

產業研究報告

Cloud to Chip - 雲端服務供應商於邊緣運算晶片布局策略分析

前言

邊緣運算具有多義性 (Polysemy)，類屬不同產業的廠商對於邊緣運算趨勢的布局與應對策略也存在差異。雲端服務供應商 (Cloud Service Providers) 關注於邊緣運算「雲端至物端」(Cloud to Thing) 的連結介面；此一關注視角，進一步被 AWS、GCP、Azure 發展成「雲至晶片」(Cloud to Chip, C2C) 的布局策略。AWS、GCP、Azure 以此策略為基礎，在 2015 年之後，提出自身邊緣運算晶片處理器發展項目，意圖將配置在物端的晶片處理器，擴展成為雲端服務的前緣。前瞻未來，「擴充 ARM 架構處理器應用」、「強化 Edge AI 推論晶片功能」、「選擇更開放化的產品生態系」將會是雲端服務供應商布局邊緣運算與晶片、處理器的三項發展重心。

施柏榮

目錄

雲端與邊緣運算技術發展背景與競合	1
雲端服務供應商於邊緣運算晶片布局	5
結論	15
附錄	20

圖目錄

圖一、雲端運算與邊緣運算整體技術階層結構	2
圖二、雲端運算與邊緣運算整體功能連結介面	6
圖三、雲端服務提供商邊緣運算晶片發展歷程 - AWS	9
圖四、雲端服務提供商邊緣運算晶片發展歷程 - GCP	11
圖五、雲端服務提供商邊緣運算晶片發展歷程 - Azure	13

表目錄

表一、雲端服務供應商邊緣運算定義內容比較	4
表二、不同業種之於邊緣運算定義內容與比較	5
表三、雲端服務供應商邊緣運算晶片布局比較	15

雲端與邊緣運算技術發展背景與競合

邊緣運算相關的技術發展與背景

回顧邊緣運算 (Edge Computing) 相關研究，如以電機電子工程師學會 (Institute of Electrical and Electronics Engineers, IEEE) 學術文獻資料庫 IEEE Xplore 的研究文獻發表來看，最早可以回溯到加拿大麥吉爾大學電腦科學學院為主要的研究者在 2007 發表的〈藉由資料庫複寫機制增強邊緣運算〉 (Enhancing Edge Computing with Database Replication)。雖然 2007 至 2013 年間，仍有少數的文獻被發表。一直要到 2014 年歐洲電信標準協會 (European Telecommunications Standards Institute, ETSI)、2015 年由 Cisco、Intel、Dell 等企業在美國成立了開放霧聯盟 (OpenFog Consortium) 之後，邊緣運算相關技術研究才進入蓬勃發展的時期。

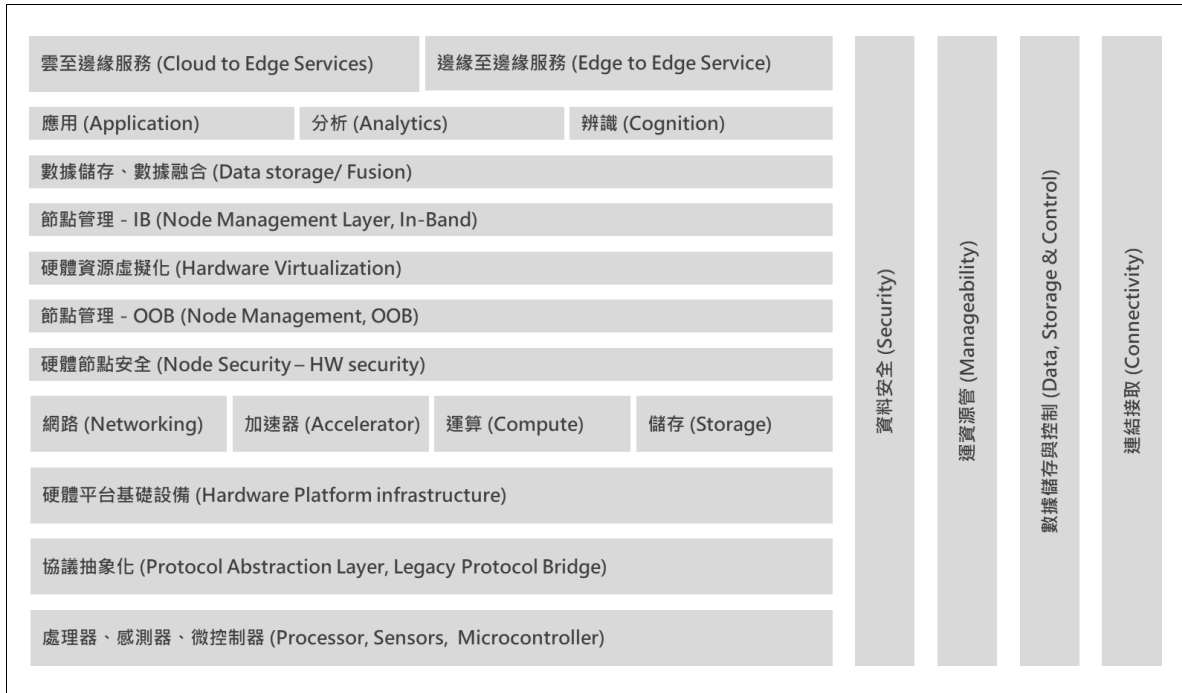
如以邊緣運算發展歷史進一步觀察，不難發現邊緣運算在 2014、2015 年初期推動階段，主要是由電信或通訊服務供應商 (Communications Service Provider, CSP) 與資訊系統與硬體供應商 (Information System and Hardware Provider) 所主導；兩種類型的服務供應商，也各自在 2017 年之後，發表有關於邊緣運算的標準，前者主要由 ETSI 提出「多重接取邊緣運算」 (Multi-access Edge Computing, MEC) 的標準，而後者則以 IEEE 與 OpenFog 主導，提出「IEEE 1934-2018 霧運算」標準，上述兩項，也正是目前全球有關於邊緣運算，最廣泛被採用、參考的技術標準。

值得一提的是，無論是 ETSI、OpenFog 提出有關於邊緣運算的技術標準，它們對於邊緣運算出現的背景，也多半會鎖定在 - 「邊緣運算的出現是為了補足既有雲端服務 (Cloud Service) 在運算、連結、儲存，甚至是在資訊安全功能上的缺口與不足」，甚至在 2015、2016 年前後，邊緣運算推動者，也多半會將雲端視為技術對立面。

然而，當 ETSI、OpenFog 在形塑邊緣運算的技術特徵論述之時，它們也代表背後的業者類型，試圖描繪新型態融合雲端、邊緣運算、終端處理器的一般性技術階層架構，這樣的技術階層架構，大致延伸物聯網 (Internet of Thing, IoT) 有關於雲、網、端三個主要階層架構的認知。如圖一所示，階層架構最上層為「雲端服務」，其中包含各類型的數據應用服務；中層則為「節點管理」，此一階層包括頻內 (In-Band, IB) 與頻外 (Out-of-Band, OOB) 兩種伺服器與網路設備的遠端管理工具，其大量應用硬體資源虛擬化 (Virtualization) 等技術；而最下層則涉及終端設備與模組的獨立控制，其中便包括了各類型處理器 (Processor)、感測器 (Sensor) 與微控制器。

從 ETSI、OpenFog 對於邊緣運算的技術階層架構的描繪來看，不難發現無論是電信服務供應商、資訊系統與硬體供應商，試圖以「去中心化」 (Decentralization) 的概念為核心，來重新定義整體產業生態系，並且重新定義不同資訊產品的階層位置。

圖一、雲端運算與邊緣運算整體技術階層結構



備註：雲端與邊緣運算架構會依照不同的標準組織、應用情境而有差異，此為一般性架構組成

資料來源：Industrial Internet Consortium (2019) 「The Industry Internet Consortium and OpenFog Consortium Unite」，MIC 整理，2020 年 12 月

