

ITRI IoT PaaS大數據服務平台介紹

Introduction to ITRI IoT PaaS Big Data Serving Platform

張維君

陳玉倫

蔡國手

Wei-Chun Chang, Yu-Lun Chen, Kuo-Shou Tsai

中文摘要

本論文提出一大數據服務平台，ITRI IoT PaaS。本平台設計的主要目標在於提供產業領域對蒐集來自機器或設備感測器大數據資料的處理與分析服務。許多產業領域中，機器或設備所依附的感測器會不間斷且高頻率的產生資料，這些大量資料需要空間與時間進行儲存與處理，由於量的關係，其效率與效能成為研發服務平台主要探討議題，且大量資料在輸入分析模型的前處理，對於探索智能產線提昇效能的因素，具有絕對的影響力，這是形成工業4.0的基礎概念。對於大數據資料的處理與分析，是本平台設計所挑戰議題，藉由Mesos與Spark分散式架構的整合，提出有效解決方案。驗證本平台的目標，我們採用電腦計算能力、資料寫入處理時間及大量資料查詢時間作為平台運作量測指標；相關結論提出驗證平台對大數據資料的處理與分析效能。

Abstract

A novel platform as a service (PaaS) is presented in this paper. The main objective aims to provide a service for big data preprocessing and analysis collected from machines/sensors. In main domains, massive data is collected from sensors which are associated with different machines or facilities. The data require to be stored and processed efficiently and effectively according to the big data characteristic. Volume of the data leads to main issues of effectiveness and efficiency to be explored in this paper. The main goal is to feed the data into analytic models for finding improvement factors from current management process which is also the main concept of Industry 4.0. The challenges of this paper are to deal with the manipulation and analysis of big data. The ITRI IoT PaaS is proposed to provide the services as a solution for the challenges based on the integration of distributed computing architecture of Mesos and Spark. Data access and query time, computing power, sensor connection quantity, and analytic time measurements are provided to illustrate the efficiency and effectiveness of the solution. The conclusion is presented in the end of the paper.

關鍵詞(Key Words)

物聯網(Internet of Thing ; IoT)

大數據(Big data)

工業4.0(Industrial 4.0)

雲端服務(Cloud Service)

軟體服務(Software as a Service ; SaaS)

平台服務(Platform as a Service ; PaaS)

基礎設備服務(Infrastructure as a Service ; IaaS)

虛實整合系統(Cyber-Physical Systems ; CPS)

1 · 產業的未來智能產線

因應產業科技的進化，各產業在第三次工業革命即引入科技的協助，將自動化元素引入產能線，在自動化的協助下，各產業由人力或機械力，達到自動化產能的有效提昇。接下來，產業思考精進的目標也是同樣朝著提升產能的方向與做法前進，但在進化了人力與機械至自動化的階層之後，到底有何方向是產業能精進並改善其產能的目標？這是工業4.0的啟動思緒，創造智能產線提高產能/人年。

台灣製造業97%為中小企業[1]，例如台灣工具機產業提供商的技術層級偏中低階，終端加工廠規模亦以20台以下小型工廠為大宗，產業在發展過程不論是對內或對外皆遭遇不小的阻力與挑戰。對內，產業困境在於產業平均規模偏小，對於投資企業所需資訊自動化的軟、硬體設備上缺乏經費，因此無法支持自行投資建置先進製程資訊管理平台，且後續維護與更新之費用更是一筆龐大負擔，依據一份調查[2]，以一個資訊系統的生命週期總花費計算，維護與更新約占百分之50至60，這對於一般企業是一份不小或是無法承受的負擔，另一方面即使企業願意投資進行科技改造產業的目標，然而，一個現實問題在於不易聘得跨領域之優秀資訊系統人才，由於未來發展性與待遇問題，常令這些優秀人才留駐於大型企業。對外而言，本土企業的國際化發展，其中一項最大的挑戰來自於其他國家競爭對手的快速成長[3]，在國際競合的時代，資訊科技是進入市場的基本門檻，在國際市場科技運用日益更新的壓力下，各產業邁入21世紀的主要目標是結合工業4.0[4]，創造出智能產線，提高年、人平均產值，一個最好的例子是勞斯萊斯

