

EMC PERFORMANCE OPTIMIZATION FOR MICROSOFT FAST SEARCH SERVER 2010 FOR SHAREPOINT

EMC Symmetrix VMAX, FAST VP, Microsoft Hyper-V

- Optimize scalability and performance of FAST Search Server 2010 for SharePoint
- Validate virtualization of FAST Search infrastructure
- Scale-out architecture for enterprise configurations

EMC Solutions Group

October 2011



Copyright © 2011 EMC Corporation. All Rights Reserved.

EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

The information in this publication is provided “as is.” EMC Corporation makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

For the most up-to-date listing of EMC product names, see EMC Corporation Trademarks on EMC.com.

VMware, ESX, vMotion, VMware vCenter, and VMware vSphere are registered trademarks or trademarks of VMware, Inc. in the United States and/or other jurisdictions. All other trademarks used herein are the property of their respective owners.

Part Number H8801

Table of contents

Reference architecture overview	5
Document purpose	5
Solution purpose	5
The business challenge.....	5
The technology solution.....	6
Solution architectures	7
Overview of the solution architecture	7
Configuration of the SharePoint farm	7
Configuration with physical FAST Search server farm.....	7
Physical architecture diagram	8
Configuration with virtualized FAST Search server farm	10
Virtualized architecture diagram.....	10
Hardware and software resources	11
Hardware resources.....	12
Software resources.....	13
User load profile	13
Content load profile	14
Technology overview	15
Introduction	15
Microsoft FAST Search Server 2010 for SharePoint.....	15
Microsoft Hyper-V cluster	15
EMC Symmetrix VMAX	15
EMC Symmetrix FAST VP.....	15
EMC Replication Manager	15
Summary of test results.....	16
Overview of test results.....	16
Performance test results for the physical environment	16
Performance test results for the virtualized environment.....	16
Performance test results for EMC Symmetrix FAST VP	17
Replication Manager-engaged backup and restore for FAST Search Server 2010 for SharePoint	17
Conclusion	18
Summary	18
Results.....	18
Benefits	18
References.....	19

White papers	19
Product documentation.....	19
Other documentation.....	19

Reference architecture overview

Document purpose This document describes the reference architecture of an enterprise-scale SharePoint content farm and a file share with a unique data set for indexing and searching by a Microsoft FAST Search 2010 farm (with native Microsoft SharePoint enterprise search disabled) on EMC® Symmetrix® VMAX™ storage with Fully Automated Storage Tiering for Virtual Pools (FAST VP) enabled. By focusing on the scalability testing of Microsoft FAST Search, this document provides guidance for Microsoft Hyper-V virtual machine sizing and storage profiling. It also provides assessments of the impact of FAST Search on overall SharePoint performance.

This document is not a comprehensive guide to every aspect of this solution.

Solution purpose The purpose of this solution is to build and demonstrate the function, performance, and scalability of Microsoft FAST Search Server 2010 for SharePoint enabled by EMC Symmetrix VMAX and FAST VP.

This solution proves that the EMC Symmetrix VMAX array environment is scalable and capable of accommodating a physical FAST Search server. The solution provides an intelligent array that services continually changing storage I/O workloads with the highest levels of performance. This solution consolidates FAST Search with Hyper-V to virtualize the FAST Search servers. This solution also investigates where EMC FAST VP with Flash drives can provide its value based on the observed workloads, and showcases how local protection can be accomplished with EMC Replication Manager and EMC Symmetrix TimeFinder® for the FAST Search farm and SharePoint farm.

The business challenge Microsoft FAST Search Server 2010 for SharePoint offers a new choice for enterprise-level search. Combining the power of FAST with the simplicity of SharePoint, FAST Search Server 2010 for SharePoint delivers an exceptional intranet and people-search experience and a platform for building custom search-driven applications.

The deployment of FAST Search Server 2010 for SharePoint in the SAN environment requires very careful planning and an understanding of both SharePoint and FAST Search servers. In many cases, an extensive consulting engagement is required.

