Website reliability, scalability, performance and ease of management are essential for any business and small to medium enterprises (SMEs) are in the middle of the same IT evolution as their larger counterparts, just on a smaller scale. Leigh Bradford from KEMP Technologies looks at what is driving demand for load balancing from the SME community.

Traditionally, server load balancing has been viewed as a problem for large enterprises and a costly and unnecessary expense for SMEs. But that was when most SMEs ran their businesses using just a couple of servers. The average number of servers used by an SME has risen from two to between 10 and 14, making server load balancing a must-have item on the IT shopping list.

In effect, load balancers are like the traffic police; they keep the traffic moving by directing and sharing internal network traffic and incoming connections across multiple servers. With back-end applications such as order processing, billing and customer management being integrated into the complete supply-chain, web-enabled applications, website reliability, scalability and performance are essential. To deal with this, server load balancers or application delivery controllers (ADCs) provide the ability to direct traffic to the best performing, most accessible servers based on factors such as concurrent connections and CPU/memory utilisation. If a server or application fails, the user is automatically re-routed to another functioning server.

If you were the coach of a sports team, would you have a team that is made up entirely of players who were specialists in defence or a team composed entirely of attacking players? Indeed not, and it is for this reason that load balancers play a vital role in any organisation where server uptime, client and staff connectivity and performance are considered essential. Load balancers are the playmakers, the midfield; the vital link in a corporate infrastructure.

Server load balancers use various techniques to distribute traffic load between two or more servers, routers, firewalls and other networked resources, to optimise utilisation and improve performance and response time. Most are capable of providing Layer 4 to Layer 7 management. Layer 4 is limited to web requests destined to TCP Port 80; therefore, no further differentiation among server groups is possible. Layer 7 switching uses
application-layer criteria to determine where to send a request to in order to provide more granular control over forwarding decisions.

Content switching is used to distribute or balance user requests to servers based on the Layer 7 payload. Most commonly, this is done by examining page content such as a URL and ‘switching’ requests to the appropriate server or group of servers. For example, www.xyz.com/images may be pointed to a server that handles graphics, while the www.xyz.com/shop may be pointed to a transaction server. In addition, if a web application makes extensive use of cookies, a Layer 7-capable load balancer can switch users based on cookie values, providing a much better model for achieving server persistence.

To further enhance and secure the user experience, more advanced server load balancers also provide secure socket layer (SSL) offload/acceleration that enables processor-intensive SSL handshake encryption/decryption processes to be removed from the servers. This offloading dramatically increases server performance, while decreasing the time and costs associated with SSL certificate management. In addition, Layer 7 capable balancers provide for application-level health checking and a more accurate means of establishing server persistence.

Another big driver for server load balancing among SMEs is the list of changes Microsoft has made to its core server architecture in Exchange 2010. Now that Exchange Client Access Server (CAS) is used to handle all client connections, it is important to ensure that email users do not suffer from poor performance and user experience by providing load balancing to automatically re-route and reconnect email users to functioning servers.

As well as the migration to Exchange 2010, many SMEs adopting Microsoft SharePoint and Lync Server technologies are discovering the need for load balancing. In fact, Microsoft increasingly recommends the use of server load balancing to optimise performance and resilience. Unlike Exchange, Lync has to support real-time VOIP traffic flows that are jitter – and latency-sensitive.

Microsoft Lync Server 2010 supports two load balancing solutions. DNS (Domain Name System) load balancing or so-called ‘round robin’ balancing is a basic software solution that works by responding to DNS requests by simply alternating these requests between servers, without consideration for matching the user IP address and its geographical location, server load or network congestion, for example. This is used to balance specific Lync network traffic such as SIP and media.

Hardware load balancing has an important role to play in Lync deployments that require high availability and need to be scalable. But in fact the term hardware load balancing is slightly misleading as it doesn’t actually have to be hardware at all. A number of virtual/software load balancers for Hyper V or VMware environments have also completed Microsoft qualification testing with Lync Server 2010, giving SMEs a range of options that are best suited to their corporate environments.

In addition, with applications increasingly being hosted across multiple data centres; global load balancing can be used to ensure users seamlessly connect to the fastest performing and geographically closest data centre for web-based applications such as Microsoft Exchange or ecommerce services to deliver optimal performance. If one site goes down, traffic is diverted automatically to the failover site for guaranteed resilience. Global load balancing enhances the end-user experience by distributing traffic to geographically dispersed data centres through advanced regional or location-based load balancing algorithms that identify traffic patterns and distribute new traffic accordingly.

While large enterprise load balancing solutions have been too costly and complex for SMEs, there is a new generation of affordable hardware or virtual load balancers or ADCs designed for SMEs. The complexity and scale of technology required to run today’s SMEs has brought with it many new challenges to deliver performance, high availability and security. SME IT is the same as enterprise IT in everything but scale and if SMEs want the functionality and quality that enterprise CIOs take for granted, load balancing has to be part of the equation.