



# Load Balancing Your Applications

Let your applications help decide where to send your clients with the most intelligent and programmable application delivery controllers available.



## KEY BENEFITS

- Implement industry-defining technologies for best-in-class application performance
- Get cloud-agnostic performance solutions that provide common infrastructure-as-code configurations no matter what platform you rely on
- Choose your weapon with F5 and NGINX offering more functionality in more application deployment models than any other cloud-native or third-party solution provider
- Leverage F5's large developer community to improve on your BIG-IP solutions or use the world's open source communities to improve your NGINX deployments and performance services

A SINGLE API ON A CORE APPLICATION SERVICE COULD END UP SLOWING QUERIES TO YOUR CLIENTS ... UNFORTUNATELY, YOU'LL ONLY FIND OUT WHEN YOUR SUPPORT TEAM STARTS GETTING CALLS.

**Increasing complexity and distribution of application functionality**—from cloud native web services to colocated microservices—are creating new ways for customers to define uptime, or worse, downtime.

For many applications, basic load balancing distribution methods like round robin or least connection solve a basic need to distribute traffic to multiple web or application servers. They prevent any one application from causing connectivity problems. Solving simple issues like traffic load or determining when a resource is available can work well—until applications outgrow them. This means that moving beyond basic load balancing features provides your DevOps and NetOps teams with the ability to reexamine what determines uptime and to plan for your application's future. After all, your best application can't retain customers if it can't scale intelligently to meet their demands.

The increased complexity of moving to automated deployments in multi-cloud and distributed service architectures pushes us past those basic load balancing distribution methods. Inter-service communication performance, interactive responses to active monitors, and side-band decision-making all become real requirements to ensure that all pieces of your application puzzle respond in a timely manner. A single API on a core application service could end up slowing queries to your clients without tripping any monitors from your traditional solutions. Your web page may render, but the secured login script might be loading an old version on some of your application servers, preventing some of your clients from successfully logging in. Unfortunately, you'll only find out when your support team starts getting calls.

At the same time, how customers and enterprises define success is evolving with how their applications perform at scale. One customer may be an end user consuming standard web pages through a browser. Another might be a service provider importing internal user data through one or more API endpoints for an HRIS integration. Internally, corporate applications may be clients and consumers of each other, with each layer of functionality having its own performance expectations and service level agreements before the service considers its resources unavailable.

Consumers demand more, and a recent survey by AppDynamics showed that 78% want compensation for poor digital experiences. Further polling showed that "49% switching suppliers due to poor digital experiences, up from 37% in 2017."<sup>1</sup> Unreliable or poorly performing applications impact revenue in more ways than just lost loyalty.

## As Companies Scale, So Does Their Need for Load Balancing

TREAT YOUR LOAD BALANCING SOLUTIONS AS YOUR APPLICATION TRAFFIC COP, DIRECTING THE RIGHT CONNECTIONS TO THE RIGHT AVAILABLE SYSTEMS.

Organizations often rely on load balancing features as a primary service for incoming connections from external users. Successfully scaled applications, however, treat each service within the application environment as not just a potential bottleneck for problems, but as an opportunity to place load balancing and other redundancy solutions in place. This creates a mesh of fault tolerance, no matter what platform your applications reside in. Container services, native or multi-cloud, private, or colocated data centers all benefit from strategically implemented load balancing and additional Application Delivery Controller services, designed to keep applications running fast, scalable, and reliable.

To understand how those services operate and distribute traffic to available infrastructure requires advanced, intelligent load balancing. Interactive monitoring allows load balancing services to understand your applications' thresholds for performance before sending clients to those resources. In turn, by leveraging BIG-IP's full proxy architecture or NGINX's reverse proxy functionality, you can protect your services from misguided side-band connections that could hurt your performance. Treat your load balancing solutions as your application traffic cop, directing the right connections to the right available systems.