[5]，全球知名的飛機引擎製造商，在邁入新世紀的企業挑戰，轉型由賣引擎改為賣監控時數，僅僅投入4%的成本進行，利用其全球售出的4600具引擎及其內的感測器(約30個)，透過物聯網與大數據資料蒐集進行即時監控，達成公司由製造業轉型為大數據服務的目標。製造業對產線的進化是提升市場競爭力的一項重要指標，這需要對製程的各項資料進行蒐集與分析，也就是將製程程序訊息化，再從中取得重要資訊反饋進行製程的進化，這過程即是所謂的智能化，上述勞斯萊斯的例子就是一個明顯的案例。對本國的製造業來說，訊息化所需的門檻過高，訊息化門檻包含以下幾個面向：軟體開發，例如持續追蹤使用記錄及機台運轉資訊，並進行對產線大量資料收集與分析；軟體測試，例如敏捷式開發與持續整合測試；營運系統，例如日誌報告、分析趨勢呈現、預測預警、事件警報、工廠製造迴授控制等高附加價值服務。目前的訊息化相關服務產品是無法滿足我國主要中小型製造業及用戶對應之市場區段，部分針對單機提供之軟體或雲端服務(針對軟、硬體服務)，亦無法有效全面提升產線智能，產業亟需整合機台物聯網與製造雲服務平台，提升台灣整體產業之上、下游整合競爭力。

在雲端服務的世代裡，軟體服務(Software as a Service) 與基礎設備服務(Infrastructure as a Service) 是大家較熟悉的兩個服務層，是一般較容易理解的類軟體及硬體層，此兩大類雲端服務在過去的發展階段獲得大多數公司與研究機構的關注與投入相當的努力，也獲得可觀的進展。至於另一個服務「平台」(PaaS)到底能為客戶做甚麼？解決甚麼問題？帶來甚麼直接或間接利益？這是建構平台服務必須要

挑戰與回答的問題。近年來，越來越多的研究團隊已開始將研發的目光聚焦於此服務。平台服務定義為提供應用程式工程師開發、執行與管理應用程式的雲端系統環境。此服務將應用程式開發的系統層(例如資料儲存與分析處理)以API介面方式提供應用程式介接使用，提高開發效率與降低開發與維護成本。

建構高效能、高容量(計算與儲存)、高穩定性與易用性高雲端服務平台系統是本研究主要挑戰的目標。我們的目的是提供客戶端軟體服務的作用平台服務。我們的解決策略是發展物聯網雲端服務平台技術，以大幅降低產業發展信息化的門檻。這個目標產生的動機是由以下幾個重要的議題，來推動我們挑戰平台建置技術：(1)沒有找到產業無法達成量產目標根本原因，使得產品瑕疵一再發生；(2)沒有採取適當預防措施預防產線錯誤，無法避免其他類似問題；(3)先進數位製造系統與其附屬感測器所產生的大數據資料處理與分析；(4)彈性與少量多樣化的商品趨勢，牽引著產線的調整達成目標；(5)應制定開發Cyber-Physical System (CPS) 應用服務平台，以凝聚產業製造優勢，朝向工業4.0概念實現邁進。透過對以上議題的分析，制定本平台研發主要目標為提供一垂直整合度高的資料收集與儲存平台，負責收集物聯網裝置的各項資料。同時提供一具有高性能的分散式運算平台，負責執行客戶端應用服務所上傳的資料分析應用程式。

以下本文對大數據資料處理與分析列於下一節；第三節定義ITRI IoT PaaS的服務架構；平台與客戶端的主要通訊介面定義在第四節；第五節採用效能分析驗證ITRI IoT PaaS的執行效能；對於本論文服務平台研究的總結列於最後一節。

2 · 大數據資料處理與分析

透過物聯網的建置與訊息傳播技術的更新，我們生活與工作的環境所產生的各項資料被有系統且快速的聚集。這是在物聯網盛行前所無法達到的資料規模數(PB規模以上)。國際研究暨顧問機構Gartner預測今年(2016)全球

有64億個物聯網裝置連線，比2015年成長30%，這個數字在2020年，會達到208億個(表一)[6]。

表 1 物聯網安裝分類表(單位: 百萬)

類別	2015	2016	2020
一般消費者	3,023	4,024	13,209
企業	1,880	2,368	7,288
總計	4,902	6,392	20,797

資料來源: Gartner[6]

這說明了物聯網的時代，伴隨而至的挑戰是對資料提供有效率的處理。在這些常用的數據資料隨著網際網路搜尋引擎的盛行，連動的引出更大量的再生資料。大數據一詞因應而生[7]來代表這個世代資料量的面向，伴隨著大量原始資料的產生，傳統的資料處理方法與模式也跟著進化式的改變，資料內蘊含著我們對生活及工作的處理與對應模式，利用資料的特性可勾勒出目前工作的優點與缺陷，再提出有效及可行的改良管理與作業，這是智慧產能的重要議題，也代表資料處理邁入新境界的一個里程碑[8]。

