



Guide gratuit:
Mettre en oeuvre une
stratégie NoSQL

CONTENTS

Résumé.....	3
Introduction	3
Pourquoi les entreprises souhaitent adopter une solution NoSQL ?	3
Le besoin de rapidité	4
Le besoin d'évolutivité	4
Le besoin de disponibilité continue	5
Le besoin d'absence de contrainte géographique.....	5
Le besoin de gérer tous les types de données	6
Le besoin de réduire les coûts	6
Quels sont les freins à l'adoption du NoSQL dans l'entreprise ?	7
Freins techniques	7
Freins non techniques	8
La technologie NoSQL est-elle faite pour vous ?	9
Applications en ligne :	9
Applications d'analyse :	10
Applications de recherche pour l'entreprise :	10
Choisir une base de données NoSQL	11
Principaux critères de sélection	11
Check-list de sélection pour l'entreprise	11
Aspects techniques	11
Exigences pour l'entreprise	12
Problématiques cloud des bases de données NoSQL	13
Opérer la transition NoSQL	14
Migrer les données vers des bases de données NoSQL	14
Conclusion	15
À propos de DataStax	15



RÉSUMÉ

Au cours des prochaines années, le marché des bases de données NoSQL est voué à s'étendre trois fois plus rapidement que le marché SQL ; on comprend donc aisément que ceux qui découvrent la technologie Big Data soient impatients de comprendre pourquoi et comment elle s'insère dans leur entreprise. Les exigences en matière de rapidité, d'évolutivité, de disponibilité continue, d'absence de contrainte géographique, de capacité à gérer tous les types de données et la réduction des coûts favorisent cette adoption de plus en plus répandue. Il existe certes des obstacles, techniques ou non, à cette adoption, qu'il est important d'examiner ainsi que les améliorations qui ont été apportées à l'écosystème NoSQL. Par ailleurs, les cas d'usage jouent de plus en plus un rôle important d'indicateurs permettant à l'entreprise de savoir quand et comment utiliser les technologies NoSQL. Ce document examine ces différents aspects et propose des conseils stratégiques de mise en œuvre pratique, notamment une check-list de sélection et des astuces de migration.

INTRODUCTION

Un article publié en 2013 par Silicon Angle décrit la croissance actuelle et prévue de la technologie NoSQL : « Selon l'analyse de David Floyer pour Wikibon (et parue au Wall Street Journal), le marché des bases de données NoSQL devrait connaître un taux de croissance annuel composé de près de 60 % entre 2011 et 2017. Ce taux ne s'élèvera qu'à 26 % pour la part SQL du marché Big Data, sur la même période. »¹

Aujourd'hui, les grandes entreprises intègrent des solutions NoSQL aux applications métier qui sont au cœur des principaux aspects de leur activité. Les responsables informatiques qui n'utilisent pas encore cette technologie sont nombreux à vouloir savoir pourquoi et comment ils peuvent mettre en place cette technologie.

Ce document présente une stratégie générale de mise en œuvre du NoSQL dans l'entreprise : il explore les principales raisons pour lesquelles les entreprises se tournent vers des solutions NoSQL et fournit des exemples d'utilisation actuelle de cette technologie. Il présente également les freins auxquels les entreprises ont pu se heurter lors de leur mise en œuvre, fournit des conseils pour choisir une base de données NoSQL et propose des méthodes pratiques de mise en œuvre de solutions NoSQL dans différents scénarios d'applications.

POURQUOI LES ENTREPRISES SOUHAITENT ADOPTER UNE SOLUTION NOSQL ?

Pour savoir comment mettre en œuvre une stratégie NoSQL dans votre entreprise, vous pouvez commencer par examiner les principales raisons ayant motivé des entreprises modernes et performantes à choisir une solution NoSQL et définir si votre entreprise répond à des motivations similaires. Les besoins suivants, sans être pour autant exhaustifs, correspondent aux principales motivations des entreprises ayant choisi de mettre en œuvre des bases de données NoSQL dans leurs applications stratégiques métier.

¹ « Oracle is in Big Trouble: Big Data is to Blame », de Maria Deutscher, Silicon Angle, mars 2013 : <http://goo.gl/ODA7g>.

