原子鐘在資料中心的作用

原子從對資料造成不利影響到帶來各種益處的轉變過程

■作者：David Chandler
Microchip 頻率與授時系統業務部產品行銷經理

利用原子鐘授時已成為資料中心不可或缺的組成部分。目前，透過全球定位系統 (GPS) 和其他全球導航衛星系統 (GNSS) 網路傳輸的原子鐘時間已使全球各地的伺服器實現了同步，並且部署在各個資料中心的原子鐘，可在傳輸時間不可用時仍保持同步。

無論是由於系統需求還是合規需求，這種出色的同步性能都至關重要，可確保每年在全球範圍內收集的資料 (以 zettabytes 為單位) 能夠得到有效儲存並用於許多應用。原子的量子性質可保持精確的時間，是確保未來能夠以更快的速度處理更多資料的關鍵所在，而具有諷刺意味的是，就在幾年前，原子的量子性質還被視為提升資料處理能力和速度的最大阻礙。

1965 年，Gordon Moore 預測積體電路上的電晶體數量每年翻一倍。這一數字最終被修改為每兩年翻一倍。隨著電晶體密度的增加，速度有了顯著提升，成本和功耗也不斷下降。

在 1965 年，人們可能很難想像，2021 年時在一個半導體上佈置 500 億個電晶體是一種現實需求，但正如半導體技術隨著時代不斷發展，應用需求也在不斷變化。手機、金融交易和 DNA 圖譜繪製等應

圖 1：極具諷刺意味的圖片：工程師試圖遵循摩爾定律

用都非常依賴微控制器每秒可執行的運算次數，而這一數位與晶片上的電晶體數量密切相關。

摩爾定律的消亡

遺憾的是，由於物理學限制，摩爾定律正在迅速走向終結。隨著晶圓製程工藝節點現已達到 10 奈米以下，電晶體的大小僅為矽原子的 10 到 50 倍左右。在這個尺度上，原子和自由電子的大小以及量子特性顯著阻礙了電晶體大小的進一步縮減。從本質上講，可以將原子視作推翻這一定律的最終原因。

儘管摩爾定律終將消亡，但是，對提高處理能力的需求卻不斷增加。隨著物聯網 (IoT)、串流影音服務、社交媒體貼文和自動駕駛汽車的出現，每天產生的資料量會繼續呈指數增長。

據估計，2021 年每天產生的資料量為 2.5 exabytes (2,882,303,761,517,120,000 位元組)。當前使用的 exabytes 資料庫每秒可處理超過 10 萬筆交易 (transaction) (一筆交易包含許多次運算)，而在可預見的將來，資料庫的規模和每秒處理的交易數將持續增長。

機器同步

資料量的這種爆炸式增長，再加上資料必須達到的寫入、讀取、複製、分析、操作和備份速度，這些因素要求資料中心架構師找到一種能夠繞過摩爾定律終結的方法。對於採用分散式資料庫的資料中心，架構師採用了橫向擴展 (horizontal scaling) 方法，即將資料庫分佈在一個叢集 (cluster) 中的多個伺服器上，而不是整個資料庫駐留在一個伺服器上。

