

LES ATOUTS DE LA BITEMPORALITÉ

LIVRE BLANC MARKLOGIC • JUIN 2015

Dans un contexte juridique où les amendes se comptent en milliards de dollars et où les actions en justice sont aussi longues que coûteuses, il est absolument essentiel de disposer d'un système central d'enregistrement étayé par une base de données solide. Cependant, l'inconvénient majeur des bases de données traditionnelles réside dans leur incapacité à conserver un historique complet des actions passées. Seule une base de données bitemporelle vous permet de mettre en place un suivi véritablement exhaustif et fiable, et de comprendre exactement à la fois la source (« ce que vous saviez ») et l'historique (« quand vous en avez eu connaissance ») des informations disponibles.



ÉVALUATION : AVEZ-VOUS BESOIN DE LA BITEMPORALITÉ ?

Avant de poursuivre votre lecture, demandez-vous si votre entreprise a besoin d'une structure de gestion des données bitemporelle. Si vous répondez « oui » à une des questions suivantes, cela signifie que vous devriez envisager le déploiement d'une solution bitemporelle.

	OUI	NON
1. Est-il vital pour votre entreprise de réaliser un suivi des moments où ont lieu des événements ou des transactions ?	✓	
2. Avez-vous déjà dû mettre à jour des données historiques ?	✓	
3. Existe-t-il un décalage entre une occurrence dans le monde réel et son enregistrement dans la base de données ?	✓	
4. Recevez-vous fréquemment des demandes provenant d'organismes de réglementation souhaitant avoir accès à des données historiques ?	✓	
5. Évoluez-vous dans un secteur où la séquence d'obtention de certaines informations est importante, notamment en droit et en informatique décisionnelle ?	✓	
6. Le stockage et l'accès à des données historiques dans votre entreprise représentent-ils un coût important et une grande complexité ?	✓	
7. La gestion et l'accès aux données historiques représentent-ils un coût élevé en termes de ressources développeurs, ou un risque de plus en plus important ?	✓	

Table des matières

Introduction	1
Le coût de travailler sans bitemporalité	
Trois types de temporalité	2
Non-temporelles	
Unitemporelles	
Bitemporelles	
Les avantages de la bitemporalité	4
Les actions que vous permet la bitemporalité	
Le besoin croissant de bitemporalité	
La bitemporalité au sein des industries	
Pourquoi l'implémentation de la bitemporalité a été difficile	7
Pourquoi la bitemporalité est adaptée au contexte actuel	8
Fonctionnalités clés de bitemporalité dans MarkLogic	
Mise en route rapide	
Plus d'informations	

INTRODUCTION

À l'heure actuelle, ce sont les bases de données qui constituent le système d'autorité pour le traitement des informations, et non pas les copies papier. Dans ce nouveau contexte, les entreprises sont tenues de conserver une image exacte de tous les événements, au moment où ils se produisent. Pour certaines industries, par exemple les services financiers, les assurances et la santé, des lois régissent même les modes de suivi et de gestion des données historiques.

Malheureusement, les bases de données traditionnelles sont incapables de fournir une image vraiment claire et précise de votre entreprise à différents moments dans le temps. En effet, les bases de données traditionnelles sont unitemporelles et ne peuvent suivre que les heures de début et de fin sur une chronologie unique. Mais, en cas de décalage entre le moment où un événement a lieu et le moment où vous vous en rendez compte, quelle en est la conséquence ? Quel temps devez-vous consigner ? Que se passe-t-il si vous vous rendez compte de devoir corriger le moment où un événement s'est produit, sans vouloir remplacer des données historiques ? Dans ces cas, une seule chronologie ne suffit pas.

Une base de données bitemporelle vous permet d'enregistrer des données et faire des recherches sur deux chronologies avec un horodatage pour le temps valide (le moment où un événement est survenu dans le monde réel - « ce que vous saviez »), et le temps système (le moment où cet événement a été enregistré dans la base de données - « quand vous en avez eu connaissance »). En effectuant un suivi des événements sur deux chronologies avec une base de données bitemporelle, vous pouvez bénéficier d'une image exhaustive et exacte de votre entreprise à un moment donné à des fins de recherche et de découverte interne, ou lorsque des organismes de réglementation procèdent à des audits.

Considérez certaines des nouvelles questions qu'une base de données bitemporelle vous permet de poser :

- Quelle était la cote de solvabilité de mon client l'an dernier au vu des informations dont je disposais au dernier trimestre ?
- Quelle était notre position concernant cette garantie avant la modification de la transaction ?
- Qu'indiquait notre informatique décisionnelle avant que nous prenions connaissance de ces nouveaux éléments informationnels ?

Une base de données unitemporelle traditionnelle vous permet d'avoir accès à la cote de solvabilité de votre client telle que vous la connaissez aujourd'hui, mais pas celle dont vous aviez connaissance hier ou le trimestre dernier.

Seule une base de données bitemporelle vous permet d'obtenir une image précise et inaltérée des données

historiques, intégrant les changements passés et présents. Une base de données bitemporelle est nécessaire pour permettre aux entreprises d'aujourd'hui d'explorer précisément les données historiques, de gérer ces données au sein des divers systèmes, de garantir une parfaite intégrité des données, et de réaliser une analyse plus poussée.

MarkLogic® est la base de données NoSQL d'entreprise la mieux à même d'enregistrer et de gérer des données bitemporelles, pour les raisons suivantes :

- **Modèle de données flexible** - Le modèle de données orienté document de MarkLogic est indifférent au schéma et peut gérer la complexité des données bitemporelles. Par contre, pour diverses raisons (contraintes d'intégrité, évolution des schémas et grand nombre de modèles de données différents), les bases de données relationnelles s'avèrent peu adaptées à ce type de problématique.
- **Fiabilité d'entreprise** - MarkLogic dispose des fonctionnalités d'entreprise dont les autres bases de données de nouvelle génération sont dépourvues. MarkLogic est une base de données éprouvée qui exécute des applications critiques au sein de centaines d'entreprises mondiales de premier plan.
- **Bitemporalité immédiatement exploitable** - La bitemporalité est une fonction intégrée à la solution MarkLogic, tandis que les autres fournisseurs imposent le déploiement d'un logiciel supplémentaire, ce qui augmente le coût et la complexité du processus.