LE BESOIN DE RAPIDITÉ

Personne ne remet en cause le fait qu'améliorer les délais de réponse des systèmes externes peut avoir un impact direct sur la satisfaction des clients et le chiffre d'affaires. Amazon, par exemple, a constaté que chaque réduction de 100 ms du délai de réponse du site se traduit par un accroissement du chiffre d'affaires de 1 %. Yahoo indique avoir constaté une augmentation de 9 % du trafic sur le site chaque fois que la performance augmente de 400 ms.²

Les experts en bases de données ont toujours rêvé d'un paramètre universel 'fast=true' et d'une base de données qui soit ultra rapide en toutes circonstances. Même si ce paramètre de configuration n'existe pas, une chose est sûre : les délais de réponse des bases de données n'ont jamais été aussi décisifs qu'aujourd'hui, dans un contexte où un simple clic peut mener chez le concurrent.

Beaucoup moins soumises aux contraintes propres aux bases de données relationnelles (SGBDR), les bases de données NoSQL peuvent assurer des performances accélérées dans de nombreux cas d'usage. L'un des avantages caractéristiques des solutions NoSQL comme Apache Cassandra est la possibilité d'écrire les données bien plus rapidement qu'un SGBDR et d'assurer un traitement des requêtes tout aussi rapide pour des volumes importants de données.

C'est précisément pour cette raison qu'eBay utilise DataStax Enterprise, développé autour d'Apache Cassandra. Argumentant que les moteurs relationnels existants étaient trop lents pour les éléments clés de ses applications externes, eBay a remplacé ses bases de données classiques par DataStax Enterprise et respecte aujourd'hui les délais de réponse de son SLA avec, par jour, 6 milliards d'opérations en écriture, 5 milliards d'opérations en lecture, et 250 To de données dans Cassandra.³

Pour augmenter encore les performances, il est possible d'utiliser des bases de données avec traitement in-memory. Tous les cas d'usage ne sont pas possibles avec ce type de base de données, mais lorsque la situation s'y prête, on peut souvent réduire les délais de réponse.

Par conséquent, DataStax Enterprise inclut une option de base de données in-memory qui permet de cibler facilement différents types de données. Les données froides peuvent être affectées à des disques en rotation, les plus chaudes à des SSD et celles nécessitant des temps de réponse E/S les plus courts possibles iront aux objets in-memory, le tout dans le même cluster de base de données.

LE BESOIN D'ÉVOLUTIVITÉ

Évolutivité et performance vont de pair. Les entreprises doivent aujourd'hui garantir des temps de réponse réduits tout en accueillant un nombre croissant d'utilisateurs et un volume croissant de données dans leurs applications métier. Les architectures scale-up classiques se sont avérées inefficaces pour pérenniser l'évolutivité de tels systèmes.

Toutefois, les bases de données NoSQL comme Apache Cassandra fournissent une architecture scale-out de type « diviser pour résoudre » qui a prouvé son efficacité en termes d'évolutivité linéaire⁴, avec l'ajout de nouveaux nœuds possible en ligne et sans interruption de l'activité.

2 <http://www.strangeloopnetworks.com/assets/images/infographic2.jpg>

3 « eBay Leveraging Cassandra to Support Growing Multistructured Data Volumes », de Jeff Kelly, 9 mai 2013, Wikibon : <http://goo.gl/LHkKB>.

4 « Benchmarking Cassandra Scalability on AWS », de Adrian Cockcroft, novembre 2011 : <http://goo.gl/G8NUa>.

La société Ooyala, éditeur de contenus vidéo et de solutions d'analyse de données pour des sociétés comme ESPN et Rolling Stone, illustre cette approche. Ooyala surveille et analyse un quart de toutes les vidéos visionnées en ligne chaque jour sur le Web, ce qui se traduit par des milliards d'événements diffusés via ses clusters de bases de données scale-out DataStax Enterprise/Cassandra. Ooyala explique que son ancienne base de données scale-up ne permettait pas d'atteindre l'évolutivité et la performance requises ; elle a donc préféré choisir Cassandra.

LE BESOIN DE DISPONIBILITÉ CONTINUE

Si des performances insuffisantes peuvent entraîner une perte de clients, rien n'est toutefois pire qu'une période d'indisponibilité. Expert du secteur informatique, le groupe Gartner indique que les temps d'indisponibilité représentent pour les entreprises un risque supérieur à celui lié aux fuites de données.⁵ Les pertes de chiffre d'affaires sont estimées à 6,5 millions par heure pour les établissements financiers, et en moyenne tous secteurs confondus à 5 600 USD/minute soit plus de 300 000 USD/h.⁶

Il existe une différence entre une approche haute disponibilité avec basculement qu'offrent les SGBDR avec leurs architectures maître-esclave et la disponibilité continue assurée par les bases de données NoSQL comme Apache Cassandra. En raison de sa conception scale-out totalement distribuée et de la prise en charge de plusieurs datacenters et zones de disponibilité cloud, Cassandra garantit l'absence de période d'indisponibilité via des copies redondantes des données et fonctions réparties sur un cluster et plusieurs sites.

