俱安全性之雲端服務行動祕書平台

江茂綸
朝陽科技大學
資訊與通訊系
mlchiang@cyut.edu.tw

林仕軒
朝陽科技大學
資訊與通訊系
s10230604@cyut.edu.tw

曾莉雅
華夏技術學院
數位媒體設計系
lytseng@cc.hwh.edu.tw

曾駿宏
朝陽科技大學
資訊與通訊系
s9930072@cyut.edu.tw

黃俊傑
朝陽科技大學
資訊與通訊系
s9930010@cyut.edu.tw

楊書泓
朝陽科技大學
資訊與通訊系
s9930090@cyut.edu.tw

劉宏軒
朝陽科技大學
資訊與通訊系
s9930102@cyut.edu.tw

楊博皓
朝陽科技大學
資訊與通訊系
s9930106@cyut.edu.tw

摘要

在現今使用者都會透過手持裝置來取得所需要的服務。這類的雲端運算服務也將成為目前的主要趨勢。然而，眾多的網路服務需求，會使得服務提供伺服器負擔過重。所以，本文利用 Hadoop 來實作雲端運算服務平台，以提高運算效能、安全及容錯的能力。

因此，本文結合再生龍這套還原系統來讓使用者可以隨時備份及回復個人檔案及系統。此外，本系統也透過 Snort 來加強其安全性。因此，本文所提出之雲端運算服務行動祕書平台可提升使用者服務效能、容錯及安全性。

關鍵字：Hadoop、封包監測引擎、Clonezilla、雲端運算、HBase。

Abstract

Nowadays, the user can obtain the requirement of service by hand-hold device. This kind of clouds service becomes a current trend. However, the multitudinous requirements of network service can make service provider overloading. Therefore, this paper proposes a cloud computing service platform by using Hadoop software to enhance computing efficiency, security and fault tolerance of the capacity.

As a result, this paper combines the recovery system (Clonezilla) to provide the user to backup and recover their files and system. Besides, the security is guaranteed by using Snort in proposed system. Therefore, the efficiency, fault tolerance, and security of cloud computing can be achieved in our proposed system.

Keywords : Hadoop、Intrusion detection、Clonezilla、Cloud Computing、HBase

1. 前言

隨著 E 化的發展，應用服務的需求不斷的在增加，但硬體設備的擴充速度遠不及服務需求增加的速度，此外，高階的伺服器與維護成本也是一筆龐大的負擔。因此，雲端運算架構透過將運算資源進行的虛擬化的整合，有效的提供較大的運算服務能力，也將是目前主要的趨勢。

此外，本文透過再生龍[11]這套由國網中心所開發的還原系統，來提供使用者進行個人電腦的備份，並可製成映像檔儲存於個人儲存空間之中，方便使用者可以隨時隨地的進行異地還原，以增加使用者的便利性。

而由於使用者都將資料放置在網路上，則安全部分也將大受考驗，由其在現在的雲端運算架構下，個人資料或運算資源皆分佈在不同地點，則運算節點則需要更進一步的防範駭客的攻擊。因此，本文將採用了封包分析引擎(Snort)[12]來監測平台封包的流量及其安全性。如有發現異常封包即進行隔離與防禦，並立即寄信向管理者反應進行後續處理的動作，另外，

也結合了分散式監測系統(Ganglia) [13]來自動進行訊息的收集，讓管理者能夠快速知道各個運算節點的狀態，並使用 rrdtool 資料庫[14]來儲存收集到的歷史數據，以方便日後進行分析及追蹤。

2. 文獻探討

2.1 Hadoop 軟體

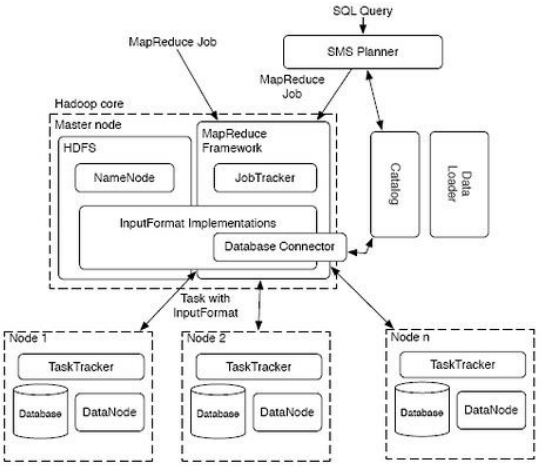
Hadoop[2][15]是一套讓使用者可以撰寫並執行海量資料應用程式的雲端運算軟體平台。不但擁有儲存與處理大量資料的配置能力外，還可透過分散式檔案系統的幫助，讓一般 PC 可以架設叢集環境，進而取得較大的運算能力。此外，每個節點每隔一段時間就會互相交換訊息，所以當某節點發生錯誤時，也能即時且自動取得備份資料。

