

ホワイトペーパー

データ収集によくある思い込み

寄稿者

Elliot Middleton

AVEVA プロダクトマネージャー

まとめ

リレーショナルデータベースは多くの用途に適していますが、時系列データに対しては最適なソリューションではありません。

データ収集アプリケーションの歴史は長く、そのため、ある種の先入観や根拠のない印象、あるいは「思い込み」が長年にわたって培われてきましたが、それらは実体や証拠がないまま永続しているように見えます。もし、あなたがリレーショナルデータベースの長所と短所をよりよく理解し、製造業向けデータ収集アプリケーションの文脈でこれらの思い込みを明らかにし、挑戦したいのであれば、この記事を読んでください。また、データ収集アプリケーションの基本を理解し、評価の際に考慮すべき点を発見し、AVEVA™ Historianがもたらす価値について期待を適切に設定する助けとなるでしょう。

思い込み 1

ストレージはとても安いので、効率は関係ない。

典型的なプロセスが実際にどれだけデータを生成するのか、よく理解しておく必要があります。あまり大規模でない5,000タグのデータ収集ですら毎秒データを記録すると、1年で1,570億個の値が生成されます。これを8バイト単位で効率的に保存すると、1年でおよそ1テラバイトになります。時系列データに対するSQL Serverのストレージ要件とAVEVA Historianのストレージ要件を比較したテストでは、必要なインデックスを含めるとその差は50:1でした。ストレージの価格が下がっているとはいえ、年間50テラバイトのデータは大変な量です。また、単にデータ量が多ければいいというものではないことも認識しておいてください。

また、ほとんどのデータ収集アプリケーションでは、データの保護が必要なため、バックアップやディスクミラーリングのためのストレージ容量も増えます。業界によっては、数年分のデータが必要な規制もあり、ストレージの必要性はさらに高まります。

思い込み 2

リレーショナルデータベースは高速だ。

ハードウェアの価格性能の向上に伴い、リレーショナルデータベースはその恩恵を受けてきました。しかし、リレーショナルデータベースは、複数のテーブルの値を一度に更新するような「トランザクション」に対して参照整合性を守るように設計されているため、大きなオーバーヘッドが発生します。

例えばハイエンドのハードウェア (Lenovo ThinkSystem SR650、112プロセッサスレッド、価格は60万ドル以上) では、SQL Server 2017は1秒あたり7,000トランザクションを実行できます。よりリーズナブルなハードウェアでこのような速度に近づけようとすると、フロントエンドのバッファリングでデータを収集し、1つのトランザクションとして多数の値をデータベースに流し込む必要があります。トランザクションを完全にサポートしていない他のデータベース、例えば、MySQLのフリーウェアのMyISAMストレージエンジンは、より費用対効果の高いスループットをサポートしますが、ごくわずかなヒストリアンアプリケーションを除いては、フロントエンドバッファが必要です。

もちろん、そもそもデータを保存する理由は、後からデータを取り出せるようにするためです。そのためには、検索のパフォーマンスも重要です。特にリレーショナルデータベースのような汎用的なソリューションでは、データを保存する際の効率 (スループットが高い) か、検索する際の効率 (検索速度が速い) かのどちらかだけでなく、両方を両立させるようにデータを整理することが可能です。汎用データベースから時系列データを効率的に検索するには、例えばスループットの高いMyISAMストレージエンジンでは利用できない「クラスター化インデックス」を使用する必要があります。

一方、時系列データ専用設計されたストレージエンジンは、データの収集と消費に関する知識を活用して、両者を効率的に保存することができますが、これはデータがより一般的なものであれば不可能です。

思い込み 3

リレーショナルデータベースは、データ収集に最適だ。

既存の技術をどのように応用するか、新しいアイデアが生まれることはよくあることです。3M社のポストイットは、既存の粘着性のない接着剤を応用したものとして有名です。

しかし、企業は10年以上前から、専用の時系列ストレージの代わりに、シンプルなりレーショナルデータベースを使うことを試みてきました。リレーショナルデータベースを使うことが最適解にもかかわらず、専用のデータ収集ソフトウェアの市場は大きく成長し続けています。

