Application delivery in uncertain times

Evolving your approach to load balancing critical systems

loadbalancer.org

"WORST CASE SCENARIOS REALLY DO HAPPEN."

COVID-19 has taught the IT industry a valuable lesson: worst case scenarios really do happen.

" When the coronavirus emerged in early 2020, it not only seriously disrupted business operations, across almost every industry sector; it also brought about a seismic shift in application usage. Employees began working from home, dependent on collaboration apps and remote access services to continue to fulfil their roles. At the same time, consumers increased their online shopping, began streaming more online entertainment services and turned to video to maintain contact with family and friends. Application usage soared, as did users' expectations for 24/7 services, yet IT teams had little or no access to their data centers.

The organizations that coped best in these extraordinary circumstances were those that had prepared for a worst case scenario, expecting it never to happen. They had load balancers in place, capable of managing the unprecedented surge in application traffic and maintaining uninterrupted services. Throughout national and local lockdowns, load balancers worked behind the scenes to keep vital collaboration tools, e-commerce platforms and communications services available for unprecedented numbers of users.

As we learn to live in new uncertain times, we cannot let our guard down. Disasters of all kinds can happen at any time. Next time, it might not be a disease, but a fire at a data center, a terrorist alert near your head office, a cyber-attack or a surge in customer interactions due to a breakdown in your supply chain. It's not what the disaster is that matters, but how ready you are. Recent events have taught organizations that they need simpler load balancing solutions that they can set up quickly, manage remotely and trust to deliver exceptional reliability for mission-critical applications in any situation. As everyone now knows, worst case scenarios can quickly become reality and we need to be ready for whatever happens next. "

Malcolm Turnbull, Co-founder, Loadbalancer.org

UNCERTAIN TIMES

The importance of effective application delivery

COVID-19 focused attention firmly on the importance of effective application delivery. During the pandemic, organizations of all sizes needed to empower their employees to work from home, with just a few days' notice, and provide large numbers of users with reliable remote access to mission-critical systems. The resulting sudden increases in application usage placed a strain on many IT infrastructures, with users reporting slow application response times and downtime. A Gartner survey of the financial services industry, published in August 2020, found that over half of business leaders considered technology infrastructure to be a weakness during the crisis.[1]

For some organizations, COVID-19 created a sudden business opportunity with high earnings potential. Ecommerce and digital services providers saw a huge uplift in demand for apps and web sites, often with many thousands of new subscribers and online customers week after week.

The video conferencing platform Zoom, for example, grew its customer base of companies with 10 or more employees by 458% at the height of the COVID-19 crisis, according to its published financial results. Zoom, and other organizations like it, needed to handle an exceptional and sustained peak in traffic and deliver consistently good application performance for previously-unimaginable numbers of concurrent users.

Right from the beginning of the pandemic, businesses were scrambling to deliver a host of new apps to enable them to continue to operate in the 'new normal'. Hospitality chains created apps to capture the details of diners and support national track and trace efforts. Other organizations introduced new apps and web services to replace face-to-face customer engagements and protect employees while continuing to deliver services to customers and achieve targets. All these solutions needed to be delivered quickly to address immediate challenges and be capable of scaling up to meet undefined demand.

Clinicians suddenly required remote access to office-based solutions to access and analyze critical patient data.

The COVID-19 crisis is far from over and the global ramifications will be both extensive and long in duration. Businesses must recover from the financial losses of lockdown, find new ways to survive in a recession and address challenges stemming not just from the pandemic but from other significant world events. According to Gartner's latest Emerging Risks Monitor, senior executives are now most concerned about further coronavirus outbreaks[2], but other issues including political uncertainty, Brexit and international trade relations are all contributing to a climate of profound uncertainty.

The emergence of new application delivery priorities

The critical role of the load balancer in these uncertain times couldn't be clearer. Supplied as hardware appliances, software-only or cloud-based solutions, load balancers balance application and network traffic between all available servers in a network, to ensure all users experience the same, consistently fast performance. If one server should fail, the load balancer will instantly and automatically reroute the traffic to an alternate server in the cluster, so that there is no interruption in the availability of the app or web service.

For a lot of applications and web sites, load balancers are essential; for others they are strongly recommended.

