SharePoint 2010 Search Performance

Technical White Paper

May 2010

Lars Fastrup
Independent Consultant and Author

Abstract

Implementing a high-end search solution requires some planning and testing efforts to ensure good performance and availability. Without proper planning of a search deployment there is a greater risk that is will be undersized or oversized. This whitepaper provides an introduction to the aspects of deploying a sound SharePoint 2010 search deployment with respect to performance requirements. More detailed information can be found on Microsoft TechNet from the links in the Resources section.

Contents

Introduction

The SharePoint Server 2010 search system is a very scalable piece of search technology that can be scaled to handling about 100 million documents per index and to handling many concurrent queries. It also offers several levels of redundancy for high availability. However, implementing a high-end search solution requires some planning and testing efforts to ensure good performance and availability. Without proper planning of a search deployment there is a greater risk that is will be undersized or oversized. An undersized system will not be able to meet the user demand while an oversized system will cost more money and resources than strictly necessary. This paper provides an introduction to the aspects of deploying a sound SharePoint 2010 search deployment with respect to performance requirements. More detailed information can be found on Microsoft TechNet from the links in the Resources section at the end of this paper.

Performance Fundamentals

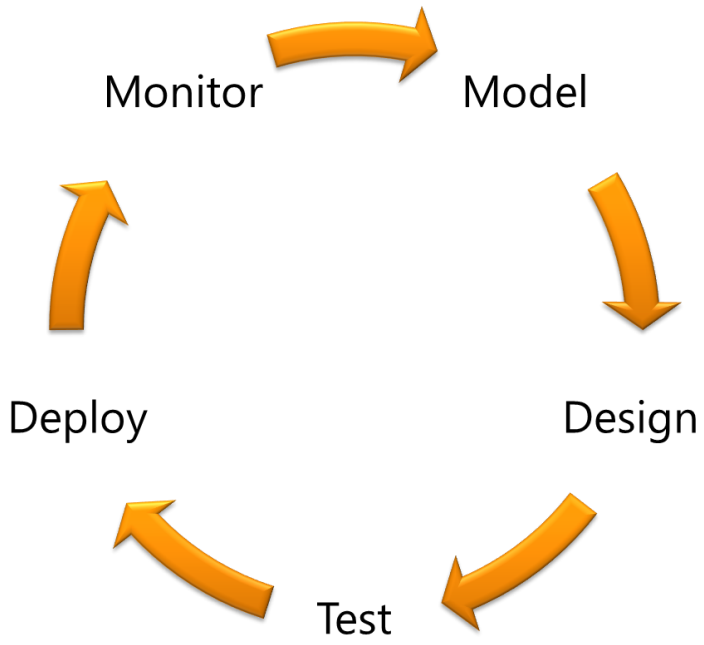
The metrics by which we define search performance generally fall into the two categories query performance and indexing performance:

- Query Performance
  - Query latency - The time it takes the search system to process a query and return the results.
  - Query rate - The number of concurrent queries that the search system can handle per second (QPS).
- Indexing Performance
  - Indexing latency - The time it takes the search system to index a document and make it available to search.
  - Indexing rate - The number of documents the search system can index per second.

Performance Management

Performance management is an extension of performance planning in the sense that a system might not always perform as planned. It can also be difficult to plan for performance as it depends on a multitude of factors like hardware, search topology, nature of the content to index, network infrastructure, etc. Consequently, it is also import to include sufficient testing and monitoring efforts to implement a sound search system. Performance management is an iterative process that includes five logical steps or phases as shown in Figure 1.

Figure 1: Performance Management Model.



