



Deploying F5 BIG-IP Virtual Editions in a Hyper-Converged Infrastructure



Justin Venezia
Senior Solution Architect

Paul Pindell
Senior Solution Architect



Contents

The Challenge	3
What is a hyper-converged infrastructure?	4
F5 and hyper-converged solutions	5
Virtual desktop infrastructure: An F5 example on a hyper-converged infrastructure	6
<hr/>	
Best Practices and Deployment Considerations	7
Sizing BIG-IP virtual editions	7
Disk space	8
Additional storage and disk considerations	9
Deployment of redundant BIG-IP virtual edition instances	10
Host placement and resources	10
Administrator-triggered and automatic VM migrations	11
Networking	11
Tuning and optimization	12
<hr/>	
Transitioning to physical BIG-IP appliances	12
<hr/>	
Conclusion	13



The Challenge

Today, IT departments are challenged in many ways to get new solutions, applications, and other IT services delivered and into production in a timely, business-friendly manner.

Constraints and challenges seen across IT organizations include:

- Budget limitations.
- Overworked IT employees.
- Specialized skill training.
- Storage, network, and computing solutions that involve multiple vendors.
- Legacy or siloed operational and support models.
- Complex architectural and implementation governance that “breaks the business.”

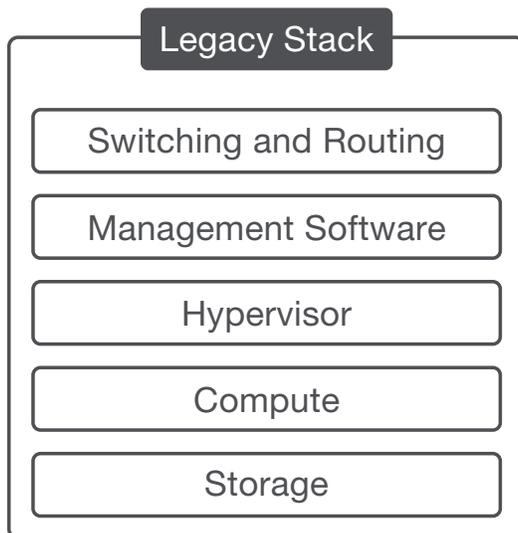


Figure 1: Legacy infrastructures often involve complexity and constraints.

Enterprises increasingly are turning to hyper-converged infrastructures to provide a modular, elastic, and manageable environment, eliminating the typical bottlenecks inherent to many IT infrastructures.



What is a hyper-converged infrastructure?

Simply put, a hyper-converged infrastructure integrates all of the major computing hardware — including storage, compute, and hypervisor — plus management software onto a single physical appliance. It's a one-stop shop with the ability to get up and running quickly.

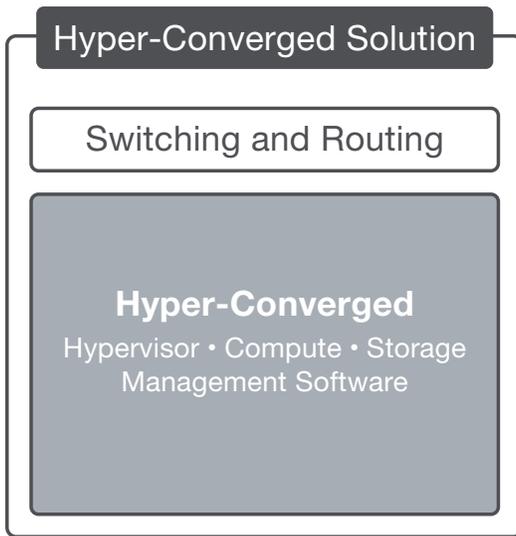


Figure 2: Hyper-converged infrastructures merge key components and management software into a single, streamlined architecture.

Additionally, a hyper-converged infrastructure is elastic and provides a method of calculating predictable scale, depending on the use case. To expand IT infrastructure capacity, all you need to do is acquire another “node” containing all the necessary components, plug it in, and you’re done. To top it off, all of the major components contained within a hyper-converged infrastructure are typically managed and orchestrated through a “single pane of glass” management interface. There is no reason to navigate through three different consoles or to contact the networking team, storage team, or virtualization team when you need to spin up a new server, host, or virtual machine. It’s all in one place.



