

DÉPASSER LE MODÈLE RELATIONNEL

LIVRE BLANC MARKLOGIC • SEPTEMBRE 2015

Les entreprises font face à une incapacité croissante à gérer les volumes massifs de données diverses, variées et changeantes avec les bases de données relationnelles utilisées ces trois dernières décennies. C'est pour cette raison que les sociétés leaders dépassent le modèle relationnel afin d'adopter une base de données nouvelle génération : MarkLogic.



ÉVALUATION DE L'ÉTAT ACTUEL

Votre entreprise a-t-elle besoin de changer, ou pensez-vous gérer parfaitement vos données ? Les réponses aux questions ci-dessous fournissent une évaluation de base permettant de répondre à ces questions. Plus vous obtenez de réponses « OUI », plus cela signifie que vos bases de données actuelles ne répondent pas aux besoins de votre entreprise.

		OUI	NON
QUESTIONS MÉTIER	1. Votre entreprise possède-t-elle des données essentielles qui ne se trouvent pas dans une base de données ?	✓	✗
	2. Y a-t-il plusieurs bases de données possédant essentiellement les mêmes données ?	✓	✗
	3. Existe-t-il plusieurs sources de données qui ne sont pas gérées de façon centrale par le service informatique ?	✓	✗
	4. Y a-t-il de gros projets informatiques qui ont dépassé le budget ou qui n'ont pas été lancés ?	✓	✗
	5. Existe-t-il des schémas de base de données si complexes que seuls quelques experts peuvent répondre de manière adéquate à des questions les concernant ?	✓	✗
QUESTIONS TECHNIQUES	6. La modélisation des données ralentit-elle ou entrave-t-elle le processus de développement des applications ?	✓	✗
	7. Existe-t-il des tableaux relationnels dans lesquels les noms de colonne ont été changés ou ont reçu une nouvelle signification « juste pour les faire fonctionner » ?	✓	✗
	8. Y a-t-il des modifications de schémas de base de données régulières chaque mois, et certains de ces changements sont-ils infructueux ?	✓	✗
	9. Perdez-vous beaucoup de temps et de ressources à essayer de trouver le moyen d'effectuer les mises à l'échelle ?	✓	✗
	10. Y a-t-il jamais eu des problèmes de performance ou des erreurs qui ont peut-être été causés par des logiciels intermédiaires complexes ?	✓	✗

Contenu

Introduction.....	1
Le monde du Big Data aujourd'hui	2
Volume	
Vitesse	
Variété	
Véracité	
Variabilité	
Se noyer dans un océan de complexité	3
Silos de données et ETL	
« Informatique de l'ombre » et failles de sécurité	
Coûts élevés, échec des projets et incapacité à innover	
Pourquoi les bases de données relationnelles ne fonctionnent pas ?	4
Les bases de données relationnelles ne sont pas conçues pour gérer le changement	
Les bases de données relationnelles ne sont pas conçues pour les données hétérogènes	
Les bases de données relationnelles ne sont pas conçues pour l'évolutivité et l'élasticité	
Les bases de données relationnelles ne sont pas conçues pour les charges de travail variées	
Une base de données nouvelle génération dédiée aux données d'aujourd'hui.....	8
Présentation des fonctions qui démarquent MarkLogic	
Les entreprises leaders sont encore plus efficaces avec MarkLogic	
Dépasser le modèle relationnel avec vos données.....	12
Prochaines étapes recommandées	
Plus d'informations	

INTRODUCTION

Au cours des dernières décennies, la technologie a évolué rapidement et chaque aspect de l'activité commerciale a changé. Aujourd'hui, les données collectées sont plus nombreuses qu'auparavant et les entreprises débordent d'idées pour développer des applications plus importantes et plus intelligentes. On pourrait donc penser que les bases de données, à savoir les emplacements où les sociétés stockent leurs ressources les plus précieuses (les *données*), ont également changé. Mais dans une large mesure, ce n'est pas du tout le cas.

La technologie dominante de stockage et de gestion des données, la base de données relationnelle, n'a pas beaucoup évolué depuis sa création au cours de la guerre froide, il y a plus de trente ans. À cette époque, les données étaient considérées comme petites, nettes, structurées et statiques car c'était la seule façon de les stocker. Cependant, les données ne se définissent pas ainsi. Dans le monde d'aujourd'hui, les entreprises affrontent la réalité, c'est-à-dire des données importantes, rapides, variées et changeantes. Les sociétés ne gèrent plus une petite poignée de systèmes, mais des centaines de systèmes et des pétaoctets de données.

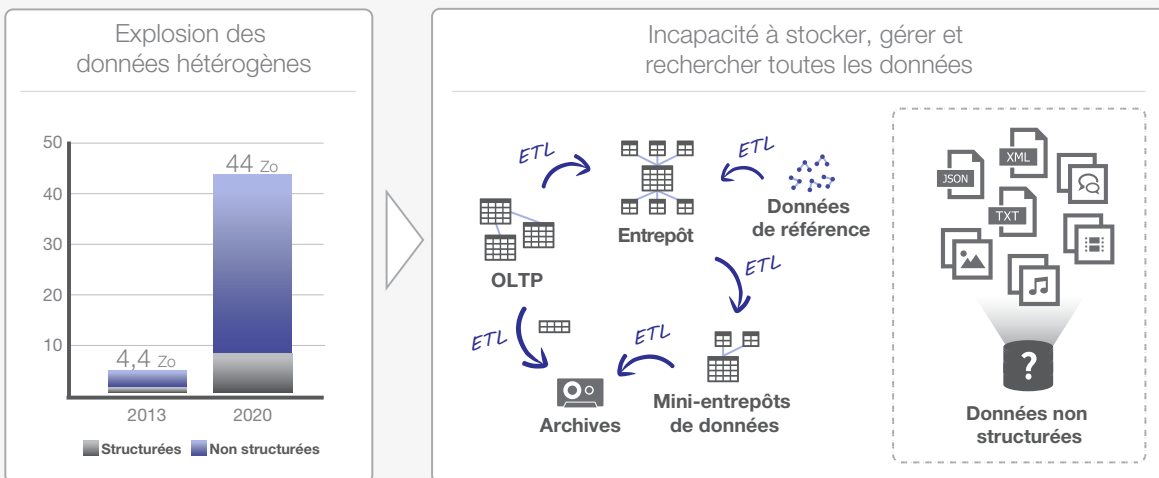
Le nouveau monde des « Big Data » crée une opportunité passionnante, mais celle-ci est trop souvent considérée comme un simple nouveau défi que les services informatiques doivent relever. Aujourd'hui, ces derniers

passent la majeure partie de leur temps à garder la tête hors de l'eau en gérant une toile complexe de silos de données et de processus ETL (extraction, transformation et chargement) fréquents destinés à déplacer les données. Tandis que les individus et les services cherchent leurs propres solutions, les « éléments informatiques de l'ombre » et les failles de sécurité commencent à apparaître. Au sein des entreprises, les coûts élevés et les longs délais de projet sont désormais la norme. Bloquées dans un état de cycles de maintenance constants, les sociétés ne peuvent pas se préoccuper de tirer le meilleur parti de toutes leurs données. Dans une certaine mesure, ces défis sont tous apparus suite à l'utilisation de bases de données relationnelles pour résoudre des problèmes pour lesquels elles n'avaient jamais été conçues.