Load balancers are a fundamental component in the IT infrastructures of medium to large organizations and have been for many years. For a lot of applications and web sites, load balancers are essential; for others they are strongly recommended. Yet, the challenges that businesses have faced during COVID-19 – together with the ongoing uncertainties of the pandemic – have put load balancers at the forefront of CIOs' and IT directors' minds. If there was previously any doubt, IT professionals now understand the importance of load balancers and are investigating the best way to use them to help them address new application delivery priorities.

GUARANTEE ZERO DOWNTIME
High availability has always been important, but with the increased reliance on home working, digital entertainment and e-services, zero downtime is now an imperative.

SIMPLIFY LOAD BALANCER MANAGEMENT
During lockdown, IT managers don't have easy access to their data centers to maintain their equipment, so now realize the value of having IT infrastructure that is reliable, resilient to disasters and easy to manage.

DEPLOY NEW APPS RAPIDLY
In an uncertain and ever-changing environment, organizations need to be able to roll out new apps quickly to keep employees productive and capitalize on evolving market opportunities.

HAVE UNRESTRAINED SCALABILITY
All organizations need to re-evaluate the load balancing provision they have for their existing applications and think carefully about the requirements for new apps, to make sure they have the capacity for exceptional and unexpected peaks in traffic.

MULTI-APPLICATION LOAD BALANCING

Customizable, but complex

There are two primary approaches to using load balancers. The first is to use a single load balancer product or platform to balance the traffic for all applications within an organization. This method, called multi-application load balancing, gives IT teams just one load balancer technology to learn and support. The large load balancer solutions that are typically used for multi-application load balancing often come with an array of supplementary features, such as added security, and can be heavily customized to handle the traffic for different applications in different ways.

The more apps that organizations try to balance on a single load balancer platform, the more difficult the load balancer becomes to support and the more vulnerable it becomes to downtime.

In very large enterprises, these multi-featured load balancing platforms can be a good option. However, organizations need to make sure that they have the specialist IT skills in-house to manage and maintain them. As more and more applications are added to large load balancer platforms, and particularly as more customizations are made, the complexity of the application delivery environment can become overwhelming.

IT staff can struggle to complete upgrades and can inadvertently cause downtime in a mission-critical application, while performing routine updates.

The biggest challenge for organizations that adopt the multi-application load balancing approach is how to cope with the proliferation of applications. The number of apps used within medium to large enterprises has grown at an exceptional pace, as is likely to grow even more rapidly in the immediate future, as companies launch new apps to respond to COVID-19.

According to Okta's 2020 Businesses @ Work (From Home) study, the average number of apps in an organization has reached 88, up 21% from three years ago, and 10% of organizations now rely on more than 200 apps.[1] The more apps that organizations try to balance on a single load balancer platform, the more difficult the load balancer becomes to support and the more vulnerable it becomes to downtime.

