

# バイテンポラルで何が できるようになるのか

MARKLOGICホワイトペーパー ・ 2015年6月

規制違反により、数千億円規模の罰金や多大な時間とコストを費やす訴訟が起これる現代において、データベースは中心的な「SoR (System of Record)」の役割を担うべきです。しかし従来のデータベースでは、過去の履歴を完全に記録できません。過去を完全かつ正確に把握できるのは、バイテンポラル機能を持つデータベースだけです。これにより「何を知っていたのか」および「いつ知ったのか」がはっきりわかります。



## 現状の確認：バイテンポラルが必要？

先に進む前に、まず自分の組織において「バイテンポラル」な（＝2つの時間軸を持つ）データ管理が必要かどうか確認しておきましょう。次の質問に対して1つでも「はい」があるのであれば、バイテンポラルソリューションを検討した方が良いと考えられます。

	はい	いいえ
1. イベントやトランザクションの発生時点のトラッキングはビジネスにとって極めて重要である	✓	
2. 履歴データを後から更新する必要性が生じたことがある	✓	
3. 現実世界での事象の発生から、データベースへの記録までに時間が空いてしまったことがある	✓	
4. 規制当局から履歴データの提供を求められることがよくある	✓	
5. 情報を得た順番が重要である業界（法務、情報機関など）で働いている	✓	
6. 組織内における履歴データの格納やアクセスが複雑でコストが膨大にかかる	✓	
7. 履歴データの管理とアクセスのために、大量の開発が必要となっている。あるいは時間の経過とともにリスクが増大している	✓	

# 目次

<b>はじめに</b> .....	1
バイテンポラルがないことによるコスト	
<b>3種類のテンポラル</b> .....	2
ノンテンポラル	
ユニテンポラル	
バイテンポラル	
<b>バイテンポラルのメリット</b> .....	4
バイテンポラルでできること	
バイテンポラルのニーズの高まり	
さまざまな業界におけるバイテンポラル	
<b>これまでバイテンポラルの導入が困難だった理由</b> .....	7
<b>バイテンポラルに取り組むべき時代が到来した理由</b> .....	8
<b>MarkLogicにおけるバイテンポラルの処理</b> .....	9
MarkLogicのバイテンポラルの主要機能	
<b>始めてみましょう</b> .....	11
次のステップ	
さらに詳しく	

# はじめに

今日、主要な「SoR (System of Record)」となっているのはデータベースです。紙ではありません。この新しい現実において、あらゆるファクト（事実）を発生瞬間から正確に把握しておくことが求められています。また履歴（過去）データのトラッキングや管理の方法が法律で定められている業界もあります（金融、保険、ヘルスケアなど）。

しかし従来のデータベースでは、異なる時点における業務の状態をすべて正確に把握することは不可能です。これは従来のデータベースは「ユニテンポラル（時間軸が1つ）」で、開始時間と終了時間しか記録できないためです。しかし例えば、何かが発生してから、それを認識するまでに時間が空いている場合はどうしたらよいでしょうか。発生時間と認識した時間のどちらを記録すべきなのでしょう。あるいは、記録した発生時間が間違っていて後から修正しなければならないけれど、過去のデータを上書きしたくない場合はどうしたらよいでしょうか。こういった場合、時間軸が1つだけでは不十分です。

バイテンポラルデータベースでは、「有効時間」（現実世界でファクトが発生した時間。「何を知っていたのか」と「システム時間」（このファクトがデータベースに記録された時間。「いつ知ったのか」）の2つの時間軸のタイムスタンプを利用して、データの格納とクエリを行います。バイテンポラルデータベースでは2つの時間軸を使ってイベントをトラッキングすることで、特定時点の業務を完全かつ正確に把握できます。これにより社内での検索や発見のみならず、規制当局による監査に対応できます。

バイテンポラルデータベースでは、例えば次のようなことを新たに確認できます。

- 我々はこの顧客の昨年の信用格付けを、前の四半期の時点でどのように把握していたのか？
- 我々はこの証券のポジションを、取引修正前にはどのように判断していたのか？
- 新しい情報の入手前、情報機関はどのようなことを言っていたのか？

従来のユニテンポラルなデータベースでは、今日の時点における顧客の信用格付けについては確認できますが、昨日

あるいは前四半期の時点における格付けは確認できません。

完全かつ修正されていない過去データを遡って確認できるのは、バイテンポラルデータベースだけです。これで過去と現在の違いも把握できます。今日の企業においてバイテンポラルデータベースは、正確な過去（履歴）データの確認、システム全体にわたるデータの管理、データの完全な整合性、より複雑な分析のために必要です。

MarkLogic® はエンタープライズ NoSQL データベースであり、以下の理由により、バイテンポラルデータの格納と管理に最適です。

- **柔軟なデータモデル** – MarkLogic のドキュメント志向データモデルはスキーマ非依存であり、リレーショナルデータベースが苦手とする複雑なバイテンポラルデータ（データの整合性の制約、スキーマの変化、異なる複数のデータモデルなど）を管理できます。
- **エンタープライズレベルの信頼性** – MarkLogic には、他の新世代データベースにはないエンタープライズ機能が備わっており、世界のリーダー企業数百社が基幹業務アプリケーションに使用している実績があります。
- **バイテンポラルは標準機能** – MarkLogic ではバイテンポラルは標準機能です。一方、他社ではこれを別のソフトウェアとして追加しなくてはならないため、コストが増し、複雑になります。

