How to Deliver App Services to Every App

Enterprises are migrating to a software-defined cloud architecture where each application has associated services—in other words, a per-app-service model. F5 provides lightweight, cost-effective load balancing services to ensure the availability of all your applications in the cloud.

White Paper

Introduction

Application growth is spiraling upward as a result of the proliferation of mobile devices and the movement to the Internet of Things (IoT). This has created challenges around application deployment time, scalability, and availability. Public clouds and web monsters like Google and Facebook have achieved fast deployment times and true agility in their data centers—and other businesses want to replicate that. Many have redesigned their data centers using a cloud and software-defined strategy to realize the benefits of application agility, automation, and ultimately, operational efficiency.

This software-defined data center (SDDC) includes fully automated provisioning and management tools, application templates, and service catalogs so application owners can "order" all the services they need to deploy, secure, and manage each application. At the same time, these enterprises are moving to a two-tier hybrid architecture that comprises a per-app-service tier using the software everywhere approach, and an application services tier that enables them to take advantage of the benefits of specialized hardware for specific services at the edge of the network.

Per-App-Service Model

As part of this software-defined data center, enterprises have already virtualized their compute and storage infrastructure. Now, they're looking to virtualize their network devices to fit in with this distributed service architecture, and provide the full breadth of services for their application service catalog.

This model (see Figure 1), which results in a per-application service instance, seeks to reduce the costs and configuration complexity associated with hardware appliances and provide rapid automated, on-demand application deployment. Having a service for each application means that it also has to scale, support granular, per-application control and monitoring, and be fully integrated with orchestration and automation systems.
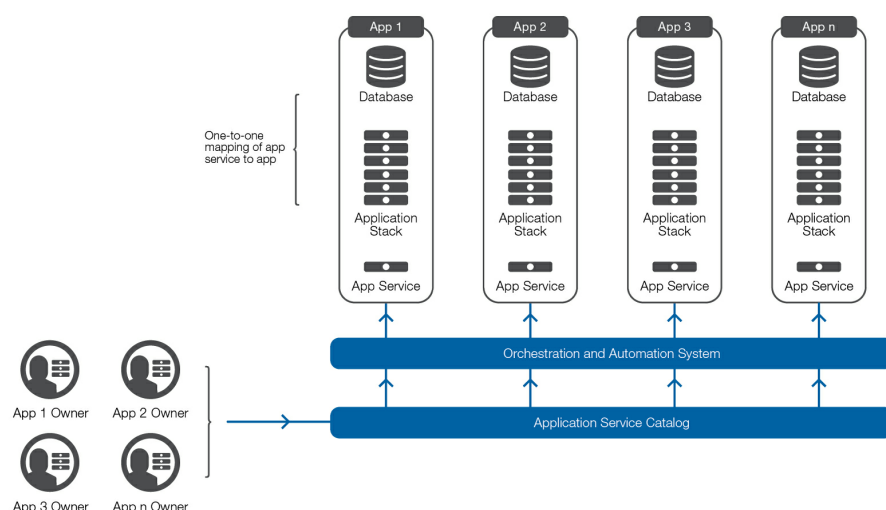
Figure 1: Network per-app-service architecture in SDDCs

While this model limits failures to individual application stacks, what is needed for each application is basic load balancing, or in some cases, advanced application delivery, to ensure the availability and scalability of all those thousands of applications. The challenge then is in providing software-based load balancing services in this dynamic virtual environment as part of that service catalog.