LE COÛT DE TRAVAILLER SANS BITEMPORALITÉ

Le fait de ne pas utiliser la bitemporalité coûte directement 25 millions de dollars à une entreprise.¹ Cela a coûté (ou peut-être sauvé) leur poste à de nombreux politiciens. Dans notre ère hyper-réglémentée, et compte tenu de la nécessité d'assurer la provenance, l'immuabilité et la gouvernance des données historiques, le coût potentiel de ne pas utiliser la bitemporalité prend des proportions considérables. Cela est particulièrement vrai dans les secteurs tels que les services financiers, où le fait de ne pas avoir une image exacte du passé s'est traduit par des amendes atteignant plusieurs milliards de dollars et par la mise en place de nouveaux règlements.

¹ Hudson Foods a rappelé un cinquième de sa production annuelle en 1997 en raison d'une infection de bactéries E.coli, qui s'est chiffrée à environ 25 millions de dollars. Sa base de données lui permettait d'accéder uniquement à une vue actualisée indiquant la source des animaux incriminés, et non pas une vue de ses données le jour où le fournisseur a traité le petit lot de viande contaminée. En conséquence, l'entreprise a dû se résoudre à rappeler tout le produit. Pour en savoir plus, lisez le livre de Richard T. Snodgrass, « *Developing Time-Oriented Database Applications in SQL* » (ch.2, 11).

TROIS TYPES DE TEMPORALITÉ

Pour comprendre le concept de bitemporalité, il faut en premier lieu comprendre comment les bases de données gèrent actuellement le temps. Il existe trois catégories élémentaires de bases de données : non-temporelles, bitemporelles et unitemporelles. Nous allons présenter ces différents types, en utilisant les exemples du moment où a été effectué le diagnostic d'allergie d'un patient, et du moment où le médecin en a été informé.

NON-TEMPORELLES

Les bases de données non-temporelles stockent des données sans aucune dimension de temps. Un fait n'est qu'un fait, sans aucun historique, et n'est considéré vrai qu'au moment présent. Les modèles de données qui ne prennent pas en charge une dimension temporelle sont appelés des instantanés.



FIGURE 1 : Une base de données non-temporelle n'enregistre pas les dimensions de temps.

Imaginez l'exemple du moment où l'allergie d'un patient a été diagnostiquée, élément d'information important tenant compte des réactions éventuellement néfastes, voire mortelles, de certains patients à des médicaments communs, la pénicilline par exemple. Avec une base de données non-temporelle, vous verriez seulement l'état actuel, qui serait soit « le patient n'a aucune allergie » ou « le patient est diagnostiqué allergique », comme illustré à la figure 1, dans laquelle la zone ombrée représente les moments où le fait est vrai.

Dans une base de données non-temporelle, vous obtenez une vue unique des données ne tenant pas compte du temps. Rien de surprenant à ce que les bases de données non-temporelles soient très rares, puisque la plupart des applications traitent des données qui varient en fonction du temps.

UNITEMPORELLES

Les bases de données unitemporelles prennent en charge le temps sur une seule dimension : *temps valide*. Pour la majorité des gens, « le temps valide » veut tout simplement dire « le temps », soit le moment où un événement est arrivé dans le monde réel. Le temps valide est suivi sur une chronologie unique pour répondre à des questions telles que : Quand l'allergie du patient a-t-elle été diagnostiquée ? Combien de patients souffrent de la même allergie ? Depuis



FIGURE 2 : Une base de données unitemporelle focalise son suivi sur le temps valide.

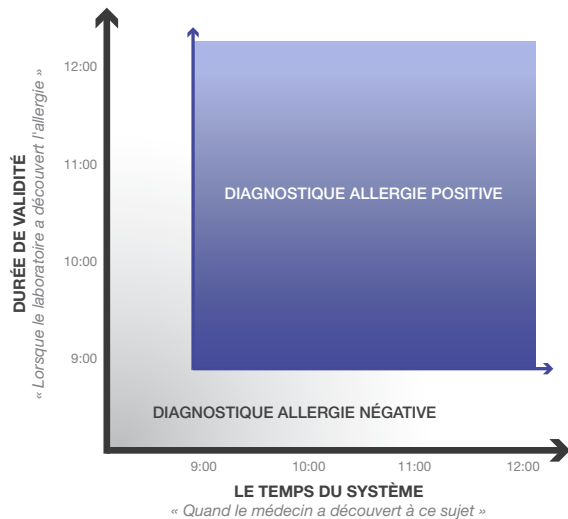
combien de temps le patient souffre-t-il de l'allergie ? Dans l'exemple du patient souffrant d'allergie, il est clair d'après le graphique de la figure 2 que son allergie a été diagnostiquée à 9h00 sur la chronologie valide.

Cependant, le temps valide ne présente qu'une image partielle. En examinant la figure ci-dessus, un observateur extérieur ne pourra pas savoir clairement quand le médecin a appris que l'allergie du patient a été diagnostiquée. Quelle serait la situation si le laboratoire avait été le premier à identifier l'allergie, mais qu'il y avait eu un décalage entre ce moment et celui où le médecin a pris connaissance des résultats du laboratoire ? Il s'agit d'une information précieuse qui n'est pas consignée dans une base de données unitemporelle. Dans cet exemple, imaginez qu'un médicament administré au patient ce jour-là ait provoqué une réaction anaphylactique. *Le médecin savait-il qu'il ne fallait pas administrer ce médicament ?* Voyons comment résoudre ce problème avec une base de données bitemporelle.