思い込み 4

SQLはリレーショナルデータベースにしか対応していない。

リレーショナルデータベースには多くの利点がありますが、一躍脚光を浴びることになったのは、パワフルなクエリー機能でした。

SQLは、ユーザーがデータの羅列から価値を引き出す方法を共通化することでそれまで必要だったプログラミング開発から、シンプルで柔軟性の高い共通言語でデータを解析する方法へと根本的に変化しました。

幸いなことに、SQLを専用のデータ収集ソフトウェアに適応させることは非常に実用的であり、リレーショナルデータベースの固有の制限を受けずにSQLの多大な利点とパワーを活用することができます。

思い込み 5

時系列データはどんなデータベースでも管理できる。

データを検索するためのSQL (Structured Query Language) の能力を使えるのだから、リレーショナルデータベースは、トランザクションデータと同様に時系列データを検索するのに適していると考えている人がいるかもしれません。SQLが持つ柔軟性を活用する意味では確かにその通りですが、それは時系列データの管理には向かない次のような根本的な仮定に基づいています。

- データレコードに固有の順序はない (実際には、時系列データは時間によって順序付けられている)、
- すべてのデータは明示的に保存されている (実際には、ほとんどのヒストリアンデータは、実際のデータの連続体からのサンプルを表しているに過ぎない)、
- すべてのデータの重要性は等しい。

これらの違いは重要です。ある測定器が「7:59:58.603」というタイムスタンプ付きの値を報告し、ユーザーがリレーショナルデータベースに「8:00:00.000」の値を問い合わせた場合、正確な時間のレコードが保存されていないため、データは返されません。

このデータベースは、時間が連続体であることを認識していません。同様に、気温が「21.0℃」で、その2分後に「23.0℃」になっていたとしても、その中間の気温がおよそ「22.0℃」であったことを推察する能力は、データベースには備わっていません。

特に製造業向けのデータ収集で最も重要なのは、定期的な作業がほとんどないということです。クライアントアプリケーションが例外を発見する唯一の方法が、測定のためにすべてのデータを照会することであるならば、サーバー、ネットワーク、クライアントといったシステム全体に大きな負荷がかかります。一方、専用のデータ収集ソフトウェアは一般的に、重要でないデータをフィルタリングする手段を持っており（連続するレコードの比較に基づいて）、クライアントアプリケーションに配信しなければならないデータの量を根本的に減らすことができます。

思い込み 6

リレーショナルデータベースでも 時系列データの管理は問題なくこなせる。

リレーショナルデータベースは、膨大な量のデータを蓄積するように設計されています。しかし、データ量の増加に伴い、クエリーの実行時間、バックアップのサイズ、その他多くの日常的な操作が必要になります。テーブルが増え続けるというパフォーマンスの問題を軽減するために、データベース管理者は、データベースからのデータのパーティ、インデックスの再構築、および関連する操作を日常的に行う必要があります。トランザクションの整合性を保護しているデータベースでは、このパーティ操作によって通常のデータベースの更新が中断されますが、これは24時間×365日の連続稼働を実行しているデータ収集のアプリケーションにとっては問題となります。

パーティされたデータが後で必要になった場合（例えば、監査や規制上の要求に対応するため）、データを復元することは簡単ではありません。一般的に推奨される方法は、必要なデータを含むデータベースのフルバックアップを、この目的のために専用の別システムに復元するか、本番システムをオフラインにして使用することです。必要なデータが1つのデータベースバックアップに含まれていない場合は、さらに問題となります。例えば、オンラインデータベースに過去30日分のデータしか保存しておらず、監査で90日分ものデータが必要な場合は、すべてのデータを手動で1つのデータベースに統合するか、3つのシステムを用意してそれぞれに30日分のウィンドウを分離するか、各バックアップを連続して調べる必要があります。

一方、専用設計されたデータ収集ソフトウェアは、データの急激な増加に対応するとともに、データのサブセットをオフラインやオンラインで簡単に取得できるように設計されています。