在這種配置下，叢集本質上運作為一台巨型機器，因此系統的大小和速度現在受到資料中心的外形尺寸而非原子大小的限制 (接招吧，原子！)。

軟體工程師現在的職業是編寫能夠實現橫向擴展的程式。但是，要使各種軟體都正常工作，所有機器都必須同步，否則會違反因果關係的概念。

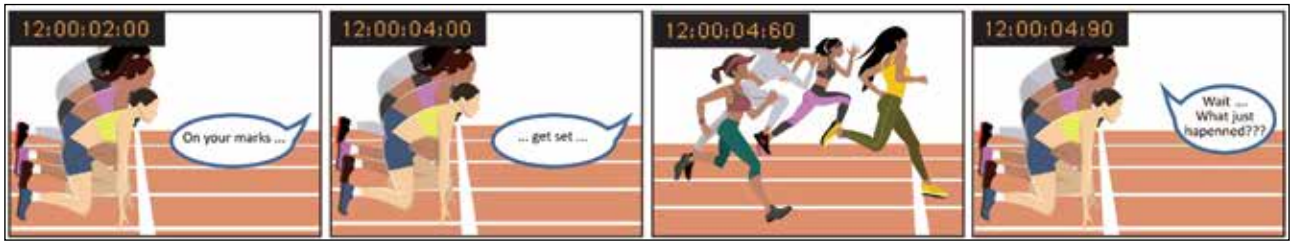
什麼是因果關係？舉個最簡單的例子。假設您用兩台攝影機來記錄 100 米短跑的圖像，每台攝影機都有自己的內部時鐘。第一台攝影機位於起跑器上。第二台攝影機位於終點線上。兩個感測器都在進行連續拍攝，並用各自時鐘的時間給每個圖像添加時間戳記。

要確定比賽中獲勝的短跑選手的正式成績，將檢查第一台攝影機的圖像以瞭解第一位選手離開起跑器時的時間點，然後用終點線上的攝影機圖像上該選手衝過終點線時的時間減去該時間戳記。

要實現此目的，兩台攝影機的同步精確度必須都達到可接受的誤差水準。如果時鐘的同步精確度只有 ±0.05 秒，那麼便無法確定成績為 9.6 秒的選手是否確實打破了 9.58 秒的世界紀錄。如果它們與體育場時鐘的同步精確度只有 ±5 秒怎麼辦？

想像一下這樣的場景：從體育場的主時鐘觀察，一場比賽正好在下午 12:00:00:00 開始。第一位選

圖 2：時鐘偏差會導致因果關係問題。在這種情況下，比賽在開始前就正式結束了。



手在下午 12:00:09:60 時衝過終點線。從體育場主時鐘的角度來看，正式比賽成績是 9.6 秒。

但是，如果第一台攝影機的時鐘正好快 5 秒，而第二台攝影機的時鐘正好慢 5 秒呢？比賽將在下午 12:00:05:00 正式開始，在下午 12:00:04:60 結束。比賽將在開始前 0.4 秒正式結束，這會打破世界紀錄並推翻物理定律，目前的紀錄保持者很有可能會無情地被所有贊助商拋棄。

將因果關係應用於資料庫

同樣的因果關係原則在資料庫中也十分重要。交易記錄更新必須按照它們發生的順序出現在資料庫中。如果您期望在透過直接取款支付每月房貸之前直接存入自己的工資，而銀行的資料庫沒有按正確的順序記錄這些交易，那麼您可能會被收取透支費。在一台機器上，因果關係錯誤很容易防止，但在多個伺服器上，每個伺服器都有自己的內部時鐘，伺服器必須同步並為每個交易加上時間戳記。

要實現此目的，必須有一個伺服器充當參考時鐘，就像體育場的時鐘，它必須採用最大程度減小每個伺服器時鐘的時間誤差的方式，將時間分配給每個伺服器。每個時間戳記的偏差 (比賽中為 ±5 秒) 形成一個時間包絡 (time envelop)，其長度為時鐘偏差的兩倍 (比賽中為 10 秒)。對於分散式資料庫，一秒內可以容納的非重疊時間包絡數量應當至少與系統預期的每秒交易數量大致相同。

概率、因果關係的關鍵性和實現成本最後都會在最終解決方案中發揮作用，但這種關係是一個很好的起點。時間戳記偏差為 ±1 毫秒的系統將具有 2 毫秒的時間包絡，一秒內最多可容納 500 個非重疊時間包絡。此系統可以支援每秒執行約 500 個交易。

NTP 和 PTP 的不足

乙太網授時技術也稱為網路時間協定 (NTP) 和精確時間協議 (PTP)，用於同步資料中心的分散式資料庫中的所有伺服器。這些協定可以確保區域網能夠以亞毫秒 (NTP) 或亞微秒 (PTP) 的偏差來分配時間，從而支持每秒執行數千 (NTP) 或數百萬 (PTP) 個交易。