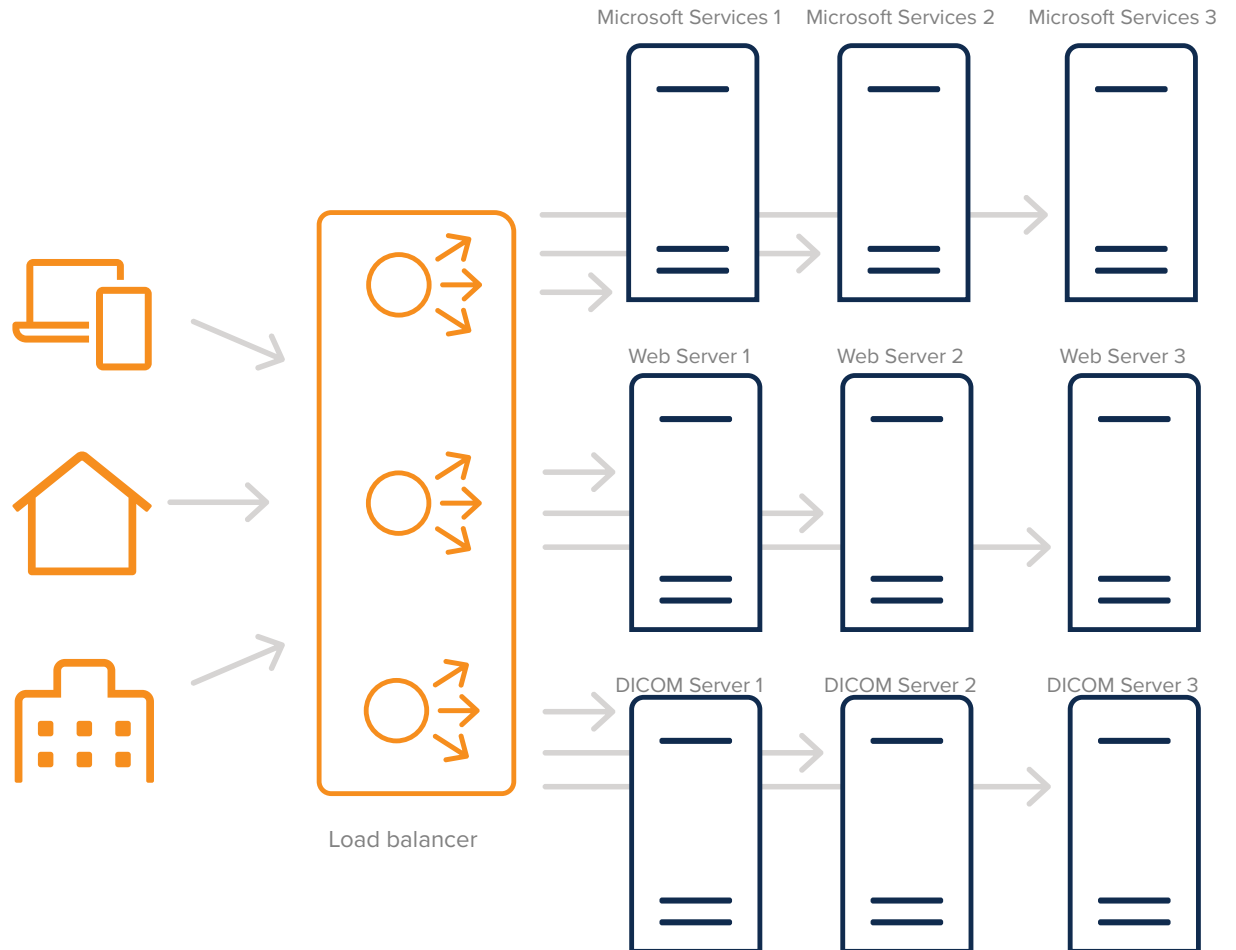1   https://www.okta.com/businesses-at-work/2020/

MULTI-APPLICATION LOAD BALANCING

Suitable for

- *Making complex application delivery customizations*

- *Accessing additional security features*

- *Being supported by load balancer specialists or external consultants*

Not recommended for

- *Balancing large numbers of applications*

- *Guaranteeing the uptime of mission-critical applications*

- *Being supported by IT teams without specialist skills*

Microsoft Services 1   Microsoft Services 2   Microsoft Services 3

Web Server 1   Web Server 2   Web Server 3

Load balancer

DICOM Server 1   DICOM Server 2   DICOM Server 3

Key considerations with multi-application load balancing:

WILL I USE ALL OF THE FEATURES I AM PAYING FOR?

The large enterprise load balancers, designed for multi-application load balancing, come with a host of additional features that are enticing. However, according to a 2018 NGINX and O'Reilly survey[1], 90% of companies only use basic load balancing, despite paying a premium cost for solutions with supplementary features.

CAN I MANAGE AND MAINTAIN A HIGH END LOAD BALANCER IN-HOUSE?

The more applications organizations put on a single load balancer, the more complicated they become to manage and maintain. Even simple upgrades that should take hours can take many months to plan and implement and will require expert skills. In an evaluation of the pros and cons of a single or dual-vendor approach for servers, Gartner Group points out that not relying on just a single vendor can mitigate risks[2] and the same can apply for load balancers.

ARE THERE ANY HIDDEN COSTS?

When considering the multi-application load balancing approach, it is important to take into account the additional and often hidden costs associated with managing and supporting a high end enterprise load balancer platform. It can cost twice as much as a new load balancer to bring an enterprise load balancer back into support, and organizations may have to hire consultants to help them implement or amend customizations, if they don't have the skills in-house. Centralized control can potentially increase security, but the result is slow deployments and difficulty making changes in the future.

DO I HAVE THE SKILLS TO MITIGATE SECURITY RISKS?

