



Multi-Tenancy Designs for the F5 High-Performance Services Fabric

F5 has transformed the traditional networking design of highly available pairs of hardware devices to create a new architecture—a multi-tenant high-performance services fabric.

White Paper
by F5



WHITE PAPER

Multi-Tenancy Designs for the F5 High-Performance Services Fabric

Introduction

The explosion in numbers of connected devices and predicted growth in machine-to-machine communications are placing new demands on today's app-centric infrastructure. Recognizing the need for more flexible capacity, improved availability, and faster time-to-value, organizations are adopting innovative solutions such as hybrid cloud computing, software-defined networking, and DevOps-influenced agile delivery techniques. The challenge this cloud-driven environment presents to IT and their suppliers is creating new application delivery architectures that augment and integrate with these underlying shifts in the infrastructure. At the same time, these architectures must still supply the layer 4–7 application services such as security, acceleration, or availability that applications require.

To address the challenge of creating a new application delivery model, F5 has transformed the traditional networking design of highly available pairs of hardware devices to create a new architecture—a multi-tenant high-performance services fabric.

What makes a services fabric?

The F5 services fabric creates a scalable, all-active container of powerful layer 4–7 application services delivered as a flexible resource pool. Designed to scale both up (through hardware and license upgrades) and out by adding more devices, the fabric offers flexible capacity, more efficient high-availability and is ready to be integrated into automation systems through a powerful API.

Fabrics can be created in both traditional data centers and in cloud environments and can be connected across many network underlays and overlays. Additionally, fabrics can contain both physical and virtual BIG-IP appliances and can accommodate services supplied by [LineRate Point Load Balancer](#) and [LineRate Precision application proxy](#) components.

With application services now delivered by a consolidated, multi-device fabric, it is essential that these services can be effectively shared across multiple applications, business units, or customers with the right levels of security and isolation to suit the organization. Flexible multi-tenancy capabilities are, therefore, a foundational component of the F5 services fabric.



WHITE PAPER

Multi-Tenancy Designs for the F5 High-Performance Services Fabric

Benefits of multi-tenancy

Multi-tenancy designs traditionally focus on the obvious benefit of cost savings that are driven by better utilization of the underlying infrastructure—the same benefits that the “first waves” of both server virtualization and cloud platforms promised. By consolidating more services onto fewer devices (physical or virtual), savings in acquisition, ongoing support, and administration can be realized. Just as in virtualization and cloud, agility is a significant benefit gained from the services fabric and multi-tenancy.

Once new services can be provisioned on to an existing infrastructure, without the need to purchase or create new components, the time to provision (and de-provision) new services drops dramatically. With a secure isolation of networks and workloads, new customers, projects, or business units can be serviced by an expandable pool of resource supplying the layer 4–7 services that their applications require.

While all this rapid (and often self-service) infrastructure provisioning brings value to the lines of business consuming the services, it must also facilitate control and monitoring for the IT department. By creating standardized services available from the services fabric, and assigning resource controls to the tenants, IT can apportion infrastructure costs to specific tenants. While chargeback is a business and organizational decision, some form of “showback” should be seen as a requirement for multi-tenant requirements. This is especially true in organizations that have a large untapped IT demand, which self-service cloud-style infrastructures will often unleash.

Delivering multi-tenancy

Organizations have different business and technical requirements for multi-tenancy depending on factors such as the nature of their businesses (e.g. regulatory or market requirements around data security), their funding models, or their wider IT strategy. F5 offers a number of different multi-tenancy models to allow customers to create an architecture that is best suited to their individual needs. Below are four multi-tenancy models, followed by suggestions on how to map each one to a particular set of business requirements.

