

大數據在銀行業運用安全性問題

吳昌紘
神通資訊科技公司
pdespd@gmail.com

摘要

金融行業是資料的產生、儲存、處理和傳送組合而成，在數位通路與虛擬銀行架構下，要如何運用大數據的分析和運用，將原先沒有相關聯的資料，串連成和客戶有價值且對銀行營運有幫助的資訊，變成虛實整合行銷的有效利器，是現今金融行業最關心的議題。從 2012 年開始，大數據已成為非常熱門的新觀念，其中最重要的不是資料量的多寡，而是如何運用，從許多的數據中找到趨勢及商機，甚至可以預測客戶的購物行為，進而貼近客戶需求，達到準確行銷的目的。然而在大數據為金融業帶來機遇的同時，也隱含著資料安全的問題。本研究目的在新科技與大數據帶給金融產業的變革同時所產生的資訊安全問題，提出可行的建議整體性方案，期望可以做為在金融產業變革下保護客戶大數據資訊安全方向的參考。

關鍵字：大數據

Abstract

The financial industry is to produce data, storage, processing and transmission combination in the digital path and virtual banking architecture, how to use the analysis and the use of big data, and the original is no associated data series into and valuable customer and on the banking helpful information into an effective marketing tool for the integration of the actual situation, the subject of today's financial industry is most concerned about. From 2012, big data has become a very popular new ideas, the most important thing is not the amount of the amount of data, but how to use, find trends and business opportunities from a number of data, and even predict the customer's shopping behavior, and then close customer demand, to achieve accurate marketing purposes. However, in the big data opportunities for the financial sector, but also implies a data security problem. This study was aimed at bringing new technologies and big data, while changes in the financial industry information security problems arising, feasible recommendations holistic program, you can expect as a reference to protect customers in the financial industry changes the direction of big data information security.

Keyword：big data

1. 簡介

因應雲端運算[1]、物聯網[2]與社群網路[3]等技術的興起，大數據[4]分析技術由此孕育而生，並建立起了迅速發展的生態體系。而且全球資料數據量呈現指數式爆炸增長，根據 IDC 發佈的研究報告，2011 年數據量的產生與複製將到達 1.8ZB，並且以每兩年增加一倍的速度成長，預計到 2020 年，全球產生的資料總量將超過 40ZB。但是，在大數據產生、收集、儲存和分析的過程中，會面臨資料加密和隱私等問題，其中有一些是傳統安全的問題，有一些則是一些新的安全問題，像是在許多的系統中，因為資料與數據十分複雜而帶來一些資料取得的安全問題，或是因為數據種類和業務類型眾多而衍生的資料整合和儲存安全疑慮，另外也因為大數據已經逐漸成為當今時代最關鍵的生產要素與產品型態，也是當前社會從工業經濟往知識經濟改變的重要特徵。大數據帶來的不只是資料量的變化，更是商業思維的轉變。尤以金融產業影響最深，在新一代的消費者的認知中，銀行已經慢慢由一個場所的概念，轉變成為無處不在的服務。時至今日，不論是辦公室、家裡或是其他地方，只要有網路或通訊可到達，都可以成為銀行服務範圍，這已是很明確的了。雖然網際網路（Internet）在 1996 年剛開始大量運用時，便有許多人認為借助這些新興科技，分行將逐漸失去功能甚至消失，但隨著這十多年來，雖然科技日新月異，人口結構與使用習慣也慢慢改變，分行和虛擬通路各自發展出新的定位與角色。且雖然在網際網路、行動應用程式（APP）[5] [6]、社群網站等在剛問世的時候或許還有些新鮮感，但對於在 2000 年後所出生的年經人而言，卻是再平常不過的事情，而只是生活的一部分罷了。所以，若是此刻正為銀行規劃策略的時候，就應該考慮上述問題，從今之後，這只是金融運作的常態。由此觀點來看，金融銀行如果還在延遲投資資訊科技方案，而不趕快打好未來發展的基礎，那麼將被時代洪流所淹沒。簡單而言，只要金融銀行推動創新和改善客戶使用的速度比不上他們採用新型科技的速度，那麼便會陷入很大的劣勢，甚至可能因為第三方業者積極創新而流失客源。而一直使用金融產業受限於政府法規或是系統開發與作業流程老舊無法配合做為藉口是行不通的。資訊時代的客戶享有更多選擇與更多的連線方式和更好、更快速的便捷服務，抗拒科技的投資不過是會讓企業落入被淘汰而已。Brett King 曾提到所謂擴散率(diffusion)是指新的想法從一個消費者傳到另一個消費者的速度[7]。自 1980 年代以來，科技無論普及率或是擴散率都變得越來越快，電話花了快 50 年才跨越關鍵多數(critical mass)的門檻[8]，而電話只花一半的時間，手機和 PC 又只花電話一半的時間，時至今日，新產品從上是到普及只需要幾個月，而不是像以前需要花費很多年的時間，簡單來說，只要你推動創新和改善客戶體驗時間的速度跟不上他採用新科技的速度，那麼你的企業便會陷入極大的劣勢，甚至會因為中介機構和第三方業者積極創新而流失客源，若還是用金融業是被管理得很嚴格或是系統與行政流程老舊無法配合做為藉口是行不通的。若是因為擔心投資報