Virtualization of the FAST Search farm is often requested by customers, but there is a lack of guidelines. Questions such as “Is it possible?”, “How can it be implemented?”, and “What kind of performance can I expect?” remain unanswered.

The technology solution

This solution demonstrates how to use EMC Symmetrix VMAX with FAST VP technology to deploy an enterprise-class Microsoft FAST Search farm for a virtualized SharePoint 2010 farm. The FAST Search farm is first deployed as physical servers, and then it is changed to Hyper-V virtual machines.

This solution differentiates EMC from other storage and infrastructure vendors because it provides guidance about enterprise-class storage and server requirements.

In this solution, EMC Symmetrix VMAX FAST VP provides excellent support for the high demand of a FAST Search server as well as a SharePoint farm. Virtualization of both SharePoint and FAST Search server farms provides valuable insight into how this can be accomplished without sacrificing search performance while minimizing the impact on SharePoint.

The backup and restore capability enabled by Replication Manager and TimeFinder provides an easy way to maintain both farms.

Solution architectures

Overview of the solution architecture

This solution tests two configurations. One is a virtualized SharePoint 2010 farm with a physical FAST Search server farm; the other is a virtualized SharePoint 2010 farm with a virtualized FAST Search server farm.

For purposes of clarity, EMC refers to the configuration with the physical FAST Search server farm as the physical architecture and the configuration with the virtualized FAST Search server farm as the virtualized architecture in this document.

Configuration of the SharePoint farm

The SharePoint farm is a key component for both the physical and virtualized architecture. The configuration is as follows:

- Three physical hosts with Hyper-V enabled.
- All SharePoint servers are virtualized on the Hyper-V hosts:
 - One admin server.
 - One application server.
 - Four Web front-end servers.
 - Two SQL servers.
- SharePoint native enterprise search is disabled and the content SSA and query SSA are deployed as the content feeding and query delivering mechanisms for the FAST Search farm. The content SSA crawls the content and feeds it into the back end of FAST Search Server 2010 for SharePoint. The FAST Search query SSA provides query results from all content sources, which are the file share content and SharePoint content in this solution.

Configuration with physical FAST Search server farm

The physical architecture of a FAST Search server farm demonstrates the typical deployment of FAST Search server configuration.

The configuration of the FAST Search server farm is as follows:

- Five physical hosts including:
 - Two document processing servers, each with 12 document processors; one document processing server (with the primary Web analyzer) serves as content distributor and the other (with the secondary Web analyzer) serves as indexing dispatcher.
 - Two index and search servers, each with six document processors
 - One FAST administration server with 12 document processors.

The query SSA and content SSA servers for the FAST Search farm are deployed as virtual machines on the SharePoint farm Hyper-V hosts.

Physical architecture diagram

Figure 1 depicts the overall architecture of the validated solution environment with a physical FAST Search server farm.