雲端服務供應商的邊緣運算認知

回顧邊緣運算技術發展背景，可以發現在技術、產品發展初期是由通訊服務供應商、資訊系統與硬體供應商所推動，而且兩者皆嘗試藉由技術參考架構、技術標準的擬定來描繪整體技術樣貌，以及相對於中心化雲端 (Centralized Cloud) 的產品生態系。然而，這樣的推動策略以及數據處理 (Data Processing) 作法，確實會與雲端服務供應商存在已久的技術架構、商業模式產生衝突；加上終端場域設備大量出現、數據生成總量增加、數位安全與資產概念的興起，雲端服務供應商也必須對此進行調整。究竟雲端服務供應商如何思考邊緣運算？便成為首要必須回應的問題。

首先，表一整理出全球主要三家雲端服務供應商 - Amazon Web Service (AWS)、Google Cloud Platform (GCP)、Microsoft Azure 有關於邊緣運算的定義，可以發現它們仍是以雲端運算架構為核心，來定義邊緣運算的技術與產品特徵，可以發現雲端服務供應商對於邊緣運算的特徵描繪，大約可歸納為三項：

雲端服務供應商邊緣運算特徵描繪 1：邊緣運算為雲端運算服務延伸功能

AWS、GCP、Azure 雲端服務供應商面對運算位置必須更靠近使用者、數據生成點的趨勢，仍然將雲端運算視為商業服務的核心。比如 AWS 描繪邊緣運算的功能時，變非常明確提出 AWS 的邊緣運算服務目標，是「將雲移到靠近端點」(Moving the Cloud Closer to the End-point)，而 Azure 也與 AWS 相同，指出邊緣運算的根本目標是在雲端與邊緣 (Cloud to Edge) 兩個位置間，達到數據傳輸的最佳化，此外 Azure 也大量利用管理容器 (Container)，將雲端的軟體支援服務降至到邊緣位置。從上述的角度來看，可以發現雲端服務供應商對於邊緣運算的基本預測、認知，揭示將邊緣運算視為雲端運算的延伸服務，或者直接認為邊緣運算是雲端運算在適應終端設備數據處理需求的「適應型態」(Adaptive Type)，其服務本質是由雲端定義的。

雲端服務供應商邊緣運算特徵描繪 2：邊緣運算與雲端形成彈性解決方案

雲端服務供應商在面對不同的應用情境 (Application) 以及不同頻寬、延遲率容忍程度等「數據特性」的需求之時，強調藉由「網路架構設計」(Network Architecture Design) 來進行彈性的設計，也正是因為如此，衍伸出了如架構師 (Architect) 等新興的專業、職業類型。如果以架構設計的角度來觀察 AWS、GCP、Azure 對於邊緣運算的認知，不難發現雲端服務供應商皆會指出雲端、邊緣結合的「彈性化架構」(Flexible Architecture)，也就是會以雲端、邊緣彈性分工的思維來設計整體架構。雲端、邊緣運算彈性架構的思維，與部分強調在地方、臨場就建立起足夠運算、儲存資源的資訊系統與硬體供應商的思維存在著相當差異；這樣的差異也體現在雲端服務供應商對於邊緣運算工作負載程度的定義，會較為聚焦在快取 (Cache)、過濾功能。

雲端服務供應商邊緣運算特徵描繪 3：邊緣運算為執行人工智慧推論端點

以雲端服務作為本體思考，也反映在人工智慧 (Artificial Intelligence ,AI) 數據訓練的模式差異。相對於部分資訊系統與硬體提供商來說，雲端服務供應商並不會特別去關注邊緣運算必須具備持續擴充的運算能力 (Computing Power)，從這樣的現象來看，雲端服務供應商也不會強調邊緣運算必須擁有 AI 訓練 (Training) 的能力，而是將邊緣運算的伺服器等裝置，視為執行 AI 推論 (Inference) 的端點。除此之外，依循此模式，AI 訓練所需的數據 (Data) 便是集中在雲端形成數據湖 (Data Lake)。從 AI 模型的訓練、推論、驗證的方法 (Method)，能夠反映出雲端服務供應商在數據處理層面與資訊系統與硬體提供商最大的特徵差異。簡言之，雲端服務供應商的整體思考架構，即使部分運算能力能夠下放到邊緣，但數據仍是採取「集中化」模式。

表一、雲端服務供應商邊緣運算定義內容比較

Company	Definition
Amazon Web Services, AWS	邊緣運算的應用程式倚靠雲端來進行處理、分析、儲存以及進行機器學習，但它會需要在數據生成位置的附近，進行一部分的處理（Processing），提供更智慧地即時回應（Real-time Responsiveness）並減少數據傳輸
Google Cloud Platform, GCP	聯網裝置數量大幅地成長，以及對於隱私、機敏、低延遲性與傳輸頻寬限制的需求，帶動了在邊緣位置執行雲端服務、雲端訓練人工智慧模型的趨勢；邊緣運算是與雲端服務相輔，提供端對端、雲端至邊緣的彈性化解決方案
Microsoft, Azure	智慧邊緣運算系統是不斷擴展的連結系統、設備整合，這些系統與設備可以在靠近使用者、數據或兩者的情境之下，蒐集、分析數據，使用者可以藉由操作快速，且具有系統關聯性的應用程式，來實現即時回應與體驗服務

資料來源：AWS (2020) 「AWS for the Edge」、Google Cloud (2020) 「Edge TPU」、Microsoft Azure (2020) 「The future of computing」，MIC 整理，2020 年 12 月

不僅電信與通訊服務供應商、資訊系統與硬體提供商、雲端服務供應商曾經對於邊緣運算進行定義，包含資訊軟體服務提供商及半導體設計與製造商，也曾針對邊緣運算的功能、定位進行描述，如表二所整理。然而，藉由整理上述三點雲端服務供應商之於邊緣運算特徵的認知，可以發現相對電信或通訊服務供應商、資訊系統與軟體與硬體供應商而言，雲端服務供應商同樣允許放置在邊緣、臨場的資訊設備，如同服务器等擁有更佳的運算能力，來進行數據的「預處理」（Pre-Processing）來藉此降低對於雲端運算的壓力。不過，相對於資訊系統與軟體以及硬體供應商從邊緣運算延伸出來的「IT 系統整合」（IT System Integration）甚至是整合性更強的「超融合基礎建設」（Converged Infrastructure）等地方系統整合概念，雲端服務供應商一般情境下，對於邊緣運算工作負載的認知，並非持續擴張，而是藉由雲端運算的需求來定義。

除此之外，值得注意的是，無論是雲端服務供應商的何種商業模式皆是以「數據」的運算、儲存為主要來源，而且雲端運算所提供的商業智慧（Business Intelligence）、數據管理（Data Management Service）以及用 AI 為基礎的高階預測與分析服務，都同樣仰賴集中化資料庫（Centralized Database），這與其他業種最大的差異點。整體而言，藉由雲端運算的服務需求來設定邊緣運算的運算能力，以及其服務核心，仍是強調雲端話、集中化的數據處理，是雲端服務供應商有關邊緣運算的認知之上，相對於其他業種來說，最為明確的認知特徵與差異所在。

表二、不同業種之於邊緣運算定義內容與比較

Industrial Classification	Definition
雲端運算服務供應商	邊緣運算與智慧終端設備，皆為雲端運算服務延伸的端點，必須倚靠穩定的網路通訊環境，多數採雲、網、端的「垂直型運算架構」，將雲端服務平台所具備的運算、儲存與 AI 推論等功能，放置到靠近使用者或數據生成位置
電信通訊服務供應商	邊緣運算屬於下世代行動通訊的重要環節，邊緣運算的執行位置多半與蜂巢網路的基站 (Base Station)、小型基站 (Small Cell) 相重疊，邊緣運算主要目標是希望在行動網路的邊緣，提供雲端運算以及強化 IT 環境的能力
資訊系統軟體供應商	邊緣運算非常強調藉由軟體定義網路 (Software-Defined Networking, SDN) 與虛擬化 (Virtualization) 等資源管理技術的解決方案，隨著硬體與應用的複雜化，藉由軟體集中化管理技術，可以提供邊緣端的設備效率
資訊系統硬體提供商	邊緣運算是接近使用者、數據生成點的伺服器、資料中心、路由器、閘道器等資訊硬體設備所形成的邊緣運算節點 (Edge Nodes) 構成的運算架構，藉由分散式運算的技術，在邊緣端建立如雲端的資源池 (Resource Pool)
半導體設計與製造商	邊緣運算為具備足夠運算能力的晶片組 (Chipset)，或者可以是搭載這些晶片組的終端設備；藉由晶片組運算能力的提高，可以降低對於雲端運算的依賴，同時強化晶片資訊安全設計，也可強化整體雲端運算環境的安全性

