Splunk® Enterprise 8.2.0

数据导入

生成时间：2021 年 5 月 24 日，14:21

Table of Contents

简介

什么数据可以建立索引？

Splunk 平台可以**索引**任何类型的数据。特别是，Splunk 平台可以索引任意的 IT 流、计算机和历史数据，例如 Microsoft Windows 事件日志、Web 服务器日志、实时应用程序日志、网络源、**指标**、变更监视、消息队列、归档文件等。

Splunk Cloud 中的数据来源类型

Splunk Cloud 提供了一些用于配置多种类型数据导入的工具，其中的许多工具只用于特定的应用程序需求。Splunk Cloud 还提供了一些用于配置任意数据导入类型的工具。通常，可以按下列方式对 Splunk Cloud 输入进行分类：

- 文件和目录
- 网络事件
- Windows 数据来源
- HTTP 事件收集器 (HEC)
- 指标

文件和目录

很多数据直接来自文件和目录。您可以使用**通用**和**重型**转发器来监视这些文件和目录并将它们发送到 Splunk Cloud。作为最佳实践，在您想要监视文件和目录的每台计算机上安装通用转发器，并将该数据发送到重型转发器，然后重型转发器将数据发送到 Splunk Cloud。要监视文件和目录，请参阅“获取文件和目录的数据”。

网络事件

您可能希望从网络端口收集数据，例如来自运行 syslog 的计算机的网络数据。要在 Splunk Cloud 中执行此操作，请使用重型或通用转发器来收集网络数据，然后将该数据发送到 Splunk Cloud。要从网络端口获取数据，请参阅“从 TCP 和 UDP 端口获取数据”。

Windows 数据来源

要将数据从 Windows 源导入 Splunk Cloud，请在您的通用转发器上安装 Splunk Add-on for Windows。在这种情况下，您可以使用部署服务器将 Splunk Add-on for Windows 传送到您要监视的 Windows 计算机。该加载项将收集数据并将其发送到 Splunk Cloud。

有关将 Windows 数据导入 Splunk Cloud 的更多信息，请参阅 Splunk Cloud *《管理员手册》*中的“将 Windows 数据导入 Splunk Cloud”。

HTTP 事件收集器

在 Splunk Cloud 中，您可以使用 HTTP 事件收集器直接从有 HTTP 或 HTTPS 协议的数据来源获取数据。有关更多信息，请参阅“HTTP 事件收集器端点”。

指标

您还可以从技术基础设施、安全系统和业务应用程序中获取指标数据。有关更多信息，请参阅“指标”。

Splunk Enterprise 中的数据来源类型

由于 Splunk Enterprise 在本地，您可以直接将数据导入实例，也可以使用通用或重型转发器将数据导入。一般来说，您可以将 Splunk Enterprise 输入分类如下：

- 文件和目录
- 网络事件
- Windows 数据
- 其他数据来源

文件和目录

您可以使用文件和目录**监视器**输入处理器从文件和目录中获取数据。要监视文件和目录，请参阅“获取文件和目录的数据”。

网络事件

您可以为来自任何网络端口的数据建立索引，例如，来自 syslog-ng 或任何其他通过 TCP 协议传输数据的应用程序的远程数据。它还可以为 UDP 数据新建索引，但是为了提高可靠性，应尽可能改用 TCP。

Splunk Enterprise 也可以接收 SNMP 事件和远程设备触发的告警并为其建立索引。

要从网络端口获取数据，请参阅本手册中的“从 TCP 和 UDP 端口获取数据”。

要获取 SNMP 数据，请参阅本手册中的“将 SNMP 事件发送到您的 Splunk 部署”。

Windows 数据

Windows 版本的 Splunk Enterprise 支持广泛的 Windows 特定输入。您可以使用 Splunk Web 配置以下 Windows 特定输入类型：

- Windows 事件日志数据
- Windows 注册表数据
- Windows Management Instrumentation (WMI) 数据
- Active Directory 数据
- 性能监视数据

要在非 Windows 的 Splunk Enterprise 实例上为 Windows 数据新建索引并进行搜索，您必须首先使用 Windows 实例来收集数据。请参阅“确定如何监视远程 Windows 数据的相关注意事项”。

有关在 Splunk Enterprise 中使用 Windows 数据的详细说明，请参阅本手册中的“监视 Windows 数据”。

其他数据来源

Splunk Enterprise 可以直接收集以下数据来源：

- 您可以使用 HTTP 事件收集器直接从有 HTTP 或 HTTPS 协议的数据来源获取数据。请参阅“HTTP 事件收集器端点”。
- 您还可以从技术基础设施、安全系统和业务应用程序中获取指标数据。请参阅“指标”。
- 您可以监视先进先出 (FIFO) 队列。请参阅“监视先进先出 (FIFO) 队列”。
- 您可以从 API 及其他远程数据接口和消息队列获取数据。请参阅“脚本式输入”。
- 您可以定义自定义输入功能以扩展 Splunk Enterprise 框架。请参阅“模块化输入”。

开始使用数据导入

既然您知道 Splunk 平台可对哪些类型的数据建立索引，您就可以开始将数据导入 Splunk 平台。请参阅“开始使用数据导入”。

开始使用数据导入

如需开始将数据导入您的 Splunk 部署中，请通过配置一个输入使您的部署指向某些数据。您可通过多种方式导入数据。对于最直接的选项，请使用 Splunk Web。对于 Splunk Cloud 部署，您可能需要配置重型转发器或通用转发器以将数据发送到您的 Splunk Cloud 实例。

您还可以下载并启用一个应用，如 Splunk App for Microsoft Exchange 或 Splunk IT Service Intelligence。请参阅“使用应用和加载项导入数据”。

在您配置好输入或启用应用之后，您的 Splunk 部署会存储并处理指定数据。您可以转到“搜索和报表”应用或主应用页面，并开始详细浏览您已收集的数据。

了解您的需求

在开始向部署添加输入之前，请先问自己以下问题：

问题	文档
我想要为哪种数据建立索引?	什么数据可以建立索引?
是否有相关应用?	使用应用导入数据
数据位于何处? 位于本地还是远程?	我的数据位于何处?
我是否应该使用转发器来访问远程数据?	使用转发器导入数据

