

分散型システムのタイミング要件に関する白書

データストレージがネットワーク エッジに移行してレイテンシが抑えられ、アプリケーションが分散型システムモデルに移行して容量と速度が増すにつれ、明確な同期要件がデータセンター設計要素として不可欠になっています。大規模な分散データベースは今やペタバイトおよびエクサバイトの規模になり、数千 TPS(トランザクション / 秒) [1] を実行し、世界中の複数の拠点にデータの複製を保管しておく必要があります。複製間のデータの一貫性を管理する事で、データセンター内のサーバー間だけでなく数千マイル離れたデータセンター間でも同期させる必要があります。特定の分散型システム、アプリケーション、データベースに必要なタイミング精度のレベルは、タイミング アーキテクチャを計画する際に考慮する必要がある複数の要因によって決まります。本書ではその要因を複数検討し、クロック要件を定量化するための 1 秒あたりの時間エンベロープの概念を紹介し、分散型外部同期クロックを使った各種ソリューションについて説明します。

データの一貫性と高精度時間のための要件

データの一貫性とは、特定の時点においてレコードがデータベースまたはアプリケーションの全ての複製で同じ値になる必要がある事です。様々なニュアンスを持つ別名 (因果的一貫性、局所的一貫性、大域的一貫性、強力な一貫性、最終的一貫性) がありますが [2]、基本的な原理は全て同じです。特定のタイムスタンプの時点でデータベース全体で値が一貫している事です。たとえ実行後の照合によって一部の複製で値を変更して値を一貫させたに過ぎない場合であってもです。

1 台のコンピュータの複製されていないインスタンスにデータベースやシステム全体が格納されている場合、これは簡単な作業です。複数の複製を持つグローバルな分散型システムでも、通信の遅延がなく、絶対時間に各サーバが確実にアクセスできるのであれば、これはまだ簡単な作業です。しかし、通信は遅延する事があり、確実な絶対時刻は存在しません。物理的な距離、通信経路の技術、非対称性、各拠点でのクロックの不確実性等の問題は全て軽減策を必要とする課題を生み出します。

データベースとアプリケーションをネットワーク エッジにグローバルに分散させる事でユーザとデータベースの間の通信遅延 (レイテンシ) を最小化できます。例えば、バージニア州アッシュバーンのデータセンターに全ての複製が配置されているデータベースがあるとします。カリフォルニア州サンノゼのユーザがサーバと通信する必要があります。理論上、2 つのサイト間で情報を転送する際の最高速度は光速の 12.7 ms です。データベースから値を読み取るという単純な操作でもメッセージの往復が必要です。つまり、遅延は約 25 ms です。2 つの施設間を専用の光ファイバで結んだとしても、ファイバの屈折率、実際のファイバルート、システム内の他の遅延等の要因によって性能はさらに制限され、1 往復で約 75 ms から 100 ms かかります。

レイテンシを改善するにはデータベースの複製をサンノゼのサーバに配置できます。しかし、これでは一貫性の管理が複雑になってしまいます。同じような時間枠で両方の場所で同一または関連レコードが更新された場合、照合してどちらが実際に先に発生したのかを判断する方法が必要になります。レコードロックを適用するだけでは問題は解決しません。なぜなら、アッシュバーンがレコードをロックする必要性を認識したとしても、サンノゼがこのレコードがロックされた事を知るまでに少なくとも 12.7 ms はかかるからです。一貫性の問題を最低限に抑えるために、両方のデータセンターが外部クロックに同期している場合、そのクロックに従ってトランザクション (Note 1) にタイムスタンプを付け、実行後の照合アルゴリズムによって非一貫性を解決できます。

アルゴリズムが絶対的な確実性を以ってトランザクションを順序付ける事ができる理論上の最高の時間粒度は 2 つの場所の CU(クロック不確実性) のレベルの 2 倍に制限されます。システムが ± 100 ms の CU でタイムスタンプを付けられる場合でも、そのシステムが明確に判別できるのは 200 ms 以上のタイムスタンプを付けた 2 つのイベントの正しい順序だけです。

本書ではこの間隔を時間エンベロープと言います。図 1 の最初のグラフではエンベロープが重なっていないため、イベントの実際の順序を判別できますが、2 番目のグラフでは解決できない非一貫性が発生しています。エンベロープが重なっているため、どちらのトランザクションが先に発生したかを確実に判別する方法がありません。他のシステムの問題によってエンベロープのサイズが大きくなる事もありますが、技術の進歩に伴い、クロックの不確実性はイベントの順序を判別するために使う事のできる理想条件での最小時間エンベロープを常に決定します (線形化とも言います)。

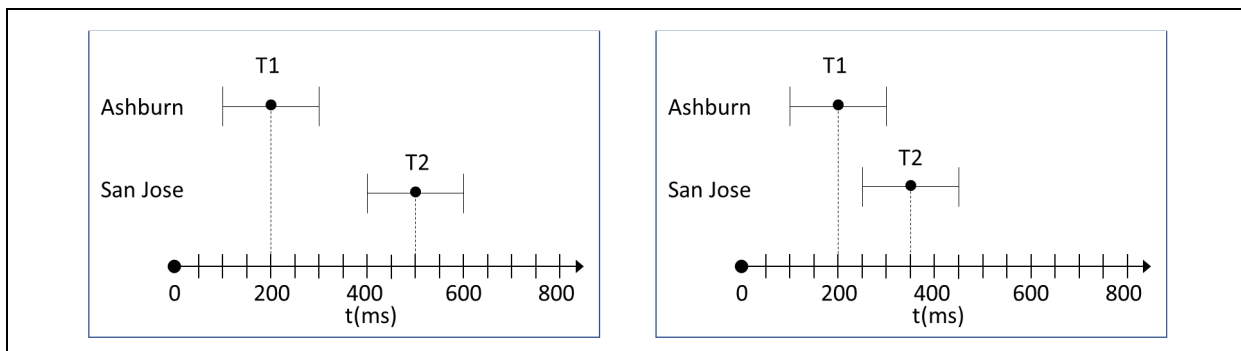


図 1: 最初のグラフは、トランザクションのタイムスタンプ (300 ms 間隔) が時間エンベロープ (200 ms) よりも大きいため、解決可能な一貫性の例です。2 番目のグラフは、トランザクションのタイムスタンプ (150 ms 間隔) が時間エンベロープ (200 ms) より短いため、解決できない非一貫性を示します。

関連レコードまたは因果関係のあるレコードの各集合を固有の色で可視化します。異なる色のエンベロープは重なってもかまいませんが、同じ色のエンベロープは重なってはなりません。時間エンベロープが重なる複数のトランザクションが異なる無関係なレコードで動作する場合、実際の時間順序を解決できなくても非一貫性は生じません。これを図 2 に示します。この図では R4 と R5 のトランザクションのみが問題です。しかし、エンベロープの期間が長くなるにつれ、1 つの時間エンベロープで 2 つの同一または関連レコードが影響を受けて、トランザクションがロールバックされた時にシステムの性能が低下する可能性が高くなります。コンセンサス方式、確率分析、特定の複製に優先順位を付ける方法を使うと片方のトランザクションだけをロールバックできますが、ロールバックのたびにシステムが非効率になり、ロールバック回数が管理不能なレベルに達した結果、レコード間の同期が破綻する可能性があります。コストとのトレードオフを管理する必要があります。

