Hardware Load Balancing for Optimal Microsoft Exchange Server 2010 Performance

A comprehensive F5® solution readies application infrastructure for a successful implementation of the re-engineered Microsoft Exchange Server 2010.

by Lori MacVittie
Technical Marketing Manager

Contents

Introduction

According to a study conducted by Ferris Research in 2008[i] Microsoft Exchange Server holds approximately 65 percent market share in email and communications across all organizations. In healthcare organizations with more than 5,000 employees, it enjoys 75 percent penetration; in telecommunications organizations with at least 1,000 employees, it does even better, garnering a 90 percent penetration rate. Small businesses, too, are particularly drawn to Exchange Server: In the Ferris Research survey, nearly all organizations with up to 49 employees currently use Exchange Server 2007.

With such a broad distribution across organizations of different sizes and throughout multiple industries, the effect of core changes on the Exchange Server 2010 architecture is significant when it comes to migration. It is no longer possible to simply replace existing installations and migrate mailboxes from one version to another. Instead, changes to Exchange Server 2010 architecture make it necessary to employ a migration strategy that includes re-evaluation of the supporting network architecture.

Exchange Server 2010

What's Changed

Normalization of user connectivity is the biggest change in the architecture of Exchange Server 2010 for which organizations need to prepare. In previous versions of Exchange Server, users might or might not connect directly to mailbox servers, depending on their particular client. Exchange Server 2010 no longer permits direct access to mailbox servers regardless of client type. Now, all client access is brokered through the Client Access server role.

The Client Access server role supports services for mailboxes, public folders, calendar items, the global address list, and related data. Also new to the Client Access server role in Exchange Server 2010 is RPC Client Access, which provides traditional "native" access to Exchange Server mailboxes via Messaging API (MAPI), but it moves the connectivity point from the Mailbox server role to the Client Access server role.

These changes, along with new requirements regarding the use of load balancing—and hardware load balancers specifically—to deploy Exchange Server 2010 have a significant effect on the application infrastructure.

Effect on Application Infrastructure

The changes in the internal architecture of Exchange Server 2010 mean that even internal users must be routed through a Client Access server role in order to access email. Such a requirement might necessitate network-level changes, such as new or modified routes and VLAN configurations, as well as new policies on firewalls. Furthermore, Exchange Server 2010 now requires load balanced Client Access server role implementations for internal connections. Microsoft now recommends a hardware load balancing [ii] solution rather than a software solution in all deployment scenarios requiring high availability. This is a change from previous recommendations that based the use of hardware load balancers on the number of CAS servers or Exchange Server roles deployed on a single machine. In essence, Microsoft's recommendation moves hardware load balancing to a required core component of a highly available Exchange Server 2010 deployment.

Microsoft has engineered Exchange Server 2010 for high scalability and efficient deployment, and it recommends that multi-role servers be employed for optimal scalability. The recommendation to utilize hardware load balancing solutions when scaling Client Access server roles comes from the ability to intelligently route requests at the application layer. This capability is common to what is often referred to as the modern load balancer, an Application Delivery Controller (ADC). An ADC offers additional application-focused features and functions beyond simple load balancing that can be leveraged to further improve the reliability, performance, and security of the applications it delivers. This includes the ability to apply other optimizations—such as caching, compression, TCP connection optimization, and SSL offload—that increase availability, performance, and security for Exchange Server, making hardware load balancers a natural fit in an Exchange Server environment.

Organizations employing multiple nodes to support a large user base might require changes to the network architecture, upgrading infrastructure, or investing in additional infrastructure to provide the same level of reliability and performance as previous Exchange Server installations.

Scale Out or Scale Up?

The decision whether to scale up (larger hardware) or out (load balanced multiple servers) must be made by the individual organization. Decision makers should consider information provided by Microsoft that comes from the company's unique understanding of the architecture of Exchange Server 2010 in large-scale deployments.