我希望如何处理索引数据？	Splunk 知识是什么？

添加新输入

要添加数据，请执行下列高级步骤：

1. 新建一个测试索引并添加几个输入。所有已添加到测试索引的数据都将计入用于许可授权的最大每日索引量。
2. 在将数据提交到测试索引之前，预览并修改为您的数据建立索引的方式。
3. 查看您用“搜索和报表”应用添加的测试数据。问自己以下这些问题：
    - 您所看到的数据是否符合预期？
    - 该默认配置是否适用于您的事件？
    - 数据是否丢失或混杂？
    - 是否获得了最佳结果？
4. 如有必要，对您的输入和事件处理配置进行调整，直到事件符合您的预期为止。
5. 如有必要，删除测试索引中的数据，然后重新开始。
6. 当您准备好为数据建立永久索引时，请配置输入以使用默认 `main` 索引。

您可以重复这项任务，继续添加其他输入，在此过程中让自己熟悉导入数据的过程。

索引自定义数据

Splunk 平台可以为任何时间序列数据建立索引，通常无需其他配置。如果您已从自定义应用程序或设备获取日志，先使用默认配置对其进行处理。如果您未获得想要的结果，可以通过稍作调整来确保软件正确地对您的事件建立索引。

请参阅“事件处理概览”和“索引如何工作”，以便您决定如何让 Splunk 平台处理您的数据。

然后，请考虑以下几种收集数据的方案：

- 您数据中的事件不止一行吗？请参阅“配置事件换行”。
- 您的数据是否属于不常见的字符集？请参阅“配置字符集编码”。
- Splunk 平台是否无法正确决定时间戳？请参阅“时间戳分配如何工作”。

有关配置数据输入和将数据导入 Splunk 平台的更多信息

有关探索和进一步配置数据的一些方法，请参阅下表：

任务	文档
配置输入	其他数据导入方式
将数据添加到 Splunk 部署	您要如何添加数据？
尝试添加测试索引	使用测试索引测试输入
添加来源类型	将正确的来源类型分配给您的数据
配置事件处理	Splunk Enterprise 如何处理您的数据
从 Splunk 部署中删除数据	删除索引数据并重新开始
使用默认索引配置输入	配置输入以使用默认索引

我的数据位于本地还是远程？

如果您使用的是 Splunk Cloud 或在云端运行 Splunk Enterprise，则所有建立索引的数据均为远程数据。如果您在本地使用 Splunk Enterprise，您的数据是本地数据还是远程数据取决于许多因素：

- Splunk Enterprise 实例所在的操作系统。
- 数据存储在哪个物理磁盘上？
- 已连接到 Splunk Enterprise 实例的数据存储类型。
- 您是否需要执行验证或其他中间操作，才能访问其中包含您想要新建为目录的数据的数据存储。

本地数据

本地资源是指您的 Splunk Enterprise 实例可以直接访问的固定资源。您可访问本地资源及其中的任何内容，无需附加、连接或执行其他任何中间操作（如验证或映射网络驱动器）。如果您的数据位于此类资源上，则该数据为本地数据。

以下是本地数据的一些示例：

- 台式电脑、笔记本电脑或服务器主机上的硬盘或固态驱动器上存储的数据。
- 已通过高带宽物理连接（计算机可于开机时访问）永久安装的某个资源上的数据。
- RAM 磁盘上的数据。

远程数据

远程数据是指不满足本地资源定义的任何资源。存在于这类资源上的数据即为远程数据。

以下是远程资源的一些示例：

- Windows 主机上的网络驱动器。
- Active Directory 架构。
- NFS 或 *nix 主机上其他基于网络的装入点。
- 大多数基于云的资源。

远程数据的例外情况

在某些情况下，资源可能被视为远程资源，但实际是本地资源：

- 主机上的某个卷已通过高带宽物理连接（如 USB 或 FireWire）永久安装。由于计算机在开机时能安装该资源，Splunk Enterprise 因此将其视为本地资源，即使理论上与该资源的连接稍后可能会断开。
- 主机上的某种资源已通过高带宽网络标准（如 iSCSI 或光纤存储区域网络）永久安装。因为该标准将此类卷视为本地块设备，所以此类资源被视为本地。

使用转发器将数据导入 Splunk Enterprise

Splunk **转发器**会获取数据并将数据发送至一个索引器。转发器需要的资源最少，而且对性能影响很小，因此通常驻留在生成数据的计算机上。

例如，如果您有很多 Apache Web 服务器将生成数据，并且您希望集中搜索这些数据，您可以在 Apache 主机上设置转发器。转发器可以获取 Apache 数据并将其发送至您的 Splunk Enterprise 部署，方便您为数据建立索引；数据会在此过程中合并和存储，成为可供搜索的数据。由于转发器占用的资源空间减少，因此对 Apache 服务器性能的影响微乎其微。

同样，您可以在员工的 Windows 台式计算机上安装转发器。这些转发器会将日志和其他数据发送至您的 Splunk Enterprise 部署，您可以在部署中查看整个数据，以追踪恶意软件或其他问题。Splunk App for Windows Infrastructure 依赖此类部署。

转发器执行的操作

从远程计算机中获取数据。与原始网络源不同，转发器具有以下功能：

- 标记元数据（数据来源、来源类型和主机）
- 缓冲数据
- 压缩数据
- 使用 SSL 安全性
- 使用任何可用的网络端口
- 本地运行脚本式输入

转发器通常不会为数据建立索引，只会把数据发送至 Splunk Enterprise 部署，由该部署为数据建立索引并执行搜索。Splunk Enterprise 部署可以处理来自很多个转发器的数据。有关转发器的详细信息，请参阅*《转发数据手册》*或*《通用转发器手册》*。

在大多数 Splunk Enterprise 部署中，转发器都是作为数据的主要使用者。大型 Splunk Enterprise 部署中可能会有数百个甚至数千个转发器在获取数据，并将数据转发到其他地方进行整合。

如何配置转发器输入

以下是为 Splunk Enterprise 配置转发器输入步骤的高级概览。

1. 配置一个 Splunk Enterprise 主机以接收数据。
2. 决定要在一个含数据的主机上安装哪种转发器。

- 您既可以使用重型转发器，也可以使用通用转发器；重型转发器为完整的 Splunk Enterprise 实例，转发功能呈开启状态，而通用转发器则为单独的安装软件包。
- 您使用的转发器类型取决于主机的性能要求以及您是否需要转换导入 Splunk Enterprise 的任何数据。