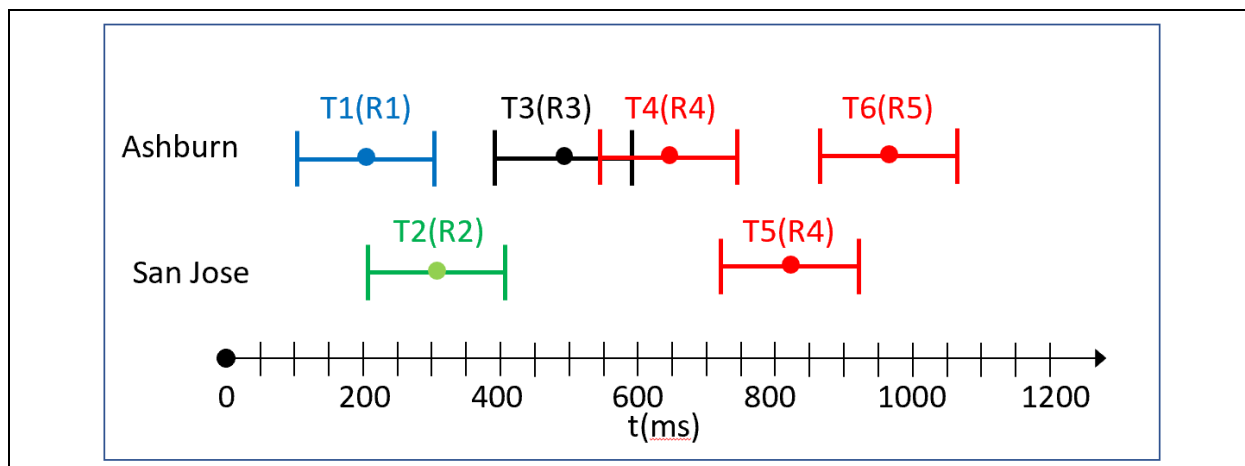


図 2: R1、R2、R3、R4 が完全に独立したレコードで互いに影響し合う事がない場合、順序は重要ではないため、エンベロープの重なりを許容してもリスクは最低限に抑えられるかあるいは全くありません。R4 の操作 (アッシュバーンで 650 ms、サンノゼで 825 ms) は同じレコードのコピーに対するもので、時間エンベロープが重なるため、問題が発生します。R5 がレコード R4 の関数 (例えば $R5 = 2R4 + 6$) の場合、812 ms と 975 ms の 2 つのトランザクションは因果関係のあるレコードが重なったものであるため、問題になります。

タイムスタンプと時間エンベロープを使っても、ソフトウェアに可能な操作は非一貫性を事後的に調整する事だけです。CU = ± 2 ms でサイト間のレイテンシが 12.7 ms のシステムでは、実行後のアルゴリズムによってのみ一貫性を確保できます。実行後の同期にかかる全般的な負担を最小限に抑えるため、現地用のシャードにデータベースの一部を保存し、複数の場所で需要が発生する確率の高いレコードのみを複製する事もあります。

システム要件と機器仕様の間の関係

システム設計者は通常 TPS を指定します。クロックベンダーは CU を指定します。上記の説明から非一貫性が生じない簡単な式を導出できます。上記の ± 100 ms のシステムの場合、重ならない時間エンベロープを最大 5 つ作成できます (図 3 参照)。

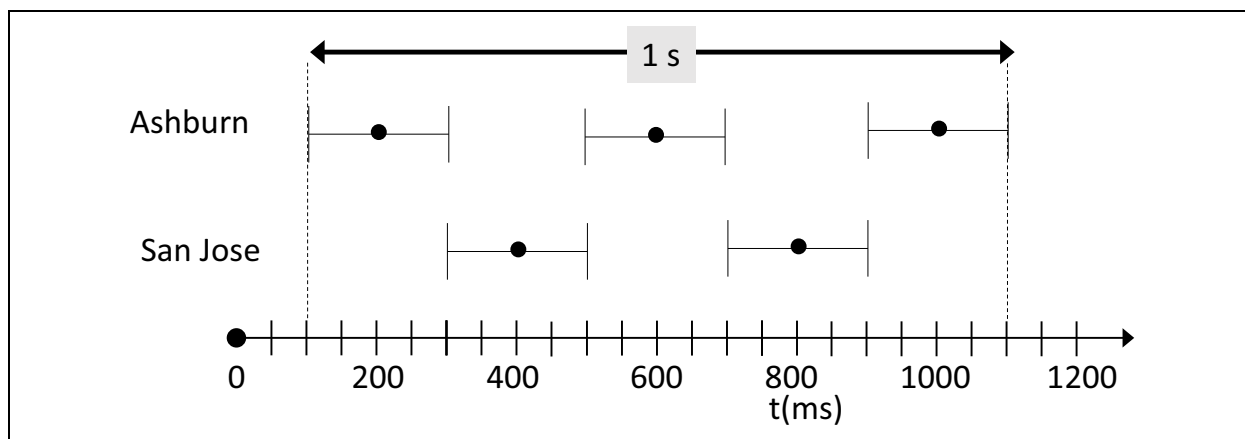


図 3: 不確実性 ± 100 ms のクロックの場合、明確な分解能のための各時間エンベロープは 200 ms です。この結果、1 秒間に最大 5 つの時間エンベロープが発生します。

ここから、TPS の要件を CU の関数として定量化して解決不可能な非一貫性をなくしたのが式 1 の不等式です。

式 1:

$$TPS < \frac{1}{2CU}$$

式 1 はシステムにとって正しいクロックを決定するための基礎となる原理です。しかし、実際のアプリケーションを考慮して TPS と CU を修正し、クロック要件を (システムの劣化を招く) アンダースペックにする事も (システムに不必要なコストをかける) オーバースペックにする事も避ける必要があります。

TPS に関する注意事項

1. **平均 TPS とピーク TPS。**一部のデータベースは書き込みコマンドを一定のレートで実行します。製造ライン用にセンサの値を指定の間隔で記録する産業用プロセス制御データベースは一定レートのシステムの例です。ピークトランザクション レートがあるアプリケーションもあります。金融取引アプリケーションでは、市場が開いた直後と閉まる直前がピークとなる場合があります。これらの期間では TPS が平均 TPS の 100 倍になる事もあります。多くのシステム仕様書では 1 日あたりのトランザクション数を 86,400 秒で割った値を平均 TPS としています。ただし、CU は 1 日平均のレートではなくピークのレートで設計する必要があります。
2. **解決不可能な非一貫性の重大性。**多くのアプリケーションでは、非一貫性がユーザに与える影響は煩わしさを除けばほとんどないかもしれません。写真のメタデータを含むクラウドベースの写真アプリケーションであれば、少数のキーワードまたはタグのロールバックを余儀なくされる非一貫性があっても大きな影響をおそらく受けないでしょう。しかし、アプリケーションによっては、多額の金銭的または法的コスト、規制基準 [3、4] が発生するものもあり、救急医療サービスや軍の判断に影響を及ぼすアプリケーションであれば人命への影響も懸念されます。
3. **因果関係のある 2 つのレコードが同じ時間エンベロープに異なる場所で操作される確率。**分散型データベースが顧客住所を格納し、各地域が現地用のシャードで顧客住所を管理している場合、東京にいるユーザがロンドンの顧客住所をニューヨークにいるユーザと同時に更新する確率はゼロです。一方、複数のレーダ画像を統合してまとめておく軍事的なアプリケーションがあるかもしれません。レーダが設置されている各拠点は全体の航空写真に寄与します。複数のレーダが高速移動する同一のターゲットを追跡している場合、システムはほぼ同時にデータを報告する可能性があるため、高速移動ターゲットに関する同時書き込みの確率を下げるためにエンベロープサイズを最小にする必要があります。

これらの各注意事項に対応するためにスケール係数を使って式 1 を変更できます。

式 2:

$$S_{pk} \times S_c \times S_{pr} \times TPS < \frac{1}{2CU}$$

S_{pk} = 平均値に対するピーク値に使うスケール係数。この値はピーク TPS を平均 TPS で除算する事で直接求められます。

S_c = 非一貫性の重大性に使うスケール係数。アプリケーションに基づく主観的なリスク値です。非一貫性を許容しないシステムは 1 より大きい S_c を選択できます。その場合、より小さな CU が必要です。非一貫性を許容しても影響が少ないシステムでは S_c 値 0.01 を選択できます。これにより、CU 要件が緩和されます。

S_{pr} = 同一エンベロープ内の同一または因果関係のあるレコードに書き込む確率に使うスケール係数。上記の顧客データベースのように地域的に分離されているデータが一般的なアプリケーションでは、このスケール係数がきわめて低くなる可能性があります。つまり、0.001 という低いスケールを使う可能性があります。一方、発生確率が高い場合はスケール係数を 1 に設定すべきです。