How to Choose a Lightweight Load Balancing Service Solution

If basic load balancing services are all that is required, any potential solution must then meet these five requirements (see Figure 2):

- Basic yet powerful capabilities
- Lightweight footprint
- Orchestration- and automation-friendly
- Multiple deployment methods
- Cost-efficient

Requirement	Benefit
Basic yet powerful load balancing service	• High availability • SSL offload • Health monitoring
Lightweight footprint	• Efficient resource use • Higher service density
Orchestration- and automation-friendly	• On demand automated deployment • Agile application scaling
Multiple deployment methods	• Deployment across multiple environments and data centers
Cost-efficient	• Subscription-based consumption models • 24x7 support included

Figure 2: Load balancing service requirements for a successful SDDC

Basic load balancing capabilities should include L4 (TCP) and L7 protocol (HTTP/HTTPS) support; act as both a reverse and transparent forward proxy; include multiple load balancing algorithms; support persistence so that traffic gets directed back to the same server; and provide SSL offload, high availability (HA), and performance and health monitoring.

With a lightweight footprint, a load balancing service should efficiently use the commodity server's CPU, memory, and storage resources to provide the highest possible density and to scale appropriately, as well as to minimize virtual machine resource usage.

It should easily integrate into orchestration and automation systems to truly achieve the benefits of software-defined architecture: fully automated deployment via orchestration systems, and the ability to scale the load balancing service on demand.

Some enterprises may have multiple virtualization environments and data centers that require the solution to be deployed as a virtual machine across different hypervisors or on bare metal to maximize performance. These methods offer flexibility and provide consistent service across any infrastructure.

Advanced traffic management capabilities include:

- Broader protocol support (UDP, SIP, SPDY)
- Dynamic routing (BGP, RIP, OSPF, ISIS, BFD support)
- DDoS protection and access management
- DNS and global server balancing
- Web performance optimization (intelligent client caching, adaptive compression, image optimization)

Lastly, cost-efficiency ties directly into total cost of ownership (TCO). How the solution is licensed and supported over the lifecycle of the service should be considered. In the cloud utility model, enterprises should have not only infrastructure flexibility, but also cost flexibility—making a subscription-based consumption license model ideal.

Introducing the F5 LineRate Platform

F5 is the leader in both physical and virtual Application Delivery Controllers, with a broad portfolio of Software-Defined Application Services™ that are delivered via a high-performance services fabric. The F5 portfolio now includes LineRate, a network application services platform designed from the ground up for SDDCs. The LineRate platform includes F5® LineRate® Point™ Load Balancer, a cost-effective software product that is based on the scalable LineRate Operating System (LROS) architecture. Point Load Balancer is a lightweight virtual load balancer that provides basic yet powerful application availability services.

LineRate Point Load Balancer for the SDDC and Cloud

Point Load Balancer provides basic yet powerful load balancing features with built-in HA clustering support that ensures reliability and uptime. For instance, it includes sub-second node-to-node failover and self-healing internal features to ensure that if a script, process, or other issue causes an internal error, recovery is rapid (milliseconds) and automatic. Support for multiple persistence methods—cookie, source IP, and application—is standard.

Securing data passing over the Internet has never been more important or more challenging. Protecting your data—and that of your customers—requires the latest SSL/TLS security measures. Point Load Balancer delivers high-performance, scalable SSL offload with next-generation elliptical curve cryptography (ECC) and perfect forward secrecy (PFS).

Designed for automated deployment and easy manageability, Point Load Balancer features a REST-based API, CLI and GUI interfaces, and connectors to third-party orchestration systems such as Chef and Puppet. You can monitor all aspects of health and performance with visibility into thousands of metrics via the REST API or SNMP.

Point Load Balancer is very lightweight, and supports stateless PXE or standard disk install. It can be deployed as a virtual machine (via VMware/KVM/Amazon hypervisors) or on bare metal commercial-off-the-shelf (COTS) servers.

