash ou basé SSD. Par ailleurs, exécuter Spark avec FlashBlade procure trois avantages par rapport au système HDFS et aux autres systèmes traditionnels de stockage de fichiers et d'objets.

- Premier avantage : les performances 100 % flash. Au-delà des chiffres, les performances de FlashBlade autorisent la consolidation, même à grande échelle.

- Deuxième avantage : l'efficacité du stockage. Nous avons conçu FlashBlade autour d'un système efficace d'effacement du codage qui permet d'atteindre une redondance de N + 2 avec une charge minimale de répliques supplémentaires. Le déploiement HDFS classique comprend au moins trois copies de chaque donnée (notamment en raison des taux de panne extrêmement élevés des disques durs dans leur ensemble).

- Troisième avantage : la conception du système et la simplicité qu'il apporte. Ces avantages font partie de l'approche globale qu'applique Pure à tous ses produits. Nous voulions un système de stockage qui soit extrêmement simple à installer, configurer, utiliser et développer. Si vous avez un ensemble de données et qu'il vous faut davantage de capacités, il vous suffit d'ajouter une lame à l'architecture de châssis modulaire. Aucune nouvelle répartition manuelle des données sur bande (data striping) n'est nécessaire. Rien qui nécessite l'intervention de l'utilisateur. Vous n'avez pas à brancher de câbles supplémentaires. Vous n'avez rien d'autre à faire que glisser une nouvelle lame offrant 40 ou 50 To de capacité effective supplémentaire.

Tous ces avantages sont à portée de main des clients qui exécutent Spark avec leurs données dans FlashBlade. Un cluster Spark peut maintenant être exécuté sur des serveurs de calcul dénués de toute exigence de stockage particulière. Et ce cluster peut être dimensionné comme nécessaire pour maximiser le parallélisme potentiellement accessible pour une charge de travail donnée.

**TT :** Pouvez-vous nous donner un exemple de ce modèle avec un cas d'utilisation réel ?

**BG :** Une chose particulièrement intéressante dans ce modèle de déploiement, c'est qu'il permet d'exécuter, juste à côté, un autre cluster sur le même ensemble de données. Très souvent, les clients ont un cluster Spark de grande dimension qui exécute un pipeline de données de production. Une charge de travail particulière accède à ces données, il peut s'agir de données de streaming avec de nouvelles ligne de journal en provenance de différents capteurs ou, dans notre cas, d'enregistrements de fabrication ou de données télémétriques des clients.

Vous pouvez également accéder aux enregistrements anciens en même temps. Vous avez donc des pipelines de données complexes qui sont exécutés dans un environnement de production ; ils doivent en permanence être allumés, en état

Récemment, nous avons effectué un test comparatif entre le déploiement d'un entrepôt de données sur site dans Spark exécutant FlashBlade et un cluster Spark de volume similaire avec un système de stockage HDFS sur Amazon. Nous nous sommes aperçus qu'en moyenne, les requêtes étaient exécutées environ 3 fois plus rapidement avec FlashBlade et jusqu'à 6 fois plus vite pour les opérations particulièrement gourmandes en données.

Parmi les aspects novateurs de FlashBlade, il y a notamment le fait que nous créons et concevons conjointement tous les logiciels avec le matériel, jusqu'aux composants flash NAND de premier niveau. Ainsi, nous pouvons surveiller l'état et les performances du moindre bit de mémoire flash, à chaque niveau de lecture et écriture. Nous avons désormais accès à de toutes nouvelles formes de données, que nous pouvons utiliser pour optimiser certains des algorithmes internes et construire des modèles prédictifs.

de marche et ils exigent un débit plutôt élevé. Toutes ces données peuvent être conservées sur FlashBlade.

En même temps, vous pouvez avoir une équipe de science des données chargée des analyses exploratoires, à laquelle vous pouvez attribuer son propre cluster Spark indépendant. Comme il ne s'agit plus que d'un groupe d'ordinateurs sur un réseau haut débit, l'équipe peut effectuer ses travaux exploratoires directement sur le même ensemble de données.

Je parle d'un ensemble de données de plusieurs pétaoctets. Vous ne pouvez pas forcément vous permettre d'en faire une copie.  Ou bien vous ne souhaitez pas en faire une copie et avoir la charge supplémentaire d'exécuter une infrastructure de stockage entièrement séparée pour l'équipe de science des données. Vous pouvez donc exécuter la même classe d'algorithmes que celle que vous exécuteriez en production, mais en effectuant des expériences plus avancées.