備註：資訊系統軟體、資訊系統硬體提供商，兩個不同業種的廠商或利害關係人，在多數情境之下會採取共同協作的模式，或者係由同一個資訊系統整合商 (Systems Integrator) 進行系統性整合

資料來源：MIC，2020 年 12 月

雲端服務供應商於邊緣運算晶片布局

雲端至晶片的運算布局策略藍圖

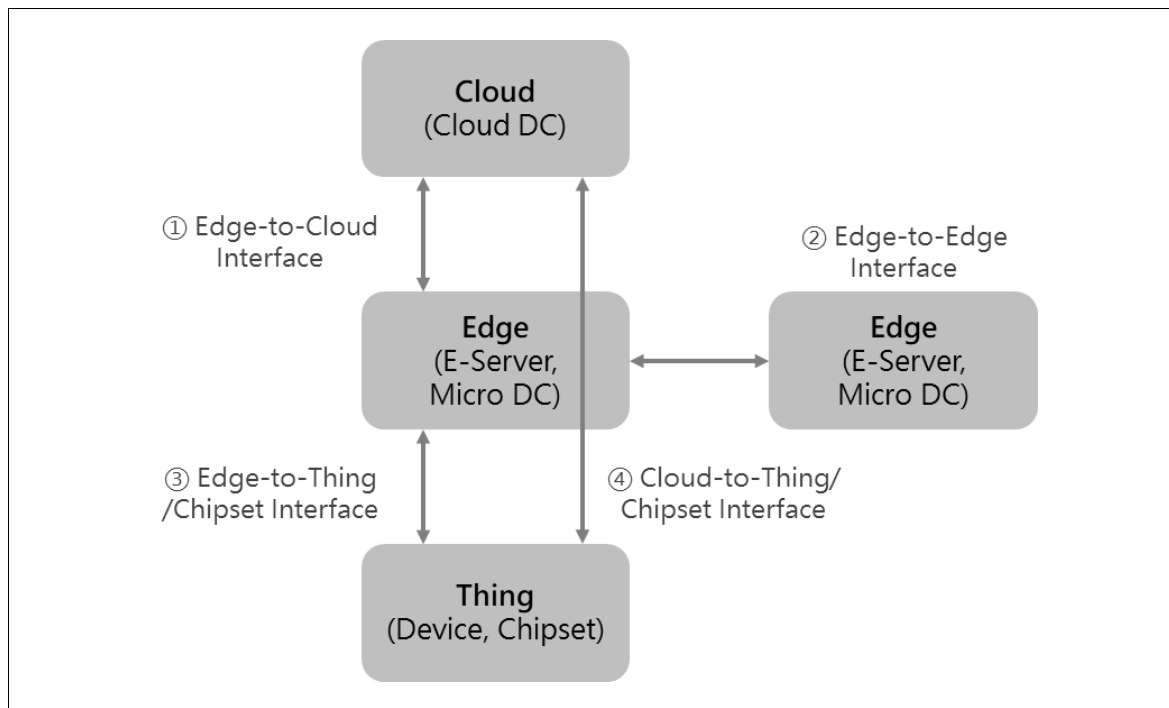
彙整邊緣運算的技術發展背景、雲端服務供應商的邊緣運算認知，以及不同業種之於邊緣運算的比較與特徵比較，再次顯現出邊緣運算的多義性 (Polysemy)，而雲端服務供應商在此競合格局之中的角色變顯得特別；因為對於雲端服務供應商來說必須同時面對其他業種的競爭之外，也必須讓自身的雲端服務產品更為彈性化，如此一來才足以因應使用者對即時回應 (Real-Time Response)、低延遲 (Low-Latency) 甚至是回應新型態資訊安全、數位資產保護 (Digital Asset Protection) 的挑戰。

因此，在回答「雲端服務供應商如何思考邊緣運算」問題之後，接續而來的問題是：雲端服務供應商面對邊緣運算之時，又提出什麼樣的布局「策略」與「作法」？

有關於此一部份，擔任美國普渡大學工程學院 (Purdue University College of Engineering) 院長以及在 2019 年 12 月被任命為美國國務卿科學技術顧問的蔣濛 (Mung Chiang)，在 2016 年發表的〈霧運算與物聯網：研究機會概述〉 (Fog and IoT: An Overview of Research Opportunities) 一文所提出的雲端運算與邊緣運算的四種介面 (Interface)，將有助於再以「網路架構設計」、「功能分工」為視角，提供我們一個健全的思考框架，用以釐清雲端服務供應商在邊緣運算的布局策略。

觀察雲端與邊緣運算的功能連結介面，可以發現共有：第一，邊緣至雲端 (Edge to Cloud)、第二，邊緣至物端 (Edge to Thing)、第三，邊緣至邊緣 (Edge to Edge) 與第四，雲端至物端 (Cloud to Thing) 四種功能介面，如圖二所示。

圖二、雲端運算與邊緣運算整體功能連結介面



備註：在未細部區隔邊緣運算本身的产品分層，如邊緣伺服器、微型資料中心的产品分層情境之下，邊緣運算主要可分為 4 種介面 (Interface)；如細部區隔邊緣運算分層，則會衍生出 7 種介面

資料來源：Mung Chiang et al. (2016) 「Fog and IoT: An Overview of Research Opportunities」，MIC 整理，2020 年 12 月

第一種邊緣至雲端介面主要處理邊緣設備如地方伺服器、微型資料中心與雲端之間的數據傳輸，比如 AWS 在 2018 年所發表的 Snowball Edge 裝置，便可以歸屬於此一介面的產品型態。第二種邊緣至物端的介面，則是產品生態相對多元、豐富的介面，包括應用乙太網路、無線通訊技術，在邊緣的伺服器、閘道器與終端設備之間的數據

傳輸網路，都屬於此一介面的產品發展樣態。而在第三種邊緣至邊緣的介面，則主要處理水平層級的數據傳輸，處於這類型介面的產品也多半會應用分散式運算 (Distributed Computing)、資料分散式服務 (Data Distribution Service, DDS) 進行資料的同步化，相當倚賴軟體面的資料管理能力。而在第四個介面，則強調雲端與終端設備、晶片組的直接介接，比如 Google 於 2017 年所發表的 Titan 資訊安全晶片，強調藉由晶片的資訊安全設計來強化雲端的安全性，也屬於此一介面的範疇。

第四種「雲端至物端」 (Cloud to Thing) 是雲端服務供應商採用的策略，這可以從三個面向來作證：首先，Amazon、Google、Microsoft 在 2015、2016 年之後，皆投入於晶片處理器、晶片組，或者是加速器 (Accelerator) 的自主研發與設計，而這樣的合作，並非單純的企業轉投資策略，三家雲端服務供應商皆試圖將自主開發的處理器或晶片組鑲嵌到自家的雲端運算技術架構之中，以融入原本的雲端生態系；其次，在 Amazon、Google、Microsoft 三個雲端服務供應商所建構的雲端環境，都將具有智慧功能的處理器單元，視為其雲端基礎架構的一部分，最為明確的案例為 Google 在 2015 年發表 Cloud TPU (Cloud Tensor Processing Units) 解決方案，該方案非常清晰地提出雲端與終端 TPU 之間的系統架構以及任務分工模式。最終，雲端服務供應商與「半導體設計與製造商」兩者建立起策略聯盟，比如 2016 年 ARM 提出「晶片至雲」 (Chip to Cloud, C2C) 策略，此種策略內涵，與雲端服務供應商所提的「雲端至物端」或「雲至晶片」 (Cloud to Chip, C2C) 的策略並無二致。

彙整上述所論，分別在「產品研發投資」、「產品生態系建構」、「異業策略聯盟」三個層面，明確地指出全球主要的雲端服務供應商 - Amazon、Google、Microsoft 皆是以「雲端至物端」作為基礎介面，再深化到更加細部的處理器、晶片組的設計，建立起一套「雲至晶片」的總體發展策略，期望建立更為全面的雲端環境系統架構。在「雲至晶片」策略的指引之下，雲端服務供應商不僅可以應對來自資訊系統硬體、軟體提供商的競爭，以提升自身對於整體物聯網市場、數據分析市場的掌握能力。