- Scaling out provides the following at low cost:
    - Large mailboxes
    - High availability
    - Rich feature set

- Scaling up:
    - Increases risk that an outage or failure will affect more users
    - *Usually* costs more, and can force feature decisions due to
  hardware choices [ii]

It is noted that scaling up usually costs more; however, there also are costs associated with scaling out, particularly if an organization does not currently take advantage of a load balancing solution. Organizations that have already invested in a load balancing solution will find the costs of scaling out significantly lower than scaling up even if upgrades or deployments of additional functionality are required.

Migration versus Cutover

When organizations determine it is time to make the move to Exchange Server 2010, it is often too complex to support both the existing and upgraded installations. Some organizations will therefore choose to simply "cut off" the old system and move to the new one overnight. This is a perilous process that often incurs additional support costs as users are unable to access Exchange resources.

Most organizations generally prefer a phased migration approach in which batches of users are migrated from existing Exchange mailboxes to the new infrastructure. This, too, comes with administrative costs and potential infrastructure issues, but is less likely to cause a disruption in service and allows organizations enough time to ensure the deployment is stable at each phase of the migration.
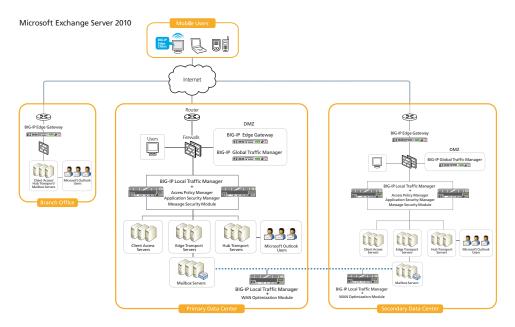
F5 Solutions for Exchange Server 2010

The F5 solutions for Exchange Server 2010 focus on providing security, availability, acceleration, and secure remote access to internal and external users of Exchange Server 2010. It is designed to simplify the process of scaling out Exchange Server 2010 based on Microsoft recommendations for highly available deployments. Not every deployment will require the use of all F5 components. Secure remote access, acceleration, message security, and global load balancing are

optional components that, while enhancing the overall user experience, security, and availability of email services, are not required to meet Microsoft recommendations.

Deployment architecture for complete F5 solution for Exchange Server 2010

The Deploying F5 with Microsoft Exchange Server 2010 guide includes detailed configuration assistance for each F5 solution component.

F5 Solution Components

BIG-IP Local Traffic Manager

With its core load balancing support, the F5 BIG-IP® Local Traffic Manager™ (LTM) Application Delivery Controller addresses the minimum requirement for deployment of Exchange Server 2010. BIG-IP LTM provides basic load balancing as well as advanced load balancing features that are necessary for some architectures in which Exchange Server 2010 might be deployed. In a recommended deployment, BIG-IP LTM load balances traffic for Client Access server roles and for incoming mail destined for Exchange Server 2010 Edge Transport server roles. This way, mail can be routed to Edge Transport server roles without interfering with the native routing built into both SMTP and Exchange Server 2010 that manages communication between different Exchange Server 2010 environments and from Edge Transport to Hub Transport server roles.

Beyond simple load balancing support for Exchange Server 2010, BIG-IP LTM can also improve application performance through features such as persistence (server affinity), connection optimization, and custom application control. Advanced health monitoring options provide a variety of mechanisms for evaluating Exchange Server 2010 components to ensure high availability of the entire Exchange Server 2010 infrastructure.

The minimum requirement to meet Microsoft recommendations for a highly available Exchange Server 2010 implementation is the deployment of BIG-IP LTM for load balancing. All other components of this solution, while certainly recommended by F5 to increase resiliency, security, and performance of Exchange Server 2010 implementations, are optional.

BIG-IP WAN Optimization Module

The combination of BIG-IP® WAN Optimization Module™ (WOM) with iSessions—a symmetric, optimized network tunneling feature of the BIG-IP platform—provides a secure tunnel through which optimized data can be exchanged with remote sites. When moving Database Availability Groups (DAGs) across data centers, BIG-IP WOM ensures that they are transported quickly and securely, making the process much less time consuming.