## バイテンポラルがないことによるコスト

ある企業では、バイテンポラルがないことで 2500 万ドルの損失が発生しました。<sup>1</sup> バイテンポラルがないために職を失った政治家はたくさんいます（逆に失わなくて済んだ政治家もたくさんいるでしょう）。極度な規制があり、また履歴データの出自やガバナンスの管理およびデータの変更不能性などが求められる時代において、バイテンポラルがないことに起因する損失の可能性はさらに大きくなります。これは特に金融サービスなどの業界で顕著です。規制において、過去を正確に把握できない場合の罰金が数十億ドル規模になっており、これは今後さらに拡大すると思われます。

1 ハドソンフーズ社は、O157 の発生により 1997 年の年間出荷量の 1/5 をリコールしました。この損害は 2500 万ドルと推定されています。この会社のデータベースでは個々の牛の入荷元に関する現時点の情報しか確認できず、汚染された少量の牛肉を業者が処理した日のデータを確認できませんでした。このため全商品をリコールしなければならなくなりました。詳細については、Richard T. Snodgrass 著『Developing Time-Oriented Database Applications in SQL』（第2章）を参照のこと。

## 3種類のテンポラル

バイテンポラルを理解するには、まず今日のデータベースによる時間の扱い方を理解しておく必要があります。データベースは、時間の観点から「ノンテンポラル」「ユニテンポラル」「バイテンポラル」の3種類に分類できます。これからそれぞれを説明していくにあたって、「ある患者がアレルギーと診断されたタイミング」と「医師がこれについて認識したタイミング」を例として取り上げます。

### ノンテンポラル

ノンテンポラルデータベースでは、時間軸を持つデータは格納しません。ファクトは単なるファクトであり、過去は存在せず、これが現時点では真であるということしかわかりません。時間軸がないデータモデルは単なる「スナップショット」です。



図1: ノンテンポラルデータベースには時間に関する情報が何も含まれない

患者がアレルギーと診断された例を考えてみましょう。アレルギーの有無は、ペニシリンなどの一般的な処方による副作用（あるいは最悪の場合、死亡の可能性も）に関する重要な情報です。ノンテンポラルデータベースでは、現時点での状態しかわかりません。つまり図1のように「この患者にはアレルギーがない」か「この患者はアレルギー陽性と診断されている」しかわかりません（このファクトが真である部分が四角で表現されています）。

ノンテンポラルデータベースでは、時間に無関係なデータの1つの像/見え方だけが提示されます。当然ながら、ノンテンポラルデータベースはあまり使用されません。というもほとんどのアプリケーションでは、時間とともに変化するデータを扱っているからです。

### ユニテンポラル

ユニテンポラルでは、1つの時間軸を扱います。これは「有効時間」(valid time)と呼ばれます。ふつう「時間」と言えばこの「有効時間」のことです。つまり現実世界



図2: ユニテンポラルデータベースは有効時間(実際の時間)のみを記録

で実際に何が起こった時間です。有効時間は1つの時間軸でトラッキングされ、「この患者がアレルギーと診断されたのはいつか」「これと同じアレルギーを持つ患者は何人いるのか」「この患者がアレルギーになってからどのくらい経つのか」のような問いに答えます。今回の例では、図2のグラフを見るとこの患者がアレルギーと診断されたのは、「有効時間」の午前9時だということがわかります。

しかしここで問題なのは、有効時間では全体像の一部しか把握できないことです。この図を第三者が見た場合、「この患者がアレルギーだと診断された」のを医師が知ったのはいつか、ということにははっきりしません。例えば、検査機関がアレルギーを確認した後、この結果を医師が実際に知るまでに時間が経っている場合も考えられます。この重要な情報は、ユニテンポラルデータベースには記録されません。このため例えば、この患者にある薬物を投与したためにアナフィラキシーショックが起きてしまった場合、この薬物を投与してはいけなことを医師が知らなかったかどうかは把握できません。次に、こういった問題をバイテンポラルデータベースがどのように解決するのかを見ていきましょう。

### バイテンポラル

バイテンポラルデータベースでは、「有効時間」と「システム時間」という2つの時間軸に沿ってイベントのタイムスタンプを記録します。「有効時間(valid time)」は現実世界でイベントが発生した時間です。「システム時間(system time)」(「トランザクション時間」とも呼ばれる)は、イベントがデータベースに記録された時間です。この2つの時間軸は、図3のようになります。この例で言うと、「有効時間」は検査機関がアレルギーを発見した時間であり、「システム時間」は医師がこれを認識し、カルテに記録した時間です。

ユニテンポラルデータベースは、「有効時間とシステム

時間は必ず同じである」という誤った前提に基づいており、その結果、貴重な情報が失われてしまいます。図3のように、有効時間とシステム時間が同じになる場合もあります。しかしそのようなことも、バイテンポラルデータベースがない限り確認できません。バイテンポラルデータベースでは2つの時間軸で記録するので、正確な記録を取ることができます。