Aujourd'hui, les entreprises ne peuvent pas s'appuyer uniquement sur le modèle relationnel universel. Motivées par le besoin de changer, elles adoptent de nouveaux types de bases de données. MarkLogic est à l'avant-garde de ce changement générationnel et fournit une base de données mieux adaptée à toutes les données actuelles. MarkLogic possède un modèle de données plus souple de stockage, de gestion et de recherche de gros volumes de données diverses, tout en conservant également toutes les fonctionnalités d'entreprise nécessaires aux sociétés. C'est cette combinaison unique qui a permis aux entreprises leaders de dépasser le modèle relationnel et d'exploiter encore plus de données qu'auparavant.

POURQUOI CHANGER ?

Les entreprises font face à une incapacité croissante à gérer leurs données diverses, variées et changeantes. C'est uniquement en adoptant de nouvelles approches que les sociétés pourront résoudre ce problème et réduire leurs risques, développer plus rapidement des applications plus intelligentes et tirer le meilleur de leurs données.



« Dans l'ensemble, les « V » des Big Data peuvent se résumer en une vérité : *les données actuelles sont importantes, rapides, complexes et changeantes.* »

LE MONDE DU BIG DATA AUJOURD'HUI

Limitées par les technologies de l'époque, les données avaient tendance à toutes se ressembler. Elles arrivaient lentement et de façon ordonnée dans les centres de données, et elles étaient bien organisées sous la forme de données tabulaires qui remplissaient des lignes et des colonnes liées dans des tableaux préconfigurés. Le rythme des changements était quasi nul, à la fois pour l'industrie et l'informatique, et cela convenait à tout le monde. Mais c'étaient les années 1980, et la donne a radicalement changé.

Le monde des Big Data dans lequel nous évoluons de nos jours est bien caractérisé par les trois « V » souvent évoqués, le volume, la vitesse et la variété. En outre, deux autres « V » sont de plus en plus importants, la véracité et la variabilité. Dans l'ensemble, tous ces « V » peuvent se résumer en une vérité : *les données actuelles sont importantes, rapides, complexes et changeantes.*

VOLUME

L'univers numérique augmente de 40 % par an et devrait croître de 4,4 zettaoctets en 2013 à 44 zettaoctets en 2020 (un zettaoctet équivaut à 1 billion de gigaoctets).¹ Les dossiers papier ne sont plus le système d'archivage, ce sont les bases de données... ce qui signifie qu'elles stockent *tout*. Les entreprises actuelles doivent désormais gérer plus de données aux formats encore plus variés sur un nombre de systèmes encore plus grand, et réaliser cette gestion avec efficacité, de façon sécurisée et à peu de frais. Le coût du stockage des données continue de diminuer, et les consommateurs et les organismes de régulation s'attendent à ce que les sociétés stockent vraiment tout.

VITESSE

Tout est plus rapide dans le monde d'aujourd'hui. Les données sont créées et se modifient plus rapidement. Et les questions posées aux données changent également plus rapidement, afin de répondre aux nouvelles exigences d'entreprise établies pour gérer les transformations rapides dans la dynamique de marché, la nouvelle direction, les services sur demande ou les acquisitions et les scissions. De nos jours, les décisions sont prises en quelques minutes, pas en jours, et les données qui accompagnent

ces décisions doivent être fournies dans le bon format, avec une latence réduite et une efficacité plus grande. Que ce soient des données relatives à des événements sportifs ou de détection de fraudes bancaires, le besoin d'obtenir les données en *temps réel* n'est plus un souhait, mais une exigence. Le délai de développement des applications est également plus rapide et se mesure en semaines, pas en années. Et ces applications doivent faire face à d'innombrables utilisateurs, qui supportent de moins en moins l'attente, sont moins fidèles et souhaitent de plus en plus bénéficier de produits personnalisés.

VARIÉTÉ

La variété est l'un des plus grands défis de tous les « V ». Les données d'aujourd'hui sont bien plus variées, ou hétérogènes, qu'elles ne l'étaient : quelque 20 % sont structurées (par exemple, les données tabulaires et transactionnelles) et 80 % non structurées (par exemple, les documents, les textes, les e-mails, les images et les vidéos).² Les nouvelles sources de données non structurées sont assurément problématiques. Dans une étude, selon 64 % des entreprises, la raison principale d'envisager une nouvelle approche des Big Data est la diversité, la nouveauté et le flot des sources de données qu'elles doivent désormais gérer.³ Mais la variété des données structurées est peut-être encore plus problématique, car les entreprises s'efforcent de gérer les innombrables formats, tailles et types des données structurées qui augmentent rapidement en volume et se transforment. Les nouvelles applications, les fusions et acquisitions, et la réutilisation des données font partie des raisons communes qui entraînent la variété des données structurées.

VÉRACITÉ

La véracité concerne l'authenticité des données ou l'intégrité des données. Les données sont des ressources très précieuses, et les entreprises mettent tout en œuvre pour s'assurer que leurs données sont précises et non corrompues de quelque manière que ce soit. C'est pour cette raison qu'il est de plus en plus important de tracer les données, ou leur cycle de vie, c'est-à-dire de connaître leur origine (leur provenance), leur historique récent (comment elles ont été modifiées et par qui), leur maintien (combien

1 IDC. *Digital Universe*. Avril 2014 <<http://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm>>

2 Khan et al. *Big Data: Survey, Technologies, Opportunities, and Challenges*. Scientific World Journal, 2014 <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4127205/#B53>>

3 New Vantage Partners. *Big Data Executive Survey: Themes & Trends*. 2012. <<http://newvantage.com/wp-content/uploads/2012/12/NVP-Big-Data-Survey-Themes-Trends.pdf>>

« D'après les études, les silos de données constituent l'obstacle numéro un à la réussite des projets Big Data. »

de temps elles doivent rester disponibles) et leur intérêt (les données qui fournissent la meilleure réponse). En plus, les sociétés doivent posséder de fortes politiques de gouvernance des données afin de préserver l'accès des utilisateurs aux données à un niveau granulaire. Ces problématiques deviennent toutes plus importantes pour des raisons d'audit et de conformité, et pour fournir des analyses plus avancées.

VARIABILITÉ

La variabilité concerne les variations de signification des données en fonction du contexte. Elle a été examinée par un analyste senior de Forrester, Brian Hopkins, qui l'a définie comme « la variabilité des significations dans la langue naturelle et la façon d'utiliser la technologie des Big Data pour les résoudre ». ⁴ Prenons le mot « sub » en anglais. Se réfère-t-il à un sous-marin ou à un sandwich Subway® ? Ce problème dépasse même la langue naturelle. Il existe également des différences dans la façon dont les utilisateurs et les modélisateurs de données décrivent les entités de base. C'est le cas par exemple de l'État de « Caroline du Nord » qui parfois apparaît sous la forme « Caroline N » ou plus simplement « CN ». Comment une base de données peut-elle savoir qu'on se réfère ici à la même chose, ou même comprendre la notion d'« État » ? Il est simple pour n'importe qui de saisir la signification à partir du contexte, mais les bases de données peinent davantage face à ces défis sémantiques. Les données de plus en plus nombreuses impliquent des variations de description de plus en plus grandes pour les personnes, les lieux et les choses, et le problème s'amplifie donc encore davantage.

SE NOYER DANS UN OCÉAN DE COMPLEXITÉ

Le monde des Big Data d'aujourd'hui doit être considéré comme une opportunité énorme, mais il est trop souvent et simplement considéré comme un défi supplémentaire. Les services informatiques actuels passent la plupart de leur temps à garder la tête hors de l'eau et, s'ils ne s'attaquent pas directement à la complexité, l'entreprise tout entière pourrait bien couler.