F5 and hyper-converged solutions

With many companies taking a “virtualization first” approach to their IT services, hyper-converged solutions provide an opportunity to simplify how the IT staff manages and deploys infrastructure. Virtual machines (VMs), as well as storage and compute resources, benefit from the resiliency of hyper-converged IT platforms.

What’s missing from a hyper-converged infrastructure is a way to ensure applications running on the platform are scalable, highly available, intelligently monitored, and secure. Whether you need to federate application access between global data centers, provide secure remote access to business critical applications, or build a highly scalable and resilient application platform, F5® BIG-IP® products can help.

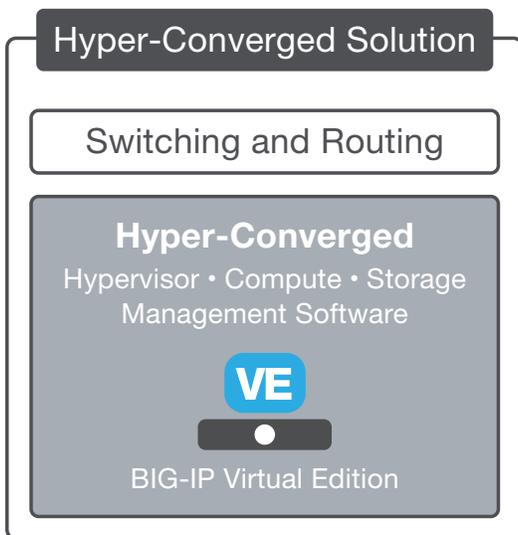


Figure 3: BIG-IP products can ensure the scalability, availability, and security of applications running on a hyper-converged infrastructure.

BIG-IP virtual editions work with the leading hyper-converged platforms to extend familiar F5 benefits — including locally and globally aware intelligent traffic management, acceleration, firewall services, and access management — to a hyper-converged IT environment.

Companies such as Nutanix, SimpliVity, Scale Computing, NIMBOX, and Pivot3 are some of the leading vendors of hyper-converged appliances. While many of these vendors serve a common goal—unifying management, storage, compute, and hypervisor—each vendor has individual strengths. Evaluate each to determine the right choice for your organization’s business and technical requirements. Regardless of your selection, you’ll need to also address L4-L7 services like those provided by F5 products.



Virtual desktop infrastructure (VDI): An F5 example on a hyper-converged infrastructure

It's quite simple to incorporate BIG-IP products into a hyper-converged platform. BIG-IP® virtual editions (VE) run as a virtual machine on one or more of the nodes in your hyper-converged infrastructure deployment. One or more BIG-IP virtual editions can be deployed for high availability (HA) or to segment various application services on different BIG-IP virtual edition instances if required.