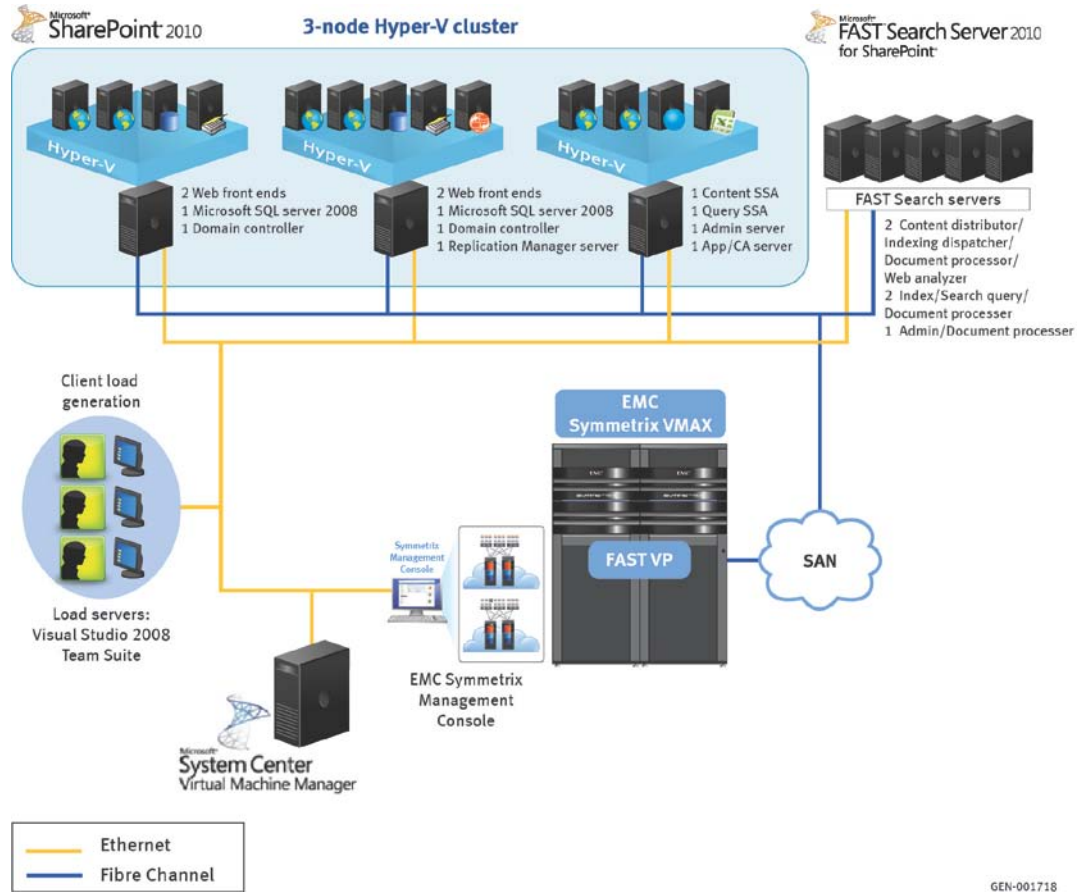


Figure 1. Physical architecture

Figure 2 shows the configuration of the physical FAST Search servers.