AWS、GCP、Azure的發展案例

從 Amazon、Google、Microsoft 發展路徑來觀察，全球三家主要雲端服務提供商，在其各別所轄的 AWS、GCP、Azure 的雲端運算服務之中，皆能夠找出「雲至晶片」 (Cloud to Chip, C2C)，或者「晶片至雲」 (Chip to Cloud, C2C) 的策略思考。除此之外，也不難發現雲端服務供應商，幾乎是在「邊緣運算」概念甫形成的 2015、2016 年前後，就開始規劃布局底層的晶片、處理器，並且積極找尋合作的對象。

然而，AWS、GCP、Azure 布局於晶片、處理器的具體「作法」，並未全然地相同，這部分除了涉及到晶片、處理器的研發能力之外，更為重要的是涉及對於邊緣運算、

數據服務的商業營運認知。以下，我們將再次依循蔣濛 (Mung Chiang) 2016 年所提出的邊緣運算第四種「雲端至物端」的介面定義，將布署於邊緣運算、終端設備的晶片與處理器，定義為「邊緣運算晶片」 (Chips for Edge Device)，以此來觀察 AWS、GPC、Azure 實現「雲至晶片」策略的細部作法，來進行綜合比較與分析。

Amazon : AWS

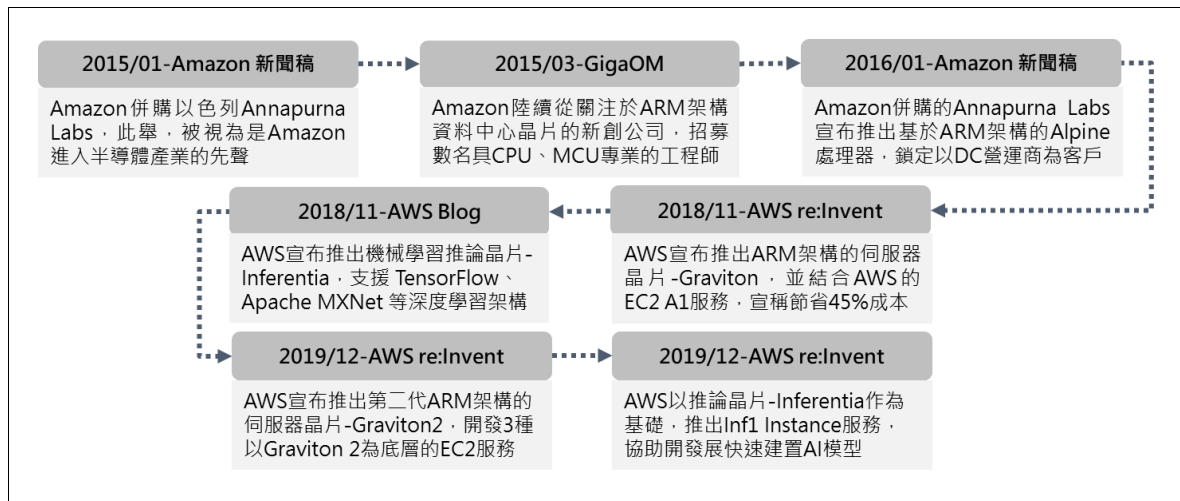
AWS 邊緣運算晶片處理器發展項目

圖三整理出 Amazon 在邊緣運算晶片發展的歷程。Amazon 於 2015 年 1 月份併購以色列 Annapurna Labs 於 AWS 旗下，此次半導體公司的併購案被視為 Amazon 跨足半導體事業的先聲。除了 Annapurna Labs 之外，Amazon 在 2015 年 3 月亦從美國新創公司 - Calxeda 招募了數名具備有資料中心晶片硬體設計能力的半導體工程師，不斷充實 Amazon 自身於半導體設計的研發能力。藉由 2015 年的併購、招募事件觀察，Amazon 在 2015 年之後便有非常明確規劃，試圖將半導體融入到 AWS 服務體系，2016 年之後 Amazon 又更加投入於此。

2016 年 1 月，Amazon 併購的 Annapurna Labs 宣布推出第一款基於 32 位元 ARMv7 以及 64 位元 ARMv8 的架構的 - Alpine 系列處理器，按 Amazon 於新聞稿的說法，Alpine 被預期配置在居家聯網、媒體串流影音裝置；Amazon 意圖以此應對物聯網設備運算能力有限、低功耗等問題。除了 Alpine 此一種聚焦物聯網裝置的處理器，2018 年 11 月 AWS 宣布推出同樣基於 ARM 架構的伺服器晶片處理器 - Graviton。此一處理器，在 Amazon 的新聞稿中，提出是基於 AWS 自身 EC2 服務需求所打造，它除了是 Intel 處理器之外的另一種選擇，相對於原先採用的 Intel 系列處理器，將更加適合 AWS 的雲端環境，宣稱可以有效提升工作負載的效果、降低客戶端的成本之外，也可以強化雲端服務安全性；2019 年 11 月份，在 Graviton 的研發基礎之上，AWS 則再次宣布推出 Graviton2 的處理器，相對於 Graviton 的處理器，Graviton2 更加適用於應用程式伺服器、視訊編碼、叢集運算等更高工作負載的應用上。

除了預期放置在居家裝置的 Alpine、雲端伺服器的 Graviton 之外，2018 年 11 月 AWS 宣布發表名為 - Inferential 的 AI 推論 (Inference) 的特殊應用積體電路晶片 (Application Specific Integrated Circuit, ASIC)；與 Graviton 相同，也是由 AWS 直接發表，預計可以支援 TensorFlow、Apache MXNet 與 PyTorch 深度學習架構。依據 AWS 新聞稿的陳述，Inferential 具備高傳輸、低延遲的推論效率，能夠執行複雜的 AI 模型，此外，Inferential 也可以支援 AWS EC2 的彈性化推論服務，此舉，無疑協助 AWS 的雲端服務，大幅度拓展到更高階的 AI 模型與數據預測服務。

圖三、雲端服務提供商邊緣運算晶片發展歷程 - AWS



資料來源：AWS (2019) 「AWS Inferentia」、Yevgeniy Sverdlik (2018) 「Amazon Web Services Unveils Custom Machine Learning Inference Chip」、James Hamilton (2019) 「AWS Inferentia Machine Learning Processor」、MIC 整理，2020 年 12 月

AWS 邊緣運算晶片處理器發展特徵

Amazon 對於邊緣運算晶片的布局，符合「雲至晶片」的發展策略。在 2016 年之後 Amazon 的推動重心，明顯從 Amazon 的母公司過渡到 AWS，此外，無論是配置在終端設備的 Alpine，或是配置在雲端伺服器的 Graviton、Graviton2，甚至是專注於執行 AI 模型與推論的 Inferentia，其晶片處理器的功能，皆是由雲端需求來定義，並成為 AWS 更大型如 AWS IoT Greengrass、AWS EC2 生態系不可或缺的環節。

如以「晶片能力取得」、「商業營收選擇」兩個個層面來看 Amazon 或者 AWS 的發展特徵，則可以更加凸出 AWS 投入於邊緣運算晶片處理器的細緻思考與作法。

首先，在晶片、處理器「晶片能力取得」層面，Amazon 採取企業併購、獵才方式，逐步計畫在企業內部，培育源於自身的晶片硬體設計的能力，而且從 Amazon 在 2015、2016 年併購、獵才的對象來看，兩者都是具備 ARM 架構開發能力的企業或專業工程師；而會有此種選擇，也反映出 Amazon 也希望藉由自身晶片的研發投入，稀釋掉長期以來由 Intel 主宰伺服器、資料中心的 x86 指令集架構晶片之市場情境。

其次，在發展晶片、處理器「商業營收選擇」層面，從 Amazon 所提出的自我研發晶片來看，Amazon 所關注的商業服務可以分為兩個項目：第一，如 Alexa 等智慧終端裝置及其應用情境，如智慧家庭、語音串流裝置，所延伸出的新興商業營收來源；第二，AI 智慧服務與應用，Amazon 在 2016 年 11 月宣布推出 AWS Greengrass 軟體預覽版本之後，開始擴大關注 AI 帶來的商業收益，AWS Greengrass 強調可以