2.2 HBase 資料庫

HBase[4]是由 BigTable[5]所衍生出來支援 Hadoop 的資料庫。以圖一[10]來說明，其有別於一般資料庫系統所使用的 row-key 儲存方式，主要是以 column-family 的儲存方式來表現。

一個 column family[6]就是一個 column qualifier(Column1、Column2)的集合，裡面可有很多組 qualifier，這些 qualifier 可以視需要隨時新增，而不用重新設定整個資料表。

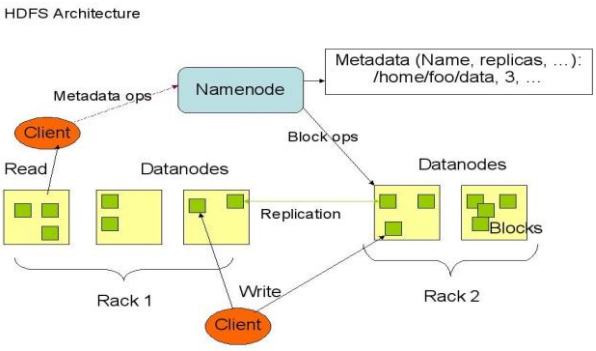
而 HBase 底層則是使用了分散式的檔案系統 HDFS (Hadoop Distributed File System)[16]來進行儲存；其主要是透過將一個資料表拆成很多份後，再由不同的伺服器負責該部份的存取，藉此達到高效能的存取效率。此外，HBase 也同時提供 Map/Reduce[7]程式設計模式、Java 函式庫、php 及 Thrift 等介面。



圖一、Hbase 與 Hadoop 資料儲存架構圖[10]

2.3 Hadoop Distributed File System (HDFS)

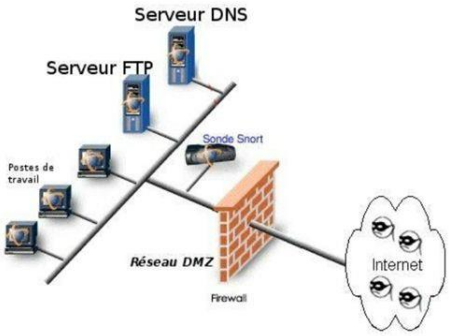
Hadoop Distributed File System (HDFS) 將分散的儲存資源整合成一個具容錯能力、高效率且超大容量的儲存環境。而其架構為 master/slave，並由 Name node 及 data nodes 兩種角色所組成；Name node 負責檔案系統中各個檔案屬性權限等資訊 (metadata, namespace) 的管理及儲存；而 data node 通常由數以百計的運算節點所擔任，並將一個資料檔切割成數個較小的區塊來進行儲存，而每一個區塊還會有數份副本存放在不同節點中，若有其中一個節點損壞時，檔案系統中的資料還能保存無缺，其架構如圖二所示。因此，name node 需要紀錄每一份檔案存放的位置，並負責協調 data node 及進行錯誤的回復。



圖二、HDFS 架構圖[9]

2.4 封包分析引擎(Snort)

Snort[12]是一套開放原始碼的網路入侵預防軟體與網路入侵偵測軟體，是目前全世界最廣泛使用的入侵預防與偵測軟體，其架構如圖三所示。



圖三、Snort 架構圖[12]

而 Snort 主要是經由蒐集網域內的封包並依設置的規則來進行分析，進而檢測出異常封包及異常行為，然後產生警告，並且可以做出相對應的防禦和防範。其工作模式有三種：偵測模式(Sniffer mode)，主要是擷取網域內的封包，並顯示在終端螢幕上。封包記錄模式(Packet logger mode)，可將擷取到的封包記錄並儲存。線程模式(Inline mode)，則是將擷取到的封包依設置的規則（Rules）進行分析，判斷是否有網路異常行為的出現。