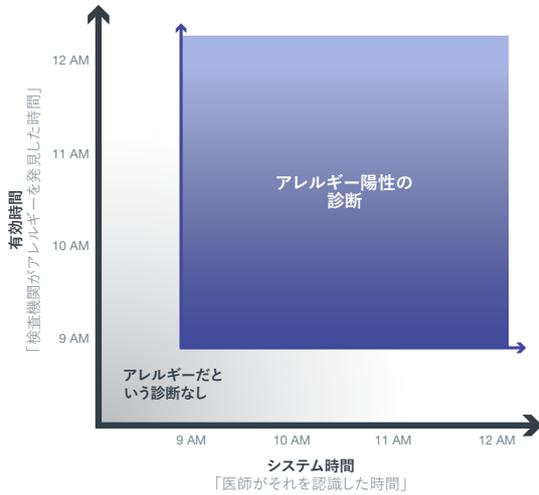


図3: バイテンポラルデータベースは「有効時間」と「システム時間」の両方を記録

患者のアレルギーの例で言うと、医師が実際にアレルギーを認識したのは10時30分です。これは検査機関が検査を行い、この患者にはアレルギーがあると診断してから1時間半後のことです。検査機関がこの患者のアレルギーを発見したのは午前9時ですが、この情報は午前10時30分になるまで医師に伝わりませんでした。これは有効時間とシステム時間の間にタイムラグがあるということであり、図4のように表現できます。

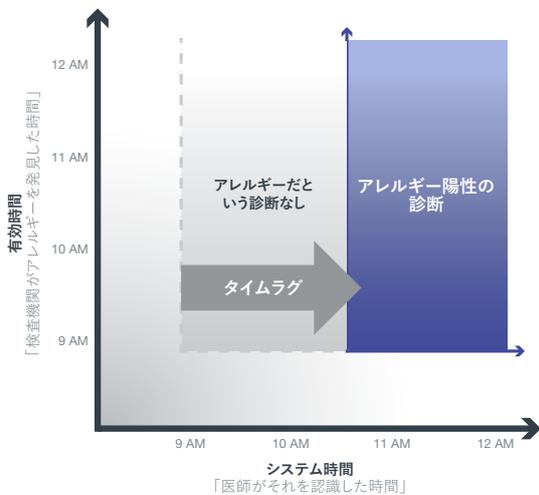


図4: バイテンポラルデータベースでは、情報のタイムラグを記録

この例をさらに進めましょう。同日の午前11時30分、医師のところに検査機関から電話がきました。「検査のやり方を間違っていたことが判明した」というのです。結果は実は陰性で、この患者にはアレルギーはありませんでした。この修正は図5のように表現できます。バイテンポラルデータベースでは、履歴データを簡単に修正できます。またこれはどのデータも上書きしません。

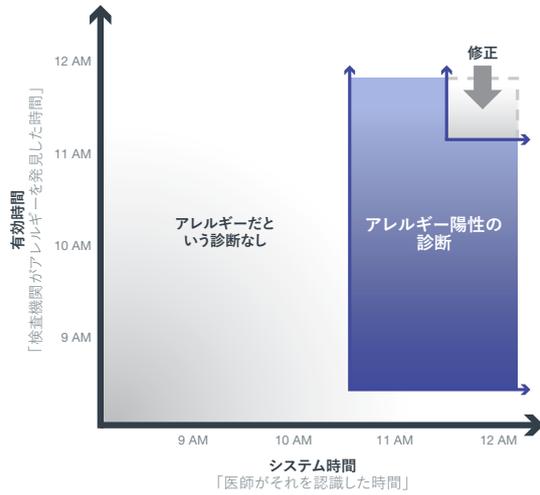


図5: バイテンポラルデータベースでは、データを上書きせずに修正を記録

図5を見ると、以下のことが確認できます。

- **午前10時30分（システム時間）以前**、医師はアレルギーのことを知らなかった。
- **午前10時30分（システム時間）**、医師はこの患者にアレルギーがあることを記録する。これは検査機関で午前9時（有効時間）にすでに発見されていたこと。
- **午前11時30分（システム時間）**、検査機関と医師はこの患者にはアレルギーがなかったという間違いに気づき、このレコードを更新した。

2つの時間軸でトラッキングすることで、遡ってあるイベント（出来事）の真の姿を把握できます。これは間違いを理解し、これを避けるために非常に重要です。というのも医師の判断は、ある時点において彼が知っていたこと / 知らなかったことによって変わってくるからです。病院においては、薬物アレルギーは命にかかります。このため患者が診断された時間、また病院側がこの情報を知った時間の正確な記録は極めて重要です。

ユニテナポラル	バイテナポラル
この患者がペニシリンアレルギーだということを検査機関が認識したのはいつか。	この患者がペニシリンアレルギーだということを検査機関が認識したのはいつか、ならびに、このアレルギーを医師が認識したのはいつか。
銀行の相手方が売り注文をキャンセルしたのはいつか。	銀行の相手方が売り注文をキャンセルしたのはいつか、ならびに、トレーダーがこのキャンセルを認識したのはいつか。
12月4日の取引に関してどのようなリファレンスデータが存在していたのか。	この取引に関してトレーダーが12月4日時点で実際に持っていたリファレンスデータはどのようなものか。
現在の雇用記録によるとジョンが保険の適用内となったのはいつか。	2012年の雇用記録によるとジョンが保険の適用内となったのはいつか。

表1: 複数の例におけるユニテナポラルとバイテナポラルの比較

このアレルギー診断の例はシンプルですが、同じ考え方は、金融取引、保険加入、住宅購入などあらゆるデータに適用できます。これら3つの例のいずれの場合でも、正確な真実の姿を把握するためには、有効時間ならびにシステム時間の「開始日」と「終了日」を記録しておく必要があります。