酬率(ROI)而在進行觀望[9]，那便顯得非常危險了，現在的新科技通常只需要幾個月就能形成主流並改變消費者的行為，若是你所在的銀行一定需要等到眼見為憑，進而形成主流的時候，或是等別家公司證明某項新興服務有好的投資報酬率才願意跟進，再花費 1~2 年開發，那麼等到你開發出的產品上線後，通常已經比競爭者晚了很多年。若是公司採用跟隨者的策略，那麼結果只有被市場淘汰。數位金融 Bank 3.0 的重點在於改變，新興的科技力量與創新的商業模式正不斷的顛覆產業的規則。面臨變化萬千的潮流，銀行可以拒絕改變，因為改變讓人不安，也會打破原有的價值體系；但是，也可以選擇提早接受必然的改變，善用科技，勇於創新，積極的進入客戶的日常生活當中，適時的為其提供銀行的新功能與優良服務，在不斷改變的環境當中找出受惠和獲利方法，這樣才能確保銀行在新世代中存活並致勝。另外由於網路攻擊的手法日益變化，透過分析的過程中可不斷演進，而使得資安防護更趨完整。

本研究目的在巨量環境下數位金融 Bank 3.0 帶給產業的變革同時，所產生的資訊安全問題進行探討，並提出可行的建議方案，期望可以做為產業在 Bank 3.0 的模式下處理與改善安全管理相關問題時的參考。

2. 文獻探討

(a) 大數據（Big data）顧名思義，當資料量龐大到資料庫系統無法在合理時間內進行儲存、運算、處理，分析成能解讀的資訊時，就稱為大數據。資料來源的多樣性和資料數據量的爆炸性成長促使大數據的出現且受到廣泛的應用。大數據的分析與利用也不再侷限在資訊技術產業，也衍生到民生、經濟、政府、產業界等領域，像是物流倉儲、氣象及生態分析、健康監護醫療等。而且大數據幾乎無法使用大多數的資料庫管理系統處理，而必須使用「在數十、數百甚至數千台伺服器上同時平行運行的軟體」。一般來說，大數據涵蓋的範圍很廣，定義也各家歧異。最早由 Gartner 公司的分析師 Doug Laney 在 2001 年發表的「3D Data Management: Controlling Data Volume, Velocity, and Variety.」一文中挑明了資料處理的三個關鍵挑戰－資料量、速度、多樣性[10]，並在 2012 年 Doug Laney 給予大數據一個全新的定義：「大數據是大量、高速或類型多變的資訊資產，它需要全新的處理方式，去促成更強的決策能力、洞察力與最佳化處理。」