Search System Modeling

This is the first phase in the performance management cycle where you analyze and collect the requirements for the search system that you are planning to deploy. Important considerations include:

- Estimate the expected workload and dataset:
  - Number of items in the search index.
  - Data change rate; no need to index every 10 minutes if content is mostly static.
  - Average and peak number of queries per second.
- If possible collect usage data from legacy search system to help estimate the dimensions of the new SharePoint 2010 search system.
- Analyze bandwidth and responsiveness of all content sources – how fast can the SharePoint crawlers access and download content for indexing? This is particularly important to know when you are planning to index external content sources like Windows file shares, external Web sites and third-party repositories like Documentum and FileNet.
- Setup performance targets for:
  - Average query latency at the max. estimated query rate.
  - Average time for new and modified documents to appear in search results (index freshness)
- Specify reliability targets (e.g. 99.99% uptime = 52.6 min downtime / year):
  - Specify overall uptime of search databases
  - Specify overall uptime of query servers
  - Specify overall uptime of index servers

Search System Design

The next phase in the performance management cycle is the design of an architecture that is likely to meet the requirements established in the modeling phase. Consider the following aspects in your design:

- Design a starting point architecture that:
  - Is likely to meet your performance and availability targets for the expected demand.
  - Is scalable by adding more hardware and/or more search components
  - Does not violate any SP2010 boundaries
- Use Microsoft case studies as a reference.
- Consider virtualization of search servers.
- Select appropriate hardware.