## バイテナポラルのメリット

簡単にいうと、バイテナポラルにより時間の管理が向上します。バイテナポラルを代替できるものではありません。例えばテナポラルバージョンングなどでは、履歴（過去）データをシームレスに扱い、柔軟にクエリを実行して把握することはできません。バイテナポラルはあらゆる組織で活用できる重要な機能ですが、特に規制・訴訟リスクが増加している業界（金融、保険、ヘルスケアなど）ではバイテナポラルのニーズが高まっています。これらの業界では、新しい規制、訴訟、頻繁かつ詳細な監査、コンプライアンス違反の罰金などにおいて、過去のあらゆる行為に関する説明責任があります。履歴管理に優れた組織は、リスクを削減し、無傷のまま監査を終えることができます。

## バイテナポラルでできること

- **規制や監査への対応** – 過去を正確に把握し、厳しさを増す透明性やアカウントビリティの要件を満たす
- **リスクの管理** – 真の履歴データを分析することで、リスクモデルやビジネスインテリジェンスを改善する
- **コスト削減** – シンプルなアーキテクチャにより履歴データの重複を解消することでコストや運用上のリスクを削減

## バイテナポラルのニーズの高まり

一般的に言って、規制対応改善のニーズは拡大しています。特に金融サービスなどの業界においてはそのインパクトは重大です。近年、大手銀行は莫大な罰金を科せられ、規制の圧力も高まっています。2009年以降、欧米の銀行は1280億ドルの罰金を規制当局に支払っています。罰金・課徴金の最大は2014年の650億ドルで、これは2013年比で40%増です。<sup>2</sup>

今日、規制当局はより積極的であり、詳細まで厳しく要求しています。ボストン・コンサルティング・グループのリスク管理および規制のグローバル統括責任者であるジェロルド・グラスホフ氏によると、規制圧力は今や銀行の主要課題となっています。「今や運用モデル、製品、法的リスクを変えていく必要があります。…規制問題ほどビジネスモデルに影響を及ぼしているものはありません。これは戦略上の最大のチャレンジなのです」。ビジネス手法の変化に対応するために、銀行はIT部門ならびにデータ管理手法を変更し、透明性を高めるようにしています。<sup>3</sup>

他の業界でも規制の圧力は高まっています。例えばヘルスケア業界には医療過誤の問題があり、この影響を1兆ドルと推定するレポートもあります。<sup>4</sup> いくつかの理由で過誤が発生したのかを把握することは、医療判断の改善や医療過誤の回避のために重要です。またヘルスケア業界全体における不正や濫用のコストの増大を考慮する必要があります（米国では想定820億

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

3 Boston Consulting Group, "Building the Transparent Bank", Dec. 2014 <[https://www.bcgperspectives.com/Images/Building\\_the\\_Transparent\\_Bank\\_Dec\\_2014\\_tcm80-177814.pdf](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>>

# 「私たちは極端に規制が厳しい時代にいるのです。これはだまってやり過ぎてせるものではありません」

ベンジャミン・ロースキー氏  
 ニューヨーク州金融サービス管理局<sup>5</sup>

～2720億ドル)。<sup>6</sup> 残念ながら患者の安全、医療過誤の訴訟、不正、濫用に関するコストは増加する一方で、さらに複雑化しています。

バイテンポラルなデータ管理の導入により、リスク削減、透明性改善、競争力強化を進められます。

## さまざまな業界におけるバイテンポラル

### 金融サービス

大手銀行はバイテンポラルを活用して、データ活用を改善し、業務に影響する法律や規制の改正に対応しています。例えば、バイテンポラルを使うことにより、取引（発生と修正の時間）を正確に記録できます。取引後の相手方との照合において、取引が更新されることはよくあります。ユニテンポラルデータベースでは、更新によって過去のデータが上書きされます。これは個々のトレーダーだけでなく会社全体に対しても大きなリスクとなります。バイテンポラルでは、取引のライフサイクル全体を正確に把握できます（相手方の名称変更、取引ID、金額変更など）。

### 保険

保険業界では、バイテンポラルで、時間の経過に伴う保証範囲の変化を明確に把握できるようになります。過去に遡った変更の場合でも、データを上書きしません。

保険会社は、過去の任意の時点における保険の適用範囲を確認できます。また保険会社は、ある人が任意の時点において社員であったかどうかについて、また任意の時点でこれを保険会社がどう認識していたのかについて正確に把握する必要があります。

ユニテンポラル	バイテンポラル
トレーダーのポジションはどのようなものだったと我々は理解しているのか。また取引実行時点においてトレーダーが知りえた情報はどのようなものだと考えられるか。	取引実行時のトレーダーの正確なポジションはどうだったのか。取引実行時点に参照可能なリファレンスデータは正確にはどのようなものだったのか。
この顧客の昨年の信用格付けはどうだったのか。	前の四半期の時点で我々が把握していた、この顧客の昨年の信用格付けはどうだったのか。
11時の取引時点での市場エクスポージャーはどのくらいだったのか。	11時の取引時点での市場エクスポージャーはどのくらいだったと、11時30分の時点で認識していたのか。
発表した時点で、我が社の利益はどのくらいだったのか。	発表した時点で、我が社の利益はどのくらいだと我々は考えていたのか。