By deploying BIG-IP WOM on BIG-IP LTM, organizations can simplify their architecture by eliminating the need to employ separate WAN optimization controllers to enhance the transfer of large data files such as DAGs between locations.

BIG-IP Global Traffic Manager

BIG-IP® Global Traffic Manager™ (GTM) provides cross-site and data center redundancy, failover, and load balancing. BIG-IP GTM is particularly adept at collaborating with BIG-IP LTM to enforce performance requirements on Exchange Server 2010 in multi–data center deployments by choosing the site that best fits the needs of the user, especially when the user is traveling or at a remote location. For global organizations, the IP geolocation capabilities of BIG-IP GTM can further assist in building an optimized, global Exchange Server infrastructure based on user-specific location. These options enable more sophisticated deployments that are not only highly available but also highly localized and specialized based on the location of the users and the Exchange Server components.

BIG-IP Message Security Module

BIG-IP® Message Security Module™ (MSM) provides reputation-based, perimeter anti-spam functionality that significantly reduces the volume of spam processed by Exchange Server 2010 Edge Transport server roles, reduces the amount of storage required to comply with retention policies, and improves performance of Exchange Server 2010 by eliminating unnecessary messages from mailbox stores. These benefits mean fewer Exchange Server Edge Transport server roles must be deployed, which results in a need for fewer physical servers and lower costs associated with maintaining critical email infrastructure.

BIG-IP Access Policy Manager

BIG-IP Access Policy Manager™ (APM) is a dynamic authentication and authorization management solution built on the BIG-IP core platform. Combined with BIG-IP LTM, BIG-IP APM removes the time and complexity barriers often associated with Exchange migration by allowing migration to occur over time with no interruption to service. Because BIG-IP APM integrates with Active Directory (AD), only authenticated user sessions are allowed access to corporate resources, eliminating security risks associated with remote user access.

BIG-IP APM continues to add value after migration to Exchange Server 2010 is complete by continuing to perform authentication duties in the DMZ, thus preventing access to corporate resources to any but those with authorized access. By providing a single, unified point of access (a single URL) for all remote users of Outlook Web Access, ActiveSync, and Outlook Anywhere regardless of device, location, or network, a combined BIG-IP LTM and BIG-IP APM solution reduces administrative overhead and simplifies the process of securing Exchange components from unauthorized remote access.

BIG-IP Edge Gateway

F5 BIG-IP® Edge Gateway™ offers accelerated remote access support to Exchange Server 2010 via secure connections (including HTTPS, POP3S, or IMAPS, depending on choice of web browser or email client).

Edge Gateway contains further guidance on the implementation of endpoint security checks in addition to the configuration of accelerated remote access to email via Microsoft Office Outlook and Outlook Web Access. Endpoint security checks can assist in the enforcement of corporate policies regarding client security—

such as requiring anti-virus software and scanning for virus infections before permitting access to corporate resources.

This level of visibility and contextual awareness gives administrators flexibility in designing access policies based on location, device, or user, and it enables finer-grained control over access to corporate resources.

Edge Gateway further simplifies management of and access to corporate Exchange Server 2010 components by providing a single URL through which all remote users access Outlook Web Access, ActiveSync, and Outlook Anywhere regardless of device, location, or network.

F5 Management Pack

The F5 Management Pack for Microsoft System Center Operations Manager 2007 is a software plug-in that provides comprehensive monitoring for a range of F5 devices. The information produced and aggregated by the F5 Management Pack for Microsoft System Center Operations Manager can be used for trending and analysis, maintenance, diagnostics, and recovery actions.

For Exchange Server 2010 integration, the F5 Management Pack for Microsoft System Center Operations Manager can be combined with the Exchange Server 2010 Management Pack, to build up an aggregated (roll-up) model to manage the health of the Exchange Server 2010 distributed application environment. A typical use-case scenario for implementing this aggregated health model would be to map a group relationship between the Client Access server roles and the corresponding BIG-IP LTM pool members, using a distributed application health model in System Center Operations Manager. The F5 Management PRO Pack for SCVMM also includes support for Live Migration and other Enterprise Private Cloud scenarios.

