

A Comprehensive Review of OpenStack Compatible SDN Solutions

李慧蘭 陳敏 劉德隆

財團法人國家實驗研究院國家高速網路與計算中心
{ gracelee, minchen, tlliu}@narlabs.org.tw

摘要

隨著雲端運算的崛起，資料中心的基礎架構出現重大的演變。OpenStack 開放式的雲端運作平台被廣泛佈署於資料中心的私有雲或公有雲架構。對於多租戶的網路控管和網路虛擬化的需求，傳統網路已不堪勝任，新一代的軟體定義網路 (SDN) 具高彈性和高擴展性，提供快速變動的網路需求來因應新型態的雲端資料中心。因此，在建構以 OpenStack 為基礎架構來提供服務的資料中心時，應用軟體定義網路來優化網路配置是當前發展趨勢，本論文綜合評述目前以軟體定義網路技術應用於 OpenStack 架構中的解決方案和比較其優劣。

關鍵詞： OpenStack、Neutron、SDN、軟體定義網路

Abstract

With the rise of cloud computing, the infrastructure of data center has changed dramatically over the past few decades. OpenStack is most widely deployed open source cloud computing software in data center and private clouds or public clouds. Multi-tenancy data centers and network virtualization represent a complexity and extremely challenging networking environment, we can't effectively manage them by using the legacy network. However, the Software-Defined Network (SDN) technology is the perfect solution for these problems. SDN helps cloud-based data centers to achieve high flexibility and scalability with the platform that can efficiently handle the demanding network needs of present and future growth. Therefore, it is the trend that applies SDN to OpenStack helps data centers transform their networks. In this paper, we comprehensive review of OpenStack compatible SDN solutions, and explore the pros and cons of four types of currently solutions.

Keywords: OpenStack、Neutron、SDN.

1. 介紹

根據最新 Forbes 的預測[1]，自2016年到2020年在 IT-as-a-Service 預算中雲端運算的花費將成長83%；另外報導，雲端運算技術成為建置資料中心的主流且加速企業從資料集中的傳統資料中心轉移到雲端資料中心的趨勢，到2020年將會達56%[2]。IaaS 廠商 Amazon 提供的雲端服務 AWS，2016年到2017年同期比較其營運收入成長

23% [3]。在雲端資料中心興起的風潮下，更凸顯虛擬化平台上的整合型網路面臨新的挑戰。

在眾多的開放式雲端平台中，OpenStack 被認為將是引領 IaaS 服務的新趨勢，其營造的商業模式到2020年會成長34%[4]。其強調開放式的架構，可以用來建置私有雲、公有雲和混合雲等，把儲存(Cinder)、網路(Neutron)和計算(Nova)、設備和資源等組織起來。然而，OpenStack 在計算節點上所建立的虛擬主機(VM)其網路功能卻顯不足，雲端運算中多租戶(multi-tenancy)的虛擬網路路徑分配和資源調度之需求增加，而且大資料在不同租戶或應用程式之間的傳輸負荷提高[5]。除此之外，受限於路由支援度並不完善、防火牆的使用諸多限制、缺乏進階網路功能管理和擴展性受到局限等問題。

軟體定義式網路(Software Defined Network；SDN)適合運作在網路變動性高的環境中，尤其是雲端資料中心[6]。SDN把控制層(control plane)集中在控制器，透過可程式化可以快速的滿足網路服務佈署、動態更新網路狀態、自動供裝網路功能、集中式派發策略到網路設備和網路虛擬化[7][8]。SDN控制器的北向可程式化介面(API)可以和 OpenStack Neutron 整合，南向介面可以操控計算節點內的虛擬機器網路連線。適足以強化 OpenStack 網路功能。

本篇文章在第二節將說明計算節點內虛擬機器之間的網路橋接的兩種模式 Linux bridge 和 Open vSwitch(OVS)，並探討 OpenStack Neutron 專責提供網路服務的運作架構。第三節描述目前應用 SDN 技術與 OpenStack 整合的解決方案。第四節將比較分析這些方案的優劣。第五節為結論和未來發展。

2. OpenStack Networking

在我們論述應用 SDN 於 OpenStack 平台之前，我們必須了解虛擬主機彼此之間的網路連線和對外的網路連線。位於計算節點的 hypervisor 透過軟體建立虛擬交換器供虛擬機器連線，目前可以在 Linux 核心供裝 bridge 和 Open vSwitch 進行網路封包交換。然而，大部分的網路管理工作是由網路節點的 Neutron Server 所負責。以下將討論 Open vSwitch 和 Neutron 的運作架構。