2.5 Thrift

Thrift[9]源自於 Facebook 之手，在 2007 年 Facebook 提交 Apache 基金會將 Thrift 作為一個開源專案，對於當時的 Facebook 來說創造 Thrift 是為了解決 Facebook 各系統間大資料量的傳輸通訊以及系統語言環境不同而 需要跨平臺的特性。所以 Thrift 可以支援多種程式語言，例如: C++, C#, Cocoa, Erlang, Haskell, Java, Ocami, Perl, PHP, Python, Ruby, Smalltalk 等。因此，Thrift 可以作為二進位的高性能的通訊中介軟體，來支援資料(物件)序列化和多種類型的 RPC 服務。

此外，Thrift 適用於程式對程式之間的靜態資料交換，但需要事先確定好溝通之間的資料結構，而當資料結構發生變化時，必須重新編輯 IDL 檔，進行代碼生成，再編譯載入的流程，跟其他 IDL 工具相比較可以視為是 Thrift 的弱項，然而，Thrift 適用於搭建大型資料交換及存儲的通用工具，對於大型系統中的內部資料傳輸相對於 JSON 和 xml 無論在性能、傳輸大小上有明顯的優勢。

2.6 Ganglia

Ganglia[13]監控軟體主要是用來監控系統性能的軟體，如：Cpu、Memory、硬碟利用率，I/O 負載、網路流量情況等，通過統計圖表便可見到每個運算節點的工作狀態，對於系統的調整、資源分配等有很大的助益。而由於 Ganglia 是分散式的監控系統，有兩個 Daemon, 分別是：用戶端 Ganglia Monitoring Daemon(gmond) 和服務端 Ganglia Meta Daemon(gmetad)，還有 Ganglia PHP Web Frontend 組成。

此外，Ganglia 可用來測量數以千計的運算節點，透過每台運算節點來發送度量資料（如處理器速度、記憶體使用量等）來傳入主機的層次結構中，以便進行觀察及監控。

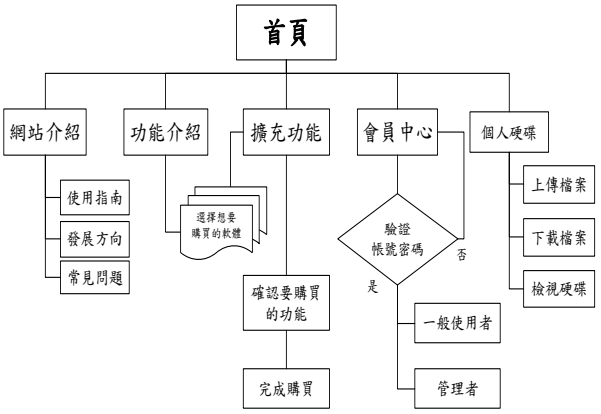
2.7 Clonezilla

Clonezila Live[11]再生龍是一款由國網中心研發的全中文介面的硬碟備份軟體，包含還原程式以及作業系統，不但可以製作、還原硬碟，還可及還原再生多種作業系統，包含 Linux (ext2, ext3, ext4, reiserfs, reiser4, xfs, jfs)、Mac OS (HFS+)、微軟 Windows (fat, ntfs)等。而這些檔案系統可以只備份有資料的區域來節省備份時間與硬碟空間。而其他不支援的檔案系統，也可採用全部複製的方式來進行處理。因此，本文將結合遠端備份功能至所提之雲端運算架構來提供使用者更可靠的雲端運算服務。

3. 系統架構

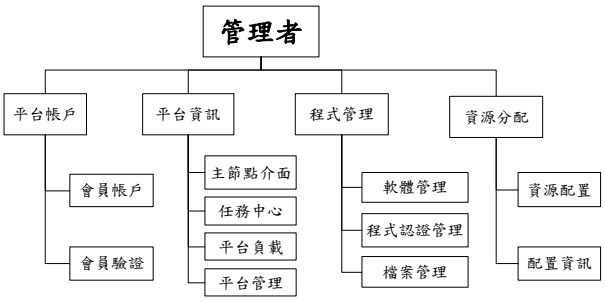
本文所提之雲端服務行動祕書平台主要

可分為二層架構；網頁使用端及網頁管理者端。此外，為因應現有雲端運算平台的安全問題及錯誤回復機制，也將結合入侵檢測系統 Snort 和備援還原系統 Clonezilla，來達到平台的安全性及穩定性，其架構如圖四和圖五所示。