遺憾的是，即使憑藉這些解決方案可以繞過原子帶來的摩爾定律消亡，物理學仍以光速的形式在分散式資料庫的道路上設置了另一個障礙。

試想一下，一個使用 PTP 進行準確同步的分散式資料庫在加州聖約瑟運行，每秒可輕鬆執行 100,000 個交易，且不會產生任何因果關係問題。一位資料庫架構師正坐在自己位於紐約的辦公室裡，他的老闆要求他更新大量記錄。

這名架構師希望能夠充分利用其新資料庫並展示系統的能力。他計畫每秒執行 100,000 個交易。

為了根據請求更新記錄，他創建了一個簡單的交易，即僅當第一個記錄的值大於第二個記錄時，才會將第一個記錄的值與第二個記錄相加。如要達到這一目的，他必須對這兩個記錄發出讀取請求。然後，他在紐約的本地機器對這些值進行比較，然後在需要時向第二個記錄發送寫命令。

完成此操作後，他想要接著執行下一個交易，即將第三個值與新的總和進行比較。如果新的總和大於第三個記錄，那麼將使用第三個記錄替換總和。他想對 600 萬條記錄重複此操作。由於資料庫每秒能夠處理 100,000 個交易，他認為此任務將在大約一分鐘內完成。他告訴老闆，他將在五分鐘內更新記錄，然後離開去喝杯咖啡。

喝咖啡的時候，他讀到一個故事，內容是新的百米短跑成績是負 0.4 秒，這違背了物理定律，並且之前的紀錄保持者因為失去了所有的代言費正在起訴體育場負責人。架構師自顧自地笑了起來，認為體育場應該聘請他作為同步專家。

五分鐘後他回到辦公桌前，沮喪地發現他的資料庫更新只完成了不到 1,500 個交易。他難過地意識到自己的錯誤，並準備將自己的履歷發給那個體育場，他希望他的 PTP 部署不會出現同樣的問題。

問題出在哪裡？光速將紐約和聖約瑟之間理論上最快的資料傳輸速度限制在 13.7 毫秒。

圖 3：光速對兩點之間的資料傳輸速度施加了理論上的限制

距離問題

遺憾的是，現實世界的交易處理速度甚至更慢。即使兩個地點之間有專用的光纖網路，光纖的折射率、光纖的實際路徑和其他系統問題也會延長傳輸時間。因此，僅僅從紐約傳輸一次，就需要 40 到 50 毫秒的時間才能到達聖約瑟。

但是，此交易中有四個獨特的操作。有兩個可以同時發生的讀取操作，隨後必須將它們發送回紐約。往返過程需要 80 到 100 毫秒。然後，在對兩個值進行比較後，就會發出寫入操作，並且必須發回寫入確認以指示寫入操作已完成，然後才能開始下一個交易。

突然之間，資料庫每秒能否執行 100,000 個交易已無關緊要，因為距離將系統每秒的處理能力限制為不超過 5 個交易。要完成 600 萬個交易，此系統需要 13 天的時間，這樣便有足夠的時間再喝幾杯咖啡，甚至更新一份簡歷。這種延遲稱為通信延遲。

規避延遲

但就像摩爾定律一樣，資料庫架構師想出了規避延遲的方法。在使用者附近創建資料庫副本，這樣他們便可隨意使用資料，而不必將信號發送到全國各地。

定期比較和協調複製以確保一致性。在協調過程中，交易時間戳記用於確定交易的實際順序，並且當存在不可協調的差異時 (例如交易時間包絡重疊時)，有時會回捲記錄。減少時鐘偏差可以減少複製的實例中不可協調的差異數量，因為時間包絡增多會減少重疊的概率。這可提高效率並降低資料損壞概率。