在末能夠連接到網路、間接雲的邊緣運算情境之下，也可在裝置上執行 AWS Lambda 函數，而 2018 年 11 月發表 Inferential 推論晶片之後，更可在邊緣端的裝置之上，運行自雲端訓練的 AI 模型與推論，成功讓 AWS Greengrass 軟體擴及到邊緣運算。

彙整 Amazon 的邊緣運算晶片布局的作法，可歸納為二：第一，強調晶片處理器的自主研發能力，期望藉此來提升 AWS 雲端服務效率之外，也希望藉此來稀釋 Intel 的影響力；第二，在邊緣端，因應未來 AI 市場的興起，布局低功耗、現場即時反應能力的 Inferentia 晶片處理器，擴充雲端的 AI 部屬能力至邊緣運算端點，據以因應愈來愈多的聯網設備未能連結上雲端，或者處於「間歇雲」情境的數據處理需求。

Google : GPC

GCP 邊緣運算晶片處理器發展項目

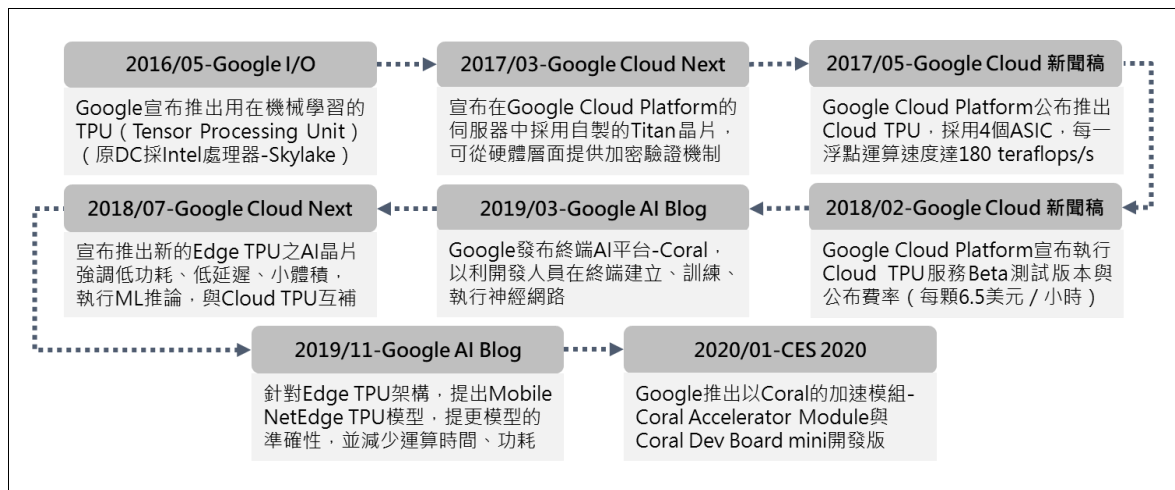
圖四整理出 Google 在邊緣運算晶片的發展歷程。Google 在 2016 年 5 月，宣布推出第一款應用於機器學習 (Machine Learning) 的張量處理器單元 - TPU (Tensor Processing Unit)，而 TPU 也多半被視為 Google 進入晶片處理器市場的代表案例。如果就 Google 對於 TPU 新聞稿的陳述，事實上 Google 在 2015 年時，便在 GCP 的資料中心進行試驗，雖然在 2016 年 5 月對外宣布，但一直要到 2018 年 2 月時，才在既有 GCP 平台上，提出第三方使用 TPU 的收費方案。藉由 TPU 案例來觀察，Google 採取自主研發模式，而且起初就有明確的布局重心。

在 TPU 的發展基礎之上，2018 年 7 月 GCP 在年度舉辦的 Google Cloud Next 會議宣布推出 - Edge TPU 的晶片處理器，Edge TPU 相對於雲端資料中心使用的 TPU，可以滿足較低的功耗，同時也有較小的體積，也因為預期配置在邊緣運算設備之內，因此，也有較低的延遲率；而最為重要的是，Google 強調 Edge TPU 可在邊緣運算的設備之內執行機器學習推論，與 GCP 的 Cloud TPU 進行互補；2019 年 11 月，Google 又再接再續提出 Mobile NetEdge TPU 模型，意圖更加優化 Edge TPU 架構。從 Google 推出 TPU 的發展歷程來看，可以發現到 Google 投入於 TPU 的核心思考有二：第一，提高雲端資料中心的效率，尤其是對於提高 Google 對於巨型資料機器學習能力的掌握程度；第二，2018 年之後，將 TPU 轉向「微型化」(Tiny) 發展，此一部份也屬於 Google 最為核心的能力之一，亦即盡可能保有運算效率的前提下，讓積體電路單元的體積，可以更加適應於更小型化的邊緣運算設備、裝置之內。

然而，除了明確扣合 GCP 雲端服務體系的 TPU 之外，2017 年 3 月 Google 也宣布推出 Titan 晶片處理器，Google 指出 Titan 可以在硬體層級，便提供加密驗證機制，此舉可以從底層硬體的多重安全性設計，確保雲端資料中心整體安全性。除此之外，Google 在 2019 年正式宣布推出名為 Coral 的 AI 平台，並且整合了多種開發板模組

與加速器模組(Accelerator Module)，這些開發板、加速器模組嵌入了 Edge TPU，未來在開發板大量進入到各類型的應用之中，Edge TPU 也將協助 GCP、Coral 等 Google 大型雲端與數據服務，進一步下放至邊緣運算以及智慧終端的位置之上。

圖四、雲端服務提供商邊緣運算晶片發展歷程 - GCP



資料來源：Google Cloud (2020) 「Cloud TPU」、Google Cloud (2020) 「Titan 安全金鑰」、Azalia Mirhoseini et al., (2020) 「Chip Placement with Deep Reinforcement Learning」、Coral (2020) 「Coral USB Accelerator」、MIC 整理，2020年12月

GCP 邊緣運算晶片處理器發展特徵

Google 針對不同層級所開發出的 TPU，如 Cloud TPU、Edge TPU，甚至是針對 Edge TPU 開發出來的 Mobile NetEdge TPU 模型，充分演繹出「雲至晶片」的策略思考，其背後，不僅是由補足 GCP 的雲端服務需求，來定位晶片處理器的功能設計，其中也將晶片處理器視為執行雲端服務的端點，此舉，也將確保 Google 可以建立雲至端的產品生態系，確保終端生成的數據，能夠穩定地傳輸至 GCP 的雲端環境之中。

如以「晶片能力取得」、「商業營收選擇」兩個個層面來解析 Google 或者 GCP 的發展特徵，則可以更加凸出 Google 投入於邊緣運算晶片處理器的細緻思考與作法。

首先，在晶片、處理器「晶片能力取得」層面，採取自我投資研發的模式，而且鎖定應用在 TensorFlow 機器學習框架的 ASIC，在能力的孕育上，起初是以 GCP 的資料中心作為試驗場域，待晶片處理器與 GCP 的雲端平台、TensorFlow 框架驗證之後，才對外向第三方提供使用服務。在此一過程之中，Google 有效整合了雲端服務平台、AI 與數據分析服務，以及晶片硬體設計的團隊，完成不同層級 TPU 的開發與驗證。

其次，在發展晶片、處理器「商業營收選擇」層面，相對 AWS、Azure 而言，較晚進入雲端服務市場的 Google，在商業營收選擇上，早在 2015、2016 年就觀察到 AI 在雲端市場的發展潛力，因此甚早就將 AI 視為目標的商業營收來源。在此一選擇思維之下，Google 推出 Cloud TPU、Edge TPU 與 Mobile NetEdge TPU 模型，就各自代表了不同的內涵；Cloud TPU 確保 GCP 可以用有足夠運算能力，來應用大量出現的數據，而 Edge TPU 的設計，則是 Google 嘗試確保 GCP 甚至 Coral 生態系可以進入到邊緣端、終端。換言之，藉由 TPU 的晶片處理器開發，可以串聯、整合雲端到終端的生態，能夠保全雲端平台收益之外，也可擴充到可期的 AI 應用服務。

