Communicating With Sam

Scaling WebSphere: Mission Critical Communications

Sam Johnston

Question:

Currently we have implemented a standalone WebSphere Express for iSeries application to support our e-business presence. While this application has allowed us to get started quickly, our e-business requirements are evolving due to business growth and a shift in transactions towards self-service. To support this trend and maintain if not increase profitability, we would like to improve our operational efficiency and strengthen customer loyalty while Web enabling more of our business processes. How do we ensure that WebSphere and our network performance are optimized to deliver the demanding level of service this will require?

Answer:

The Websphere family of products has evolved to become a leading software platform for meeting the most challenging e-business requirements. There are a number of Websphere products that will allow you to choose the right business model by flexibly scaling how far to Web enable your business. These tools broadly fall into three categories:

- Foundations & Tools
- Business Integration
- Business Portals

Based on your situation, we will focus our discussion on the foundation and networking aspects of the environment. The key piece of the foundation is the WebSphere Application Server (WAS). The WAS comes in multiple configurations to meet specific business needs. If you are seriously working on becoming an e-business then it is likely that you will need to migrate from your current Express platform to the Network Deployment (ND) version of WAS. This version extends the server infrastructure to include full support of J2EE technologies including Enterprise JavaBeans and Java Message Service, cluster capabilities (on the same or multiple machines), edge services, and high availability. These are key features required to build mission critical WEB applications.

It is crucial to ensure the iSeries server has the capacity to handle the additional workload in the form of WEB enabled applications and planned increases in transaction volumes. You need to remember that WebSphere is the middle tier in a three-tiered e-business environment, linking the HTTP server to the business data and application logic. Ongoing iSeries performance reporting and management is a must in this environment to ensure acceptable response times.

Highly available systems and networks need to have at least two of everything to deliver the adequate redundancy. The redundancy, both logically and physically, allows work to continue in the event of a single component or service failure. WAS ND includes crucial logical capabilities such as advanced cross-domain, multi-domain, availability and failover capabilities. This means you can architect a solution to eliminate single points of failure, while still benefiting from a single administration repository for maintaining control of all servers and processes within an extended environment. In other words, robust, while still simple to operate.

WebSphere has several Edge Components, which are essential if you are to extend the capabilities out into your network. These features are valuable in creating environments capable of reacting to dynamic high volume workloads. In an iSeries environment, the architecture is such that these services run on a separate server.

- Load Balancer: WAS ND has a general-purpose load balancer for dynamic routing and balancing of HTTP requests amongst multiple Web servers.
- Site Selector: This is used with the load balancer to act as a DNS server when the load exceeds the capacity of a single load balancer.
- The Caching Proxy component adds SSL termination, authentication, caching and routing functions.

They key here is to understand that as your environment grows, there are many new layers you will need to add to your WebSphere architecture to ensure that it scales effectively and with granularity.

While so far we have focused on tools associated with the WebSphere platform, it is crucial that you understand the impact on your network if it is to support the scaling your solution.

In scaling your WebSphere applications to the customer, it is also crucial that your network performance does not become the bottleneck. It is essential that you have the same levels of redundancy and load balancing engineered into the network infrastructure.

▶

Initially, at a minimum, it is mission critical that you provision your network to support redundant routing, switching and firewall services. As your solution continues to scale, and becomes a high capacity e-commerce site, the solution should also incorporate load balancing of the key network components that will increase throughput while also providing redundancy.

The architecture of your solution should include, but not be limited to:

- Firewall load balancing
- Content switching for intelligent load balancing of available application or web servers
- Content switching to support GSLB to provide redundancy of application engines across geographies (excellent support for region wide power outages).

Figure 1 is an example of the network architecture that could be deployed to create the level of redundancy, availability and throughput performance via load balancing.

The key message is not that your network architecture will look exactly as shown, but rather that it is crucial to consider the impact on the network from the outset. If designed correctly, your network strategy can be as scaleable and flexible as the WebSphere platform in terms of application support. If you integrate the network component within the overall architecture, you should be able to add pieces as needed while minimizing the re-work. You can ensure this by following two simple rules: first, make sure your WebSphere team includes a network expert who can champion the architecture, and secondly, make sure that your initial deployment strategy includes an end-to-end architecture of the transactional servers and network touch points, including understanding the scaling points and paths.