⁴ Hopkins, Brian. « Blogging From the IBM Big Data Symposium - Big Is More Than Just Big, » 2011 <http://blogs.forrester.com/brian_hopkins/11-05-13-blogging_from_the_ibm_big_data_symposium_big_is_more_than_just_big>

SILOS DE DONNÉES ET ETL

D'après les études, les silos de données constituent l'obstacle numéro un à la réussite des projets Big Data. ⁵ C'est une évidence en examinant la plupart des diagrammes complexes de l'architecture des entreprises. Ils révèlent d'anciens systèmes incompatibles enchevêtrés avec d'autres systèmes hérités, d'où une architecture complexe et fragile dans laquelle les données ne peuvent pas être partagées ni utilisées. Dans la majeure partie des entreprises, seuls quelques spécialistes peuvent encore identifier et comprendre le lien de l'architecture aux règles commerciales complexes de la société. Par conséquent, il n'est pas surprenant que, pour la plupart des initiatives de stratégie d'entreprise, la majeure partie du temps soit employée uniquement pour identifier et profiler les sources de données. ⁶

Les silos de données n'appartenaient pas intentionnellement au schéma, elles résultaient de solutions à court terme. La plupart des bases de données sont conçues uniquement pour prendre en charge une certaine application ou un certain type de données. Pour extraire les données de ces bases de données et les utiliser à d'autres fins dans une autre base de données, un processus ETL (extraction, transformation et chargement) doit intervenir pour s'assurer qu'elles correspondent au schéma de la nouvelle base de données cible. Les processus ETL interviennent souvent dans la plupart des entreprises et créent de nouveaux silos de données à chaque fois.

Avec la prolifération des silos, il est de plus en plus difficile de les gérer et de les relier. Les développeurs finissent par utiliser un code de maintenance pour « bricoler » et relier les différentes applications, et ainsi éviter la véritable source du problème. Cela crée cependant encore plus de complexité et au final, un élément arrête de fonctionner, les développeurs se frustreront et démissionneront, ou les nouveaux projets sont tellement ralentis que leur progression perd tout espoir.

⁵ Oracle. *IT Assessment Complexity Survey*. 2015 <<http://www.oracle.com/us/corporate/features/it-complexity-assessment-survey-2281110.pdf>>

⁶ Boris Evelson. *Boost Your Business Insights By Converging Big Data And BI*. Forrester, 25 mars 2015 <<https://www.forrester.com/Boost+Your+Business+Insights+By+Converging+Big+Data+And+BI/fulltext/-/E-RES115633>>

« Les fournisseurs de bases de données relationnelles proposent toujours un produit des années 1990 aux utilisateurs, qui utilise un code écrit dans les années 1980 conçu pour résoudre des problèmes des années 1970 dans un domaine apparu dans les années 1960. »

« INFORMATIQUE DE L'OMBRE » ET FAILLES DE SÉCURITÉ

Le contrôle des données d'entreprise s'éloigne de plus en plus de la direction, car les employés et les services règlent leurs propres problèmes au moyen de logiciels qui ne se sont pas supervisés ou gérés par un service informatique centralisé. La plupart des directeurs pensent que leurs entreprises utilisent une petite dizaine d'applications « informatiques de l'ombre », mais c'est plus souvent une petite centaine. Dans une enquête, des sociétés utilisaient 923 services Cloud différents, un nombre incroyable, et seulement 9,3 % d'entre elles répondaient aux exigences de sécurité d'entreprise.⁷ Ce changement est le résultat direct de l'absence de réaction perçue face aux besoins de l'entreprise, ce qui crée des risques et des manques d'efficacité majeurs pour toute la société.

Cela survient pendant que le coût d'une faille de sécurité continue de grandir et que les cybercriminels recourent à des attaques de plus en plus sophistiquées. La réputation d'une entreprise peut être gravement entachée par une seule violation de sécurité, et une fuite de données peut se révéler également très coûteuse. Une étude a montré qu'un seul incident de cybersécurité peut coûter en moyenne 5,4 millions de dollars à une entreprise, ou 188 \$ par document.⁸ Malheureusement, la protection des données est encore plus difficile à réaliser que jamais en raison de la prolifération des silos de données qui créent plus de points d'entrée, de vulnérabilités et de fuite des données.

COÛTS ÉLEVÉS, ÉCHEC DES PROJETS ET INCAPACITÉ À INNOVER

Il est regrettable de savoir que de nombreux projets informatiques ne respecteront pas les délais et dépasseront le budget. Les coûts élevés et l'échec des projets représentent la norme et, en réalité, la moitié des projets informatiques avec des budgets de plus de 15 millions de dollars dépasse de 45 % le budget, prend 7 % de retard et propose 56 % de fonctionnalités en moins que prévu. Pis encore, environ 17 % des projets informatiques se déroulent tellement mal qu'ils peuvent menacer l'existence

même de l'entreprise.⁹ Ces problèmes se rencontrent malgré les longues planifications, les énormes ressources et les grandes équipes d'esprits brillants qui travaillent sur ces projets.

Lorsque les équipes travaillent sur des projets qui dépassent le budget et n'atteignent pas tous les objectifs, elles ne passent pas de temps sur les projets innovants qui sont absolument essentiels à la réussite de la société. Comment une entreprise peut-elle tirer profit des Big Data si elle ne peut y consacrer aucune ressource ? Aujourd'hui, 95 % de toutes les dépenses liées aux bases de données sont utilisées pour les bases de données relationnelles, qui stockent seulement environ 20 % des données d'entreprise.¹⁰ Il reste alors uniquement 5 % à dépenser sur la gestion des 80 % de données d'entreprise restants. Sans changement, les directeurs et les services informatiques géreront au final tout un éventail d'anciens systèmes et de responsabilités, et seront dépassés par des concurrents qui consacreront plus de ressources à l'innovation et à la transformation de l'entreprise.

POURQUOI LES BASES DE DONNÉES RELATIONNELLES NE FONCTIONNENT PAS ?

La plupart des défis actuels liés aux Big Data et la complexité conséquente peuvent être imputés aux bases de données relationnelles. Il n'y a rien de fondamentalement mauvais concernant les bases de données relationnelles : c'est simple, elles n'ont jamais été conçues pour gérer les données d'aujourd'hui. C'est pourquoi l'ancien directeur du gouvernement fédéral, Vivek Kundra, a déclaré en 2009 que « cette notion où l'on envisage les données dans une base de données relationnelle et structurée est morte ».

