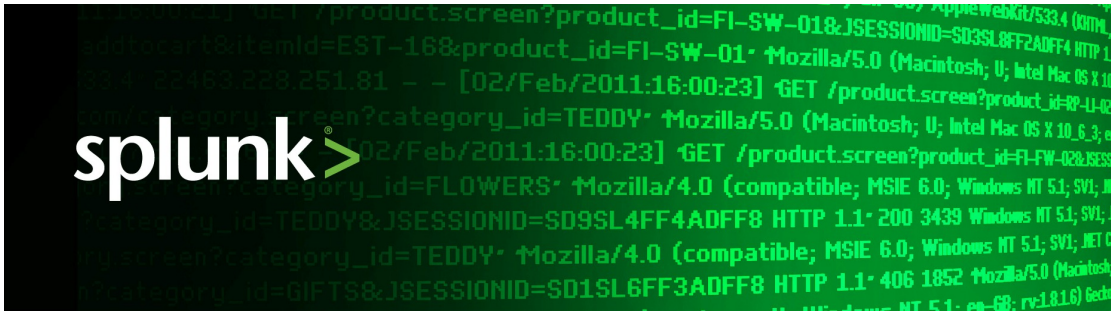


Splunk® Enterprise 7.2.0

容量规划手册

生成时间：2018 年 10 月 17 日，上午 11:14

Table of Contents

简介

Splunk Enterprise 容量规划简介

您可以对 Splunk Enterprise 进行扩展以满足几乎任何容量要求。要利用这种扩展能力需要进行规划。本手册介绍了 Splunk Enterprise 部署的高级别硬件指导，同时介绍了 Splunk Enterprise 如何在不同的情况下使用硬件资源。

作为 Splunk Enterprise 6.2 以及之后版本的新内容，本手册代替 *《安装》*和 *《分布式部署》*手册中有关容量规划的指导。它提供有关参考硬件和性能检查表的信息，以确定根据您的需求应该在何时以及如何扩展您的部署规模。

在您决定 Splunk Enterprise 的硬件之前，请参阅下列信息：

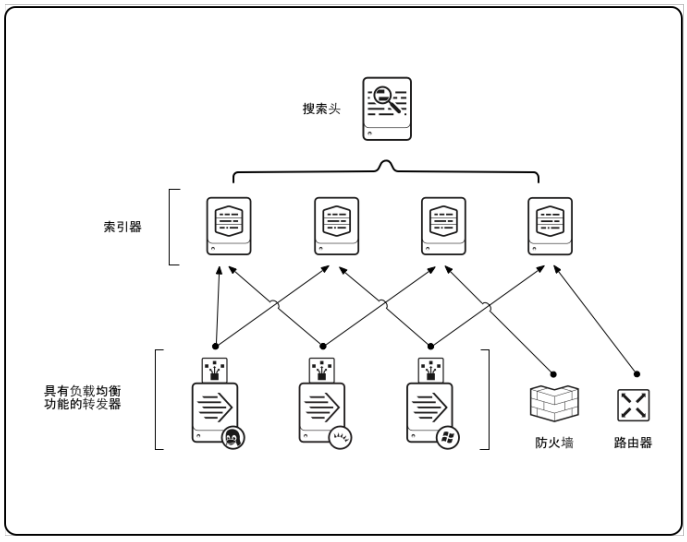
- 有关 Splunk Enterprise 安装中元素的描述，请查看本手册中的“Splunk Enterprise 部署的各个组件”。
- 阅读 Splunk Enterprise 部署维度，了解这些维度如何影响性能，以及如何将性能最大化。
- 在本手册的“参考硬件”中了解有关 Splunk Enterprise 部署的基本构建块的信息。

硬件容量规划

Splunk Enterprise 部署的各个组件

最简单的部署是您在计算机上首次安装 Splunk Enterprise 时默认获得的部署：一个可处理索引和搜索的独立实例。您可在实例上登录 Splunk Web 或 CLI 并配置数据导入，以收集计算机数据。然后您使用同一实例搜索、监视、告警并报告传入数据。

您还可在多个计算机上部署专门的 Splunk Enterprise 实例以处理负载和可用性要求。这些专门的实例称为“组件”。本部分介绍了组件类型。参阅*《分布式部署》*手册，特别是以下主题：使用 Splunk Enterprise 组件调整部署规模。



索引器

Splunk **索引器**为本地和远程数据提供数据处理和存储，并托管主要 Splunk 数据存储区。有关更多信息，请参阅*《管理索引器和群集》*手册中的“索引如何工作”。

搜索头

搜索头是分发搜索到索引器（在此上下文中称为“搜索节点”）的 Splunk Enterprise 实例。搜索头可以是专用或非专用，这取决于它们是否还执行索引。除常见的内部索引外，专用搜索头没有任何自己的索引。相反，它们会整合并显示来自远程搜索节点的结果。

要配置跨索引器池搜索的搜索头，请参阅*《分布式搜索手册》*中的“分布式搜索是什么”

转发器

转发器是转发数据到远程索引器进行数据处理和存储的 Splunk 实例。在大部分情况下，它们不会自己索引数据。请参阅*《转发数据》*手册中的“关于转发和接收”主题。

部署服务器

Splunk Enterprise 实例也可作为**部署服务器**。部署服务器是一款用来将配置、应用和内部更新分发到各组 Splunk Enterprise 实例的工具。您可以使用部署服务器将更新分发到大多数 Splunk 组件：转发器、非群集索引器和非群集搜索头。请参阅*《更新 Splunk Enterprise 实例》*手册中的“关于部署服务器和转发器管理”。

功能一览

函数	索引器	搜索头	转发器	部署服务器
索引	x			
Web		x		
直接搜索		x		
转发至索引器			x	

部署配置	x		x	x

索引复制和索引器群集

索引器群集是配置为复制彼此数据的一组索引器，这样系统便会保留所有数据的多个副本。此过程称为**索引复制**。通过保留数据的多个相同副本，索引器群集能够防止数据丢失，同时还便于数据搜索。

Splunk Enterprise 群集功能会自动从一个索引器故障转移到下一个索引器。这意味着，如果一个或多个索引器出现故障，可继续为传入数据新建索引，且索引数据继续保持可搜索状态。

除了能提高数据可用性，群集还具备当您调整部署规模时应该考虑的其他功能，例如，轻松地跨群集中所有索引器协调配置更新的功能。此外，群集内置了分布式搜索操作。请参阅*《管理索引器和索引器群集》*手册中的"关于群集和索引复制"。