WTPS(重み付けした TPS) は、3 つのスケール係数と TPS を 1 つの変数に統合するために使う事ができます。

表 1 に、複数の仮想システム要件と WTPS の評価プロセスを示します。

表 1: アプリケーションと WTPS 要件

ユースケース	解決不可能な非一貫性の重大性	短時間に異なる場所で関連レコードが更新される確率	平均 TPS	S_{pk}	S_c	S_{pr}	WTPS
軍事作戦のデータベース	高: 誤った報告データは人命の損失につながる可能性がある。	高: 統合された戦場の作戦は複数の情報源から更新される。	10,000	10	10	1	1,000,000
金融取引企業	高: 取引の損失は経済的に大きな影響を及ぼす可能性がある。	高: 特定の時間帯に発生する大量の取引要求と株式ニュースによって個別のターゲットが生じる。	1,000	10	10	1	100,000
多国籍銀行のデータベース	高: トランザクションが失われると法規制の問題が生じる可能性がある。	中: 複数の異なる場所で口座にアクセスする必要がある。	1,000	1	10	0.1	1,000
医療データベース	高: 誤った報告データは人命の損失につながる可能性がある。	低: 個々の患者データを複数の異なる場所で同時にやり取りする事はほぼない。	1,000	1	10	0.01	100
産業プロセス制御	中: 欠落した少数のデータポイントは平均化される可能性が高い。	低: センサからデータを一定の間隔と特定の場所で取得する。	10,000	1	1	0.01	100
ソーシャルメディアの写真アプリケーション	低: ユーザ体験に影響するものの、再入力する事で解決できる。	低: ほとんどの写真のメタデータはアカウントの所有者によってのみ更新される。	10,000	10	0.01	0.01	10
営業連絡先データベース	低: 最悪の場合でもユーザに必要な作業はデータの再入力である。	低: 更新は複数の異なるサイトで発生する可能性があるものの、近接した時間帯に更新される可能性はそれほど高くない。	100	1	0.1	0.01	0.1

CU に関する注意事項

式 1 の右辺 $1/2CU$ は、クロックの不確実性に基づいて利用可能な時間エンベロープの数を表しています。この量を TPS と比較するためのより直感的な方法が EPS (1 秒あたりのエンベロープ) という用語を使う事です。TPS を WTPS 値に置き換えて EPS を使うと、式 1 を式 3 に書き換えられます。

式 3:

$$WTPS < EPS$$

この式を使うと、より単純に要件を可視化できます。

複数のサイトの EPS を一貫させるためには、外部クロックに同期する必要があります。通常、UTC が外部クロック源として使われ、トランザクションが発生する機器に到着するまでに様々な方法で配信されます。図 4 に、データセンター機器に時間を配信するための最も一般的な 2 つの経路を示します。

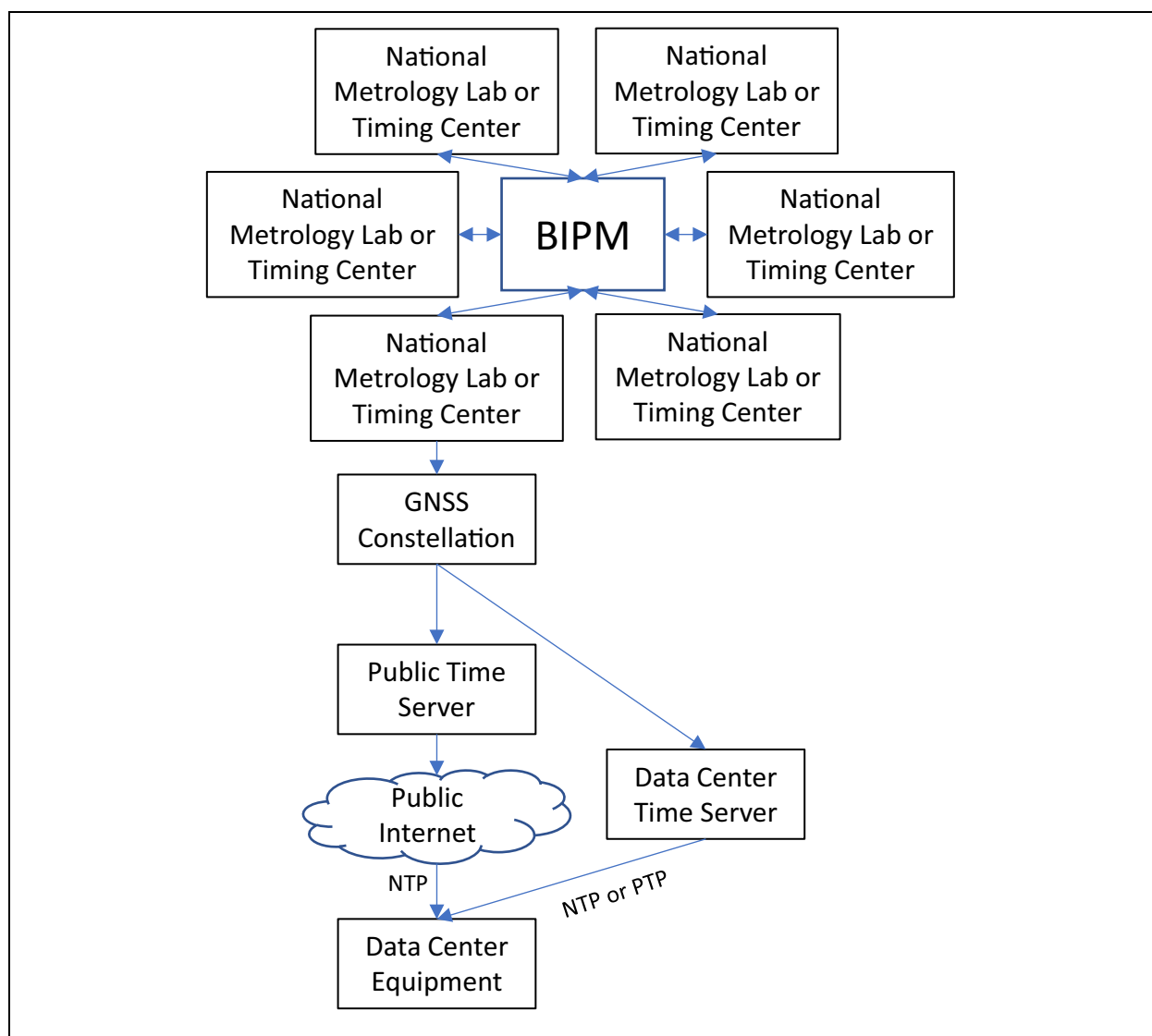


図 4: データセンターのタイミングで一般的な経路。どちらの経路でも、時間は BIPM が発行する UTC から生成されて GNSS コンステレーションに送信されます。次に、パブリックまたはプライベートタイムサーバがパブリック インターネットまたはローカルエリア ネットワークを通じてデータセンター機器に時間を配信します。

図 4 ではどちらの経路も GNSS ネットワークを使って UTC に同期します。左の経路は、GNSS ネットワークとデータセンター間のステップにパブリック ネットワークのタイムサーバを使っています。パブリック NTP (Network Time Protocol) サーバはインターネット上でパケットベースのメッセージを介して UTC 時間をパブリックな領域に無料で配信します。各種サーバ用に公開されている CU には 100 μ s から 10 ms の幅があります。これは、高い EPS を必要とせず (50 未満)、セキュリティ上の懸念が重大ではなく、精度に対する規制要件が比較的緩い存在せず、異常なタイムスタンプが引き起こす解決不可能な非一貫性のリスクを許容できる分散型システムの同期に使う外部クロックとしては許容できます。しかし、パブリック NTP サーバを利用する事には大きなリスクがあります。