大數據時代的來臨帶來無數的機遇，但是與此同時個人或機構的隱私權也極有可能受到衝擊，大數據包含各種個人資訊資料，現有的隱私保護法律或政策無力解決這些新出現的問題。有人提出，大數據時代，個人是否擁有「被遺忘權」，被遺忘權即是否有權利要求資料商不保留自己的某些資訊，大數據時代資訊為某些網際網路巨頭所控制，但是資料商收集任何資料未必都獲得使用者的許可，其對資料的控制權不具有合法性。2014 年 5 月 13 日歐盟法院就「被遺忘權」（right to be forgotten）一案作出裁定，判決 Google 應根據使用者請

求刪除不完整的、無關緊要的、不相關的資料以保證資料不出現在搜尋結果中。這說明在大數據時代，加強對使用者個人權利的尊重才是時勢所趨的潮流[11]。

(b) 數位金融 Bank 3.0，銀行業的興起距今已有數百年，一直依循著人們到銀行存款、借款、外幣業務與購買金融商品等模式，但在近年，隨著科技與網路發展所帶動的電子金融興起，其金融服務突破了時間和空間的限制，對於銀行的經營模式產生很大的影響。客戶的行為已經改變，當全世界的人都將手機當成銀行帳戶使用時候時，未來的銀行將不只是一個地方，而是一種行為，客戶需要的不是銀行本身，而只是需要其功能而已。藉由科技的進步，消費者的行為已經由電腦連結網際網路到目前的行動式裝置，像是行動電話或平板電腦等，加上支付行為行動化，演進為無現金的行動錢包等[12]，銀行和客戶的互動方式發展至無時無刻均可使用的境界，此時的銀行產業將徹底被改變。由於科技演進的快速，雲端運算、大數據、行動支付等近年來已陸續被運用在產業與人們的日常生活當中，而在未來銀行的金融服務需要透過一種新的與客戶在基於以上各種的新技術與觀念中進行互動，方可在此數位時代中提供打動客戶心理的服務。

而金融業的內部資料來源包括壽險、銀行、產險、證券與投信等單位的客戶資料，這些資料的收集包括營運交易、客服系統、CRM、電話行銷[13]、郵件行銷與行動服務等多元管道的資訊整合，在取得客戶同意下，針對其整體需求進行分析與服務。

數位金融應以客戶為中心，以客戶的角度來進行提升服務效率的規畫交易流程與作業，而非以金融業本身的便利為主要考量。例如：減少客戶不必要的填寫表單作業、自動的取得客戶基本資料並同步相關系統等，以提升客戶滿意度。另外金融產業必須認知系統數位自動化雖然可有效降低服務成本，但本質上並非為了降低成本而進行數位化，不然將造成錯誤的決策與公司經營方向的偏差，甚至流失既有的客戶族群。

(c) 雲端運算 (cloud computing) 根據維基百科定義，雲端運算是一種基於網際網路的運算方式，通過這種方式，共享的軟硬體資源和資訊可以按需求提供給電腦和其他裝置。使用者不再需要了解「雲端」中基礎設施的細節，不必具有相應的專業知識，也無需直接進行控制。雲端運算描述了一種基於網際網路的新的 IT 服務增加、使用和交付模式，通常涉及通過網際網路來提供動態易擴充功能而且經常是虛擬化的資源。在雲端運算中可以實現 IT 資源的自動化管理和分配，降低 IT 管理的複雜性，提高資源使用效率，大數據的技術主要用於解決大規模的資料運算與處理探勘等問題，而雲端運算為一種資料儲存、運算的能力等。大數據結合雲端運算為未來資訊科技發展的必然趨勢。

(d) 大數據的安全威脅：除了和傳統環境相同都會遇到的非授權訪問、資訊洩漏或竊取、傳輸過程中破壞資料完整性、阻斷式攻擊與網路病毒外，比較大的區別是在於資料庫，因為資料量的因素，所以大數據的儲存通