圖四、 網頁使用者端架構圖

根據圖四，網頁使用端提供使用者在註冊之後進行登入，並可享有個人基本的儲存空間來進行上傳、下載及檔案的檢視。此外，使用者可以至擴充頁面選購所需軟體，便可直接在平台上進行操作，並可透過備份功能輕鬆的將整台電腦備份至其它的儲存節點，並可隨時進行檔案的修改及還原。



圖五、網頁管理者端架構圖

而管理者可以於平台上進行平台帳戶管理、平台資訊管理、程式管理及資源分配等功能。其功能說明如下所示。

- 平台帳戶管理功能：可供進行會員驗證、會員資料管理等操作。
- 平台資訊管理功能：可於平台上公開資訊之管理，並可透過 Snort 來偵測異常封包，再將有異常的封包進行分析。若確定為攻擊時， Snort 會優先封鎖此 IP，再寄信告知管理者。此外，本平台還可結合 Ganglia 套件，清楚掌握平台上的流量、CPU 或硬碟等等的狀態。
- 程式管理功能：管理者可以在此頁面進行供應商上傳軟體程式的掛載與卸除，經過管理者認證後，即可進行販售及使用。
- 資源分配功能：管理者可透過本功能來進行平台上各項資源的分配，並可直接在網頁上開啟/關閉平台所使用的軟體。

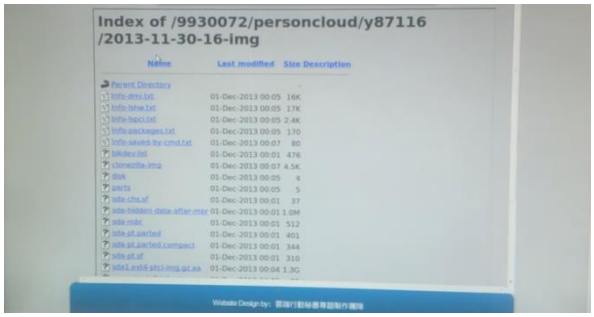
4. 實作雲端服務行動祕書平台結果

4.1 網頁使用者端

使用者於登入後，可於平台上直接使用程式和個人存放空間，讓使用者減輕儲存空間的負擔及方便進行使用，其結果如圖六所示。

圖六、使用端平台圖

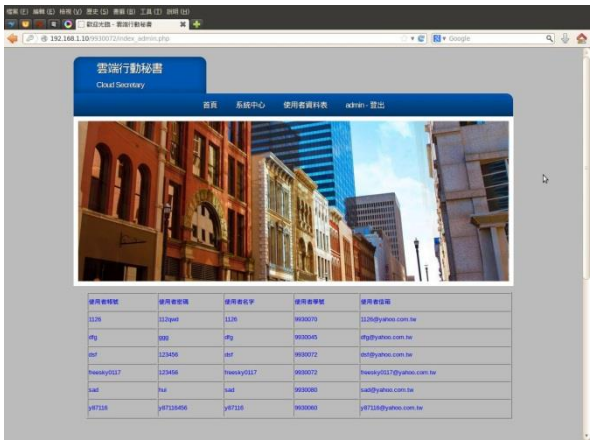
而為了提高容錯性，則結合 Clonezilla 讓使用者可以快速備份檔案並隨時隨地的於異地還原，其實作結果如圖七所示。



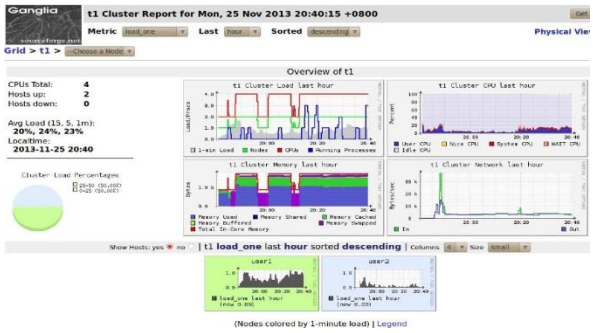
圖七、再生龍備份成功圖

4.2 網頁管理者端

透過本系統，管理者可以於平台上直接進行管理，包括其平台流量及硬體設備狀況等資訊，其管理介面如圖八及圖九所示。



圖八、管理者端平台圖



圖九、Ganglia 管理頁面圖

最後本文也針對安全性來建置 Snort，對於意圖危害此平台的 IP 進行偵測及防護，並結合信件通知來方便管理者去管理及強化平台安全，其實作結果如圖十和圖十一所示。