Splunk Enterprise 部署维度

Splunk Enterprise 部署有多个维度。这些方案确定单个参考计算机是否可以处理索引和搜索负载。

在一些情况下，单个参考计算机可以有效地收集、存储和搜索数据。在其他情况下，可以考虑向 Splunk Enterprise 部署中添加计算机以提高性能。以下是会对 Splunk Enterprise 性能有显著影响的项目列表。

- **传入数据量。**您发送到 Splunk Enterprise 的数据越多，它需要越多时间处理数据直到可以搜索、报告并生成告警的事件。
- **索引数据量。**当存储在 Splunk Enterprise 索引中的数据量增加的时候，存储数据并提供搜索结果所需要的 I/O 带宽也会增加。
- **并发用户的数量。**如果有一次使用 Splunk Enterprise 实例的人数超过一人，则该实例需要更多资源以便这些用户执行搜索及新建报表和仪表板。
- **保存的搜索数量。**如果计划调用许多保存的搜索，Splunk Enterprise 需要其他容量才能立即有效地执行这些搜索。在给定的时间周期内，更多的搜索次数需要更多的资源。
- **使用的搜索类型。**与保存的搜索数量几乎一样重要的是，您对 Splunk Enterprise 实例运行的搜索类型。这里有几种类型的搜索，每个搜索将影响索引器响应搜索请求的方式。
- **您是否运行 Splunk 应用。**Splunk **应用**和解决方案拥有独特的性能、部署和配置注意事项。如果计划运行应用，请考虑您所使用的应用资源要求。有关更多信息，请参阅应用文档。

这些维度如何影响整体性能？

尽管这些因素会影响您的 Splunk Enterprise 部署的基本大小要求，但是单独处理每个问题无法保证获得部署的最高性能。您必须通过试验了解这些因素如何在特定应用程序中彼此相关。

例如，如果您的 Splunk Enterprise 部署调用少量索引，但是具有大量并发用户，则资源要求与具有少量并发用户和大量日常索引数据量的设置完全不同。此外，随着用户计数和索引的数据量增加，您必须跨多个服务器分发环境，以保持类似的性能级别。搜索类型使得问题变得复杂，因为一些搜索受可用 CPU 资源限制，而其他的搜索则取决于磁盘子系统的速度。

何时应扩展 Splunk Enterprise 部署？

您必须了解本主题中描述的部署维度如何适用于您的特定用例。回答以下问题，然后参阅本手册的性能检查表以确定何时应增加更多硬件资源：

- 您每天预计索引多少数据？
- 您需要保留多少数据以及保留多长时间？
- 您预计有多少用户会在任何时间同时搜索数据？
- 您是否计划使用一次以上的某些特定搜索？
- 您是否希望或需要使用 Splunk 应用以显示或操作数据？

想要安装时运行良好的关键在于在部署进程早期制定计划，以考虑最初的硬件资源支出和部署扩展后的资源增加。

传入数据如何影响 Splunk Enterprise 性能

参考 Splunk Enterprise 索引器可在较短时间内索引大量数据：超过每秒 20 MB 数据或每天 1.7 TB。假定服务器仅处理数据而不进行其他任何操作，则会达到该索引级别。

由于 Splunk Enterprise 实例不仅仅进行索引，可将该数字看作索引器的最大吞吐量。性能变化取决于传入数据的大小和数据量。大型事件会减慢索引性能。随着事件的大小增加，索引器将使用更多系统内存以处理和索引它们。

如果需要较单个索引器所提供的更大的索引容量，可添加索引器到部署以应对增加的需求。

参阅本章中的主题以了解其他因素如何影响性能数据。

索引的数据如何影响 Splunk Enterprise 性能

在 Splunk Enterprise 耗尽数据并放入索引之后，这些索引将增长并占据磁盘空间。随着索引增长和可用磁盘空间减少，Splunk Enterprise 将花费更多时间索引传入数据，因为索引器的磁盘子系统会花费更多时间查找空间以存储数据。

该增长也会影响搜索。在单个索引器上，磁盘吞吐量拆分为索引（这是持续的）和搜索请求（这将基于用户计划的请求中断）。随着索引增长，搜索速度会变慢，因为磁盘子系统需要处理搜索请求，并且还需要处理时间更长的传入数据存储请求。根据搜索类型的不同，这些请求类型会耗费很多的 I/O 资源。

并发用户如何影响 Splunk Enterprise 性能

参考索引器需要为用户调用的每个搜索（只要搜索是活跃的）专门提供一个可用的 CPU 核心。如果多个用户登录和运行搜索，可用的 CPU 核心的数量会迅速耗尽。

这些数字假定该 CPU 收到登录或搜索请求时处于空闲状态。这不考虑其他系统请求或 Splunk Enterprise 使用来索引数据的 CPU 核心。如果正在处理任何其他系统请求，则负载将在其他可用 CPU 之间进行拆分。

随着 CPU 核心满载，索引器上的所有活动将减慢，因为计算机在索引、搜索和处理在线用户之间拆分处理时间。仅其他索引器可以增加所有三个 Splunk Enterprise 操作功能的容量。

保存的搜索/报表如何影响 Splunk Enterprise 性能

在参考索引器上，**保存的搜索**或报表在执行期间使用约 1 个 CPU 核心和指定的内存量。它的行为类似于临时搜索。随着磁盘子系统扫描**索引**以提取数据，保存的搜索还会临时增加磁盘 I/O 量。

同时执行的每个其他保存的搜索会使用一个其他的 CPU 核心。此消耗情况与操作系统以及 Splunk Enterprise 索引和存储进程的 CPU 使用分开。

如果要执行命令的已保存搜索超过可以处理的数量，则它们会一直在队列中等待，直到进行处理。当系统达到保存的搜索最大排队数量时，Splunk Enterprise 还会发出警告。当搜索等待队列时，搜索结果将更慢返回。

添加**搜索头**将提供其他 CPU 核心以运行更多并发搜索。对于因为添加搜索头而导致的搜索负载和并发的增加，添加索引器能帮助调整规模。添加 RAM 到现有计算机有助于并发搜索，但不会提供其他搜索容量。

搜索类型如何影响 Splunk Enterprise 性能

您可以对存储在 Splunk Enterprise 索引中的数据调用四种搜索类型。每种搜索类型均以不同方式影响索引器。

下表汇总了不同搜索类型。对于密集搜索和稀疏搜索，Splunk Enterprise 根据匹配事件的数量来衡量性能。对于超稀疏搜索和罕见搜索，则基于总索引量来衡量性能。

