

Архитектуры поиска для SharePoint 2016

Общие сведения

Архитектура поиска в Microsoft® SharePoint® Server 2016 состоит из компонентов и баз данных, которые согласованно работают для выполнения операций поиска. Все компоненты располагаются на серверах приложений, а базы данных — на серверах баз данных.

Взаимодействие с компонентом службы поиска

Процессы обхода и компонентов

Архитектура обхода и обработки контента включает **компонент обхода**, **базу данных обхода** и **компонент обработки контента**. Оба компонента поиска можно масштабировать в зависимости от объема обхода и требований к производительности.

- О компоненте обхода**
 - Компонент обхода отвечает за обход источников контента. Он передает элементы (как фактический контент, так и связанные с ним метаданные), обход которых выполнен, компоненту обработки контента.
 - Компонент обхода вызывает соединители и обработчики протоколов, взаимодействующие с источниками контента для получения данных. Вы можете развернуть несколько компонентов обхода, выполняющих обход одновременно.
 - Компонент обхода использует одну или несколько баз данных обхода для временного хранения сведений об элементах и отслеживания журнала обхода.

- О базе данных обхода**
 - База данных обхода содержит подробные сведения об отслеживании и данные журнала элементов, для которых выполнен обход.
 - В этой базе данных хранятся такие сведения, как время последнего обхода, его ИД и тип обновления во время последнего обхода.

- О компоненте обработки контента**
 - Компонент обработки контента располагается между компонентом обхода и компонентом индексирования. Он обрабатывает элементы, обход которых выполнен, и передает их компоненту индексирования.
 - Компонент обработки контента преобразует элементы, обход которых выполнен, в артефакты, которые можно включить в индекс поиска с помощью таких операций, как разбор документа и сопоставление свойств.
 - Как компонент обработки контента, так и компонент обработки запросов выполняют лингвистическую обработку. К примерам таких операций при обработке контента относятся определение языка и извлечение объектов.
 - Он записывает сведения о ссылках и URL-адресах в базу данных ссылок.

Процессы индексирования и запросов

Архитектура индекса и запросов включает **компонент индексирования**, **раздел индекса** и **компонент обработки запросов**. Все эти элементы можно масштабировать в зависимости от объема контента и запросов, а также требований к производительности.

- О компоненте индексирования**
 - Компонент индексирования — это логическое представление реплики индекса. В архитектуре поиска необходимо подготовить по одному компоненту индексирования для каждой реплики индекса.
 - Компонент индексирования получает обработанные элементы из компонента обработки контента и записывает их в файл индекса.
 - Компонент индексирования получает запросы из компонента обработки запросов и предоставляет в ответ наборы результатов.
 - Запросы отправляются в реплики индекса через компонент обработки запросов. Система направляет входящие запросы в реплики индекса, балансируя их нагрузку.

- О разделе индекса**
 - Раздел индекса — это логическая часть индекса поиска. Индекс поиска — это совокупность всех разделов индекса.

- О компоненте обработки запросов**
 - Компонент обработки запросов располагается между внешним интерфейсом поиска и компонентом индексирования.
 - Компонент обработки запросов анализирует и обрабатывает поисковые запросы и результаты поиска.
 - Как компонент обработки запросов, так и компонент обработки контента выполняют лингвистическую обработку. К примерам таких операций при обработке запросов относятся разбиение на слова и выделение корней.
 - Когда компонент обработки запросов получает запрос от внешнего интерфейса поиска, он анализирует и обрабатывает запрос, чтобы попытаться оптимизировать точность, полноту и релевантность. Затем обработанный запрос отправляется компоненту индексирования.
 - Компонент индексирования возвращает набор результатов, основанный на обработанном запросе, компоненту обработки запросов, который, в свою очередь, обрабатывает набор результатов, прежде чем возвращать его внешнему интерфейсу поиска.

Администрирование поиска

Администрирование поиска состоит из **компонента администрирования поиска** и соответствующей базы данных.

- О компоненте администрирования поиска**
 - Компонент администрирования поиска отвечает за выполнение ряда системных процессов, необходимых для поиска.
 - Этот компонент выполняет подготовку, то есть добавляет и инициализирует дополнительные экземпляры других компонентов поиска.

- О базе данных администрирования поиска**
 - В базе данных администрирования поиска хранятся данные конфигурации поиска, такие как топология, правила обхода, правила запросов и сопоставления между свойствами для обхода и управляемыми свойствами.

Роли сервера

Веб-сервер

- Размещает веб-части поиска и страницы веб-частей для ответов на поисковые запросы.
- В выделенных фермах поиска для поиска в корпоративной среде эта роль необязательна, так как веб-серверы во внешних фермах связываются с серверами, которые обслуживают запросы напрямую.
- Эта роль необходима для ферм, включающих другие возможности SharePoint Server 2016.
- В небольших фермах эту роль также можно использовать на сервере с ролью сервера приложений.

Сервер приложений с компонентами поиска

- Размещает все компоненты поиска, если настроен только один сервер. В противном случае на нем хранятся связанные с сервером компоненты, настроенные администратором.
- Хранит весь индекс поиска, если настроен только один раздел индекса. В противном случае на нем хранятся связанные с разделами части индекса, настроенные администратором.
 - Компонент обработки запросов направляет входящие запросы в реплики индекса.
 - Каждая реплика индекса является компонентом индексирования.
 - Для каждой фермы должен быть настроен как минимум один раздел индекса.