2.1 大數據資料面臨的挑戰

對大數據資料處理的挑戰已定義為四個主要面向[9][10]: 量(Volume)、速度(Velocity)、多變性(Variety)、價值(Value)，也就是「4Vs」。大多數的機構或企業對大數據資料的處理仍著重於前三個面向，對於第四個則有較多不同的見解與定義。藉由這四個V的導引，可了解目前的挑戰在於如何處理大量的資料達到TB(10^{12})，甚至於是PB(10^{15})或更高。這樣規模的資料量在當前許多的領域是常見的，所以第一個挑戰在於處理這樣高量的資料群。下一個挑戰在於資料的處理速度，這方面較與產業領域相關，例如國家安全領域，對於恐怖活動資料的處理相對來說就是越快越好，但對於部分產業領域並不在意資料的即時處理，但也必須在合理的處理時間範圍內，例如工具機產業，由於大量資料本質上就已具有處理的挑戰，因此如何在合理時間內處理產業終端感測器或設備的資料，對於服務平台的架

構設計是一個有趣的挑戰。第三個挑戰是來自於資料多樣性的特質，傳統是資料庫設計是針對將產業資料先進行結構化分析後產生資料綱要，後續處理即依據此綱要進行，但進入大量資料的世代，資料來源不僅大且多樣化，非結構化資料集是其中重要的一部分，為使平台的使用度提高，處理此類資料成為必然的一個挑戰。最後的挑戰是來自於資料所賦予的價值開發，若是少了這一挑戰，前面的挑戰所具備的意義就不大，因為一群原始資料如同一堆原料般，無進行加工處理就不成產品，因此構成了平台設計價值的另一個重要挑戰。除了上述的挑戰外，對部分特殊領域會有更深度的挑戰需要提出與討論，但並不在本論文的範圍內，因此省略不進行更深入的討論。

2.2 資料分析方法

有品質的原始資料，容易導向高效率與高價值的資訊，資料的不良品質情況包含不完整性、雜訊與不一致性等[11]，資料的處理意味著在原始資料寫入資料庫前，必須進行前置處理，以達到高品質的目標。主要著墨於資料寫入前處理，這在討論資料分析前也屬於重要的一部分。

前面對大數據資料的挑戰討論中，提到第四個挑戰，也就是如何處理原始資料後提高其附加價值，重要的一個面向即是對資料進行分析以獲取有用的資訊再加利用，這就是資料的增值服務。目前業界採用的分析方法較早期由谷歌 (google) 企業提出的一個程式模型 MapReduce[12]，由於其企業特性常須面對使用者要求在大量的資料裡搜尋目標資料，其架構為一組映對 (map) 函數來並行處理大量原始資料，再由另一組組合 (reduce) 函數進行運算合併的作業，這樣的分析架構，在本質上已具有平行處理的能力，然而，這是針對網際網路型搜尋型態的作業，假設前提是針對大型資料具有淺平而規律的結構，以及分布於數個機器與儲存裝置上，對於需大量運算類型的分析工作，則不具有任何的優勢。

由於 MapReduce 對於資料的中介處理時須與儲存裝置進行大量的輸出入作業，因此對

於資料處理會造成缺乏效率的問題，為解決這個問題加州柏克萊大學的 AMP 實驗室開發一個開源叢集運算架構 Spark[13][14]，最重要的特性是將需要處理的資料以及中介資料皆存放於記憶體中，對於處理的效率與速度有顯著的效能提昇。

2.3 雲端平台服務

目前各家通訊與科技大廠皆積極投入平台服務的開發，例如：

■ IBM Bluemix[15]: 以 Cloud Foundry 為基底，建構而成的開放式雲端服務平台，主要在提供 IBM 軟體使用權給網際網路與行動開發人員，進行快速應用程式開發的環境，目標在大幅減少創建和配置應用程式所需開發時程。

■ 甲骨文雲端平台 (Oracle Cloud Platform) [16]: 針對合作夥伴及產品用戶的開發人員，簡化開發應用程式週期，提供一致化服務平台，拓展甲骨文軟體服務 (SaaS) 開發應用。

■ 亞馬遜 (Amazon) PaaS[17]: 是各企業中，最早大規模投入雲端服務建置領域，其服務平台主要提供應用程式資訊的儲存與訊息的傳遞，主要目標是進行企業雲端服務的垂直整合，為其客戶提供更完整的服務。