Netflix, désignée plus grande application cloud au monde⁷, utilise Cassandra pour assurer l'absence de période d'indisponibilité à ses clients ; 95 % de ses données sont stockées dans Cassandra. Lorsqu' Amazon a subi une panne importante en 2012, Netflix n'a jamais déploré aucune fausse note : « Nous n'avons pas eu besoin de faire quoi que ce soit. Cassandra a acheminé les requêtes autour de la zone non disponible et lorsque tout est revenu à la normale, l'anneau était réparé. »⁸

LE BESOIN D'ABSENCE DE CONTRAINTE GÉOGRAPHIQUE

Presque toutes les entreprises performantes bénéficient d'une présence mondiale ; il est donc fondamental que leurs données soient disponibles rapidement sur de nombreux sites. Fondés sur une conception maître-esclave, les anciens SGBDR ont du mal à assurer des lectures rapides sur plusieurs sites. Ils ne peuvent pas non plus offrir la simple fonctionnalité d'écriture « write-anywhere », si importante pour de nombreuses entreprises.

Les bases de données NoSQL comme Cassandra peuvent aisément répartir les données sur plusieurs datacenters et zones de disponibilité cloud. Par ailleurs, en raison de son architecture pair-à-pair, Cassandra permet des opérations de type « read/write-anywhere », ce qui assure une réelle indépendance pour les données en termes de contrainte géographique.

Les sociétés comme Adobe apprécient beaucoup cette fonctionnalité offerte par Cassandra. Pour son application marketing cloud, Adobe exécute son cluster de base de données DataStax Enterprise/Cassandra entre deux datacenters pour garantir à ses clients une lecture/écriture rapide, où qu'ils se trouvent.

5 « Gartner's state of cloud security: Outages are bigger risk than breaches », de Brandon Butler, 14 novembre 2012 : <http://goo.gl/mpnm3>.

6 « Confronting System Downtime », Evolgen : <http://goo.gl/jWQTy>.

7 « The biggest cloud app of all: Netflix », de Steven J. Vaughan-Nichols, ZDnet, avril 2013 : <http://goo.gl/KQRKr>.

8 « Post-mortem of October 22, 2012 AWS degradation », Netflix Tech Blog, octobre 2012 : <http://goo.gl/X20sp>.

LE BESOIN DE GÉRER TOUS LES TYPES DE DONNÉES

Aujourd'hui, la diversité des types de données constitue un défi pour les bases de données relationnelles classiques et l'une des principales raisons pour lesquelles les entreprises se tournent vers des solutions NoSQL. Les bases de données NoSQL comme Cassandra proposent un modèle de données bien plus flexible qui convient sans problème aux données structurées, semi-structurées et non structurées, tout en conservant performance et efficacité en termes de stockage.

La NASA utilise Cassandra pour ses applications de sécurité qui gèrent le suivi de tous les patches matériels et logiciels dans le monde entier. Il s'agit de données structurées et non structurées. La NASA a conclu que le modèle de données flexible de Cassandra lui avait permis d'insérer des données bien plus naturellement qu'avec leur précédent SGBDR. Par ailleurs, les temps de réponses ont également été réduits pour la récupération des données.

LE BESOIN DE RÉDUIRE LES COÛTS

L'adoption du NoSQL dans l'entreprise s'explique également par un dernier facteur, et non le moindre : la réduction des coûts. Le coût reste encore très souvent un élément rédhibitoire pour les SGBDR, en particulier pour certains nouveaux appareils mainframe de type scale-up. Les solutions NoSQL comme DataStax Enterprise coûtent habituellement 70 à 80 % moins cher que les anciens systèmes relationnels et sont conçus pour fonctionner sur des équipements standard rentables.

Constant Contact, une société qui répond aux besoins marketing de nombreuses petites entreprises, a réalisé d'énormes économies en se tournant vers DataStax, il y a deux ans. Ayant besoin de redimensionner ses systèmes mais ne disposant pas du budget nécessaire pour l'offre onéreuse de son précédent SGBDR, la société a choisi une solution scale-out et DataStax Enterprise : elle a ainsi pu mettre en œuvre son nouveau système dans un délai 3 fois plus court qu'avec son ancien SGBDR et pour un coût 90 % inférieur à celui de sa base de données précédente.