常是採用非關聯式 (NOSQL) 資料庫進行儲存[14] [15] [16]，雖然其優點在於資料的可擴展性和可用性，但是其結構化查詢能力也較為薄弱，且因為資料是分散放置在不同地理位置與不同伺服器中，用來實現較佳的資料查詢與容錯備份，但相對而言也比較難定位進而保護。另一方面網路安全也是大數據的一個重要方向，現有的安全防護機制套用在於大數據架構上，並無法達到完善，而且如何由成千上萬的安全事件與紀錄中找到有價值和最需要很快速處理的問題也是過程中重要的方向。

(e) 大數據的安全防禦：依據大數據應用的生命週期，在資料取得的階段，必須保障資料在傳輸時候的機密性保護，在資料儲存的階段，隱私保護與如何提高備份的效率是關注的焦點，而在資料探勘階段，主要需注意在非關聯式資料庫架構中的計算和處理問題，最後需要考慮的是近年來常見的 DDoS 與 APT 網路攻擊防護[17] [18] [19] [20]。

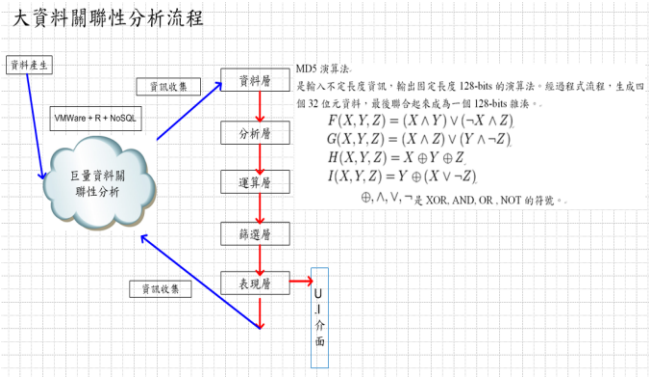
(e) 大數據與安全檢測技術：大數據的技術發展為各行各業發展奠定了堅實的基礎，但同時也使得資訊安全變得更加複雜。在大數據時代，可以將資料檢測和大數據兩者互相結合，用以發現潛在的威脅，保護客戶的資產與資訊。其主要流程約可分為三個步驟，大數據收集、大數據探勘與資訊安全檢測。

(f) 社群媒體 (Social Media)：社群媒體的誕生可以追溯到 1978 年，電子佈告欄 (BBS) 的發明，這是第一個可以和朋友分享資訊和活動的電子平台，而臉書 (Facebook) 的產生在 2003 年，一直到現在成為最大的社群媒體，網友們可透過網頁分享自己的近況、相片和影音等，也可接收到朋友即時的回應。而企業也可以透過社群媒體在與客戶的互動策略中，取得行銷、銷售、服務、客服、與了解客戶行為想法等方面，均是兼具效率及成本效益的工具，由 Bain & Company 的研究報告發現，若客戶使用社群媒體和企業互動，其消費金額會比其他客戶高 20% ~ 40%。但必須避免只有行銷業務人員才可以使用臉書的錯誤，企業面對社群媒體會有兩種直覺的反應，第一種是直接將傳統廣告和公關的方式套用在社群媒體上，第二種是限制或關閉內部員工連線社群媒體的管道，以避免員工對外發言可能產生的風險，或是花費過多時間在非工作上的錯誤想法。這些做法將會導致企業無法真正去了解社群媒體的真正力量和影響，但隨著員工的工作環境由傳統桌上型電腦轉換到行動裝置的同時，企業要透過關閉對外部社群媒體的聯繫來避免風險，事實上是無法達到效果的，反而會使得員工改用非公司管理的環境來連結社群媒體。

(g) MD5 (MD5 Message-Digest Algorithm)：一種被廣泛使用的密碼雜湊函式，可以產生出一個 128 位元的雜湊值 (hash value)，用於確保資訊傳輸完整一致。MD5 由 Ronald Rivest 設計，於 1992 年公開，用以取代 MD4 演算法。這套演算法的程式在 RFC 1321 中被加以規範[21] [22] [23] [24] [25] [26]。