搜索类型	描述	参考索引器吞吐量	性能影响

密集	返回给定时间范围内给定数据集的一个大比例匹配结果（10% 或更多）。密集搜索通常会先占用服务器的 CPU，因为解压缩存储在 Splunk Enterprise 索引中的原始数据需要一定的开销。密集搜索的示例包括仅使用通配符字符的搜索，或搜索任何索引。
	
	示例：
	
	`*`
	
	`index=m …| stats count by fieldA`
	
	`index=a sourcetype=b …| timechart count by myfield`
		每秒最多 50,000 个匹配事件。	受 CPU 限制

稀疏	与密集搜索相比，返回给定时间范围内给定数据集的较少结果（任何位置都在 0.01 到 1% 范围内）。	每秒最多 5,000 个匹配事件。	受 CPU 限制
超稀疏	返回每个索引**数据桶**中与搜索匹配的少数结果。超稀疏搜索耗费很多的 I/O 资源，因为索引器必须浏览索引的全部数据桶才能找到结果。如果您的索引器上存储了大量数据，这就需要很多数据桶，因此超稀疏搜索可能需要很长时间才能完成。	每个索引数据桶最多 2 秒。	受 I/O 限制

罕见	与超稀疏搜索相类似，但需要借助**布隆过滤器**，该过滤器可帮助消除那些不匹配搜索请求的索引数据桶。罕见搜索在任意位置返回结果的速度均可达超稀疏搜索的 20 到 100 倍。	每秒 10 到 50 个索引数据桶。	受 I/O 限制

Splunk 应用如何影响 Splunk Enterprise 性能

单个 Splunk Enterprise 索引器可以同时运行多个应用。Splunk Enterprise 包括在同一时间运行的若干个应用。

然而，更复杂的应用提供高级视图（需要使用在后台运行的摘要和加速搜索）。应用需要的后台处理越多，您越有可能必须跨多个计算机分发处理负载。

许多应用需要设计分布式 Splunk Enterprise 部署。无论是提取数据并发送数据到单个中央实例的通用转发器，还是连接在一起并用于报表、仪表板或告警的许多索引器和搜索头，Splunk 应用经常都需要一个以上服务器实现企业的最大性能和潜能。

Splunk 应用对资源需求的影响

如果您使用的 Splunk 应用或解决方案将通过执行大量保存的搜索来获取知识，单一服务器 Splunk Enterprise 实例会被塞满。多个搜索很快就会耗尽索引器上的可用 CPU 资源。请参阅本手册中的“容纳许多同时进行的搜索”。

当您安装应用或解决方案时，请阅读相应应用或解决方案文档中所述的系统要求。如果未提供相关信息，请联系该应用或解决方案的作者，以获取有关正确运行该应用所需内容的信息。

Splunk Enterprise 如何计算磁盘存储

在高级别，Splunk 将如下计算总磁盘存储：

```
( Daily average indexing rate ) x ( retention policy ) x 1/2
```

由于经过压缩处理，Splunk Enterprise 会以原始大小约一半的空间存储原始数据。在包含 500GB 可用磁盘空间的卷上，您可以以 5GB/天的索引速率存储近六个月的数据，或以 100GB/天的速率存储十天的数据。

如果需要其他存储，则可选择更多本地磁盘（需要频繁搜索时）或附加或网络存储（可接受偶尔搜索时）。对于每 GB 更低成本优先于即时搜索返回的长时间搜索，可以接受通过 NFS 或 SMB/CIFS（服务器信息块/通用互联网文件系统）的低延迟连接。

重要提示：通过广域网 (WAN) 连接的共享安装或磁带等备用存储不适合 Splunk Enterprise 操作的存储选择。

评估存储要求

本主题介绍了如何评估 Splunk Enterprise 索引大小，以计划您的存储容量要求。

Splunk Enterprise 为您的数据新建索引时，会新建两种主要类型的文件：包含压缩形式的原始数据 "rawdata" 文件，以及指向该数据的索引文件。（它还会新建几个元数据文件，这不会使用太多空间。）只需少量实验，您就可以评估给定传入数据量所需的索引磁盘空间。

通常，压缩的原始数据文件为传入预索引原始数据大小的 10%。关联的索引文件大小约为原始数据文件的 10% 至 110%。数据的唯一术语数量对该值产生影响。

根据数据特性的不同，您可能希望微调分段设置，如同《数据导入手册》中的“关于分段”所述。

了解空间需求的最佳方式是通过索引数据的代表性样本进行试验，然后检查 `$SPLUNK_HOME/var/lib/splunk/defaultdb` 中产生的目录大小。

在 *nix 系统上，遵照这些步骤

一旦索引完数据样本：

1.转到 `$SPLUNK_HOME/var/lib/splunk/defaultdb/db`。

2.运行 `du -ch hot_v*` 并查看最后 `total` 行以查看索引大小。

在 Windows 系统上，遵照这些步骤

1.从 Microsoft TechNet 下载 du 实用工具。

2.从下载的 ZIP 文件提取 `du.exe` 并放入您的 `%SYSTEMROOT%` 或 `%WINDIR%` 文件夹。

注意：您还可将它放在您 `%PATH%` 中的任何地方。

3. 打开命令提示符。

4. 从命令提示符，转到 `%SPLUNK_HOME%\var\lib\splunk\defaultdb\db`。

5. 运行 `del %TEMP%\du.txt & for /d %i in (hot_v*) do du -q -u %i\rawdata | findstr /b "Size:" >> %TEMP%\du.txt`。

6. 打开 `%TEMP%\du.txt file.`。您将看到 `Size: n`，这是每个找到的 `rawdata` 目录的大小。

7. 合计这些数字，找出压缩的持久化原始数据大小。

8. 接下来，运行 `for /d %i in (hot_v*) do dir /s %i`，其摘要是索引的大小。

9. 添加该数字到总持久化原始数据数量。

这是您已索引样本的索引和关联数据的总大小。现在，您可用它外推随时间变化的 Splunk Enterprise 索引和 `rawdata` 目录的大小要求。

问答

有什么问题吗？请访问 Splunk Answers，查看其他 Splunk 用户有哪些与数据大小相关的问题和解答。

调整您的 Splunk Enterprise 部署规模

分布索引和搜索

本主题介绍分发您的 Splunk 平台部署组件的原因。

分布式索引和搜索概念

为 Splunk 平台设计一个可调整规模的架构需要 Splunk 实例角色的知识，以及它们打算如何调整规模。

最常见的两种角色是搜索头和索引器。它们代表承担管理用户对象、搜索、分析和数据存储职责的角色。

搜索头负责：

- 托管用户。
- 存储用户新建的对象。
- 计划搜索和告警。
- 通过仪表板和视图提供可视化反馈信息。
- 执行访问控制。

索引器负责：

- 从转发器接收数据流。
- 分析数据。
- 将数据写入数据桶。
- 维护数据桶。
- 接受来自搜索头的搜索请求。
- 搜索数据桶并将结果以流化方式传回给搜索头。