Inventées à la fin des années 1970, les bases de données relationnelles prirent la relève des systèmes centraux hiérarchiques pour devenir la base de données de choix au début des années 1990. Les bases de données relationnelles répondaient parfaitement aux besoins des

7 Skyhigh. *Cloud Adoption and Risk Report - Q1 2015*. Juin 2015 <<http://info.skyhighnetworks.com/rs/skyhighnetworks/images/WP%20CARR%20Q1%202015.pdf>>

8 Ponemon Institute. *2013 Cost of Data Breach Study: Global Analysis*. Symantec. 2013 <https://www4.symantec.com/mktginfo/whitepaper/053013_GL_NA_WP_Ponemon-2013-Cost-of-a-Data-Breach-Report_daiNA_cta72382.pdf>

9 McKinsey & Company. « Delivering large-scale IT projects on time, on budget, and on value, » octobre 2012 <http://www.mckinsey.com/insights/business_technology/delivering_large-scale_it_projects_on_time_on_budget_and_on_value>

10 Carl Olofson. *Worldwide Database Management Systems 2014–2018 Forecast and 2013 Vendor Shares*. IDC, juin 2014 <<http://www.idc.com/getdoc.jsp?containerId=248952>>

« Cette notion où l'on envisage les données dans une base de données relationnelle et structurée est obsolète. »

- Vivek Kundra, Federal CIO (Chargé du budget informatique au Ministère du budget des États-Unis), 21 juillet 2009 ; Open Government and Innovations Conference

débuts de l'informatique. Elles permettaient de découpler les applications à partir des données et d'utiliser moins de code personnalisé, et donnaient aux utilisateurs plus de contrôle sur la consultation des données grâce à SQL, le langage de requêtes standard.

Au cours de leurs 40 ans d'existence ou presque, les bases de données relationnelles ont continué de s'améliorer et les écosystèmes entourant chaque produit de fournisseur se sont développés, mais le modèle fondamental de gestion des données n'a pas changé. Le fait est que les fournisseurs de bases de données relationnelles proposent toujours un produit des années 1990 aux utilisateurs, qui utilise un code écrit dans les années 1980 conçu pour résoudre des problèmes des années 1970 dans un domaine apparu dans les années 1960.¹¹ Aujourd'hui, les entreprises ont besoin de beaucoup plus que ce que les bases de données relationnelles peuvent offrir, et les sections suivantes en débattent.

LES BASES DE DONNÉES RELATIONNELLES NE SONT PAS CONÇUES POUR GÉRER LE CHANGEMENT

Les bases de données relationnelles organisent les données en tableaux de lignes et de colonnes, comme les feuilles de calculs sous Microsoft Excel. Chaque ligne représente une entrée unique et chaque colonne décrit des attributs uniques. Une colonne est choisie comme clé principale pour identifier de façon unique chaque ligne dans le tableau.

Ainsi, si vous créez une base de données relationnelle pour des clients et les produits qu'ils ont commandés, vous pourriez commencer par la création d'un tableau « Clients » avec une colonne intitulée « ID client » à utiliser comme clé principale. Ensuite, vous créeriez d'autres colonnes pour chaque attribut de chaque client, comme par exemple « Prénom », « Nom » et « Adresse », en définissant le type de données à stocker dans chacune d'elles.¹² Puis, vous relieriez l'« ID client » à un autre tableau, « Commandes », qui contiendrait les informations relatives aux achats d'un client. Chaque ligne du tableau « Commandes » aurait son propre identifiant unique et également une référence à la clé principale du tableau « Clients ».

11 C'est en 1969 qu'Edgar « Ted » F. Codd a publié son célèbre document en interne chez IBM. Le document fut rendu public en 1970. (E.F. Codd, « A Relational Model of Data for Large Shared Data Banks. » Communications of the Association for Computing Machinery, Vol 13 No 6 Pgs. 377-387)

12 Dans les années 1980, les bases de données relationnelles limitaient les noms de colonne à 8 caractères et devaient être d'une seule casse. Par conséquent, les noms de colonne étaient « fnom » ou « Inom ». Aujourd'hui, la norme SQL limite les noms de colonne à 30 caractères.

Vous poursuivez ce processus de création de différents tableaux, en vous assurant que vos créations répondent à toutes les exigences d'entité et de référentiel, que tout est correctement « standardisé » pour éviter les colonnes répétées, que les colonnes sont toutes dépendantes de leur clé principale et qu'aucun tableau ne duplique d'information. Ces exigences maintiennent la cohérence des données et assurent la rapidité des requêtes, la marque même du modèle relationnel. Ce processus de conception du modèle de données, ou schéma de données, implique l'intervention d'une équipe dédiée qui doit décider des tableaux qui seront créés et des noms à donner aux colonnes. C'est un processus important et le résultat final est souvent et fièrement représenté sous la forme d'un grand diagramme entité-relation (ERD, entity-relationship diagram), qui est imprimé et accroché à une place de choix dans le couloir.

Les problèmes de cette approche sont doubles.

Premièrement, le processus peut prendre des mois, voire des années, en fonction de la taille de la base de données. Les schémas relationnels sont complexes, et toute la modélisation doit être réalisée *avant* de charger les données ou de développer l'application. Deuxièmement, si un changement est nécessaire après le développement des applications sur la base de données, ce processus requiert du temps, des ressources et peut prendre à nouveau plusieurs mois ou années. Le modèle relationnel est comparable à un écosystème de forêt tropicale sensible et complexe, dans lequel un petit changement peut provoquer des effets néfastes et des répercussions en chaîne dans la base de données et la pile applicative. Même un simple changement comme l'ajout ou le remplacement d'une colonne dans un tableau peut coûter jusqu'à un million de dollars.¹³

Aujourd'hui, les changements sont fréquents et la modélisation des données est un défi majeur en raison du temps et des ressources que nécessitent les bases de données relationnelles. Chaque année, des milliards de dollars sont dépensés dans la modélisation des données et les processus ETL afin de créer et de recréer des modèles plus « parfaits » qui ne changeront jamais. Mais ils changent toujours.

13 Selon le client d'une entreprise de technologie leader du classement Fortune 100, le coût pour l'ajout d'une colonne peut prendre jusqu'à une année et coûter plus d'un million de dollars. Pour d'autres projets de modélisation des données plus complexes impliquant la gestion de données permanentes, des délais encore plus longs dépassant 5 ans ont été signalés.

« Il est surprenant de constater que 95 % des dépenses totales des bases de données concernent les bases de données relationnelles, alors que celles-ci sont conçues uniquement pour gérer les 20 % des données qui sont structurées. »

LES BASES DE DONNÉES RELATIONNELLES NE SONT PAS CONÇUES POUR LES DONNÉES HÉTÉROGÈNES

Étonnamment, 95 % des dépenses totales des bases de données concernent les bases de données relationnelles, alors que celles-ci sont conçues uniquement pour gérer les 20 % des données qui sont structurées.¹⁴ Les entreprises font face à une incapacité croissante de gérer ces données structurées, tandis que les autres 80 %, les données non structurées, deviennent complètement orphelines en dépit de l'énorme valeur qu'elles renferment.

Les sociétés avaient coutume de stocker uniquement certaines données transactionnelles clés et quelques éléments fondamentaux concernant leurs clients. De nos jours, cependant, les entreprises ne peuvent plus choisir les données qui leur plaisent uniquement. Elles doivent tout stocker. Le coût pour l'infrastructure réalisant cela est désormais raisonnable, et les sociétés peuvent tirer parti de cette opportunité pour réduire les risques et diminuer les coûts. Les clients, les partenaires et les organismes de régulation attendent également des entreprises qu'elles stockent tout dans un format utilisable dont ils pourront aussi tirer profit.

La quantité grandissante de données structurées est problématique pour les bases de données relationnelles car la structure de chaque source de données est différente. Les modifications nécessaires pour gérer une nouvelle source de données, comme nous l'avons déjà indiqué, sont lourdes et complexifient les schémas. C'est également le cas lorsque les nouvelles données représentent le même domaine ou les mêmes concepts.