3. 實驗設計與系統架構

有鑑於硬體處理速度越來越快的情況下，在本文中將提出一個當資料在本地端收集完成後，為了備援的原因所以在本地端先進行加密且切割成數個檔案後再進行傳送到世界各地的公有雲 (Public Cloud) 當中，之後如有需要使用到此資料的時候，再從 Public Cloud 中將資料各個片段取回進行組合並進行解密。其中在此使用常見的 MD5 演算法進行資料加解密，而加密後檔案切割和組合的方式採用處理速度較快的 Merge Sort 進行實作，如圖一，透過切割後將各個小片段資料隨時放置到不同地區的 Public Cloud 中，而欲進行組合的時候採用相同的方式進行，將資料片段由各地收集回來後進行重組與解密運算。透過上述方式加密且進行分割資料片段來加強安全性，其中 Public Cloud 環境使用 VMWare 進行模擬建構，加快處理速度，在本文中 MD5 加密演算法與 Merge Sort 使用 C 語言撰寫與實現，模擬架構如圖二。



圖一：分析流程

```
Top-down implementation [edit]
Example C-like code using indices for top-down merge sort algorithm that recursively splits the list (called runs in this example) into sublists until sub-list size is 1. Then merges those sublists to produce a sorted list. The copy back step could be avoided if the recursion alternated between two functions so that the direction of the merge corresponds with the level of recursion.

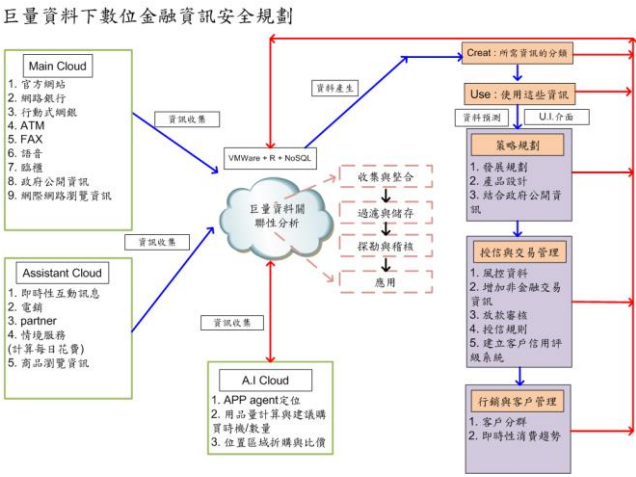
/* array A[] has the items to sort; array R[] is a work array */
TopDownMergeSort(A[], R[], n)
{
    TopDownSplitMerge(A, 0, n, R);
}

// begin is inclusive; end is exclusive (A[end] is not in the set)
TopDownSplitMerge(A[], begin, end, R[])
{
    if (end - begin < 2)
        return; // consider it sorted
    // then merge them and return back up the call chain
    int mid = (begin + end) / 2;
    TopDownSplitMerge(A, begin, mid, R); // split / merge left half
    TopDownSplitMerge(A, mid, end, R); // split / merge right half
    TopDownMerge(A, begin, mid, end, R); // merge the two half runs
    CopyArray(R, begin, end, A); // copy the merged runs back to A
}

// left half is A[begin:mid-1]
// right half is A[mid:end-1]
TopDownMerge(A[], begin, mid, end, R[])
{
    int i = begin, j = mid, k = end;
    while (i < j && j < k)
    {
        // If left run has elements and is <= existing right run head.
        if (A[i] <= A[j])
            R[k++] = A[i++];
        else
            R[k++] = A[j++];
    }
    // while there are elements in the left or right runs
    for (; i < j || j < k;)
    {
        // If left run has elements and is <= existing right run head.
        if (i < j && A[i] <= A[j])
            R[k++] = A[i++];
        else
            R[k++] = A[j++];
    }
}

CopyArray(R[], begin, end, A[]);
for (i = begin; i < end; i++)
    A[i] = R[i];
}
```