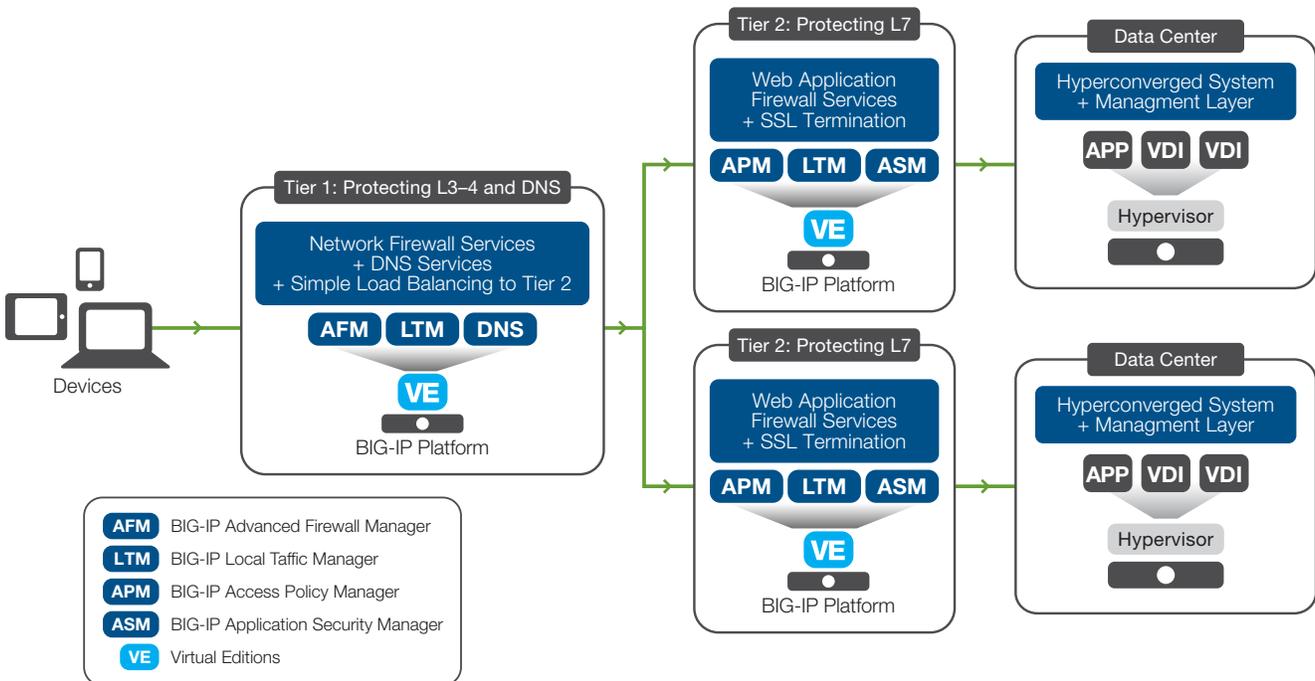


Figure 4: BIG-IP virtual editions can be deployed for high availability.

For example, you could run two BIG-IP virtual editions in each of two data centers — which is typical of an HA, active/passive deployment scenario. Each pair of BIG-IP instances can provide various application delivery services, including the functionality of BIG-IP products such as BIG-IP® Advanced Firewall Manager™ (AFM) or BIG-IP® Access Policy Manager® (APM). At Tier 1, the HA pair of BIG-IP virtual editions provides DNS protection services, as well as global traffic management and network firewall protection. At Tier 2, the BIG-IP virtual editions provide protection for the applications themselves, including application-level firewalls and additional intelligent traffic management.



Best Practices and Deployment Considerations

When BIG-IP virtual editions are deployed to hyper-converged platforms, many of the same principles and practices apply as for deployment to industry-leading hypervisors. Before deployment, take a look at the [hypervisor-specific deployment/installation guides](#) for BIG-IP virtual editions on f5.com. These guides provide additional information for any unusual or platform-specific configurations.

Sizing BIG-IP virtual editions

Selecting the correct amount of virtual CPU and memory for a BIG-IP virtual editions instance is important for optimal performance and scalability. The number of BIG-IP modules enabled on each BIG-IP virtual edition instance will also dictate the virtual machine's required CPU and memory configuration. If you need to change the size of the VM, all you have to do is add memory and/or compute resources.

While there are a few exceptions, typical sizing guidelines include:

- Ensure at least 2 GB of virtual RAM (vRAM) are allocated for each virtual CPU (vCPU).
- The recommended minimum size is 2 vCPU with 4 GB or less of vRAM. At this minimum, the BIG-IP instance cannot run more than two BIG-IP modules.
- A 4 vCPU platform with more than 4 GB but less than 8 GB of vRAM cannot run more than three BIG-IP modules.
- An 8 vCPU platform with more than 12 GB of vRAM (16 GB are recommended) can run all BIG-IP modules.

It's also important not to forget that the type of BIG-IP product license and the amount of licensed throughput will dictate the virtual CPU configuration. Find more information on f5.com in "[Overview of BIG-IP VE license and throughput limits.](#)"



Disk space

When you download a BIG-IP virtual edition instance, you can choose from three options. The size and capabilities of your BIG-IP instance can be determined by locating the following key words in the image template name:

- LTM_1SLOT. This option requires only about 7 to 8 GB of disk space. It's suitable for a system provisioned with BIG-IP LTM only. There is only one slot available and you cannot upgrade or apply a hotfix to the system by default.
- LTM. This option requires 31 GB of disk space. It's suitable for a system provisioned with BIG-IP LTM only. By default, there are two slots available for upgrade or hotfix.
- ALL. This option requires 124 GB of disk space. It's suitable for a system provisioned with multiple BIG-IP modules. Although there are two slots available for upgrade or hotfix by default, there should be sufficient space to install a third slot.