彙整 Google、GCP 的邊緣運算晶片布局的作法，可歸納為二：第一，自主投入 TPU 的研發，先從雲端資料中心進行試驗，再微型化至邊緣運算的設備、開發板進行運行，建構一個以 TPU 為基礎的「雲至晶片」AI 服務生態系；第二，除了聚焦在 AI 應用的 TPU 之外，Google 也鎖定可以提高硬體安全性的 Titan 晶片，藉此來強化邊緣設備能夠進行多重安全性驗證，如同為 GCP 進行安全「過篩」一般，維持雲端安全性。

Microsoft : Azure

Azure 邊緣運算晶片處理器發展項目

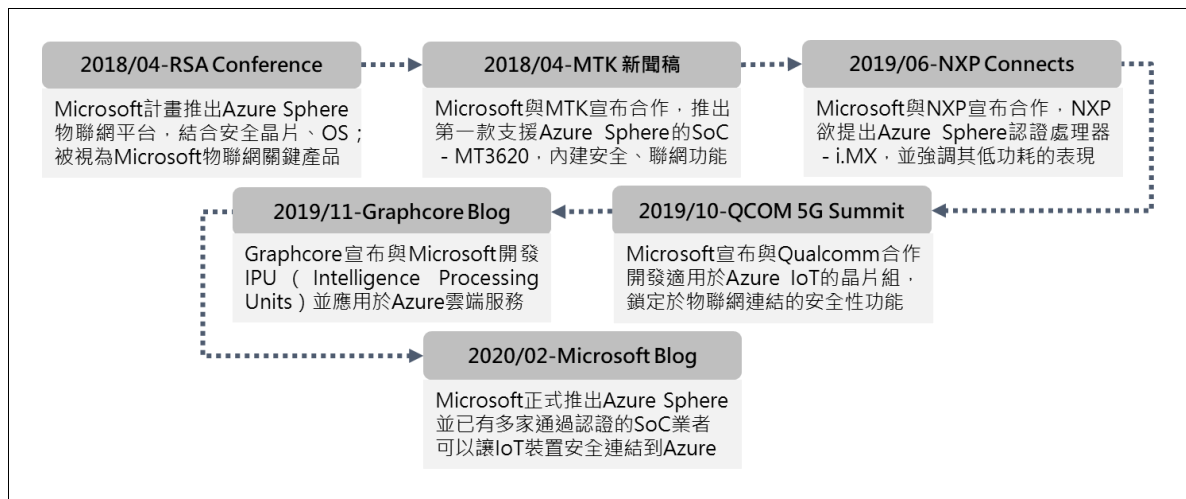
圖五整理出 Microsoft 在邊緣運算晶片發展的歷程。Microsoft 有關邊緣運算晶片、處理器的發展，最早可以回溯到 2018 年 4 月，Microsoft 在 RSA Conference 會議中，宣布推出 Azure Sphere 物聯網平台，而且幾乎是在同一時間點，宣布與聯發科技 (MediaTek, MTK) 合作推出名為 MT3620 的系統單晶片 (System on a Chip, SoC)；在 MTK 對於 MT3620 的新聞稿陳述中，便明確指出 MT3620 是與 Microsoft 共同合作的成果，且在系統單晶片開發階段，就以依循 Azure Sphere 相容性與維護 Azure Sphere 的安全性為主要的設計準則。

除了上述與 MTK 的合作經驗之外，2019 年 6 月 Microsoft 宣布與 NXP 進行合作，同樣推出符合 Azure Sphere 物聯網平台認證的晶片處理器 - i.MX。NXP 的 i.MX 與 MTK 的 MT3620 皆是基於 ARM 架構進行設計，宣稱具有高效率、低功耗的特徵。時間至 2019 年 10 月，Microsoft 則進一步與 Qualcomm 進行策略合作，共同推出應用於 Azure Sphere 物聯網的晶片組，強化物聯網連結至雲端的安全性。藉由上述 Microsoft 與 MTK、NXP、Qualcomm 的合作經驗來看，Microsoft 對於邊緣運算晶片、處理器的布局，主要採用與 IC 設計商「策略聯盟」的合作模式，Microsoft 提供 Azure Sphere 的「認證」給予這些 IC 設計商。雖然，Microsoft 並未提出自身品牌的晶片處理器，但藉由這種模式，Azure Sphere 物聯網平台獲快速拓展；這些

IC 設計商關注的應用場域，包括智慧家庭控制、串流影音媒體、Edge AI 與影像辨識，都成為 Azure Sphere 擴及到的場域。Microsoft 開闢出一條獨特的發展路徑。

除了物聯網應用之外，同樣值得注意的是 Microsoft 在 2019 年 11 月與英國聚焦在機器學習的半導體新創公司 Graphcore 共同開發名為 - 智慧處理單元(Intelligence Processing Unit, IPU)晶片處理器。在 Graphcore 的新聞稿陳述之中，IPU 處理器，擁有更好的機器學習工作負載能力，預期配置在 Microsoft Azure 雲端資料中心內。除了擁有較佳的工作負載能力之外，數個 IPU 也可擴充成為大型的運算系統，這舉，也可以協助 Azure 提升對於巨型資料、圖像資料的處理能力。未來 IPU 的應用預期將會更加強化，也可預期如同 Google Cloud TPU，廣泛為第三方提供運算服務。

圖五、雲端服務提供商邊緣運算晶片發展歷程 - Azure



資料來源：Microsoft Azure (2020) 「Azure Sphere」、Graphcore (2019) 「Microsoft and Graphcore collaborate to accelerate artificial intelligence」、MediaTek (2018) 「MT3620」，MIC 整理，2020 年 12 月

Azure 邊緣運算晶片處理器發展特徵

Microsoft 雖然相對於 Amazon、Google 而言，關注於晶片、處理器的時間稍晚，而且 Microsoft 於晶片處理器的布局，也並非投入自我品牌，而是採用與 IC 設計商合作的模式來進行；然而，即使如此，仍可以發現 Microsoft 與 Amazon、Google 相同，Microsoft 仍是以「雲至晶片」作為核心的策略思考，比如 Microsoft 提供 Azure Sphere 物聯網平台的「認證」予 IC 設計商進行開發，就是一個明確的驗證。如以「晶片能力取得」、「商業營收選擇」兩個個層面來解析 Microsoft 或者 Azure 的發展特徵，則可以凸出 Microsoft 投入於邊緣運算晶片處理器的細緻思考與作法。

首先，在晶片、處理器「晶片能力取得」層面，Microsoft 與 IC 設計商的合作經驗，Microsoft 妥善運用雙方的協作時機，也同時培養自身的晶片設計與解決方案團隊，尤其是在前端、後端晶片的物理階層區塊的功能整合，形成一個更為全面的基礎建設即服務 (Infrastructure as a Service, IaaS)。藉由 Microsoft 在晶片處理器的投入，不難發現未來的晶片處理器設計，朝向硬體元件軟體化整合的趨勢，將更加明確。

其次，在發展晶片、處理器「商業營收選擇」層面，相對於 AWS、GCP 而言，Azure 鎖定在一般性的物聯網市場，採認證而非自我品牌的模式，這意味著 Microsoft 並非以硬體、模組的販售作為潛在的收益來源，而是藉由大量認證，讓 IC 設計或者資訊硬體設備製造商應用，此舉，可以大量將符合 Azure Sphere 規格的開發板、模組等，擴散到各個應用領域之中，一方面可以讓 Azure 雲端平台能獲取穩定的數據來源，另一方面，Microsoft 也可以與下游廠商，建立起一個從雲端到物端的生態系，在此生態系之中，不同廠商收益來源的矛盾，可以盡可能漸少，並且持續擴充。除此之外，AI、高效率運算 (High Performance Computing, HPC) 亦是其鎖定的營收來源。

觀察 Microsoft、Azure 的邊緣運算晶片布局的作法，以更加嚴格的定義，Microsoft 在硬體層面，並未在邊緣運算晶片處理器進行布局，主要是以軟體層面思考邊緣運算的內涵；不過，值得一提的是，Microsoft 與 Graphcore 共同開發的 IPU 處理器，基於提供終端設備更為即時、智慧化的服務，IPU 未來將有非常大的潛力朝向微型化發展，往邊緣運算的位置移動，成為延伸 Azure 在 AI 推論、高效率運算的端點。

