Internet Sites Search Architectures

for SharePoint Server 2016

Overview

This model illustrates a medium-size Internet Sites farm architecture. Architecture requirements can vary depending on the composition of the data that is crawled (size of items and formats). This example illustrates the type of search components needed and how many of each. Use this example as a starting point for planning your own search environments. For more information about search processes and how search components interact, see Search Architectures for SharePoint® Server 2016 (https://go.microsoft.com/fwlink/p/?linkid=832554).

Search components

Index

Index component
The index component is the logical representation of an index replica.

Index partitions
- You can divide the index into discrete portions, each holding a separate part of the index.
- An index partition is stored in a set of files on a disk.
- The search index is the aggregation of all index partitions.

Index replicas
- Each index partition holds one or more index replicas that contain the same information.
- You have to provision one index component for each index replica.
- To achieve fault tolerance and redundancy, create additional index replicas for each index partition and distribute the index replicas over multiple application servers.

Query processing

Query processing component
Analyzes and processes search queries and results.

Admin

Search administration component
Runs system processes that are essential to search. There can be more than one search administration component per Search service application, but only one component is active at any given time.

Crawl

Crawl component
Crawls content based on what is specified in the crawl databases.

Content processing

Content processing component
Carries out various processes on the crawled items, such as document parsing and property mapping.

Analytics

Analytics processing component
Carries out search analytics and usage analytics.

Search databases

Search admin DB

Search administration database
Stores search configuration data. Only one search administration database per Search service application.

Crawl DB

Crawl database
Stores the crawl history and manages crawl operations. Each crawl database can have one or more crawl components associated with it.

Link DB

Link database
Stores the information extracted by the content processing component and also stores click-through information.

Analytics DB

Analytics reporting database
Stores the results of usage analytics.

Performance considerations

Application servers hosting search components

THINGS TO CONSIDER	REASONING
Cache	The query and its results are cached with Windows Server AppFabric, in key-value pairs: the query being the key and the results being the value. For each query there is an approximate 50% cache ratio. This means that if you have a usage pattern of 200 queries per second, about 100 queries will be sent to the search index and the other 100 queries will be cached. Results from the cache have lower query latency than those from the search index. For example, results for front-page queries that are run repeatedly are likely to be cached.
Continuous crawl	We recommend that you enable continuous crawl with an interval of 1 minute instead of the default interval of 15 minutes. You can enable continuous crawl on SharePoint content sources only.
Anonymous access	With anonymous access, users don't have to use credentials to log into a SharePoint Internet site. In addition, anonymous queries are cached, so they are cheaper because of lower query latency. You must enable anonymous access in two locations: on the web front-end and on the site itself.
Query latency	Query latency is influenced by caching, anonymous access, and by other factors such as the number and complexity of query rules that are applied and triggered. Also, consider the disks on which the search index is stored; a disk that has multiple spindles can improve the access speed of the disk and reduce query latency.

Search topologies for Internet Sites

Topologies for Internet Sites differ from other search topologies in the following ways:
- Web servers are incorporated in the topology for Internet Sites to serve user requests. This is different in search topologies for Enterprise Search where Web servers are not part of the dedicated search farm.
- The query processing component and index components reside on Web servers to make maximum use of the available hardware resources and to simplify scaling out the search topology.

Hardware requirements

These requirements apply to the medium Internet Sites search topology. The minimum listed RAM requirements for a server that hosts a search component is the total required amount of RAM for that server. For example, if you are hosting a content processing component, a search administration component and a crawl component on one server, the total amount of minimum required RAM for that server is 24 GB.

Minimum hardware requirements for application servers

SEARCH COMPONENT ON THE PHYSICAL SERVER	RAM	HARD DISK	PROCESSOR
Index component	48 GB for each server in the farm that hosts an index component, a query processing component, and the Web front-end.	500 GB additional disk space, preferably a separate disk volume/partition.	ALL COMPONENTS: 64-bit, 4 cores minimum, 8 cores recommended.
Analytics processing component	24 GB for each server in the farm that hosts an analytics processing component, a crawl component, and/or a search administration component.	300 GB additional disk space, preferably a separate disk volume/partition. This disk space is necessary for local processing of analytics data before it is written to the analytics reporting database.	
Crawl component		80 GB for your system drive. You must have sufficient space for the base installation and sufficient space for diagnostics such as logging and debugging, for creating memory dumps, and for other operations. For production use, you also need additional free disk space for regular operations. Maintain twice as much free space as you have RAM for production environments.	
Content processing component	See the requirements listed for the analytics processing component.		
Query processing component	See the requirements listed for the index component.		
Search administration component	See the requirements listed for the analytics processing component.		

Minimum hardware requirements for database servers

HARDWARE COMPONENT	REQUIREMENTS
Processor	64-bit, 4 cores for small deployments
RAM	8GB for small deployments
	16GB for medium deployments
Hard Disk	80GB for system drive

Medium Internet Sites farm (~85 Page views per second)

This farm is intended to provide a fault-tolerant SharePoint Server 2016 search farm topology that is optimized for a corpus that contains 3,400,000 items. The example farm processes 100-200 documents per second, depending on the language, and it accommodates 85 page views per second and 100 queries per second.

This topology is optimized for physical hardware, but you can deploy it on virtual machines as well.