

2021

# The Load Balancing Buyer's Guide

## Overview

Digital transformation means that businesses are relying on their IT networks and applications more to connect with their employees and customers. These businesses need to optimize the Application Experience [AX] through application availability, agility, and performance.

Load balancing technology is a key piece of the IT architecture to create the ideal AX topology. Load balancers ensure that applications are always available through robust application pooling and advanced health checking. Finally, load balancers improve application performance with acceleration including encryption/decryption, and content offload technologies that reduce server loads.

Load balancing is a technology that intersects the world of traditional IT networking with modern application development. As such, it can be hard to navigate the requirements and expectations for load balancing technology for any given business' application.

## How to use this buyer's guide

This step by step guide is designed to help navigate users through the process of understanding their application requirements, identifying their load balancing needs to meet these requirements, and selecting the right set of features to optimize the AX.

Using the checklist and additional resource recommendations, this guide is a comprehensive document to assist any organization of any size, select the right load balancer to fit their requirements.

## Step 1: Identify the business goals of the application

The type of load balancer that is required depends on how the application behaves and what the business requirements are for the application's use. It is important to understand how the application will be used within the business before selecting the required load balancing requirements.

Load balancing solutions should enable organizations to cost-effectively scale their operations, while ensuring high availability. This creates an outstanding AX and therefore customer experience (CX). According to Forrester Research, 91% of enterprises say speed and agility in load balancing is critical. Since the load balancer is the proxy for the application, its speed and agility translate to application speed and agility.

Remote working and the need to support 24x7 access to business critical applications and data has been the key topic for most IT teams in 2020. ZK Research has shown that half of all organizations receive insufficient notification and assistance with mission-critical application issues causing unnecessary downtime. Moving forward, the lessons learned about downtime, application lock out and security threats along with slashed budgets have heightened the need for a load balancing solution that is secure, quick to deploy, easy to manage and at a value point that fits future IT budgets.

### Answer the following questions to help identify your needs.

Identify the criteria that are important each application in question.

1. Do you require an application that is continuously available? What is the acceptable length of time that an application is down due to faults or maintenance?
2. Is site redundancy required? For what reasons?
  - a. Disaster recovery
  - b. Active-active site availability
  - c. Geo-location for optimized regional content
3. Where will your application reside?
  - a. On site data center
  - b. Private cloud (virtualized server infrastructure)
  - c. Public cloud (AWS, Azure)
4. Does the application usage vary over time?
  - a. Steady usage
  - b. Continuously increasing
  - c. Spikes at certain events (holidays, start of school, etc.)
5. Are changes made to the application often?
  - a. Changes made every month or less
  - b. Changes made between one and 6 months
  - c. Changes made every 6 months or greater
6. Do you need to support an identical test and or developer environment?
7. How is the application delivered? Is it HTTP/web-based? If not, what protocol is used?
8. Is the application encrypted? If so, what are the encryption standard requirements?
9. Do you have service level agreements (SLA) in place for your minimum AX?
  - a. What is the acceptable downtime and mean time to resolution (MTTR) for any given problem?
  - b. How is it determined when an application is running 'too slow'?
10. Is your application and its data secure? Are there any security compliance requirements?

## Step 2: Understand the features and capabilities you require from a load balancing solution.

The type of load balancer that is required depends on how the application behaves and what the business requirements are for the application's use. It is important to understand how the application will be used within the business before selecting the required load balancing requirements.

### Infrastructure

Where you put the application and how the infrastructure is supported can influence the type of load balancer that you require.

- Available as a hardware appliance for private data centers
- Available as software for virtual hypervisor environments
  - VMWare
  - KVM
  - Xen
  - Hyper-V
  - Other
- Available as a bare metal image
- Available in the public cloud marketplace
  - AWS
  - Azure
  - Other
- Integration with container orchestration such as Kubernetes or Ansible

### Health Checking

Health checks are performed by the load balancer to ensure that the application is working and responding as expected. Moving up the OSI stack means more advanced health checks that provide more feedback. This intelligence usually requires more advanced load balancing technology.

- Layer 3 ICMP health checks (ping ECHO/ECHO reply)
- Layer 4 TCP handshake or UDP port availability

- Layer 7 HTTP status code (i.e., 200 OK or 503 Server Unavailable)
- Layer 7 HTTP custom content (look for specific ASCII response such as 'Down for Maintenance')
- Layer 7 for any TCP/UDP port

### Availability & Scalability

Load balancers need to be resilient enough to manage application connections even when they fail themselves. They can also offer application scalability by providing more load balancing resources as application demand increases.