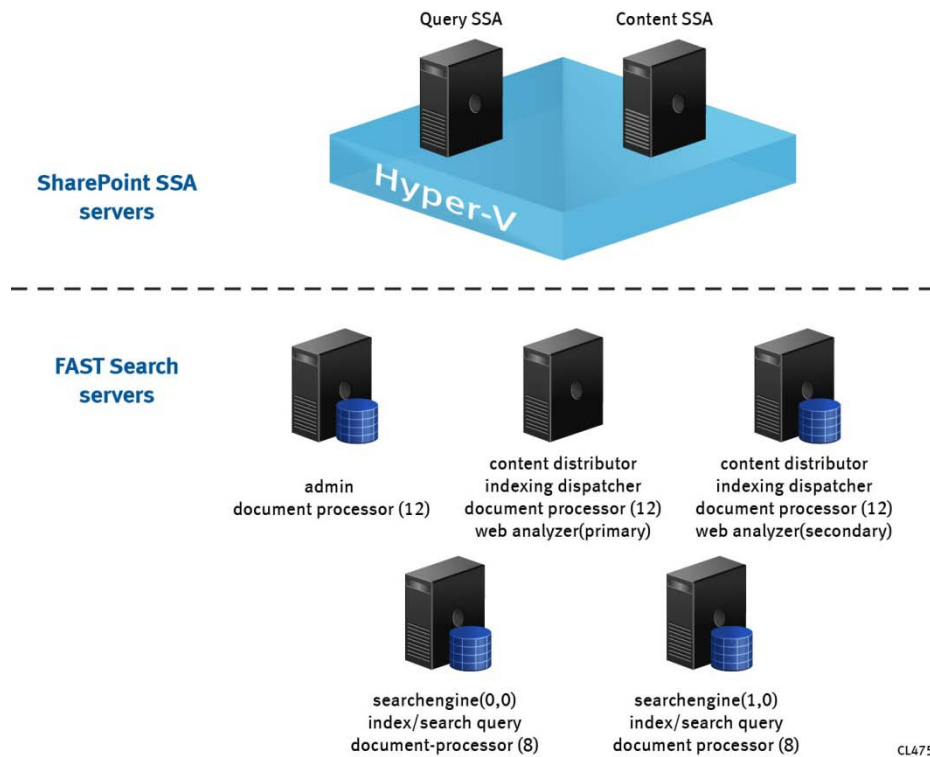


Figure 2. Physical FAST Search servers configuration

Configuration with virtualized FAST Search server farm

Before moving the FAST Search farm into the virtualized environment, the following factors need to be considered:

- **The CPU number limitation of the guest virtual machine**
The CPU number of the Windows 2008 R2 Hyper-V virtual machine is limited to four. It is suggested to set the CPU number for each FAST Search server virtual machine to four to get the best indexing and query performance in the virtualized environment.
- **FAST Search server license cost**
The license of the FAST Search Servers 2010 for SharePoint is based on the number of FAST Search servers. Adding more servers means higher cost. Similar numbers of FAST Search servers should be considered when moving from the physical to the virtualized environment for the FAST Search server farm.

The virtualized architecture of the FAST Search server farm demonstrates a virtualized FAST Search server configuration.

The configuration of the FAST Search server farm is as follows:

- Two physical hosts with Hyper-V enabled:
 - Two document processing virtual machines, each with eight document processors (Hyper-V virtual machine has a limit of four processors; virtual machines cannot scale more document processors).
 - Two index and search virtual machines.
 - One FAST administration virtual machine, with six document processors and a primary Web analyzer.
- The query SSA and content SSA servers for the FAST Search farm are deployed as virtual machines on the SharePoint farm Hyper-V hosts.

Virtualized architecture diagram

Figure 3 depicts the overall architecture of the validated solution environment with a virtualized FAST Search server farm.

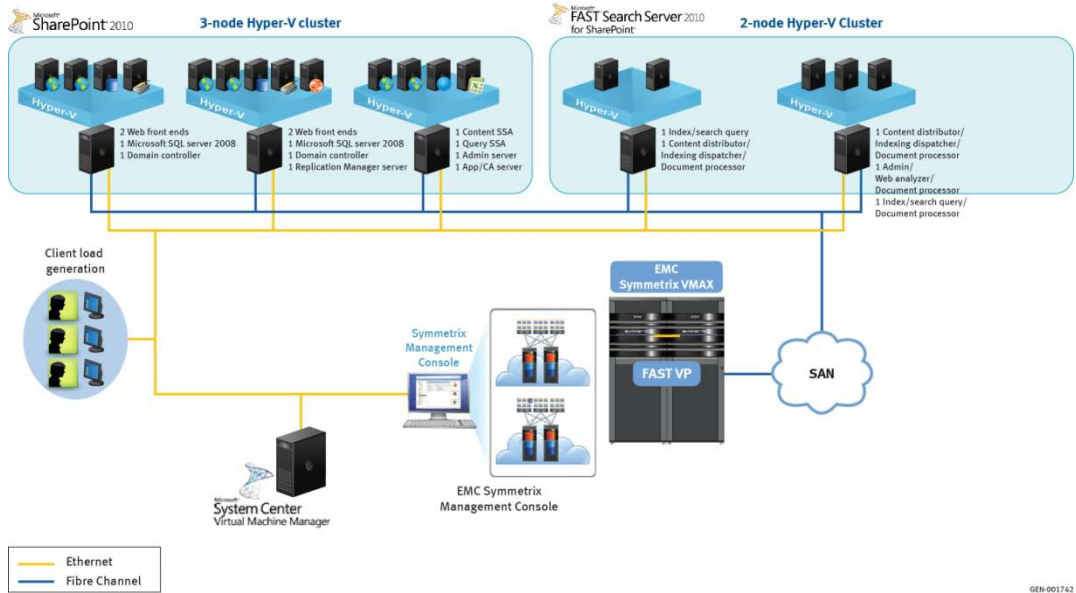


Figure 3. Virtualized architecture

Figure 4 shows the configuration of the virtual FAST Search servers.