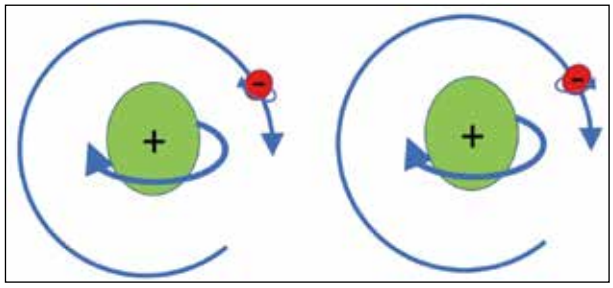
但現在，時間戳記不僅在每個資料中心內部必須做到精確，在不同的資料中心之間也必須精確，這些資料中心可能相隔數千英里，並透過雲相互連接。由於需要一個偏差極低且在兩個地點均可隨時獲得的外部參考，因此這項任務變得愈加困難。

下至原子級別

此時，資料庫架構師以前的敵人“原子”登場。當原子忙於廢除摩爾定律時，其亞原子粒子卻在忙於自旋。原子核內的中子和質子一直在旋轉，而與此同時電子則一邊忙於圍繞原子核公轉，一邊自旋。這類似於地球在繞太陽公轉的同時自轉。

電子可以圍繞自身的軸順時針或逆時針自旋。考慮到人體內約有 7*10^27 (7 後面有 27 個零) 個原子，所有亞原子粒子都在我們體內自旋，令人驚訝的是我們並沒有一直頭暈目眩。(注：亞原子粒子並不是真的在忙著自旋和公轉，它們實際上是在忙著給我們提供概率波函數和磁相互作用，這會讓我們獲得類似於它們進行自旋和公轉時的結果。但是，

圖 4：具有核和價電子的概念性原子，具有核自旋、電子自旋和軌道自旋



如果想到所有的自旋會讓您頭暈目眩，那麼試圖理解量子物理學的現實肯定會更令人厭惡。)

如果電子吸收特定精確頻率的微波輻射，繞電子軸的自旋方向會改變。如果地球上發生這種情況，太陽會突然從東方落下，從西方升起！

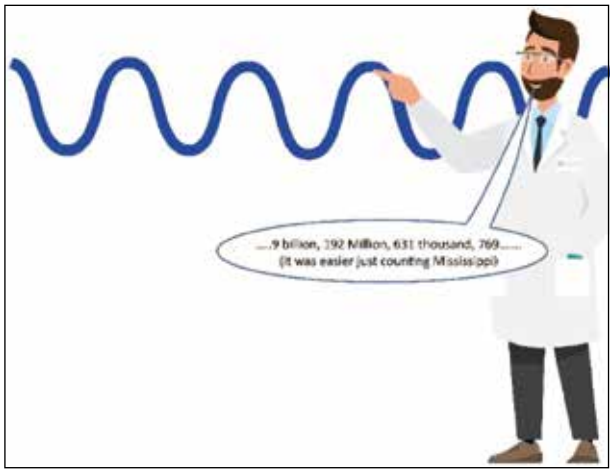
原子鐘這種機器專門用於檢測電子自旋狀態，然後透過微波輻射改變方向。頻率變化取決於元素、同位素和電子的激發態。

在機器確定頻率 (即所謂的超精細躍遷頻率) 後，便可將週期確定為頻率的倒數，這樣便可計算週期數來確定經過的時間。國際上對秒的定義是誘導銫原子軌道外層內電子的超精細躍遷所需的 9,192,631,770 個輻射週期。

原子鐘是世界上最穩定的商用時鐘。一副紙牌大小的原子鐘稱為晶片級原子鐘 (CSAC)，其 24 小時內的漂移為百萬分之一秒，而冰箱大小的原子鐘稱為氫微波激射器，其 24 小時內的漂移僅為十兆分之一秒。巧合的是，十兆分之一也大約是氫原子半徑與百米短跑選手和現已失業的紐約資料中心架構師身高的比值。

憑藉這些原子鐘提供的精確度，可以為在東京、倫敦、紐約、西非的廷巴克圖或世界其他任何地方的資料中心運行的分散式資料庫提供大約 50 萬到 500 億個非重疊時間包絡。