思い込み 7

リレーショナルデータベースか専用設計された データ収集ソフトウェアしか選択肢がない。

ほとんどのデータ収集ソフトウェアは、リレーショナルデータベースの固有の限界に対処するために完全に独自の技術を使用するか、独自のエンジニアリングコストを削減するためにリレーショナルデータベースを完全に活用するかのどちらかであることは事実ですが、AVEVA Historianは実際に両方の世界の最良のものを提供します。AVEVA Historianは、比較的静的な設定データを管理するために強固なリレーショナルスキーマに依存していますが、Microsoft SQL Serverのネイティブなトランザクションストレージエンジンとクエリープロセッサを独自の拡張機能で拡張し、時系列データに対する制限に対応しています。

Microsoft SQL Serverをベースに構築することで、完全に独自のソリューションに比べて、セキュリティや管理が容易なソリューションを提供しますが、ヒストリアンに必要な基本的な機能については妥協しません。

思い込み 8

製造業に特化したアプリケーションである必要はない。

正しく設計されたデータ収集ソフトウェアは、リレーショナルデータベースの領域の外側にある、産業アプリケーションの過酷な使用環境に対処するための仕組みを提供します。

データをどのように利用しようと考えていますか？データ分析のために工業値に変換する必要がありますか？もしそうなら、そのデータを取り出すSQLクエリーはかなり複雑になります。計測器とデータ収集の信頼性は100%ですか？それとも計測器のエラーを含むデータを「やりくり」する必要がありますか？汎用データベースは確かにデータを保存できますが、「データ品質」の概念を計算に組み込むようには設計されていません。また、「リットル/分」で測定された信号を、バッチや一日の「総リットル」に変換するといった、一般的に必要とされるルーチンの時系列計算を単純に実行することはできません。

思い込み 9

リレーショナルデータベースは 効果的なデータ収集ソフトウェアではない。

データ収集ソフトウェアは、リアルタイムのデータを継続的に収集し、そのデータの注目すべきサブセットを保存し、そのデータから意味のある情報を抽出する手段を提供するという、いくつかの関連する機能を備えています。データ収集ソフトウェアとはアプリケーション、いわばデータの応用・利活用を表すのに対し、リレーショナルデータベースは技術を表します。

時系列データにリレーショナルデータベース技術を使用することには確かに大きな課題がありますが、アプリケーションがリレーショナルデータベースを使用しているからといって、それがデータ収集ソフトウェアではないというわけではありません。根本的な技術の選択（リレーショナルデータベースか専用ファイルか）に注目するのではなく、必要な機能に注目してください。そのためには、アプリケーションの全体像の理解が必要ですし、単なるデータ保存以上の機能が求められます。

思い込み 10

すべてのデータは同じ重要性和有効性を備えている。

リレーショナルデータベースでは、保存された値はまさに「値」であり、常に有効であると想定されています。しかし、何千ものデータポイントから何百万ものサンプルをプロセスから収集する場合、そこに一部の情報に誤りや欠落があることは避けられません。測定器の問題で値が範囲外になっていたり、通信が途絶えていたり、データが誤っていたりすることがあります。

データ収集ソフトウェアでは、保存されたデータポイントには、値とタイムスタンプが関連付けられているだけでなく、データの品質を示す意味もあります。例えば、ある機器のデータポイントを、その機器の通常の動作範囲外で保存すると、その値とともに特定の連続の品質指標が保存されることとなります。これらの指標は、単にデータベースの別の列ではなく、サンプルの固有の特性です。これらの指標は、検索したり、計算に含めたり、オペレーションやエンジニアリングの担当者に異常の可能性を警告するために使用することができます。

値を集計する場合（例えば、過去1時間の平均気温を計算する場合）、データ収集ソフトウェアはこのデータ品質を計算結果に反映させ、オプションで疑わしいデータをフィルタリングし、データが欠けていたり無効と判断された場合には外挿することができなければなりません。このような現実世界の異常を正しく処理しないと、結果としてのレポート、ビジネスシステムの統合、意思決定が正しく行われなくなります。リレーショナルデータベースだけでは、このような機能はありません。