- 为此平台或含数据的主机架构下载 Splunk Enterprise 或通用转发器。
- 把转发器安装到主机上。
- 在主机上启用转发并指定一个目标
- 为您想要从主机中收集的数据配置输入。如果转发器为完整的 Splunk Enterprise 实例，您可以使用 Splunk Web。
- 确认来自转发器的数据会到达接收索引器。

有关如何对转发和接收进行配置的详细信息，请参阅*《转发数据手册》*或*《通用转发器手册》*

以下是您可在转发器上配置数据导入的主要方式：

- 在初始部署转发器时指定输入。
- 如为 Windows 转发器，请在转发器安装过程中指定通用输入。
- 如为 *nix 转发器，请在安装结束后直接指定输入。
- 使用 CLI。
- 编辑 inputs.conf 文件。
- 安装其中包含您所需的输入的应用或加载项。
- 使用 Splunk Web 配置输入和部署服务器配置，以把配置后产生的 inputs.conf 文件复制到转发器。

转发器拓扑和部署

- 有关转发器的信息，包括使用案例、典型拓扑结构及配置，请参阅*《转发数据》*手册中的“关于转发和接收”。
- 有关如何部署通用转发器的详细信息，包括如何使用**部署服务器**来简化把配置文件和应用分布到多个转发器的过程，请参阅*《通用转发器》*手册中的“转发器部署拓扑结构示例”。

使用应用和加载项导入数据

Splunk **应用**和**加载项**可以扩展功能，并简化数据导入 Splunk 平台部署的过程。从 Splunkbase 下载应用。

应用通常以特定数据类型为目标，可处理从配置输入到生成有用的数据视图的整个过程。例如，Splunk App for Windows Infrastructure 提供了用于 Windows 主机管理的数据导入、搜索、报表、告警和仪表板。Splunk App for Unix and Linux 为 Unix 和 Linux 环境提供相同内容。还有各式各样处理特定应用程序数据类型的应用，包括以下应用和加载项：

- Splunk DB Connect
- Splunk Stream
- 适用于 Amazon Web Services 的 Splunk 加载项

更多有关获取和安装应用的信息

请转到 Splunkbase 以浏览可下载的大型应用集。请经常查看 Splunkbase，因为上面会不断添加新应用。

更多有关应用的信息，请参阅*《管理员手册》*中“应用和加载项是什么？”。有关如何下载和安装应用的信息，请参阅“从哪里获取更多应用和加载项”。

有关如何新建您自己的应用的信息，请参阅*《开发用于 Splunk Web 的视图和应用》*手册。

其他数据导入方式

您可以通过多种方法将数据导入 Splunk 平台实例。最佳方式取决于数据位置、数据量、您的基础设施和安全需求以及您使用数据的目。

评估需求

回答以下问题有助于帮您确定将数据导入 Splunk 实例的最佳方式。

问题	注意事项
我想要为哪种数据建立索引？	您想要建立索引的数据类型会影响数据导入方式。例如，如果您想要从专门的应用程序中导入数据，您可能想要使用使用 HTTP 事件收集器 (HEC)。另一方面，如果您想要引入 Windows 数据，您可能使用应用帮助您导入数据。请参阅“可为哪些数据新建索引？”。
是否有相关应用？	Splunk 和许多第三方开发者提供的应用可促进和改善数据引入。如果有一个应用可以用于您想要导入数据类型，您可以在配置和调整通用转发器上的输入时节省大量时间。如果存在可用于想要导入的数据类型的应用，请使用有关应用。
数据位于何处？	对于 Splunk Cloud 实例，数据始终是远程数据，这意味着您必须使用通用转发器或 HEC 才能将数据索引到 Splunk Cloud。对于 Splunk Enterprise 实例，数据可能是本地数据，也可能是远程数据。请参阅“我的数据位于本地还是远程？”。

我是否需要使用转发器来访问远程数据？	如果您有 Splunk Cloud 实例，您可能需要使用转发器。请参阅“使用转发器将数据导入 Splunk Cloud”。
我希望如何处理索引数据？	请参阅*《知识管理器手册》*中的“什么是 Splunk 知识？”。

添加数据

如需添加新的数据类型到您的 Splunk 平台实例，请配置一个数据导入。您可以使用以下方法配置数据导入：

- **应用**。您可以使用多种**应用**，这些应用为各种用例提供预先配置的输入、视图和知识对象。有关更多信息，请参阅“使用应用导入数据”。
- **Splunk Web**。您可以使用 **Splunk Web** 配置一些输入。可以从 Splunk Web 主页访问“添加数据”页面。另外，在上载文件之后，您可以预览并调整 Splunk Cloud 必须为文件建立索引的方式。请参阅“将正确的来源类型分配给您的数据”。
- **转发器**。如果数据是远程数据，您可以配置**转发器**以将数据从外围计算机发送到您的 Splunk Cloud 实例。对于非 Splunk Cloud 安装，您可以使用这些转发器将数据发送到中央**索引器**。根据操作系统，您可以在转发器安装时指定一些输入。请参阅“使用转发器将数据导入 Splunk Enterprise”。

还有其他方式可以为 Splunk Enterprise 导入数据。请参阅“将数据添加到 Splunk Enterprise”。

使用应用导入数据

Splunk 应用和**加载项**可以扩展功能，并简化数据导入 Splunk Cloud 平台部署的过程。

应用通常以特定数据类型为目标，可处理从配置数据输入到生成有用的数据视图的整个过程。例如，Splunk App for Windows Infrastructure 提供了用于 Windows 主机管理的数据导入、搜索、报表、告警和仪表板。Splunk App for Unix and Linux 为 Unix 和 Linux 环境提供相同内容。大多数 Splunk 应用直接与 Splunk Cloud 配合使用，而其他应用可能需要您将它们安装在通用或重型转发器上才能将数据发送到 Splunk Cloud 实例。

您可以在 Splunkbase 上下载这些应用：

- Splunk App for Windows Infrastructure
- Splunk App for Unix and Linux

您还可以下载应用来处理特定类型的应用程序数据。下面是几个示例：

- Splunk DB Connect
- Splunk Stream
- Splunk Add-on for Amazon Web Services

使用 Splunk Web

您可以从 Splunk Web 主页或通过选择**设置** > **数据导入**来添加数据导入。

- 在 Splunk Web 主页中，单击**添加数据**。
- 选择**设置** > **添加数据**。
- 再从**设置**下拉列表的**数据**部分选择**设置** > **数据导入**。