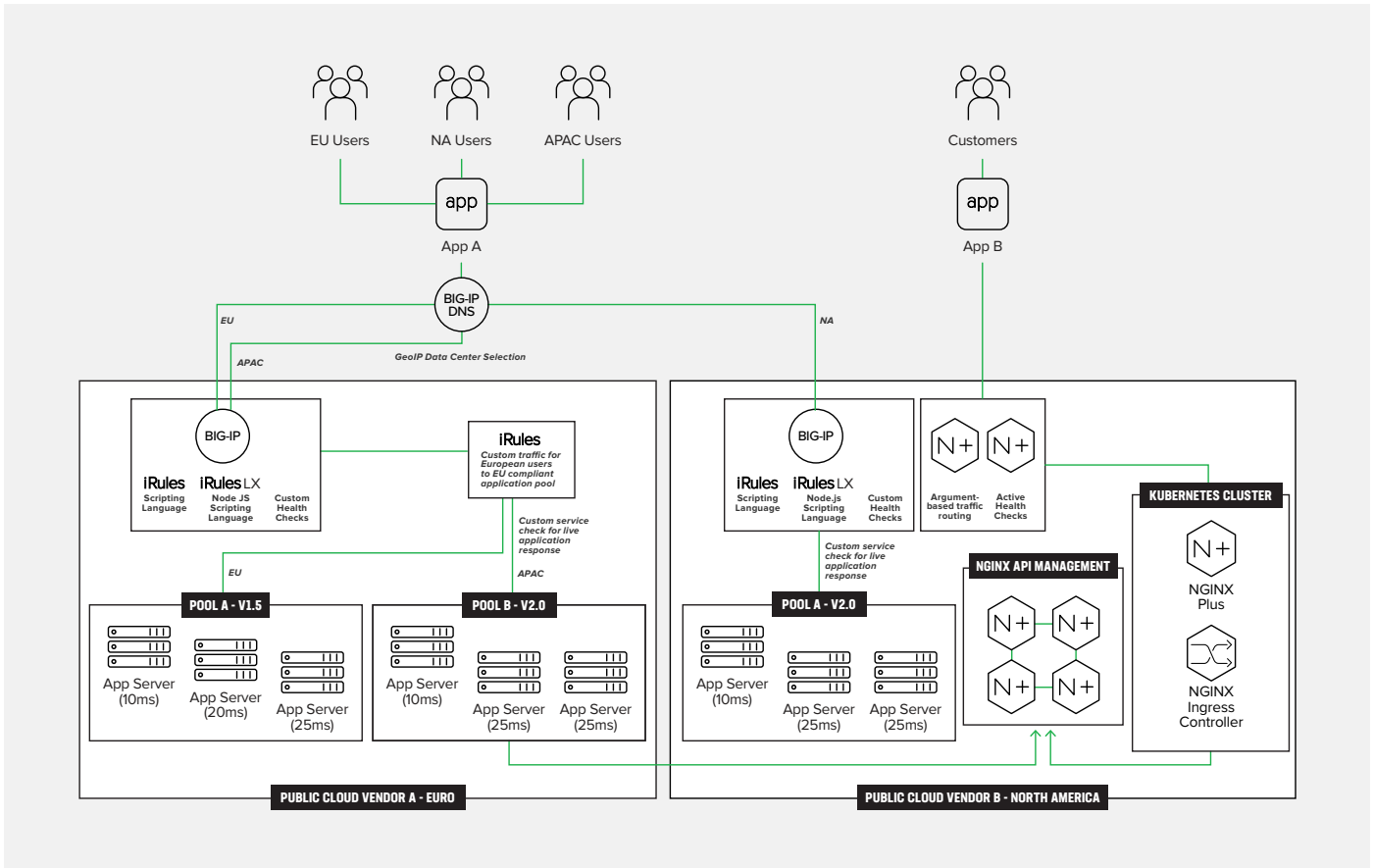
Performance (55%) and availability (40%) topped consumers' "frustration" points that could easily be mitigated by distributing traffic to increase application reliability and uptime<sup>2</sup>.