結論

雲端服務供應商的布局策略比較

由於雲端服務供應商的布局，與母公司的發展相互扣合，本研究嘗試以 AWS、GCP、Azure 三個雲端服務供應商為主體，整理其布局內容與比較資訊如表三，並針對 AWS、GCP、Azure 布局的邊緣運算晶片處理器的「產品項目」(What 層面)以及三者的「布局策略」為何(How 層面)來進行比較分析。期望藉由這兩個層面分析，綜合理解雲端服務供應商的思維之外，也嘗試去凸顯出不同雲端服務供應商在產品、商業服務思考上的差異，藉此，提供台灣主要如雲端服務設備、半導體設計與製造商相關廠商，擘劃相關產品與服務之思考。

表三、雲端服務供應商邊緣運算晶片布局比較

	Amazon, AWS	Google, GPC	Microsoft, Azure
核心服務	Amazon EC2	Google Cloud Platform	Azure Sphere
核心論述	AWS for the Edge	Cloud IoT Edge	Chip to Cloud
聯網架構	雲端到晶片中心化架構	雲端到晶片中心化架構	雲端到晶片中心化架構
雲端產品	Graviton, Graviton2	TPU	IPU
邊緣產品	Inferential	Edge TPU	無
物端產品	Alpine	Titan	無(採認證模式)
晶片能力	併購具備 ARM 架構能力的企業，進行自主設計研發	提供 GCP 雲端資料中心為試驗，進行自主設計研發	採用安全性認證、IC 設計廠商的策略性合作模式
AI 方案	推出自主設計的機器學習推論晶片 Inferentia，支援如 Tensor Flow、Apache MXNet 等深度學習架構	在雲端訓練 Tensor Flow 機器學習模型；然後再配置 Edge TPU 的設備上執行機器學習推論(Inference)	與 Graphcore 研發 IPU 的 AI 加速器(平行處理架構)用於 Azure 雲端伺服器，解決巨型資料量運算需求
合作夥伴	NVIDIA、ARM 等	Broadcom、Samsung、ARM 等	MTK、NXP、ARM、Qualcomm 等

資料來源：MIC，2020 年 12 月

AWS、GCP、Azure 邊緣運算晶片布局產品比較

倘若先論雲端資料中心的晶片處理器布局，可以發現到 AWS、GCP、Azure 三者皆嘗試建構新的雲端資料中心、伺服器所用的晶片處理器。GCP 大約在 2015 年便提出應用在雲端資料中心的 TPU，AWS 則是在 2018 年提出 Graviton，而 Azure 則是在 2019 年提出 IPU；雖然三者提出的時程略有差異，但有兩項共同特徵：第一，TPU、Graviton、IPU 的提出，皆是為了提升自身雲端服務的運算效率，尤其針對大型資料與影像資料的處理能力，而且相關晶片處理器的設計，會依循自身的雲端環境需求來設定功能準則；第二，AWS、GCP 皆尋求 ARM 的架構進行設計開發，此舉背後，蘊含著稀釋 Intel x86 架構對於雲端資料中心處理器的主導、降低雲端成本的思考。

回到非雲端、降至邊緣運算的晶片處理器布局，AWS、GCP、Azure 三者開始出現差異化發展，AWS、GCP 皆在 2018 年提出用以執行雲端模型、推論的晶片處理器，如 AWS 的 Inferential 與 GCP 的 Edge TPU，不過前者強調可以支援多種機器學習的架構，而後者則明確鎖定在 Google 本身的 Tensor Flow 學習架構；而且相對於 AWS，Google 更為明確希望以晶片處理器為基礎，建構一個從雲端至邊緣端的產品生態，這也是何以應用至雲端資料中心的處理器命名為 Cloud TPU，而應用在邊緣運算端點的處理器則命名為 Edge TPU，本身就蘊含濃厚的雲端、邊緣端分工思考。值得注意的是，Azure 未在邊緣運算布局硬體晶片處理器，探索背後原因，可能與 Azure、Intel 長期性的合作關係有關，因此在產品布局上與 AWS、GCP 明確不同。

最終在物端，亦即配置在終端設備的晶片處理器，AWS 在 2015 年由於推出 Alexa 智慧虛擬助理 (Virtual Assistant)，先於 GCP、Azure 提出應用於智慧終端裝置的 Alpine 晶片處理器。GCP 則是到 2018 年才推出強調安全性的 Titan 晶片處理器，希望藉由底層晶片的加密、驗證方法，來確保雲端環境的安全性。Azure 雖然未推出自我品牌的晶片處理器，但一般來說，會認定 2018 年與 MTK 共同開發的 MT3620 是 Azure 關注於設備中端所用晶片處理器的先聲。藉由 AWS、GCP、Azure 三者，對於物端晶片處理器產品投入來看，AWS、GCP、Azure 三者的產品投入的思考與三者希望投入的應用市場有關，因此呈現較為多元化的發展樣貌；然而，卻可歸納出兩項共同特徵：第一，物端、終端設備晶片必須滿足低功耗、小體積的電子設計需求；第二，物端、終端設備的晶片處理器必須強化軟體、硬體的安全性設計，期望藉此來降低整體雲端環境遭受資訊洩漏、攻擊的風險，將其視為安全過濾的前緣。

藉由雲端，以及廣義上非雲端的邊緣運算、終端設備的晶片處理器布局，非常清晰地反映出 AWS、GCP、Azure 皆是將聯網底層的晶片處理器，視為是整體雲端服務的環節；如以雲端服務供應商的角度來看，以雲端服務需求設計的處理器，即是雲端。然而，發展歷程相對 Azure 而言，並未與 Intel 建立緊密合作關係的 AWS、GCP，在晶片處理器研發，有較高的自主布局意識，對於產品生態系的主導能力亦較強力。

AWS、GCP、Azure 邊緣運算晶片布局策略比較

依循 AWS、GCP、Azure 的布局產品比較分析，在取得晶片處理器能力的策略以及方法上，三者也顯現出非常大的差異。AWS 主要採取併購、獵才的模式，以在內部孕育晶片設計能力，不過在 2018 年以前發表的處理器產品，皆保留了被併購對象的品牌，比如 Annapurna Labs；這也意味著 AWS 對於晶片處理器的布局，除了補充雲端服務的功能之外，也希望藉由併購、開發，拓展 AWS 的晶片處理器事業板塊。

與 AWS 相同，強調晶片處理器能力自主化的 GCP，其能力孕育的策略則是採取全然的自我投資、研發，並且在產品的原型驗證階段，就以 GCP 的資料中心作為試驗；再以這些經驗為基礎，開發出 Cloud TPU、Edge TPU 兩項代表性晶片處理器產品。不過，值得再次陳述的是，AWS、GCP 皆選擇以 ARM 架構作為自主開發的策略，在兩者的推動之下，ARM 架構與 x86 架構的產業競爭態勢，也值得後續持續觀察。

然而，Azure 與 AWS、GCP 兩者布局策略就存在著極大的差異。Azure 有關於晶片處理器的布局策略，主要是採取提供認證的模式，這樣的模式可以收取到兩個效益：首先，這些取得 Azure 認證的 IC 設計與資訊硬體製造商，可以作為拓展 Azure Sphere 物聯網平台的前緣，Azure 的生態系也可以保有相對較高的開放性、應用性，比如 Azure 與 Qualcomm 的合作，便利於 Azure 進入 5G 的應用市場；其次，間接布局，而非直接投入自主研發的方模式，可以免除與 Intel 等合作對象的市場競爭矛盾。