À mesure des progrès constatés, vous pouvez déployer ces nouveaux algorithmes dans votre environnement de production. En ce qui nous concerne, chez Pure, nous avons des centaines de personnes qui dépendent de la qualité des analyses que nous exécutons dans nos propres pipelines de données. Par conséquent, nous cherchons constamment des moyens de faire plus avec ce que nous avons. En revanche, on ne peut pas se permettre d'interruptions parce qu'il s'agit de production.

Spark fournit de nouvelles opportunités innovantes d'utilisation des données à tous les niveaux, pour les entreprises telles que la nôtre. FlashBlade peut devenir un élément central de votre mode de développement, pas uniquement en ce qui concerne les pipelines de production, mais également pour les pipelines de science des données qui peuvent exister grâce au découplage du stockage et du calcul, ainsi qu'aux performances et à la simplicité qu'offre FlashBlade.

**TT :** Pourquoi ne pas exécuter Spark et autres outils d'analyse similaires sur le cloud ? C'est plus ou moins le modèle dominant jusqu'à présent. Pourquoi chercher à installer Spark sur site ?

**BG :** D'abord il faudrait clarifier ce que vous entendez par « cloud ». Pure a eu de nombreux contacts avec des entreprises SaaS et avec des entreprises qui créent leurs propres clouds privés. Il s'agit de fournisseurs de clouds en tous genres et jusqu'à présent, tous les avantages qui ont été avancés sont précisément ceux pour lesquels ces entreprises déploient de plus en plus les produits Pure pour différentes parties de leur infrastructure de données. Spark et l'analyse de données sont au centre de tout ça.

En général, quand on parle de cloud, il s'agit en fait de fournisseurs de cloud public (et ils sont très bons dans leur domaine). Dans bien des cas, le déploiement flexible dont vous bénéficiez à un coût de démarrage très intéressant présente de nombreux avantages. Les clouds publics sont aussi un excellent moyen pour commencer à expérimenter cette nouvelle classe d'outils analytiques. Tout comme Spark, ils peuvent suivre une courbe d'apprentissage accélérée. Et un petit cluster éphémère, placé dans une infrastructure de cloud public, peut constituer un excellent point de départ pour comprendre comment une entreprise peut générer de la valeur grâce à Spark.

Cependant, lorsque l'ensemble de données commence à prendre de l'ampleur, que les besoins de traitement augmentent et que l'on commence à percevoir de réels avantages, nous avons constaté qu'exécuter ces charges de travail sur une infrastructure dédiée sur site présente de sérieux atouts (ce que confirme la grande majorité de nos clients). En découplant le stockage haute performance (dans notre cas FlashBlade) des fonctions de calcul générales, un déploiement de cloud privé ou sur site peut gagner l'agilité d'un cloud public sans y sacrifier ses performances ni encourir le risque de dépassement sévère des coûts.

Récemment, nous avons effectué un test comparatif entre le déploiement d'un entrepôt de données sur site dans Spark exécutant FlashBlade et un cluster Spark de volume similaire avec un système de stockage HDFS sur Amazon. Nous nous sommes aperçus qu'en moyenne, les requêtes étaient exécutées environ 3 fois plus rapidement avec FlashBlade et jusqu'à 6 fois plus vite pour les opérations particulièrement gourmandes en données.

Pourtant, les coûts liés au fonctionnement d'un cluster HDFS toujours allumé dans un environnement de cloud public deviennent rapidement prohibitifs lorsque la taille du cluster est élevée. Le cloud public est excellent pour permettre une expansion soudaine puis s'adapter aux besoins. Alors que, pour de nombreux déploiements, on cherche des interfaces objets moins chères, telles que S3 dans l'écosystème Amazon, Les performances font défaut pour les tâches d'analyse itérative.