En outre, même si Constant Contact utilise des centaines de nœuds DataStax Enterprise, aucun administrateur dédié n'est nécessaire pour gérer les clusters de la base de données. Il suffit d'une équipe en temps partiel chaque semaine pour s'occuper de la base de données en parallèle des autres systèmes.

Les entreprises ont certainement d'autres motivations pour mettre en œuvre une solution NoSQL dans leur infrastructure informatique, mais les facteurs décrits dans ce document sont les plus fréquemment cités par les clients DataStax. Un autre problème se pose ensuite lorsque vous définissez une stratégie NoSQL : les obstacles que vous risquez de rencontrer lors de cette transition.

QUELS SONT LES FREINS À L'ADOPTION DU NOSQL DANS L'ENTREPRISE ?

Même si les solutions NoSQL sont extrêmement prometteuses pour de nombreux cas d'usage, les entreprises satisfaites de leur solution NoSQL admettent que certains aspects doivent être bien examinés avant de déployer cette technologie. Ces aspects appartiennent à deux catégories : les freins techniques et les freins non techniques.

Freins techniques

Voici certaines des principales contraintes techniques auxquelles les sociétés performantes ont été confrontées avec leur solution NoSQL :

- ☑ **Différents modèles de données :** On ne pourra jamais assez insister sur le fait que le principal problème technique des entreprises est la transition mentale du modèle de données relationnelles au modèle de données NoSQL. La réussite des projets tient au fait que l'équipe informatique a correctement ou non modélisé les données pour la base de données NoSQL afin d'optimiser ses capacités. Il est donc essentiel que les experts en bases de données soient formés et se familiarisent bien avec le nouveau modèle de données NoSQL dans la base de données choisie.
- ☑ **Manque de sécurité :** En 2012, InformationWeek a publié un article spécial intitulé « NoSQL Equals No Security » (NoSQL ou l'absence de sécurité).⁹ Dans cet article, l'auteur déplore que le manque de fonctionnalités de sécurité dans les bases de données NoSQL risque de remettre en cause leur utilisation dans des environnements nécessitant des règles strictes de protection des données. Certaines bases de données NoSQL ne possèdent aucune mesure de sécurité, mais DataStax Enterprise offre des fonctionnalités de ce type, répondant à la grande majorité des exigences de sécurité de l'entreprise.¹⁰
- ☑ **Prise en charge des transactions ACID :** Le fait que la plupart des bases de données NoSQL ne prennent pas en charge les transactions de niveau ACID pose problème à certains experts informatiques. Il est vrai que si votre application cible requiert des transactions imbriquées complexes nécessitant des restaurations et des points de sauvegarde, une base de données NoSQL ne conviendra peut-être pas. Il faut toutefois noter qu'une base de données NoSQL telle que Cassandra offre atomicité, durabilité et isolation (AID), la cohérence (C) étant variable : éventuelle ou stricte, en fonction des besoins spécifiques de l'application ou de l'opération, avec possibilité de transactions pour les lots.

9 « Why NoSQL Equals No Security », de Michael Davis, InformationWeek, mars 2012 : <http://goo.gl/4E5Ac>. 7

10 Voir : « What's New in DataStax Enterprise 3.0? », <http://goo.gl/GW1vb>.

Freins non techniques

Les principaux problèmes non techniques auxquels les entreprises sont confrontées lors d'une mise en œuvre NoSQL sont les suivants :

