

巨量資料即時查詢技術研究與效能評估

謝俊偉 Chun-Wei Hsieh

鍾依芳 Yi-Fang Chung

葉文宏 Wen-Hung Yeh

摘要

巨量資料 (Big Data) 所涉及的資料量規模十分龐大，是現行的資料庫技術所無法輕易處理的資料量，對現行資料庫技術而言是一大挑戰。雖然Apache Hadoop [3] 開源專案，開放了分散式檔案系統(HDFS)、Map/Reduce[3]、HIVE[4]、HBase[5] 等技術，為巨量資料的儲存、處理、分析提供了一個分散式運算解決方案，但HIVE或Map/Reduce機制，無法滿足資料分析師或資料科學家與巨量資料即時互動的需求。本研究的目的，即在探討目前在Hadoop平台上，數個具備即時互動查詢機制的開源技術發展現況，並針對其中一種技術解決方案，Cloudera Impala，進行深入的分析與效能評估。透過實際測試評估發現，Impala效能相較HIVE 0.10版本，效能可以提升到5倍以上，甚至能夠達到20倍。另外也發現目前Impala的一些缺陷，如資料表JOIN順序會影響效能表現，因此在設計時需要特別留意。

關鍵詞

巨量資料

開源專案

即時互動查詢

impala

## 基於 Hadoop 多叢集環境的工作分派系統及方法

黃俊翔 Chun-Hsiang Huang  
林修民 Hsiu-Ming Lin  
黃瀨瑩 Jing-Ying Huang  
蔡慶堂 Ching-Tang Tsai

### 摘 要

Apache Hadoop 是目前主要的巨量資料處理開放原始碼解決方案之一，使用者可以基於 Apache Hadoop 提供的分散式運算處理框架 MapReduce 來開發巨量資料應用程式。由於 Hadoop 無法將異質硬體設備整合在單一叢集環境，因此企業需要根據設備規格來建置多套 Hadoop 叢集環境；而多叢集環境的運算工作管理將成為 Hadoop 平台管理者面臨的挑戰之一。在這篇論文我們提出一個基於 Hadoop 多叢集環境的工作分派之系統及方法，能夠根據執行程式特徵、待處理資料特性、與電腦叢集的動態行為，選擇最佳的執行環境。這個系統可以降低不同運算特性的工作之排程等待時間，有效的加快運算分析的速度，並提升整體資源使用效率。該系統係由包括叢集監控模組、叢集特徵模組、工作資料分析模組、工作程式分析模組以及執行環境選擇模組所組成。提供 Hadoop 多叢集環境的工作分派的方法係透過掌握叢集特徵、監控叢集運作情形、分析運算資料特性與程式運算特性等影響參數，進而運算比對找出最合適的叢集，再將用戶工作，包含用戶程式與輸入資料派送到最適合的叢集中執行。

### 關鍵詞

Hadoop、MapReduce、Job dispatching、Job profiling、Scheduler

林逸修 Yi-Hsiu Lin  
謝萬蒲 Wan-Pu Hsieh  
周自強 Tzu-Chiang Chou  
許士榮 Shih-Jung Hsu  
何佩玲 Pei-Ling Ho  
鄭復榕 Fu-Rong Cheng  
吳冠峯 Kuan-Feng Wu

### 摘 要

中華電信公司擁有涵蓋全台灣最多的 PSTN(Public Switched Telephone Network)公眾電話網路，其中包含大量的 SS7(Signaling System No. 7)信號資料可供研究與分析，除可偵測異常通信問題與異常通信型態外，還可發展相關應用服務。在過去沒有開源雲端環境下，由於信號資料量過於龐大，無論是使用關聯式資料庫，或是使用 Memory Disk 方式來嘗試實作，均無法提供有效的解決方案。因此，在雲端環境發展成熟之後，嘗試使用 HBase/Hadoop 雲開源架構，始能提供呼叫訊務資料彙集分析功能完整解決方案。本實作係收集公司中區 PSTN 網路上 110 與 119 呼叫之 SS7 訊務資料，並將 110/119 之信號資料彙集分析功能建置於 TL 雲端平台中，利用開源雲端技術，節省自行購置設備之成本。經由分散大量資料與分析運算任務，除提供使用者快速查詢功能，更可收容錯之功效。