2.1 Open vSwitch(OVS)

Linux 支援兩款虛擬交換機，快速簡單封包轉發的 linux bridge 模式和可以依 L2-L4進階進行封包傳送的 OVS 模式。就處理效能、集中管理能力、對外的管理介面(OpenFlow、OVSDB)支援度

和即時反應網路變動能力等各方面而言，OVS 均優於 bridge 模式[9]。

OVS 架構如圖1所示。當 VM1 要和 VM2 通訊，第一個封包(圖1紅線所示)會送至 user space 中的 ovs-vswnitchd。ovs-vswnitchd 是 Open vSwitch 系統程式(daemon)，用來管理和控制虛擬交換器，接著詢問 ovsdb-server 相關的介面、埠號資訊以作為封包轉發的依據。ovsdb-server 是輕量級的資料庫伺服器，提供 ovs-vswnitchd 查詢以下資訊：橋接設定檔、QoS 設定檔、openflow 控制器設定資訊和 netflow 設定資訊等等。最後自動產生網路流表(flow table)送到 kernel space。後續封包(圖中藍線所示)進到 OVS 均會比對 fast path 中 flow table 直接轉送出去。

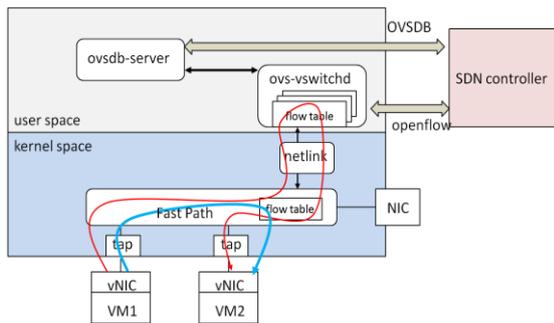


圖1 Open vSwitch 架構

2.2 Neutron Server

Neutron Server[10] 分為兩層，一是前端(frontend)：執行具 WSGI(Web Server Gateway Interface)介面的 HTTP 伺服器作為 REST API 層。另一是後端(backend)：具 pluggable 特性，以 plugin 的方式實現各式網路服務。

REST API 是 HTTP 伺服器的前端，負責把外部請求的服務交由後端適當的 plugin 作相對應的處理。Core API 中定義：基本的網路設定、埠號和子網段，屬於 Neutron 標準的 API。其他擴增的網路服務則是 Extension API，例如：LBaaS、FWaaS 和 VPNaaS 等。如圖2所示。

Plugin API 接收 REST API 請求之後，載入相關驅動程式並以 RPC 通訊機制呼叫位於本地端或計算節點端的 Agent。Agent 的工作是實現 plugin 所要求的任務。這些 plugin 中最核心的是 core plugin ML2(Modular Layer2) 包含了許多基礎的 layer2 的網路技術。ML2 包含 type drivers 和 mechanism drivers；type drivers 定義了支持的網路拓模類型：Flat、VLAN、VXLAN 和 GRE 等。Mechanism drivers 處理 type drivers 所建立的資訊，確認正確的啟動指定的網路橋接機制，如 OVS、Linux bridge 或特定廠商的驅動軟體。Mechanism drivers 可以利用 L2 Agent 經由 RPC 直接和外部設備或控制器溝通。而 core plugin 之外的其他 plugin 則稱為 service plugin，

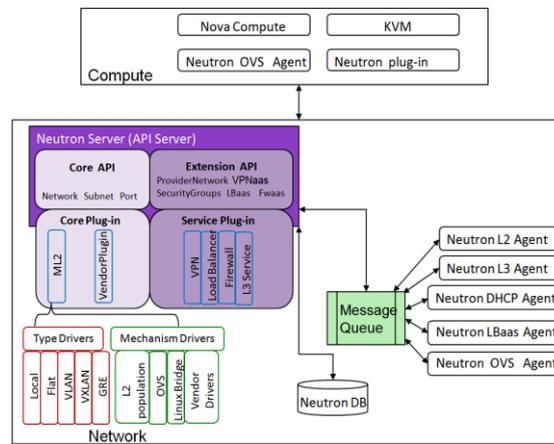


圖2 Neutron 元件

圖2中 Message Queue[11] 提供 Neutron Server 與其他 Agent 或服務之間高可靠度且允許非同步的訊息傳遞和交換。OpenStack 支援以下三種 Message Queue：RabbitMQ、Apache Qpid 和 ZeroMQ，均是輕量級、俱高處理能力但低延遲性的訊息系統。