搜索头的任务主要受 CPU 限制。随着更多的用户和更多的应用添加到搜索头，并发搜索负载会迅速攀升并达到极限。该极限表示以搜索头的 CPU 核心衡量的跨所有用户和应用的总搜索负载。

添加搜索头到部署可增加总的 CPU 资源，增加环境中所支持的总的搜索并发以及活跃用户和应用的数目。

索引器的任务主要受 I/O 限制。随着更多的转发器添加到网络，要接受更多的并发数据流，并且在写入前要分析更多的数据。此外，搜索请求需要 I/O 访问和处理器时间来分析、收集和返回所请求的数据。当数据流的量增加和并发搜索请求攀升地更高时，索引器会达到极限。该极限表示以索引器的 I/O 容量衡量的所有搜索头的总搜索负载和来自转发器的索引负载。

添加索引器到部署可增加总的 I/O 容量和可用来保存数据的存储空间，减少每个索引器负载上的数据量，并通过将搜索负载分布到更多的索引器上来降低其影响。

调整 Splunk 平台规模

典型的 Splunk 平台部署规划基于 2 点：每天的索引数据量和评估的搜索负载。用户计数通常被用作搜索负载的替代指标。例如，一个具备管理员级别搜索并发权限的活跃用户可以在 Splunk 平台部署中与多个更低级别角色的用户维持同样的负载。

大多数的 Splunk 执行是建立在搜索数百 GB 数据的少数用户和一些应用上。在这种情况下，添加索引器是调整规模的首选方法。执行索引器群集时适用同样的规则。

当搜索头获得更多用户的时候，CPU 的局限性就会变得明显，此时搜索可能被跳过，用户会体验到更慢的搜索结果速度。向您的分布式部署中添加另一个搜索头不能保证提高搜索性能。随着用户计数的增加，必须添加索引器以保持搜索性能。有关调整规模的指导表格，请参阅本手册中的“性能建议摘要”。

在对于用户计数为 50 或更多的部署进行规划时，可考虑执行搜索头群集以吸收高级别用户，同时增加冗余的搜索层。

调整规模的性能

索引器获取数据时，它们会将数据存储在**数据桶**中，数据桶是索引的单独元素。随着传入数据的增加，数据桶数量也会随之增加。随着数据桶数量的增加，索引器必须通过“滚动”数据桶来管理它们，为新传入的数据留出空间。此过程会占用 I/O 周期，这会减少搜索请求获取事件的可用资源。对于拥有少量数据的索引数据桶，其影响是显而易见的。

避免配置多个组成小数据桶的索引。例如，使用 `maxDataSize` 数据桶设置，可参阅 Splunk Enterprise *《管理员手册》* 中的 `indexes.conf.example`。

搜索的数量和类型也会影响索引器性能。大多数搜索类型会使用索引器的磁盘子系统，但有一些会使用更多的 CPU。有关同时进行搜索的信息，请参阅本手册中的“容纳并发用户和搜索”。

如果分配给索引器的硬件超过参考计算机规格，可考虑查看并执行并行化设置中的一种设置来提高特定用例的性能。

使用监视控制台跨 Splunk 平台环境监视和跟踪资源的使用情况。有关详细信息，请参阅*监视 Splunk Enterprise* 中的“关于监视控制台”。

并发用户和并发搜索对性能的影响

Splunk Enterprise 部署中最大的性能因素是：

- 并发用户的数量。
- 并发搜索的数量。
- 所使用的搜索类型。

提交搜索请求的用户将在每个索引器中使用一个 CPU 核心直到该搜索完成。用户提交的其他任何搜索也占用一个 CPU 核心。您可以调整一台计算机能运行的全局并发搜索数。请参阅 Splunk Enterprise*《搜索手册》* 中的“实时搜索和报表的预期性能和已知限制”。

用户调用的搜索类型还会影响硬件资源使用情况。请参阅“搜索类型如何影响 Splunk Enterprise 性能”。

如何最大程度地提高搜索性能

要适应运行多个并发搜索的资源开销，添加额外索引器，并最大化索引器可用的物理内存。索引器在搜索操作中会执行大量工作，如识别所请求的数据、从磁盘读取数据、解压缩数据、筛选数据和输出流报告。

例如，如果某搜索占用了 200MB 的内存，另外还有 48 个并发搜索请求约需要 10GB 的内存以满足不包括操作系统要求的搜索负载。可用内存量是要监视的一个重要资源。尽管索引器性能随着并发搜索任务对 CPU 使用率的增加而逐渐下降，但当所有可用物理内存耗尽时，索引器性能便会显著下降。

搜索性能：基本场景

所有搜索的总运行时间会随着索引器中可用 CPU 核心数量的减少而增加。例如，对于一个没有负荷且拥有 12 个可用核心的索引器，所收到的第一批待处理的搜索可于较短时间内完成。在此场景中，所有搜索都将于 10 秒内完成。

12 个并发搜索：一个有 12 个核心的索引器且没有数据已建立索引。

并发搜索数	/ 可用核心数	= 每个核心的搜索数	单个搜索所需的秒数	= 完成所有搜索大约花费的时间（秒）
12	12	1	10	10

当同时要运行 48 个并发搜索时，完成所有搜索所需的总时间会大幅增加。

48 个并发搜索：一个有 12 个核心的索引器且没有数据已建立索引。

并发搜索数	/ 可用核心数	= 每个核心的搜索数	单个搜索所需的秒数	= 完成所有搜索大约花费的时间（秒）
48	12	4	10	40

由于索引器在搜索操作中会执行大量工作，如识别所请求的数据、从磁盘读取数据、解压缩数据、筛选数据和输出流报告，所以最佳方式是添加索引器以减少返回所有搜索结果所需的总时间。

部署更多索引器可以提高核心的数量，从而在出现多个并发搜索时缩短完成所有搜索所需的时间。当可用的核心数超过并发搜索数时，Splunk Enterprise 可使用这些核心来进行维护操作，也可以处于闲置状态。