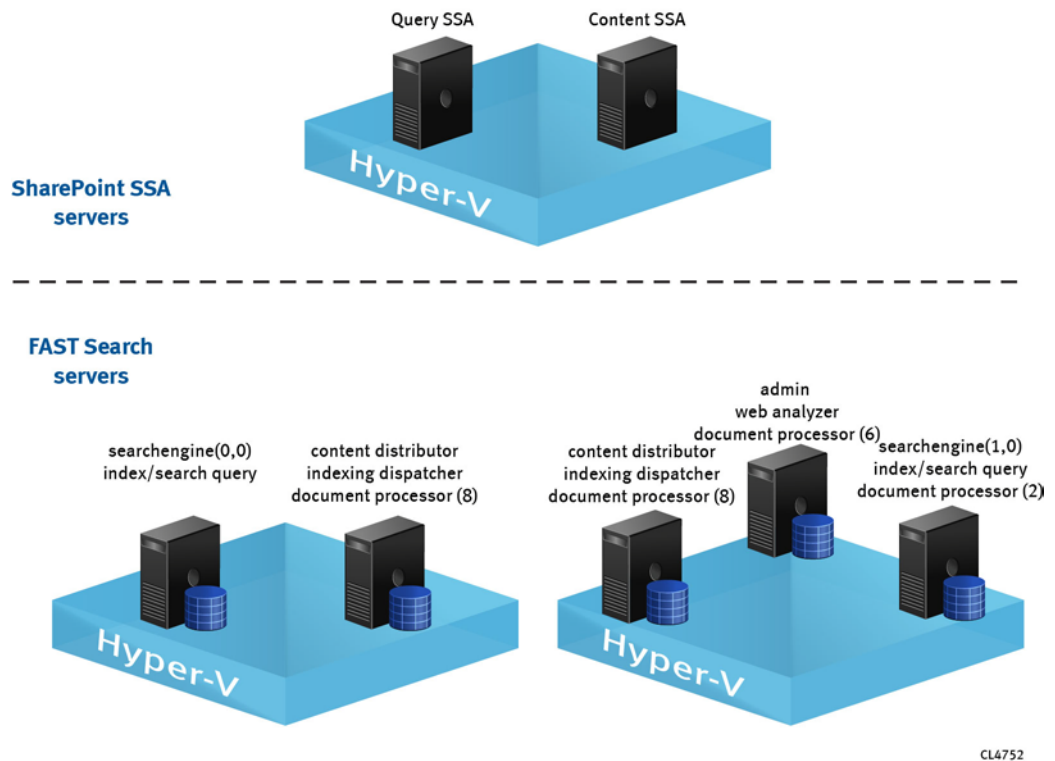


Figure 4. Virtual FAST Search servers configuration

Hardware and software resources

Hardware and software resources for both physical and virtual architecture are listed in the following tables.

Hardware resources

Table 1 lists the details of the hardware resources used in this solution.