您可以在**添加数据**页面上选择不同的选项来导入数据。单击图标可转到定义您要上载、监视或转发数据的页面。有关更多信息，请参阅以下主题：

- 上载数据
- 监视数据

有关如何在 Splunk Web 中添加数据的更多帮助信息，请参阅“您想如何添加数据？”。

将数据添加到 Splunk Enterprise

借助 Splunk Enterprise，您可以使用 Splunk Web 或 Splunk 应用添加数据。除了这些方法，您还可以使用以下方法。

- **Splunk 命令行界面 (CLI)**。此方法可用于将数据导入 Splunk Enterprise。您可使用 CLI 配置大多数导入类型。您还可以在重型转发器上使用 CLI 来将数据导入 Splunk Cloud。
- **inputs.conf 配置文件。**使用 Splunk Web 或 CLI 指定您的输入时，详细内容保存在 Splunk Enterprise 索引器和重型转发器实例上的**配置文件**中。此选项在 Splunk Cloud 上不可用，但您可以使用重型转发器将数据直接发送到您的

Splunk Cloud 实例。您可以直接在索引器和重型转发器上编辑配置文件，一些高级数据导入需求可能需要您对其进行编辑。

Splunk Enterprise**添加数据**页面有数据导入的附加选项：

- 上载数据
- 监视数据
- 转发数据

使用 CLI

在 Splunk Enterprise 和通用转发器上，可以用 Splunk **CLI** 来配置多数输入。从 shell 或命令提示符导航到 $SPLUNK_HOME/bin/ 目录，并使用 ./splunk 命令。例如，下列命令会把 /var/log/ 添加为数据导入：

```
splunk add monitor /var/log/
```

有关 CLI 的更多信息，包括如何获取命令行帮助，请参阅*《管理员手册》*中的“关于 CLI”。

编辑 inputs.conf 配置文件

在 Splunk Enterprise 和通用转发器上，可以编辑 input.conf 文件来配置您的输入。您可以使用文本编辑器新建或修改此文件，您可以在其中为每个输入添加一个**段落**。可以向 $SPLUNK_HOME/etc/system/local/ 中或 $SPLUNK_HOME/etc/apps/<app name>/local/ 中您的自定义应用程序目录中的 input.conf 文件添加段落。

要配置数据导入，可向其段落添加键/值对。可在输入段落中设置多个设置。如果您未指定某个设置的值，Splunk Enterprise 会使用默认设置值。input.conf 文件中所有设置的默认值都在 inputs.conf 配置规范文件中。请参阅*《管理员手册》*中的“inputs.conf 规范文件”。

如果您之前未使用过配置文件，请在添加输入前参阅“关于配置文件”。

inputs.conf 配置文件段落示例

以下配置示例将指示 Splunk Enterprise 在 TCP 端口 9995 上侦听来自任何远程主机的原始数据。Splunk Enterprise 使用远程主机的 DNS 名称设置数据的主机。它将来源类型 log4j 和来源 tcp:9995 分配到数据。

```
[tcp://:9995]
connection_host = dns
sourcetype = log4j
source = tcp:9995
```

有关如何配置特定输入的信息，请参阅本手册中涉及该特定输入的主题。例如，要配置文件输入，请参阅“使用 inputs.conf 监视文件和目录”。

有关每个数据导入的主题均介绍了该输入可用的主要属性。可用属性的完整列表，包括属性及几个示例的描述，请参阅*《管理员手册》*中的“inputs.conf 规范文件”。

如需开始把数据导入您的 Splunk 部署中，请通过配置一个输入使您的 Splunk 部署指向某些数据。配置输入的方法有很多种。最简单的方法是使用 Splunk Web。

您还可以下载并启用一个应用，如 Splunk App for Microsoft Exchange 或 Splunk IT Service Intelligence。有关更多信息，请参阅“使用应用导入数据”。

应用上下文如何决定 Splunk Enterprise 写入配置文件的位置

当您通过 Splunk Enterprise 上的 Splunk Web 添加输入时，软件会将该输入添加到 input.conf 配置文件的副本中。应用程序上下文，亦即您当前正在其内配置前述输入的 Splunk 应用，决定着 Splunk Enterprise 写入 inputs.conf 文件的位置。

例如，如果您直接从“搜索”页面导航到“设置”页面，然后再添加一项输入，Splunk Enterprise 将把该输入添加到 $SPLUNK_HOME/etc/apps/search/local/inputs.conf，因为 Splunk Enterprise 位于搜索和报表应用中。

添加输入时请确认自己位于想要的应用上下文内。如需介绍配置文件工作方式的背景信息，请参阅 Splunk Enterprise*《管理员手册》*中的“有关配置文件”。

另请参阅

- 转到 Splunkbase 以安装和下载应用。
- 有关什么是应用的详细信息，请参阅*《管理员手册》*中“应用和加载项”。特别是，请参阅“从哪里获取更多应用和加载

项”以下载和安装应用。
- 有关如何新建您自己的应用的信息，请参阅*《开发用于 Splunk Web 的视图和应用》*手册。

Splunk® Enterprise 如何处理您的数据

Splunk® Enterprise 可获取数据并为其**建立索引**，把数据转换为**事件**形式的可搜索知识。**数据管道**显示建立索引期间操作数据的进程。这些进程组成了**事件处理**。将数据处理成事件后，您可以把事件与**知识对象**关联起来，以增加其实用性。

数据管道

传入数据通过数据管道移动。有关更多详细信息，请参阅“数据如何通过 Splunk 部署移动：*《分布式部署手册》*中的“数据如何通过 Splunk 部署：数据管道”。