表2: 金融サービスにおけるバイテンポラル

ユニテンポラル	バイテンポラル
災害による保険料への影響はどのように推定されていたのか。	遡ってデータを調整する前は、災害による保険料への影響はどのように推定されていたのか。
診断時に保険金受取人は保険の対象となっていたのか。	法律施行前、診断時に保険金受取人は保険の対象となっていたのか。
この出来事の発生時点においてこの会社の社員だったか。	記録に基づくと、この出来事の発生時点においてはこの会社の社員だったということになっているのか。

表3: 保険におけるバイテンポラル

5 James Sterngold. "For Banks, 2014 Was a Year of Big Penalties." Wall Street Journal, 2014.

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

MarkLogicのバイテンポラルにより、柔軟な関係付けができ、またデータ（営業終了後の情報だけでなく、営業中の情報など）から得られた新しい価値をさまざまな顧客グループに即座に提供できます。これまではこれを完全に行うことはできませんでした。…実際、MarkLogicのバイテンポラルによって、顧客は新しい分析や豊かなコンプライアンス管理機能を利用できるようになるでしょう」

パオロ・ベリッツォーリ氏、ブロードリッジ・ファイナンシャル・ソリューションズ、アーキテクチャおよびグローバルテクノロジーグローバル主任

### ヘルスケア

ヘルスケアでは関係者（医療従事者、保険会社、製薬・バイオテクノロジー企業など）すべてが巨大な問題に直面しています。ヘルスケアのIT部門はコスト削減と結果の改善を実現するために、バイテンポラルなどの機能を活用することで、一人の患者が複数の病院で受けたすべての治療を把握できます。また有害事象の調査ツールも改善できます。また保険会社は治療費の請求コードを受け取った際に、各患者ごとに履歴全部を遡って確認できます。また保険の適用範囲を遡って変更したとしても、過去の情報が損なわれることはありません。また製薬企業やバイオテクノロジー企業にとっても、リサーチとビジネスの両方においてバイテンポラルで意思決定を強化できるというメリットがあります。

### 警察と情報機関

バイテンポラルを使うことで、あるイベントに関して集められたバラバラなファクト（事実）から全体像を作り出せます。全体像が得られることで、政府機関は動機をもっと理解できるようになり、将来の事象の予測を改善できます。捜査においても、バイテンポラルを使うと、過去のある時点において特定の判断をした理由を確認できます。これは未解決事件を再捜査する場合に特に便利です。

ユニテンポラル	バイテンポラル
この薬が処方された時点でこの患者のカルテはどのようなものだったのか。	ラボの結果によってこのカルテが更新される前、この薬が処方された時点でこの患者のカルテはどのようなものだったのか。
2010年6月の時点において、この患者への利用が認められている治療法は何だったのか。	2010年6月の時点でこの患者への利用が認められている治療法として、2010年8月の時点で我々が理解していたものは何だったのか。
追加投資をした時点で、臨床検査の結果はどのようなものだったのか。	臨床検査の結果の更新前、追加投資をした時点で臨床検査の結果はどのようなものだったのか。

表4：ヘルスケアにおけるバイテンポラル

ユニテンポラル	バイテンポラル
この判断をした際には何が起っていたのか。	この判断をした際には何が起っていたかと我々は考えていたのか。
この出来事はいつ発生したのか。	この出来事はいつ発生したのか。またこれを記録したのはいつか。
あのような行動に出た理由を、今どのように考えているのか。	あのような行動に出た理由を、その当時はどのように考えていたのか。

表5：警察と情報機関におけるバイテンポラル

「リレーショナルが提供するバイテンポラルはあまり受け入れられていません。というのも時間の経過とともに、データのかたちも変わることが普通だからです。…RDBMSではスキーマの変化に対応できません」

グローバルな投資銀行

## これまでバイテンポラルの導入が困難だった理由

ここまで読んできたみなさんは、「もしバイテンポラルがそんなに重要ならば、なぜこれまで聞いたことがなかったのか」と思うでしょう。ここ20年、テンポラル(時間)データに関する研究論文は数千件書かれてきています。また、バイテンポラルも1990年初頭から専門家たちによって議論されていますが、未だにあまり知られていません。

バイテンポラルにビジネス的なメリットがあることは明らかです。しかし業務部門のアナリストのほとんどは、バイテンポラルデータの存在すら知りません。というのもほとんど実績がないからです。ここで問題なのは、リレーショナルデータベースでは、バイテンポラルの実装と保守が複雑すぎて、メリットを上回ってしまうということです。事実、ただのテンポラルデータでさえリレーショナルデータベースで扱うことは極めて困難なのです。

残念ながら、リレーショナルデータベースでバイテンポラルを簡単に扱えるよう努力しても、従来のツールでは実現は困難です。複雑なバイテンポラルをリレーショナルデータベースで扱えるエキスパートは世界でもほんの数人しかいないでしょう。ここでは詳細な解説はさておき、リレーショナルデータベースがバイテンポラルのデータモデリングに適していない主な理由をいくつか挙げておきます。

- **整合性の制約** - リレーショナルデータモデルでは、参照整合性および実体整合性に制約があり、定義されたスキーマは容易に変更できません。またテンポラルデータ特有の制約もあります。例えば、表内の子行に含められる有効期間は、親行で定義した有効期間内でなければなりません。リレーショナルテーブルにバイテンポラル列を追加すると、リレーショナルデータモデルが大きく損なわれる可能性があります。