■ 研華 WISE-PaaS[18]: 是國內少數具有能力開發平台的企業，憑藉著其工業電腦全球市場的佔有率第一[19]，進行其與客戶端的垂直整合開發平台，主要採堆積木方式，透過其夥伴合作方案，協助客戶端開發人員共同加速打造專屬物聯網應用軟體，初期為無償開放使用，已於 2015 年採收費制 (每位客戶: 2 萬美元/年)[20]。

綜觀上述的全球及本國大型企業對於服務平台的研發現況，我們可以歸納出幾個關於服務平台研發的重要面向，第一就是平台跨領域的服務是一個巨大的議題與挑戰，對於這些企業的研發成果也無法對平台服務範圍的廣度與深度有效的涵蓋，大部分仍是以企業既有客戶群進行垂直整合為出發點；其次，平台的研發，目前仍較著重於大量資料的寫入與取出的處理，對於增值產出的部分仍缺乏較具體的成果；另外，對於其他傳統產業如工具機產業

領域的適用服務平台闕如，這也激發本論文對於傳統產業打造適性的服務平台的研究動機，而非僅考量一般通用於類似社交平台與搜尋引擎等網際網路型服務領域。

3 · ITRI IoT PaaS架構

本平台建構主要目標為提供一資料收集與儲存平台，負責收集物聯網裝置的各項資料。同時提供一具有高性能的分散式運算平台，負責執行應用軟體服務所上傳的資料分析應用程式。植基於分散式的架構，將工作進行分流，配置不同伺服器達到運算負載平衡的並行處理，期能對多感測器資料處理與分析的有效性能表現。平台架構及其內含各模組的設計分述於以下各小節內。

3.1 平台架構

本平台架構採分散式架構(圖1)設計，主要包含三個模組，分別為註冊模組、資料收集模組與分散式運算模組。

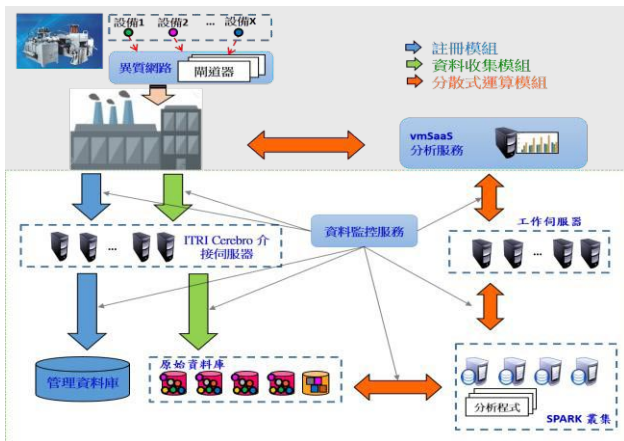


圖 1 ITRI IoT PaaS 架構圖

ITRI IoT PaaS平台佈署後可經由營運支援介面透過註冊模組所提供的應用程式介面(APIs)來進行註冊使用者與服務，註冊之後，才可以提供客戶端進行機器設備的註冊，在資料收集工作方面，已註冊且認證過的機器設備與感測器可使用資料收集模組所提供的APIs，將資料傳送至平台資料庫儲存。當資料開始寫入後，即可開始進行資料分析，例如訓練一個商品退貨(RMA)預測模型，此時雲端軟體服務程式可以透過分散式運算模組所提供的APIs將資料分析程式上傳至平台，並使用此平台所提供

的分散式運算環境來進行運算，當運算完成，可以使用API將預測模型運算(可進行即時或批次的預測運算)結果取回(如圖2)，其中“ACL模組”目前僅提供基本使用權限控制，亦可用來延伸與雲端垂直服務層使用權限進行整合，達到對雲端所有服務一致的控制。

3.2 註冊模組

平台的資料註冊主要是針對機器設備與感測器，為便於對資料的後續收集與管理，將一企業組織定義為三個階層，亦即廠區(Company)、機器設備(Machine)與感測器(Sensor)。註冊模組對於任一階層物件的識別，以下列組合定義進行。