12 个并发搜索：四个索引器每个有 12 个核心且没有数据已建立索引。

并发搜索数	/ 可用核心数	= 每个核心的搜索数	单个搜索所需的秒数	= 完成所有搜索大约花费的时间（秒）
12	48	1.默认情况下，单个搜索无法利用多个核心。	10	10

48 个并发搜索：四个索引器每个有 12 个核心且没有数据已建立索引。

并发搜索数	/ 可用核心数	= 每个核心的搜索数	单个搜索所需的秒数	= 完成所有搜索大约花费的时间（秒）
48	48	1	10	10

搜索性能：对数据建立索引场景

在主动部署中，当搜索到达时，系统处于非空闲状态。如果索引器每天获取 150GB 的数据，则在可用核心中最多使用 4 个核心来新建这些数据的索引。可用核心数越少，则返回所有搜索结果所需的时间越长。

12 个并发搜索：一个 12 核心的索引器，有 8 个可用核心。

并发搜索数	/ 可用核心数	= 每个核心的搜索数	单个搜索所需的秒数	= 完成所有搜索大约花费的时间（秒）
12	8	2，因为每个核心一直处于使用状态直到前一个搜索完成。	10	20

48 个并发搜索：一个 12 核心的索引器，有 8 个可用核心。

并发搜索数	/ 可用核心数	= 每个核心的搜索数	单个搜索所需的秒数	= 完成所有搜索大约花费的时间（秒）
48	8	6	10	60

48 个并发搜索：四个 12 核心的索引器，有 8 个可用核心。

并发搜索数	/ 可用核心数	= 每个核心的搜索数	单个搜索所需的秒数	= 完成所有搜索大约花费的时间（秒）
48	32	2，因为每个核心一直处于使用状态直到前一个搜索完成。	10	20

如果索引器的数量较少但每个索引器的核心数较多，则会减少完成搜索所需的总时间，但也会向搜索操作提供较少的聚合 IOPS 从而降低扩展效率。

添加索引器可减少任意一台计算机上的索引负载。另外，您可以减少搜索时间，降低高搜索并发率的影响以及 I/O 和内存之间争用资源的影响。

性能参考

参考硬件

参考硬件规格是限定 Splunk 平台范围和调整其规模时所用的基线。它是处理搜索和索引建立负载的性能指南。

用于单实例部署的参考主机规格

下面的要求代表 Splunk Enterprise 部署的基本构建块。

- Intel x86 64 位芯片架构
- 12 个 CPU 核心，每个核心速率大于或等于 2Ghz
- 12GB RAM
- 标准 1Gb 以太网 NIC、另一个用于管理网络的 NIC 选件
- 标准 64 位 Linux 或 Windows 分发

另有两个参考规格提供了较高性能和搜索并发能力。这些规格将于本主题后面的内容中进行介绍。

磁盘子系统

参考计算机的磁盘子系统应能够处理大量平均**每秒输**入/输出**操**作 (IOPS)。

IOPS 是硬盘驱动器可以生成多少数据吞吐量的测量方法。因为硬盘驱动器以不同速度读取和写入，所以有多种 IOPS 数量用于磁盘读取和写入。平均 IOPS 是这两个数字的混合。

硬盘驱动器可以生成越多平均 IOPS，它在给定时间周期可以索引和搜索的数据越多。尽管许多变量项目考虑了硬盘驱动器可以生成的 IOPS 量，但是以下是最重要的元素：

- 它的旋转速度（每分钟转速）。
- 它的平均延迟（旋转盘片半周所花费的时间量）。
- 它的平均搜索时间（检索请求的数据块所花费的时间量）。

可生成最高 IOPS 的驱动器具有高转速和低平均延迟和搜索时间。每家驱动器制造商均会提供这一条信息，另外还有一些制造商会提供更多信息。

有关 IOPS 及如何计算的相关信息，请参阅“获得 Symantec 的 Connect Community 相关 IOPS 信息”。

此规格使用了 8 个 146 GB、15,000 RPM 串行连接 SCSI (SAS) HD 独立磁盘冗余阵列 (RAID) 1+0 故障容错方案。每个硬盘驱动器能够提供平均约 200 IOPS 的速度。组合的阵列能够提供平均略超过 800 IOPS 的速度。

没有足够的磁盘 I/O 是 Splunk 基础架构中最常见的限制。为在对数据建立索引时获得最佳结果，在配置您的硬件之前请查看磁盘子系统的要求。

最大性能

最大性能单独测量索引和搜索性能，不代表典型 Splunk 使用案例的组合负载。如要查看带搜索负载索引的参考计算机的性能建议，请参阅“性能建议摘要”。

索引性能

- 最多每秒 20MB（每天 1700GB）的原始索引性能，如果不出现搜索或其他索引相关的活动的话。

搜索性能

- 对于密集搜索，每秒最多 50,000 个事件。
- 对于稀疏搜索，每秒最多 5,000 个事件。
- 对于超稀疏搜索，每索引数据桶最多 2 秒。
- 对于带布隆过滤器的罕见搜索，每秒 10 至 50 个数据桶。

要了解搜索类型以及它们如何影响 Splunk Enterprise 性能的信息，请参阅“搜索类型如何影响 Splunk Enterprise 性能”。

用于分布式部署的参考主机规格

由于活跃用户的数量会随着数据写入率的提高而有所增加，因此架构需求将从单个事例转换为分布式的 Splunk Enterprise 环境。搜索头和索引器角色具有唯一的硬件建议。

专用搜索头

搜索头将比索引器更一致地利用 CPU 资源，但不会对索引要求高速磁盘吞吐量或较大的本地储存池。

- Intel 64 位芯片架构
- 16 个 CPU 核心，每个核心速率大于或等于 2Ghz。
- 12GB RAM
- 2 个 300GB、10,000 RPM SAS 硬盘，采用 RAID 1 配置
- 1Gb 以太网 NIC、另一个用于管理网络的 NIC 选件
- 64 位 Linux 或 Windows 分发

当搜索处于活跃状态时，单个搜索请求最多使用 1 个 CPU 核心。当除了用户运行的临时搜索之外还需要配置搜索头时，必须考虑到计划的搜索。更多的活跃用户和更高的并发搜索负载要求额外的 CPU 核心。

如要查看搜索优先级顺序的排列方式，请参阅*《报表手册》*中的主题"配置计划的报表优先级"。关于扩展搜索性能的信息，请查阅"如何最大化搜索性能"。

索引器

在分发索引过程中，Splunk 平台可将数据处理的规模调整为数 TB 一天。当添加更多索引器时，可将搜索请求的工作和数据索引的工作分发给这些索引器。这会大幅提升性能。