## The Architectural Components: How F5 Approaches Load Balancing

[F5 BIG-IP Local Traffic Manager \(LTM\)](#) includes static and dynamic load balancing to eliminate single points of failure. Application proxies give you protocol awareness to control traffic for your most important applications. BIG-IP LTM also tracks the dynamic performance levels of systems in a group, ensuring that your applications are not just always on, but also are easier to scale and manage.

[F5 BIG-IP DNS](#) takes load balancing across applications and applies it globally, ensuring that your applications are on and responding to your customer's needs. F5 BIG-IP DNS distributes DNS and user application requests based on business policies, data center and cloud service conditions, user location, and application performance. The BIG-IP platform delivers F5's high-performance DNS services with visibility, reporting, and analysis. It hyper-scales and secures DNS responses geographically to survive DDoS attacks. It delivers a real-time DNSSEC solution. And it ensures high availability of global applications in all cloud environments.

NGINX Plus builds on the popular NGINX open source solutions, combining load balancing, web server, and content caching solutions. NGINX Plus has exclusive enterprise grade features beyond what's available in the open source offering, including session persistence, configuration via API, and active health checks. Deliver easy-to-implement traditional load balancing features, handle more traffic with resource optimized web server features, and use the same caching technology that powers the world's largest CDN. Simplify your architecture while reducing costs with the only all-in-one load balancer, API gateway, sidecar proxy, content cache, and web server.



**Figure 1:** F5 and NGINX can be applied across your application infrastructure to provide intelligent, secure, and scalable load balancing services anywhere you need reliable performance.

The F5 iRules scripting language—F5's traffic scripting interface—enables programmatic analysis, manipulation, and detection of all aspects of the traffic in your networks. Customers routinely implement security mitigation rules, support new protocols, and fix application-related errors in real time. With robust and flexible iRules, you can easily and rapidly develop solutions that you can confidently deploy across multiple applications.

## KEY FEATURES

### Industry-Leading Load Balancing

Define protocol or application layer routing decisions. Or take the programmatic approach and define load balancing solutions tailored to your applications' responsiveness and architecture.

### SSL Offload

Implement SSL termination/offload, which provides access to encrypted application traffic to make intelligent load balancing decisions.

### Programmatic Traffic Decision Making

Take the visibility and control BIG-IP provides and immediately act on it using iRules, F5's event-driven scripting language. Or leverage NGINX Plus's robust JavaScript module for scripting and advanced configurations.

### API Management and Rate Limiting

Define and publish APIs quickly while delivering high-performance load-balanced processing for north-south and east-west traffic. Protect APIs from volumetric attacks so your services can stay responsive to customer and partner integrations.

F5 makes it easy to integrate app services directly into your container management solution, regardless of where it's hosted, allowing the security and network teams to keep application delivery services up to date without slowing down development.

The intelligence of BIG-IP can be deployed and updated in coordination with container events. F5 provides a free container integration solution—the [Container Ingress Services \(CIS\)](#) component, which runs as a container and subscribes to cluster management events. CIS combines event awareness with native container platform tools like config maps and annotations to configure Ingress services via the BIG-IP REST API.

Or add NGINX to your Kubernetes Container Solution with [NGINX Ingress Controller for Kubernetes](#). This solution provides enterprise-grade delivery services for Kubernetes applications, with benefits for users of both NGINX Open Source and NGINX Plus. With the NGINX Ingress Controller for Kubernetes, you get basic load balancing, SSL/TLS termination, support for URI rewrites, and upstream SSL/TLS encryption. NGINX Plus users also get session persistence for stateful applications and JSON Web Token (JWT) authentication for APIs.

## Conclusion

F5 wrote the book on load balancing and we're still writing. There's a reason 48 of the Fortune 50 use F5 in their infrastructure. F5 and NGINX provide the most scalable and reliable load balancing solutions for your application plans. It doesn't matter if you're managing a top-performing enterprise solution or a scrappy application on the rise. Make F5 part of your application lifecycle and we can grow with you.

To learn more, contact your [F5 representative](#), or visit [F5.com](#).

<sup>1</sup>The App Attention Index 2019, found at <https://www.appdynamics.com/blog/news/app-attention-index-2019/>

<sup>2</sup>Ibid.