BITEMPORELLES

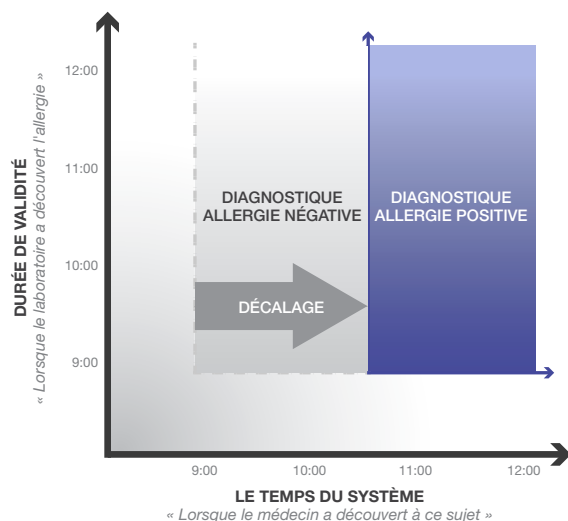
Une base de données bitemporelle enregistre des horodatages pour obtenir des événements sur deux dimensions de temps : *temps valide* et *temps système*. Le temps valide correspond au moment où un événement s'est produit dans le monde réel. Le temps système (parfois appelé « temps transaction ») correspond au moment où l'événement est enregistré dans la base de données. Ces deux dimensions de temps sont représentées graphiquement sur les deux axes dans la figure 3. Dans cet exemple, le temps valide représente le moment où le laboratoire a identifié l'allergie, et le temps système représente le moment où le médecin a pris connaissance de l'information et l'a consignée dans son dossier.

Les bases de données unitemporelles partent de l'hypothèse erronée que le temps valide est toujours égal au temps système, et font ainsi l'abstraction d'informations précieuses. Dans certains cas, comme l'illustre la figure 3, le temps valide est égal au temps système. Mais, la seule manière de le savoir est de disposer d'une base de données bitemporelle. Une base de données bitemporelle enregistre le temps séparément sur deux dimensions, pour vous permettre d'avoir en main des informations précises.



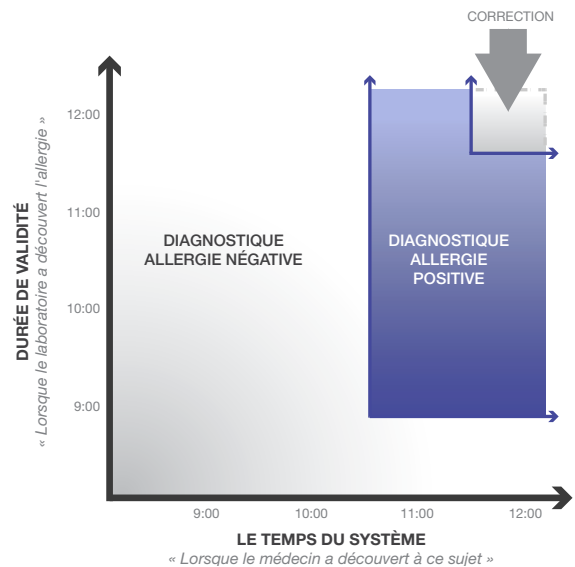
En utilisant l'exemple du patient souffrant d'allergie, imaginez que le médecin ait pris connaissance de l'allergie à 10h30, soit une heure et demie après que le laboratoire ait effectué les tests et identifié que le patient souffrait d'une allergie. Le laboratoire a noté que le patient souffrait d'une allergie à 9h00, mais cette information n'est pas parvenue au médecin avant 10h30. Cela représente un décalage entre le temps système et le temps valide, comme dans la figure 4.

FIGURE 4 : Une base de données bitemporelle suit les décalages d'information.



En allant plus loin avec cet exemple, imaginez que plus tard le même jour, à 11h30, le médecin reçoit un appel d'un agent du laboratoire l'informant qu'il vient de se rendre compte que les résultats des tests sont incorrects. Le résultat du laboratoire était en fait négatif : le patient n'a pas d'allergie. Cette correction est illustrée à la figure 5. Avec une base de données bitemporelle, il est facile d'apporter des corrections à des données historiques sans écraser les données antérieures.

FIGURE 5 : Une base de données bitemporelle suit les corrections sans écraser les données.



En examinant la figure 5, nous pouvons déterminer les faits suivants :

- **Avant 10h30 (temps système)**, le médecin n'avait pas connaissance de l'allergie
- **À 10h30 (temps système)**, le médecin inscrit dans le dossier du patient qu'il ou elle souffre d'une allergie, identifiée par le laboratoire à 9h00 (temps valide)
- **À 11h30 (temps système)**, le laboratoire et le médecin découvrent l'erreur et mettent à jour les dossiers en indiquant que le patient ne souffre pas d'une allergie

Grâce à cette chronologie suivie sur deux axes, il est maintenant possible d'obtenir une image fidèle des événements. Cela peut être extrêmement utile pour comprendre et éviter les erreurs, car les décisions du médecin peuvent facilement être reliées aux informations dont il disposait et dont il ne disposait pas à tout moment donné dans le temps. Dans un cadre hospitalier, les allergies médicamenteuses peuvent être mortelles, et il est donc essentiel d'avoir des informations précises sur le moment où le patient a été diagnostiqué et sur le moment où les soignants ont pris connaissance du diagnostic.

L'exemple du diagnostic d'allergie peut sembler par trop simpliste, mais le même concept peut être appliqué à n'importe quel type de données, qu'il s'agisse de l'occurrence d'une transaction financière, de la souscription à une assurance, ou de l'achat d'une maison. Dans tous ces cas, la DATE DE DÉBUT et la DATE DE FIN pour le temps valide et le temps système peuvent être suivies dans le but d'obtenir l'image la plus fidèle de la réalité.

TABLEAU 1 : Comparaison des bases de données unitemporelles et bitemporelles pour divers exemples.

UNITEMPORELLES	BITEMPORELLES
Quand les résultats de laboratoire ont-ils indiqué que le patient souffrait d'une allergie à la pénicilline ?	Quand les résultats de laboratoire ont-ils indiqué que le patient souffrait d'une allergie à la pénicilline, et quand le soignant a-t-il été informé de l'allergie ?
Quand l'ordre de vente a-t-il été annulé par la contrepartie de la banque ?	Quand l'ordre de vente a-t-il été annulé par la contrepartie de la banque, et quand l'opérateur boursier a-t-il appris cette annulation ?
Quelles étaient les données de référence concernant les événements de transaction du 4 décembre ?	De quelles données de référence l'opérateur boursier disposait-il en réalité le 4 décembre ?
Quand Jean est-il devenu éligible pour une couverture d'assurance, comme l'indique maintenant son dossier ?	Quand Jean est-il devenu éligible pour une couverture d'assurance, comme l'indiquait son dossier en 2012 ?