<C, M, S> 其中，

C：廠區物件

M：機器設備物件

S：感測器物件

客戶端採用此組合定義與ITRI IoT PaaS對其所屬的各階層物件進行註冊作業。

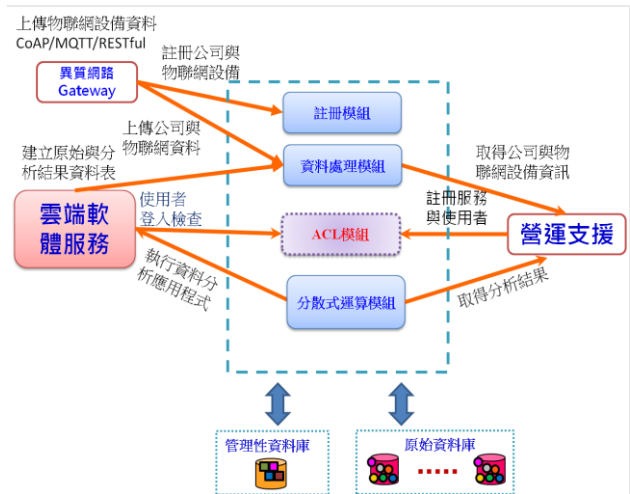


圖 2 平台模組與周邊關係圖

本模組作業資料以一結構化資料庫(MySQL)儲存相關管理性資料，註冊模組的資料庫作業程序以標準SQL語法進行，選用此類資料庫為管理性資料本身即具結構化特性，且資料量(MB)在作業能力範圍內，效能成果佳。

3.3 資料收集模組

儲存與處理來自於感測器終端設備的大資料量是本平台研發的主要目標，因此對於原始資料的特性須經由分析後設計其綱要結構，以決定最佳儲存與處理效能的目標資料庫，由於各產業結構與設

備性質大異其趣，為求最具彈性化設計，也回應上述討論有關大數據資料的多樣性挑戰，因此，在資料綱要設計上採用以下的基本資料元設計：

```
{“key”, “value”[; ...]}
```

Key: 資料屬性名稱

Value: 原始資料值

註：“[...]”表示資料屬性數量可由使用者自行定義，不需要預先定義資料庫綱要(schema)。

對於這類的原始資料，半結構或非結構性是主要特質，模組提供資料以即時或批次方式將感測器資料寫入平台內資料庫，由於資料來源的特性，因此，對於此類資料本模組採用PostgreSQL資料庫作為原始資料的儲存目的地。

3.4 分散式運算模組

為達到對大量資料進行有效即時或批次分析，分散並行運算架構是最適合的選擇，採用Mesos[21]作底層工作節點管理加上Spark的分散運算架構建立分散式計算環境(圖3)，Mesos是一個開放原始碼的叢集管理系統，管理分散應用程式間的資源隔離與分享。

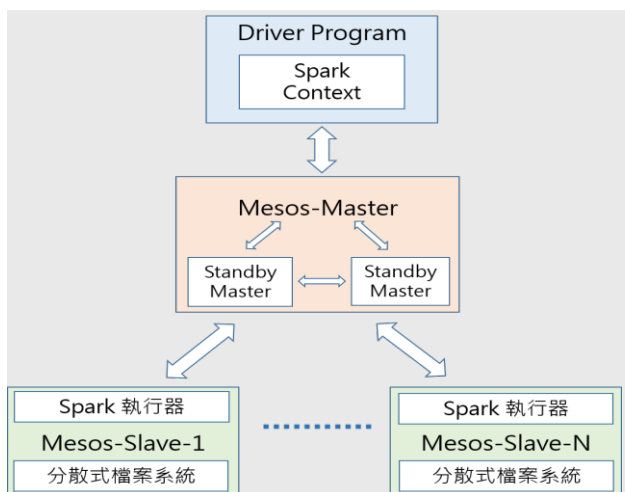


圖 3 spark分散式運算架構圖

由於Spark對於運算資料塊皆存放於記憶體內部，因此在分析成效上有顯著的效率[14]，另一方面，對於叢集內多工作節點的管理，Mesos採用架構具自動產生master管理的智慧機制，所有分散式節點共用一分散式檔案系統。這符合對大數據資料分析的另一挑戰，

目標在建立一分散式計算環境，提供平台使用者進行有效的分析後將結果反饋回終端裝置進行智慧管理後提高產能。

4 · 平台通訊介面設計

本平台對應來自終端感測器的大量資料寫入/讀取處理與智慧分析的服務是由一群高效率的伺服器介面(以下簡稱API)設計，此API集提供終端裝置與平台的通訊需求，藉由高效率與安全的通訊傳輸，達到對產生的大量資料做有效的處理與應用。本設計規範研析RESTful API發行之全球資訊網軟體架構設計風格，考量企業客戶對來自於物聯網大量資料所需的存取與資料分析需求，訂定ITRI IoT PaaS存取API介面規範。使用者依循本規範建置物聯設備端資料存取應用程式及快速地導入業界規範的大數據資料分析方法：如Hadoop/Spark，快速而有效率的建立與存取其所屬基本資料及物聯設備所產生的大量資料；並可針對大量數據資料分析提供擴充功能，有效提高企業產能。本設計規範之資料存取API服務用於物聯設備端與雲端資料儲存平台做自動化介接、即時作業監控、提供資料集存取與導入巨量信息分析程式進行資訊萃取等功能。企業客戶可根據本規範提供的API介面定義，進行相關應用程式介面開發，以提供資料使用者或開發者一致性API，進行取得、搜尋或分析所需的資料集資料，取得有益信息反饋物聯設備進行產能修正，有效提升產能。