RabbitMQ 和 Apache Qpid 是採用 Advanced Message Queuing Protocol (AMQP) 的 RPC 平台，以集中式或非集中式佇列伺服器的方式提供點對點的溝通。然而，ZeroMQ 是透過 TCP socket 直接點對點溝通。

3. 應用 SDN 於 OpenStack 之整合方案

SDN 是將控制層以軟體操控的方式集中控管的網路方案。OpenStack 是最多人使用且發展最成熟的開源雲端運算管理平台。本節以開源軟體為例說明應用 SDN 在 OpenStack 以作為虛擬資源和虛擬網路管理的工具。

3.1 MidoNet

MidoNet[12] 是一分散式具軟體定義虛擬網路功能的基礎設施即服務 (Infrastructure as a Service；IaaS) 平台。可以在實體網路上運載多個虛擬網路，即 overlay network 架構，如圖3所示，底層實體設施以實體線路介接，由 hypervisor 所建構的虛擬機之虛擬埠 (vPort)，串接到虛擬交換器和虛擬路由器，基於底層 IP 連線已經建立，各虛擬機埠口之間可以建立多個邏輯連線，也就是圖中的上層邏輯拓模圖。

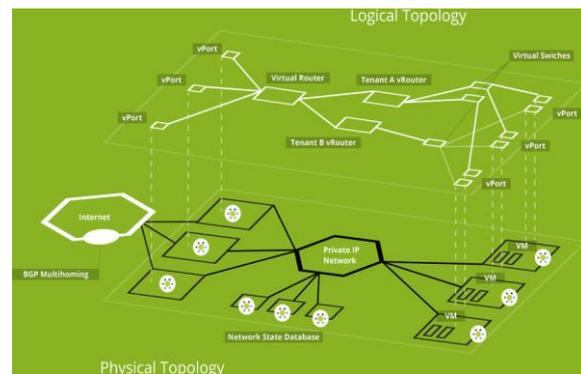


圖3 MidoNet overlay network[11]

MidoNet 應用於 OpenStack 的運作架構如圖4所示，主要元件如下：

- MidoNet Neutron Plugin：取代 OpenStack 中的 ML2 Plugin。
- MidoNet CLI 和 MidoNet API：利用 OpenStack keystone 套件進行管理身分認證和授權 MidoNet API 使用。另外也提供 CLI 命令列介面執行網路組態管理。
- MidoNet Agent：又稱為 Midolman 安裝於網路封包會流經的所有虛擬網路節點，包含所有計算節點和架構中的軟體式開道器。
- MidoNet Network State Database(NSDB)：利用 ZooKeeper 和 Cassandra 建置叢集式資料庫，前者用於儲存虛擬網路拓模圖和實體網路拓模圖。後者用來儲存資料流狀態資訊、NAT 位址連結表和網路設定檔等。然而，原生 Neutron 所使用的是 MySQL/MariaDB 資料庫；在建置時為了確認兩端資料庫一致，避免同步期間拓模圖異動立即更新至 ZooKeeper。因此，再開始備份之前設定 MidoNet API 為 read-only 模式。備份完成之後再改回 read/write 模式。

Midolman 是分散式架構中的 SDN 控制器，直接控制 OVS 運作，通過 linux netlink 操作 kernel space 中的核心模組 openvswitch.ko，以建立到虛擬機器的資料通道。

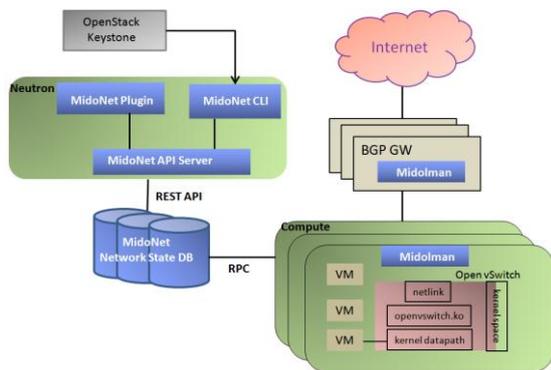


圖4 MidoNet 與 OpenStack 運作架構圖

3.2 Contrail/OpenContrail

Contrail 是一套以 SDN 實現網路功能虛擬化 (NFV) 軟體，以 MPLS L3VPN 和 MPLS EVPNs 技術作到 L3 和 L2 的 overlay network 架構。原為商業版軟體，由 Juniper 公司支援服務；之後，開源成為免費軟體稱為 OpenContrail[13]。架構圖如5所示，有一獨立的控制器系統，主要負責控制層，可分成三類節點：