### 關鍵詞

PSTN, SS7, Hadoop, HBase

林威廷 Wei-Ting Lin  
黃奕欽 I-Ching Huang  
沈允中 Yun-Chung Shen  
謝靖慈 Ching-Tsu Hsieh  
朱煜煌 Yu-Huang Chu  
徐浩然 Hao-Jan Hsu

### 摘 要

由於現有傳統網路設備屬於封閉型設備，沒有開放介面供使用者設定，客戶的服務需求無法即時被滿足，這樣的網路架構不僅限制了 IP 網路的技術發展，更無法滿足當前雲端運算和大數據應用趨勢。軟體定義網路(Software-Defined Networking)作為一種最新的網路架構概念，SDN 將網路設備抽象化，首先將網路設備的控制層及轉發層分離，控制層上提供應用層，如此一來，使得網路設備軟體擁有可程式化之能力，藉由應用層開發即時應用程式，讓 SDN 網路可以配合服務需求，動態修改設定。本文主要介紹 SDN 開源軟體及如何利用這些開源軟體來實現雲端資料中心內及資料中心間虛擬網路架構，並實現及驗證。

### 關鍵詞

SDN (Software Defined Network)

OpenFlow

OpenStack

## 如何利用開源軟體建立 IP PBX

鄭旭宸 Shiu-Chen Cheng

曾宇仲 Ju-Dzung Dzeng

施銘原 Ming-Yuan Shih

陳伯濤 Bor-Tauo Chen

陳尚逸 Sam-Yei Chen

許明風 Ming-Feng Hsu

### 摘 要

Asterisk 是一個非常有名的開源軟體，在許多的技術論壇都可見到相關的討論。由於其本身的設計相當的靈活，可與多種不同的平台進行介接，因此研究它或在它上面開發的附掛元件亦為數眾多。本文將對 Asterisk 作一個較為全面性的介紹，包含基本的安裝、設定，其本身內建的一些特殊服務，以及一些較為有名的 Asterisk 周邊附掛應用進行介紹，最後，本文會說明一下我們對此技術目前的應用方式及架構，並以一實例介紹 Asterisk 多元化的應用。

### 關鍵詞

Asterisk 一個廣泛使用的 IP PBX 開源軟體

AGI (Asterisk Gateway Interface, Asterisk 閘道器介面)

AMI (Asterisk Manager Interface, Asterisk 管理介面)

FOP2 (Flash Operator Panel 2)

GPL (Gnu General Public License, GNU 通用公眾授權條款)

IP PBX (Internet protocol Private Branch Exchange, IP 私有交換機)

OSS (Open Source Software, 開源軟體)

## 開源模組化居家照護閘道器

朱國馨 Kuo-Hsing Chu

陳柏文 Po-Wen Chen

李彥良 Yen-Liang Lee

### 摘要

『開源模組化居家照護閘道器』，目的是為居民提供更有效率之健康守護生活方式，在此裝置中整合了備有無線傳輸功能的生理訊號感測裝置，如血壓機、血糖機等，透過這些資訊對慢性病患進行長期的觀察，使家屬能即時掌握病人的生理狀況，並且做進一步的關心與提醒。

本系統以 **Linux 2.6** 為基礎，將應用程式原碼透過交叉編譯的方式移植到開源模組化閘道器中，使它成為居家生理資訊中控系統。在生理訊號傳輸方面，透過藍芽 (**Bluetooth**) 的傳輸方式來簡化生理訊號量測數值的傳輸流程，此外為了確保每一筆生理資訊不會因為網路故障而無法成功記錄，特別加入了小型嵌入式資料庫，使量測資訊暫存於系統中，直到網路修復後再傳送至遠端資料庫中儲存，家屬或醫護人員即可藉由這些量測資訊對使用者的生理狀況有更進一步的了解，並提供更妥善完整的照護。