- Для каждой фермы должен быть настроен как минимум один компонент поиска.
- На одном сервере приложений не могут размещаться несколько компонентов поиска одного типа.
- Чтобы обеспечить избыточность, добавляйте компоненты поиска на отдельные серверы.

Сервер базы данных

- Размещает базы данных поиска.
- На нем могут размещаться другие базы данных SharePoint Server 2016.
- Поддерживает зеркальное отображение и кластеризацию.
- Чтобы повысить производительность и емкость, вы можете добавить диски на сервер базы данных или добавить серверы баз данных (в зависимости от проблемы).

Компоненты поиска

Нет

В архитектурах SharePoint Server 2016 для поиска в корпоративной среде компоненты поиска не размещаются на веб-серверах. В архитектурах поиска для веб-сайтов это не так. В этом случае компонент обработки запросов и компоненты индексирования размещаются на веб-серверах, обеспечивая максимально эффективное использование имеющихся аппаратных ресурсов и упрощая масштабирование топологии поиска.

Индекс
Компонент индексирования
Компонент индексирования — это логическое представление реплики индекса.

Разделы индекса

- Индекс можно разделить на фрагменты, каждый из которых отвечает за его отдельную часть.
- Раздел индекса хранится в виде набора файлов на диске.
- Индекс поиска — это совокупность всех разделов индекса.

Обработка запросов
Компонент обработки запросов
Анализирует и обрабатывает поисковые запросы и результаты.

Администратор

Обход контента
Компонент обхода
Выполняет обход контента на основании сведений, указанных в базах данных обхода.

Обработка контента

Аналитика
Компонент обработки аналитики
Выполняет анализ поиска и использования.

БД администрирования поиска
База данных администрирования поиска
Хранит данные конфигурации поиска. У каждого приложения службы поиска может быть только одна база данных администрирования поиска.

БД обхода
База данных обхода

- Хранит журнал обхода контента.
- Управляет операциями обхода контента.
- С каждой базой данных обхода может быть связан один или несколько компонентов обхода.

БД ссылок
База данных ссылок
Хранит данные, извлеченные компонентом обработки контента, а также сведения о выборе рекламных объявлений.

БД аналитики
База данных отчетов аналитики
Хранит результаты аналитики по использованию.

Реплики индекса

- Каждый раздел индекса содержит одну или несколько реплик с одинаковыми данными.
- Для каждой реплики индекса необходимо подготовить по одному компоненту индексирования.
- Чтобы обеспечить отказоустойчивость и избыточность, создайте дополнительные реплики индекса для каждого его раздела и распределите эти реплики индекса между несколькими серверами приложений.

Компонент администрирования поиска

- Выполняет системные процессы, необходимые для поиска.
- Одно приложение службы поиска могут обслуживать несколько компонентов администрирования поиска, но они не могут быть активными одновременно.

Компонент обработки контента
Выполняет различные операции с элементами, для которых выполнен обход (например, разбор документа и сопоставление свойств).

Процессы аналитики

Архитектура аналитики состоит из **компонента обработки аналитики**, **базы данных отчетов аналитики** и **базы данных ссылок**.

- О компоненте обработки аналитики**
 - Компонент обработки аналитики выполняет два вида анализа: анализ поиска и аналитику по использованию. Этот компонент использует полученную информацию для повышения релевантности поиска, создания отчетов по поиску, а также формирования рекомендаций и прямых ссылок.
 - Анализ поиска позволяет извлекать такую информацию, как ссылки, число щелчков элемента, текст привязки, связанные с пользователями данные и метаданные, из базы данных ссылок. Эта информация важна для релевантности.
 - Аналитика по использованию позволяет проанализировать сведения из журнала использования, полученные с интерфейсного веб-сервера через хранилище событий. Аналитика по использованию создает отчеты об использовании и статистические отчеты.
 - Результаты анализа будут добавлены в элементы индекса поиска. Кроме того, результаты аналитики по использованию хранятся в базе данных отчетов аналитики.

- О базе данных ссылок**
 - В базе данных ссылок хранятся сведения, извлеченные компонентом обработки контента. В ней также хранятся сведения о том, сколько раз пользователи выбрали результат на странице результатов поиска. Эти данные хранятся в необработанном виде и анализируются компонентом обработки аналитики.

- О базе данных отчетов аналитики**
 - В базе данных отчетов аналитики хранятся результаты аналитики по использованию. Кроме того, в ней хранятся статистические сведения об анализе. SharePoint использует эти данные для создания статистических отчетов Excel.

- О хранилище событий**
 - В хранилище событий хранятся события использования, зафиксированные на внешнем интерфейсе, например количество просмотров элемента.
 - Эти события использования хранятся в виде файлов журналов на сервере приложений, на котором размещается компонент обработки аналитики.

Пример топологии поиска

Универсальная отказоустойчивая ферма для поиска в корпоративной среде

Эта ферма призвана предоставлять полностью отказоустойчивую виртуальную среду для SharePoint Server 2016, включающую поиск. На этой иллюстрации представлен пример фермы среднего размера для поиска в корпоративной среде, содержащей около 80 млн элементов в индексе поиска.

Примечание. Этот пример не относится к топологиям поиска для веб-сайтов.



Иллюстрация взаимодействия с компонентом поиска