Option	Disk Space Required (GB)	Module Support	Slots
LTM_1SLOT	7-8	BIG-IP LTM only	1
LTM	31	BIG-IP LTM only	2
ALL	124	All BIG-IP modules	2-3

Choose the image that meets your needs: BIG-IP LTM only or multiple modules. You can add additional space to the BIG-IP instance if you choose additional modules, but there's always a risk when making this type of change; see the discussion of thin or thick provisioning below.



Additional storage and disk considerations

Although the BIG-IP system is primarily network and CPU dependent, both the input/output operations per second (IOPS) requirements of the virtual machine's storage and the type of disk configured (thin versus thick) can vary and should not be ignored. In a case of storage contention, BIG-IP system performance could be affected. General guidelines for sufficient storage include:

- **IOPS.** IOPS requirements may vary based on which BIG-IP modules are enabled. For example, BIG-IP® Application Acceleration Module® (AAM) uses disk capabilities on the BIG-IP instance to assist with traffic acceleration. Therefore, the IOPS requirements for this combination will be higher than for BIG-IP LTM alone. To ensure adequate storage performance, F5 recommends monitoring the storage IOPS performance of the BIG-IP virtual edition instance using the licensed modules that will be used by your organization.
- **Thin versus thick disks.** There have been several studies and debates on the use of thin or thick disks — when to use or not use them, which performs better, what types of workloads benefit or suffer based on the disk type, the speed and type of disk array (SSD versus SCSI), oversubscription of thin-provisioned data stores, and so on. In some cases, thin provisioning may be your only option. For instance, using a network file system (NFS) data store on VMware vSphere allows only for thin provisioning.

For BIG-IP virtual editions, F5 recommends thick provisioning to ensure the BIG-IP instance has ample disk capacity. This mitigates the risk of the BIG-IP system running out of disk space. If you need to use thin-provisioned storage, make sure there are controls in place to avoid running out of disk space.

Regardless of the thin or thick virtual disk choice, don't ignore the need to ensure you have enough storage IOPS to support the BIG-IP instance.



Deployment of redundant BIG-IP virtual edition instances

The best way to get redundancy with BIG-IP application delivery services is to have two instances of the BIG-IP virtual edition deployed. A typical configuration would make one BIG-IP instance active while the other is in standby mode. In the event a host or BIG-IP instance fails, the passive node will immediately take over the work of the active node with minimal or no interruption in service.

While hypervisors like ESXi and others offer a high availability option to power virtual machines on other hosts in the event of a host failure, this operation takes time. Therefore, your BIG-IP instance will be down while it powers itself back on (which could be several minutes). Redundant deployment is a more reliable option.

If you do deploy more than one virtual machine, ensure that controls are in place to prevent a pair of BIG-IP virtual edition instances from running on the same host. Many hypervisors offer methods to prevent this from happening.

Host placement and resources

One of the benefits of virtualization is the effective management and sharing of compute and memory resources on a host or server. The BIG-IP system depends on ample CPU and memory resources to efficiently process network traffic and application access requests. Any significant resource contention with other VMs may reduce traffic passing through BIG-IP virtual editions.

For this reason, F5 recommends that the BIG-IP instance be placed on a host with ample compute and storage resources. Placing a BIG-IP instance on a busy host may reduce the performance of either the BIG-IP system or the other applications collocated on the same hyper-converged appliance.

Similarly, for the BIG-IP system to have ample CPU and memory resources, administrators should ensure proper reservations are configured for the virtual machine. Oversubscription of memory and CPU for a BIG-IP instance is strongly discouraged. Depending on the hypervisor, these settings may need to be manually configured or may be automatically configured and adjusted when the VM is deployed.

The general guidelines for resource reservation include:

- Reserve 100 percent of memory and CPU capacity for the BIG-IP instance.
- For each virtual CPU, reserve a single physical core.

RECOMMENDED PRACTICES

Deploying F5 BIG-IP Virtual Editions in a Hyper-Converged Infrastructure



For example, if the processor speed of a server is 3 GHz and there are two virtual CPUs/4 GB of virtual RAM allocated to the virtual machine, you will need to reserve 6000 MHz of CPU and 4 GB of RAM.