Point Load Balancer is economically priced and may be purchased via F5's Volume Licensing Subscription (VLS) program that is specifically designed to meet the SDDC per-app-service cost efficiency requirements.
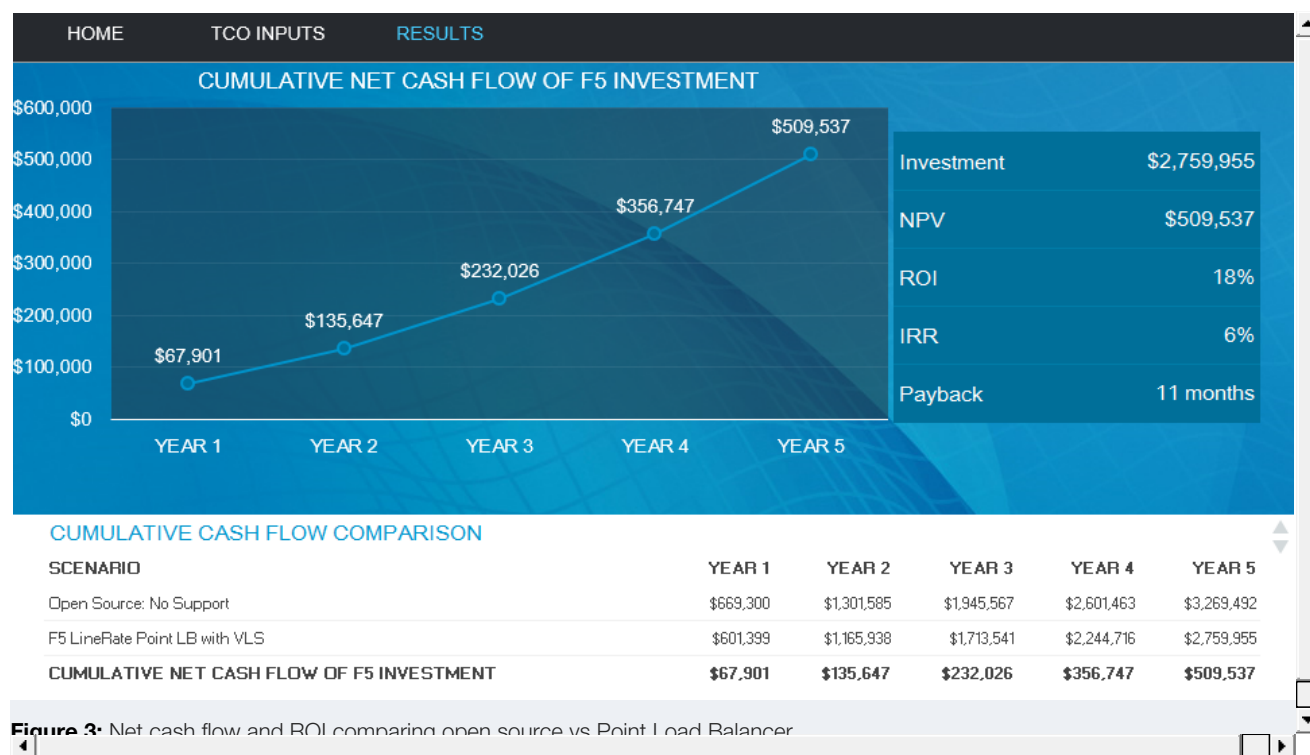
Challenges with Open Source Software for Load Balancing

Enterprises can develop their own solution using open source tools, like HAProxy and Nginx, or rely on "free" add-ons from vendors that do not have specific expertise in load balancing technologies. While appearing to meet the cost-efficiency requirement for at-scale cloud deployments, these types of solutions often end up being more costly in the long run. They tend to be more difficult to implement, manage, and scale than commercially supported solutions. The main challenge is being able to fully integrate them with your orchestration and automation systems to make it work in a large enterprise environment.

You also must consider security patches and other updates that are part of maintenance and support. Based on its engineering experience and customer feedback, F5 estimates that it could take three to four senior software engineers to develop, implement, maintain, and support the open source tools. Factoring in that operating cost and comparing LineRate's VLS pricing (assuming 400 Point Load Balancer instances with a one-year subscription and fully burdened cost of four software engineers over a five-year period), Point Load Balancer has a better TCO with an ROI of 18% (see Figure 3).