Virtualization Support

It is important to note that Exchange Server 2010 is not "virtualization aware" [iii]. In testing, the hypervisor adds approximately 12 percent of processor overhead, which needs to be accounted for when sizing Exchange Server 2010 implementations.

In addition to providing availability, scalability, and performance improvements for Exchange Server 2010, BIG-IP LTM can further improve the efficiency of Exchange Server 2010 when deployed in a virtualized environment. The use of connection optimization features such as OneConnect™ in BIG-IP LTM improves

88 percent of IT organizations improved virtual machine density by 10 to 40 percent on a typical server with F5.

Source: TechValidate
TVID: 975-FFD-F8D

the efficiency of TCP connection management in Exchange Server 2010 and can increase the capacity of virtualized applications.

Virtual machine density improvements with F5

Using BIG-IP LTM optimization features can further improve the density of virtual machines deployed on a single, physical server by increasing efficiency and reducing the impact of the overhead associated with virtualization.

Deploying Exchange Server 2010 in a virtual environment does not change the architectural requirements in any way; load balancing for Client Access server roles deployed in multiple roles and in implementations of eight or more will still require hardware load balancing services, whether those servers are virtual or physical. BIG-IP LTM supports both virtual and physical deployments of Exchange Server 2010—as well as combinations thereof—with equal alacrity.

Conclusion

With the release of Exchange Server 2010, Microsoft has re-engineered the architecture of its enterprise-class email and communications services to better support scalability, reliability, and high availability. But these changes have consequences on existing installations, and Microsoft recommendations regarding the use of hardware load balancers have been made after extensive internal testing using a variety of high-availability techniques.

Migration of corporate mail services from one version of Exchange Server to another does not happen overnight. Maintaining two completely separate deployments is difficult enough without needing to potentially maintain multiple application delivery components (each with their own configuration and management needs) as well. Leveraging an F5 solution enables a simpler management and deployment infrastructure capable of simultaneously supporting both Exchange 2003/2007 and 2010 deployments during migration and enabling a smoother transition to a unified access and application delivery architecture that better supports the more unified Exchange Server 2010 architecture.

Microsoft IT has published its own architectural white paper describing how its teams architected and deployed a high-availability Exchange Server 2010 implementation leveraging hardware load balancing. The paper, "Exchange Server 2010 Design and Architecture at Microsoft: How Microsoft IT Deployed Exchange Server 2010," highlights the need for a robust Application Delivery

Controller in Exchange Server deployments that supports a variety of persistence methods across the different client access types.

In general, the addition of a load balancing solution might require some changes to network and application infrastructure. The F5 solution for Exchange Server 2010 helps make the implementation of a Microsoft-recommended compliant deployment as painless as possible by providing step-by-step guidance on an F5-tested configuration of all F5 solution components.

A complete load balanced F5 implementation supporting Exchange Server 2010 can enhance the performance, availability, reliability, and security of the organizational email infrastructure—protecting both capital and operational investments.

[i] http://www.ferris.com/2008/01/31/email-products-market-shares-versions-deployed-migrations-and-software-cost/

[ii] Getting the Most out of Microsoft Exchange Server 2010: Performance and Scalability [UNC309]

[iii] Microsoft Exchange Server Virtualisation: Does It Make Sense? [UNC03-IS]

http://www.slideshare.net/louisgohl/unc309-getting-the-most-out-of-microsoft-exchange-server-2010-performance-and-scalability

http://download.microsoft.com/download/8/5/D/85D61478-8719-4219-96BA-E5C53DD4F436/0941_ExchangeServer2010ArchitectureTWP.docx

F5 Networks, Inc.   401 Elliott Avenue West, Seattle, WA 98119   888-882-4447   www.f5.com