Le nombre croissant de données non structurées présente aussi un problème pour les bases de données relationnelles. Les lignes et les colonnes d'une base de données relationnelle sont idéales pour stocker des ensembles de valeurs, mais la plupart des informations sont composées d'éléments bien différents. Prenez par exemple le dossier médical d'une personne. Il inclut des **valeurs** (nom, date de naissance), des **relations** (aux membres de sa famille ou aux prestataires de soins, aux symptômes et aux médicaments), des **données géospatiales** (adresses), des **métadonnées** (provenance, attributs de sécurité), des **images** (scanners) et des **textes libres** (notes des médecins, transcriptions).

Maintenant, imaginez que vous insériez toutes ces données dans une feuille de calcul Microsoft Excel. Cette tâche nécessiterait beaucoup d'ingéniosité et de nombreux choix complexes. *Les gros blocs de texte doivent-ils être séparés ou réunis dans une cellule du tableau ? Comment stocker les nouvelles sources de données qui se présenteront ultérieurement ? Combien de colonnes utiliser pour les métadonnées ? Comment apprécier les relations entre les différentes entités ? Comment gérer la structure au sein du document ? Quels index doivent être créés ? Que faire si je désire filtrer les données avec un élément qui n'est pas défini par une ligne ou une colonne ?*

En dépit de tout le travail effectué et des compromis utilisés pour que le modèle relationnel fonctionne pour tout le monde, le fait est qu'il n'a pas été conçu pour les données hétérogènes.

LES BASES DE DONNÉES RELATIONNELLES NE SONT PAS CONÇUES POUR L'ÉVOLUTIVITÉ ET L'ÉLASTICITÉ

Aujourd'hui, les entreprises possèdent des millions d'utilisateurs et des pétaoctets de données. Elles exécutent leurs applications dans le Cloud afin de proposer du contenu dynamique à des millions de bureaux, tablettes et appareils mobiles dans différents emplacements géographiques. Pour gérer cette nouvelle réalité, les sociétés ont besoin de l'évolutivité (l'ajout de capacité pour plus de données et plus d'utilisateurs) et de l'élasticité (la *facilité* avec laquelle le système évolue, généralement la capacité à se réduire lorsque la demande utilisateur disparaît).

Malheureusement, l'évolutivité des bases de données relationnelles représente un véritable défi. Les bases de données relationnelles sont conçues pour s'exécuter sur un seul serveur, afin de maintenir l'intégrité des mappages tabulaires et d'éviter les problèmes de l'informatique distribuée. Dans cette conception, lorsqu'un système a besoin d'évoluer, les clients doivent acheter des matériels propriétaires plus grands, plus complexes et plus onéreux, avec plus de puissance de calcul, de mémoire et de stockage. Les mises à niveau constituent également un défi car la société doit suivre un processus d'acquisition très long, et souvent basculer le système hors ligne pour effectuer véritablement les modifications. Tout cela intervient alors que le nombre d'utilisateurs continue de croître, provoquant de plus en plus de tension et augmentant les risques sur les ressources sous-provisionnées.

14 IDC, juin 2014

« Le problème ? Le modèle relationnel fait apparaître des éléments complexes que les services informatiques doivent prendre en compte, car il n'a pas été conçu pour livrer des informations à différents groupes d'utilisateurs, de la bonne façon et au bon moment. »

Pour gérer ces difficultés, les fournisseurs de bases de données relationnelles ont élaboré toute une série d'améliorations. Aujourd'hui, l'évolution des bases de données relationnelles leur permet d'utiliser des architectures plus complexes, qui s'appuient sur un modèle « maître-esclave » où les « esclaves » sont des serveurs supplémentaires qui peuvent gérer des données répliquées ou des données « éclatées » (divisées et partagées sur de multiples serveurs ou hôtes) pour alléger la charge de travail du serveur maître. D'autres améliorations comme le stockage partagé, le traitement en mémoire, une meilleure utilisation des répliques, la mise en cache distribuée et de nouvelles architectures relationnelles ont permis de rendre les bases de données relationnelles plus évolutives. Mais la réalité est qu'il n'est pas difficile de trouver un système avec un point unique de défaillance.¹⁵

Les améliorations apportées aux bases de données relationnelles impliquent également des coûts élevés et de gros compromis. Par exemple, lorsque les données sont distribuées dans une base de données relationnelle, elles sont généralement basées sur des requêtes prédéfinies afin de maintenir les performances. En d'autres termes, la flexibilité est sacrifiée au profit des performances. En outre, les bases de données relationnelles ne sont pas conçues pour être réduites, elles sont fortement *non* élastiques. Dès que les données ont été distribuées et qu'un espace supplémentaire a été attribué, il est presque impossible d'« annuler la distribution » de ces données.

LES BASES DE DONNÉES RELATIONNELLES NE SONT PAS CONÇUES POUR LES CHARGES DE TRAVAIL VARIÉES

Les « charges de travail variées » se réfèrent à la capacité de gérer à la fois des charges de travail opérationnelles et analytiques. Les charges de travail opérationnelles englobent les transactions commerciales journalières qui surviennent en temps réel, notamment les achats effectués par un grand nombre de clients. Les charges de travail analytiques regroupent les opérations de stratégie d'entreprise et d'extraction de données, par exemple lorsqu'un analyste souhaite examiner un groupe d'achats sur une période spécifique.

Au milieu des années 1990, une scission apparut entre les bases de données optimisées pour les charges de travail

opérationnelles, ou systèmes OLTP (*online transaction processing*, traitement de transactions en ligne), et les bases de données optimisées pour les charges de travail analytiques, ou systèmes OLAP (*online analytical processing*, traitement analytique en ligne). Dans les systèmes OLTP, les données sont modélisées pour être optimales pour l'application développée sur elles, ce qui nécessite des transactions constantes et rapides. Dans les systèmes OLAP, les données sont modélisées pour être optimales pour le partage et le découpage, dont les agrégats et les tendances. Très vite, des modèles sophistiqués furent développés, les experts s'entendant et ne s'entendant pas sur les meilleures façons de modéliser les données pour différents scénarios. C'est à ce moment que les « schémas en étoile », les « schémas en flocon de neige » et les « hypercubes OLAP » entrèrent dans le jargon des modélisateurs de données.

Malheureusement, la scission entre les systèmes opérationnels et analytiques contribua à la création de mini-entrepôts de données disparates, d'entrepôts de données, d'unités de stockage de données de référence et d'archives qui se sont multipliés par nécessité. Les données des systèmes opérationnels furent déplacées via ETL dans un entrepôt de données central, conçu pour être l'entrepôt de toutes les décisions commerciales. Il échoua cependant lorsqu'il ne réussit pas à répondre aux nouvelles et différentes questions qui apparaissaient. Un autre processus ETL fut donc utilisé pour déplacer un certain sous-ensemble de données dans un mini-entrepôt de données. D'autres systèmes furent définis pour accueillir les données de référence. Puis, un système d'archive conserva toutes les données historiques de tous les systèmes. À chaque fois qu'une nouvelle question devait être posée ou une nouvelle application développée, un nouveau et meilleur modèle était créé... et aucun modèle ne se ressemblait jamais. Ce qui se limitait à un simple schéma et quelques bases de données se chiffra bientôt en centaines.

C'est l'une des raisons pour lesquelles la plupart des services informatiques actuels utilisent la majorité de leur temps et de leur argent à gérer simplement la myriade de systèmes dans l'entreprise. Le problème est le suivant : le modèle relationnel fait apparaître des éléments complexes face auxquels les services informatiques doivent répondre, car il n'a pas été conçu pour livrer des informations à différents groupes d'utilisateurs, de la bonne façon et au bon moment.