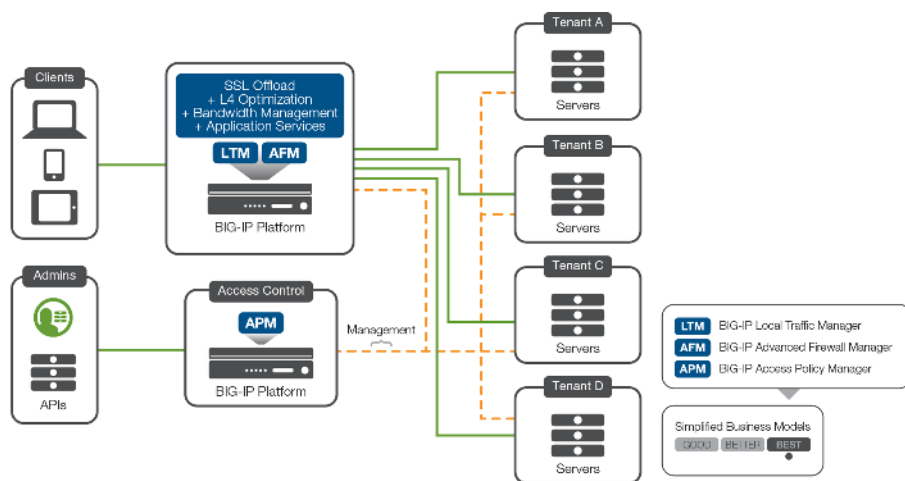
Route domains and administrative partitions

Route domains and administrative partitions, two core BIG-IP technologies, can be used to create a flexible multi-tenant design that offers both application traffic isolation and administrative separation between tenants. Route domains and administrative partitions are available across all BIG-IP platforms.



WHITE PAPER

Multi-Tenancy Designs for the F5 High-Performance Services Fabric



Route Domains create strictly defined address spaces within a network. Each route domain contains IP address spaces, routing information, and VLANs. IP address spaces can be duplicated between domains, allowing easy reuse of RFC 1918 private addressing for multiple customers or projects. Route domains can be strictly isolated from one another, or have explicitly controlled access between them. This allows a common “front end” to be presented to an access network but with services running within dedicated tenant spaces. Although system resources are not explicitly dedicated, each domain can be rate-limited by connections or throughput to provide some resource constraint. This design allows for the most efficient use of system resources since each domain will consume only the resources it requires.

Administrative Partitions ensure that specific users are granted access to only the partitions for which they are authorized. This is in addition to the role-based access that restricts users to specific operations. With administrative partitions, configuration objects are placed into specific partitions that only authorized users can access. While this design does have some limits in terms of number of objects and partitions, it is quite capable of maintaining many hundreds of administrative partitions and route domains, making it a suitable candidate for larger-scale multi-tenancy.

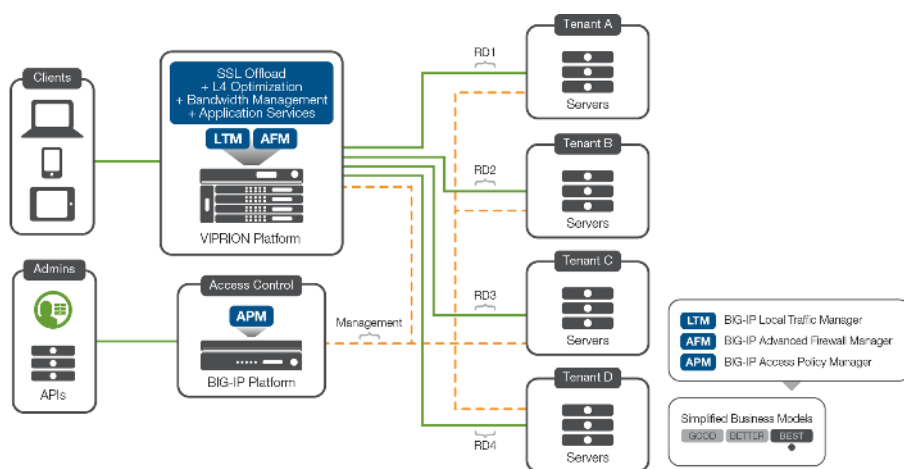


WHITE PAPER

Multi-Tenancy Designs for the F5 High-Performance Services Fabric

vCMP

Virtual Clustered Multiprocessing (vCMP) creates multiple, isolated instances of the BIG-IP software on a single F5 hardware platform. Each instance has its own CPU, memory, and disk, and can take advantage of multiple assigned CPU cores using the same clustered multiprocessing¹ design used on all F5 platforms. BIG-IP “guests” run on F5 vCMP-enabled hardware using a standards-based, purpose-built hypervisor that provides robust security and isolation.²



Each vCMP guest can run an isolated, independent version of the BIG-IP software, allow differentiation of services used by tenants, and be upgraded or rebooted without disruption to other guests. In addition, on VIPRION platforms, guest instances can be dynamically scaled to consume more resources as required. Different numbers of cores can be assigned to guests running on a platform to allow the creation of tenants with differing capacity. Also, guests can be assigned to dedicated networks or share common networks.