在此提醒您，以下提供了一个参考索引器规格：

参考主机规格

- Intel 64 位芯片架构。
- 12 个 CPU 核心，每个核心大于或等于 2GHz。
- 12GB RAM。
- 能够提供至少 800 平均 IOPS 速度的磁盘子系统。有关详细信息，请参阅主题"磁盘子系统"。
- 1Gb 以太网 NIC，以及另一个用于管理网络的 NIC 选件。
- 64 位 Linux 或 Windows 分发。

Splunk 引入了两种新规格，提供额外的 CPU 核心以实现更佳的索引性能和搜索并发能力，从而提升用户体验。

单个的索引器与一组索引器有相同的磁盘 I/O 带宽要求。

中档规格

中档规格类似于基本参考规格。此规格提升了分布式 Splunk Enterprise 部署中的索引容量和搜索并发能力。

- Intel 64 位芯片架构
- 24 个 CPU 核心，每个核心速率大于或等于 2GHz
- 64GB RAM
- 能够提供至少 800 平均 IOPS 速度的磁盘子系统
- 1Gb 以太网 NIC，以及另一个用于管理网络的 NIC 选件
- 64 位 Linux 或 Windows 分发

高性能规格

高性能规格是在中档规格上的进一步提升。

- Intel 64 位芯片架构
- 48 个 CPU 核心，每个核心速率大于或等于 2GHz
- 128GB RAM
- 能够提供至少 1200 平均 IOPS 速度的磁盘子系统
- 固态磁盘 (SSD) 子系统作为热/温**索引数据桶**的最低要求
- 1Gb 以太网 NIC 以及另一个 NIC 选件
- 64 位 Linux 或 Windows 分发

较高性能规格的磁盘子系统信息

当索引器为搜索检索数据时会进行大量磁盘搜索和批量读取操作。每日数据量较高时，本地磁盘可能无法为想要快速搜索的时间期间提供经济有效的存储。在这些情况下，快速连接存储或网络存储，如光纤存储区域网络 (SAN)，能为每个索引器提供所需的 IOPS。

在规划储存基础设施时要注意以下几个关键点：

- 磁盘越多（特别是主轴越多），索引性能越好。
- 整个系统的总吞吐量很重要。
- 特定系统中磁盘与磁盘控制器的比例应该较高一些，与配置数据库主机的情况类似。

索引器与搜索头的比例

对索引器可以支持的搜索头数量或搜索头可以搜索的索引器数量没有实际的限制。用例决定基础架构中需要什么样的 Splunk 实例角色（搜索头或索引器）来调整规模同时保持性能。有关调整规模的指导表格，请参阅"性能建议摘

要”。

群集部署的网络延迟限制

具有搜索头或索引器群集的 Splunk 环境，其群集和群集节点间必须具备快速、低延迟的网络连接。这对于在多个站点分布群集的环境尤为重要。

对于索引器群集节点，网络延迟不得超过 100 毫秒。较高的延迟将极大地减慢索引性能并阻碍群集节点故障恢复。

网络延迟对群集部署操作的影响。

网络延迟	群集索引时间。1 TB 数据	群集节点恢复时间
< 100 ms	6202 s	143 s
300 ms	6255 s (+ 1%)	1265 s (+ 884%)
600 ms	7531 s (+ 21%)	3048 s (+ 2131%)

延迟影响依单个配置不同而有所区别。

对于搜索头群集，延迟不得超过 200 毫秒。较高的延迟将影响搜索头群集选择群集管理员的速度。

与您的网络管理员确认将支持群集 Splunk 环境的网络可满足或超出这些延迟指南。如果将延迟减少到这些级别以下不易实现，请联系 Splunk 支持团队或专业服务团队，讨论调整群集节点群集超时设置以处理增加的延迟。

高级 Splunk 应用需求

相较于本主题所描述的参考规格，高级 Splunk 应用具有更高的硬件资源要求。运行高级应用（例如 Enterprise Security 或 Splunk IT 服务情报）的 Splunk Enterprise 实例必须有磁盘子系统，可产生至少 1200 平均 IOPS。

为高级应用架构部署之前，请查看用于调整和硬件建议的应用文档。以下列出了部分高级 Splunk 应用示例及其推荐硬件规格。

- Splunk Enterprise Security
- Splunk IT Service Intelligence
- Splunk App for PCI

虚拟硬件

Splunk 支持在虚拟主机环境中使用其软件。虚拟机管理程序（如 VMWare）上的索引器（具备满足其中一个硬件规格的预留资源）可处理数据的速度比布置在裸主机上的索引器慢约 10% 到 15%。在虚拟主机环境中的搜索性能与裸机上的性能近似相当。

虚拟主机提供的性能是不考虑与同一物理主机或存储阵列上共享的其他活跃虚拟主机争用资源的最佳方案。它也不考虑某些供应商的特定 I/O 增强技术，如直接 I/O 或原始设备映射。

有关如何在 VMWare 虚拟机上运行 Splunk Enterprise 的信息，请参阅 splunk.com 中的“在虚拟环境中部署 Splunk Enterprise”。

Splunk Enterprise，在 cloud 中自我管理

在 cloud 中运行 Splunk Enterprise 是在裸机硬件上本地运行的替代方式。Splunk Enterprise 在基于 cloud 的基础架构上提供与裸机硬件上的类似性能。取决于供应商和用于配置 cloud 实例的技术，可用资源可能会比 OS 报告的更少。

如果您在 Amazon Web Services (AWS) 实例上运行 Splunk Enterprise：

- AWS 测量虚拟 CPU (vCPU) 中弹性计算 Cloud (EC2) 实例的 CPU 功率，而不是真实 CPU 中的功率。
- 每个 vCPU 是大多数 AWS 实例类型上的 Intel Xeon 核心的一个超线程。请参阅 AWS 网站上的“AWS | Amazon EC2 | 实例类型”。
- 作为核心的超线程，vCPU 被用作核心，但物理核心必须在物理核心处理的其他 vCPU 的其余工作量之间安排工作量。

对于索引和数据存储，请注意以下事项：

- 如果您选择使用弹性块存储 (EBS)，您选择的 EBS 卷类型决定了您可获得的性能量。
- 不是所有的 EBS 卷类型都有处理 Splunk Enterprise 操作所需要的 IOPS。
- “配置 IOPS”和“磁性”EBS 卷类型为获得索引和搜索所需要的 IOPS 提供了最佳选择。请参阅 AWS 网站上的“EBS - 产品详细信息”。
- 不是每一个 EC2 实例类型均为您需要的 EBS 卷提供网络吞吐量。要确保该带宽，您必须将实例作为“EBS-优化”启动，或者选择一个能提供最小 10Gb 带宽的实例类型。请参阅 AWS 网站上的“Amazon EC2 实例配置”。

