搜索体系结构
适用于 SharePoint 2016

概述

Microsoft® SharePoint® Server 2016 中的搜索体系结构由为了执行搜索操作而紧密合作的各个组件和数据库组成。所有组件都驻留在应用程序服务器上，所有数据库都驻留在数据库服务器上。

搜索组件交互

爬网和组件处理

爬网和内容处理体系结构包括**爬网组件**、**爬网数据库**和**内容处理组件**。这两个搜索组件都可以爬网量和性能要求为依据进行扩展。

1 关于**爬网组件**
- 爬网组件负责对内容源进行爬网，将已爬网项（包括实际内容及其关联的元数据）传递到内容处理组件。
- 爬网组件调用为了检索数据与内容源进行交互的连接器或协议处理程序。可部署多个爬网组件以便同时爬网。
- 爬网组件使用一个或多个爬网数据库来临时存储已爬网项的信息，并跟踪爬网历史记录。

A 关于**爬网数据库**
- 爬网数据库包含已爬网项的详细跟踪信息和历史记录。
- 此数据库存储上次爬网时间、上次爬网 ID 以及上次爬网期间的更新类型等信息。

2 关于**内容处理组件**
- 内容处理组件位于爬网组件和索引组件之间，负责处理已爬网项，并将这些项馈送给索引组件。
- 内容处理组件通过执行文档分析和属性映射等操作，将已爬网项转换成可在搜索索引中添加的项目。
- 内容处理组件和查询处理组件均执行语言处理。在内容处理期间，语言处理示例包括语言检测和实体提取。
- 内容处理组件将链接和 URL 的相关信息写入链接数据库。

索引和查询处理

索引和查询体系结构包括**索引组件**、**索引分区**和**查询处理组件**，所有这些都可以内容量、查询量和性能要求为依据进行扩展。

4 关于**索引组件**
- 索引组件是索引副本的逻辑表示。在搜索体系结构中，必须为每个索引副本预配一个索引组件。
- 索引组件接收来自内容处理组件的已处理项，并将这些项写入索引文件。
- 索引组件接收来自查询处理组件的查询，并提供返回结果集。
- 查询通过查询处理组件发送到索引副本。系统将传入查询路由到索引副本，并执行负载平衡。

关于**索引分区**
- 索引分区是整个搜索索引的逻辑部分。搜索索引汇总了所有索引分区。

5 关于**查询处理组件**
- 查询处理组件位于搜索前端和索引组件之间。
- 查询处理组件分析和处理搜索查询和结果。
- 查询处理组件和内容处理组件均执行语言处理。在查询处理期间，语言处理示例包括分词和词干分解。
- 当查询处理组件收到来自搜索前端的查询时，它会分析并处理查询，以尝试优化精准率、召回和相关性。然后，它会将处理后的查询提交到索引组件。
- 索引组件会根据处理后的查询向查询处理组件返回结果集。反过来，查询处理组件会先处理此结果集，然后将其发送回搜索前端。

搜索管理

搜索管理由**搜索管理组件**及其相应的数据库组成。

6 关于**搜索管理组件**
- 搜索管理组件负责执行搜索所需的多个系统处理。
- 此组件执行预配，即添加和初始化其他搜索组件的其他实例。

D 关于**搜索管理数据库**
- 搜索管理数据库存储搜索配置数据，如拓扑、爬网规则、查询规则和已爬网属性与已托管属性之间的映射。

服务器角色

Web 服务器
- 托管搜索 Web 部件和 Web 部件页，以响应搜索查询。
- 在企业级搜索的专用搜索服务器场中，此角色不是必需的，因为位于远程场的 Web 服务器与直接处理查询的服务器联系。
- 对于包括其他 SharePoint Server 2016 功能的场来说，此角色是必需的。
- 在小型场中，可以在服务器上将此角色与应用程序服务器角色共享。

带有搜索组件的应用程序服务器
- 仅在配置了一个服务器的情况下托管所有搜索组件。在其他情况下，保留与服务器相关联的组件，由管理员进行配置。
- 仅在配置了一个索引分区的情况下托管整个搜索索引。在其他情况下，保留与索引分区相关联的索引部分，由管理员进行配置。
  - ❑ 查询处理组件将传入的查询路由到索引副本。
  - ❑ 每个索引副本都是一个索引组件。
  - ❑ 每个场至少必须配置一个索引分区。
- 每个场至少必须配置一个索引组件。
- 在一台应用程序服务器上，不能有多个类型相同的搜索组件。
- 在各个服务器上单独添加搜索组件可以实现冗余。

数据库服务器
- 托管搜索数据库。
- 可以托管其他 SharePoint Server 2016 数据库。
- 可以执行镜像或群集操作。
- 为了提升性能和扩大容量，应考虑将磁盘添加到数据库服务器，或考虑添加数据库服务器（具体视瓶颈而定）。

搜索组件

无

在 SharePoint Server 2016 的企业级搜索体系结构中，搜索组件并未托管在 Web 服务器上。

这与 Internet 站点搜索体系结构有所不同。在后一体系结构中，查询处理组件和索引组件驻留在 Web 服务器上，以最大限度地利用可用硬件资源，并简化搜索拓扑扩展。

索引

索引组件
索引组件是索引副本的逻辑表示。

索引分区
- 可以将索引划分为各个独立部分，每个保留索引的不同部分。
- 索引分区存储在磁盘上的一组文件中。
- 搜索索引汇总了所有索引分区。

索引副本
- 每个索引分区保留一个或多个包含相同信息的索引副本。
- 必须为每个索引副本预配一个索引组件。
- 为了实现容错和冗余，请为每个索引分区创建其他索引副本，并在多个应用程序服务器上分布索引副本。