Table 1. Hardware resources

Equipment		Quantity	Configuration	
Symmetrix VMAX		1	Four engines 32 front-end directors 32 back-end directors 942 disks of the following types: 200 GB Flash drives, 600 GB 10,000 rpm FC, 450 GB 15,000 rpm FC, 1 TB 7,200 rpm SATA Cache: 248 GB	
SharePoint Servers	Hyper-V hosts	3	Four sockets, quad cores, 128 GB memory	
	Virtual machines	12	2 x SQL server – 4 vCPU/32 GB 4 x WFEs – 4 vCPU/6 GB 2 x DC – 2 vCPU/4 GB 1 x APP/CA – 2 vCPU/4 GB 1 x Query SSA – 4 vCPU/8 GB 1 x Content SSA – 4 vCPU/8 GB 1 x Replication Manager server virtual machine – 2 vCPU/4 GB	
FAST Search servers	Physical		5	Two sockets, six cores, 48 GB 2 x Indexing/Search service 1 x FAST administrator 2 x Document processors with Web analyzers
	Virtual	Hyper-V hosts	2	Two sockets, six cores, 48 GB
		Virtual machines	5	2x Indexing/Search service 4 vCPU/16 GB 1x FAST administrator with Web analyzer and document processor 4 vCPU/16 GB 2x document processors 4 vCPU/8 GB
FC switches		2	8 Gb FC switches (minimum of 24 ports)	
Network switches		1	1 Gb IP switch (24 ports)	
Systems center virtual machine manager (SCVMM)		1	Quad cores/8 GB	

Software resources

Table 2 lists the software resources used in this solution.

Table 2. Software resources

Title	Number of licenses	Version
Microsoft Windows Server (Hyper-V)	5	2008 R2
Microsoft Windows Server	22	2008 R2
Microsoft SQL Server	2	2008 R2
Microsoft FAST Search server	5	2010 SP1
Microsoft SharePoint	8	2010 SP1
EMC Replication Manager	1	5.4
EMC PowerPath®	11	5.5
EMC Solution Enabler	12	7.3
EMC Symmetrix Management Console	1	7.2
Knowledge Lake Virtual Studio Team System (VSTS) test code	1	1.3

User load profile

During validation, EMC used a Microsoft heavy-user load profile to determine the maximum user count that the SharePoint 2010 Server farm could sustain while ensuring that average response times remained within acceptable limits. Microsoft standards state that a heavy user performs 60 requests per hour.

Table 3 shows the required response times for Microsoft SharePoint 2010 user operations.

Table 3. Required response times

Test type	Details	Required response time
Browse	User browse	Less than 3 seconds
Search	Unique value search	Less than 3 seconds
Modify	Browse and metadata modify	Less than 3 seconds

Table 4 shows the content load profile used during solution validation

Table 4. Content load profile

Item	Value
Content type	doc, docx, pdf, xls, xlsx, ppt, pptx, txt
Average content size	1.65 MB
SharePoint – Total data	4 TB (not externalized)
SharePoint – Document size range	200 KB - 50 MB
SharePoint – Number of documents	2,000,000
SharePoint – Total site count	200
SharePoint – Content database size (each)	Up to 200 GB
SharePoint – Total site collections	20
SharePoint – Sites for each site collection	10
SharePoint – Total user count	22,000
SharePoint – User concurrency	10 percent
File share – Total data	5 TB
File share – Number of documents	3,000,000

Technology overview

Introduction

This section introduces the following technology components that are used in this solution:

- Microsoft FAST Search Server 2010 for SharePoint
- Microsoft Hyper-V cluster
- EMC Symmetrix VMAX
- EMC Symmetrix FAST VP
- EMC Replication Manager

Microsoft FAST Search Server 2010 for SharePoint

Microsoft FAST Search Server 2010 for SharePoint provides significant enhancements for enterprise search capabilities.

FAST Search Server 2010 for SharePoint can be used to resolve various search problems. It provides an enhanced general productivity search and also improves high-value search applications. High-value search applications drive measurable returns on investment (ROI) by helping a specific set of people obtain the most from a specific set of information. Common examples include product support applications, research portals, and customer record locators.

Microsoft Hyper-V cluster

Microsoft Hyper-V on Windows Server 2008 R2 enables you to create a virtualized server computing environment. Hyper-V provides a software infrastructure and basic management tools that you can use to create and manage a virtualized server computing environment. This virtualized environment can be used to address business goals aimed at improving efficiency and reducing costs.