| HOME | TCO INPUTS | RESULTS |

CUMULATIVE NET CASH FLOW OF F5 INVESTMENT

Investment	$2,759,955
NPV	$509,537
ROI	18%
IRR	6%
Payback	11 months

CUMULATIVE CASH FLOW COMPARISON

SCENARIO	YEAR 1	YEAR 2	YEAR 3	YEAR 4	YEAR 5
Open Source: No Support	$669,300	$1,301,585	$1,945,567	$2,601,463	$3,269,492
F5 LineRate Point LB with VLS	$601,399	$1,165,938	$1,713,541	$2,244,716	$2,759,955
CUMULATIVE NET CASH FLOW OF F5 INVESTMENT	**$67,901**	**$135,647**	**$232,026**	**$356,747**	**$509,537**

Figure 3: Net cash flow and ROI comparing open source vs Point Load Balancer.

For solutions developed in-house, there is typically no dedicated support to address problems that arise. Support is provided by developer teams with minimal experience. While both HAProxy and Nginx offer paid versions that include limited support, compared to those licensing costs, Point Load Balancer with VLS pricing still has better TCO given the standard 24x7 support, which includes all software patches and updates. In addition, there are feature differences that impact meeting the requirements stated earlier, namely, the ability to work with orchestration and automation systems. Out of the box, Point Load Balancer features a proper REST API (meaning that the LineRate REST API was built in at the beginning—not "bolted on") and has better manageability (GUI, CLI) and operational visibility than these open source solutions. Ultimately, open source options could lead to more complex deployments and maintenance, pushing OpEx costs up over time.

Two-Tier Hybrid Application Services Architecture

Enterprises typically require more than basic load balancing services for each application. In today's business environment, it's critical to protect your network and applications against the multi-layer spectrum of attacks. A defense-in-depth approach guards against both network and application level attacks with services such as distributed denial-of-service (DDoS) mitigation, firewall services at scale, and high-volume SSL acceleration, which typically is suited for the edge of the network. That's where a two-tier hybrid architecture comes into play. The first tier manages all traffic entering the network and data center based on overall business and security policies. For these services that deal with high volumes and require the highest performance and scalability, dedicated, purpose-built hardware can be more cost-effective than commodity computing infrastructure.