- 精度に関するリスク：
 - インターネットのタイムサーバで時間が正しい事を保証できない。
 - 多くのインターネットタイムサーバにおいてクロックソースまたはクロックソースの変更の情報が開示されていない。
 - ローカルエリア ネットワーク内からファイアウォールを経由してパブリック タイムサーバに到るまでの遅延が非対称になる。
- 信頼性に関するリスク：
 - フォルト発生時にインターネットのタイムサーバは SNMP トラップまたはアラームを送信しない。
 - インターネットのタイムサーバに対するリクエストの量は大きく変化し、トラフィックが多い時間帯には精度がさらに低下する可能性がある。
 - インターネットのタイムサーバからの持続的な応答を保証できない。
- セキュリティに関するリスク：
 - 双方向トラフィックのためにファイアウォール ポート 123 を開放する必要がある。
 - パブリック タイムサーバのタイムソースの制御にリスクがある。
 - インターネットのタイムサーバ用の認証方法がないため、中間者がタイムスタンプを操作する可能性がある。

図 4 の右側の経路はパブリック NTP とインターネットをバイパスする経路です。これにより、上記のリスクを全て排除するだけでなく、より高い EPS 値を実現できます。プライベート タイムサーバは GNSS コンステレーションから時間を直接取得し、プライベート NTP サーバまたは PTP (Precision Time Protocol) グランドマスタとして LAN 経由でデータセンター内に時間を配信します。Microchip 社はこれらのニーズに応えるために各種フォームファクタのソリューションを提供しています。Microchip 社製 SyncServer® S600 High Performance NTP Time Server [5] と TimeProvider® TP4100 PTP Grandmaster [6] はどちらも 1U ラックマウント ソリューションで正確な時間を提供し、セキュアなローカルエリア ネットワークに配信するソリューションです。これらの機能豊富なソリューションは GPS、Galileo、GLONASS、Beidou、その他のコンステレーションから時間を生成する 72 チャンネルの GNSS (グローバル ナビゲーション衛星システム) レシーバを実装しています。



図 5: Microchip 社製 TimeProvider TP4100 PTP Grandmaster と SyncServer S600 High Performance NTP Time Server。

どちらの製品もローカル タイムソースを提供し、ネットワークを介してパケットベースのメッセージでこの時間を配信します。ただし、ローカル タイムソースは実機のタイムスタンプの最終的な精度に寄与する一要素に過ぎません。これ以外にタイムスタンプの精度に影響する要因として、タイムサーバとタイムクライアント間のルータとスイッチの数、タイムスタンプに使うハードウェアとソフトウェアがあります。全ての要因を考慮すると、システムは最終的にタイムソースからのタイムスタンプ精度をさらに 20 ns から 5 μ s 低下させる可能性があります。システムの精度低下を含めると、式 3 の右辺は式 4 のように書き換えられます。

式 4:

$$EPS = \frac{1}{2(CU_C + CU_N)}$$

CU_C = 専用クロック リファレンスのクロックの不確実性。

CU_N = ネットワーク機器等のハードウェアのクロックの不確実性。

多くのシステムは最終的なタイムスタンプの精度をマイクロ秒以下の範囲にし、バージニア州アシュトン、カリフォルニア州サンノゼ、オーストラリアのシドニーにあるサーバと分散アプリケーションの全ての機器で生成されたタイムスタンプが UTC に対してだけでなく互いに対してマイクロ秒以下の精度を備える事を目指しています。

S600 と TP4100 は ITU の PRTC(プライマリ リファレンス タイムクロック) 要件に対応するように設計されています。ITU の仕様は代表値または RMS 値ではなく最大値です。表 2 に、最先端のローカルエリア タイム ネットワークに対する S600、TP4100、パブリック NTP の比較を示します。他の全ての不正確性のソースはタイム クライアントでの最終的なタイムスタンプ精度に 100 ns 寄与しています。この表には関連する EPS も記載されています。

表 2: 各種外部クロック ソリューションの性能比較

参照 クロックの タイプ	性能を実現する技術	UTC に対する ソースの 精度	先端 ネットワーク タイム スタンプ精度	最先端 EPS	S600	TP4100
パブリック NTP	パブリック インターネットを介した パケットベースの時間	2.00E+06 ns	2.00E+06 ns	250	–	–
PRTC-A	シングルバンド GNSS レシーバ	100 ns	200 ns	2.50E+06	X	X
PRTC-B	シングルバンド GNSS レシーバ	40 ns	140 ns	3.57E+06	–	X
ePRTC	マルチバンド GNSS レシーバと セシウム原子時計	30 ns	130 ns	3.85E+06	–	X

データシートにソース精度が 15 ns RMS 未満と記載されている場合、仕様の解釈に関する一般的な問題が発生します。これは標準偏差 1 つ分に相当する RMS 値であり、表 2 または ITU 仕様に記載されている最大値ではありません。

ラックマウント ソリューションに加え、Microchip 社は小型ソリューションも提供しています。例えば、Tekron ブランドの TTM 01-G DIN-rail mounted Compact GNSS Clock [7] は 32 チャンネルの GPS および GLONASS レシーバを備えた NTP および PTP 機能を提供しています。本製品は完全な PRTC-A 対応ではありませんが、100 ns 未満の代表値を提供できます。



図 6: Microchip Tekron TTM 01-G DIN Rail-Mounted Compact GNSS Clock

式 3 の左辺 (WTPS) はシステム要件です。式 3 の右辺 (EPS) は機器の容量です。各エンベロープには重み付けをしたトランザクションを 1 つしか入れられません。表 1 に記載した各種アプリケーションの WTPS と表 2 に記載した機器の EPS の値を比較したグラフを図 7 に示します。システムを設計する場合、特定のアプリケーションを示す赤いバー (WTPS) はシステムの容量を示す黒いバー (EPS) より小さくなければなりません。このグラフから、軍事、金融、取引のアプリケーションではパブリック NTP は十分に高い EPS を提供できない事が分かります。その他の 4 つのアプリケーションではパブリック NTP が十分な EPS を提供します。ただし、セキュリティまたはその他の理由からパブリック NTP は不適切な選択肢かもしれません。