EMC Symmetrix VMAX

With the strategy of building simple, intelligent, and modular storage, EMC Symmetrix VMAX with Enginuity™ version 5875 incorporates a new, highly scalable Virtual Matrix Architecture™ that enables Symmetrix VMAX arrays to grow seamlessly and cost-effectively from an entry-level configuration into the world's largest storage system. Symmetrix VMAX offers more efficiency, scalability, and security.

EMC Symmetrix FAST VP

Fully Automated Storage Tiering for Virtual Pools (FAST VP) builds on and extends the capabilities of Virtual Provisioning to provide a new form of FAST for Symmetrix at the sub-LUN level. FAST VP can automatically plan and implement non-disruptive changes to storage allocations and optimize the use of a storage configuration composed of different drive technologies. The goals of the optimizations are to take full advantage of the low response times of the Flash drives and to provide the most cost-effective use of a mixed drive type configuration while providing the best performance.

EMC Replication Manager

EMC Replication Manager manages point-in-time replication through a centralized management console. Replication Manager coordinates the entire data replication process—from discovery and configuration to the management of multiple application-consistent, disk-based replicas. With Replication Manager, the right data can be put in the right place at the right time—on demand or based on schedules and policies that can be easily defined. This application-centric product provides simplified replica management with application consistency.

Summary of test results

Overview of test results

This solution proved that a FAST Search farm could be virtualized while keeping search performance comparable to a physical FAST Search farm. In testing, the crawl rate in the virtual FAST Search farm was acceptable with limitations caused by Hyper-V CPU constraints.

Symmetrix FAST VP automated the storage tiering and improved the search performance by 20 percent.

The FAST Search farm as well as the SharePoint farm can be fully protected with Replication Manager with minimum impact on the SharePoint farm. The restore and recovery operations for the whole FAST Search farm with 5 TB of storage took less than 15 minutes.

Performance test results for the physical environment

The following is a summary of test results for the FAST Search server farm in the physical environment:

- From a storage perspective, the disks and thin pools performed well. During the full crawl testing, the disk pool had an average of 60 percent utilization, with an 80 percent maximum. Total disk IOPS averaged 3,500 on the Symmetrix VMAX back end for full crawls and was around 1,900 for the query testing.
- The FAST Search crawl rate was 150 GB/hour with five servers, 60 available CPUs, and 52 document processors.
- Search response time was less than 1 second, supporting more than 22,000 users with 10 percent concurrent access.

Performance test results for the virtualized environment

The following is a summary of the test results for the FAST Search server farm in the virtualized (Hyper-V-enabled) environment:

- The storage performance for the virtualized environment was similar to the storage performance of the physical environment.
- The FAST Search crawl rate was 80 GB/hour with three Hyper-V servers, five virtual machines, 20 available CPUs, and 24 document processors.
- The search response time for the virtualized environment was the same as the search response time for the physical environment (less than 1 second).

The shift from a physical FAST Search Server for SharePoint 2010 to a virtualized computing infrastructure makes the management and deployment of FAST Search servers much easier while maintaining comparable search performance. With Symmetrix VMAX and FAST VP, virtualized FAST Search Server 2010 for SharePoint can have the expected crawl performance and can maintain similar search performance as physical FAST servers.

The number of document processors in the FAST Search server farm can have a major impact on the crawl performance. The SharePoint performance test is closely related to the number of Web front-end servers in the SharePoint farm. Compared with the physical FAST Search server environment, in the virtualized FAST Search environment:

- The number of physical servers reduced by 60 percent from five to two.

- The number of FAST Search servers remained the same as in the physical environment.
- The number of document processors reduced by approximately 58 percent, from 52 to 24, mainly because of the Hyper-V vCPU limits (four for each virtual machine).
- The full crawl performance was reduced by about 40 percent.
- The browse/search/modify test results were similar to the physical environment.