LES AVANTAGES DE LA BITEMPORALITÉ

En deux mots, la bitemporalité améliore votre gestion du temps. La bitemporalité est la seule solution capable de fournir une vue homogène et flexible des données historiques pouvant faire l'objet de recherches. Même le suivi de version temporel ne produit pas cette qualité de résultats. La bitemporalité est une capacité essentielle dont toute entreprise peut bénéficier, et pour laquelle il existe un besoin croissant, surtout dans les industries devant faire face à une augmentation des pressions réglementaires et du nombre d'actions en justice, à savoir les services financiers, les assurances et la santé. Dans ces secteurs, les entreprises doivent tenir un registre plus précis de la totalité de leurs actions passées, dans ce nouveau contexte : nouvelles lois et nouvelles possibilités d'actions en justice, audits plus fréquents et plus approfondis, et amendes plus lourdes en cas de non-conformité. Les entreprises qui gèrent mieux leurs données historiques sont en mesure de réduire leurs risques et d'aborder sereinement leurs audits.

LES ACTIONS QUE VOUS PERMET LA BITEMPORALITÉ

- **Gérer les réglementations et les audits** - Produisez une image précise du passé pour répondre aux exigences accrues de transparence et de responsabilité
- **Gérer les risques** - Créez de meilleurs modèles de risque et améliorez l'informatique décisionnelle en analysant de vraies données historiques
- **Réduire les coûts** - Simplifiez l'architecture et réduisez le coût et le risque opérationnel que représente le stockage redondant des données historiques

LE BESOIN CROISSANT DE BITEMPORALITÉ

Il devient globalement de plus en plus important de mieux gérer les enjeux réglementaires, même si l'impact de cet impératif touche particulièrement certains secteurs, notamment les services financiers. Les grandes banques ont été pénalisées par des amendes record au cours des dernières années, et subissent des pressions accrues liées aux réglementations. Depuis 2009, les banques aux États-Unis et en Europe ont versé plus de 128 milliards de dollars aux organismes de réglementation, et 2014 a été une année record, avec 65 milliards de dollars de pénalités et d'amendes, soit environ 40 % de plus que 2013.²

À l'heure actuelle, les organismes de réglementation sont plus intrusifs, ils appliquent les lois plus rigoureusement et réalisent des analyses d'entreprises en entrant vraiment dans les détails. Selon Gerold Grasshoff, responsable international de la gestion des risques et de la réglementation pour le Boston Consulting Group, les pressions réglementaires représentent désormais un enjeu majeur pour les banques. « Il faut dès maintenant modifier votre modèle d'exploitation, changer vos produits, et convertir les risques juridiques... Les exigences réglementaires ont vraiment le pouvoir de chambouler les modèles opérationnels. C'est là le plus grand défi stratégique. » Pour s'adapter à l'évolution des modes d'exploitation, les banques sont obligées de modifier leurs approches de la gestion de l'informatique et des données, dans un souci de transparence accrue.³

D'autres secteurs sont également confrontés à un renforcement des pressions réglementaires. Dans le domaine de la santé, par exemple, citons le problème lié aux erreurs médicales, que certains rapports quantifient aux alentours d'un milliard de dollars.⁴ Il est essentiel de savoir quand et comment des erreurs se sont produites,

² James Sterngold, « For Banks, 2014 Was a Year of Big Penalties », 30 décembre 2014 <<http://www.wsj.com/articles/no-more-regulatory-nice-guy-for-banks-1419957394>>

³ Boston Consulting Group, « Building the Transparent Bank », décembre 2014 <https://www.bcgperspectives.com/Images/Building_the_Transparent_Bank_Dec_2014_tcm80-177814.pdf>

⁴ Andel, Davidow, Hollander, Moreno. « The economics of health care quality and medical errors. » Journal of Health Care Finance 39(1):39-50 (2012) <<http://www.ncbi.nlm.nih.gov/pubmed/23155743>>

« Nous sommes à une époque où les lois sont appliquées très vigoureusement, et dans un contexte super-réglementé. Cette situation est là pour durer. »

Benjamin Lawsky,
Surintendant des services financiers, État de New York⁶

dans le but d'améliorer la prise de décisions médicales et d'éviter les fautes professionnelles. Il convient également d'envisager le coût croissant de la fraude et des abus dans le secteur de la santé, représentant entre 82 et 272 milliards de dollars aux États-Unis.⁵ Malheureusement, le coût général et la complexité des aspects tels que la sécurité des patients, les actions en négligence professionnelle ainsi que les fraudes et abus ne cessent qu'augmenter.

En mettant en œuvre la gestion des données bitemporelle, les entreprises peuvent prendre une initiative audacieuse en vue de diminuer le risque, d'améliorer la transparence et d'obtenir un avantage concurrentiel afin de prendre l'ascendant sur la concurrence.

LA BITEMPORALITÉ AU SEIN DES INDUSTRIES

SERVICES FINANCIERS

La bitemporalité aide les grandes banques à mieux gérer leurs données et à s'adapter aux changements des lois et réglementations ayant une incidence sur les modes d'opération. Par exemple, la bitemporalité étaye les structures en offrant un enregistrement précis des transactions à mesure qu'elles se produisent et sont modifiées. Une fois les transactions effectuées, elles font l'objet d'un rapprochement avec des contreparties, et des mises à jour sont souvent à intégrer avant la fin du jour de cotation. Avec une base de données unitemporelle, les mises à jour écrasent les données historiques, ce qui peut causer des risques considérables pour des opérateurs boursiers individuels et des entreprises entières. La bitemporalité donne une image exacte de l'ensemble du cycle de vie d'une analyse boursière, notamment en cas de changement des noms des contreparties ou des identifiants de transaction, ou en cas de correction de prix.

ASSURANCES

Dans le secteur des assurances, la bitemporalité permet de déterminer clairement la couverture au fil du temps. En conséquence, vous êtes certains que les données ne sont jamais écrasées, même en cas de changements rétroactifs.