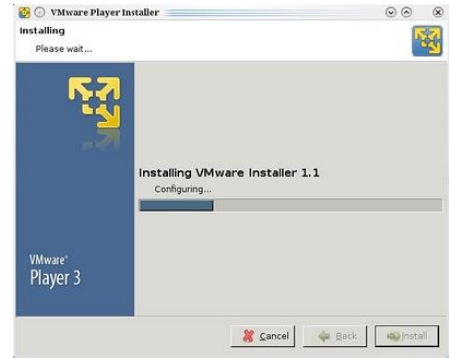
圖二：Merge Sort (wiki)



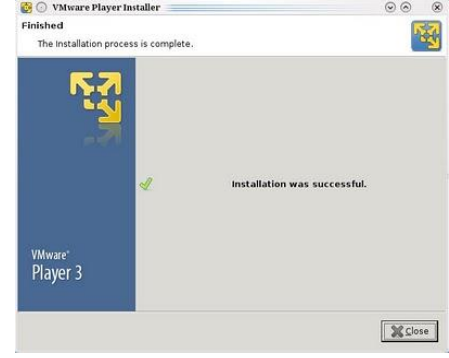
圖三：大數據下數位金融資訊安全規劃



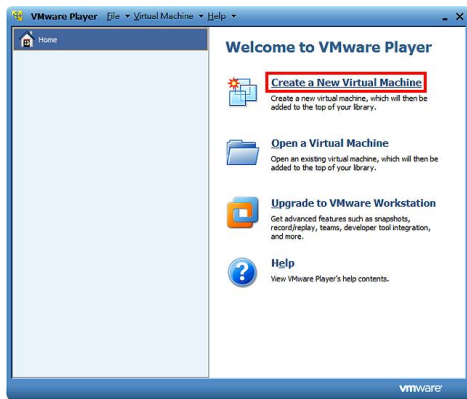
圖四：下載虛擬環境架設軟體



圖五：軟體安裝畫面



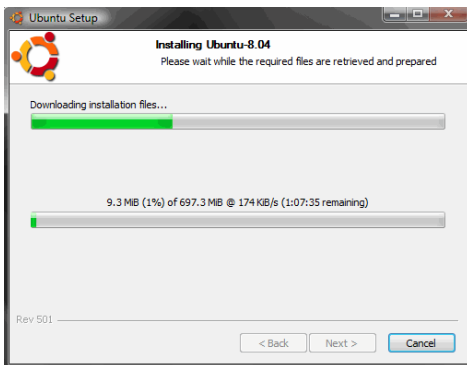
圖六：完成軟體安裝



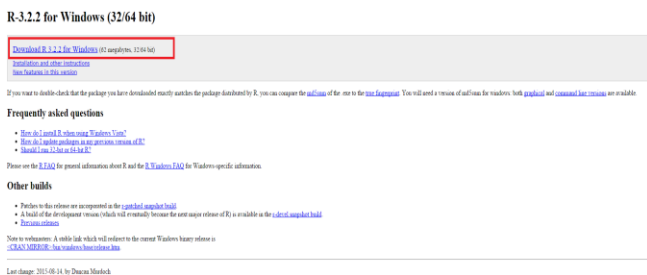
圖七：建立一個新的虛擬機器



圖八：安裝 Windows 2003 Server



圖九：安裝 Ubuntu



圖十：下載 R 語言 package 檔案

Double click on the installer to select the installation language.



圖十一：選擇安裝的語言體系

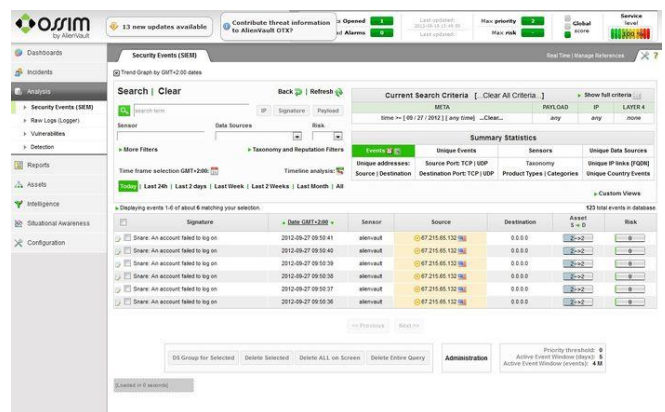
Click on Finish to complete the installation.



