Discover Technologies

Search Scalability in SharePoint 2013

Author: Natalya Voskresenskaya
SharePoint Architect

Date: June 2015

Contents

Discover Technologies

Search Scalability in SharePoint 2013

SharePoint 2013 has evolved into a fully matured content, records and documents management system. One of the key pillars of the SharePoint platform is search. It is pervasive throughout the SharePoint platform and many key components in SharePoint 2013 are powered by search, with the inclusion of several of the out-of-the-box web parts that are fully dependent on the search to provide an enriched user experience.

Enterprise search was enhanced with the release of SharePoint 2010 and was deployable using two options, SharePoint 2010 search with 100 million documents limit and FAST Search for SharePoint 2010 as a high-end Enterprise Search. With the release of 2013, the FAST search technology has been fully integrated with the default search experience so customers no longer have to decide which deployment option will meet their needs. Microsoft has taken the major features of FAST Search for SharePoint 2010 and integrated them natively with Enterprise search. This means that core features in relation to scalability and functionality from the FAST Search platform, are now built into SharePoint 2013 Search.

To increase the flexibility and scalability of search, Microsoft has separated the features into three core components which allow organizations to scale as needed to meet user demands.

- **Crawl**
  Crawl component is responsible for discovering items that need to be searchable, processing the content and metadata and passing content to the indexer component.
- **Index**
  The index component receives the processed items from the content processing component that is part of the crawler system and writes them to the search index. Index also handles incoming queries, retrieves information and sends results to the query processing component.
- **Query**
  The query processing component analyzes and processes incoming search queries from the Web front end and result sets sent from the index component.

When determining the initial search architecture or the scalability strategy, the following dimensions of search need to be taken into consideration:

Crawl Load

Crawl load is determined by the index freshness requirement and the content volume. The index freshness is defined by the time it takes for newly created or updated content to become searchable. SharePoint Search 2013 includes a few major improvements from its predecessor that are extremely important improvements that eliminate limitations that previously existed.

The crawl component is now split into 2 different processes, one being responsible for discovering or crawling the content, while the other component performs content processing which includes metadata extraction and property assignments. The crawler identifies changes to the content by comparing the latest crawl timestamp, which lives in the Crawl Databases, to the event cache table of the associated content database. Once the changes are returned from the system, the crawler passes them to the content processor and repeats the process again.

In the previous version of SharePoint, there was a one to many relationship, which would allow you to associate on crawl database with many crawl components. This architecture presented a number of issues that could potentially degrade system performance due to contention for database resources. SharePoint 2013 has removed the crawl component to crawl db limitation. You can

now scale out crawl components to increase the number of documents processed per second and multiple databases can be added to distribute the crawler read/write database activity. Now, crawl components will also automatically communicate with all crawl databases, if there is more than one. These improvements mitigate the following issues which were commonly seen in SharePoint 2010:

- Extensive communications to the crawl database, causing numerous database locks leading to degraded SQL performance.
- Skewed database sizes due to the unique relationship with the crawl database and crawl component mapping, which did not allow the crawl components to distribute their entries across different crawl databases.

The crawl component and databases architecture allows organizations to easily scale the architecture out to handle large volumes of content. By adding more crawl components, organizations will increase the number of documents crawled and processed simultaneously which is critical when it comes to crawling large bodies of content.

Microsoft recommends one crawl database for every 20 million items crawled. For example if the requirement is to crawl about 60 Million items, 3 crawl databases will be required.

The crawl component often uses a high percentage of the available processor resources, so consideration should be given to increasing CPU cores or adding additional crawl components when the CPU becomes a bottleneck.

Content Volume and Index

Index component is logically located in between the crawler and query components. It receives data from the content processing component and adds it to the index where it's later retrieved, to query component's request.

Index Component

SharePoint 2013 Search Indexes consist of partitions, where each partition represents a complete copy of the index. Partitions also can be mirrored to represent a complete replica of other partitions for failover purposes. Microsoft recommends that each index partition (one partition or replica per server) can hold up to 10 million items. When the number of items in one partition exceeds the 10 million items, you might start seeing index and query latency issues.

Increasing a number of partitions also accommodates a high demand on query size and results. You can create multiple index partitions on the same server but the performance will still be bound by the actual server resources. Other factors to consider are the network load between the query components and the index.

Query Volume

Since Search is a major component behind the scenes for additional functionality within SharePoint 2013, query volume is not only defined by how many queries end-users execute while searching for information, it also includes the number of queries and result set volumes that the SharePoint platform issues and receives to satisfy the end user demand on other functionalities driven by search. Regularly analyzing query administration reports and other search activity reports is crucial in determining the actual query volume to ensure your system is appropriately sized.

In reference to scaling out query components, the number of CPU cores should be taken into consideration, one CPU core can serve up to four queries per second. Adding additional query components to other servers for high availability, as both the primary

and the replica partitions allow you to can increase performance. The number of query components should be increased as the index grows, the smaller index partitions are, the more queries per second you can expect.

SharePoint 2013 collects search performance measurements in the Crawl Health Reports and Query Health Reports. These reports are useful in determining if the search needs to be scaled out and which dimension of search need to be scaled.

To reduce the processing time for queries, consider adding more partitions to the index. More partitions will create smaller size partitions and each partition would respond faster to queries. Creating too many partitions will not produce the best performance though, as the query processing component has to merge the responses from each partition to produce a result set, a merge takes more time when the index has more partitions. When you add more partitions on a running installation, the index repartitions itself.

Microsoft had invested in building an Enterprise search platform with the acquisition of FAST Search and initially introduced this highly scalable platform as an ad-on product for SharePoint 2010. Search within SharePoint 2013 is featuring full integration of FAST and true scalability of Enterprise search.

For information on Discover Technologies products and services, contact us. We'd love to hear from you!

***Phone**: 800-368-2344*
***Office**: 13241 Woodland Park Road, Suite 610, Herndon, VA 20171*

***E-mail (General Inquiries)**: info@discovertechnologies.com*
***E-mail (Sales Inquiries)**: Brian Vernon, bvernon@discovertechnologies.com*
***Web**: DiscoverTechnologies.com*