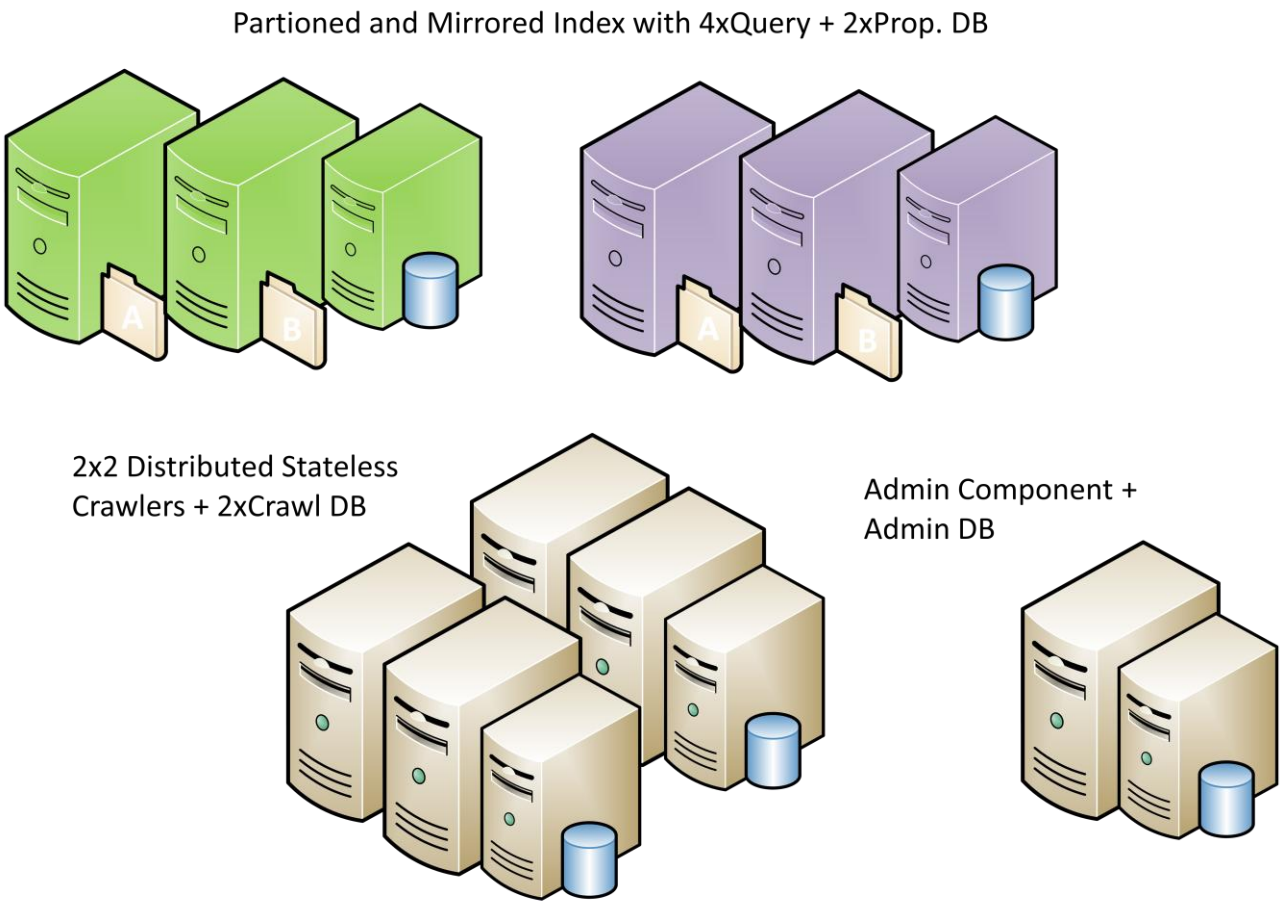
SharePoint 2010 Search Architecture

The SharePoint Server 2010 search engine introduces a new and highly componentized deployment architecture that is very flexible and highly scalable. The available components that you must learn how to deploy include an Administration Component, Administration Database, Query Components, Crawl Components, Property Databases and Crawl Databases. This componentization of the search engine offers the following features and benefits:

- **Index Partitioning** enabling a search index to be partitioned across multiple query servers, which will in turn work in parallel on delivering the results for each query. This enables deployment architectures with sub-second query latency up to about 100 million indexed items.
- **Index Mirroring** enabling query failover by cross mirroring the search index on the query servers (passive mirroring) or mirroring it to a parallel set of query servers (active mirroring).
- **Multiple Stateless Crawlers** offering good crawl performance and high availability of crawls. Stateless refers to the fact the crawlers are redundant and they do not keep a copy of the index on the server as was the case with the index server in MOSS 2007. Consequently, crawlers have a low disk space requirement.
- **Multiple Crawl Databases** for scalable crawl performance. Supports native SQL mirroring for failover.
- **Multiple Property Databases** for scalable query performance. Supports native SQL mirroring for failover.

Figure 2 shows a sample deployment with a partitioned and mirrored search index, multiple property databases, multiple crawlers and multiple crawl databases.

Figure 2: Sample SharePoint 2010 Search deployment.



The sample architecture in Figure 2 is good for indexing and querying a dataset of up to 20 million documents with sub-second query latency and good crawl performance. The exact crawl rate is hard to predict here as it depends on many factors like document type distribution, average document size, network infrastructure and hardware specifications of the crawl servers and crawl database servers. Also, there are no good data available at the time of this writing.

Search Application Boundaries

A SharePoint Server 2010 farm can host one or more search service applications where 20 is the recommended maximum. Each search application has its own physical search index and it is recommended that it stays within the boundaries listed in table 1.

Table 1: Software boundaries that applies to each Search Application.

Limit	Maximum value
Number of index partitions	20
Number of items per index partition	10 million
Number of items across all index partitions	100 million

Number of crawl components	16
Number of crawl databases	10
Number of property databases	10
Number of query components	64
Number of content sources	50
Number of crawl rules	100

Scale-out Decision Points

The following tables provide some scalability insight based on the number of documents that you plan or expect to index, and other important metrics.

Table 2: Scalability based on number of items in the search index.

# of items in search index	Actions
0-1 million	All Search components can coexist on one or two servers.
1-10 million	Move crawl components to a dedicated server; keep the query component on the Web server.
10-20 million	Add another crawl server, split index into two partitions distributed across two query servers.
20-40 million	Add query servers and index partitions. Add another crawl database, associated with a new crawler on each crawl server. Also, add another property database to a second database server, and mirror both property databases.
40-100 million	Isolate each topology layer into "server groups" in which each role is deployed to its own servers. Each server group can then be scaled out to meet specific requirements for the components in that role.

Table 3: Scalability based on other important metrics.

Metric	Actions
Crawl rate	To improve index freshness, add crawl servers, crawlers and crawl databases.
Query latency	To improve query response time, add query servers and index partitions. If poor query response time is caused by database load, move property database to a dedicated database server or optimize existing server if possible.
Availability of query functionality	Deploy redundant query servers, multiple query components for each index partition, and use clustered or mirrored database servers to host crawl and property databases.
Availability of indexing functionality	Use multiple crawlers on redundant crawl servers, and add crawl databases. Crawlers associated with a given crawl database can be distributed across crawl servers for availability and load distribution.