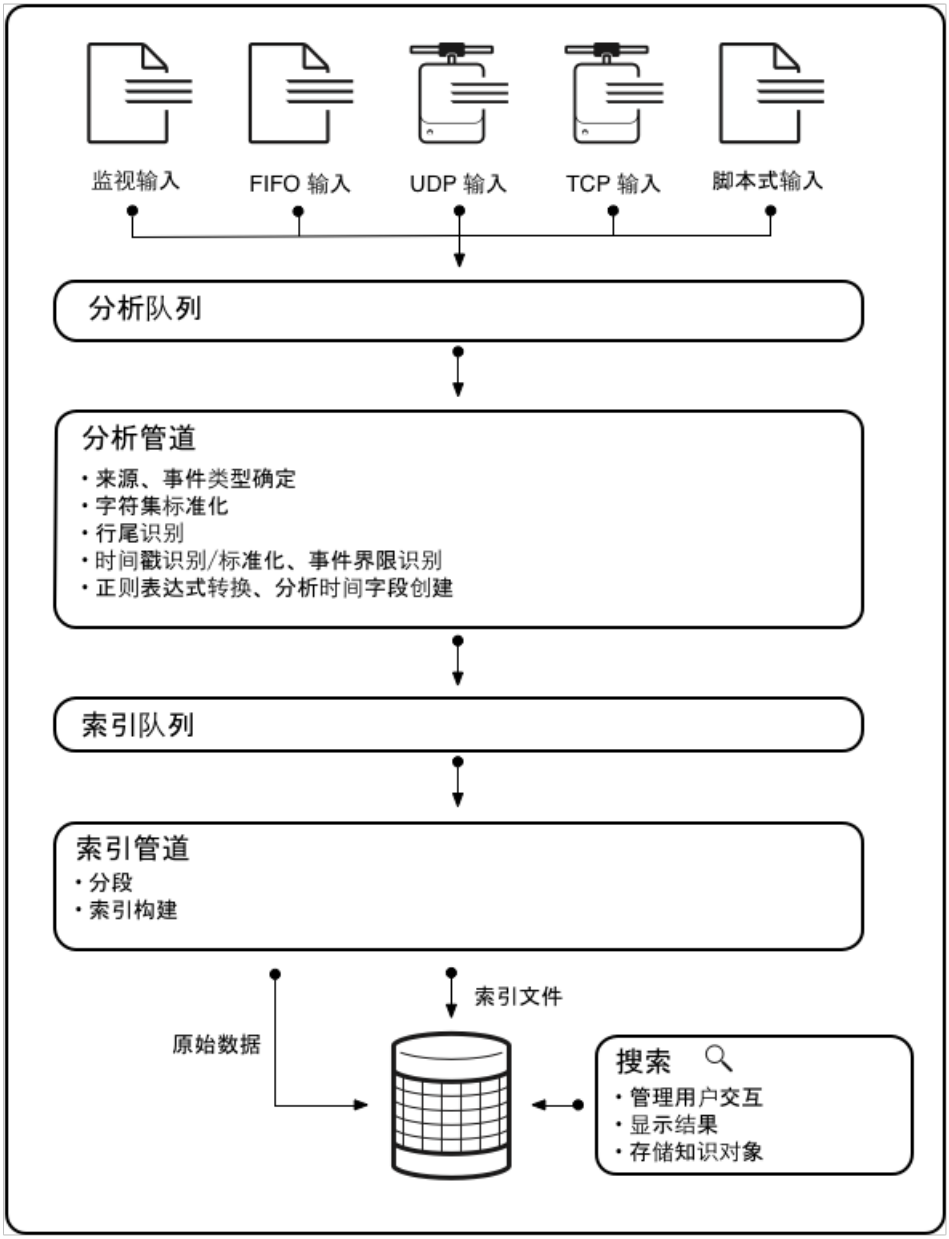
每个 Splunk® Enterprise 处理**组件**都驻留在三个典型的处理层其中一个上：数据导入层、索引层和搜索管理层。这些层共同支持发生在数据管道中的进程。

当数据沿着数据管道移动时，Splunk® Enterprise 组件将数据从其外部原始来源（例如日志文件和网络源）转换为封装有价值知识的可搜索事件。

数据管道包括以下几段：

- **输入**
- **分析**
- **索引**
- **搜索**

下图显示了数据管道中的主要步骤。在数据导入层，Splunk® Enterprise 使用来自各种输入的数据。接着，在索引层，Splunk® Enterprise 会对数据进行检查、分析及转换。然后，Splunk® Enterprise 会获取分析后的事件并将其写入磁盘上的索引中。最后，搜索管理层可管理用户如何访问、查看及使用索引数据的所有方面。



事件处理

事件处理将出现在两个阶段：分析和新建索引。所有数据以大数据块的形式通过分析管道进入。分析期间，Splunk 平台会将这些数据块分为若干事件。然后它将这些事件传递到执行最终处理的索引管道。

在分析和建立索引期间，Splunk 平台会转换数据。您可配置其中大部分进程以根据需要进行调整。

Splunk 平台将在分析管道中执行大量操作。除了相关信息外，以下表格还提供一些示例：

操作	相关信息
为每个事件提取一组默认字段，包括 host、source 和 sourcetype。	关于默认字段
配置字符集编码。	配置字符集编码
使用换行规则识别行尾。还可以使用 Splunk Web 中的“设置来源类型”页面交互修改行尾设置。	配置事件换行 将正确的来源类型分配给您的数据
标识或新建时间戳。Splunk 软件在处理时间戳的同时会识别事件界限。可以使用 Splunk Web 中的“设置来源类型”页面交互修改时间戳设置。	时间戳分配如何工作 将正确的来源类型分配给您的数据

根据您的配置使数据匿名。在此阶段，您可以用掩码显示敏感数据（如信用卡号或社会保险号）。	使数据匿名
根据您的配置，将自定义元数据应用到传入事件。	动态分配默认字段

在索引管道中，Splunk 平台会执行其他处理。例如：

- 将所有事件分段，然后基于段执行搜索。您可以确定分段的级别，它将影响索引和搜索速度、搜索功能以及磁盘压缩效率。请参阅“关于事件分段”。
- 构建索引数据结构。
- 将原始数据和索引文件写入磁盘，其中将执行后索引压缩。

分析管道与索引管道之间的区别主要对于转发器非常重要。重型转发器能够在本地完全分析数据，然后将分析后的数据转发到接收索引器上执行最终的索引建立。通用转发器在特定案例（如处理结构化数据文件）中提供最低限度的分析。其他分析将在接收 Splunk® Enterprise 索引器上进行。

有关事件及其在索引新建过程中所发生情况的信息，请参阅“事件处理概述”。

使用知识对象增强并优化事件

在将数据转换为事件之后，您可以将其与知识对象（如事件类型、字段提取和报表）相关联，以使事件变得更加有用。有关管理 Splunk 软件知识的信息，请参阅*《知识管理器手册》*（从“什么是 Splunk 知识？”开始）。