Performance test results for EMC Symmetrix FAST VP

Symmetrix VMAX FAST VP was enabled in the virtualized environment to improve the storage resource utilization:

- Symmetrix FAST VP automatically tiered the data for FAST Search volumes to the appropriate virtual pool during every query performance test session.
- Symmetrix FAST VP improved the search performance by 20 percent.

Replication Manager-engaged backup and restore for FAST Search Server 2010 for SharePoint

With a Replication Manager-engaged TimeFinder snap, the entire SharePoint farm, with a total of 9 TB of data, can be backed up in 15 minutes. The FAST Search server farm, with a total of 5 TB of FAST Search volumes, can be backed up in 8 minutes.

With a Replication Manager-engaged TimeFinder snapshot, the SharePoint content database (4 TB) and file share data (5 TB) can be restored and recovered in 90 minutes. The FAST Search server farm, with a total of 5 TB of FAST Search volumes, can be restored within 13 minutes.

After restoring the SharePoint content databases and FAST Search server farm, the incremental crawl rate was the same as the full crawl rate before the restore. The farms were healthy and working properly.

Conclusion

Summary

The Symmetrix VMAX array is highly scalable and easily capable of accommodating the storage performance and capacity requirements of Microsoft FAST Search server. It can fully support a FAST Search farm in a physical or virtualized (Hyper-V-enabled) environment.

Results

The following results were achieved in this solution:

- Designed a reference architecture for FAST Search servers on physical servers and on Hyper-V virtual servers running against a virtualized SharePoint farm.
- Validated the reference architecture and methodology against the solution objectives and search service-level agreement (SLA) goals.
- Proved that FAST Search servers can be virtualized with acceptable crawl performance and comparable query performance with reduced host server numbers and maintenance costs, while adding server-level resiliency with the Hyper-V cluster.
- Demonstrated the value of FAST VP, which automates the storage tiering and improves search performance.
- Profiled the impact of local replication by using Replication Manager and TimeFinder snap jobs to provide protection for both SharePoint and FAST Search farms. The user experience of the farm operation (browse/search/modify) remained the same with Replication Manager replication. The restore operation of both SharePoint content and FAST Search farm was simple and fast.

Benefits

This solution provides the following benefits:

- Symmetrix VMAX can fully support FAST Search Server 2010 for SharePoint in both physical and virtual environments.
- The moving of FAST Search servers into a virtual environment reduces the number of physical host servers and leverages all of the management advantages of a virtualized environment while achieving comparable query performance to a physical environment and acceptable crawl performance.
- FAST VP maintains or improves the query performance of FAST Search while reducing the total cost of ownership.
- EMC Replication Manager ensures that the environment is protected. The SharePoint content and FAST Search farm can be backed up and restored with minimal impact on end users.

References

White papers

For related information, see the white papers (available on emc.com and EMC Powerlink) listed below:

- *Reduce Costs and Optimize Microsoft SQL Server Performance in Virtualized Environments with EMC Symmetrix VMAX*
- *FAST VP for EMC Symmetrix VMAX—Theory and Best Practices for Planning and Performance*
- *EMC Symmetrix with Microsoft Hyper-V Virtualization*

Product documentation

For additional FAST Search information, refer to:

- *FAST Search Server 2010 for SharePoint Capacity Planning* (<http://www.microsoft.com/download/en/default.aspx>)
- *Deployment Guide for FAST Search Server 2010 for SharePoint* (<http://www.microsoft.com/download/en/default.aspx>)
- *FAST Search Server 2010 for SharePoint on TechNet* (<http://technet.microsoft.com/en-us/enterprisesearch/default>)

Other documentation

For additional SharePoint information, refer to:

- *Capacity Planning and Sizing for Microsoft SharePoint Products and Technologies* (<http://technet.microsoft.com/en-us/enterprisesearch/default>)