¹⁵ Par exemple, Oracle RAC est une base de données relationnelle « groupée » qui utilise un système de fichiers en grappes, mais il existe toujours un sous-système de disque partagé en dessous.



Le résultat de l'architecture relationnelle traditionnelle est une perte de performance et plus d'opportunités pour du code erroné. »

LES BASES DE DONNÉES RELATIONNELLES SONT DISPARATES POUR LE DÉVELOPPEMENT D'APPLICATIONS MODERNES

Les applications modernes sont développées avec des langages de programmation orientés vers l'objet, notamment Java, JavaScript, Python et C#, pour n'en citer que quelques-uns. Ces langages traitent les structures de données comme des « objets » qui contiennent des données et du code (comme les attributs et les méthodes). Malheureusement, cette gestion des données est très différente de celle adoptée par bases de données relationnelles, d'où une disparité d'impédance entre la base de données et la programmation d'applications.

Pour contourner la disparité d'impédance, les développeurs utilisent une technique baptisée mapping objet-relationnel (ORM, object-relational mapping), un mappage actif-actif bidirectionnel entre les objets dans la couche de l'application et les données qui sont représentées dans le schéma de la base de données relationnelle. Avec ORM, les développeurs d'applications utilisent les règles commerciales et la logique, pour créer des vues des données qui correspondent le mieux du point de vue du développement de l'application. Avec cette approche, les bases de données sont vues plus simplement comme le lieu où les données sont conservées et où les procédures stockées sont gardées. Un large éventail d'outils ORM est disponible, ce qui permet de simplifier le développement des applications avec des bases de données relationnelles. Entre autres outils ORM, citons Hibernate for Java, ActiveRecord for Ruby on Rails, Doctrine for PHP et SQLAlchemy for Python.

Malheureusement, ORM est également considéré comme une médiocre solution de rechange pour un problème systémique avec des bases de données relationnelles. ORM a été baptisé le « Viêt-Nam de l'informatique », car il « représente un bourbier qui démarre bien, devient de plus en plus complexe avec le temps et emprisonne rapidement ses utilisateurs dans un engagement qui n'a pas de point de démarcation clair, de condition de victoire évidente et de stratégie de sortie manifeste. »¹⁶ De nombreuses autres publications ont continué de montrer qu'ORM était plus dommageable qu'utile.¹⁷

¹⁶ Ted Neward. Billet de blog sur « The Blog Ride », 26 juin 2006 <<http://blogs.tedneward.com/2006/06/26/The+Vietnam+Of+Computer+Science.aspx>>

¹⁷ Consultez *OrmHate* par Martin Fowler <<http://martinfowler.com/bliki/OrmHate.html>>, *Object-Relational Mapping Is the Vietnam of Computer Science* par Jeff Atwood, *ORM Is an Anti-Pattern* par Laurie Voss <http://seldo.com/weblog/2011/08/11/orm_is_an_antipattern>, *ORM is An Offensive Anti-Pattern* part Yegor Bugayenko <<http://www.yegor256.com/2014/12/01/orm-offensive-anti-pattern.html>> et bien d'autres.

Au lieu de préserver les aspects intéressants des données dans un objet, ORM choisit d'extraire les données, de les démanteler et d'ajouter encore plus de gestion dans le processus. Et cela se produit après que les données aient déjà été séparées dans différents tableaux via le processus de normalisation. Si l'on reprend l'exemple du dossier médical d'une personne, envisagez toutes les données différentes du dossier qui doivent être séparées dans les tableaux d'une base de données relationnelle. Après avoir été « morcelées » dans les tableaux, les données doivent être rassemblées pour être affichées ou agrégées dans la couche de l'application, puis être présentées à l'utilisateur. Cela impose beaucoup de gestion, un mappage important et une structure personnalisée ou de nombreux liens afin d'obtenir des vues matérialisées de l'entité commerciale (par exemple, un formulaire ou un document).

Le résultat de l'architecture relationnelle traditionnelle : une perte de performance et plus d'opportunités pour du code erroné. Dans les cycles de développement d'applications rapides d'aujourd'hui, où les utilisateurs demandent plus d'interactivité et de réactivité, le modèle relationnel révèle ses défauts. Plutôt que devoir trouver des solutions de rechange pour le modèle relationnel disparate, les développeurs adoptent de nouveaux modèles qui impliquent moins d'abstraction et révèlent des performances plus élevées.

UNE BASE DE DONNÉES NOUVELLE GÉNÉRATION DÉDIÉE AUX DONNÉES D'AUJOURD'HUI

MarkLogic est une base données NoSQL (« Not only SQL », pas seulement SQL) qui se trouve à l'avant-garde d'un changement vis-à-vis des bases de données relationnelles universelles. De nombreuses fonctions de MarkLogic en font le choix logique des données d'aujourd'hui, mais MarkLogic possède quatre éléments qui le rendent vraiment unique :

1. Un **modèle de données flexible** pour stocker les données variées, changeantes et disparates d'aujourd'hui
2. **Capacités de recherche intégrées** pour tirer le meilleur des données à n'importe quel moment
3. **Évolutivité et élasticité** pour gérer les volumes de données massifs et changeants
4. **Fonctionnalités d'entreprise** nécessaires pour exécuter des applications d'entreprise stratégiques

« Le premier projet MarkLogic a duré 60 jours. La durée estimée avec la technologie existante était de 3 000 jours. »

Paolo Pelizzoli, responsable de l'architecture sur le plan mondial, opérations et technologie mondiales chez Broadridge Financial Solutions

PRÉSENTATION DES FONCTIONS QUI DÉMARQUENT MARKLOGIC

MODÈLE FLEXIBLE DE GESTION DES DONNÉES

MarkLogic est une base de données multimodèles conçue pour stocker de manière native et effectuer des recherches rapides dans des triplets JSON, XML et RDF, des données géospatiales et des fichiers binaires volumineux (comme des images ou des vidéos). Cette capacité permet à MarkLogic d'être bien plus habile dans la gestion d'un éventail encore plus large de types de données et de structures de données que les bases de données relationnelles. Elle lui permet par ailleurs de simplifier la gestion des changements dans le modèle de données à mesure de la transformation des données.

MarkLogic recourt principalement aux deux formats de documents JSON et XML pour stocker ses données. Contrairement aux tableaux d'une base de données relationnelle, les documents sont plus lisibles pour l'utilisateur et fournissent une approche plus naturelle à la modélisation des données riches, variables et complexes avec lesquelles travaillent les entreprises d'aujourd'hui. En utilisant un modèle de données plus riche, les sociétés peuvent tirer le meilleur de leurs données.

Les documents sont également mieux adaptés au développement d'applications modernes, car ils évitent le problème de la disparité d'impédance. Le modèle des documents permet aux développeurs de maintenir l'intégrité des données dans chaque niveau d'une pile applicative. Par exemple, les développeurs peuvent utiliser JSON dans la base de données, la couche de l'application et l'interface utilisateur. C'est une approche plus flexible qui correspond bien à la quantité croissante de JavaScript utilisé dans le développement des applications Web modernes.

Autre avantage du modèle de données de MarkLogic : aucun schéma, ou structure, n'a besoin d'être défini avant de charger les données dans MarkLogic. Le terme de cet attribut est « indifférent au schéma ». MarkLogic permet aux utilisateurs de stocker des documents aux schémas différents et de modifier des schémas spécifiques sans perturber les autres. Cette puissante capacité permet aussi aux utilisateurs de combiner rapidement des tableaux de données relationnels avec différents modèles, le tout dans MarkLogic.