如何将数据导入 Splunk 部署

您要如何添加数据？

将数据添加到您的 Splunk Cloud 实例或 Splunk Enterprise 部署的最快方法是使用 Splunk Web。

“添加数据”页面

要将数据添加到 Splunk 平台，请按照以下步骤访问 Splunk Web 中的“添加数据”页面：

1. 登录 Splunk Web，将显示主页。
2. 单击**添加数据**以访问“添加数据”页面。
   如果您的搜索头是**搜索头群集**的一部分，则不会显示“添加数据”页面。有关更多信息，请参阅 Splunk Enterprise *《分布式搜索》*手册中的“关于搜索头群集化”。
3. 访问“添加数据”页面后，从以下三个选项中选择一个，以通过 Splunk Web 将数据导入您的 Splunk 平台部署：
   - 上载
   - 监视器
   - 转发

上载

“上载”选项允许您上载文件或文件的归档，以便为其建立索引。选择“上载”选项时，Splunk Web 将打开上载进程的页面。有关更多详细信息，请参阅“上载数据”。

监视器

对于 Splunk Cloud 部署，您可以使用 HTTP 事件收集器监视文件和目录。对于 Splunk Enterprise 安装，“监视”选项允许您监视一个或多个文件、目录、网络数据流、脚本、事件日志（仅在 Windows 主机上）、性能指标或任何其他 Splunk Enterprise 实例可访问的计算机数据类型。选择“监视”选项时，Splunk Web 会加载一个启动监视进程的页面。请参阅“监视数据”。

转发

如果您在 Splunk Cloud 环境中，使用转发器是导入数据的最常用方法。“转发”选项允许您把数据从**转发器**接收到您的 Splunk Cloud 部署中。选择“转发”选项时，Splunk Web 会将您带到一个从转发器启动数据集合进程的页面。请参阅“转发数据”。

“转发器”选项需要其他配置。仅在单实例 Splunk Cloud 环境中使用此选项。

Guided Data Onboarding

Guided Data Onboarding (GDO) 功能还可提供端到端指导，以将选择数据来源导入特定的 Splunk 平台部署。

从 Splunk Web 主页中，单击**添加数据**查找数据导入指南。在此，您可选择数据来源和配置类型。然后查看图像、高级步骤和文档链接，这有助于您设置和配置数据来源。

您可以单击 Splunk Enterprise 文档网站中的**添加数据**选项卡查找所有 Guided Data Onboarding 手册。

上载数据

“上载”页面允许您将文件从计算机上载到 Splunk® Enterprise 实例。您可以通过在 Splunk Web 的 Splunk 主页中单击“上载”按钮来上载文件。

Windows 事件日志文件和上载功能的注意事项

您可以将任何文件上载到 Splunk Enterprise 或 Splunk Cloud，但您从另一台 Windows 计算机导出的 Windows 事件日志 (.evt) 和 Windows 事件日志 XML (.evtx) 文件无法使用该上载功能，因为这些文件包含的信息是生成这些文件的 Windows 计算机所特定的。如果不变更格式，其他计算机将无法处理这些文件。有关处理这类文件时所受限制的更多信息，请参阅“索引已导出的事件日志文件”。

“上载”页面

1. 在此页面上通过以下任意一种方法上载数据：

- 把您想要为其建立索引的文件从桌面拖入**将您的数据文件拖到这儿**区域。
- 单击**选择文件**并选择您要建立索引的文件。

- Splunk® Enterprise 接下来会根据该文件的类型加载并处理该文件。
- 文件加载完成后，单击**下一步**。

下一步

设置 sourcetype

监视数据

如果要监控 Splunk® Enterprise 实例上的文件和网络端口，请导航到 Splunk Web 中的“监视”页面：

1. 在 Splunk Web 系统栏中，单击**设置**。
2. 选择**添加数据**页面。
3. 单击**监视**页面。

“监视”页面

在“监视”页面中，选择您希望 Splunk® Enterprise 监视的数据类型。Splunk Enterprise 会先列出默认的输入，再列出转发而来的输入，最后才是实例中的任何模块化输入。

“监视”页面仅显示您可以监视的数据来源类型，而您可以监视的数据来源类型取决于您使用的 Splunk 部署类型。如果您正在运行 Splunk Enterprise，该页面还会向您显示运行实例的平台。有关 Splunk® Enterprise 可以监视的内容的更多信息，请参阅“数据来源类型”。

添加数据导入

某些数据来源仅在特定操作系统上可用。例如，Windows 数据来源仅在运行 Windows 的计算机上可用。Splunk Cloud 无法直接监视 Windows 输入，因为它不在 Windows 上运行，但您可以将数据从运行 Windows 的通用转发器转发到 Splunk Cloud。

1. 单击来源即可选定该来源。页面更新以您已选择的来源为基础。
2. 遵照屏幕提示选择您想要监视的来源对象。
3. 单击**下一步**即可继续到“添加数据”进程的下一步。

如果您在执行这些步骤时遇到问题，则登录的 Splunk 用户帐户可能没有添加数据或查看您想要添加的数据来源的权限。

后续步骤

将正确的来源类型分配给您的数据

转发数据

使用“转发数据”页面，您可以选择和控制已连接到 Splunk 平台实例的转发器。您可以配置转发器，以便将数据发送到实例。

您在“添加数据”页面上单击**转发**按钮时，会出现此页面。在以下情况下会出现此页面：

- 您使用 Splunk Enterprise 的单个实例作为索引器和部署服务器。
- 您使用免费试用 Splunk Cloud 部署，并已配置通用转发器以作为部署客户端连接到 Splunk Cloud 实例。

如果这些场景之一适合您的情况，那么您可以按照本主题后面介绍的“使用‘选择转发器’页面以使用转发器定义和填充服务器类”过程来管理可用转发器。

如果您的 Splunk 部署中有多个进行索引的计算机，那么此页面并没有用。请参阅*《更新 Splunk Enterprise 实例》*中“关于部署服务器和转发器管理”以了解部署服务器和如何用部署服务器管理转发器配置以发送到多个索引器。