- **時間に伴うスキーマの変化** - リレーショナルモデルにバイテンポラルを追加するとかなり複雑になります。アーキテクチャや構造自体の変化もテンポラル(時間的)です。新しい列の追加により時間的な次元(ディメンション)が含まれるようになった場合、あるいは新規データの読み取り時に新規テーブルを作成した場合、スキーマが変更されます。スキーマの変更やそれに伴うアプリケーションの変更は、バイテンポラルでなくても複雑です。
- **複数のデータモデル** - 変化するスキーマの扱いは困難です。複数のデータモデルやデータサイロにまたがる変化するスキーマが複数あり、これを「single source of truth」に集約しなければならない場合は、さらに困難です。データ統合自体たいへんな作業ですが、これにバイテンポラルが加わると複雑さは爆発的に増大します。
- **パフォーマンスの劣化** - 通常バイテンポラルクエリにより、読み取り/書き込みのパフォーマンスは劣化します。これは、あらゆるクエリにおいて考慮すべき時間軸がもう1つ増えるからです。またデータは複数の表(場合によっては複数のサーバー)にまたがっていることが一般的であり、クエリをシンプルにしてパフォーマンスを向上させようとしても、複数の表に散在するバイテンポラルデータの複雑さやパフォーマンスの問題を解消できません。
- **ベンダーロックイン** - ベンダーのなかにはバイテンポラルを改善しはじめたところもあります。しかしながら、昔のSQL標準化の際と同様に、各ベンダーは追加機能を独自のシンタックスで実装しており、そのコストを上乗せしてきます。

リレーショナルデータベースでバイテンポラルを扱おうとした結果、次善の策を発見することがよくあります。一般的な反応としては以下のものがあります。

「極めて普遍的な『時間』と、時間の経過に伴う企業内での変化をモデル化しようとする場合（静的かつ固定的な状態は稀であり、あまり意味のないものですが）、正直言ってSQLはこういった時間の経過に伴う変化や情報の効率的なモデリング、クエリ、修正に関して、かなりお粗末なことしかできません」

Richard Snodgrass 『Developing Time-Oriented Database Applications in SQL』<sup>7</sup>

- 「スローリーチェンジングディメンションを使えばいいのでは」と思うかもしれませんが。最近、スローリーチェンジングディメンション（SCD）のタイプ2を使ってバイテンポラルデータを近似的に扱う試みがなされていますが、この手法の問題点はすでに詳細に記述されています。<sup>8</sup> SCDは有効なテンポラルデータを近似的にしか扱えず、多くの矛盾が発生します。これは後から把握・修正するのが困難です。またすべてを適切に設計した場合でも、クエリのパフォーマンスは遅く、結果にも再現性がない可能性が高いです。
- 「スナップショットを頻繁に取ればいいのでは」と思うかもしれませんが。このアプローチは「テンポラルバージョンニング」と呼ばれ、バイテンポラルの扱いに関してはより一般的です。というのほとんどの企業はすでに月次あるいは週次でデータのスナップショットを取っているからです。このアプローチは安定しており予測可能です。しかし残念なことに、このアプローチでは重複データが大量に発生するためストレージのコストが肥大し、依然として多くの情報が失われてしまいます（スナップショット間のギャップがあるため）。またたとえスナップショットを頻繁に取ったとしても、ほとんどの業種において規制監督庁を満足させられません。規制監督庁およびデータアナリストは、特定の問いに素早く答える必要があるため、このようなギャップは望ましくありません。
- 「監査ログを使えばいいのでは」と思うかもしれませんが。しかし、ログはイベント情報のトラッキングには便利ですが、バイテンポラルに対しては容易かつ素早くクエリを実行できないため不十分です。また変更不可の（イミュータブルな）レコード保持のための基準も満たせません。

バイテンポラルでしか、時間を管理して過去データを素早くかつシームレスに確認したり、任意の時点で素早くクエリを実行したり、現在のデータと過去のデータの両方を業務活用したりできません。

## バイテンポラルに取り組むべき時代が到来した理由

MarkLogicはエンタープライズ NoSQL データベースであり、実際の業務におけるバイテンポラルデータの格納や管理に必要な柔軟性があります。この際、複雑なクエリにおいてもパフォーマンス、データの完全性、セキュリティを犠牲にしません。MarkLogicは、バイテンポラル機能を持つ唯一のエンタープライズ NoSQL データベースです。

MarkLogicはスキーマ非依存であり、データをドキュメントとして管理します。つまり、データベースの利用中

MarkLogicのバイテンポラルの利点	MarkLogic	他のデータベース
スキーマ非依存で、スキーマの変化や複数のデータモデルの変更に対応	✓	✗
コーディングと運用がシンプル	✓	✗
短期間での価値実現	✓	✗
スケラビリティ、弾力的な拡張・縮退、ストレージコストの削減	✓	✗

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

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

ずっと準拠しなければならないような厳密なスキーマは不要です。

MarkLogic では、後から新規データソースを統合する場合でも読み込み前に複雑な ETL は不要です。リレーショナルデータベースでは列 (DATE 列など) の追加が面倒ですが、その心配もありません。