圖十二：完成安裝



圖十三：下載 OSSIM

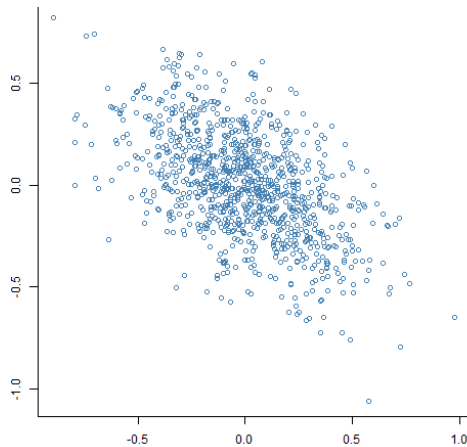


圖十四：OSSIM 顯示畫面

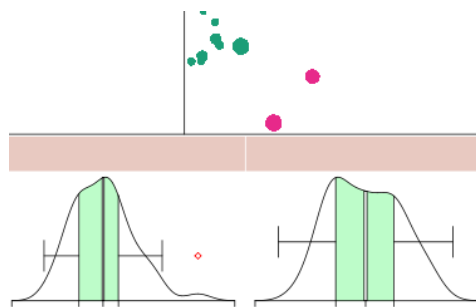


```
[root@server ~]# sudo chkconfig --list | grep mongod
mongod          0:off  1:off  2:off  3:on   4:on   5:on   6:off
[root@server ~]# systemctl stop mongod
tcp             0      0  0.0.0.0:27017  0.0.0.0:*        LISTEN
```

圖十五：完成 NoSQL



圖十六：系統產生資料



圖十七：系統產生資料

4. 結論

一般消費者認為，可以透過新科技使用金融服務是理所當然的，就像是使用 ATM 或到分行臨櫃辦事相同。只要是身為金融業的一員，都是別無選擇的，若是目前仍是未積極地導入這些科技的話，便趕不上消費者的期待和消費模式的變化了，便只能等著被大浪所淹沒。因為各種跡象顯示，在不久的將來，金融就會像是資訊科技公司，而金融業務則會像是一般民生用品一樣的透過科技來提供給消費者。而如何在公司在很安全的即時掌握營運績效的第一手資料，資料的收集方式與安全的防護是基本的門檻，許多的公司通常受限於資訊設備的提供，資料的管理與資安人才的缺乏，經常抱怨因此無法達到，但也無可奈何，其實要做好資料收集，除了要有良好且穩定的硬體外，也必須要有好的資安技術與人才投入，才能運用發揮，企業營運與管理階層如果希望隨時都有最新且有用的報表，透過上述的資料收集及分析可達到，進而在安全的環境下掌握企業未來發展的關鍵。