- Control Node，實現控制層功能，透過 IF-MAP 接收 Configuration Node 傳遞之網路拓模和網路配置要求，藉此進行網路設備控制。IF-MAP 是網路存取控制介面的標準，可以相互溝通異常流量和攻擊，俱聯合防護網路安全能力。支援的南向協議有：XMPP、BGP 和 NETCONF。XMPP[14]是以 XML 串流為基礎的開放式即時通訊協定，能夠和 vRouter 之間建立即時通訊連線，並且傳送和接收即時訊息。Control Node 叢集之間以 BGP peering 建立連線；和其它實體網路設備之間則使用 BGP 或

NETCONF 協議。並不支援 OpenFlow 協定。

- Configuration Node，負責提供北向介面 REST API 與 OpenStack Neutron plugin 整合。因此，外部可透過 REST API 配置和取得網路資源，另外以 IF-MAP 協定和 Control Node 進行資料傳遞。
- Analytics Node，負責收集、分析和展示網路資訊，可以有助於網路除錯和了解網路使用狀態。該節點支援叢集建置，資料庫之間以分散式同步。

OpenContrail 不使用 OVS，而是自行開發另一套 vRouter 負責資料層虛擬網路封包轉送。因此在 OpenStack Neutron 需安裝 OpenContrail plugin，並於在每一個計算節點安裝 vRouter，其中位於 user space 的 vRouter Agent 透過 XMPP 協定接收來自 Control Node 網路配置政策，再把封包轉送規則的網路流表(flow table)送往位於 kernel space 的 vRouter forwarding plane，據此轉發封包。

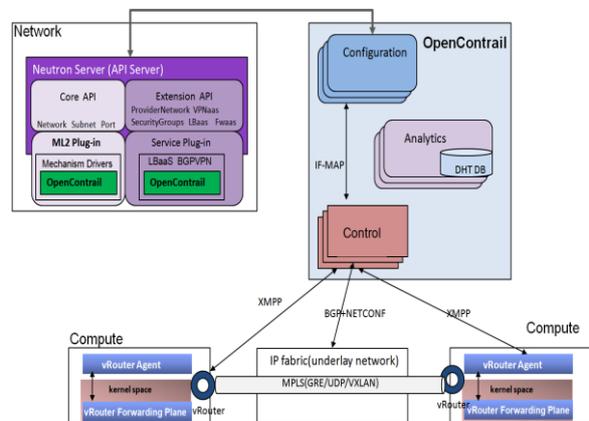


圖5 OpenContrail 與 OpenStack 運作架構圖

3.3 OpenDayLight

OpenDayLight[15]是 SDN 控制平台，架構上採用功能模組化且層次化分明，並通過各式 plugin 實現功能。如圖6所示。應用層和控制層之間透過北向 API 介面交換訊息，控制層和網路基礎設施層之間則是利用南向 API 介面控制網路設備。在 OpenDayLight 控制器內使用 JAVA 的動態模組化系統：OSGI 框架，易於安裝、啟動、停止和卸載模組；核心元件中的模型驅動的服務抽象層 (Model-Driven Service Abstraction Layer；MD-SAL)，可以讓不同模組使用統一的數據結構，作為南向 API 和北向 API 的溝通介面。MD-SAL 是使用 Yang Model 數據模型語言對相關請求和通知建立模型，再經由內部 plugin 對該模型查詢 MD-SAL 內的 data store，該 data store 內儲存網路參數、操作資訊和內部元件交換數據等，接著根據這些參數動態的進行路由調配，達到北向 plugin 和南向 plugin 訊息交換。