資料存取應用程式API介面主要提供一般結構化資料內容存取，不包括資料內容應用層面解析，如複雜數學邏輯運算篩選、應用領域所代表之意義判斷、欄位之間相關性等。

4.1 介面規範準則

■語法規則: RESTful API的規劃，主要目標是要讓企業客戶可以HTTP GET、POST、DELETE方法，存取ITRI IoT PaaS之資料。API呼叫回傳內容格式則以Json為主，API服務路徑採用URI通用語法組成(圖4): 分為服務網址(scheme+host)、資源路徑(API service path)和服務選項(API Query Options)

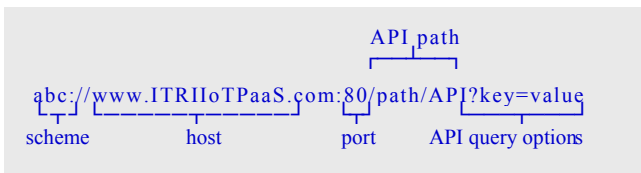


圖 4 URI通用語法結構示意圖

其中，

服務網址：平臺上提供該類別 API 應用服務之協定(目前僅支援 HTTP)及主機網址/網域名稱。

資源路徑：接續於服務網址後，指定某一 API 服務項目路徑名稱。

服務選項：接續於API服務項目路徑後，針對某一API服務，指定所欲進行作業資料參數。

■命名規則：定義API介面的命名原則，提供企業客戶於開發平台應用程式可依循之API呼叫，命名規則依據下列格式定義提供API服務。

{/}{VERSION}{/}{OBJECT}[{/]{SUBOBJ}...

其中，

{/}：必要項目

[]：選擇項目

...：允許多重項目

範例：

“/v2.1/session”：API 版本 2.1，向API伺服器取得作業會期期間使用之辨識碼。

4.2 API 介面類型

主要提供兩大類型服務，即資料異動與取出，分別敘述如下。

■資料異動：

POST：用於新增一企業級客戶、部門廠區(company_uuid)、用於新增及修改機器/感應器的基本管理資料及感測資料、分析服務應用程式的作業部署與啟動。企業自建私有物聯網雲端資料儲存平台，則可用於建立分公司/單位帳號使用

DELETE：刪除機器/感應器的感測資料。

■資料取出

GET：用於向ITRI IoT PaaS取得資料庫的資料，包括會期識別碼、企業客戶列表、機

器/感應器的管理資料、機器/感應器的感測資料、分析服務應用程式的結果、分析服務應用程式的狀態及物聯網雲端資料儲存平台運作紀錄。

5. 平台實作與測試

實現ITRI IoT PaaS服務平台的建置，我們採用最輕量的架構進行平台設計與實行，主要目的在取得平台的壓力參數，作為擴增的架構參考資料。

5.1 系統測試環境

本平台測試實驗的機器及環境以工研院資通所內的基礎設備為主，其主要參考參數資料如表2。測試環境佈署一網際網路伺服器(nginx + uwsgi + Flask)，對應與處理使用者端的Restful API 服務要求，其次，安裝管理性資料庫(MySQL 5.6)與感測器原始資料庫(PostgreSQL 9.5)，分散式分析運算叢集佈署包含一Spark 驅動器、一Mesos-Master及2台Mesos-slave，有關分散式檔案系統供運算叢集使用採glusterDFS。

表 2 ITRI IoT PaaS 系統實作環境參數表

IaaS	工研院資通所 VMs
作業系統	Ubuntu 14.04
CPU cores	4
RAM	8GB
Network	1Gbps

5.2 平台執行成效

由於設計與建置一雲端平台服務對於其它雲端垂直服務具有複雜的依存關係，這可由第2.4節中所討論的方向可得到驗證，因此在業界或學界尚未有統一的量測性效能因子做各平台的比較，因此對平台服務的基準驗證是一項困難的工作，有鑑於此，我們對於本平台佈署完成後進行測試驗證以設計的功能性與時間效能做為參考。在功能完整性方面，平台的設計目標在提供高效能與高容量的資料處理與分析，藉由APIs (參考第4節)的定義與對個別功能測試成功的項目列於表3，列表APIs可涵蓋對雲端平台服務的功能項目。