TABLEAU 2 : La bitemporalité dans les services financiers

AVANT LA BITEMPORALITÉ	APRÈS LA BITEMPORALITÉ
Quelle était la position de l'opérateur boursier, et quelles informations étaient à sa disposition au moment où la transaction a été exécutée ?	Quelle était la position exacte de l'opérateur boursier lorsque la transaction a été exécutée, et quelles références exactes étaient disponibles au moment où la transaction a été exécutée ?
Quelle était la cote de solvabilité de notre client l'an dernier ?	Quelle était la cote de solvabilité de notre client l'an dernier, selon les informations dont nous disposons au trimestre dernier ?
Quelle était l'exposition de notre marché au moment de la transaction à 11h00 ?	Quelle était l'exposition de notre marché au moment de la transaction à 11h00, tel que nous en étions informés à 11h30 ?
Quel était le bénéfice de l'entreprise quand nous avons communiqué nos recommandations ?	Quelles étaient nos estimations du bénéfice de l'entreprise au moment où nous avons communiqué nos recommandations ?

TABLEAU 3 : La bitemporalité dans le secteur des assurances

AVANT LA BITEMPORALITÉ	APRÈS LA BITEMPORALITÉ
Quelle était l'incidence estimée du sinistre sur les primes d'assurance ?	Quelle était l'incidence estimée du sinistre sur les primes d'assurance, avant le réajustement rétroactif des données ?
Le bénéficiaire disposait-il d'une couverture au moment du diagnostic ?	Le bénéficiaire disposait-il d'une couverture au moment du diagnostic, avant la ratification du projet de loi ?
L'employé travaillait-il pour l'entreprise lorsque l'événement s'est produit ?	L'employé travaillait-il pour l'entreprise lorsque l'événement s'est produit, tel que l'indiquait votre dossier à ce moment-là ?

5 Berwick, Hackbarth. « Eliminating waste in US health care. » JAMA 307(14):1513-6 (2012) <<http://www.ncbi.nlm.nih.gov/pubmed/22419800>>

6 James Sterngold. « For Banks, 2014 Was a Year of Big Penalties. » Wall Street Journal, 2014.



La bitemporalité de MarkLogic offre la flexibilité de mise en corrélation et de création de valeur supplémentaire rapide pour les données (en fournissant des informations en cours de journée, et pas seulement en fin de journée) à un groupe de clients diversifié, capacité qui n'était pas pleinement concrétisée auparavant. En fait, la bitemporalité MarkLogic offrira la toute nouvelle possibilité pour nos clients d'effectuer d'autres analyses et de mettre en œuvre des capacités beaucoup plus riches dans le domaine de la gestion de la conformité. »

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

La compagnie d'assurance peut toujours avoir un historique de la couverture d'assurance à n'importe quel moment dans le passé. Un assureur peut aussi vouloir connaître le statut d'un employé et avoir besoin d'une image précise du moment où l'employé travaillait effectivement pour une entreprise à tout moment, tel que ces informations étaient disponibles à tout moment dans le temps.

SANTÉ

Le secteur de la santé est confronté à de considérables défis pour tous les intervenants, notamment les prestataires, les instances de remboursement ainsi que les sociétés pharmaceutiques et biotechnologiques. La bitemporalité est un composant des améliorations dans le domaine de l'informatique médicale permettant de réduire les coûts et d'améliorer les résultats en procurant aux prestataires une image plus précise de l'historique d'un patient, sachant que différentes équipes dirigent le traitement, et un outil d'enquête optimisé pour analyser les événements indésirables. De plus, quand les instances de remboursement reçoivent les codes de facturation pour les procédures, elles sont en mesure d'avoir un suivi de l'historique complet pour chaque patient. Même si des modifications de la couverture d'assurance ont été effectuées de façon rétroactive, aucune partie de l'historique n'est perdue. Les compagnies pharmaceutiques et biotechnologiques bénéficient également d'avantages, car elles sont en mesure d'utiliser la bitemporalité pour améliorer la prise de décisions concernant à la fois la recherche et les aspects purement commerciaux.

DROIT ET INFORMATIQUE DECISIONNELLE

La bitemporalité permet de broser un tableau exhaustif, même si des faits disparates sont collectés de manière fragmentée avant et après certains événements. Les organismes gouvernementaux ont accès à une image plus complète, ce qui leur permet de mieux comprendre les motivations, voire de prévoir les événements futurs. Grâce à la bitemporalité, les agents d'application de la loi chargés d'audits sont en mesure de demander pourquoi vous avez opté pour une stratégie spécifique. Cette capacité est particulièrement utile lorsque des audits sont relancés à partir de dossiers non résolus.

TABLEAU 4 : La bitemporalité dans le secteur de la santé

AVANT LA BITEMPORALITÉ	APRÈS LA BITEMPORALITÉ
À quoi ressemblait le dossier du patient lorsque le médicament a été prescrit ?	À quoi ressemblait le dossier du patient lorsque le médicament a été prescrit, avant que les résultats du laboratoire y aient été intégrés ?
Quelle était la détermination de couverture pour ce patient en juin 2010 ?	Quelle était la détermination de couverture pour ce patient en juin 2010, tel que nous la connaissions en août 2010 ?
Qu'indiquaient les résultats d'essai clinique lorsque vous avez effectué l'investissement supplémentaire ?	Qu'indiquaient les résultats d'essai clinique lorsque vous avez effectué l'investissement supplémentaire, avant la mise à jour des résultats de recherche ?

TABLEAU 5 : La bitemporalité dans le droit et l'informatique décisionnelle

AVANT LA BITEMPORALITÉ	APRÈS LA BITEMPORALITÉ
Quel était le contexte lorsque nous avons pris la décision ?	Quelle était notre perspective du contexte lorsque nous avons pris la décision ?
Quand l'événement s'est-il produit ?	Quand l'événement s'est-il produit, et quand a-t-il été consigné ?
Pourquoi pensons-nous actuellement avoir appliqué cette stratégie ?	Que pensons-nous quand nous avons appliqué cette stratégie ?

« Les offres de bitemporalité relationnelle ne sont pas adoptées sur une grande échelle car la forme des données évolue au fil du temps et les systèmes de gestion des bases de données relationnelles ne sont pas en mesure de prendre en charge cette évolution du schéma. »

Banque d'investissement internationale

POURQUOI L'IMPLEMENTATION DE LA BITEMPORALITÉ A ÉTÉ DIFFICILE

À ce stade, vous vous posez probablement la question : « Si la bitemporalité est si importante, pourquoi n'en ai-je jamais entendu parler ? » Malgré la publication de milliers d'articles de recherche sur les données temporelles au cours des vingt dernières années, et bien que des thèses d'experts portent sur ce thème depuis le début des années 1990, la bitemporalité demeure relativement peu connue.