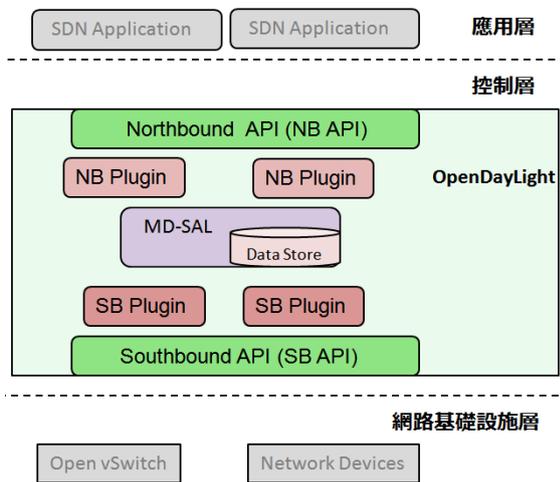


圖6 OpenDayLight 架構圖

以圖7為例說明 OpenStack 與 OpenDayLight 整合運作架構，圖中計算節點內虛擬主機的網路服務將交由 OpenDayLight 接管。在 OpenStack Neutron 內安裝 ODL Mechanism Driver，再透過 REST 介面向位於 OpenDayLight 的 Neutron Northbound 提出服務請求。NetVirt[16] 是 OpenDayLight 中支援 OpenStack Neutron API 的網路虛擬化應用，可謂 Neutron 服務供應者，可提供雲端多租戶虛擬化網路環境，支援軟體交換器 OVS 虛擬化、硬體交換器 VTEP 虛擬化和在虛擬的環境對網路服務功能虛擬化(SFC)。NetVirt 透過 MD-SAL 與北向 Neutron Northbound plugin 和南向 OVSDB/OpenFlow plugin 彼此傳遞訊息達到操控實體和虛擬交換器，作到兼容 SDN 網路設備和傳統網路設備的控管。

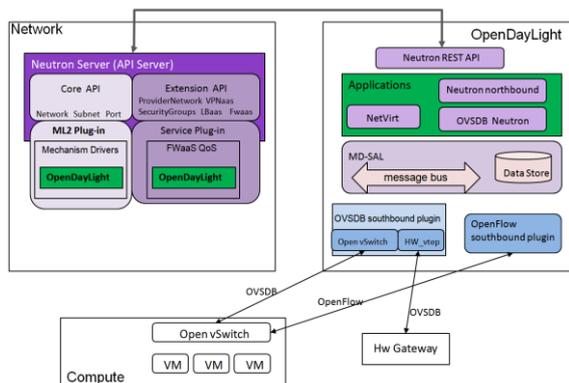


圖7 OpenDayLight 與 OpenStack 運作架構圖

3.4 Open Virtual Network (OVN)

在 OpenStack 中 Neutron Server 透過 message queue 的 RPC 與眾多 agent 通訊，造成效能低落問題，且耗費許多系統資源在維護 namespace。因此開發 OVS 的同一個團隊提出 OVN[17]，以 OVN 的 ML2 plugin 和 Service plugin 掛載到 Neutron Server，取代 OVS plugin。建置架構是以資料庫記錄實體網路資訊和邏輯網路資訊的方式，減少 namespace 的維護成本並實現 OVS 軟體定義式的虛擬化網路架構；在 OVN 資料庫和不同元件之間的通訊以 OVSDB 為管理協定，降低 RPC 通訊需求。OVS 是 OpenStack 預設的 backend，而 OVN 是以 OVS 為基礎開發，被視為和 OpenStack 有良

好的整合。

圖8是 OVN 架構，以 OVN Southbound DB 為集中式控制中心，紀錄實體網路和邏輯網路的執行狀態，連接 ovn-northd 和 ovn-controller，是所有計算節點 ovn-controller 的控制層協調者，支援分散式的叢集架構以因應擴展性需求。

ovn-northd 負責同步 OVN Southbound DB 和 OVN Northbound DB 的資料庫，把 northbound DB 的資訊轉成 southbound DB 的格式。

OVN Northbound DB 是和 OpenStack 的整合入口處，記錄邏輯網路埠號、邏輯交換器和邏輯路由器等資訊。

ovn-controller 安裝於每個計算節點的 hypervisor 上，北向連接 OVN Southbound DB，向其註冊底層的 hypervisor 資訊、開道和虛擬網卡介面(VIF)資訊；南向以 OpenFlow 控制器的角色連接 ovs-vswnchd，將 VIF UUID 轉換成 OpenFlow 埠號，亦即將邏輯 flow 轉換成實體 flow。另一南向連接 ovnsdb-server 是以 OVSDB 協定通訊。