### 關鍵字

開源模組化閘道器

嵌入式資料庫

PSTN

OPENWRT

## OpenCV 之使用授權與應用

林多常 Duo-Chang Lin

簡大為 Da-Wei Jain

### 摘 要

OpenCV (開源計算機視覺程式庫)採用 BSD 使用授權，這讓 OpenCV 使用者可以幾無拘束的將其應用在學術研究，或是打包置於商業專案裡運用，它以 C/C++ 語言撰寫並做了很大程度的優化，可在 Windows、Linux、Android，以及 Mac Os 等作業系統上運作，並有發展 Matlab、Python、Ruby 以及 Java 等使用介面。另被設計成可很有效率快速的運算並適合應用在即時系統運用上，亦支援多重核心處理器的應用。目前使用者遍布全球，函式庫內含 2500 條以上的演算法，自上線以來已被下載超過 600 萬次，共有 4 萬 7 千人參與其網路社群。另外 OpenCV 可被應用的領域亦非常廣泛，例如常被用來實做衛星或網路地圖的拼接、影像掃描校正、醫學影像雜訊消除、物件搜尋辨認、保全入侵偵測、製造檢測系統、相機校正、軍事應用、無人飛行器、無人車具、無人船舶潛艇、機器人視覺視統、人臉偵測、車牌辨識、文字辨識、追蹤系統等等之應用。現今在影像處理的領域裡，所有基礎的演算法幾乎都可以在 OpenCV 內找到。

### 關鍵字

OpenCV

開源計算機視覺程式庫

CMake

BSD License

MIT License

zlib License

FLTK

人臉辨識

劉秋宗 Chiou-Tzung Liou  
孫漢傑 Han-Chieh Sun  
賴弘文 Hung-Wen Lai  
陳秋玉 Chiu-Yu Chen  
謝琬汝 Wan-Ju Hsieh

### 摘 要

中華電信近年為成為一可信賴的資通訊公司，已積極投入客戶資訊安全服務的建置。近年來由於行動及網頁應用服務開發的蓬勃發展，被視為白箱檢測模式的原始碼掃描服務，由於驗測涵蓋性廣，因此總是被指定成為軟體應用服務驗收項目的必要條件。本文分析目前市面上常用的開源及商用原始碼掃描服務模式及使用者目前的需求，提出一個導入第三方驗證服務同時整合加密上傳工具及雙因子登入認證方法的可信賴雲端原始碼掃描服務模式，提供給使用者一個更完整可靠的原始碼掃描服務環境。

### 關鍵字

白箱檢測 (White-Box Testing)  
程式碼審查 (Code Review)  
原始碼檢測 (Source Code Analysis)

## 開放原始碼軟體再利用之影響因素探討與實務分析

莊志如 Chih-Ju Chuang

董元昕 Yuan-Hsin Tung

### 摘 要

在強調軟體使用自由性的趨勢下，開放原始碼軟體成為資訊軟體業的熱門話題，也成為在資訊應用上逐漸興起的一股新勢力。相較於一般的專屬軟體，開放原始碼是以社群開發力量為後盾，並透過不同的授權條款，將社群的集體智慧發揮至極致。本文主要探討目前企業應用原始碼軟體時遭遇的困難，影響原始碼軟體再利用之可能因素，推動開放原始碼軟體再利用的管理機制，並且透過問卷調查及實際案例的分析，探討在企業原始碼軟體再利用過程中的影響因素，並提出解決方案。

### 關鍵字

開放原始碼軟體  
軟體再利用

許長裕 Chang-Yu Hsu

李國豪 Kuo-Hao Lee

### 摘要

行動網路設備可能因效能改善、穩定性、除錯、參數調整等因素，而進行相關升版改接。如果電信公司無事先告知客戶升版改接作業，不僅可能造成公司營收上的損失，更因為客戶無法獲得服務而造成商譽上更大的無形損失。本篇論文將探討當於行動通信網路規劃升版改接作業時，如何有效率且精準地通報維運同仁與專案經理。最後，由專案經理於升版改接作業施工前，告知相關企業客戶。

### 關鍵字

泛歐數位式行動電話系統 (Global System for Mobile Communications, 簡稱 GSM)

通用封包無線服務技術 (General Packet Radio Service, GPRS)

通用移動通訊系統 (Universal Mobile Telecommunications System, 簡稱 UMTS)

新一代行動網路維運管理系統 (New Generation Mobile Operations Support Systems, 簡稱 NMOSS)

企客關係管理系統 (Business Customer Relationship Management, 簡稱 BCRM)

網路改接及障礙通報系統 (Network Rearrangement and Fault Tasks Notification Management, 簡稱 eRAM)

障礙申告資訊系統 (Trouble Report Information System, 簡稱 TRIS)

新世代網路整合監視系統 (簡稱 ARGUS)

新一代全業務客服中心應用系統 (Call center Application System 2.0, 簡稱 CAS2.0)

高速封包存取 (High Speed Packet Access, 簡稱 HSPA)

組態管理 (Configuration Management, 簡稱 CM)

障礙管理 (Fault Management, 簡稱 FM)

服務品質管理 (Service Quality Management, SQM)

客戶服務品質管理 (Customer Quality Management, CQM)