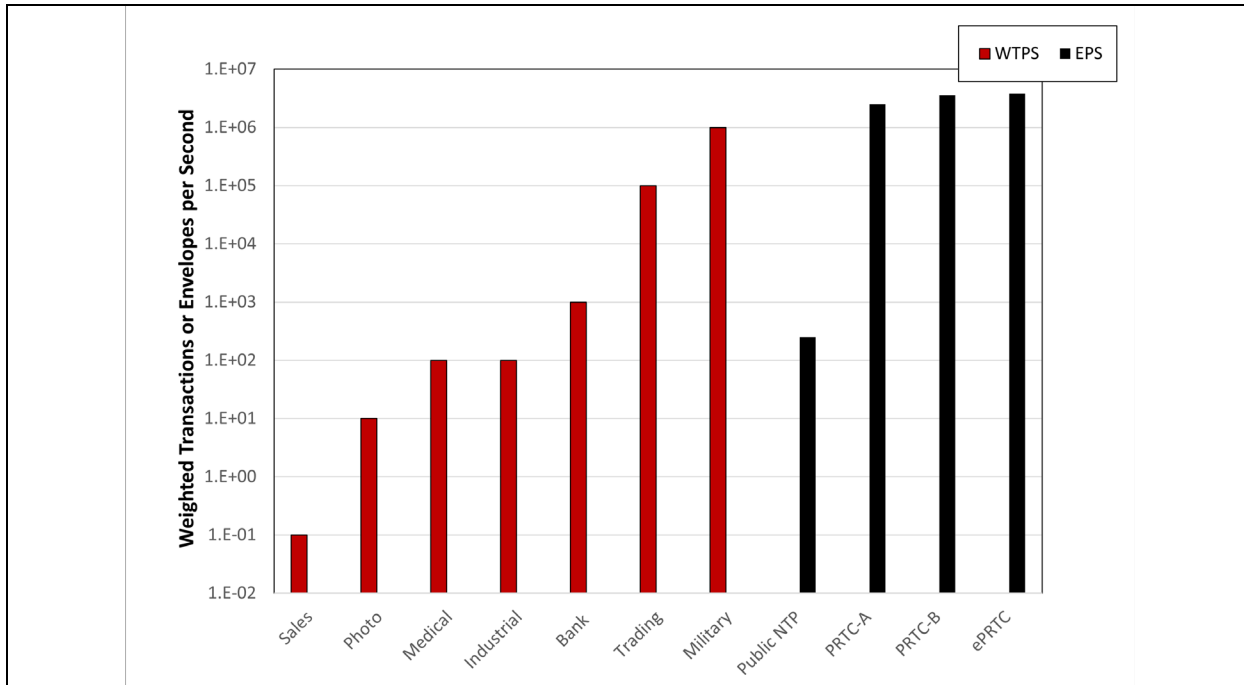


図 7: アプリケーションの WTPS 要件と機器の EPS 容量の比較

外部の高精度時間セキュリティ ソリューション

上記の全てのクロックは UTC への時間トレーサビリティを提供するために GNSS レシーバに依存しています。GNSS 信号は非常に弱いため、妨害の影響を受けやすくなっています。これは、他の信号のエネルギーによってレシーバが GNSS 信号を読み取れない場合に起こります。妨害信号には、悪意のある個人が意図的に発生させるものと類似の周波数帯で動作する信号が意図せずに発生させるものがあります。GNSS 信号は、悪意を持ってシミュレートした GNSS 信号を送信する事で意図的になりすます事も可能です。GNSS のなりすましと妨害は、米国の CISA (Cybersecurity and Infrastructure Security Agency) によって重要なインフラストラクチャに対する大きな脅威と認識されています。2020 年に発令された大統領行政命令 13905 号 [8] は、米国の連邦機関や重要なインフラストラクチャ事業者に対してなりすましと妨害に関するリスクを軽減するための措置を講じるように求めています。妨害となりすましは、そのための技術の普及に伴って発生頻度が高まっています。GNSS 拒否とも呼ばれるこれらのイベントの間、NTP サーバまたは PTP グランドマスタは UTC から離れるため、クロックの不確実性が増大します。

なりすましを発見できずにいる場合、第三者がデータセンターの時間を意図的に操作し、データへの不正アクセスおよびデータの再シーケンスが可能になるかもしれず、セキュリティに関するリスクがさらに増大します。

Microchip 社の Bluesky™ GNSS ファイアウォール [9] ハードウェア ソリューションはなりすましと妨害を検出し、シミュレートした GPS 信号をこれらの停止中に提供できます。または、時間がネットワークに送信される前にレシーバから信号を消去します。また、これらのイベントの間、代替のタイムソースを通過するように設定する事も可能です。



図 8: Microchip 社の Bluesky™ GNSS ファイアウォール

このハードウェア ソリューションに加え、TP4100 と S600 は最近リリースされた「Bluesky Inside」ソフトウェアを含む同種の妨害およびなりすまし防止ソフトウェアを同梱しています。

外部の高精度時間回復ソリューション

外部の高精度タイムソースをデータセンターへセキュアに配信する場合、EPS は外部のタイムソースの精度によって決まります。ただし、外部のタイムソースが失われると、そのデータセンターは内部にあるクロックまたは外部の代替タイムソースに頼らざるを得なくなります。外部クロックは様々な理由で失われます。パブリック NTP の場合は

インターネットの停止、GNSS ベース時間の場合は太陽フレアもあります。また、正しくグランドされていないアンテナから送信した信号が意図せずに近くのアンテナの GPS 信号を妨害する等の単純な原因もあります。機能が停止すると、ローカル タイミングが外部タイムソースからずれるため、不確実性が高まって EPS が低下します。このため、解決不可能な非一貫性が発生する確率が急速に上がり、サイトが同期されなくなるためにグローバル データベース全体が直ちに危険にさらされる可能性があります。タイミングがずれる量は現在のローカル クロックソースの安定性によって異なります。Microchip 社は S600 と TP4100 用に多数の水晶およびルビジウムベースの内蔵クロック オプションを用意し、タイムソースの代替時間基準として使う事ができる 5071A [10] とタイムセシウム 4310B [11] を含む外部セシウム原子標準時計を提供しています。



図 9: Microchip 社の 5071A および 4310B セシウム原子時計

図 10 に、GNSS にロックした状態の各種クロックタイプを備えた複数の Microchip 社ソリューションの EPS、GNSS 拒否から 24 時間後の EPS 劣化、100 ns の不確実性をタイムスタンプに加える最先端タイムネットワークでの GNSS 拒否から 1 週間後の EPS 劣化を示します (Note 2)。ローカル水晶振動子オシレータのバックアップを使った精度 2 ms のパブリック NTP も参考用に追加しました。このグラフは、GNSS にロックされている場合、Microchip 社の全てのソリューションがパブリック NTP と比較して 3 桁高い EPS の改善を実現している事を表しています。ローカルクロックの違いが EPS に与える影響も明らかです。ローカル TCXO はロックしない場合、直ちにパブリック NTP レベルの性能に落ちますが、セシウム原子時計では EPS は丸 1 週間ほとんど劣化しません。データセンターの開発者は EPS の減少の影響を判別し、適切なリファレンスで設計する必要があります。それに伴い、多くの開発者は GNSS 拒否の間でもパブリック NTP または他のバックアップ外部タイムソースを選択できます。

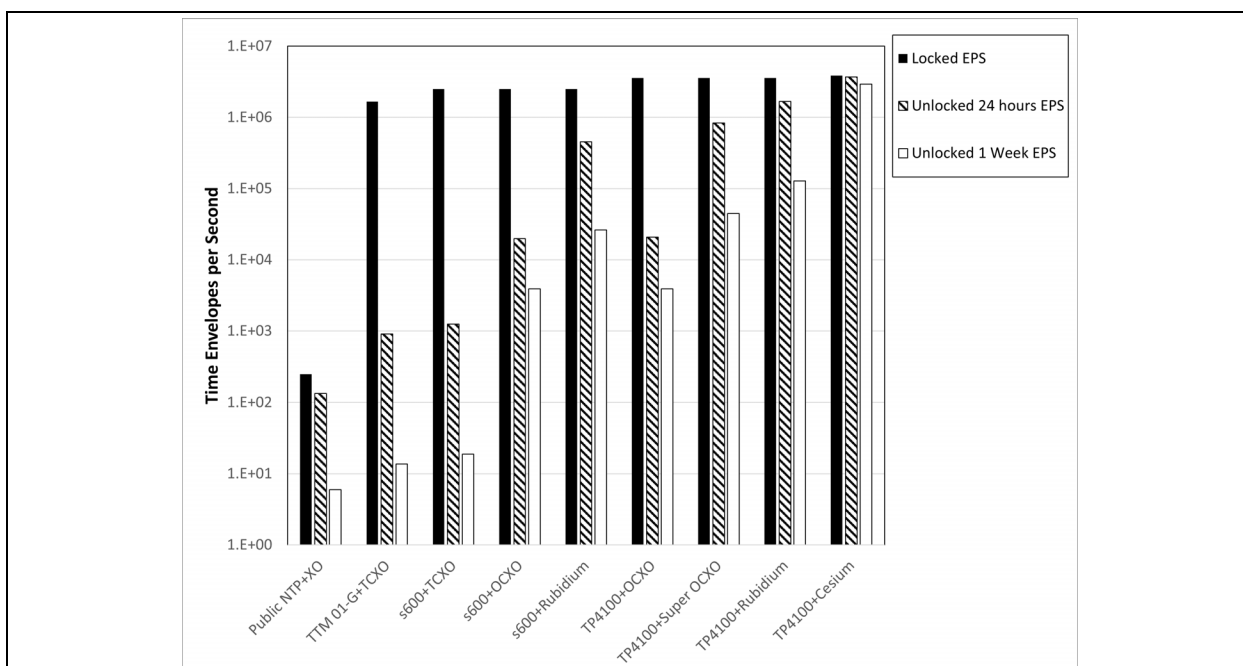


図 10: GNSS 同期中および GNSS 拒否中の各種外部クロック オプションの EPS

図 11 と図 12 は、アンロック (GNSS 拒否) 状態での 24 時間と 1 週間の動作におけるアプリケーションの WTPS と EPS の比較です。どちらのグラフも、アンロック状態でアプリケーションの WTPS 要件を確実に満たすにはローカルオシレータの選択が重要であることを示しています。GNSS がたった 1 日なくなっただけでも、銀行アプリケーションでは少なくとも OCXO が必要になり、取引および軍事アプリケーションではルビジウムが必要になります。1 週間後、ほとんどのアプリケーションでは原子時計が必要になります。