圖 5：單位“秒”是透過計算銫超精細透射輻射頻率的 9,192,631,770 個週期來定義的

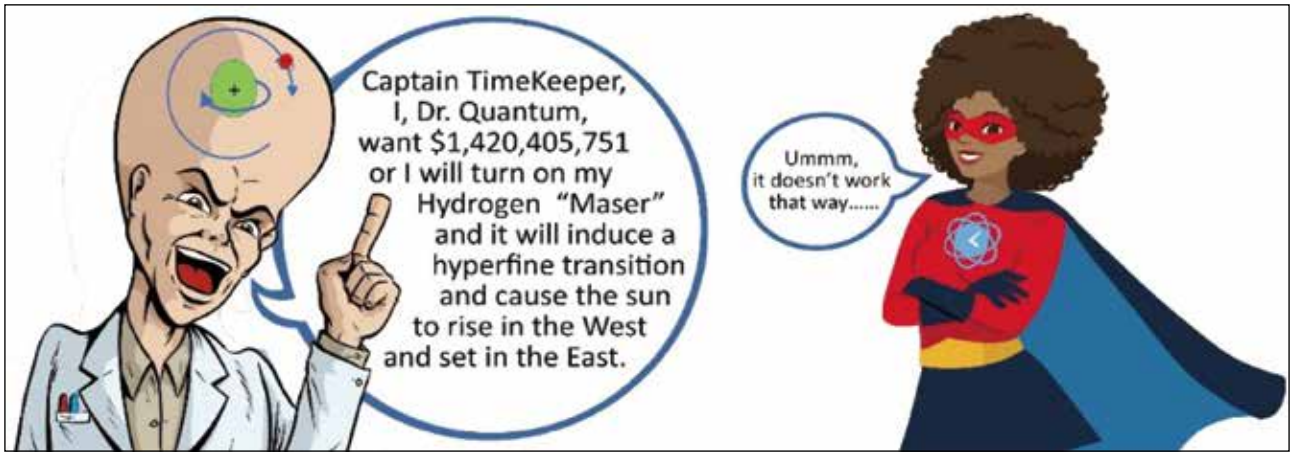


時間的分配

時間如何從這些原子鐘到達所有資料中心？世界協調時間 (UTC) 是透過衛星、光纖網路甚至互聯網分配的全球時間。UTC 本身源自位於世界各地的國家實驗室和授時站的一系列高精確度原子鐘。UTC 的提供組織會收到一份報告，其中載明了源自這些時鐘的 UTC 時間以及它們各自與計算出的 UTC 的偏移量。然後，這些實驗室和其他設施將時間傳送到世界各地。

UTC 報告每月公佈一次，告訴這些國家實驗室他們在上一個月與 UTC 的微小時間偏移量。從技術上講，直到事發一個月後，我們才知道準確的時間偏差。更糟糕的是，由於地球自轉和我們與可觀測恒星的相對位置的變化，UTC 會定期增加額外的秒數，即躍遷秒。雖然這可使地球與宇宙保持一致，但它會引起資料中心和 100 米短跑成績的混亂。