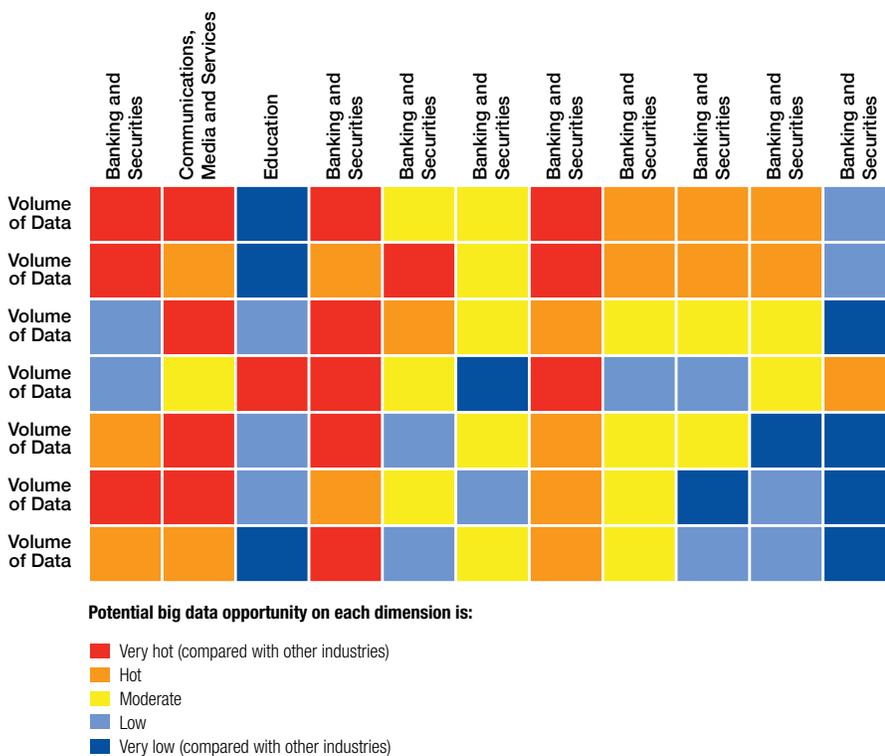
- ✓ **Trouver du personnel expérimenté** : Heureusement, ce problème devient moins prégnant puisque les entreprises sont aujourd'hui plus nombreuses à déployer des solutions NoSQL et que les éditeurs proposent de nombreux cours de formation. Pour les employés qui opèrent la transition vers le NoSQL, la courbe d'apprentissage sera différente en fonction de la technologie choisie. Certaines bases de données NoSQL, comme Cassandra, facilitent la courbe d'apprentissage car l'interface linguistique principale est quasi identique au SQL.
- ✓ **Guerre technique « de religion »** : Dans de nombreuses entreprises, les départements informatiques se livrent une sorte de guerre de religion autour de la technologie. Des alliances se forment pour certaines technologies et toute tentative d'apporter une nouvelle pierre à l'édifice rencontre résistance. Les défenseurs des anciennes technologies pensent souvent qu'elles sont encore capables de gérer n'importe quelle tâche. Face à cette situation, il convient d'appliquer une méthodologie orientée données incluant une analyse à la fois technique et non technique, qui sera examinée en toute objectivité par des experts informatiques.
- ✓ **Questions liées à la viabilité des éditeurs** : Les responsables informatiques qui s'engagent à long terme pour une technologie veulent être certains que les fournisseurs de cette technologie les suivront dans la durée. Les petites startups dont les clients sont peu nombreux émettront naturellement des objections et des inquiétudes tandis que les sociétés en plein essor avec une clientèle bien assise, accompagnées d'une équipe de support bien rodée, offriront la tranquillité d'esprit recherchée.

LA TECHNOLOGIE NOSQL EST-ELLE FAITE POUR VOUS ?

Nous savons maintenant pourquoi les entreprises adoptent de plus en plus la technologie NoSQL et quelles sont les problématiques qu'elles doivent résoudre. Il reste toutefois une question importante : Comment déterminer si la technologie NoSQL vous convient ? L'idéal est de commencer par analyser ce que peut offrir le NoSQL à votre entreprise, par branche, puis par cas d'usage.

Selon une perspective globale, le groupe Gartner a établi dans le tableau suivant une vision intéressante des avantages que peut procurer la technologie NoSQL à divers secteurs d'activité, selon les caractéristiques clés des Big Data (vitesse, variété, volume, complexité, etc.) :

Figure 1 – Carte thermique sectorielle du Big Data, groupe Gartner



Source: Gartner (July 2012)

Lors de l'évaluation des cas d'usage applicatifs possibles avec la technologie NoSQL, il convient de considérer en priorité les grandes catégories d'applications, à savoir en ligne, analyse et recherche (une même application pouvant bien entendu appartenir à une combinaison des trois catégories).

La liste suivante, non exhaustive, présente les cas d'usage actuellement compatibles avec la technologie NoSQL, notamment Apache Cassandra, Hadoop et Solr :

Applications en ligne :

- ✓ flux de séries chronologiques (données financières ou autres données chronologiques)
- ✓ sillage de données/capteurs/appareils

- ✓ applications transactionnelles distribuées
- ✓ diffusion de contenu multimédia
- ✓ vente en ligne (par ex. transactions, paniers d'achat)
- ✓ analyse de données en temps réel
- ✓ capture et analyse de données des médias sociaux
- ✓ analyse des flux de clics sur le Web
- ✓ systèmes transactionnels avec nombreuses opérations d'écriture