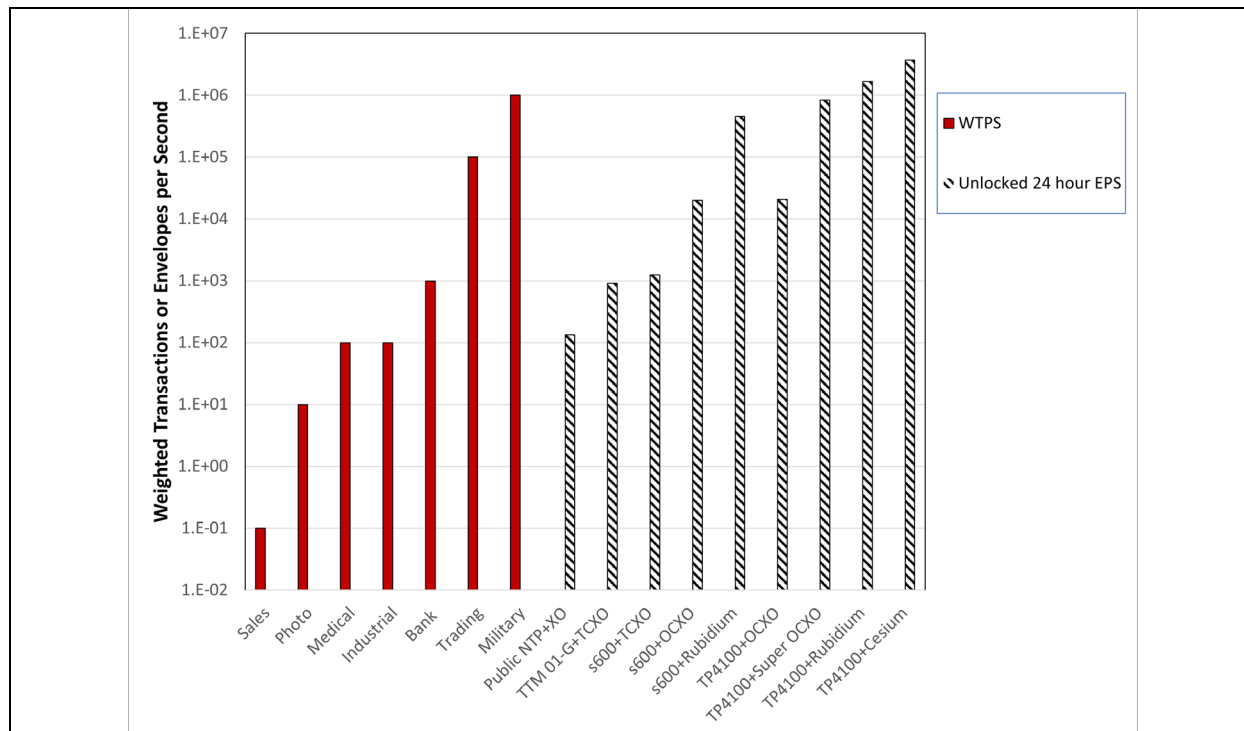


図 11: アプリケーションの WTPS 要件と 24 時間アンロック状態にある機器の EPS 能力の比較

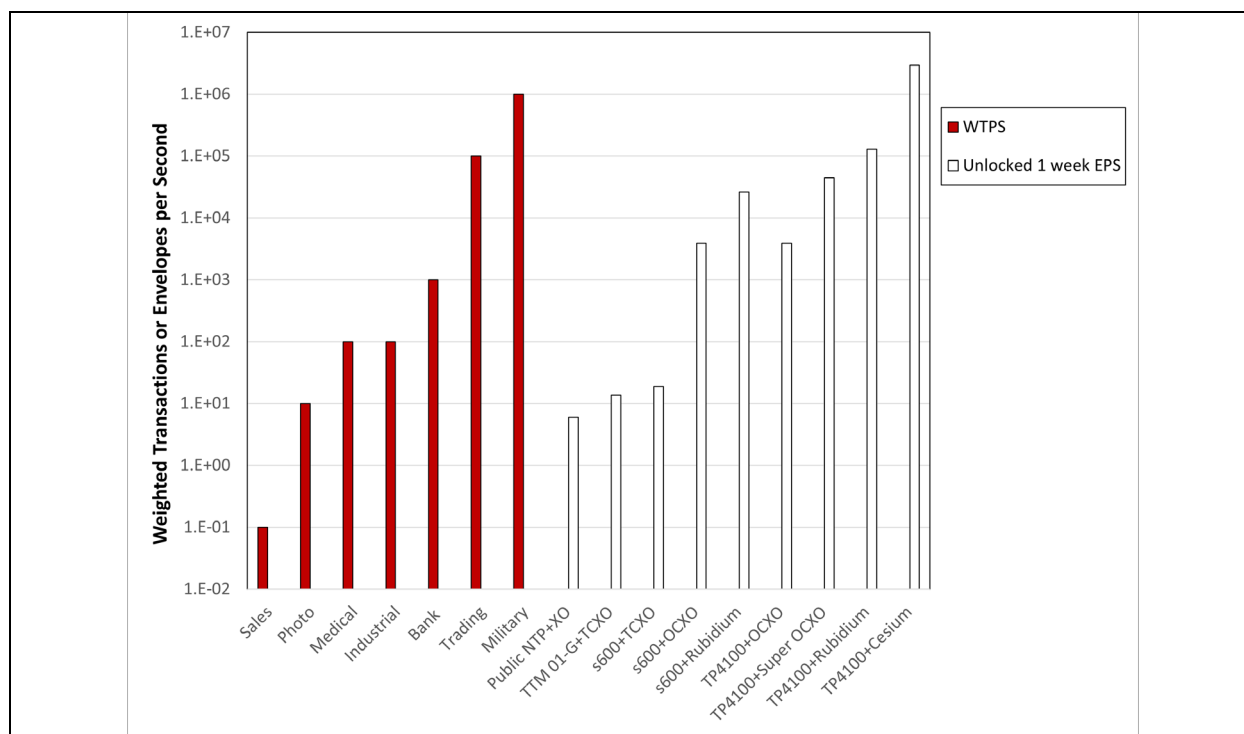


図 12: アプリケーションの WTPS 要件と 1 週間アンロック状態にある機器の EPS 能力の比較

仮想プライマリ リファレンス クロック

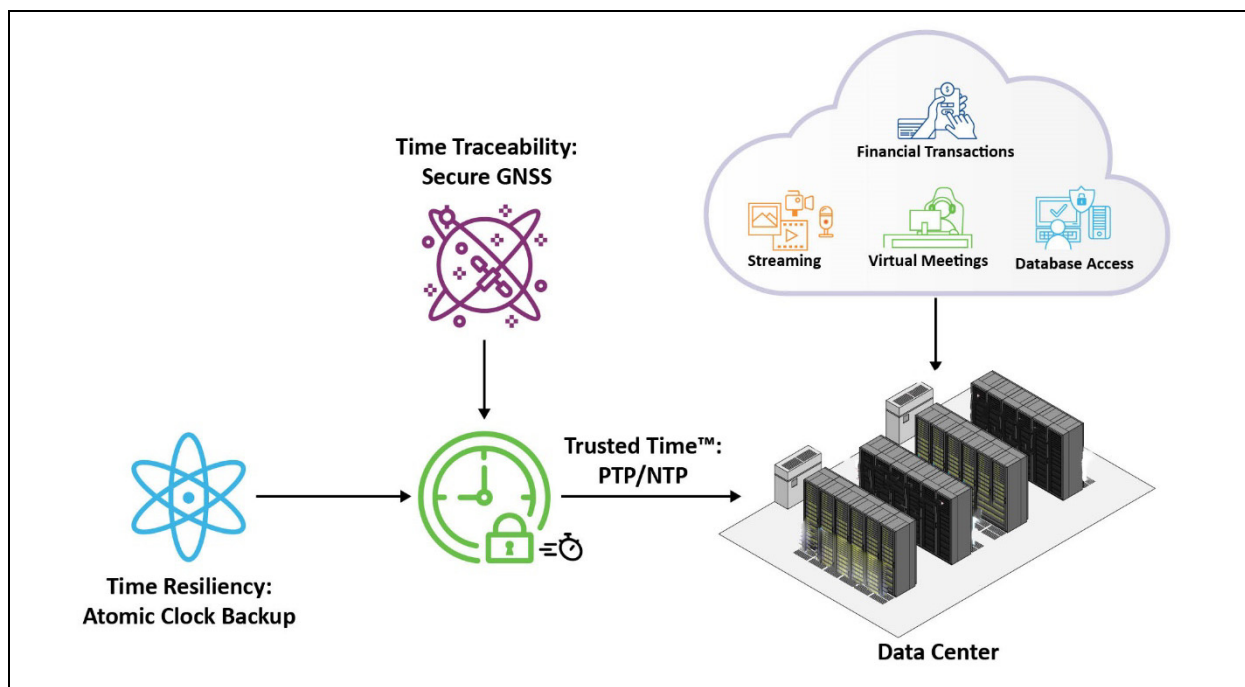


図 13: 仮想プライマリ リファレンス クロックの概念図

大容量と速度が要求される分散型システムモデルの要求に応えるため、Microchip 社は正確でセキュアかつレジリエントな時間をターンキー ソリューションで配信する「仮想プライマリ リファレンス クロック」を開発しました [12]。S600 または TP4100 タイムサーバを内蔵ルビジウム クロックまたは外部セシウムクロックに組み合わせ、Bluesky Firewall ハードウェアまたはソフトウェアで保護する事で高精度タイムソースを各データセンターに配置して全般的な EPS を高め、外部タイムソースが 1 日または数週間失われた場合でも重大な一貫性リスクなしに数百万の TPS を発生させる事ができます。

まとめ

世界中に分散したデータをサポートするための適切なタイミング アーキテクチャを決定する作業は、TPS(トランザクション / 秒)、レプリケーション間のデータの解決不可能な非一貫性の重大性、短時間に複数の場所でレコードまたは関連レコードが更新される可能性等、複数の要因によって決まります。これらの要因を分析すれば、WTPS を決定して適切な機器を指定し、データ内の必要なレベルの一貫性を保証できます。

Note

Note 1: 本書はトランザクションが1つの操作で構成されている事を前提としています。ほとんどのデータベーストランザクションは複数の操作を伴います。本書の原理は有効ですが、トランザクションの操作回数を考慮して値をスケールリングする必要があります。

2: クロックの精度は温度が一定である事を前提としています。値は公表値、実証データ、モデルによる予測値から採用しました。アンロック値は、データシートに指定した期間、クロックが安定した状態で UTC を参照している事を前提としています。

参考資料

1. <https://storage.googleapis.com/pub-tools-public-publication-data/pdf/65b514eda12d025585183a641b5a9e096a3c4be5.pdf>
2. <https://www.cs.princeton.edu/courses/archive/spr11/cos461/docs/lec24-strong.pdf>
3. <https://www.esma.europa.eu/document/guidelines-transaction-reporting-order-record-keeping-and-clock-synchronisation-under-mifid>
4. <https://www.nyse.com/publicdocs/nyse/markets/nyse/rule-interpretations/2017/NYSE%20Number%2017-02%203.8.pdf>
5. <https://www.microsemi.com/product-directory/enterprise-network-time-servers/4117-syncserver-s600>
6. <https://www.microsemi.com/product-directory/carrier-grade-ntp-ptp-ieee-1588-grand-masters/4422-timeprovider-4100>
7. <https://tektron.com/products/compact-gnss-clock-ttm01g/>
8. <https://www.federalregister.gov/documents/2020/02/18/2020-03337/strengthening-national-resilience-through-responsible-use-of-positioning-navigation-and-timing>
9. <https://www.microsemi.com/product-directory/gps-instruments/4398-bluesky-gps-firewall>
10. <https://www.microsemi.com/product-directory/cesium-frequency-references/4115-5071a-cesium-primary-frequency-standard>