对于转发，请注意，您的 Cloud 基础架构邻近您的转发器会对整个环境的性能有重大影响。

有关在 AWS 中运行 Splunk Enterprise 的建议，请参阅 splunk.com 上的“在 Amazon Webservices 上部署 Splunk Enterprise”。

Splunk Cloud

Splunk 将它的计算机数据平台和获许可的软件作为名为 Splunk Cloud 的订阅服务提供。当您订阅服务时，您可以购买容量以索引、存储和搜索您的计算机数据。Splunk Cloud 对您的基础架构规格进行提取摘要，并在您所购买的容量上提供高性能服务。

要了解 Splunk Cloud 的其他信息，请访问 Splunk Cloud 网站。

在合作基础架构上部署 Splunk 软件的注意事项

许多硬件供应商和云供应商致力于新建参考架构和介绍如何在基础架构中部署 Splunk Enterprise 和其他 Splunk 软件的解决方案指导。为了您使用方便，Splunk 维持单独的页面，在此页面上，Splunk Technology Alliance Partners (TAP) 可以提交符合甚至超出文档参考硬件标准规格的参考架构和解决方案指导。请参阅 Splunk 网站的 Splunk Partner 解决方案页面。

Splunk 与 TAP 并用以确保他们的解决方案满足标准时，Splunk 不支持任何特定硬件供应商或技术。

确定何时调整您的 Splunk Enterprise 部署规模

在考虑何时调整及如何调整环境之前，评估需要建立索引的数据量和会搜索该数据的用户数。

性能问卷

该问卷假定您有一个基于参考架构（在“用于单实例部署的参考计算机”主题中有介绍）的单实例 Splunk Enterprise 部署。这些指南会帮助您决定何时分发您的 Splunk 平台部署。

问题 1：您每天是否需要索引超过 2GB 的数据？

问题 2：您是否有超过两位用户同时登录？

如果您对问题 1 和 2 都回答**否**，则您的 Splunk 平台实例可与其他 Splunk 平台服务共享用于“分布式部署的参考计算机”。

如果您对问题 1 或 2 回答**是**，则继续问题 3。

注意 在 Windows 操作系统上部署 Splunk Enterprise 时，不要使用提供 Active Directory 或 Exchange 服务或运行计算机虚拟化软件的主机。这些服务一般会耗费很多 I/O 资源，从而会降低 Splunk Enterprise 的索引和搜索性能。

问题 3：您每天是否需要索引超过 300GB 的数据？

问题 4：您是否需要超过四位并发用户？

如果您对问题 3 和 4 都回答**否**，则运行在参考计算机上的单个专用 Splunk Enterprise 实例可提供足够的资源来处理索引和搜索的工作负载。转到问题 5。

如果您对问题 3 或 4 回答**是**，则扩展 Splunk Enterprise 部署到多台计算机，以处理不断增加的索引和搜索需求。转到问题 5。

问题 5：您是否需要超过 600GB 的总存储空间？

请参阅“Splunk Enterprise 如何计算磁盘存储”。

如果回答**否**，则单个参考架构的专门计算机应能够处理索引和搜索的工作负载，但您可考虑为计算机添加额外存储以应对因保存更高而导致的磁盘使用率的增加。转到问题 6。

如果您回答**是**，则扩展 Splunk Enterprise 部署到多台计算机，以处理不断增加的索引和搜索需求。转到问题 6。

问题 6：您是否希望新建或运行执行超过 8 个并发的保存搜索的 Splunk 应用、告警或解决方案？

问题 7：您是否需要为少量结果（不到 1%）搜索大量数据？

如果您对问题 6 和 7 回答**否**，则可能此次的 Splunk Enterprise 部署中不需要多个索引器。

如果您对问题 6 或 7 回答**是**，则扩展 Splunk Enterprise 部署到多台计算机，以处理不断增加的索引和搜索需

求。

性能建议摘要

“每日索引量”表格汇总了性能检查表给出的性能建议。下表显示了在 Splunk Enterprise 中索引和搜索数据所需的参考计算机的数量，这取决于并发用户计数和实例索引的数据量。

满足参考硬件需求的索引器可以在支持搜索负载时每天最多可插入 300GB。要查看当前参考硬件的规格，请参阅本手册中的“参考硬件”。

此表格仅为指导原则。根据用例修改相关数字。如果您需要帮助确定和调整 Splunk 平台环境，请联系 Splunk 销售代表或“专业服务”。

每日索引量						
	小于 2GB/天	2 到 300 GB/天	300 到 600 GB/天	600GB 到 1TB/天	1 到 2TB/天	2 到 3TB/天
总用户数：小于 4	1 个合并实例	1 个合并实例	1 个搜索头， 2 个索引器	1 个搜索头， 3 个索引器	1 个搜索头， 7 个索引器	1 个搜索头， 10 个索引器
总用户数：最大 8	1 个合并实例	1 个搜索头， 1 个索引器	1 个搜索头， 2 个索引器	1 个搜索头， 3 个索引器	1 个搜索头， 8 个索引器	1 个搜索头， 12 个索引器
总用户数：最大 16	1 个搜索头， 1 个索引器	1 个搜索头， 1 个索引器	1 个搜索头， 3 个索引器	2 个搜索头， 4 个索引器	2 个搜索头， 10 个索引器	2 个搜索头， 15 个索引器
总用户数：最大 24	1 个搜索头， 1 个索引器	1 个搜索头， 2 个索引器	2 个搜索头， 3 个索引器	2 个搜索头， 6 个索引器	2 个搜索头， 12 个索引器	3 个搜索头， 18 个索引器
总用户数：最大 48	1 个搜索头， 2 个索引器	1 个搜索头， 2 个索引器	2 个搜索头， 4 个索引器	2 个搜索头， 7 个索引器	3 个搜索头， 14 个索引器	3 个搜索头， 21 个索引器

转发器与索引器之间的比例

Splunk Enterprise 索引器负责从内部和外部数据源（例如转发器）接收数据流，并在本地为该流建立索引。为数据建立索引需要大量的磁盘 I/O 带宽以及一些计算资源。当您考虑一个索引器能处理多少个转发器的数据时，索引容量仍然是最关注的问题。

