

Search Architectures for SharePoint 2016

Overview

The search architecture in Microsoft® SharePoint® Server 2016 consists of components and databases that work cohesively to perform the search operation. All components reside on application servers and all databases reside on database servers.

Search component interaction

Crawl and component processes

The crawl and content processing architecture includes the **crawl component**, **crawl database** and **content processing component**. Both search components can be scaled out based on crawl volume and performance requirements.

- 1 About the **crawl component**
 - The crawl component is responsible for crawling content sources. It delivers crawled items – both the actual content as well as their associated metadata – to the content processing component.
 - The crawl component invokes connectors or protocol handlers that interact with content sources to retrieve data. Multiple crawl components can be deployed to crawl simultaneously.
 - The crawl component uses one or more crawl databases to temporarily store information about crawled items and to track crawl history.
- A About the **crawl database**
 - The crawl database contains detailed tracking and historical information about crawled items.
 - This database holds information such as the last crawl time, the last crawl ID and the type of update during the last crawl.
- 2 About the **content processing component**
 - The content processing component is placed between the crawl component and the index component. It processes crawled items and feeds these items to the index component.
 - The content processing component transforms crawled items into artifacts that can be included in the search index by carrying out operations such as document parsing and property mapping.
 - Both the content processing component and the query processing component perform linguistics processing. Examples of linguistics processing during content processing are language detection and entity extraction.
 - The content processing component writes information about links and URLs to the link database.

Index and query processes

The index and query architecture includes the **index component**, **index partition**, and **query processing component**, all of which can be scaled out based on content volume, query volume, and performance requirements.

- 4 About the **index component**
 - An index component is the logical representation of an index replica. In the search architecture, you have to provision one index component for each index replica.
 - The index component receives processed items from the content processing component and writes those items to an index file.
 - The index component receives queries from the query processing component and provides results sets in return.
 - Queries are sent to the index replicas through the query processing component. The system routes and load balances the incoming queries to the index replicas.

About the index partition

- An index partition is a logical portion of the entire search index. The search index is the aggregation of all index partitions.

5 About the query processing component

- The query processing component is between the search front-end and the index component.
- The query processing component analyzes and processes search queries and results.
- Both the query processing component and the content processing component perform linguistics processing. Examples of linguistics processing during query processing are word-breaking and stemming.
- When the query processing component receives a query from the search front-end, it analyzes and processes the query to attempt to optimize precision, recall, and relevancy. The processed query is then submitted to the index component.
- The index component returns a result set based on the processed query back to the query processing component, which in turn processes that result set before sending it back to the search front-end.

Search administration

Search administration is composed of the **search administration component** and its corresponding database.

- 6 About the **search administration component**
 - The search administration component is responsible for running a number of system processes that are essential to search.
 - This component carries out provisioning, which is to add and initialize additional instances of the other search components.
- D About the **search administration database**
 - The search administration database stores search configuration data, such as the topology, crawl rules, query rules, and the mappings between crawled and managed properties.

Server roles

Web server

- Hosts Search Web Parts and Web Part pages for answering search queries.
- In dedicated search farms for Enterprise Search, this role is not necessary because Web servers at remote farms contact servers that serve queries directly.
- This role is necessary for farms that include other SharePoint Server 2016 capabilities.
- In small farms, this role can be shared on a server with the application server role.

Application server with search components

- Hosts all of the search components if only one server is configured. Otherwise, it holds components associated with the server, as configured by the administrator.
- Holds the entire search index if only one index partition is configured. Otherwise, it holds portions of the index that are associated with the index partitions as configured by the administrator.
 - The query processing component routes incoming queries to index replicas.
 - Each index replica is an index component.
 - At least one index partition must be configured per farm.
- At least one of each search component must be configured per farm.
- You cannot have multiple search components of the same type on one application server.
- Add search components on separate servers to provide redundancy.

Database server

- Hosts search databases.
- Can host other SharePoint Server 2016 databases.
- Can be mirrored or clustered.
- To increase performance and capacity, consider adding disks to the database server or adding database servers (depending on the bottleneck).

Search components

None

In SharePoint Server 2016 for Enterprise Search architectures, search components are not hosted on Web servers.

This is not the case with Internet Sites search architectures. Here, 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.

Index

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

- 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.

Query Processing

Query processing component
Analyzes and processes search queries and results.

Admin

Search administration component
Runs system processes that are essential to search.