11. <https://www.microsemi.com/product-directory/cesium-frequency-references/4114-csiii-model-4310b>
12. <https://www.microchip.com/en-us/products/synchronization-and-timing-systems/virtual-primary-reference-time-clock>

Microchip 社製品のコード保護機能について以下の点にご注意ください。

- Microchip 社製品は、該当する Microchip 社データシートに記載の仕様を満たしています。
- Microchip 社では、通常の条件ならびに動作仕様書の仕様に従って使った場合、Microchip 社製品のセキュリティ レベルは、現在市場に流通している同種製品の中でも最も高度であると考えています。
- Microchip 社はその知的財産権を重視し、積極的に保護しています。Microchip 社製品のコード保護機能の侵害は固く禁じられており、デジタル ミレニアム著作権法に違反します。
- Microchip 社を含む全ての半導体メーカーで、自社のコードのセキュリティを完全に保証できる企業はありません。コード保護機能とは、Microchip 社が製品を「解読不能」として保証するものではありません。コード保護機能は常に進化しています。Microchip 社では、常に製品のコード保護機能の改善に取り組んでいます。

本書および本書に記載されている情報は、Microchip 社製品を設計、テスト、お客様のアプリケーションと統合する目的を含め、Microchip 社製品に対してのみ使う事ができます。それ以外の方法でこの情報を使う事はこれらの条項に違反します。デバイス アプリケーションの情報は、ユーザの便宜のためにのみ提供されるものであり、更新によって変更となる事があります。お客様のアプリケーションが仕様を満たす事を保証する責任は、お客様にあります。その他のサポートは Microchip 社正規代理店にお問い合わせ頂くか、<https://www.microchip.com/en-us/support/design-help/client-support-services> をご覧ください。

Microchip 社は本書の情報を「現状のまま」で提供しています。Microchip 社は明示的、暗黙的、書面、口頭、法定のいずれであるかを問わず、本書に記載されている情報に関して、非侵害性、商品性、特定目的への適合性の暗黙的保証、または状態、品質、性能に関する保証をはじめとするいかなる類の表明も保証も行いません。

いかなる場合も Microchip 社は、本情報またはその使用に関連する間接的、特殊的、懲罰的、偶発的または必然的損失、損害、費用、経費のいかににかかわらず、また Microchip 社がそのような損害が生じる可能性について報告を受けていた場合あるいは損害が予測可能であった場合でも、一切の責任を負いません。法律で認められる最大限の範囲を適用しようとも、本情報またはその使用に関連する一切の申し立てに対する Microchip 社の責任限度額は、使用者が当該情報に関連して Microchip 社に直接支払った額を超えません。

Microchip 社の明示的な書面による承認なしに、生命維持装置あるいは生命安全用途に Microchip 社の製品を使う事は全て購入者のリスクとし、また購入者はこれによって発生したあらゆる損害、クレーム、訴訟、費用に関して、Microchip 社は擁護され、免責され、損害をうけない事に同意するものとします。特に明記しない場合、暗黙的あるいは明示的を問わず、Microchip 社が知的財産権を保有しているライセンスは一切譲渡されません。

Microchip 社の品質管理システムについては www.microchip.com/quality をご覧ください。

商標

Microchip 社の名称とロゴ、Microchip ロゴ、Adaptec、AVR、AVR ロゴ、AVR Freaks、BesTime、BitCloud、CryptoMemory、CryptoRF、dsPIC、flexPWR、HELDO、IGLOO、JukeBlox、KeelQ、Kleer、LANCheck、LinkMD、maxStylus、maxTouch、MediaLB、megaAVR、Microsemi、Microsemi ロゴ、MOST、MOST ロゴ、MPLAB、OptoLyzr、PIC、picoPower、PICSTART、PIC32 ロゴ、PolarFire、Prochip Designer、QTouch、SAM-BA、SenGenuity、SpyNIC、SST、SST ロゴ、SuperFlash、Symmetricom、SyncServer、Tachyon、TimeSource、tinyAVR、UNI/O、Vectron、XMEGA は米国とその他の国における Microchip Technology Incorporated の登録商標です。