- Active/Standby high availability between pairs of load balancers
- Stateful connection failover to prevent the loss of existing active application connections
- Clustering beyond 1+1 availability to deliver high performance scalability
- Disaster recovery (DR) active/standby ability for multiple sites
- Active/Active site availability through dynamic global server load balancing (GSLB)

### Session Persistence

Sometimes the connections need to be maintained to a specific server because of local content associated with the session. It is also possible that different applications need access to the same local session specific information. Session persistence solves this problem. Different parameters can be used to maintain session persistence.

- Client IP address
- TLS (SSL) SessionID
- Web/HTTP cookie and/or session ID
- Port Following for multiple applications (i.e., webstore and checkout sites)

### Scheduling Methods

Load balancers distribute traffic across pools of servers using various algorithms. The best algorithm to use depends on the behavior of the application. Round robin is good for connections of a predictable fixed length while least connections is good for variable length connections. Other methods provide benefits for different server capabilities (weighted) or when connections need to be distributed based on client IP address (hash).

- Round Robin
- Weighted Round Robin
- Least Connection
- Weighted Least Connection
- Source-IP Hash

### Encryption Features

When applications use encryption, they are usually relying on SSL, now known as TLS. The more current versions are more secure with better and more reliable encryption methods. The current version of TLS is 1.3. Some applications require older version support for compatibility reasons. Certificates are used to manage the encryption keys and are generally unique for each application instance. Load balancers can manage the certificates and encryption to offload the heavy performance requirements from the application servers.

- Configurable TLS (1.0, 1.1, 1.2, 1.3) and SSL (2.0, 3.0)
- Support for RSA and elliptical curve encryption (ECC) algorithm standards
- OCSP certificate validation
- Server Name Identification (SNI) support
- Support for up to 1,000 TLS (SSL) certificates
- Automated TLS (SSL) certificate chaining
- Certificate Signing Request (CSR) generation
- STARTTLS mail protocols (POP3, SMTP, IMAP)
- FIPS 140-2 encryption certification

### Management

The IT operational team needs to support the load balancing solution. They will need to make sure that the

load balancer has an easy to use and secure management interface. The load balancer also needs to integrate into other operational management tools.

- Web User Interface (WUI)
- RESTful and PowerShell APIs
- Integration with VMware vRealize Orchestrator
- Context based help
- Analytics providing real time display of performance and availability
- Application templates for standardized configurations
- Configuration backup
- Comprehensive logging and reporting
- SNMPv2 and SNMPv3 support
- Authenticated NTP
- Diagnostic shell with in-line packet tracing

### Security

Inherently, the load balancer provides application and network security. As a reverse proxy, it only allows connections to specific IP addresses on designated TCP/UDP ports similar to a traditional network firewall. Other security capabilities can enhance the application and IT infrastructure security profile. The load balancer can also provide enhanced application authentication for more robust application security.

- Permit/Deny extended Access Control Lists (ACL)
- IPsec tunnel support
- Attack protection for SYN flood and other layer 4 and layer 7 attacks
- Microsoft TMG replacement
- Application pre-authentication
- Multi-Domain authentication & single sign on (SSO)
- Custom login forms
- Two factor authentication (2FA)
- SAML, Active Directory, RADIUS & LDAP authentication

**Web Application Firewall (WAF)**

As a reverse proxy, load balancers have the opportunity to provide layer 7 content protection, especially for HTTP/S and web content. The WAF can implement policies to protect websites against common and specific threats.

- Real-time application threat mitigation
- Daily rule updates
- OWASP Top 10 vulnerability protection
- PCI-DSS support

**Global Server Load Balancing (GSLB)**

GSLB offers site availability through active DNS monitoring. There are differences in how GSLB services work and it is important to obtain solutions that meet your application requirements.

- Disaster recovery (DR) failover for backup sites
- Active/active configuration for site availability
- IP reputation to block malicious and unwanted clients
- Geo-location for site direction based on physical location and/or distance to site
- Customizable IP mapping of clients to primary sites
- Advanced checking of server and data centre/cloud health and status

Creating a clear understanding of the features required will allow you to pick the right solution that won't under perform or offer additional non-useful features to maintain.

## Step 3: Licensing Options

What you want to pay versus what the CFO wants to pay for a load balancing solution may differ. Using step 1 and step 2 before looking at pricing will provide much needed arguments to get the purchase complete. There are two key items that you need to identify when determining the whether the price you find is acceptable for your project's requirements.

First, you need to determine the method that you will consume and pay for the load balancing services. There are multiple ways to obtain the solution you need.