Storage Architecture

The SharePoint Server 2010 search system is a very disk intensive application in terms of throughput, especially in scenarios with many concurrent users and a large non-static volume of content to index. The crawl databases and the property databases of a high-end search deployment will depend on a high performance storage system to avoid unnecessary bottlenecks. The best medicine for performance bottlenecks in the search databases is a sound storage architecture where each database ideally has its own spindles in a RAID10 disk array to work with. However, having many separate disk arrays does not come cheap. Use the following priorities to decide in which order to separate the search databases to their own disk arrays.

Table 4: Prioritization of the separation of search databases.

Priority	Database File	RAID	Optimization
1	Temp Database Data	RAID 10	Write
2	Temp Database Log	RAID 10	Write
3	Crawl Database Log	RAID 10	Write
4	Crawl Database Data	RAID 10	Read / Write
5	Property Database Log	RAID 10	Write
6	Property Database Data	RAID 10	Read / Write

Hardware Requirements

The following table outlines the required and recommended hardware specifications for the different components in a search deployment.

Table 5: Required and recommended search deployment hardware specifications.

Component	Required	Recommended	Virtual
Database Server	64-bit CPU, 4 cores 8 GB RAM	64-bit CPU, 8-16 cores 32GB RAM Disks in RAID 5 for system files Fast disks in RAID 10 for search databases	Not recommended
Index Server	64-bit CPU, 4 cores 8 GB RAM	64-bit CPU, 8-16 cores 16-32GB RAM Disks in RAID 5	Ok
Query Server	64-bit CPU, 4 cores 8 GB RAM	64-bit CPU, 4-8 cores 16GB RAM Disks in RAID 5	Ok

Search System Testing

The third phase in the performance management cycle is testing. This is the point where you should test and optimize the search architecture before it is deployed to production. Consider the following aspects in your work:

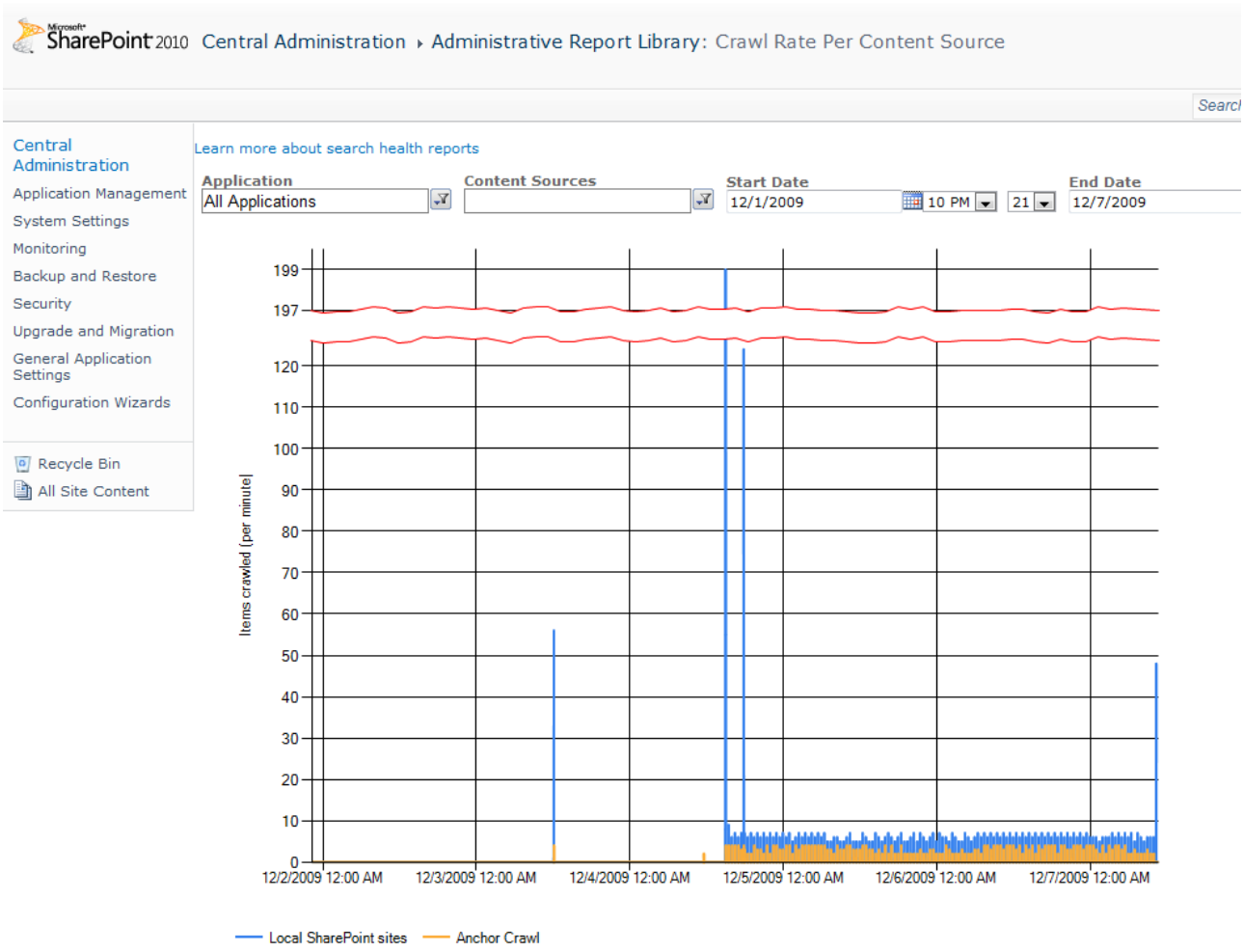
- Create a repeatable test plan.
- Create a test environment from the architecture you have designed.
- Migrate or replicate existing content to the test environment for realistic testing of indexing performance. If applicable, setup indexing of external content as well.
- Use load test tools to simulate expected demand and to automate test procedures, e.g. the upcoming SharePoint Load Test Kit 2010 and Visual Studio 2010 Ultimate.
- Analyze results, identify bottlenecks and optimize the architecture. Retest if needed.