表 3 API 功能測試表

API	測試目標
GET_SESS	合法企業帳號認證的取得
ADD_COMP	企業用戶註冊廠區
QRY_COMP	企業用戶查詢廠區資料
ADD_MACH	註冊廠區內的機器設備與感測器
GET_MACH	查詢廠區機器設備與感測器資料
ADD_RAW_SNGL	感測器感測資料單筆寫入
QRY_RAW_SNGL	感測器感測資料單筆查詢
ADD_RAW_BULK	感測器感測資料多筆寫入
QRY_RAW_BULK	感測器感測資料多筆查詢
ADD_DAPP	上傳分析程式到Spark叢集
RUN_DAPP	Spark叢集執行分析程式
GET_DAPP_RESULT	取得分析程式執行結果

針對部分應用程式支援介面(API)的執行時間分析列於圖5。經更深入作時間分析，約70%的時間花費在訊息接收後的解密運算，這可提供做為未來改善執行時間，增進伺服器服務更多Restful API 服務需求的數量，目前測試單一網際網路伺服器(參考表2)的HTTP 服務要求負載測試，在完成單一“GET_SESS”API(參考表3)的標準測試下，目前可達1800次/秒的水準。

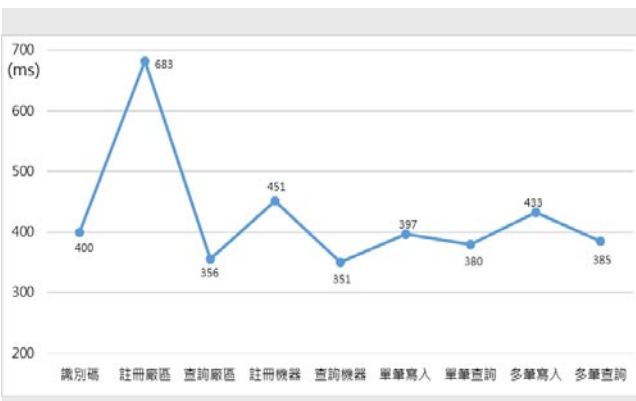


圖 5 支援應用程式介面(APIs)執行時間

對資料處理與分析的平台服務時間效能測試列於表4。

表 4 資料處理時間效能表

註冊	識別碼取得時間(0.4秒)
上傳(單筆)	每個API呼叫時間(0.4~0.5秒)
上傳(多筆)	一次寫入1000筆資料(2.5秒)
查詢(多筆)	50萬筆資料(20秒內)
分析	載入190萬筆資料(5秒內)

其中對於資料查詢的時間效能與傳統關聯式資料庫的比較如圖6。並行式資料存取架構突破傳統硬體架構存取效能，以分散式儲存技術(業界標準Spark並行式資料計算叢集)增快大量資料存取與平行分析效能，資料載入時間比傳統式SQL 伺服器減少80%(如圖6)。

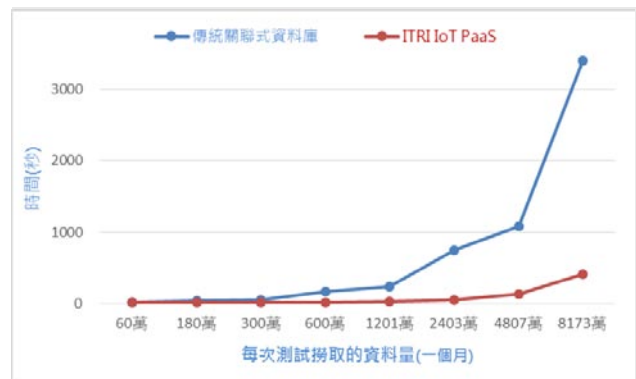


圖 6 查詢資料(50萬筆)時間比較表

6. 結論

本服務平台目標在於有效為客戶端解決終端設備大量資料的處理與分析，從而推導出智慧管理策略，有效提高產能。在資料量與多樣性方面以分散式儲存架構與非結構化資料庫的設計提供解決方案。在速度上，對於來自客戶端的API執行要求，可達到毫秒級的執行時間，由於提供加解密的保護機制，佔據執行的大多數時間，在未來有更精進的空間；資料查詢速度與傳統式資料庫的比較(請參考5.2)，有顯著的效能。在附加價值上，提供分散式計算的Spark環境加上分散式檔案系統的可擴充性，對於進行原始資料分析的規模與複雜度可再擴增，提升其結果的價值與效能。本平台的設計強調產業線引進工業4.0的