Applications d'analyse :

- ✓ analyse du comportement des acheteurs
- ✓ analyse de conformité/respect réglementaire
- ✓ recommandations clients
- ✓ détection des fraudes
- ✓ analyse des risques
- ✓ analyse des campagnes de programme de vente
- ✓ analyse de la chaîne logistique
- ✓ analyse des flux de clics sur le Web en batch

Applications de recherche pour l'entreprise :

- ✓ recherche générale sur le web
- ✓ recherche à facettes pour la vente sur le Web (catégories)
- ✓ recherche/priorisation et mise en avant des résultats
- ✓ recherche et analyse de journaux d'applications
- ✓ recherche et analyse de documents (par ex. PDF, MS Word)
- ✓ recherche géospatiale
- ✓ recherche de biens immobiliers et de propriétés
- ✓ recherches dans les médias sociaux

CHOISIR UNE BASE DE DONNÉES NOSQL

Le site web nosql-database.org fournit la liste de plus de 150 bases de données NoSQL différentes. Comment procéder pour réduire cette liste à une sélection de candidats susceptibles de convenir aux cas d'usage de vos applications ?

Principaux critères de sélection

Les différentes bases de données NoSQL se distinguent par de nombreuses fonctionnalités différentes. Voici quelques critères permettant d'affiner votre recherche pour des déploiements spécifiques :

- ✓ **Le modèle de données** : Commencez par déterminer le type de données à stocker et son format de début/fin. Les bases de données NoSQL diffèrent beaucoup selon le modèle de données utilisé (par ex. « wide row », document, graphe, etc.) ; une non compatibilité entre le modèle de données d'une solution NoSQL et l'application cible peut remettre en cause la réussite d'un projet.
- ✓ **Les attentes en termes de scalabilité des données** : Réfléchissez ensuite à la croissance prévue pour l'application et à la scalabilité nécessaire pour les données. Certaines bases de données NoSQL sont en mémoire principale et ne permettent pas une croissance horizontale sur plusieurs machines, tandis que d'autres comme Cassandra offrent une scalabilité linéaire sur de nombreuses machines.
- ✓ **Le modèle de distribution des données** : Il convient de déterminer l'échelle de distribution des données, si le modèle inclut plusieurs régions géographiques, s'il est utilisé à des fins de récupération après sinistre, ou autres. Il faut aussi définir si les opérations aussi bien de lecture que d'écriture doivent être possibles sur des sites distribués. Certaines bases de données NoSQL utilisent des architectures maître-esclave (même si elles sont appelées « primaires/secondaires ») qui n'autorisent que les opérations de lecture tandis que les architectures pair-à-pair autorisent à la fois lecture et écriture.

CHECK-LIST DE SÉLECTION POUR L'ENTREPRISE

Vous trouverez ci-après une check-list plus détaillée pour les bases de données NoSQL. Elle permet de s'interroger sur les aspects techniques et sur l'activité elle-même de l'entreprise afin de choisir la base de données NoSQL qui convient le mieux à votre cas d'usage.

Aspects techniques

- ✓ La base de données NoSQL peut-elle servir de source de données principale pour l'application en ligne concernée ?
- ✓ La base de données NoSQL peut-elle servir de source de données d'analyse et/ou être facilement liée à des processus Hadoop ?
- ✓ La base de données NoSQL peut-elle gérer des logiciels de recherche pour l'entreprise ou s'y intégrer de manière transparente ?
- ✓ La base de données NoSQL permet-elle d'isoler les processus en ligne, d'analyse et de recherche au sein d'une même application ?

- ✓ Quel est le niveau de sécurité de la base de données NoSQL, par exemple, en cas de perte de données stratégiques ? Les opérations d'écriture sont-elles suffisamment durables pour que les données soient protégées ?
- ✓ La base de données NoSQL fournit-elle un ensemble solide de fonctionnalités de sécurité ?
- ✓ La base de données NoSQL est-elle tolérante aux pannes (autrement dit, ne présente pas de point unique de défaillance) ? Assure-elle une disponibilité continue ?
- ✓ La base de données NoSQL peut-elle aisément répliquer les données entre des datacenters identiques et multiples, ainsi que des zones de disponibilités cloud différentes ?
- ✓ La base de données NoSQL offre-t-elle des fonctionnalités de type « read/write-anywhere » ?
- ✓ La base de données NoSQL nécessite-t-elle ou non des couches spécifiques de mise en cache ?
- ✓ La base de données NoSQL offre-t-elle une option de calcul in-memory ?
- ✓ La base de données NoSQL est-elle capable de gérer les Big Data et d'assurer des performances élevées quel que soit le volume de données ?
- ✓ La base de données NoSQL offre-t-elle une scalabilité linéaire par exemple pour l'ajout de nouveaux nœuds ?
- ✓ Est-il possible d'ajouter de nouveaux nœuds ou de les supprimer en ligne (autrement dit sans impact sur les processus de l'entreprise) ?
- ✓ La base de données NoSQL gère-t-elle les principales plateformes et les principaux langages de développement ?
- ✓ La base de données NoSQL peut-elle fonctionner sur des équipements standard sans exigences particulières ?
- ✓ La mise en œuvre et la maintenance de la base de données NoSQL sont-elles facilitées pour les grands déploiements ?