如果您有 Splunk Cloud 部署，则此页面不可用。您可以安装本地部署服务器以同步转发器配置，这样您就不必手动管理转发器。

使用“转发数据”页面的前提条件

如需使用“转发数据”页面配置数据导入，您必须至少将一个转发器配置为部署客户端。如果您尚未将转发器配置为部署客户端，该页面会通知您它找不到部署客户端。

如需将重型转发器配置为部署客户端，请参阅*《更新 Splunk Enterprise 实例》*手册中的“配置部署客户端”。

如需将通用转发器配置为部署客户端，请参阅 Splunk*《通用转发器》*手册中的“将通用转发器配置为部署客户端”。您可以在 Windows 计算机上安装转发器期间，将该转发器配置为部署客户端。请参阅 Splunk*《通用转发器》*手册中的“通过安装程序安装 Windows 通用转发器”。

使用“选择转发器”页面以使用转发器定义和填充服务器类

您从“添加数据”页面中选择**转发数据**后，将出现“选择转发器”页面。

您可定义**服务器类**并将转发器添加至这些类中。服务器类是基于架构或主机名称等的主机逻辑分组。

此页面将显示您已配置用于转发数据并用作此实例的部署客户端的转发器。如果您还未配置任何转发器，则该页面会建议您进行配置。

通过以下过程，您可以为已将自己报告为此 Splunk 平台实例的部署客户端的转发器设置服务器类。

1. 在**选择服务器类**中，单击下列选项中的任意一个。
   - 单击**新建**以新建服务器类，或当现有服务器类与您要为其配置数据输入的转发器组不匹配时。
   - 单击**现有**以使用现有服务器类。
2. 请在**可用主机**窗格中选择您想要此实例从中接收数据的转发器。转发器从**可用主机**窗格移动到**选择主机**窗格。

   服务器类必须包含特定平台的主机。例如，您不能将 Windows 和 *nix 主机放入相同的服务器类。

3. （可选）您可通过单击**添加所有**链接添加所有主机，或通过选择**删除所有**链接删除所有主机。
4. 如果您在“选择服务器类”中选择**新建**，则为服务器类输入唯一名称。否则，请从下拉列表选择您想要的服务器类。
5. 单击**下一步**。“选择来源”页面显示对您已选择的转发器有效的来源类型。
6. 选择您要转发器将数据发送到此实例的数据来源。
7. 单击**下一步**。

下一步

修改输入设置

将正确的来源类型分配给您的数据

来源类型是 Splunk 平台分配给所有传入数据的**默认字段**之一，它决定 Splunk 平台如何在索引创建期间设置数据格式。通过将正确的来源类型分配给您的数据，数据的索引版本将以所希望的方式呈现，并且采用适当的**时间戳**和**事件**换行。

您可以使用 Splunk Web 中的“设置来源类型”页面确认 Splunk 平台按照您希望的方式为数据建立索引。

为数据分配来源类型

Splunk® Enterprise 提供了许多预定义的来源类型，并尝试根据数据的格式为您的数据分配正确的来源类型。在一些情况下，您可能需要手动为数据选择其他预定义的来源类型。在其他情况下，您可能需要使用自定义的事件处理设置新建一个来源类型。

在“设置来源类型”页面上，您可以查看 Splunk® Enterprise 将如何对基于预定义来源类型应用程序的数据建立索引。您可交互式地修改设置并将这些修改保存为新的来源类型。

按照“设置来源类型”页面上的以下步骤，确保为数据分配正确的来源类型：

1. 使用默认的事件处理配置查看数据未进行任何更改时的外观。
2. 应用其他来源类型查看其是否能提供更可取的结果。
3. 修改时间戳和事件换行设置，以提高索引数据的质量并将修改保存为新来源类型。
4. 新建来源类型。

如果您使用 Splunk Enterprise，则可以将任何新的来源类型保存到 props.conf 配置文件中。您稍后可以将这些新来源类型分布到部署中的各索引器上，以供全局使用。请参阅“在 Splunk Enterprise 中分布来源类型配置”。

某些来源类型（例如“日志到指标”类别中的来源类型）无法预览。有关详细信息，请参阅本主题后面的“关于日志到指标来源类型类别”。

有关来源类型及来源类型为何如此重要的更多信息，请参阅“来源类型为何重要”。

关于 Log to Metrics 来源类型类别

“日志到指标”类别中的来源类型是特殊的来源类型。Splunk 平台使用这些来源类型在引入时将日志事件转换为指标数据点。

如果您选择此类别中的来源类型，将在“设置来源类型”页面左侧显示一组**指标**控制。有关日志到指标转换和**指标**设置的更多信息，请参阅*《指标》*手册中的“在 Splunk Web 中设置引入时日志到指标的转换”。

当您将日志到指标来源类型应用于某个输入时，将无法预览该输入的数据。

为数据分配来源类型

当“设置来源类型”页面打开时，Splunk® Enterprise 基于您指定的数据选择来源类型。您可以接受该来源类型或按照以下步骤更改来源类型。

1. 检查预览窗格以了解 Splunk® Enterprise 将如何为数据建立索引。查看事件换行和时间戳。
2. （可选）单击**查看事件摘要**可查看事件摘要。
   Splunk Web 会在新的窗口中显示该事件摘要。请参阅“查看事件摘要”。
3. 如果数据以您希望的方式显示，则单击**下一步**以转到**输入设置**页面。否则，请从以下任一选项中选择：
   - 选择一个现有来源类型以更改数据格式。请参阅“选择一个现有来源类型”。
   - 手动调整时间戳、分隔符和换行，然后将更改保存为新的来源类型。请参阅“调整时间戳和事件换行”。

选择一个现有来源类型

如果数据未以您希望的方式显示，请查看现有来源类型是否已修复该问题。

如果 Splunk® Enterprise 可检测出来源类型，则会在来源类型：<sourcetype> 下拉列表中显示该来源类型。如果无法确定源类型，则会显示“来源类型：系统默认”。

1. 设置**来源类型：<sourcetype>** 下拉列表可查看来源类型类别列表。每个类别包含该类别内的来源类型列表。
2. 请将鼠标悬停在最能代表您的数据的类别上。
   当您执行此操作时，在该类别下的来源类型会显示在下拉列表中。
3. 请选择最能代表您的数据的来源类型。
   Splunk Web 将更新数据预览窗格，以显示新来源类型下的数据外观。您可能需要滚动以查看类别中的所有来源类型。
4. 在预览窗格中再次查看您的数据。如果现有来源类型不适用于您的数据，则必须手动调整时间戳、分隔符和事件换行。请参阅“调整时间戳和事件换行”。
5. 如果结果符合您的要求，请单击**下一步**以转到**输入设置**页面。