The customization features offered by large enterprise load balancers can be very useful, enabling organizations to specify exactly how traffic is directed to meet a business requirement. However, over reliance on customization can create

security weaknesses, such as the one identified in F5's iRules in August 2019.[3] Similarly Citrix had to address a "critical flaw" in its application delivery controllers in January 2020.[4] As all multi-application load balancer implementations are unique, organizations will need to be able to detect and mitigate any security risks themselves.

HAVE I ACCOUNTED FOR PEAKS IN TRAFFIC?

It is important to understand the potential peaks in traffic for each individual application supported in a multi-application load balancing environment. If resources are not allocated correctly for each app, a peak in traffic in one app could adversely affect the performance of several others or lead to the failure of a critical business system.

WHAT ARE THE PROCESSING POWER REQUIREMENTS?

In a recent test performed by a storage vendor, a standard load balancer product only required 1 CPU and 2GB RAM, while the more complex enterprise load balancer required more than 4 CPU and 16GB of RAM to perform the same function and remain stable. The extra resource required for the large load balancer platform added significantly to the overall cost of the solution, whilst providing no extra benefit for the customer, as the extra features included would never be used.

WILL I BE ABLE TO RESIST THE URGE TO SWEAT THE ASSET?

When you have purchased your new enterprise load balancer, and sized it appropriately with capacity for unforeseeable, exceptional peaks in application usage, you need to have the discipline to keep this capacity ring fenced. Post procurement, organizations commonly make the mistake of looking at their usage data, thinking they have excess capacity and adding more applications to sweat the asset. This can overload the platform and lead to application downtime when peaks occur.

1   https://www.nginx.com/resources/library/survey-2018-new-infrastructure-for-a-multi-cloud-world/
2   https://www.gartner.com/en/documents/2503843/when-two-server-vendors-are-better-than-one
3   https://www.finextra.com/pressarticle/79455/serious-security-issue-in-f5s-big-ip-could-lead-to-cyber-breaches-en-masse
4   https://www.forbes.com/sites/kateoflahertyuk/2020/01/20/critical-citrix-fix-now-availableheres-how-to-apply-it/

FINANCIAL ORGANIZATION

A large financial services organization in the UK had two data centers and adopted the multi-application load balancing approach to balance all its traffic for hundreds of applications. Using the single load balancer platform it selected, it configured load balancing capabilities in different ways for different applications and security zones within virtual instances of the product. This resulted in a multi-application load balancer comprising up to twenty virtual high availability pairs across the two data centers. Some virtual systems were dedicated to specific departments or applications, while others were used cross platform to provide services to many applications.

Over the course of a few years, the financial services organization began to encounter a series of challenges with managing its multi-application load balancer environment. In particular, routine load balancer updates – for new features, patches or security updates for example – became overly complex and time-consuming. As the organization needed to upgrade the core operating system, all virtual instances of the load balancer – and therefore all business applications – would be impacted simultaneously. This was deemed too high a business risk, so each virtual instance was upgraded and tested individually, before going into production, which meant that load balancer software upgrades took up to eighteen months of effort to complete. In some cases, the upgrade required significant re-architecting of the usage of the individual virtual instances on each chassis, and additional investment, as the upgrade needed more memory or CPU resource than had previously been available.

The load balancer used by the organization had an extensive range of additional features, beyond load balancing. These features enabled the organization to add additional security layers or other functionality which was not available in the native applications themselves. This solved a range of business problems, but meant that ongoing upgrades of either system were then intrinsically linked. When business requirements changed, it took technical specialists from multiple teams a significant amount of time to adapt the applications, leading to added costs and project delays.

PER-APPLICATION LOAD BALANCING

Flexible, but resilient

The alternative approach to multi-application load balancing is per-application load balancing. This involves using a dedicated load balancer (or more commonly a pair of dedicated load balancers) for each application. This approach is particularly recommended for mission-critical applications, as a pair of load balancers can be tuned specifically to ensure the high availability of the priority application, without having to take the needs of other applications into account. Far easier to set up, manage and maintain, they are also less likely to be compromised by human error.

The reduced cost of hardware, together with the availability of virtual and cloud-based load balancers, makes the per-application approach very viable.