參考文獻

- [1]. Shaikh, F.B. Dept. of Comput. & Technol., SZABIST, Islamabad, Pakistan, "Security threats in cloud computing" International Conference for Internet Technology and Secured Transactions (ICITST), 2001
- [2]. Wikipedia, Internet of Things, https://en.wikipedia.org/wiki/Internet_of_Things
- [3]. Wikipedia, 社群網路, <https://en.wikipedia.org/wiki/社群網路>
- [4]. Tiwari, A.K. Dept. of Comput. Sci. & Eng., Maharishi Arvind Coll. of Eng. & Res. Center, Jaipur, India, "A review on Big Data and its security" International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), 2015
- [5]. Joorabchi, M.E. Univ. of British Columbia, Vancouver, BC, Canada, "Real Challenges in Mobile App Development" ACM / IEEE International Symposium on Empirical Software Engineering and Measurement, 2013
- [6]. Microsoft, 什麼是行動應用程式?, <https://azure.microsoft.com/zh-tw/documentation/articles/app-service-mobile-value-prop-preview/>
- [7]. Brett King, diffusion, <http://www.amazon.com/Bank-3-0-Banking-Somewhere-Something/dp/1118589637>
- [8]. Telephone, Telephone, <https://en.wikipedia.org/wiki/Telephone>
- [9]. Investopedia, Return On Investment, <http://www.investopedia.com/terms/r/returnoninvestment.asp>
- [10]. Doug Laney, "3D Data Management: Controlling Data Volume, Velocity, and Variety.," Gartner, 2001
- [11]. EU court, "right to be forgotten", <http://www.theguardian.com/technology/2014/may/13/right-to-be-forgotten-eu-court-google-search-results>. 2014.
- [12]. Varshney, U. Georgia State Univ., Atlanta, GA, USA. , "Mobile payments", Electrical Engineering, IEEE Computer Volume:35, Issue:12, Dec. 2002.
- [13]. Overman, D.J. CPenney Catalog Marketing, Brookfield, WI, USA. , "Telemarketing", Electrical Engineering, Communications Magazine, IEEE Volume:28, Issue:6, Dec. 2002.
- [14]. Wikipedia, NoSQL, <https://en.wikipedia.org/wiki/NoSQL>
- [15]. Tudorica, B.G. Dept. for Econ. Math. & Econ. Inf., Pet.-Gas Univ. of Ploiesti, Ploiesti, Romania. , "A comparison between several NoSQL databases with comments and notes", Roedunet International Conference RoEduNet, 2011.
- [16]. Jing Han PCN&CAD Center, Beijing Univ. of Posts & Telecommun., Beijing, China. , "Survey on NoSQL database", Pervasive Computing and Applications (ICPCA), 2011.
- [17]. Kupwade Patil, H. Nanthealth, Dallas, TX, USA. , "Big Data Security and Privacy Issues in Healthcare", IEEE International Congress on Big Data (BigData Congress), 2014.
- [18]. Venkata Narasimha Inukollu, "SECURITY ISSUES ASSOCIATED WITH BIG DATA IN CLOUD COMPUTING", International Journal of Network Security & Its Applications (IJNSA), Vol.6, No.3, May 2014.
- [19]. N, Gonzalez, Miers C, Redigolo F, Carvalho T, Simplicio M, de Sousa G.T, and Pourzandi M., "A Quantitative Analysis of Current Security Concerns and Solutions for Cloud Computing.", Athens:2011., pp 231 – 238, Nov. 29 2011- Dec. 1 2011.
- [20]. F.C.P, Muhtaroglu, Demir S, Obali M, and Girgin C., "Business model canvas perspective on big data applications", Big Data, 2013 IEEE International Conference, Silicon Valley, CA, Oct 6-9, 2013, pp.32 - 37.
- [21]. Ahmad, I. VMware Inc., Palo Alto, CA, USA., "An analysis of disk performance in VMware ESX server virtual machines", IEEE International Workshop on Workload Characterization, 2003. WWC-6. 2003.
- [22]. Wikipedia, R programming language, [https://en.wikipedia.org/wiki/R_\(programming_language\)](https://en.wikipedia.org/wiki/R_(programming_language))
- [23]. Sujata Kumari, Rahul Abhishek, B.S. Panda., "Intelligent Computing Relating to Cloud Computing", International Journal of Mechanical Engineering and Computer Applications (IJMCA), Vol 1, Issue 1, February 2013.
- [24]. Yang Zheng Sch. of Inf. Sci. & Eng., East China Univ. of Sci. & Technol., Shanghai, China., "VM scheduling strategies based on artificial intelligence in Cloud Testing", IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2014.
- [25]. Wikipedia, MD5, <https://en.wikipedia.org/wiki/MD5>
- [26]. Wikipedia, merge sort, https://en.wikipedia.org/wiki/Merge_sort