Exigences pour l'entreprise

- ✓ La solution NoSQL est-elle proposée par une entité commerciale ?
- ✓ L'entité commerciale fournit-elle un support et des services 24h/24 et 7j/7 ?
- ✓ La solution NoSQL prévoit-elle une documentation professionnelle en ligne ?
- ✓ La solution NoSQL possède-t-elle des clients de référence dans de nombreux secteurs ?
- ✓ La base de données NoSQL possède-t-elle une structure tarifaire intéressante ?
- ✓ En open source, la base de données NoSQL possède-t-elle une communauté open source active ?

PROBLÉMATIQUES CLOUD DES BASES DE DONNÉES NOSQL

La quantité d'informations qui réside actuellement uniquement dans le cloud est faible, mais cela va bientôt changer. Une récente étude menée par le groupe d'analyse IDC, spécialiste du marché des technologies de l'information, estime que le cloud computing représente moins de 2 pourcent des dépenses informatiques actuelles, mais que d'ici 2015, près de 20 pourcent de toutes les informations seront « touchées » (stockées ou traitées) dans un cloud.¹¹ IDC prédit par ailleurs que la même année, jusqu'à 10 pourcent de toutes les données seront conservées dans un cloud.¹²

Le cloud offre de nombreuses promesses : élasticité transparente et scalabilité, disponibilité plus élevée, distribution des données simplifiée, gérabilité simplifiée, etc. Toutefois, si de nombreux éditeurs indiquent que leur base de données est « compatible cloud », cela signifie souvent simplement que vous pouvez facilement en installer et exécuter une instance auprès d'un fournisseur cloud. La question la plus importante à se poser est la suivante : la base de données offre-t-elle tous les avantages, ou la plupart d'entre eux, d'une exécution dans le cloud ?

Qu'il s'agisse d'un ancien SGBDR ou d'une base de données NoSQL, les questions à se poser sont les mêmes pour pouvoir véritablement tirer parti des avantages du cloud computing. Nous vous recommandons de vérifier les points suivants auprès de tout fournisseur de base de données NoSQL que vous envisagez de choisir pour le cloud :

- ☑ La base de données fournit-elle une élasticité transparente, avec possibilité d'expansion ou de réduction, sans période d'indisponibilité ?
- ☑ Est-il possible d'étendre les capacités de manière horizontale dans le cloud, et si oui, quels en seront les avantages ?
- ☑ La base de données peut-elle facilement utiliser plusieurs zones de disponibilité du fournisseur cloud de manière à assurer une disponibilité continue en cas de défaillance d'une ou plusieurs zones ?
- ☑ La base de données offre-t-elle des fonctions de sécurité qui protègent les données dans le cloud ?
- ☑ L'éditeur NoSQL fournit-il des outils permettant de gérer et de superviser la base de données auprès du fournisseur cloud ?
- ☑ Quelle est la différence de coût sur une période de 3 à 5 ans selon qu'on utilise la base de données NoSQL localement ou dans le cloud ?

11 « Extracting Value from Chaos », de John Gantz et David Reinsel, IDC, juin 2011, <http://idcdocserv.com/1142>.