In the past, it was cost prohibitive to have separate load balancers for each application. Now, the reduced cost of hardware, together with the availability of virtual and cloud-based load balancers, makes the per-application approach very viable, particularly when the reduced costs of support and management are taken into account. If it is company policy to standardize on single vendor solutions, to simplify the IT estate, then organizations can buy multiple load balancers from the same vendor and configure a pair to support each critical application. Furthermore, as it is much easier to configure load balancers when you only have to take into account the needs of a single application, this approach enables new apps to be launched much more quickly.

Load balancers installed for single applications don't generally come with as many supplementary features as the much more expensive load balancer platforms commonly used for multi-application load balancing. This, however, does not mean that their core load balancing capabilities are inferior. Easy-to-use load balancers still come with advanced, intelligent load balancing capabilities and are capable of routing and rerouting traffic in a fraction of a second to deliver exceptional user experiences.

Strictly speaking, the per-application approach involves allocating one pair of load balancers to one mission-critical application, but standard load balancers do have the sophisticated functionality needed to guarantee high availability for two or more applications simultaneously. Therefore, the definition of the per-application approach can be widened to encompass two mission-critical applications or a small cluster of similar applications.
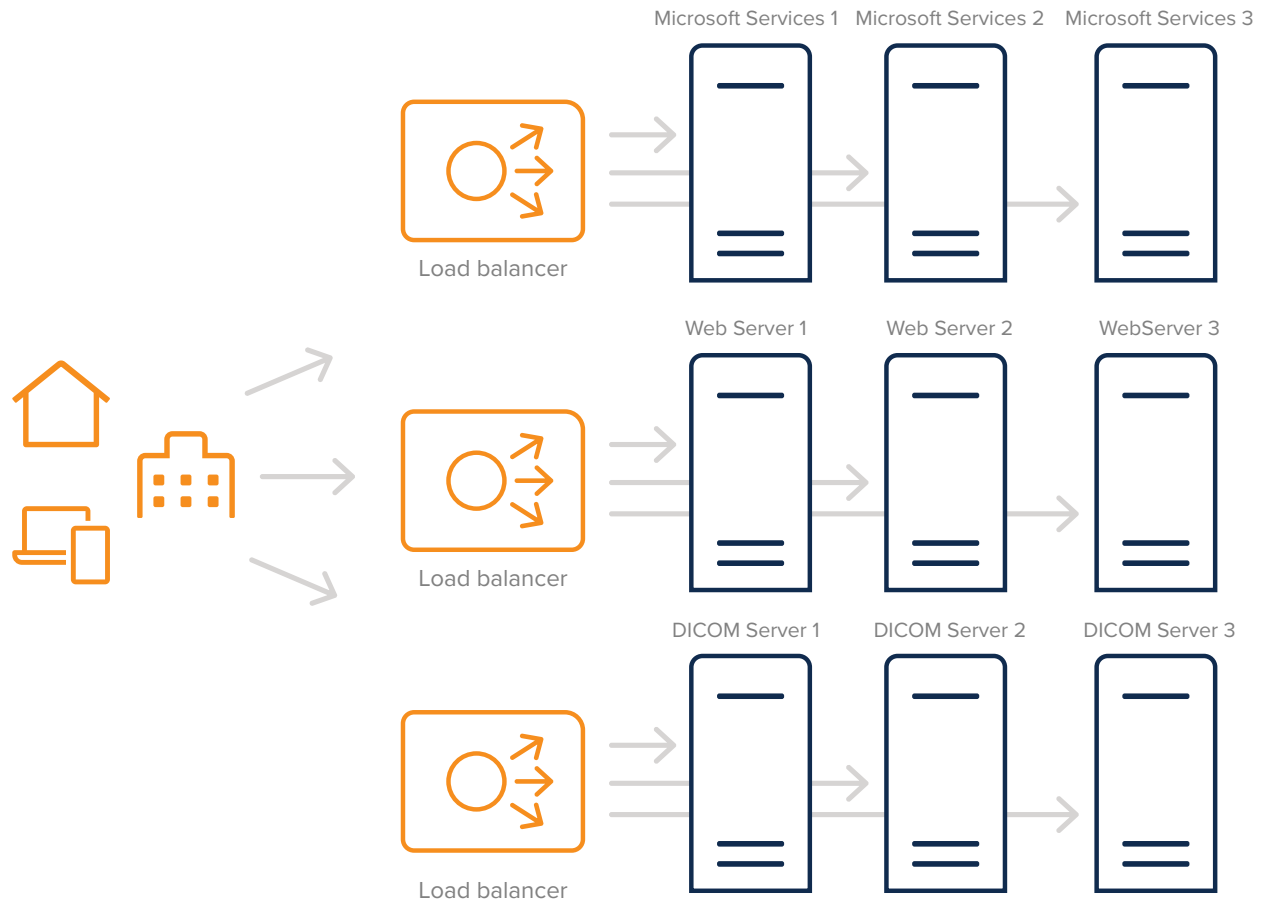
PER-APPLICATION LOAD BALANCING