查询处理

查询处理组件
分析并处理搜索查询和结果。

管理

搜索管理组件
- 执行搜索所必需的系统处理。
- 每个 Search Service应用程序可以有多个搜索管理组件，但在任何指定时间，只有一个可用。

爬网

爬网组件
根据爬网数据库中指定的内容对内容进行爬网。

内容处理

内容处理组件
对已爬网项执行各种处理（如文档分析和属性映射）。

分析

分析处理组件
执行搜索分析和使用情况分析。

搜索管理 DB

搜索管理数据库
存储搜索配置数据。每个 Search Service 应用程序只有一个搜索管理数据库。

爬网 DB

爬网数据库
- 存储爬网历史记录。
- 管理爬网操作。
- 每个爬网数据库可以有一个或多个与之相关联的爬网组件。

链接 DB

链接数据库
存储由内容处理组件提取的信息，也存储单击后到达信息。

分析 DB

分析报表数据库
存储使用情况分析结果。

搜索拓扑示例

适用于企业级搜索的多用途容错场

此场旨在为 SharePoint Server 2016（包括搜索）提供完全容错的虚拟环境。此示例示意图展示的是中型企业级搜索服务器场，搜索索引中大约有 8000 万个项。

注意：此示例不适用于 Internet 站点的搜索拓扑。

分析处理

分析体系结构由**由分析处理组件**、**分析报表数据库**和**链接数据库**组成。

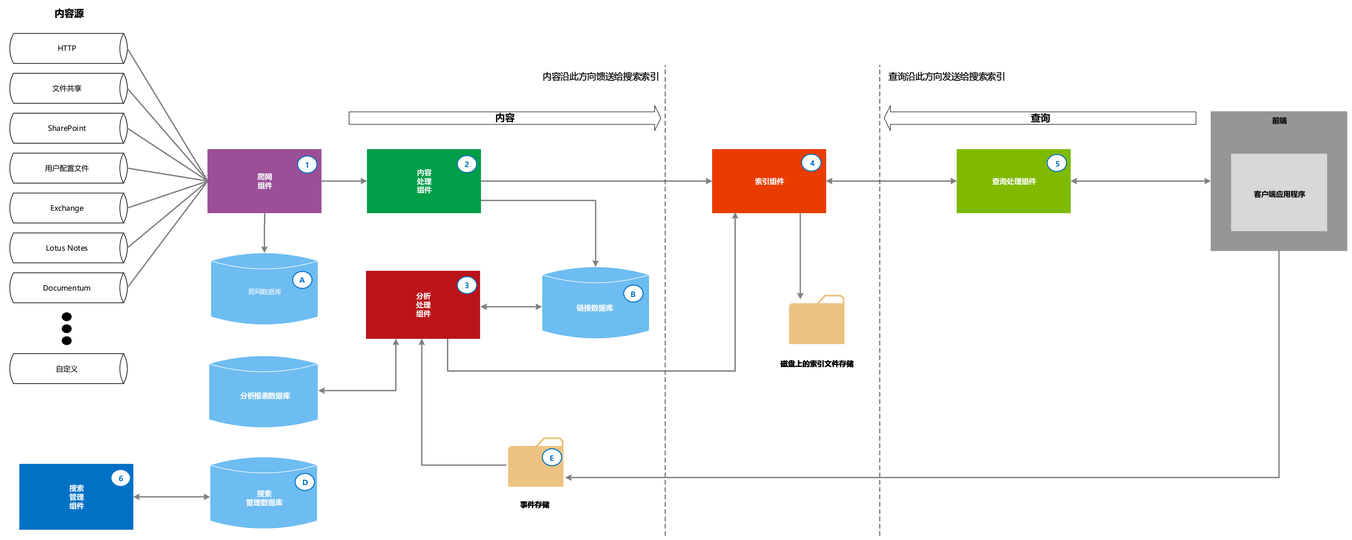
3 关于**分析处理组件**
- 分析处理组件执行两种分析：搜索分析和使用情况分析。此组件使用这两种分析提供的信息来提升搜索相关性、创建搜索报告并生成建议和深层链接。
- 搜索分析是指从链接数据库提取信息，例如链接、项的单击次数、定位标记文本、用户相关数据和元数据。此信息对于相关性很重要。
- 使用情况分析是指分析通过事件存储从前端接收的使用情况日志信息。使用情况分析可生成使用情况和统计信息报告。
- 分析结果将添加到搜索索引中的项内。此外，使用情况分析结果存储在分析报表数据库中。

B 关于**链接数据库**
- 链接数据库存储由内容处理组件提取的信息。此外，还存储有关搜索单击的相关信息，以及用户单击搜索结果页中的搜索结果的次数。所存储的信息未经处理，将由分析处理组件执行分析。

C 关于**分析报表数据库**
- 分析报表数据库存储使用情况分析结果。
- 此外，分析报表数据库还存储分析统计信息。SharePoint使用此类信息创建Excel报表，以显示不同的统计信息。

E 关于**事件存储**
- 事件存储保留在前端捕获的使用事件，如项的查看次数。
- 这些使用事件作为日志文件存储在托管分析处理组件的应用程序服务器上。

搜索组件交互示意图



SharePoint

© 2016 Microsoft Corporation。保留所有权利。若要发送有关此文档的反馈，请发送电子邮件至以下地址：ITSPdocs@microsoft.com。

Microsoft