MarkLogic possède également des fonctions de base de données graphique car il stocke de manière native des triplets RDF, le langage de la sémantique. À haut niveau, la sémantique est un modèle de données qui relie deux entités (des personnes, des lieux ou des choses) en fonction de leur relation et forme un triplet. Lorsqu'ils sont reliés, les triplets constituent un graphe lisible par les machines, qui peut être utilisé pour déduire des faits nouveaux. MarkLogic peut stocker des centaines de milliards de triplets aux côtés des documents JSON ou XML, ou même à l'intérieur.

MarkLogic est la seule base de données d'entreprise qui combine une unité de stockage de documents et une unité de stockage de triplets. Cette capacité unique accélère et facilite la modélisation des données dans le format qui semble le plus judicieux et améliore la valeur que les entreprises peuvent tirer de leurs données. Dans l'ensemble, le modèle de données de MarkLogic est bien plus flexible que le modèle relationnel. Il propose une plate-forme permettant de développer plus rapidement des applications plus intelligentes, et plus de flexibilité pour gérer les changements qui se présentent.

Les améliorations par rapport aux bases de données relationnelles peuvent être spectaculaires : « Le premier projet MarkLogic a duré 60 jours. La durée estimée avec la

COMPARAISON DES BASES DE DONNÉES NOSQL

Suite à la nécessité de changement urgent, les nouvelles technologies de gestion des données ont connu un essor spectaculaire au cours des dernières années, proposant toutes d'options plus efficaces pour gérer les données d'aujourd'hui. Il existe de nombreuses bases de données NoSQL open source de différents types, dont des unités de stockage de documents, des unités de stockage de graphes, des unités de stockage de colonnes et des unités de stockage de valeurs clés. Il existe des similitudes générales entre elles, notamment la capacité à mettre à l'échelle le matériel de base, mais chaque type de base de données est vraiment très différent. Si vous cherchez à mieux comprendre le paysage NoSQL, téléchargez l'e-book, *Enterprise NoSQL for Dummies*, disponible gratuitement sur info.marklogic.com/nosql-for-dummies.html.

« MarkLogic est conçu pour gérer le volume, la variété et la vitesse du Big Data comme aucune autre solution NoSQL, *ET* il possède les fonctions d'entreprise qui ont assuré la fiabilité des bases de données relationnelles de la génération précédente. »

technologie existante était de 3 000 jours », a déclaré Paolo Pelizzoli, responsable de l'architecture sur le plan mondial, opérations et technologie mondiales chez Broadridge Financial Solutions.

CAPACITÉS DE RECHERCHE INTÉGRÉES

Pour rechercher des données rapidement et précisément, une base de données a besoin d'index. Les index des bases de données sont semblables aux index que l'on trouve à la fin des livres : ils fournissent une liste d'informations dans le livre rapidement consultables, ce qui évite de passer en revue l'ouvrage tout entier. Avec la plupart des bases de données, l'indexation est généralement considérée comme une tâche secondaire du stockage des données. L'indexation est un processus complexe via lequel les utilisateurs doivent définir les index qui doivent être créés pour répondre à des questions, la nature des implications de performance de chaque index et la façon dont les index seront gérés. Puis, pour obtenir des recherches en texte intégral, des index en texte intégral sont nécessaires. Avec les bases de données relationnelles, cela implique d'ajouter des logiciels supplémentaires qui doivent être configurés et gérés en même temps que la base de données.

MarkLogic travaille différemment, car le produit possède une formidable fonction d'indexation et des capacités intégrées de recherche en texte intégral. MarkLogic répertorie le contenu et la structure des données lorsqu'elles sont chargées, et possède de nombreux index (par exemple, index de plages, index de triplets, index géospatiaux) qui peuvent être activés et désactivés. Les index de MarkLogic facilitent les réponses à la fois des requêtes simples et sophistiquées grâce à un éventail de langages de requêtes : JavaScript, XQuery, SPARQL (le langage de requêtes pour la sémantique) et SQL bien évidemment. Un exemple de requête sophistiquée serait de « trouver tous les gains et les classements des athlètes professionnels avec lesquels Michael Jordan a joué pendant sa carrière. Limitez les résultats aux athlètes vivant à New York et qui ont été cités dans des sources d'informations fiables après janvier 2015. Classez les résultats des candidats par pertinence. »

Répondre à de telles questions multidimensionnelles n'est pas trivial, car ce serait extrêmement difficile ou impossible d'y parvenir avec une base de données relationnelle. Une base de données relationnelle aurait

des difficultés à modéliser les relations entre Michael Jordan et les athlètes avec lesquels il a joué, et à trouver les mentions spécifiques dans les sources d'informations, qui seraient des documents au format texte. Une base de données relationnelle ne pourrait pas réaliser également de classement par pertinence à la manière d'un moteur de recherche tel que Google ; elle renverrait une liste de résultats basée sur un ordonnancement simple de valeurs.

D'autre part, ce type de requête sophistiquée est le type même que MarkLogic peut gérer avec peu de code. Les difficultés liées à la pertinence et au texte intégral sont résolues par le modèle de données riche de MarkLogic et par ses index puissants, conçus pour répondre aux mêmes types de questions que quiconque poserait avec SQL dans une base de données relationnelle et bien plus encore. Même si les questions changent, cela ne pose aucun problème. MarkLogic est prêt à gérer les nouvelles requêtes inattendues qui se présentent, sans que les utilisateurs n'aient à reconfigurer les données et les index comme dans une base de données relationnelle. En plus, MarkLogic renvoie les résultats avec des temps de réponse inférieurs à la seconde, avec des centaines de téraoctets de données dans un système dans lequel les données sont gérées constamment et de façon fiable.

ÉVOLUTIVITÉ ET ÉLASTICITÉ

MarkLogic n'est pas limité par les architectures de serveur unique, et il a été conçu pour une mise à l'échelle massive dans les systèmes distribués. MarkLogic se développe « horizontalement », ce qui signifie qu'il s'exécute sur plusieurs serveurs reliés, chacun partageant une partie de la charge de travail. Avec cette approche, MarkLogic peut fonctionner avec des centaines de serveurs, des pétaoctets de données et des milliards de documents...et continuer de gérer des dizaines de milliers de transactions par seconde. Et il peut réussir tout cela sur du matériel de base peu coûteux fonctionnant dans n'importe quel environnement, que ce soit sur site ou dans un environnement Cloud tel qu'Amazon Web Services.

La mise à l'échelle massive est impressionnante, mais le facteur encore important est probablement la *flexibilité* de MarkLogic. L'architecture unique de MarkLogic permet d'ajouter ou de supprimer rapidement et facilement des nœuds dans un cluster, afin que la base de données reste en phase avec les besoins de performance sans

« MarkLogic est éprouvé dans des systèmes stratégiques au sein du Département de la Défense des États-Unis, de grandes banques d'investissement, de payeurs de soins de santé, d'organismes médias mondiaux et de nombreux autres secteurs dans lesquels la réussite n'est pas facultative. »

provisionnement excessif coûteux. Il n'y a pas d'éclatement complexe des données ou de solutions de rechange architecturales, les données sont automatiquement rééquilibrées dans un cluster lorsque des nœuds sont ajoutés ou supprimés. C'est l'une des raisons pour lesquelles MarkLogic est si simple d'utilisation au niveau administratif.