AgileSwitch、APT、ClockWorks、The Embedded Control Solutions Company、EtherSynch、Flashtec、Hyper Speed Control、HyperLight Load、Libero、motorBench、mTouch、Powermite 3、Precision Edge、ProASIC、ProASIC Plus、ProASIC Plus ロゴ、Quiet-Wire、SmartFusion、SyncWorld、Temux、TimeCesium、TimeHub、TimePictra、TimeProvider、TrueTime、ZL は米国における Microchip Technology Incorporated の登録商標です。

Adjacent Key Suppression、AKS、Analog-for-the-Digital Age、Any Capacitor、AnyIn、AnyOut、Augmented Switching、BlueSky、BodyCom、Clockstudio、CodeGuard、CryptoAuthentication、CryptoAutomotive、CryptoCompanion、CryptoController、dsPICDEM、dsPICDEM.net、Dynamic Average Matching、DAM、ECAN、Espresso T1S、EtherGREEN、GridTime、IdealBridge、In-Circuit Serial Programming、ICSP、INICnet、Intelligent Paralleling、IntelliMOS、Inter-Chip Connectivity、JitterBlocker、Knob-on-Display、KoD、maxCrypto、maxView、memBrain、Mindi、MiWi、MPASM、MPF、MPLAB Certified ロゴ、MPLIB、MPLINK、MultiTRAK、NetDetach、Omniscient Code Generation、PICDEM、PICDEM.net、PICKit、PiCtail、PowerSmart、PureSilicon、QMatrix、REAL ICE、Ripple Blocker、RTAX、RTG4、SAM-ICE、Serial Quad I/O、simpleMAP、SimpliPHY、SmartBuffer、SmartHLS、SMART-I.S.、storClad、SQL、SuperSwitcher、SuperSwitcher II、Switchtec、SynchroPHY、Total Endurance、Trusted Time、TSHARC、USBCheck、VariSense、VectorBlox、VeriPHY、ViewSpan、WiperLock、XpressConnect、ZENA は米国とその他の国における Microchip Technology Incorporated の商標です。

SQTP は米国における Microchip Technology Incorporated のサービスマークです。

Adaptec ロゴ、Frequency on Demand、Silicon Storage Technology、Symmcom はその他の国における Microchip Technology Incorporated の登録商標です。

GestIC は、その他の国における Microchip Technology Germany II GmbH & Co. KG (Microchip Technology Incorporated の子会社) の登録商標です。

その他の商標は各社に帰属します。

© 2022, Microchip Technology Incorporated and its subsidiaries.

All Rights Reserved.

ISBN: 978-1-6683-0765-6

各国の営業所とサービス

南北アメリカ

本社
2355 West Chandler Blvd.
Chandler, AZ 85224-6199
Tel: 480-792-7200
Fax: 480-792-7277
技術サポート:
<http://www.microchip.com/support>
URL:
www.microchip.com

アトランタ
Duluth, GA
Tel: 678-957-9614
Fax: 678-957-1455

オースティン, TX
Tel: 512-257-3370

ボストン
Westborough, MA
Tel: 774-760-0087
Fax: 774-760-0088

シカゴ
Itasca, IL
Tel: 630-285-0071
Fax: 630-285-0075

ダラス
Addison, TX
Tel: 972-818-7423
Fax: 972-818-2924

デトロイト
Novi, MI
Tel: 248-848-4000

ヒューストン, TX
Tel: 281-894-5983

インディアナポリス
Noblesville, IN
Tel: 317-773-8323
Fax: 317-773-5453
Tel: 317-536-2380

ロサンゼルス
Mission Viejo, CA
Tel: 949-462-9523
Fax: 949-462-9608
Tel: 951-273-7800

ローリー, NC
Tel: 919-844-7510

ニューヨーク, NY
Tel: 631-435-6000

サンノゼ, CA
Tel: 408-735-9110
Tel: 408-436-4270

カナダ - トロント
Tel: 905-695-1980
Fax: 905-695-2078

アジア / 太平洋

オーストラリア - シドニー
Tel: 61-2-9868-6733

中国 - 北京
Tel: 86-10-8569-7000

中国 - 成都
Tel: 86-28-8665-5511

中国 - 重慶
Tel: 86-23-8980-9588

中国 - 東莞
Tel: 86-769-8702-9880

中国 - 広州
Tel: 86-20-8755-8029

中国 - 杭州
Tel: 86-571-8792-8115

中国 - 香港 SAR
Tel: 852-2943-5100

中国 - 南京
Tel: 86-25-8473-2460

中国 - 青島
Tel: 86-532-8502-7355

中国 - 上海
Tel: 86-21-3326-8000

中国 - 瀋陽
Tel: 86-24-2334-2829

中国 - 深圳
Tel: 86-755-8864-2200

中国 - 蘇州
Tel: 86-186-6233-1526

中国 - 武漢
Tel: 86-27-5980-5300

中国 - 西安
Tel: 86-29-8833-7252

中国 - 厦門
Tel: 86-592-2388138

中国 - 珠海
Tel: 86-756-3210040

アジア/太平洋

インド - バンガロール
Tel: 91-80-3090-4444

インド - ニューデリー
Tel: 91-11-4160-8631

インド - プネ
Tel: 91-20-4121-0141

日本 - 大阪
Tel: 81-6-6152-7160

日本 - 東京
Tel: 81-3-6880-3770

韓国 - 大邱
Tel: 82-53-744-4301

韓国 - ソウル
Tel: 82-2-554-7200

マレーシア - クアラルンプール
Tel: 60-3-7651-7906

マレーシア - ペナン
Tel: 60-4-227-8870

フィリピン - マニラ
Tel: 63-2-634-9065

シンガポール
Tel: 65-6334-8870

台湾 - 新竹
Tel: 886-3-577-8366

台湾 - 高雄
Tel: 886-7-213-7830

台湾 - 台北
Tel: 886-2-2508-8600

タイ - バンコク
Tel: 66-2-694-1351

ベトナム - ホーチミン
Tel: 84-28-5448-2100

欧州

オーストリア - ヴェルス
Tel: 43-7242-2244-39
Fax: 43-7242-2244-393

デンマーク - コペンハーゲン
Tel: 45-4485-5910
Fax: 45-4485-2829

フィンランド - エスポー
Tel: 358-9-4520-820

フランス - パリ
Tel: 33-1-69-53-63-20
Fax: 33-1-69-30-90-79

ドイツ - ガーヒンク
Tel: 49-8931-9700

ドイツ - ハーン
Tel: 49-2129-3766400

ドイツ - ハイムブロン
Tel: 49-7131-72400

ドイツ - カールスルーエ
Tel: 49-721-625370

ドイツ - ミュンヘン
Tel: 49-89-627-144-0
Fax: 49-89-627-144-44

ドイツ - ローゼンハイム
Tel: 49-8031-354-560

イスラエル - ラーナナ
Tel: 972-9-744-7705

イタリア - ミラノ
Tel: 39-0331-742611
Fax: 39-0331-466781

イタリア - バドヴァ
Tel: 39-049-7625286

オランダ - ドリュウネン
Tel: 31-416-690399
Fax: 31-416-690340

ノルウェー - トロンハイム
Tel: 47-7288-4388

ポーランド - ワルシャワ
Tel: 48-22-3325737

ルーマニア - ブカレスト
Tel: 40-21-407-87-50

スペイン - マドリッド
Tel: 34-91-708-08-90
Fax: 34-91-708-08-91

スウェーデン - ヨーテボリ
Tel: 46-31-704-60-40

スウェーデン - ストックホルム
Tel: 46-8-5090-4654

イギリス - ウォーキンガム
Tel: 44-118-921-5800
Fax: 44-118-921-5820