La bitemporalité a une valeur considérable pour le monde des affaires. Pourtant, la plupart des analystes d'entreprise ne savent même pas qu'ils peuvent demander des données bitemporelles, car elles sont rarement produites. En effet, la complexité de la mise en œuvre et de l'exploitation des données bitemporelles dans des bases de données relationnelles l'emportent souvent sur les avantages. En fait, le traitement de données temporelles ordinaires dans une base de données relationnelle peut à lui seul représenter un défi majeur.

Malheureusement, malgré les efforts pour rendre les données bitemporelles plus faciles à gérer dans les bases de données relationnelles, la bitemporalité demeure un objectif impossible à atteindre en utilisant des outils traditionnels. Le nombre d'experts dans le monde capables de gérer les complexités inhérentes aux implémentations bitemporelles avec des bases de données relationnelles est probablement limité à un cercle très restreint de spécialistes. Sans trop entrer dans les détails de la modélisation de données bitemporelles, nous présentons certaines des principales raisons pour lesquelles les bases de données relationnelles conviennent mal à la gestion des données bitemporelles :

- **Évolution du schéma** - L'ajout de données bitemporelles à un modèle relationnel s'avère d'une extrême complexité. Les modifications architecturales et structurelles sont elles-mêmes temporelles, et lorsque de nouvelles colonnes sont ajoutées avec des dimensions temporelles, ou de nouveaux tableaux sont créés dès l'ingestion de nouvelles données, le schéma s'en trouve modifié. La gestion d'un changement de schéma et les modifications qui en résultent dans le code d'application représentent déjà des projets complexes, même avant d'essayer d'y ajouter des données bitemporelles.
 - **De multiples modèles de données** - La gestion de l'évolution du schéma constitue un important défi, mais imaginez maintenant la tâche que représente le traitement de multiples schémas en évolution sur plusieurs modèles de données et silos de données, puis leur regroupement en une source unique de vérité. L'intégration des données est en soi une tâche coûteuse, mais en y ajoutant des données, la complexité s'accroît de façon exponentielle.
 - **Baisse de performance** - Généralement, les performances de lecture et d'écriture diminuent nettement car les requêtes bitemporelles doivent considérer l'axe supplémentaire du temps pour chaque requête, et les données se trouvent généralement sur plusieurs tableaux voire sur plusieurs serveurs. Des tentatives de simplification des requêtes et d'amélioration des performances ont été engagées, mais elles n'ont pas été suffisamment loin et n'ont pas réussi à éliminer la complexité inhérente et les problèmes de performances provoqués par la présence de données bitemporelles dans plusieurs tableaux.
 - **Dépendance vis-à-vis d'un fournisseur unique** - Certains fournisseurs ont commencé à mettre en œuvre des améliorations appliquées à la bitemporalité. Toutefois, comme cela s'est produit par le passé avec la mise en place des normes SQL, chaque fournisseur les met en œuvre différemment en appliquant sa propre syntaxe, et ajoute à son tarif le coût de cette fonction.
- Souvent, la réponse aux défis de mise en œuvre de la bitemporalité dans une base de données relationnelle consiste à trouver une solution alternative appropriée. Voici certaines des réponses fréquentes.

« Malgré la quasi-universalité du temps et la nature fluctuante de l'entreprise en cours de modélisation (une configuration statique et rigide rare et inintéressante), force est de constater que SQL n'est pas à la hauteur quand il s'agit de prendre en charge des aspects changeants au fil du temps, ou de fournir des constructions pour modéliser, interroger ou modifier efficacement ces informations. »

Richard Snodgrass, *Developing Time-Oriented Database Applications in SQL*⁸

- « **Il me suffit d'utiliser des dimensions à évolution lente** » - L'utilisation des dimensions à évolution lente issues de la modélisation dimensionnelle de « type 2 » afin d'obtenir une approximation des données bitemporelles a été mise à l'épreuve au cours des dernières années, et les problèmes liés à cette approche sont clairement documentés.⁷ L'utilisation des dimensions à évolution lente se borne à donner une approximation des données temporelles valides et se traduit par de nombreuses incohérences qui sont ensuite difficiles à identifier et encore plus à corriger. D'ailleurs, même si la conception tient la route, les performances des requêtes resteront probablement lentes et les résultats risquent de ne pas être reproductibles.
- « **Je peux prendre des instantanés à intervalles réguliers** » - Cette approche, également appelée « versionnement temporel », est l'un des arguments les plus fréquents contre la bitemporalité, car la plupart des entreprises prennent déjà régulièrement des instantanés hebdomadaires ou mensuels de leurs données. Cette approche est stable et prévisible. Malheureusement, elle génère des quantités massives de données redondantes, des coûts de stockage importants, et encore beaucoup de pertes d'informations en raison des écarts entre les instantanés. Même si la prise d'instantanés s'effectue fréquemment, les organismes de réglementation dans la plupart des secteurs rechignent de plus en plus à accepter cette approche. Les organismes de réglementation et les analystes de données ont des questions spécifiques, exigent des réponses rapides, et n'apprécient guère les écarts.

- « **Je peux compter sur mes journaux d'audit** » - Bien qu'utiles pour un suivi des informations d'événement, les journaux ne suffisent pas pour la bitemporalité parce qu'ils ne peuvent pas être interrogés facilement et rapidement, et ne répondent pas aux normes de conservation d'archives immuables.

La bitemporalité est la seule approche de gestion du temps fournissant un moyen transparent et rapide d'analyser les données historiques, de les interroger à la volée à n'importe quel moment, et de les travailler sur un plan opérationnel, tout comme vos données les plus récentes.

POURQUOI LA BITEMPORALITÉ EST ADAPTÉE AU CONTEXTE ACTUEL

En tant que base de données d'entreprise NoSQL, MarkLogic offre la souplesse nécessaire pour faire du stockage et de la gestion de données bitemporelles une réalité pratique, sans aucun compromis en termes de performances en cas de requêtes complexes, ou en termes de résilience et de sécurité. MarkLogic est également unique, dans le sens où il s'agit de la seule base de données d'entreprise NoSQL offrant des capacités de bitemporalité.