彙整上述所言，嘗試比較 AWS、GCP、Azure 對於邊緣運算晶片處理器的布局策略，可以發現影響其策略的選擇有三：第一，雲端服務供應商對於「未來市場」的認定，比如 GPC 就明確將 AI 視為關鍵的收益來源，因此打造出適用於 Tensor Flow 學習架構的 Cloud TPU、Edge TPU；第二，雲端服務供應商與合作夥伴的競爭合作態勢，Azure 的布局策略便深受 Microsoft、Intel 長期合作的影響，便是一個例子；第三，雲端服務供應商對於「產品生態系」的構想，這是其中最為隱微，卻又是最為關鍵的決策因子，這部分可以細分為兩種類型的討論，一者是產品生態系的開放程度，比如採取模組、開發板認證的模式，便與直接提出自有產品的模式大有不同，另一者則是涉及到不同業種「分潤模式」的界定，這部分意味雲端服務供應商，與 IC 設計商、資訊設備製造商之間的利潤分配課題，對此，GCP、Azure 就呈現出明顯的差異。

雲端服務供應商的未來布局方向

如果以 AWS、GCP、Azure 布局「非雲端」的邊緣運算晶片處理器的時間點來看，「雲至晶片」(Cloud to Chip, C2C) 的發展策略，並非是雲端服務供應商因應電信或通訊服務供應商、資訊系統與硬體供應商提出「邊緣運算」的應對之道，而是期望

將雲端服務深入到終端的長期战略布局，無論是 AWS、GCP、Azure 其根本的想法皆相似，必須藉由晶片處理器的布局，來確保雲端服務的功能、收益可以更加優化。雲端服務供應商，在面對傳統高效率運算 (HPC) 與新興 AI 模型的商業課題，前瞻未來，雲端服務供應商之於邊緣運算與晶片處理器的布局，應可歸納三個主要方向：「擴充 ARM 架構處理器應用」、「強化 Edge AI 推論晶片功能」、「選擇更開放化的產品生態系」。上述三者也分別代表著雲端、邊緣運算、物端，三個階層的方向。

擴充 ARM 架構處理器應用

藉由 AWS、GCP、Azure 三個雲端服務供應商在 2015 年之後，對於晶片處理器的布局來看，ARM 架構處理器的應用，成為衍生出的重要子題之一。其中，最為明顯的案例為 Amazon 與 AWS，不僅強化對於具有 ARM 架構設計能力的企業進行併購，更在 2018、2019 年，分別提出名為 Graviton、Graviton2 的雲端晶片處理器，而 Graviton3 更有可能在 2021 年時被提出。除了 AWS 之外，與 Intel 有長期合作關係的 Microsoft，其企業內部也不乏採用 ARM 架構晶片處理器應用於自身 Azure 雲端伺服器、資料中心的討論；此舉，不難發現會對 Intel 等 x86 架構處理器帶來威脅。

不僅是在雲端的層級，比如取得 Azure Sphere 認證，並且預期配置在終端設備的 MT3620 也是採取 ARM 的架構進行設計。前瞻未來，以 AWS 做為主力的雲端服務供應商的需求帶動之下，基於 ARM 架構晶片處理器應會獲得到更多的應用與青睞，不過，這樣的擴充態勢，反映在處理器的市場採用比例上，仍是一個相對緩慢的過程，不僅是雲端服務供應商自製晶片的進度，其他包括半導體精密製程，與 Nvidia 併購 ARM 之後，對於 ARM 的整體規劃藍圖，都會是影響此一發展方向的變因。

強化 Edge AI 推論晶片功能

AI 及其衍伸出的數據分析服務收益，皆是 AWS、GCP、Azure 大量投以關注的潛力市場。過去，具有較佳運算能力的雲端，過去面對到兩個挑戰：首先，傳統大型數據、資料量的 AI 訓練模式，需要有穩定的數據來源，才能夠建立起龐大的 AI 訓練數據湖 (Data Lake)；其次，智慧終端的興起，這些終端裝置，也必須在沒有雲端的支援之下，獨立執行 AI 的智慧化功能。在這兩個主要的挑戰之下，雲端服務提供商開始提出「邊緣運算人工智慧」(Edge AI) 的解決方案；能夠配置在邊緣運算的設備，或是嵌入終端裝置執行 AI 推論的晶片處理器，就成為 Edge AI 不可或缺的環節。

Google 在 2018 年發表的 Edge TPU，以及 AWS 同樣在 2018 年發表的 Inferential 都屬於 Edge AI 推論晶片的範疇，其中，Google 更是建構起 Cloud TPU 與 Edge TPU

的完整產品生態系，也顯示出整合硬體、軟體兩端，掌握 AI 整體市場的戰略企圖心。前瞻未來，雲端服務供應商預期會更強化 Edge AI 的推論晶片、模組、開發板功能，而有兩個發展方向：第一，硬體層面朝向更微型化、低功耗的方向演進；第二，晶片硬體功能，將更進一步與軟體堆疊 (Software Stack) 架構結合成為全面解決方案。

選擇更開放化的產品生態系

雲端服務供應商面對的挑戰，已不全然在運算效率的提升，隨著 AI 與數據分析服務的興起，雲端服務供應商轉而思考的是，如何藉由大量、有價的數據，優化雲端服務的分析模型、演算法 (Algorithm)。然而，雲端演算法的孕育與優化，有賴在不同應用場域，可以獲取到大量、穩定、多元化的數據，因此，如何建構一個「開放化」的產品生態系，並且讓更多終端設備產生的數據，能夠傳輸到自身的雲端平台之中，就成為 AWS、GCP、Azure 三個雲端服務供應商，必須持續去面對的發展課題，而這樣的課題，也反映在 AWS、GCP、Azure 對於邊緣運算晶片處理器的布局策略。

Azure 針對 IC 設計商、資訊硬體設備商提供「驗證」的模式，可以被視為硬體層面產品生態系「開放化」的案例；在軟體層面，AWS 在 2018 年推出，宣稱能夠支援 TensorFlow、Apache MXNet 與 PyTorch 深度學習架構，則是另一種「開放化」的表徵；也正如前述，產品生態系的開放與否，其實涉及到的是不同業種的「分潤」與商業營收範圍的「界定」。前瞻未來，不同的雲端服務供應商，皆會以不同的方式，去建構一個更為開放的產品生態系，而這也是台灣廠商必須持續關注的議題所在。

附錄

英文名詞縮寫對照表

AI	Artificial Intelligence
ASIC	Application Specific Integrated Circuit
AWS	Amazon Web Services
C2C	Cloud to Chip
CSP	Communications Service Provider
DDS	Data Distribution Service
EC	Edge Computing
EC2	Elastic Compute Cloud
ETSI	European Telecommunications Standards Institute
GCP	Google Cloud Platform
HPC	High Performance Computing
IaaS	Infrastructure as a Service
IB	In-Band
IEEE	Institute of Electrical and Electronics Engineers
IIC	Industrial Internet Consortium
IoT	Internet of Thing
MEC	Multi-access Edge Computing
OOB	Out-of-Band
SDN	Software-Defined Networking
SoC	System on a Chip
TPU	Tensor Processing Units

中英文名詞對照表

小型基站	Small Cell
分散式運算	Distributed Computing
加速器	Accelerator
多義性	Polysemy
即時回應	Real-Time Response
延遲率	Latency
物聯網	Internet of Thing
基站	Base Station
張量處理器單元	Tensor Processing Unit
軟體定義網路	Software-Defined Networking
軟體堆疊	Software Stack
晶片組	Chipset
虛擬化	Virtualization
開放霧聯盟	OpenFog Consortium
資料中心	Data Center
資料分散式服務	Data Distribution Service
運算能力	Computing Power
網路架構設計	Network Architecture Design
數位資產保護	Digital Asset Protection
數據湖	Data Lake
歐洲電信標準協會	European Telecommunications Standards Institute
邊緣節點	Edge Nodes
邊緣運算	Edge Computing
邊緣運算人工智慧	Edge AI
霧運算	Fog Computing



發行所	財團法人資訊工業策進會 產業情報研究所(MIC)
地址	台北市 106 敦化南路二段 216 號 19 樓
電話	(02)2735-6070
傳真	(02)2732-1353
全球資訊網	https://mic.iii.org.tw
會員服務專線	(02)2378-2306
會員傳真專線	(02)2732-8943
E-mail	members@micmail.iii.org.tw
AISP 會員網站	https://mic.iii.org.tw/aisp

以上研究報告經 MIC 整理分析所得，由於產業變動快速，並不保證上述報告於未來仍維持正確與完整，引用時請注意發佈日期，及立論之假設或當時情境。
著作權所有，非經 MIC 書面同意，不得翻印或轉載