The second tier manages the application stack inside the data center, which leverages the highly scalable, flexible virtual app service architecture and service catalogs to deliver both basic load balancing and advanced services on a per-application basis. Advanced services can include advanced, app-specific traffic management, application firewalling, access management, and non-HTTP traffic. Enterprises can select the virtual platform and services that match their business and technical needs. This two-tier hybrid architecture (see Figure 4) offers the best of both worlds: hardware where it's needed and software flexibility closer to the app.
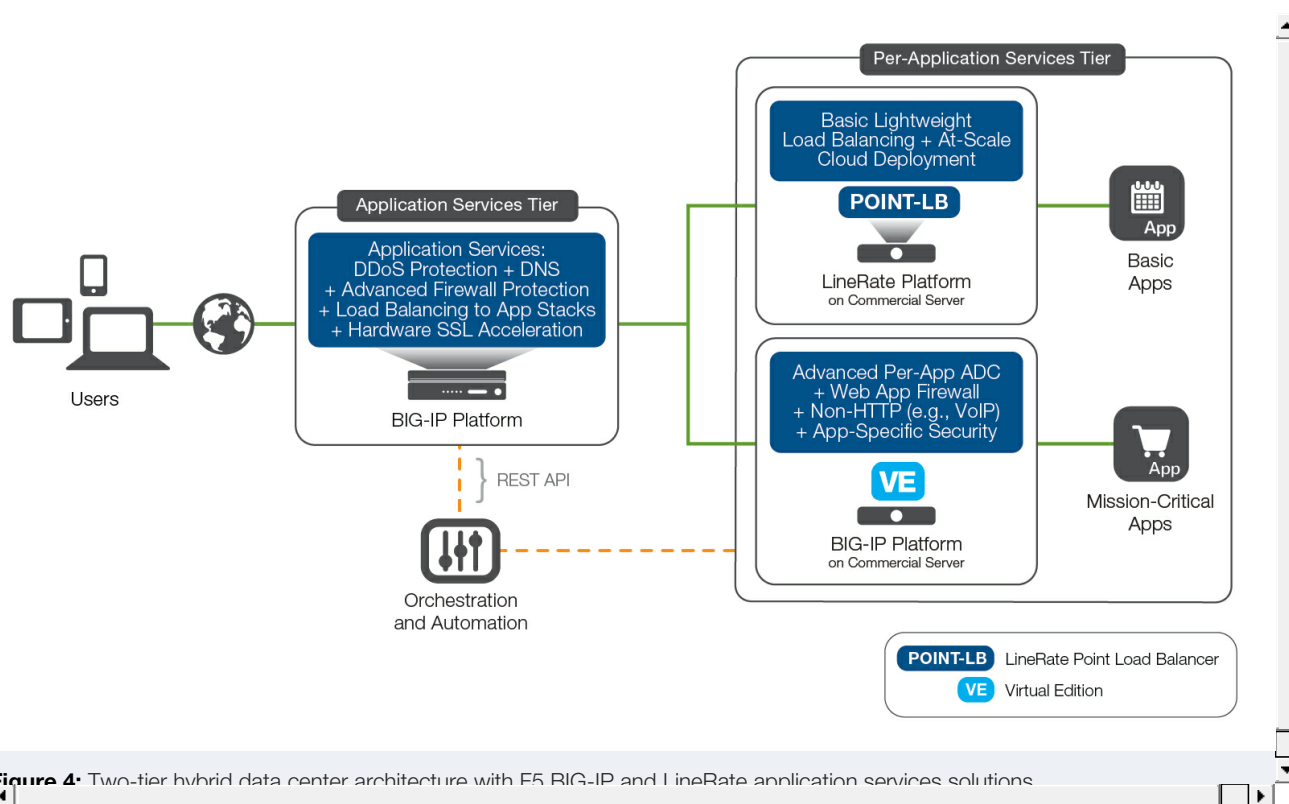
Figure 4: Two-tier hybrid data center architecture with F5 BIG-IP and LineRate application services solutions

A two-tier hybrid architecture also offers a clean separation of functionality and control. Application teams can have full administrative control of their application tier services but with isolation from other tenants, which creates failure boundaries in the event of misconfiguration or other human error. Network operations teams control the overall network and front door services.

F5 High-Performance Services Fabric

The F5 Synthesis™ architectural vision enables enterprises and service providers to deliver application services without constraints. F5 provides a catalog of Software-Defined Application Services, delivered via a high-performance services fabric, that enables applications to be fast, available, and secure on any device, whether on-premises, in the cloud, or both. F5 Synthesis moves organizations away from managing devices and toward managing services. In the first tier of the solution, where hardware is needed, F5 offers high-performance VIPRION® and BIG-IP® hardware platforms based on the market-leading ScaleN™ architecture that can scale to a throughput of hundreds of gigabits per second, and hundreds of millions of concurrent connections.

At the second application services tier, F5 offers LineRate for basic load balancing and BIG-IP virtual editions for advanced traffic management, web application firewall, and access and identity management services. This fabric is programmable and can be managed and automated with orchestration systems. Standardizing on F5 as a strategic data center partner for both front door and application services provides operational efficiency and lower TCO as enterprises have one set of services to support across the solution stack.

Conclusion

Enterprises are moving to a software-defined, private cloud data center model for agility, operational efficiency, and a self-service approach to deploying applications and associated services. They are utilizing a two-tier hybrid services architecture to get the benefits of specialized hardware for front door network services and scalable software for application stack–specific services. For basic load balancing at the application stack tier, F5 LineRate Point Load Balancer provides the ideal lightweight, scalable solution with standard 24x7 support. F5's broad portfolio of Software-Defined Application Services ensures that all applications are fast, available, and secure. These highly programmable services integrate with the orchestration and automation systems of a software-defined data center.