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 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.
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

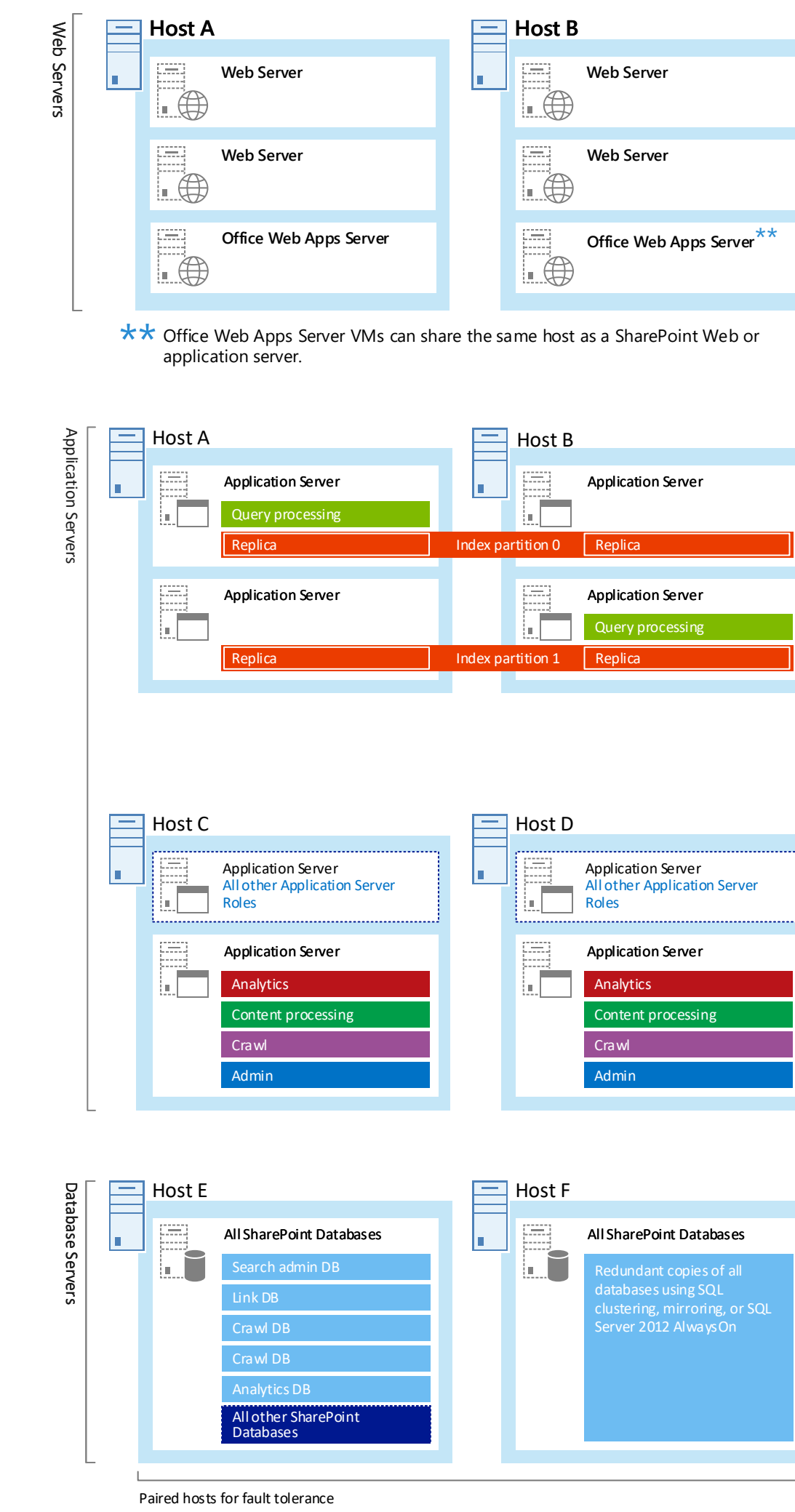
Analytics reporting database
Stores the results of usage analytics.

Example search topology

All-purpose fault tolerant farm for Enterprise Search

This farm is intended to provide a fully fault-tolerant, virtual environment for SharePoint Server 2016 including search. This illustration is an example of a medium Enterprise Search farm with approximately 80 million items in the search index. The hosts in this farm use eight CPU cores per virtual machine.

Note: This example does not apply to search topologies for Internet Sites.



Analytics processes

The analytics architecture consists of the **analytics processing component**, **analytics reporting database** and **link database**.

- 3 About the **analytics processing component**
 - The analytics processing component performs two types of analyses: search analytics and usage analytics. This component uses information from these analyses to improve search relevance, create search reports, and generate recommendations and deep links.
 - Search analytics is about extracting information such as - links, the number of times an item is clicked, anchor text, data related to people, and metadata – from the link database. This information is important to relevance.
 - Usage analytics is about analyzing usage log information received from the front-end via the event store. Usage analytics generates usage and statistics reports.
 - The results from the analyses are added to the items in the search index. In addition, results from usage analytics are stored in the analytics reporting database.
- B About the **link database**
 - The link database stores information extracted by the content processing component. In addition, it stores information about search clicks; the number of times people click on a search result from the search result page. This information is stored unprocessed, to be analyzed by the analytics processing component.
- C About the **analytics reporting database**
 - The analytics reporting database stores the results of usage analytics. In addition, the analytics reporting database also stores statistics information from the analyses. SharePoint uses this information to create Excel reports that show different statistics.
- E About the **event store**
 - The event store holds usage events that are captured on the front-end, such as the number of times an item is viewed.
 - These usage events are stored as log files on the application server that hosts the analytics processing component.

Illustration of search component interaction