Search System Monitoring

Even with near perfect planning and testing you will often still meet unexpected conditions in a production environment as well as changing demand on the search system. Consequently, it is also very important to monitor your search deployment very closely and improve it accordingly.

- Monitor the live search system to
    - Identify and address performance bottlenecks
    - Address changes in demand like increasing size of search index and increasing query latency
- Use the crawl log to verify that all content is successfully added to the index.
- Use performance counters to monitor CPU, Memory, Network and Disk usage.
- Use search administration reports to monitor crawl rate and query latency.
- Use the upcoming SPDiag 2010 tool to monitor system health and usage.

Figure 3: Monitoring the crawl-rate using the search administration reports.


SharePoint 2010
Central Administration ▸ Administrative Report Library: Crawl Rate Per Content Source
Central Administration
Application Management
System Settings
Monitoring
Backup and Restore
Security
Upgrade and Migration
General Application Settings
Configuration Wizards
Recycle Bin
All Site Content
Learn more about search health reports
Application
All Applications
Content Sources
Start Date
12/1/2009
10 PM
21
End Date
12/7/2009
Items crawled (per minute)
199
197
120
110
100
90
80
70
60
50
40
30
20
10
0
12/2/2009 12:00 AM
12/3/2009 12:00 AM
12/4/2009 12:00 AM
12/5/2009 12:00 AM
12/6/2009 12:00 AM
12/7/2009 12:00 AM
Local SharePoint sites
Anchor Crawl

Resources

Please consult the following resources on Microsoft TechNet for more in-depth information about performance and capacity management for SharePoint Server 2010:

Capacity Planning and Sizing for Microsoft SharePoint 2010 Products and Technologies

SharePoint Server 2010 Software Boundaries

United States
445 Hamilton Avenue, SUITE 1102
White Plains, New York 10601
United States of America
Tel | 914.220.8395

Europe
Vester Voldgade 106,
DK-1552 København V
Denmark
Tel | +45 2161 6500

Web | www.BA-Insight.net
Email | info@ba-insight.net