转发器（一个索引器能从其接收数据）的数量取决于几个因素：

- 计算机上的 CPU 核心数。核心数应该满足或超过“参考标准”。
- 计算机上磁盘主轴的数量。主轴的数量应该满足或超过“参考标准”。
- 索引器是否运行在 Windows 或 *nix 上。
- 要转发到索引器的数据量。
- 索引器是否也会用作部署服务器。

对于 *nix 索引器，转发器与索引器之间的比例

Splunk Enterprise 使用以下设置来为能连接到 *nix 索引器的转发器数量提供指导：

- 一个索引器（具备 8 个核心，7GB RAM，采用 RAID 0 配置的 4 x 420GB 磁盘，运行 64 位 Linux OS）。
- 以 100Mb/s 或更快速度传输的高速局域网 (LAN)。
- 所有的通用转发器发送未经预先处理的数据。

在这些情况下，索引器能够处理至少 2,000 个转发器的数据，通常能处理多达 5,000 个转发器的数据。

当服务器被配置为接受大量的 Unix 文件描述符（通常是索引器可以接受的转发器数量的三到四倍）时，性能最佳。

注意：这些数字仅用于指导。取决于索引器、转发器和网络的配置，结果会有所不同。

并行化设置

在 Splunk Enterprise 中，可使用新设置来提高搜索和索引性能。

谁能使用这些设置

并行化设置被设计来提高 Splunk Enterprise 中特定组件的性能。并行化功能的目的是使得具有多余的 CPU 核心和 I/O 容量的客户能充分利用他们的硬件提高跨索引层的性能。您可以使用这些设置来分配 CPU 资源，以用于 Splunk 平台环境中最广泛的用途，方法是将索引器调整以满足其要求。

设置摘要

设置	描述
批处理模式搜索并行化	允许批处理模式搜索在每个索引器上打开其他的搜索管道，同时处理多个数据桶。
数据模型的并行化摘要	允许计划程序在索引器上运行并发数据模型加速搜索。
报表加速的并行化摘要	允许计划程序在索引器上运行并发报表加速搜索。
索引并行化	允许在索引器和转发器上的并发数据处理管道。

如果您的 Splunk 平台环境中的索引器超过了“参考硬件”规格，您可以查看用例并增加一个并行化设置直到达到最大建议值。如果您的索引器达到或接近容量极限，更改并行化设置会对搜索和索引性能产生负面影响。所有并行化设置都需要重新启动服务使其生效。

批处理模式搜索并行化

批处理模式搜索被设计用于按数据桶而不是按时间进行搜索和返回事件数据。通过添加更多的批处理搜索管道，可同时处理多个数据桶，加快搜索结果的返回速度。使用批处理模式搜索并行化的客户会发现返回批处理模式搜索结果的速度快了一倍。

设置名称	默认	最大建议值	影响
`batch_search_max_pipeline`	1	2	乘以每个索引器上每个批处理模式搜索的搜索管道数量。

将 `limits.conf` 中的 `batch_search_max_pipeline` 设置调整为 2，乘以每个索引器上的批处理模式搜索所使用的 IO、处理开销和内存。2 倍的值提供最佳的性能增长，更高的值则会使得收益递减。有关配置的详细信息，请参阅 Splunk Enterprise *《知识管理器手册》* 中的“配置批处理模式搜索并行化”。

Splunk 管理员可使用监视控制台监视和跟踪索引器资源的使用情况。有关详细信息，请参阅*监视 Splunk Enterprise* 中的“关于监视控制台”。

并行摘要

加速搜索有两种类型：数据模型加速和报表加速。两种加速类型都会在每个索引数据桶之外的磁盘上新建搜索结果。当计划的加速搜索无法跟上索引的数据量时，延迟被引入到搜索结果中。通过允许计划程序在索引器上运行并行加速搜索，可同时处理多个数据桶，加快加速搜索结果的新建速度。使用并行化摘要的客户会发现构建加速搜索结果的速度快了一倍。

数据模型加速

设置名称	默认	最大建议值	影响
`acceleration.max_concurrent`	3	3	乘以每个索引器上每个数据模型的计划的加速搜索数量。

`datamodels.conf` 中的 `acceleration.max_concurrent` 设置默认为 3，乘以每个索引器上运行计划的加速搜索时所使用的 IO、处理开销和内存。3 倍的值提供最佳的性能增长，更高的值则会使得收益递减。有关配置的详细信息，请参阅 Splunk Enterprise *《知识管理器手册》* 中的“并行化摘要”

报表加速

设置名称	默认	最大建议值	影响

设置名称	默认	最大建议值	影响
auto_summarize.max_concurrent	1	2	乘以每个索引器上每个搜索的计划的加速搜索数量。

将 savedsearches.conf 中的 auto_summarize.max_concurrent 设置调整为 2，乘以每个索引器上运行计划的加速搜索时所使用的 IO、处理开销和内存。2 倍的值提供最佳的性能增长，更高的值则会使得收益递减。有关配置的详细信息，请参阅 Splunk Enterprise *《知识管理器手册》* 中的“使用并行化摘要加速报表摘要的新建和维护”

Splunk 管理员可使用监视控制台监视和跟踪索引器资源的使用情况。有关详细信息，请参阅*监视 Splunk Enterprise* 中的“关于监视控制台”。

索引并行化

索引并行化允许索引器维护多个管道集。管道集处理数据的过程，从接收事件流，通过事件处理，到将事件写入磁盘。通过允许索引器新建和操作多个管道，多条数据流可以使用其他的 CPU 核心来处理，加快数据分析和写入磁盘的速度，直到达到索引器的 I/O 容量极限。使用索引并行化的客户会发现索引器的持续负载的增加，或是突然从转发器接收到大量数据时索引速度快了一倍。

设置名称	默认	最大建议值	影响
parallelIngestionPipelines	1	2	乘以每个索引器上的管道数量。

将 server.conf 中的 parallelIngestionPipelines 设置调整为 2 将使用额外的 4 到 6 个 CPU 核心，以及需要 300-400 IOPS 以保持每个索引器上的索引吞吐量。另外，可用于搜索处理的 CPU 核心更少。2 倍的值提供最佳的性能增长，更高的值则会使得收益递减。有关配置的详细信息，请参阅 Splunk Enterprise *《管理索引器和索引器群集手册》* 中的“管理用于索引并行化的管道集”

Splunk 管理员可使用监视控制台监视和跟踪索引器资源的使用情况。有关详细信息，请参阅*监视 Splunk Enterprise* 中的“关于监视控制台”。