### Perpetual license

First, there is the traditional 'Perpetual License'. This is a license to purchase the solution outright, whether it is a piece of hardware or an instance of load balancing software. The license is usually based on the maximum network throughput the product can support. The perpetual license does not include support services nor software updates. The support services can also include access to additional functionality.

### Subscription service

Second, one may consider a subscription service. The business gets access to the product and support for a payment on a regular schedule, usually on an annual basis. The subscription provides full access to the solution and features without the initial upfront payment required with the purchase of a perpetual license

### Pooled licensing

Third, there is often a pooled option where a company can purchase a certain amount of capacity and then distribute it throughout their environment based on their specific needs. This is a beneficial model where there is a need for multiple load balancers and the needs can vary based on application and time. For example, a company

my acquire a 100Gbps pool and they can then distribute that into individual instances as long as the total capacity stays within 100Gbps. Capacity can be distributed and recovered as long as the pool license is valid. This is usually paid annually and includes support services.

### Metered licensing

Finally, metered licensing is an innovative model. Metered licensing offers a license capacity similar to pooled licensing, but the individual instances of the load balancer that are deployed only go against the capacity for the actual amount of application traffic being load balanced by the system.

A business can acquire a 10Gbps metered license that allows them to load balance up to 10Gbps across any number of load balancing instances. Metered licensing includes support services and is usually billed monthly based on the past month's consumption.

As you can see, there is a difference in how one pays for these options. Some will be associated with CapEx spending while others will be OpEx. It is important to find the model that is best for your environment and for the payment model for your overall business.

## Step 4: Price vs Value

Not all load balancers are priced the same and different vendors charge different amounts for certain functions and capacities. During the pricing process, it is important to stay focused on the project's requirements. An application delivery controller (ADC) is essentially a load balancer with extra capabilities. Not all load balancing projects require all of these functions. Prioritize the features that your application project needs to be successful. Do not be distracted by the shiny lights or fancy features that you probably do not need. When doing the research, compare different vendors' products and make sure you are making as equal a comparison as possible.

Looking at the top load balancing vendors; here is a price/feature comparison for a sample of their offerings.

| Feature Comparison             | Kemp<br>LM-X15 | F5<br>i2800 | Citrix<br>MPX-8905 | Kemp<br>VLM-3000 | F5<br>VE-3G | Citrix<br>VPX-3000 |
|--------------------------------|----------------|-------------|--------------------|------------------|-------------|--------------------|
| USD (MSRP)                     | 9,800          | 27,900      | 44,000             | 5,600            | 44,454      | 43,920             |
| Throughput (Gbps)              | 15             | 10          | 5                  | 3                | 3           | 3                  |
| SSL TPS (2K Keys)              | 12,000         | 4,300       | 8,000              | 4,000            | N/A         | 500                |
| Concurrent Connections (L4)    | 35M            | 14M         | N/A                | >3M              | N/A         | 1.2M               |
| Web Application Firewall (WAF) | ✓              | Add-On      | Add-On             | ✓                | ✓           | ✓                  |
| FIPS-140-2 Level 1             | ✓              | Add-On      | X                  | ✓                | Add-On      | X                  |
| Caching & Compression          | ✓              | ✓           | Add-On             | ✓                | ✓           | ✓                  |



## Step 5: Research more

Gathering this information and the requirements for your load balancing project is just the first step in the process to identify and select the best solution for your needs. Each project and application have different needs based on the business goals. It is important to understand and select the features that you need while avoiding the unnecessary extras that you may encounter.

This buyers guide provides the basics to allow you start the selection and acquisition journey. To validate your choices, you need to compare your choices with those from third party and customer review sites such as Gartner Peer Insights that offer insights on the ADC/load balancer market. Look for customers and projects that have a similar goal and/or are in the same business vertical as yourself.

The selection of any technology is not an easy process. It is essential that you ensure that the chosen solution meets the needs of your business and project. This needs to be done with an understanding that the cost, complexity, and support of the technology does not become a barrier to the effective use of the solution. We hope this buyer's guide helps your journey to selecting the optimal load balancing solution to optimize the AX for your application(s) and IT organization.

## Glossary of Useful Research

- [Gartner Peer Insights](#): Reviews for Application Delivery Controllers (ADCs) Market
- Load Balancer Trial : Try a Kemp load balancer free of charge for your set up
- [Avoiding Downtime and user productivity loss by Broadband Testing](#)
- [ADC Market Survey by ZK Research](#)
- [Per Application Delivery Drives Multi-Cloud Migration by Doyle Research](#)
- [Buyers Guide to load balancing in the Cloud](#) – Video by Azure MVP Nic Blank