バイテンポラルデータの寿命は数十年にも及びます。このため、新規データソース追加時のスキーマ変更にも素早く対応できるデータベースが必要です。MarkLogic では、新規データソースの読み込みが楽です。また矛盾がある場合 (列名「SRC\_DATE」は「CLAIM\_DATE」でなければならないなど)、MarkLogic では標準語彙に簡単に変換できます。MarkLogic では、リレーショナルデータベースにおけるモデリング制約 (実体整合性、参照整合性、非正規化など) について心配する必要はありません。これはバイテンポラルの場合も同様です。

また MarkLogic は大規模データ統合プロジェクトにおいて、リレーショナルデータベースよりも圧倒的に優れています。プロジェクトのデリバリー時間が短縮され (要件定義とデータモデリングの時間の短縮による)、プロトタイプの高品質が改善されます。大手金融サービス企業であるブロードリッジは、「MarkLogic の最初のプロジェクトは 60 日かかりました。これは従来の技術では 3000 日かかると想定されていたものです」と言っています。

## MarkLogicにおけるバイテンポラルの処理

リレーショナルデータベースの経験がある人にとっては、MarkLogic におけるテンポラル / バイテンポラルデータの扱いは馴染み深いものでしょう。一番の違いは、この情報が表の日付列ではなく、ドキュメントのタイムスタンプとして扱われるということです。MarkLogic では、バイテンポラルデータを含むすべてのデータをドキュメントとして管理します。

JSON および XML ドキュメント内に有効時間の開始 / 終了時間、またシステム時間の開始 / 終了時間が含まれている場合、これはバイテンポラルドキュメントとして扱われます。バイテンポラルドキュメントを MarkLogic に読み込む方法として、mlcp (MarkLogic Content Pump) や REST API があります。あるいは、シンプルな JavaScript 更新クエリを使えます (以下を参照)。

バイテンポラルドキュメントは MarkLogic に読み込まれた後、有効時間およびシステム時間に対してレンジインデックスが付けられた一連のドキュメントとして管理されます。有効時間軸ならびにシステム時間軸は、レンジインデックスの名前付きペア用のコンテナとなります。その後、バイテンポラルドキュメントはテンポラルコレクション (テンポラルドキュメント用の論理的なグループ) に格納されます。タイムスタンプのスキーマが異なるドキュメントには、別のテンポラルコレクションを追加できます。

