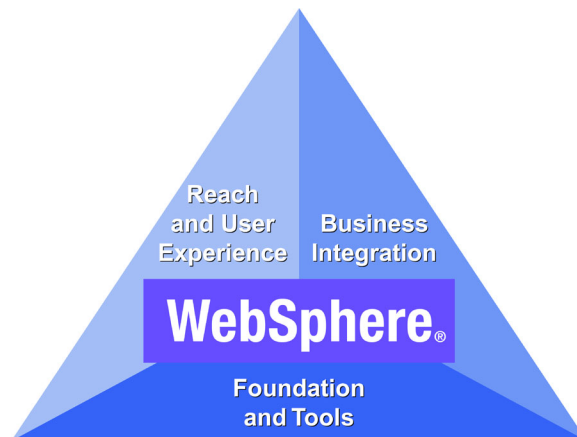# The power of edge of network technology

Edge technology in WebSphere Application Server Version 5

Laura Rademacher
IBM WebSphere Application Server

Executive overview

In today's ever-changing business environment, you need to deploy high-volume, dynamic Web sites. The notion of an infrastructure that cannot be interrupted evolved from a desire to a critical requirement for a successful, e-business. Today, application servers need to perform in a sophisticated, secure and cost-effective—yet, simplified and intelligent manner.

With the integration of edge of network technology, *IBM WebSphere® Application Server, Version 5* can meet these demands. By utilizing edge technology, *WebSphere Application Server, Version 5* can dynamically react to network pressures with security, thus helping to create an intelligent, optimized application platform that can ultimately lay the foundation for e-business grid services capabilities.

This paper will address the power of the edge of network technology in *WebSphere Application Server, Version 5* in terms of performance, availability and scalability enhancements.

Performance

What if you could take the same number of machines, and process more work through those machines than ever before? Taking that thought even further, what if you could take fewer machines, and yet process significantly more work? This could have a direct bearing on cutting costs. Faster and more consistent response times could improve customer service and customer satisfaction. The value of performance optimizations such as these can have a direct and immediate impact to the bottom line of your business.

Just imagine the efficiencies gained if application servers were taught to allocate proportionately more of the shared work to the most capable machines. What if application servers were smart enough to detect variable runtime conditions, then redirect work dynamically to the machines that are least busy?

Through the integration of IBM WebSphere Edge Server technology in *WebSphere Application Server, Version 5* (select deployment options only), these efficiencies can be achieved.

To help ensure ever-changing, accurate traffic allocation to back-end servers, an advisor can be deployed as part of the Load Balancer component. It can be as high level as periodically determining the overall status of the servers or as granular as checking specific application response times on the servers. Once the server health is determined, the advisor informs the load balancer manager function, which then sets weights for the servers to determine which server should receive new session or application requests. Through advisor code, traffic is appropriately routed to the optimal back-end server.

To dynamically—and efficiently—share workloads among WebSphere application servers, Dynamic Workload Management and clustering can be used. This Workload Management Controller (WLMC) provides end-to-end monitoring and weighting of WebSphere Application Server Servlet and Enterprise JavaBeans (EJB) engines—enabling dynamic changes in the workload distribution as the load on back-end servers changes, instead of relying on predefined static metrics.

When using the WLMC function, load balancing attributes and operations are exposed through a Java™ Management Extensions (JMX) standard interface. JMX provides a simple, lightweight way to instrument Java objects. Throughout the IBM WebSphere platform, manageable objects are exposed through JMX so that compliant management systems can manage WebSphere resources in a standard way.

For example, an external management system can define what metrics are collected for each WebSphere Application Server Cluster, and what relative weightings are used for each metric. The JMX interface to the WLMC exposes the methods for starting and stopping the controller, and attributes for setting timeouts, network communication parameters, log levels and high availability parameters. Management systems can register to be notified when the WLMC detects that load-balanced resources are not functioning, and when resources are returned to the active pool.

Additionally, to extend the load balancing capabilities beyond purely a WebSphere Application Server environment, consultant code can be used to help optimize server performance within a Cisco or Nortel infrastructure. Consultants generate server weighting metrics and distributes them to Cisco CSS 11000 switches or Nortel Alteon 180 series of switches for optimal server selection, load balancing and fault tolerance.

To further improve performance, edge of network caching technology can be deployed. Caching can be used to reduce network congestion by storing frequently accessed content so information is retrieved only once. Information can be cached depending on when it will expire, how large the cache should be, and when the information should be updated. Faster download times for cache hits mean better quality of service for customers and reduced load on back-end servers.