圖十、Snort 管理頁面圖

圖十一、Snort 警告信頁面圖

5. 結論

本文提供軟體開發商一個可以提供服務的雲端運算行動祕書平台，讓更多使用者可以透過簡單的登入及購買，即可使用其服務。而使用者也可以透過映像檔的編輯來還原受損的檔案及系統，將可有效提升系統之可靠度。然而，本文也透過安全性監控 Snort 來偵測入侵危害，以加強整體雲端運算架構系統資料的安全性。

6. 誌謝

這篇論文是國科會計畫 (NSC 102-2221-E-324-014) 研究成果的一部份，我們在此感謝國科會經費支持這個計畫的研究。

7. 參考文獻

[1] Wikipedia,“雲端運算,” ***http://zh.wikipedia.org/wiki/%E9%9B%B2%E7%AB%AF%E9%81%8B%E7%AE%97***,Apr 3,2009.

[2] jazz,“雲端運算相關技術與應用之初探,” ***http://trac.nchc.org.tw/cloud/wiki/TMUE1031*** ,Mar 17, 2011.

[3] 黃植懋,“伺服器虛擬化技術簡介,” ***http://www.cc.ntu.edu.tw/chinese/epaper/0004/20080320_4012.htm*** ,Mar 20,2008.

[4] Rafan,“HBase 介紹,” ***http://www.hadoop.tw/2008/11/hbase.html*** , Nov 16 , 2008.

[5] Wikipedia,“Bigtable,” ***http://zh.wikipedia.org/wiki/BigTable*** ,Aug 18, 2011.

[6] 周秉誼,“雲端運算平台-Hadoop” ***http://www.cc.ntu.edu.tw/chinese/epaper/0011/20091220_1106.htm*** ,Dec20, 2009.

[7] Wikipedia,“Map/Reduce,” ***http://zh.wikipedia.org/wiki/MapReduce*** ,Oct 11, 2011.

[8] Wikipedia,“Netbeans,” ***http://zh.wikipedia.org/wiki/NetBeans*** ,May 13, 2011.

[9] 林旻君,“Facebook 到底用了哪些技術?,” ***http://www.syscom.com.tw/ePaper_Content_EPArticledetail.aspx?id=140&EPID=168&j=5&HeaderName=%E7%A0%94%E7%99%BC%E6%96%B0%E8%A6%96%E7%95%8C***,Sep 11, 2011.

[10] Hadoop 之 Hbase ***http://sishuok.com/forum/posts/list/220.html***

[11] 再生龍 Clonezilla Live 再生龍 - 功能類似 GHOST 的免費硬碟備份還原軟體 ***http://moneymaker.cybertranslator.idv.tw/archives/27157***

[12] 入侵偵測系統實作-Snort ***http://green.kyu.edu.tw/gt/file/file2/128c5335ac6241b9a7a556526e8a7736.pdf***

[13] 叢集監控軟體 Ganglia 簡介 ***https://ascc.sinica.edu.tw/iascc/articals.php?_section=2.4&_op=?articalID:5134***

[14] RRD 資料庫及 RRDTool 簡介 ***http://linux.chinaunix.net/salon/200712/files/RRD_RRDTool_xa.pdf***

[15] M. N. Vora, et al., “Hadoop-HBase for Large-Scale Data,” ***International Conference on (ICCSNT)***, Dec. 26, 2011, pp. 601-605.

[16] HDFS Architecture Guide, ***http://hadoop.apache.org/docs/hdfs/current/hdfs_design.html***, Dec. 4, 2011.

[17] G. Deka, et al., “A Survey on Cloud Database,” ***IT Professional***, Vol. pp, Issue 99, Jan. 03, 2013, pp.1-6.