The visibility of the network infrastructure is controlled by the hypervisor—guests can only view and access their assigned networks. Authentication and access are controlled individually by each guest—as is the high-availability design, which may vary between guests. Since each guest is a full version of the BIG-IP system, additional multi-tenancy, using route domains and admin partitions, can be enabled within each one. This might enable a service provider (which could be an IT team with internal business consumers) to offer different levels of multi-tenancy, differentiated by levels of isolation or customer access.



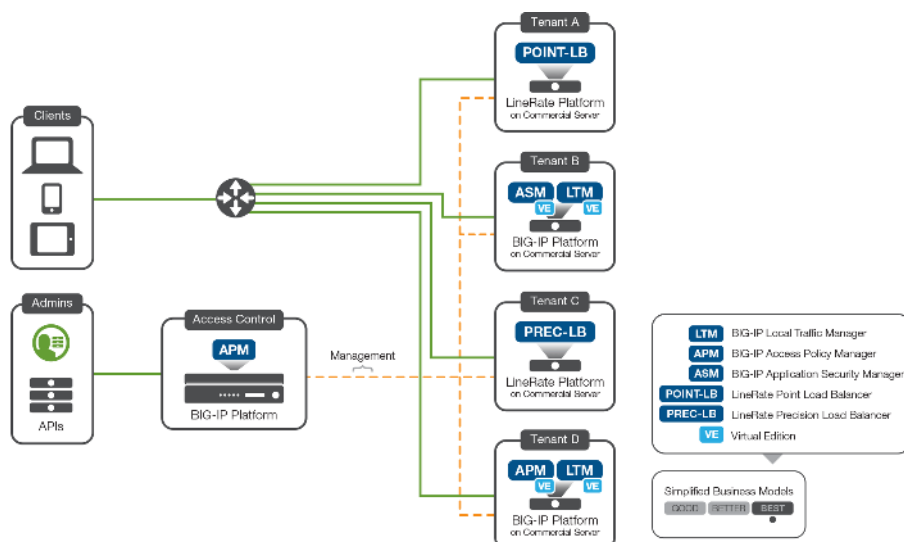
WHITE PAPER

Multi-Tenancy Designs for the F5 High-Performance Services Fabric

The maximum number of vCMP guests that can be created on a particular BIG-IP platform is determined by the available CPU cores' memory and other system resources.³ Since guests are hard-allocated CPU cores, without over-provisioning, this design promotes a smaller number of higher-capacity guests while enabling higher levels of isolation and security than other models. VIPRION chassis and selected BIG-IP models are available with vCMP technology.

Virtual Editions and LineRate

Using a general-purpose hypervisor and virtual machine management framework allows IT teams and service providers to create multi-tenancy compute environments with virtual machines that share an underlying compute infrastructure. With editions of BIG-IP and F5 LineRate⁴ software for popular hypervisors,⁵ organizations can build tenant solutions using dedicated virtual editions in the same manner that they have done for servers.



With separate virtual instances providing dedicated application services to specific tenants, the layer 4–7 services that the tenant applications require can be tightly coupled to the application servers. This creates a ring-fenced bundle of configuration that can be easily replicated or migrated to other locations for scale or availability. With a choice of supported F5 platforms, organizations can pick the best one for a tenant depending on their technology or business driver.



WHITE PAPER

Multi-Tenancy Designs for the F5 High-Performance Services Fabric

Platform	Attributes
BIG-IP Virtual Edition	Full features of the BIG-IP platform: software modules, iRules, application templates, analytics database. Higher resource requirements for disk, CPU, and memory.
LineRate Point	A simple, high-performance but lightweight load balancer. Suitable for large-scale deployments where hundreds of instances are required. Orchestration ready with a full REST API.
LineRate Precision	Extends LineRate Point into a high-performance application proxy offering a fully programmable data path using node.js with access to thousands of node.js modules.