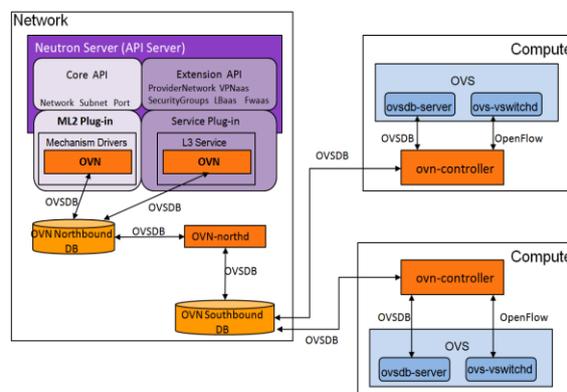


圖8 OVN 與 OpenStack 運作架構圖

4. 分析比較

MidoNet 是由日本 Midokura 公司所支持，與其商業版本 MEM(Midokura Enterprise MidoNet)最大差異是 MEM 提供 GUI 管理監控系統。OpenContrail 也有商業版本 Contrail，背後是 Juniper 公司提供服務和支援。OpenDayLight 和 OVN 是公開社群所發展的開源軟體。比較表如圖 10 所示。

OpenStack 計算節點支援多樣的 hypervisor，安裝時僅會擇一使用。MidoNet 支援的 hypervisor 是 KVM。OpenContrail 和 OVN 支援 KVM 和 Xen。OpenDayLight 與計算節點 hypervisor 不存在相依性。

MidoNet、OpenContrail 和 OVN 支援虛擬容器 docker 之網路配置。以 OpenContrail 為例，當 vRouter 建置 overlay network 之後，根據該架構產生 VRF 來作為封包轉發的依據，並且與 docker 容器連接，vRouter 是 DHCP server，docker 的虛擬網卡 veth 介面是 DHCP client 可以自動獲得 IP 配置和網路相關設定。OpenDayLight 目前支援 Kubernetes 容器管理系統，預計未來也將會支援 docker。

Support	MidoNet	OpenContrail	OpenDayLight	OVN
Hypervisors	KVM	KVM/Xen	No specific	KVM/Xen
Open source	YES	YES	YES	YES
Commercial Support	MEM (Midokura)	Contrail (Juniper)	NO	NO
SDN controller	MidoNet Agent	Contrail	OpenDayLight	ovn-controller
northbound	REST API	REST API	REST API	OVSDB
southbound	No specific	XMPP/NETCONF	OpenFlow/OVSDB	OVSDB
Docker	YES	YES	NO	YES
Kubernetes	YES	YES	YES	YES
Clustering	YES	YES	YES	YES

圖10 雲端資料中心主機之間 SDN 解決方案比較表

本文中論述的四種 SDN 開源方案與 OpenStack 整合的接入點是將 Neutron core plugin 和 service plugin，置換成各自開發的 plugin，如表 11 所示。

在 OpenStack 的網路節點和計算節點均以 OVS 作為虛擬機器之間的橋接和外部網路的介接。MidoNet 南向介面是由分散式安裝於每一個計算點的 MidoNet Agent 直接透過 kernel space 的 netlink 作封包和網路流表的比對轉發，而捨棄原來位於 user space 的 ovs-vswnched 和 ovsdb-server 對 kernel space 的控制。OpenContrail 自行發展 vRouter 有別於 OVS，分別安裝於 user space 和 kernel space，OpenContrail 控制器再透過 XMPP 協議與 vRouter 溝通，坊間 Juniper 和 Arista 交換器均支援 XMPP 協定，另外也支援 BGP 和 NETCONF 協議管理傳統網路設備。OpenDayLight 與上述兩方案迥異，維持 OVS 架構，在控制器南向 plugin 發展 OVSDB 和 OpenFlow 協議與外部 OpenStack 網路和實體網路介接。OVN 與 OVS 的整合度高，以 ovn-controller 控制。

Integrate	MidoNet	OpenContrail	OpenDayLight	OVN
Neutron core plugin	✓	✓	✓	✓
service plugin	✓	✓	✓	✓
Nova	MidoNet Agent	vRouter Agent	No specific	ovn-controller
OVS	✓	replaced by vRouter	✓	✓

圖11 OpenStack-SDN integration

5. 結論和未來發展

應用 SDN 技術整合 OpenStack 是當前趨勢，使雲端資料中心具智能管理能力。SDN 控制器北向是以掛載 plugin 至 OpenStack Neutron 的 ML2 plugin 和 service plugin 方式接手網路管理；SDN 南向介面管控計算節點上之虛擬機器網路則以 Agent 軟體介入 OVS 的運作。因此，SDN 控制器與網路節點的 Neutron 和計算節點的 OVS 在整合上有版本的高度相依性。