思い込み 11

リレーショナルデータベースにさえデータを保存すればデータ統合は簡単だ。

2つのExcelファイルを同じフォルダーに入れただけでは、たとえ生産データが含まれていたとしても、「統合」されたことにはならず何の意味もありません。

同様に、ERPシステムと生産にまつわる様々なデータを、リレーショナルデータベースに格納しても、「統合」にはなりません。確かに、すべてのデータを共通の技術で扱うことは必要な第一歩ですが、それだけのことであり、多くの場合、その後のデータ活用は思うように進みません。

思い込み 12

リレーショナルデータベースにデータを保存すると、簡単にデータを検索できる。

SQLクエリー機能は、リレーショナルデータベースを使用する大きな利点の1つですが、特定のデータベース設計が必ずしもSQLクエリーを簡単に実行できるとは限りません。中には、手動で作成したクエリーを効果的に使用できず、代わりにプログラムで生成した複雑なクエリーに頼らなければならないケースもあります。これは、ストレージの管理やポータビリティの観点からは意味のあることですが、リレーショナルデータベースを使用する本来の利点がほとんどなくなってしまいます。

思い込み 13

リレーショナルデータベースは専用のデータ収集ソフトウェアよりも低コストだ。

本格的なデータ収集ソフトウェアを導入する余裕がないと決断する前に、自分の真のニーズを理解し、幅広い選択肢を検討してみてください。実際のニーズに合わせたソリューションの価値と単純な購入価格を比較した場合、専用のソリューションの方がはるかに安価になる可能性があります。

思い込み 14

ヒストリアンが必要なのは大規模なプロセス制御だけである。

1980年代、石油プラントや製紙工場などのプロセス制御を対象とした専用のデータ収集ソフトウェアが登場しましたが、これらの業界は、そのシステムの構築のために高価なコンピューターシステムを導入できる唯一の業界でした。その後1990年代にWindows NTが急速に普及したことで、コンピューターのコストが大幅に下がり、AVEVAのIndustrial SQL Server（現在の「AVEVA Historian」）のような、新しいプラットフォーム専用開発された低価格のソリューションに門戸が開かれました。これらの新しいソリューションは、伝統的にDCSを重視するプロセス制御以外でも、その費用対効果がすぐに実証されました。

結論

リレーショナルデータベースだけでデータ収集ソリューションを構築するのは素晴らしいアイデアのように聞こえますが、そこには多くの固有の課題と限界があります。しかし、市販のソフトウェアが産業用環境に適していないというわけではありません。現在の生産情報は、工場環境の外と企業ネットワークの中の両方で必要とされています。そして、プラントデータと企業システム間のインターフェイスを提供するのに、SQLのような商業的に認められた標準ベースの技術以上の方法はありません。

AVEVA Historianは、市販の製品 (Microsoft SQL Server) を、プラントの履歴データに適合したオープンで標準的なクエリーインターフェイス (SQL) で拡張します。このインターフェイスは、IT部門がレポート作成やERPシステムへの統合のために簡単に使用することができます。

AVEVA Historian は、本稿で取り上げたすべての機能を提供します。全世界で120,000以上の導入実績があり、高い信頼性を誇るAVEVA Historianは、プラントオペレーションと企業のビジネスユーザーを同様に強化し、適切な情報を適切な人に提供し、データベース管理をプラントフロアではなく、企業のIT部門に任せることができます。

著者について

Elliot Middletonは、産業用ソフトウェアにおいて、主にプロセス向けデータ収集、統合化ビジネスシステム、BIアプリケーションで30年以上の経験を持っています。プロダクトマネージャーとして、AVEVA Historian、Historian Client、AVEVA InsightなどのAVEVAソフトウェア情報ソリューションのロードマップ策定を担当しています。顧客、ビジネスパートナー、AVEVA の開発およびサポートチームと協力しながら、市場の問題を特定し、製品の要件を定義し、機能拡張の優先順位を決定しています。2001年にAVEVAに入社する前は、AspenTech社およびIndX社で同様の役職に就いていました。バイラー大学でコンピューターサイエンスの理学士号を取得しています。

詳しくは、販売パートナーにお問い合わせください。



www.se.com/jp
www.proface.co.jp

AVEVA

Copyright © 2022 AVEVA Group plc およびその子会社がすべての著作権を保有します。
記載されているすべての製品名は、各所有者の商標です。

www.aveva.com/ja-jp/