Suitable for

- *Ensuring the availability of mission-critical applications*

- *Applying intelligent load balancing at lower cost*

- *Deploying new apps quickly*

- *Being easily supported by in-house IT teams*

Not recommended for

- *Accessing supplementary functionality beyond load balancing*

- *Having centralized control of all load balancing from a single dashboard*

Microsoft Services 1 Microsoft Services 2 Microsoft Services 3

Load balancer

Web Server 1 Web Server 2 WebServer 3

Load balancer

DICOM Server 1 DICOM Server 2 DICOM Server 3

Load balancer

Key considerations with per-application load balancing:

DO I NEED TO GUARANTEE UPTIME FOR MISSION CRITICAL APPLICATIONS?

The more critical an application is to the running of your business or the safeguarding of data, the more imperative it is for it to be protected by its own dedicated load balancer. It would be unthinkable for a vital medical diagnostic system to be taken down by mistake when a technician makes an upgrade to an email system.

WHAT DOES THE APPLICATION VENDOR RECOMMEND?

Application vendors that require the use of a load balancer with their solutions generally prefer their customers to have a dedicated load balancer product. Then, if their customers need help to address a problem or upgrade the application, they will be familiar with the set-up and can more easily make changes without putting other applications at risk.

WILL I HAVE THE AGILITY TO IMPLEMENT NEW APPS QUICKLY?

If speed of application delivery is important to you, then the per-application load balancing approach may be best. It can be significantly easier and therefore quicker to configure a load balancer specifically for a new app, than to try to adapt or customize an existing load balancer to accommodate the new solution, without jeopardizing the performance of other the other apps it already supports.

IS CENTRALIZED VISIBILITY AND CONTROL ESSENTIAL?

If it is, the multi-application approach might be best. The per-application approach is best suited to organizations that want to stay true to a decentralized approach.

WILL I GET THE THROUGHPUT I NEED?

It is essential to understand what throughput you will need, whether you opt for the multi-application or the per-application load balancing approach. To avoid future constraints and unanticipated increases in licensing costs, look for correctly sized solutions with high throughput options. Hardly any internet apps require more than 1G (unless they also stream video), while local network-area storage and object storage solutions could need 50G+.

HOW DO I MANAGE COSTS?

Standard load balancers, used for per-application load balancing, are a fraction of the cost of large enterprise load balancing platforms commonly used by large organizations to balance all the applications across the entire business. However, if you assign one pair of standard load balancers to multiple applications your costs could grow. Limit your expenditure by looking for site licenses that entitle you to deploy as many pairs of load balancers as you need for one fixed annual price.

WHAT SUPPORT WILL MY IT TEAM NEED?

Managing and maintaining per-application load balancers is much simpler than managing and maintaining a large enterprise load balancer that supports dozens or even hundreds of applications. Even so, you should check to see what level of support is available from your shortlisted load balancer vendors and make sure you can have direct access to engineers who can support you if needed.

DO I NEED A LOAD BALANCER AT ALL?

Ultimately, if a non-essential business application is available as Software-as-a-Service and your business is open to using the cloud, then this could well be a better option than using a load balancer. However, mission-critical solutions, like medical imaging, will need to stay on premise, because, despite the advent of 5G, the cloud is unreliable and is likely to remain so for the next five to ten years.

DATA STORAGE

A world-renowned environmental research institute in the UK amasses huge volumes of new data every day from remote sensors, satellites, researchers and the output of sophisticated data modelling systems. All this data needs to be backed up securely in an object-based storage system, so that it can be secured and used to raise awareness of the impacts of humankind on the global environment.