FONCTIONS D'ENTREPRISE

On croit à tort que NoSQL est réservé aux applications non sérieuses, que NoSQL est réservé uniquement aux start-ups ou qu'il accueille les données non cruciales des entreprises. Mais ce n'est simplement pas le cas avec MarkLogic.

MarkLogic possède toutes les fonctions d'entreprise majeures qui ont fait la fiabilité des bases de données relationnelles de la génération précédente, et qui sont absolument essentielles pour stocker et gérer les données d'entreprise. Voici plusieurs fonctions d'entreprise clés de MarkLogic :

- **Les transactions ACID** qui assurent la cohérence des données et évitent leur perte ou leur corruption
- **Une sécurité certifiée** qui permet à MarkLogic de s'exécuter dans les centres de données d'entreprise
- **Une haute disponibilité et la reprise après sinistre** afin que les données soient toujours disponibles
- **La surveillance des performances** pour contrôler le provisionnement et l'utilisation des ressources
- **Des outils de gestion d'entreprise** qui proposent des approches automatisées pour les tâches standard

Avec MarkLogic, toutes ces fonctions ne sont pas de simples noms dans une liste. MarkLogic les a toutes éprouvées dans des systèmes stratégiques au sein du Département de la Défense des États-Unis, de grandes banques d'investissement, de payeurs de soins de santé, d'organismes médias mondiaux et de nombreux autres secteurs dans lesquels la réussite n'est pas facultative.

LES ENTREPRISES LEADERS SONT ENCORE PLUS EFFICACES AVEC MARKLOGIC

Des centaines de sociétés ont adopté le changement et l'innovation en utilisant MarkLogic pour dynamiser leur avenir commercial. Ces exemples soulignent certaines de ces nombreuses réussites.

DIRECTION D'UN SYSTÈME COMMERCIAL OPÉRATIONNEL DANS UNE GRANDE BANQUE

Une banque d'investissement appartenant au top 5 a remplacé 20 bases de données relationnelles par une base de données MarkLogic. MarkLogic dirige désormais la plate-forme commerciale dérivée dans la banque, et gère plus de 100 000 transactions par jour et 32 millions d'échanges simultanés dans le système. Ce volume élevé génère une couverture des risques de flux de trésorerie supérieure à 100 millions de dollars. En plus des économies majeures réalisées grâce à l'utilisation de MarkLogic, la banque a également obtenu une vision globale, en temps réel, unifiée et précise de son activité commerciale dérivée.

INSCRIPTION DE MILLIONS DE BÉNÉFICIAIRES VIA LE SITE HEALTHCARE.GOV

L'agence fédérale CMS (Centers for Medicare & Medicaid Services) fournit une couverture santé à des millions d'Américains via le site HealthCare.gov et valide les conditions d'admissibilité au régime d'assurance à partir de diverses sources de données fédérales, tout en gérant des centaines de milliers d'utilisateurs simultanés, sans déplorer aucune perte de contenu. Au cours des deux premières années, le système a permis l'inscription de 12 millions d'Américains au régime d'assurance santé. Ce succès va à l'encontre de certains des échanges en matière de santé fédérale qui n'ont pas tenu leurs promesses, des échecs qui ont même conduit un État à engager des poursuites contre un fournisseur de base de données relationnelle.¹⁸

AUGMENTATION DES PERFORMANCES POUR LE SERVICE DE STREAMING iPLAYER DE LA BBC

iPlayer est le service de streaming télévisé de la BBC au Royaume-Uni. Aujourd'hui, une émission sur iPlayer peut dépasser les 3 milliards de visionnages. Pour gérer l'évolutivité massive et les exigences de performances, l'équipe de la BBC est passée de l'utilisation de la technologie relationnelle à MarkLogic en tant qu'élément principal pour stocker et livrer les métadonnées liées aux programmes de la BBC. La BBC a déjà rencontré le succès avec le développement d'une plate-forme de fourniture de contenus dynamiques sur MarkLogic pour les Jeux olympiques de 2012, et voulait également valoriser l'évolutivité et la flexibilité de MarkLogic pour iPlayer.

¹⁸ Shelby Stebens. « Oracle sues Oregon officials in healthcare website dispute, » Reuters, 27 février 2015 <<http://www.reuters.com/article/2015/02/27/us-usa-healthcare-oregon-idUSKBN0LV2LK20150227>>

« MarkLogic nous a permis d'accélérer le développement de nos produits et d'associer les utilisateurs commerciaux à l'équipe informatique. L'informatique est désormais considérée comme essentielle dans l'entreprise ; elle fait partie de la solution plutôt que du problème pour fournir la solution. »

Andrea Powell, directeur du CABI (The Centre for Biosciences and Agriculture International)

Après son implémentation, les requêtes qui nécessitaient 20 secondes auparavant avec SQL ne prenaient plus que 20 millisecondes avec MarkLogic, des ordres de grandeur bien meilleurs.

DÉPASSER LE MODÈLE RELATIONNEL AVEC VOS DONNÉES

Passer de l'ancien au nouveau monde peut sembler intimidant au préalable. C'est pourquoi il est souvent préférable de commencer par un projet plus petit, puis d'évoluer. Nous vous donnons ci-dessous quelques recommandations, qui vous aideront à planifier la transition pour utiliser NoSQL.

PROCHAINES ÉTAPES RECOMMANDÉES

CRÉER UN SENTIMENT D'URGENCE

En rendant la nécessité d'évoluer évidente pour tous, vous suscitez l'enthousiasme susceptible de faire adopter NoSQL. Une fois un sentiment d'urgence créé, il sera possible d'établir votre vision et de remporter l'adhésion.

CHOISIR L'ÉQUIPE GAGNANTE

Le succès nécessite une technologie supérieure, mais également des personnes et des processus fantastiques. Il est essentiel d'identifier à la fois les décideurs et les responsables de la mise en œuvre dans l'entreprise, et de comprendre les processus qui permettront au projet de prendre forme ou l'en empêcheront.

DÉMARRER AVEC LE BON PROJET

Choisir le bon cas d'utilisation est essentiel pour commencer. Il est souvent préférable de démarrer avec un petit projet percutant qui évitera toute perturbation inutile. Des spécialistes MarkLogic peuvent vous accompagner afin de s'assurer que votre projet correspond bien à une solution MarkLogic.

UTILISER MARKLOGIC TRÈS TÔT

Des experts MarkLogic dédiés sont disponibles et possèdent plus d'expérience combinée en matière de NoSQL que n'importe quelle autre équipe du secteur. Il est crucial de les utiliser très tôt pendant la phase de développement, lorsque l'assistance est la plus déterminante.

PLUS D'INFORMATIONS

- **e-book NoSQL pour les nuls** – Découvrez les bases de données NoSQL avec un e-book gratuit info.marklogic.com/nosql-for-dummies
- **En quoi consiste MarkLogic ?** – Découvrez les fonctions uniques de MarkLogic sur marklogic.com/what-is-marklogic
- **Livre blanc Découvrir MarkLogic** – Découvrez les raisons pour lesquelles MarkLogic est aussi puissant marklogic.com/resources/inside-marklogic-server
- **Convenir d'un rendez-vous** – Discutez de votre cas d'utilisation spécifique avec un commercial MarkLogic en nous contactant à l'adresse sales@marklogic.com