MarkLogic ne reconnaît aucun schéma spécifique, et gère les données en tant que documents. En d'autres termes, il n'est pas nécessaire de maintenir un schéma strict devant être respecté pendant toute la durée de vie de la base de données. Si vous devez intégrer une nouvelle source de données à une date ultérieure, vous n'avez pas à réaliser de processus ETL complexe avant de charger ces données dans MarkLogic. Finie la frustration d'avoir à ajouter une

Avantages de la bitemporalité dans MarkLogic	MarkLogic	Autres BD
Indifférence au schéma pour gérer l'évolution des schémas et la diversité des modèles de données	✓	✗
Codage et exploitation simplifiés	✓	✗
Délai de rentabilité raccourci	✓	✗
Évolutivité, élasticité, et coûts de stockage réduits	✓	✗

⁷ Tom Johnston. *Bitemporal Data: Theory and Practice* (Waltham, MA: Elsevier, 2014) 311 - 313.

⁸ Richard T. Snodgrass. *Developing Time-Oriented Database Applications in SQL*. Morgan Kaufmann Publishers, Inc., San Francisco, Juillet 1999. <<http://www.cs.arizona.edu/~rts/tdbbook.pdf>>

nouvelle colonne dans une base de données relationnelle, que vous ajoutiez une colonne DATE ou quoi que ce soit d'autre.

Les données bitemporelles peuvent avoir une durée de vie de plusieurs décennies, et les entreprises ont besoin d'une base de données capable de réagir rapidement afin de suivre le rythme d'évolution du schéma à mesure que de nouvelles sources de données sont ajoutées. MarkLogic facilite l'ingestion de nouvelles sources de données, et en cas de conflit (par ex., une nouvelle source de données avec comme nom de colonne « SRC_DATE » alors qu'il devrait être « CLAIM_DATE »), MarkLogic permet d'effectuer plus facilement les transformations nécessaires pour standardiser le vocabulaire. Avec MarkLogic, vous n'aurez jamais à vous soucier des contraintes liées à la modélisation des données relationnelles, notamment l'intégrité des entités, l'intégrité référentielle et la dénormalisation, même dans un contexte de gestion bitemporelle des données.

MarkLogic exécute des ordres de grandeur mieux que les bases de données relationnelles pour les grands projets d'intégration de données, en accélérant les délais de livraison des projets, en réduisant le temps consacré à la collecte des exigences et à la modélisation des données, et en améliorant la qualité des prototypes. Les responsables de Broadridge, une grande entreprise de services financiers, ont fait la remarque suivante : « Le premier projet MarkLogic a été réalisé en 60 jours, alors qu'avec la technologie existante, selon les estimations, ce processus aurait nécessité 3 000 jours ».

COMMENT FONCTIONNE LA BITEMPORALITÉ DANS MARKLOGIC

Pour les entreprises maîtrisant les bases de données relationnelles, l'utilisation des données temporelles et bitemporelles dans MarkLogic ne devrait rien avoir d'insolite. La principale différence réside dans le fait qu'au

lieu d'avoir des colonnes de dates dans un tableau, cette information apparaît maintenant sous forme d'horodatage dans les documents. MarkLogic enregistre et gère toutes les données en tant que documents, y compris les données bitemporelles.

Un document JSON ou XML est considéré bitemporel s'il comprend des horodatages pour le temps de début et de fin valide, et pour le temps de début et de fin système. Pour charger un document bitemporel dans MarkLogic, l'une des méthodes consiste à utiliser MarkLogic Content Pump, ou mlcp. Vous pouvez également utiliser l'API REST. Vous pouvez aussi charger un document bitemporel à l'aide d'une simple requête de mise à jour JavaScript, tel qu'indiqué ci-dessous.

Après le chargement de documents bitemporels dans MarkLogic, ceux-ci sont gérés comme une série de documents auxquels sont associés des indices de plages d'éléments, sur des axes de temps valide et système. Les axes de temps valide et système font office de conteneur pour une paire d'indices de plage nommée. De plus, les documents bitemporels sont stockés dans des collections temporelles, qui sont des regroupements logiques de documents temporels. Vous pouvez créer d'autres collections temporelles si vous avez des documents nécessitant un schéma différent pour les horodatages.

Après le chargement des documents initiaux dans MarkLogic, ceux-ci sont toujours conservés et jamais modifiés. Même si un document bitemporel est « supprimé », MarkLogic conserve le document, mais le temps système passe de l'infini au moment de la suppression. Le même processus s'applique aux mises à jour : les anciennes versions sont conservées et la « nouvelle » version est simplement ajoutée. MarkLogic ne permet pas non plus les mises à jour du temps de démarrage système. Une fois le temps système défini pour une collection, il continue à progresser pour mieux assurer l'intégrité des données.

FIGURE 6. Mise à jour d'un document bitemporel

```
DeclareUpdate();

var root =
{ "tempdoc" : {
  "systemStart": null,
  "systemEnd": null
  "validStart": "2014-04-03T11:00:00",
  "validEnd": "2014-04-03T16:00:00",
  "content": "certaines données, comme le prix de clôture"
}
};

temporal.documentInsert("temporalCollection", "exampledata.json", root);
```