圖 6：氫微波激射器中產生的超精細躍遷頻率為 1.420405751 GHz，將導致電子自旋反轉

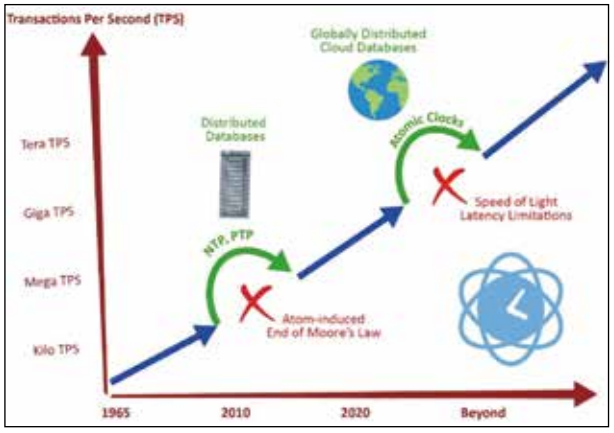


GNSS 登場

資料中心用來獲取 UTC 的常用方法有兩種：透過互聯網使用公開的 NTP 時間伺服器，以及透過衛星使用 GPS 或 GNSS 網路。雖然在分散式資料庫的早期部署期間，透過互聯網上的公共 NTP 時間伺服器進行授時很常見，但固有的性能、可追溯性和安全問題已經促使人們放棄了這種解決方案。

儘管 GPS 和其他 GNSS 通常被視為定位和導航系統，但它們實際上是精確授時系統。接收器的位置和時間取決於信號以光速從多個衛星傳輸到接收器的傳輸時間。極具諷刺意味的是，這是物理學原理引發問題的又一個案例 (此案例中是光速而不是原子)，但也有助於解決問題。

圖 7：資料庫交易速率的發展歷程以及實現和禁用的技術



這些衛星有自己的內建原子鐘，這些原子鐘與從地面站傳輸到衛星的 UTC 同步。利用這種方法獲取 UTC 可以提供 5 奈秒範圍內的時間偏差，進而實現每秒 1 億個時間包絡。

這種方法比公共 NTP 伺服器更可靠、更精確，雖然這些信號可能會被太陽風暴或蓄意的信號干擾等事件中斷，但在出現這些信號時，可以在每個單獨的資料中心放置與衛星信號同步的備份時鐘，以便在中斷期間提供所需的偏差水準。

下一步：躍遷電子

隨著未來對獲取、儲存和處理資料的需求不斷增加，我們需要具有極低偏差的新型原子鐘技術和時間傳輸系統。目前，國家授時實驗室正在開發一種新型原子鐘，用於研究電子躍過軌道層時發生的光學躍遷。這些原子鐘的頻率穩定性可達到 quintillion(10 的 18 次方) 分之一赫茲，最終將用於重新定義秒這個單位。

透過專用光纖網路或內建雷射器實現的信號傳輸已經顯著提高了傳輸精確度。憑藉這些不斷湧現的創新資料，原子和光將繼續它們之間複雜的愛恨交織關係，從而能夠以更快速度處理越來越多的資料，而不會出現一致性或因果關係問題。 CTA

Microchip 推出單對乙太網 (SPE) 元件協助推動 IIoT 邊緣和高速應用

單對乙太網 (SPE, Single Pair Ethernet) 技術正在為全乙太網 IIoT 和工業營運技術 (OT) 網路奠定基礎。這些網路採用新型同步低速乙太網邊緣設備和簡化的佈線基礎設施構建，用於延遲敏感型串流傳輸。Microchip 宣佈推出新款工業級 SPE 產品，包括更容易將邊緣 IIoT 設備連接到雲端的 10BASE-T1S MAC-PHY，以及 100BASE-T1 時間敏感型網路 (TSN) 乙太網 PHY 收發器和交換器的工業版本，可在遠距離乙太網網路中實現更高速的應用。

Microchip 新推出的 LAN8650 和 LAN8651 10BASE-T1S MAC-PHY 乙太網控制器帶有串列周邊介面 (SPI)，在為 OT 和 IT 網路的邊緣創建感測器、執行器和其他設備時，可以使用基礎微控制器 (MCU)，而不是帶有媒體存取控制器 (MAC) 的高階微控制器，因而簡化分區架構的部署。這些低速設備不需要自己的通信系統，Microchip 的 MAC-PHY 將它們連接到標準乙太網系統中，通過簡單的雙絞線直接連接到雲端。

對於需要更高頻寬的工業應用，設計人員可使用帶有整合乙太網 MAC 的微控制器。Microchip 現提供工業級版本的 LAN8770 100BASE-T1 乙太網 PHY 收發器，通過單條無護層雙絞線 (UTP) 電纜提供 100 Mbps 的發送和接收能力。