Actuellement, le cloud public ne propose pas de solution vraiment adaptée. Nous avons vu des clients essayer d'introduire de la hiérarchisation et apporter en fait encore plus de complexité pour contourner un goulot d'étranglement de stockage et de calcul. FlashBlade et cette architecture découplée de stockage/calcul vous offrent la souplesse et la capacité d'expansion soudaine que l'on trouve souvent dans le cloud public, mais en préservant les performances et les avantages en termes de coûts. Dès lors que vous connaissez la taille de votre ensemble de données et les besoins de votre production en matière d'infrastructure, nous offrons énormément d'avantages pour les applications gourmandes en données.

**TT :** Avez-vous trouvé une application pour les techniques d'intelligence artificielle dans les ensembles de données ?

**BG :** Chez Pure, nous consommons des ensembles de données massifs. Nous cherchons toujours à aller plus loin avec ce que nous avons. Et le déploiement de FlashBlade a entraîné l'entreprise dans une nouvelle dimension.

L'intégration matérielle en est un bon exemple. Parmi les aspects novateurs de FlashBlade, il y a notamment le fait que nous créons et concevons conjointement tous les logiciels avec le matériel, jusqu'aux composants Flash NAND de premier niveau. Ainsi, nous pouvons surveiller l'état et les performances du moindre bit de mémoire flash, à chaque niveau de lecture et écriture. Même dans un système de stockage à l'échelle du pétaoctet, nous pouvons collecter des données télémétriques qui nous montrent quelle forme prend l'usure ou nous indiquent les taux de correction des erreurs.

Il s'agit de nouveaux types de données, qui ne sont généralement pas accessibles, même dans un déploiement de type SSD. En concevant conjointement nos logiciels et notre matériel et en les intégrant en profondeur les uns aux autres, nous avons désormais accès à de toutes nouvelles formes de données, que nous pouvons utiliser pour optimiser certains algorithmes internes et construire des modèles prédictifs. L'exploration ne fait que commencer, mais les résultats initiaux sont très positifs : ils tendent à montrer que l'on peut faire plus avec nos propres produits en nous basant sur les données télémétriques internes que nous recevons.

**TT :** Avant de terminer, est-ce qu'il y a un sujet que nous n'avons pas abordé concernant Spark et FlashBlade et qui vous semble important ?

**BG :** Pour nous, utiliser l'interface de système de fichiers distribué la plus déployée au monde constitue un avantage de taille. Les protocoles de fichiers plus restreints, tels que le système HDFS, sont confrontés à différentes problématiques, notamment le fait qu'ils deviennent des silos de données.  C'est-à-dire que vous avez un ensemble de données, d'une échelle d'un pétaoctet par exemple, qui se trouve dans un système HDFS. Si j'ai Hadoop ou Spark ou d'autres applications dans cet écosystème, je peux accéder aux données et les traiter à partir de là. Mais si je veux y accéder en passant par mon système de fichier normal, l'opération est possible mais cela ajoute des couches de complexité.

Parmi les éléments qui font de Spark un outil idéal, il y a notamment le fait que le protocole de stockage a vraiment été découplé de l'aspect « calcul ». Cela nous permet de nous intégrer facilement, même dans des systèmes de fichiers distribués de stockage très simples et très largement déployés, tels que NFS. Nous sommes en mesure de fournir de 10 à 15 Go de débit de lecture par seconde en utilisant NFS dans un cluster Spark. Et en même temps, nous pouvons passer par une toute autre application et accéder au même ensemble de données en utilisant une méthode bien plus traditionnelle, en accédant simplement au système NFS via un environnement Linux, Unix, Mac OS ou Windows.

De manière générale, il y a un manque de consolidation dans le domaine. Les architectures traditionnelles n'ont jamais véritablement autorisé la consolidation de charges de travail de l'ordre du pétaoctet. En matière de consolidation, FlashBlade ouvre une nouvelle ère en offrant une densité, des performances et une simplicité de fonctionnement sans précédent.

**Pour en savoir plus, rendez-vous sur purestorage.com/analytics.**

© 2017 Pure Storage, Inc. Tous droits réservés. Pure Storage et le logo P sont des marques commerciales de Pure Storage, Inc.
Toutes les autres marques sont la propriété de leurs détenteurs respectifs.
Apache®, Hadoop® et Spark™ sont des marques commerciales ou déposées de l'Apache Software Foundation aux États-Unis et/ou dans d'autres pays.
L'utilisation de ces marques n'implique aucune approbation de la part de l'Apache Software Foundation.
« Python » est une marque déposée de la Python Software Foundation.