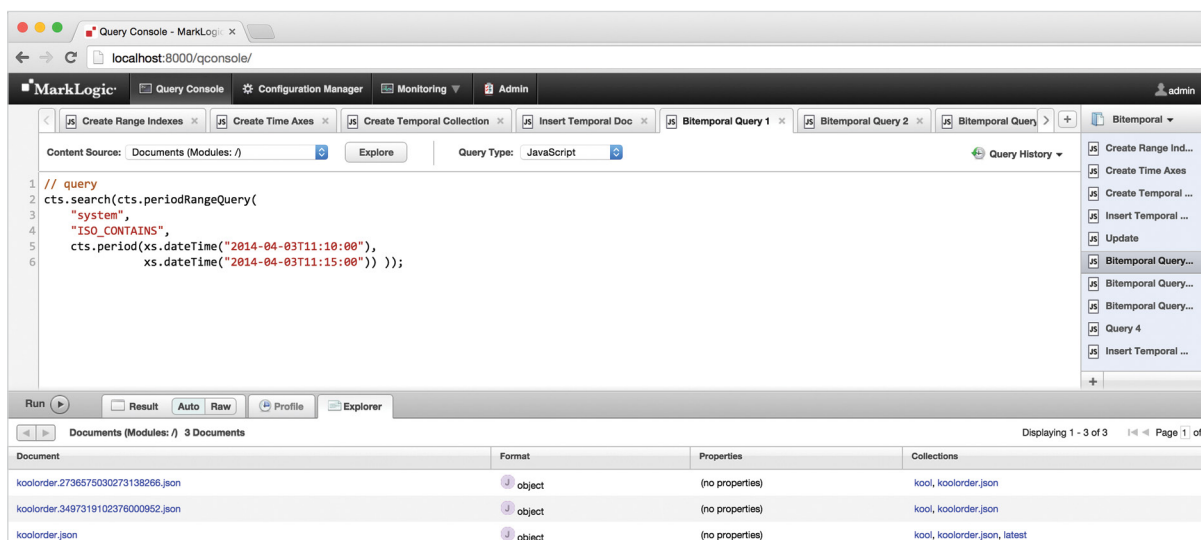


FIGURE 7. Une requête bitemporelle affichée dans la console de requête MarkLogic

Il est essentiel que le suivi de la provenance des informations s'effectue avec une gouvernance et une immuabilité totales, c'est pourquoi MarkLogic applique son modèle de sécurité aux documents bitemporels. MarkLogic est certifié par le « Common Criteria Evaluation and Validation Scheme » du « National Information Assurance Partnership » (NIAP), et utilise par défaut le contrôle d'accès basé sur les rôles (Role Based Access Control, RBAC) pour gérer l'accès aux documents. Grâce à ce niveau élevé de sécurité, les enregistrements historiques ne sont pas altérés, et les documents conservent leurs autorisations au fil du temps.

FONCTIONNALITÉS CLÉS DE BITEMPORALITÉ DANS MARKLOGIC

Insertion, mise à jour (jamais suppression) - Incorporez les documents temporels XML ou JSON avec des références à la chronologie valide à l'aide de l'API temporel ou de mlcp, et apportez des changements sans perdre aucune donnée lors de l'ajout des nouvelles versions

Requêtes temporelles complexes - Interrogez la base de données le long d'axes de temps valide et système, en utilisant les opérateurs Allen et SQL, lors des comparaisons de périodes chronologiques

Adaptation aux schémas évolutifs - Épargnez-vous les inquiétudes relatives à l'évolution de la forme du contenu au fil du temps. Contrairement aux bases de données relationnelles, MarkLogic est indépendant des données et peut gérer aisément les changements de schéma au fil du temps

Conservation du dernier temps de requête stable - Un horodatage spécial, appelé LSQT (Last Stable Query Time, dernier temps de requête stable), peut être activé afin de gérer et de coordonner les temps de début système au sein des systèmes

Association du stockage hiérarchisé – Tirez parti du stockage hiérarchisé afin de migrer aisément les données d'historique sur des niveaux de stockage moins coûteux, sans perdre votre capacité à interroger les données

Combinaison avec la sémantique - Attribuez des éléments bitemporels à des documents, qu'il s'agisse de triplets RDF ou de documents comprenant des triplets RDF, ce qui vous permet de suivre le changement des relations au fil du temps

Combinaison avec le géospatial - Obtenez la possibilité de suivre vos données dans un contexte de temps et d'espace. MarkLogic enregistre des données géospatiales, ce qui vous permet désormais de suivre précisément les changements des données géospatiales au fil du temps

Sécurité certifiée - Gérez les documents bitemporels avec la même sécurité certifiée que tous les autres documents, avec le contrôle d'accès basé sur les rôles (Role Based Access Control, RBAC) ou autres modèles de sécurité

Mise à l'échelle rapide et facile - Évitez toute inquiétude de sous-dimensionnement avec l'architecture MarkLogic à évolutivité horizontale, qui vous permet d'ajouter facilement des nœuds pour répondre à l'accroissement de la demande de données bitemporelles

« MarkLogic apporte sur le marché une technologie avancée de gestion des données, et bon nombre de ses clients et partenaires sont habitués à gérer des données complexes avec une grande agilité. En conséquence, les clients et partenaires de MarkLogic ont généralement une perspective plus mature et créative de la gestion et de l'utilisation des données que la plupart des autres utilisateurs de bases de données. »

Carl Olofson, Vice-président chargé de la recherche en logiciel de gestion des données de recherche, IDC

COMMENT DÉBUTER

La gestion du temps n'est pas facile. Si elle l'était, nous aurions probablement pu éviter des problèmes se chiffrant en milliards de dollars, notamment le passage à l'an 2000.⁹ Cela dit, la gestion du temps est une nécessité, et la bitemporalité est la solution d'avenir de la gestion du temps dans les bases de données alors que nous cherchons à maintenir un meilleur suivi de « ce que nous savions » et de « quand nous en avons pris connaissance ». MarkLogic, la meilleure base de données pour stocker et gérer les données bitemporelles, supprime les contraintes qui empêchent l'adoption de la bitemporalité.

MISE EN ROUTE RAPIDE

1. Identifier les questions auxquelles votre entreprise ne peut pas actuellement répondre
2. Identifier les avantages que présente l'ajout de la bitemporalité
3. Évaluer l'environnement actuel de gestion des données
4. Contacter MarkLogic pour parler de l'implémentation
5. Télécharger MarkLogic
6. Suivre la formation MarkLogic gratuite

PLUS D'INFORMATIONS

- **Lire la documentation MarkLogic** - Apprenez comment travailler avec les données bitemporelles dans MarkLogic sur docs.marklogic.com/guide/temporal
- **Regarder une présentation** - Écoutez un client de MarkLogic expliquer « Pourquoi les banques s'intéressent à la bitemporalité » www.marklogic.com/resources/why-banks-care-about-bitemporality/
- **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

⁹ Selon la BBC et ComputerWorld, le coût estimé de préparation et de correction du « passage informatique à l'an 2000 », ou Y2K, s'est élevé à 608 milliards de dollars, sans tenir compte de l'inflation. Pour plus d'informations : Robert L. Mitchell. « Y2K: The good, the bad and the crazy ». ComputerWorld (28 décembre 2009)
<<http://www.computerworld.com/article/2522197/it-management/y2k--the-good-the-bad-and-the-crazy.html?page=2>>