All resource management, just like the server components, is handled by the underlying virtualization framework. BIG-IP and LineRate virtual appliances are available in different throughput capacities (upgradable by license), so that tenants can be given as much capacity as they require. Virtualized environments promote the rapid creation, deployment, and destruction of virtual infrastructure, enabling methodologies such as DevOps and continuous delivery.

The virtual edition model of multi-tenancy offers near limitless scalability, as additional VE devices can be created as required. With a choice of platforms (see sidebar), organizations can select from lightweight devices with core functionality or platforms offering a rich range of functionality (such as web application firewalling, remote access management, and application layer acceleration), which enviably require more resources from the underlying infrastructure.

Managing the multiple virtual devices created by this tenant model needs to be a priority since there will likely be more separate software instances than in the route domain or vCMP models. The BIG-IQ platform⁶ can manage the creation, licensing, and management of BIG-IP virtual devices through an intuitive GUI and RESTful API.

While F5 virtual devices of all flavors offer a comprehensive range of application services, they lack the specialized hardware of the physical BIG-IP and VIPRION platforms.

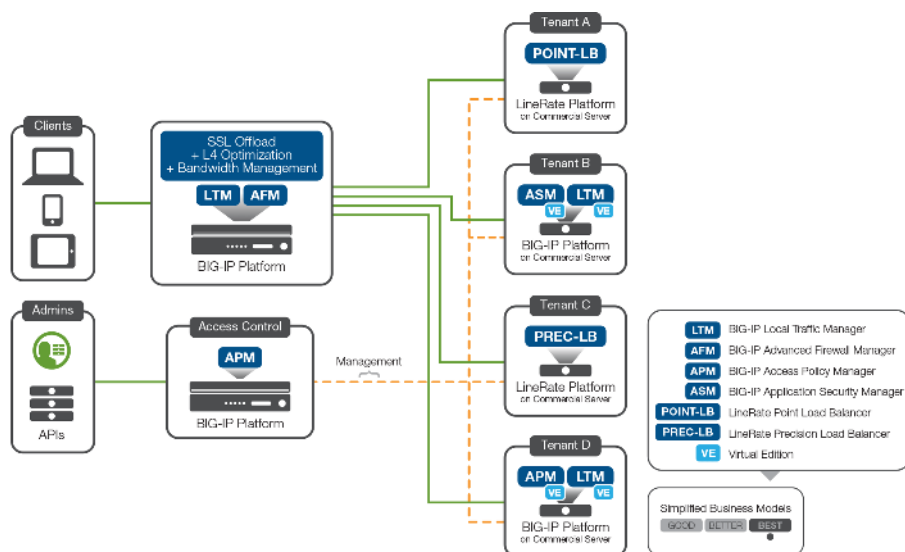


WHITE PAPER

Multi-Tenancy Designs for the F5 High-Performance Services Fabric

Hybrid

When organizations require Distributed Denial of Service (DDoS) attack mitigation, firewall services at scale, high-volume SSL session decryption, and dedicated hardware can be more cost-effective than commodity computing infrastructure. At the same time, the advantages of the highly scalable, flexible, virtual device model of multi-tenancy are highly attractive. A hybrid model offers the best of both worlds. It enables network-centric services and basic traffic separation to be handled by a high-capacity specialized hardware device, and application-centric services such as traffic management, application firewalling, or actual application logic to be handled by pools of the appropriate virtual devices.



In addition to the advantages of capacity and flexibility, the hybrid model offers a clean separation of functionality with network layer services administered on one component and application layer services on another.

Application teams can have full administrative control of all application tier devices but with isolation from other tenants, creating failure boundaries in the event of misconfiguration or other human error. By using the load balancing capabilities of the network layer device, multiple smaller application layer virtual devices can be used to scale horizontally—with application traffic further distributed beneath that layer.

This model is highly scalable since many of the BIG-IP platforms are equipped with network processing hardware that allows hundreds of Gigabits of throughput in a single device. At the application services layer below, organizations can select from the platform that suits their business and technical needs.