智能改造概念，針對相關數據資料接收處理後，供應用數位模擬模型進行機具設計與再精進，藉由分析的資訊進行實體機台控制預期將具有顯著的成效。

實作ITRI IoT PaaS在性能表現上，有顯著成效(請參考5)；在幾個議題上，包含解決更多樣資料的來源格式匯入、效能測試標準、in-memory資料庫的設計及提出分析模式架構加速開發分析程式的時程等，仍具有改善的空間，對這些議題的深入探討，可做為精進平台架構設計的參考，對產業的助益，可提昇產能/人年的效率與避免投入過高的成本，在產業進入高競爭力的國際市場時，具備優異的智能產線，這是本服務平台持續研發的目標。

參考文獻

- [1] 經濟部中小企業處, “中小企業家數,” 台北, 2015.
- [2] N. Huff, C. Glover, and M. Hillman, “How software maintenance fees are siphoning away your IT budget and How to Stop It,” 2014.
- [3] 經濟部國際貿易局, “研析我國工具機與主要競爭國產業發展現況,” 2014.
- [4] W. MacDougall, “INDUSTRIE 4.0 Smart Manufacturing for the Future,” 2014.
- [5] 辜樹仁, “勞斯萊斯不賣飛機引擎改賣「飛行時數」,” 天下雜誌, 2016.
- [6] R. van der Meulen, “Gartner Says 6.4 Billion Connected ‘Things’ Will Be in Use in 2016, Up 30 Percent From 2015,” Gartner, 2015.
- [7] C. Snijders, U. Matzat, and U.-D. Reips, “‘Big Data’: Big Gaps of Knowledge in the Field of Internet Science,” *Int. J. Internet Sci.*, vol. 7, pp. 1–5, 2012.
- [8] V. Koch, S. Kuge, R. Geissbauer, and S. Schrauf, “Industry 4.0: Opportunities and challenges of the industrial internet,” 2015.
- [9] D. Laney, “3D Data Management: Controlling Data Volume, Velocity, and Variety,” 2001.
- [10] Oracle Enterprise Architecture, “An Enterprise Architect’s Guide to Big Data,” 2016.
- [11] 曾憲雄, 蔡秀滿, 蘇東興, 曾秋蓉, and 王慶堯, “資料前置處理,” in 資料探勘, 旗標出版社, 2008, pp. 3–1 ~ 3–29.

- [12] J. Dean and S. Ghemawat, “MapReduce: Simplified Data Processing on Large Clusters,” in 6th Symposium on Operating Systems Design and Implementation, 2004, pp. 137–149.
- [13] M. Zaharia, “Spark: In-Memory Cluster Computing for Iterative and Interactive Applications,” in NIPS 2011 Workshop on Algorithms, Systems, and Tools for Learning at Scale, 2011.
- [14] J. A. Scott, *Getting Started with Apache Spark*, First. CA: MapR Technologies, Inc., 2015.
- [15] IBM, “IBM Bluemix,” 2016. [Online]. Available: <https://console.ng.bluemix.net/>.
- [16] Oracle, “ORACLE Cloud,” 2016. [Online]. Available: <https://cloud.oracle.com/home>.
- [17] Amazon, “AWS Elastic Beanstalk,” 2016. [Online]. Available: <http://aws.amazon.com/tw/elasticbeanstalk/>.
- [18] ADVANTECH, “WISE-PaaS,” 2016. [Online]. Available: <http://www.advantech.tw/embedded-boards-design-in-services/wisepaas>
- [19] Advantech, “ADVANTECH - Enabling an Intelligent Planet,” <http://www.advantech.tw/>, 2016. [Online]. Available: <http://www.advantech.tw/>.
- [20] 吳姿瑩, “研華PaaS平台改收費制,” 工商時報, 台北, 11-Feb-2015.
- [21] Apache, “Apache MESOS,” 2016. [Online]. Available: <http://mesos.apache.org/>.

作者簡介

張維君



工研院資訊與通訊研究所資料中心架構與雲端應用軟體組工程師，英國曼徹斯特大學電腦科學博士。

陳玉倫



工研院資訊與通訊研究所資料中心架構與雲端應用軟體組副組長，美國威斯康辛大學麥迪遜分校電腦科學碩士。

蔡國手



國立交通大學資訊管理研究所碩士，工研院資訊與通訊研究所資料中心架構與雲端應用軟體資深工程師。專職程式設計、偏好python語言，夢想完成百岳。