Given the criticality of the object-based storage platform, the organization installed a virtualized load balancing solution specifically for this application. The load balancer backs up tens of terabytes of data changes overnight by transferring the data from multiple Linux servers in a fast and intelligent way to remove potential bottlenecks. The selected load balancer doesn't have any throughput limits that might constrain its ability to store exceptionally high data volumes and effortlessly handles spikes in data.

PER-APPLICATION BENEFITS:

- **Delivers critical secure backup**

- **Unlimited throughput to cope with spikes**

- **Cost-effective virtual solution**

MEDICAL IMAGING

For many years, a leading supplier of medical imaging systems had advocated the per-application load balancing approach and had included load balancers as part of its diagnostic imaging solutions to ensure that medical practitioners could have uninterrupted access to patient data. However, the supplier had used a range of load balancers, from multiple vendors. This meant that different customers had slightly different installations, making it harder for the supplier to support its customers.

The medical imaging systems supplier subsequently standardized on a single load balancer vendor and worked with the load balancer vendor to create a bespoke installation guide. The per-application load balancing approach, together with the standardization on a single load balancer, enabled the supplier to accelerate the deployments of its medical imaging solutions using the same proven implementation approach for each customer. The supplier could also improve its customer service and support, as technicians have just one product and one standard network architecture to learn.

PER-APPLICATION BENEFITS

- **Uninterrupted access to data**

- **Reduced complexity improves vendor support**

- **Standardized solution means quicker deployment**

EVOLVING APPROACHES

Without doubt, the COVID-19 pandemic and current business uncertainties are changing the way that organizations view load balancing. Both the multi-application approach and the per-application approach are evolving, in response to new business concerns. Two key trends are emerging:

THE PER-APPLICATION APPROACH IS GAINING PACE WITH VENDOR SUPPORT

More organizations are adopting the per-application approach to load balance their mission-critical applications and help them maintain 100% uptime. Recent changes in working practices and consumer behavior have shown them how important it is to have IT infrastructure that is easy to manage and scale up without throughput restrictions. They don't want to potentially jeopardize the performance and availability of their most vital apps by having a complex load balancing platform that they cannot easily adapt and support.

Many vendors of mission critical systems, now proactively recommend per-application load balancing to their customers.

In addition, many vendors of mission critical systems, including suppliers of medical imaging systems and critical storage platforms, now proactively recommend per-application load balancing to their customers. These vendors need to know that their solutions will work dependably and uniformly at each customer site. They therefore recommend a specific load balancer, so that they can control the configuration and have the knowledge to support and maintain it in the future.

THE MULTI-APPLICATION APPROACH IS BECOMING MORE REALISTIC IN SCOPE

Large organizations are questioning the value they gain from large load balancer platforms. Previously, they were happy to invest in these solutions and pursue the multi-application approach whole-heartedly, because they liked the assurance of 'one product fits all', were attracted by the additional features and could afford to employ the specialists needed to support them. Years later, the employees who set up these vast load balancer platforms have moved on, their replacements lack the same knowledge, and the expensive additional features are not being used.

Now that the business outlook is more uncertain and application reliability is more critical than ever, organizations are becoming more realistic about how many applications they want to balance with a single load balancer. When organizations talk about balancing multiple applications today, they tend to be referring to smaller clusters of applications rather than all the applications they have. They recognize that it is safer to allocate separate load balancers to different parts of the business and different applications, to reduce unnecessary load balancer complexity. They also realize that they can use standard load balancers for this dual-application or clustered-application approach, which significantly reduces their capital costs.

About the company

Loadbalancer.org's mission is to ensure that its clients' businesses are never interrupted. The load balancer experts ask the right questions to get to the heart of what matters, bringing a depth of understanding to each deployment. Experience enables Loadbalancer.org engineers to design less complex, unbreakable solutions - and to provide exceptional personalised support.

To discuss your load balancer requirements with load balancer experts, contact Loadbalancer.org on:

US Headquarters

Loadbalancer.org, Inc.
4550 Linden Hill Road,
Suite 201
Wilmington, DE 19808
USA

EMEA Headquarters

Loadbalancer.org Ltd.
Compass House
North Harbour Business Park
Portsmouth, PO6 4PS
England, UK

Canada

Loadbalancer.org Appliances Ltd.
300-422 Richards Street
Vancouver, BC
V6B 2Z4
Canada

Germany

Loadbalancer.org GmbH.
Tengstraße 27
80798
München
Germany