開源的 SDN 控制器 OpenDayLight、RYU、FloodLight 和 ONOS 相繼發展支援 OpenStack 的整合套件，其整合的切入點和運作模式有相當高的相似性[18]，本文僅討論整合發展最成熟的 OpenDayLight。MidoNet 和 OpenContrail 雖然是開源軟體，但背後有商業公司的資源挹注，可提供完整的功能和完善的支援服務。相對於以上所

述，OVN 是輕量級 SDN 控制器。

台灣高品質學術研究網路 TWAREN 已經建置完成 100G 骨幹網路，接下來將串連雲端資料中心，打造即時運算的人工智慧 (Artificial Intelligence ; AI) 技術平台的基礎架構，舉凡圖像識別、語音語意識別和時間序列預測等 AI 應用均可搭建在全新的 AI 優先型資料中心。預計 AI 對未來 30 年國家競爭力是關鍵[19]，政府各部會將合作引導學、研及業界投入，積極推動。本文綜合評述各類型 SDN 技術應用在 OpenStack 雲端資料中心，以作為建置的參考依據。

參考文獻

- [1] Roundup Of Cloud Computing Forecasts, 2017. <https://www.forbes.com/sites/louisclombus/2017/04/29/roundup-of-cloud-computing-forecasts-2017>
- [2] Journey to the Cloud, <https://assets.kpmg.com/content/dam/kpmg/xx/pdf/2017/02/the-creative-cios-agenda-journey-to-cloud.PDF>
- [3] Cloud Business Drives Amazon's Profits, <https://www.statista.com/chart/9174/amazon-operating-profit/>
- [4] Where OpenStack cloud is today and where it's going tomorrow, <http://www.zdnet.com/article/where-openstack-cloud-is-today-and-where-its-going-tomorrow/>
- [5] Bitar, N., "Multi-Tenant Data Center and Cloud Networking Evolution," OFC/NFOEC 2013
- [6] R. Cziva, D. Stapleton, F. P. Tso, and D. Pezaros, "SDN-based virtual machine management for cloud data centers," in Proc. IEEE 3rd Int. Conf. Cloud Netw., 2014, pp. 388–394.
- [7] O. N. Foundation, "Software-defined networking: The new norm for networks," Open Networking Foundation, Tech. Rep., 2012.
- [8] B. N. Astuto, M. Mendonca, X. N. Nguyen, K. Obraczka, and T. Turletti, "A Survey of Software-Defined Networking: Past, Present, and Future of Programmable Networks," 2014, accepted in IEEE Communications Surveys & Tutorials To appear in IEEE Communications Surveys & Tutorials.
- [9] B. Pfaff, J. Pettit, T. Koponen, K. Amidon, M. Casado, S. Shenker. "Extending networking into the virtualization layer " ACM SIGCOMM Workshop on Hot Topics in Networking (HotNets), October 2009.
- [10] Networking architecture, <https://docs.openstack.org/security-guide/networking/architecture.html>
- [11] OpenStack website, <https://docs.openstack.org/security-guide/messaging.html>
- [12] MidoNet website: <https://www.midokura.com/midonet/openstack/>
- [13] Opencontrail architecture documentation, <http://www.opencontrail.org/opencontrail-architecture-documentation/>
- [14] Extensible Messaging and Presence Protocol (XMPP): Instant Messaging and Presence, <https://tools.ietf.org/html/rfc6121>
- [15] OpenDaylight, <https://www.opendaylight.org/>
- [16] NetVirt, <https://wiki.opendaylight.org/view/NetVirt>
- [17] Bryant, R.; Mestery, K.; and Pettit, J. OVN: Open Virtual Network for Open vSwitch, <http://openvswitch.org/support/slides/OVNVancouver.pdf>
- [18] O Tkachova, M J Salim, A R Yahya, "An analysis of SDN-OpenStack integration", Problems of Info communications Science and Technology (PIC S&T), 2015 Second International Scientific-Practical Conference
- [19] 5 年投入 160 億 建構我國 AI 創新生態環境, http://www.ey.gov.tw/News_Content2.aspx?n=F8BAEBE9491FC830&s=A3F7679A925B6C39