Administrator-triggered and automatic VM migrations

Another benefit of some hypervisors and a hyper-converged infrastructure is the ability to automatically and manually relocate virtual machines to other hosts or storage. While this feature works well for many workloads, it's not advisable for BIG-IP virtual editions. Moving an active workload that passes lots of network traffic or demands a lot of CPU will cause an interruption in service and potentially drop any connections coming through the BIG-IP virtual edition instance.

It's recommended that you also ensure controls are in place to prevent BIG-IP virtual editions from being automatically relocated to other hosts or storage media. If you have to move a BIG-IP instance to another host, it's recommended you move it when it's in an idle or stand-by state.

Networking

There are several important things to consider with networking and hyper-converged systems. Many of the vendors in this space provide two high-capacity (10 GB), high-speed network interfaces that are used for all network communications. These interfaces also support many different networking segmentation and configuration standards — such as VXLANs, VLANs, port trunking, and virtual switches — used by leading hypervisors.

Here are some simple configuration best practices to keep in mind when designing and deploying BIG-IP virtual editions on hyper-converged platforms:

- Make sure you have the minimum number of NICs required by the specific hypervisor you are deploying to.
- Remember that you'll need a minimum of two networks to get BIG-IP virtual editions up and running — one interface and network for management and the other for internal network traffic. The management network must use a different network or subnet and dedicated interface than the internal network used by the BIG-IP instance for client/server traffic.
- If using existing network links, make sure they are not oversubscribed and have ample bandwidth to handle the required network traffic volume.

RECOMMENDED PRACTICES

Deploying F5 BIG-IP Virtual Editions in a Hyper-Converged Infrastructure



- Review the hypervisor-specific deployment guide for BIG-IP virtual editions in network configurations that affect the handling of large frames, TCP offload, VMQ and NUMA spanning (for Hyper-V), and other configurations that may be needed to provide optimal performance.
- Use a dedicated network for BIG-IP HA services. It's important to keep network traffic that needs to be highly available segmented so production traffic or a busy internal network connection does not affect the device (or instance) heartbeat used to detect a failover. In addition, ensure the network connection or interface used for HA configurations is not oversubscribed.

Tuning and optimization

Monitoring the health and performance of the system is critical to successfully implementing BIG-IP virtual editions in your hyper-converged architecture. Critical elements such as CPU utilization, memory usage, and network performance should be constantly monitored for optimal performance and early problem detection. Also, some BIG-IP modules require more storage input/output (I/O) than others; regardless, storage I/O performance should not be ignored.

Administrators should pay special attention to performance characteristics and statistics of the BIG-IP virtual edition instance shortly after deployment and after adding additional modules or services (such as adding new virtual server or PCoIP proxy functionality) to ensure the capacity of the BIG-IP instance is not exceeded. It may be necessary to add additional memory, CPU, or storage, or consider separating services onto another BIG-IP instance to ensure optimal performance.

Transitioning to physical BIG-IP appliances

There may come a point where the benefits of a physical BIG-IP appliance outweigh the benefits of virtual editions. The number of access sessions, the volume of SSL transactions per second (TPS), and throughput are three typical elements that should receive attention. If use is encroaching on the upper end of the tested and published limits of BIG-IP virtual editions, F5 recommends considering a transition to a physical appliance to ensure acceptable performance and scalability of the ADC.

RECOMMENDED PRACTICES

Deploying F5 BIG-IP Virtual Editions in a Hyper-Converged Infrastructure

Conclusion

Hyper-converged platforms allow organizations to reduce some of the common inefficiencies and complexities of today's typical IT infrastructure operations and architecture. Utilizing BIG-IP virtual editions with a hyper-converged infrastructure brings end-to-end resiliency, elastic scale, and security to business-critical applications that run on hyper-converged appliances. Proper implementation and alignment with best practices for sizing, placement, resource allocation, and configuration are essential to ensure applications that receive services from BIG-IP products perform at their best, with no interruptions.

To learn more about BIG-IP products in converged and hyper-converged infrastructures (also known as integrated systems), please visit f5.com.

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

Americas
info@f5.com

Asia-Pacific
apacinfo@f5.com

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

Japan K.K.
f5j-info@f5.com

Solutions for
an application world.