12 Ibid

OPÉRER LA TRANSITION NOSQL

Comment, concrètement, opérer la transition NoSQL et mettre en œuvre votre première application ? En général, il existe trois méthodes pour mettre en œuvre une base de données NoSQL :

1. **Nouvelles applications** : De nombreuses entreprises qui se tournent vers le NoSQL choisissent une nouvelle application et partent de zéro. Cette approche permet d'éviter les problèmes de réécriture d'application, de migration de données, etc.
2. **Augmentation** : Certaines choisissent d'étendre un système existant en y ajoutant un composant NoSQL. Cela est fréquent avec les applications devenues trop encombrantes pour un SGBDR suite à des exigences d'évolutivité, de plus grande disponibilité ou autres problématiques. Certaines parties du système continuent d'utiliser le SGBDR existant tandis que les autres composants de l'application sont modifiés pour utiliser la base de données NoSQL.
3. **Remplacement intégral** : Pour les systèmes qui s'avèrent tout simplement trop onéreux pour pouvoir être conservés avec un SGBDR, ou qui montrent de nombreux signes de faiblesse, par exemple en termes de nombre d'utilisateurs simultanés, de vitesse ou de volume des données, il convient de remplacer intégralement le système par une base de données NoSQL.

MIGRER LES DONNÉES VERS DES BASES DE DONNÉES NOSQL

Qu'il s'agisse du scénario augmentation ou remplacement, il est nécessaire de migrer les données depuis le SGBDR existant vers la nouvelle base de données NoSQL. Les modalités de migration des données existantes dépendent de leur quantité.

- ☑ **Fichiers plats** : Chaque SGBDR permet d'exporter les données à partir des tables vers des fichiers plats avec délimitation. Les bases de données NoSQL comme Cassandra permettent de charger ce type de fichiers directement dans les familles de colonnes/tables.
- ☑ **Sqoop** : Il s'agit d'un utilitaire employé dans Hadoop pour déplacer les données depuis les bases de données existantes vers Hadoop. Cassandra prend également en charge l'outil Sqoop ; un développeur peut donc se connecter à un SGBDR existant et à Cassandra et envoyer les données directement dans la base de données.
- ☑ **Outils ETL** : Si la migration de données doit être techniquement plus poussée, il est possible d'utiliser des solutions ETL (extract-transform-load). De nombreux outils fournis par Jaspersoft, Pentaho et Talend assurent d'excellentes routines de transformation qui permettent de manipuler les données source exactement comme vous le souhaitez puis de les charger dans une cible NoSQL. Ils offrent également de nombreuses autres fonctionnalités comme des interfaces visuelles par pointer-cliquer, des moteurs de planification, etc. Nombre de ces outils peuvent être téléchargés et utilisés gratuitement.

CONCLUSION

Pour mettre en œuvre une stratégie NoSQL, vous devez bien comprendre les raisons pour lesquelles les entreprises performantes utilisent cette technologie, puis déterminer si ces éléments clés ou d'autres éléments caractéristiques de cette technologie peuvent présenter un avantage pour votre activité. Une fois certain que la technologie NoSQL vous convient, vous devez alors examiner en détail les pièges à éviter, étudier les critères qui vous permettront de choisir la base de données NoSQL appropriée aux cas d'usage de vos applications et définir la stratégie à utiliser pour mettre en place cette technologie.

DataStax fournit aux entreprises les logiciels, services et stratégies permettant de garantir leur réussite avec la technologie NoSQL. Avec sa solution fiable et sécurisée DataStax Enterprise, développée autour d'Apache Cassandra, ainsi qu'un support sans interruption, des conseils et des services de formation, les experts DataStax vous assurent une expérience de transition NoSQL positive et satisfaisante.

Pour en savoir plus sur Apache Cassandra, DataStax et pour télécharger des logiciels Cassandra ainsi que DataStax Enterprise, rendez-vous sur www.datastax.com ou envoyez un e-mail à info@datastax.com. L'utilisation de DataStax Enterprise Edition est entièrement gratuite dans des environnements de développement, tandis que les déploiements en production nécessitent une souscription logicielle.

À PROPOS DE DATASTAX

DataStax propose une plateforme NoSQL d'entreprise hautement évolutive permettant d'exécuter les applications métier stratégiques de nombreuses sociétés parmi les plus innovantes et les plus consommatrices de données au monde. Reposant sur la base de données open source Apache Cassandra™, DataStax fournit une plateforme totalement distribuée et continuellement disponible, dont le déploiement est plus rapide et la maintenance moins onéreuse que d'autres plateformes de bases de données.

DataStax compte plus de 400 clients dans 38 pays, notamment des entreprises leader comme Netflix, Rackspace, Pearson Education et Constant Contact, et étend ses activités sur des secteurs verticaux tels que services web, services financiers, télécommunications, logistique et gouvernement. DataStax, dont le siège social est établi à San Mateo en Californie, est soutenue par les plus grands investisseurs du secteur notamment Lightspeed Venture Partners, Meritech Capital et Crosslink Capital.

Pour plus d'informations, visitez le site www.datastax.com.