This all sounds like common sense, but unfortunately the world is littered with failed e-commerce projects where the network strategy was quickly assembled on the turn-up date! The chaos of slow or failed transactions on the go live date is no way to determine the scaling requirements of your network. TUG

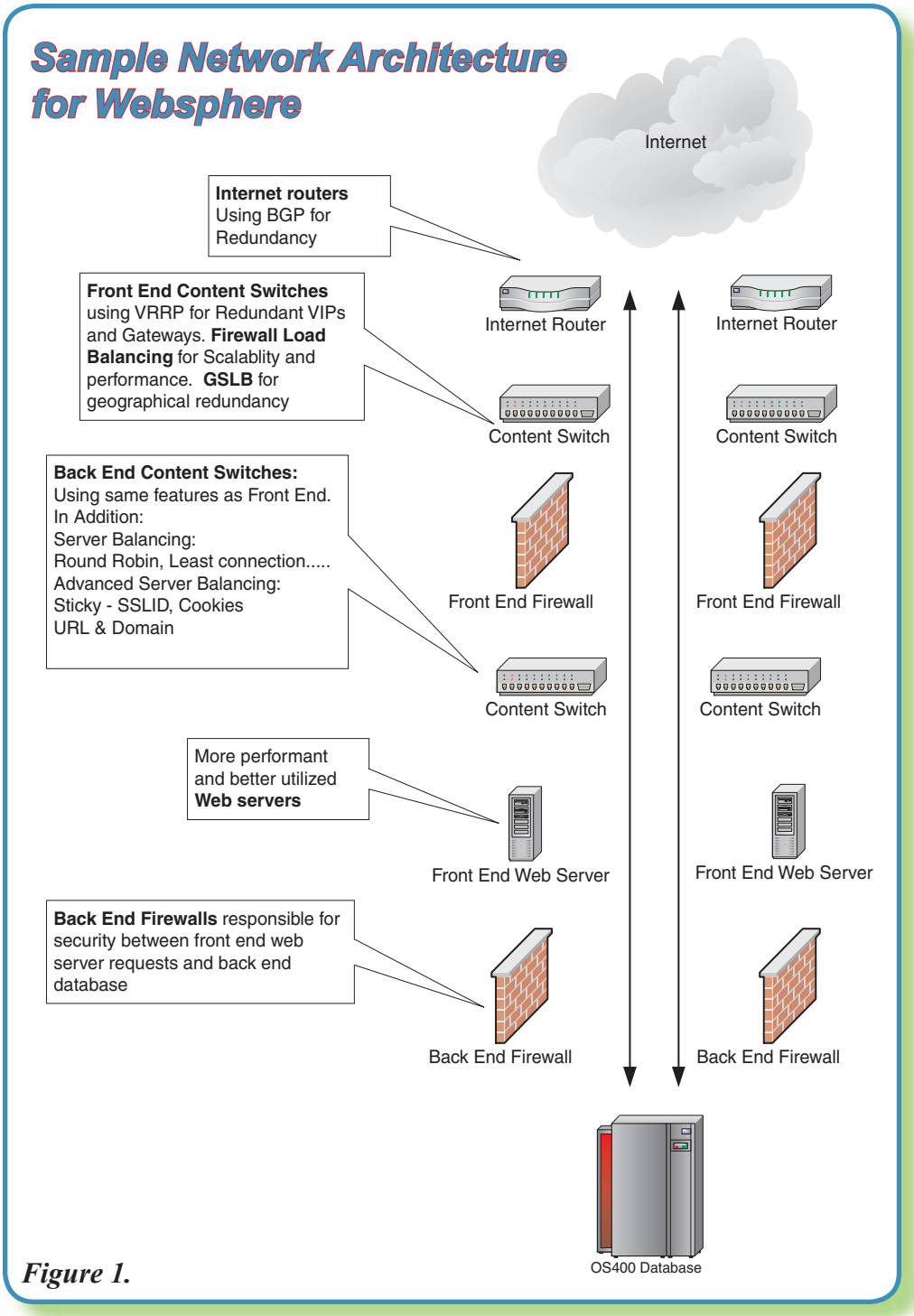

Figure 1.

Sam Johnston *is a partner and Chief Technology Officer of Intesys Network Communications Ltd., providing value-added networking and e-commerce solutions to the iSeries community. He can be reached at (416) 438-0002 or via e-mail at sjohnston@intesys-ncl.com.* *Any TUG member wishing to submit a question to Sam can forward their typewritten material to the TUG office, or to Intesys. The deadline for our next issue is Friday December 5, 2003.*