The enhanced caching capability in *WebSphere Application Server, Version 5* can improve response time by offloading back-end servers and peering links. And, in contrast to other caching proxies that can cache *static* content, the edge of network proxy server can also cache—and invalidate—dynamically-generated content from WebSphere Application Server, such as Java Servlet Program and servlet results to create a virtual extension of the application server cache into network-based caches—or to caches in the Akamai network through the implementation of Edge Side Includes technology.

Edge Side Includes (ESI) is a simple mark-up language and proposed standard for the dynamic assembly of Web page fragments, such as stock quotes and individual catalog prices. By leveraging ESI technology, dynamic content caching can be extended by moving fragments from the application server to a proxy server that resides in the network—such as Akamai's. This can enable caching to occur at a more granular level, as well as allowing you to position page composition at the most optimal location, closer to the end user. As a result, you can help improve user experiences through expedited, personalized page composition, and reduce workload on the network servers due to fragment offload to the edge. In addition, *WebSphere Application Server, Version 5* maintains control over the externally cached fragments through the ESI Invalidation Gateway.

Edge of network caching technology allows for an alternate proxy deployment option—beyond the currently available WebSphere Application Server, Version 4 caching capabilities—as well as enabling extensibility options as a result of supported edge proxy application programming interfaces (APIs).

Availability

Now, let's consider availability. Highly available systems need to have at least two of everything. This allows work to continue in the event of component failure. The more failure bypass that a system offers—including the failure of internal components—the less disruption an end user will experience when something goes wrong. By reducing unplanned outages, you can help reduce costs and make itself more accessible to customers and partners as part of the evolved business models that link the company with others.

This issue becomes ever-apparent as companies extend their network capacity and availability through the deployment of multiple WebSphere domains. To support these configurations, *WebSphere Application Server, Version 5* will include advanced cross-domain, multi-domain, availability and failover capabilities. With this integration, you can achieve the elimination of a single point of failure between dispersed data centers, while still benefiting from a single administration repository for maintaining control of all servers and processes within this extended environment.

To further enhance network availability—with security—companies can choose to implement High Availability, Light-Weight Directory Access Protocol (LDAP) Support. As an open standard, LDAP provides an extensible architecture for centralized storage and management of information required for today's distributed systems and services. In *WebSphere Application Server, Version 5* you can support secure communication amongst LDAP-replicas for enhanced user performance and system availability.

Scalability

The final optimization to consider is scale. Scaling means you can easily handle more work as your business grows beyond its traditional boundaries—embracing new customers, partners and suppliers. Effective scaling helps ensure all customers receive the service they deserve without breaking the bottom line.

Through inclusion of edge technology, you can achieve effective scaling through features such as Transactional Quality of Service (TQoS) and Content Distribution, in addition to the elements previously discussed. Through TQoS, companies can offer preferred services to priority users, groups, addresses, applications and URLs, through gold, silver and bronze classifications. Additionally, these classifications can be dynamically modified. For example, when a user switches from browsing the Web to purchasing an item, the user could be moved from a bronze to a gold service level. Using TQoS, important e-business transactions and applications can be expedited, allowing administrators the ability to easily customize and scale their WebSphere infrastructure.

In addition, effective scaling allows you to deploy new applications into large environments more easily and automatically. Through edge-based, content distribution technology, you can deploy published Web site content, including Web pages, fragments and application components, to caches and rehosting servers in the network—where it's most optimally positioned, closer to the end-user for reduced network latency. This is achieved by identifying each server as either a content distribution client (CDC) or content distribution server (CDS), facilitating communication between them to automate the placement, distribution, and synchronization of Web content. By using content distribution, you can easily distribute static, dynamic and multimedia content, as well as applications, to CDCs and

CDSs throughout the extended network for improved network utilization and enhanced user experiences.

Summary

In today's environment, it's more important than ever that you provide your employees, partners, and customers the most advanced levels of security along with superior performance. Through integration of edge of network technology in *WebSphere Application Server, Version 5* you can achieve new levels of scalability, availability and performance. With *WebSphere Application Server, Version 5* you can deploy scalable, highly available, application-optimized, cost-effective solutions—that meet today's challenges of growing your e-business, while laying the foundation for deploying e-business grid services in the future.

IBM®

e business software