元のドキュメントは、MarkLogic に読み込まれた後、そのまま保持され変更されることはありません。バイテン

```
declareUpdate();

var root =
  { "tempdoc": {
    "systemStart": null,
    "systemEnd": null,
    "validStart": "2014-04-03T11:00:00",
    "validEnd": "2014-04-03T16:00:00",
    "content": "some data, like closing price"
  }
};

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

図6: バイテンポラルドキュメントの更新

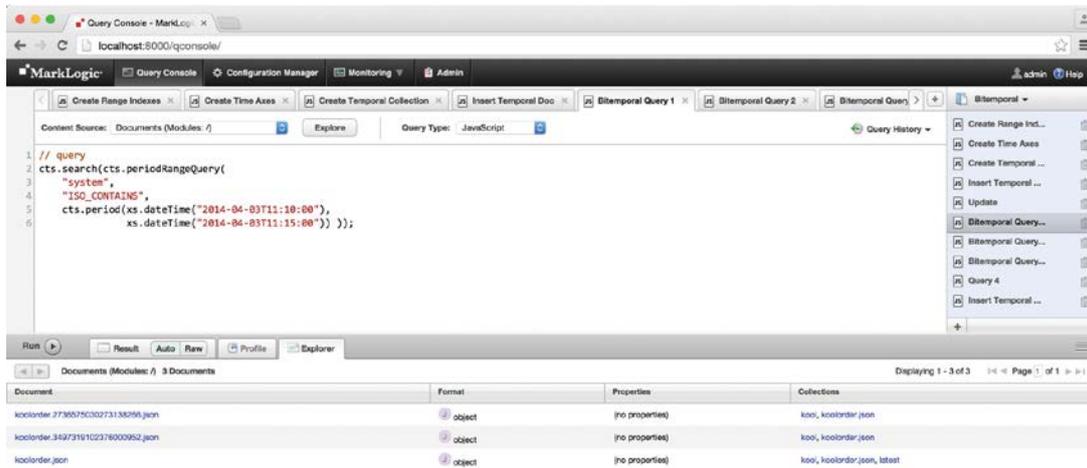


図7: MarkLogicのクエリコンソール内のバイテンポラルクエリ

ポラルドキュメントを「削除」した場合でも、システム時間を変更して（「無限」だったものを「削除の時間」に変更）このドキュメントを保持します。更新でも同様のプロセスとなります。以前のバージョンも保持され「新規」バージョンが単に追加されるだけです。MarkLogicではシステム時間の開始時間を更新できません。コレクションに対して設定されたシステム時間は継続して経過していき、これによりデータの一貫性を担保します。

情報の出自を完全なガバナンスをもってトラッキングし、またこれを変更不可にしておくことは極めて重要です。このため MarkLogic は、バイテンポラルドキュメントにセキュリティモデルを適用しています。MarkLogic は NIAP（アメリカ国家情報保証パートナーシップ）のコモンクワイテリア評価 / 検証スキーム認証を受けており、デフォルトで RBAC（ロールベースのアクセス制御）を使ってドキュメントへのアクセスを制御しています。このハイレベルなセキュリティによって、過去レコードの改ざんを阻止し、また時間が経過してもドキュメントのパーミッションを維持します。

## MarkLogicのバイテンポラルの主要機能

**挿入、更新（絶対に削除しない）** - 時間に関する JSON/XML ドキュメントを読み込む。テンポラル API や mlcp で有効時間を参照する。新しいバージョンが追加されてもデータ損失なしで変更可能

**複雑なテンポラルクエリ** - 有効時間軸とシステム時間軸に関してデータベースにクエリを実行。期間を比較する際に標準 Allen および SQL 演算子を使用

**変化するスキーマに対応** - 時間の経過とともにデータの形が変わっても問題なし。リレーショナルデータベースと違って、MarkLogic はスキーマ非依存であり、時間の経過に伴うスキーマ変更を容易に管理可能

**LSQT のメンテナンス** - LSQT (Last Stable Query Time) と呼ばれる特殊なタイムスタンプで、複数システム全体において各システムの開始時間を管理・調整可能

**階層型ストレージとの組み合わせ** - 階層型ストレージを使って、履歴データをより安価なストレージ層に移動（データのクエリは依然として可能）

**セマンティックとの組み合わせ** - バイテンポラル要素をドキュメントに割り当て。RDF トリプル自体、あるいは RDF トリプルを含むドキュメントで、時間の経過に伴う関係の変化をトラッキング可能

**地理情報と組み合わせ** - 時間と場所に関してデータをトラッキング。MarkLogic は地理情報データを格納できるため、時間の経過に伴う地理情報データの変化を正確にトラッキング

**認証済みセキュリティを活用** - 他のあらゆるドキュメントと同様にバイテンポラルドキュメントも認証済みセキュリティで管理。RBAC（ロールベースのアクセス管理）などのセキュリティモデルを使用

**素早く容易な拡張** - MarkLogic のスケールアウトアーキテクチャにより、アンダープロビジョニングの心配を払拭。バイテンポラルデータの需要が拡大した場合に、容易にノードを追加

MarkLogicには、高度なデータ管理技術を提供してきた実績があり、多くのユーザーやパートナーは、複雑なデータをアジャイルに管理しています。この結果として、MarkLogicのユーザーやパートナーはデータの管理と利用に関して、他のデータベースの利用者よりも成熟しクリエイティブです」

カール・オロフソン、IDC、データ管理ソフトウェアリサーチ担当副社長

## 始めてみましょう

時間の管理は容易ではありません。そのため、Y2Kのような数十億ドル規模の問題が発生してしまうのです。<sup>9</sup>しかし時間の管理は不可避であり、バイテンポラルこそがデータベースによる時間の管理の未来です。これは「何

を知っていたのか」および「いつ知ったのか」の記録方法の改善が求められているからです。MarkLogicはバイテンポラル導入の障害を取り除きます。またバイテンポラルデータの格納と管理に最も適したデータベースです。

## 次のステップ

1. 現在解決できていない問題を特定する
2. バイテンポラルの導入によるビジネス上のメリットを特定する
3. 現在のデータ管理環境を査定する
4. MarkLogic に問い合わせ、導入について話し合う
5. MarkLogic をダウンロードする
6. MarkLogic の無料トレーニングについて詳しく知る

## さらに詳しく

- **MarkLogic のドキュメントを読む** – MarkLogic におけるバイテンポラルデータの扱い方を説明するガイド。 <http://docs.marklogic.com/guide/temporal>
- **プレゼンテーションを見る** – MarkLogic のユーザーが「なぜ銀行がバイテンポラルに注目するのか」について説明します。 <https://www.youtube.com/watch?v=n29Gtit3lMU>
- **弊社との打ち合わせ** – 具体的な案件については、MarkLogic の営業担当までお問い合わせください。 [MarkLogic-JP@marklogic.com](mailto:MarkLogic-JP@marklogic.com)

<sup>9</sup> BBC および ComputerWorld によると、「2000 年問題」への対応コストは 6080 億ドル（インフレ考慮せず）。詳細：Robert L. Mitchell. “Y2K: The good, the bad and the crazy”. ComputerWorld (28 December 2009) <<http://www.computerworld.com/article/2522197/it-management/y2k--the-good--the-bad-and-the-crazy.html?page=2>>

© 2019 MARKLOGIC CORPORATION. ALL RIGHTS RESERVED. このテクノロジーは、米国特許番号 7,127,469B2、米国特許番号 7,171,404B2、米国特許番号7,756,858 B2、米国特許番号7,962,474 B2で保護されています。MarkLogicは、米国およびその他の国におけるMarkLogic Corporationの商標または登録商標です。本書に記載されているその他の商標は、各企業の所有物です。

マークロジック株式会社 MARKLOGIC K.K.

150-0001 東京都渋谷区神宮前1-5-8 神宮前タワービルディング 13F  
03 4540 0337 | [jp.marklogic.com](http://jp.marklogic.com) | [MarkLogic-JP@marklogic.com](mailto:MarkLogic-JP@marklogic.com)



150-0001 東京都渋谷区神宮前1-5-8 神宮前タワービルディング 13F  
03 4540 0337

[jp.marklogic.com](http://jp.marklogic.com) | [MarkLogic-JP@marklogic.com](mailto:MarkLogic-JP@marklogic.com)