WHITE PAPER

Multi-Tenancy Designs for the F5 High-Performance Services Fabric

Management

All F5 devices offer CLI, GUI, and RESTful API access, allowing manual or programmatic configuration management. Orchestration through BIG-IQ or another orchestration tool using the API is recommended to ensure reliable, repeated operations. However, when tenants need extensive unique customization, the GUI is utilized.

In multi-tenant infrastructures, segregation and control of management access is frequently a primary concern as the impact of misconfiguration may affect more services, business units, or customers. Management of systems that span multiple applications presents high-value targets for compromise and must be given adequate protection, especially when exposed to external networks.

BIG-IP Access Policy Manager (APM) offers a comprehensive suite of tools to authenticate network, API, or GUI access to devices. With features such as application tunneling, endpoint inspection, and a highly configurable visual policy editor, BIG-IP APM can be used to build appropriate security controls to protect shared infrastructures yet still enable tenant administration.

Choosing the right model

Each of the architectures described above offers a valid yet different multi-tenancy design. How should organizations select the best model for their businesses? First, they must understand their priorities in terms of key multi-tenancy attributes:

Attribute	Description
Separation and isolation	How isolated are the tenants—can the workload of one tenant affect the resources available to the other? How separate is the management access between tenants? Note that all designs offer <u>extremely robust traffic separation</u> if required.
Resource efficiency	What levels of utilization can be achieved? This is usually an inverse to the isolation metric above—the more resources that are shared between tenants, the less the over-provisioning overhead is likely to be.
Simplicity	How complex is the management likely to be and how much coordination between devices is likely to be required? How many actions will orchestration tools need to take to provision new services?
Flexibility	How many different services or tenant types can the design support? Can there be different software capabilities between tenants or different versions? How wide a range of uses will the infrastructure support?



WHITE PAPER

Multi-Tenancy Designs for the F5 High-Performance Services Fabric

With a weighting for these attributes, we can examine the four designs and score them in each category. The table below scores the architectures relative to each other—*not* relative to a particular customer’s weighting of their requirements. To get a broad and simple guide, multiply the score below by the weighting above in each category.

Design	Separation and isolation	Resource efficiency	Simplicity	Flexibility
Route domains	•	•••	•••	•
vCMP	•••	•	••	••
Virtual devices	••	••	••	••
Hybrid design	••	•	•	•••

This produces a very basic pointer and may eliminate some designs from consideration for a particular scenario. Deeper consideration will obviously be needed in order to make an informed choice of one design over another, or indeed a customized design built from components of each model.

Summary

The different multi-tenancy models offered by the F5 high-performance services fabric enable a consistent set of application services to be delivered across multiple applications, business units, or customers, but with the levels of separation and isolation tailored to suit the needs of the organization. As a result, businesses can achieve greater efficiency, lower costs, and improved agility—all of which will be required to reap the benefits from ongoing changes in the IT landscape.

¹ <http://www.f5.com/pdf/white-papers/vcmp-wp.pdf>

² <http://www.f5.com/pdf/white-papers/multi-tenancy-security-vcmp-tb.pdf>

³ <https://support.f5.com/kb/en-us/solutions/public/14000/200/sol14218.html>

⁴ <https://linerate.f5.com>

⁵ https://support.f5.com/content/kb/en-us/products/big-ip_ltm/manuals/product/ve-supported-hypervisor-matrix/_jcr_content/pdfAttach/download/file.res/VE%20Supported%20Hypervisor%20Matrix.pdf

⁶ <https://f5.com/products/big-ip>

WHITE PAPER

Multi-Tenancy Designs for the F5 High-Performance Services Fabric



F5 Networks, Inc.
401 Elliott Avenue West, Seattle, WA 98119
888-882-4447 www.f5.com

Americas
info@f5.com

Asia-Pacific
apacinfo@f5.com

Europe/Middle-East/Africa
emeainfo@f5.com

Japan
f5j-info@f5.com

©2015 F5 Networks, Inc. All rights reserved. F5, F5 Networks, and the F5 logo are trademarks of F5 Networks, Inc. in the U.S. and in certain other countries. Other F5 trademarks are identified at f5.com. Any other products, services, or company names referenced herein may be trademarks of their respective owners with no endorsement or affiliation, express or implied, claimed by F5. 0113