查看事件摘要

您可以通过单击“查看事件摘要”来查看数据示例中的事件摘要。此摘要显示以下信息：

- 示例数据的大小（以字节为单位）。
- 存在于示例中的事件数。
- 代表事件随时间的分布情况的图表。Splunk 平台使用文件内的日期戳确定显示此图表的方式。
- 示例中每个事件占用的行数的明细。

调整时间戳和事件换行

如果您未成功选择一个现有来源类型，则可手动调整 Splunk® Enterprise 对传入数据处理时间戳以及事件换行的方式。

要手动调整时间戳和事件换行参数，请使用**事件换行符**、**时间戳**、**分隔的设置**和“设置来源类型”页面左窗格上的**高级**下拉列表。当您更改设置时，预览窗格会更新。

仅当 Splunk® Enterprise 无法确定如何对文件进行换行，或您选择了未定义换行的来源类型时，才会显示“事件换行”选项卡。仅当 Splunk® Enterprise 检测到您想要导入结构化数据文件或您为结构化数据（如 CSV）选择了来源类型时，才会显示“分隔的设置”选项卡。

如需关于如何调整时间戳和事件换行的更多信息，请参阅“修改事件处理”。

要手动调整时间戳和事件换行，请按照以下步骤操作：

1. 单击**事件换行**。该列表将显示**换行类型**选项，这些选项会控制 Splunk 平台将文件划分成事件的方式。您可以从以下选项中选择：
   - **自动：**根据时间戳的位置检测事件换行。
   - **按行：**把每行划分成单个事件。
   - **正则表达式...：**使用指定的正则表达式决定换行。
2. 单击**时间戳**。列表将展开以显示提取选项。请从以下任一选项中选择：
   - **自动：**通过在文件中查找时间戳事件来自动提取时间戳。
   - **当前时间：**把当前的时间应用到所有检测到的事件上。
   - **高级：**指定时区、strptime() 中的时间戳格式和包含该时间戳的任何字段。
3. 单击**分隔的设置**可显示分隔选项。

字段	描述
字段分隔符	用于结构化数据文件的分隔符，如逗号分隔值 (CSV) 文件。
引号字符	Splunk® Enterprise 用于决定某些内容什么时候在引号内的字符。
文件序言	指示 Splunk® Enterprise 忽略结构化数据文件中一个或多个序言行或不包含任何实际数据的行的正则表达式。
字段名称	基于行数、基于逗号分隔的列表或通过正则表达式自动确定字段名称。

4. 若结果符合您的要求，请将更改保存为新来源类型，您之后可将该来源类型应用于软件为其建立索引的数据。
5. 如果要对 props.conf 进行配置更改，请单击**高级**选项卡即可显示允许您输入属性/值对的字段，这些属性/值对将直接提交至 props.conf 配置文件。有关更多说明，请参阅“在高级选项卡中更改配置”。

   在“高级”选项卡中更改配置需要您了解 Splunk 软件功能的高级知识。此处所做的更改可能会对数据的索引建立造成负面影响。在配置这些选项之前，请考虑咨询 Splunk 专业服务人员。

6. 如果结果符合您的要求，请单击**下一步**以转到**输入设置**页面。

在高级选项卡中更改配置

在“高级”选项卡中更改配置需要您了解 Splunk 软件功能的高级知识。此处所做的更改可能会对数据的索引建立造成负面影响。在配置这些选项之前，请考虑咨询 Splunk 专业服务人员。

1. 单击一个字段以编辑 Splunk® Enterprise 根据您之前的选择生成的 props.conf 条目。
2. 单击某个属性/值字段对右侧的 **X** 可将其删除。
3. 单击**新设置**可新建属性/值字段对并为 props.conf 指定有效属性和值。
4. 单击**应用设置**即可提交对 props.conf 文件所作的更改。

下一步

将正确的来源类型分配给您的数据后，请参阅“修改输入设置”。

准备用于预览的数据

“设置来源类型”页面仅对单个文件有效，且仅可访问驻留在 Splunk 部署中或已经上载在 Splunk 部署中的文件。尽管“设置来源类型”页面不直接处理网络数据或文件目录，但您能够解决这些限制。使用 Splunk Cloud 中，您可以上载任何文件进行预览。

预览网络数据

可以指示一些示例网络数据进入文件中，然后将其作为文件监视输入上载或添加。一些外部工具可以实现此操作。在 *nix 上，最常用的工具是 Netcat。

例如，如果您要在 UDP 端口 514 上监视网络设备的网络流量，则可以使用 Netcat 指示一些网络数据进入文件中。运行如下命令：

```
nc -lu 514 > sample_network_data
```

为了获得最佳结果，请在逻辑为文件大小达到 2 MB 时终止 Netcat 进程的 shell 脚本内运行此命令。默认情况下，当您预览文件中的数据时，Splunk 软件仅读取前 2 MB 的数据。

当您创建 sample_network_data 文件后，可将其添加为输入、预览数据并将任何新的来源类型分配给该文件。

预览文件目录

如果目录中的所有文件内容均相似，则您可以预览单个文件，并且结果对目录中的所有文件绝对有效。然而，如果目录中文件的数据多种多样，请务必预览表示目录中一系列数据的一组文件。请单独预览每个文件类型，因为指定通配符会导致 Splunk Web 禁用此“设置来源类型”页面。）

文件大小限制

Splunk Web 将显示“设置来源类型”页面中一个文件的前 2 MB 数据。大多数情况下，这个数量应提供足够的数据取样。如果您使用的是 Splunk Enterprise，可以通过更改 limits.conf 文件中的 max_preview_bytes 属性取样大量数据。有关 limits.conf 文件的更多信息，请参阅 Splunk Enterprise *《管理员手册》*中的“limits.conf”。

或者，您可以编辑该文件以减少大量相似数据，这样剩余的 2 MB 数据便包含原始文件中所有类型数据的表示。

修改事件处理

您可更改事件处理设置并将改进的设置保存为新的来源类型。

1. 查看事件数据，如“将正确的来源类型分配给您的数据”中所述。
2. 修改事件处理设置。
3. 查看更改效果，直到满意为止。
4. 将修改的设置保存为新来源类型。
5. 请将新来源类型应用于您的任何输入。

修改事件处理设置

如需新建来源类型，请使用事件换行和时间戳设置，然后再保存该来源类型。

在“设置来源类型”页面上，您可以使用以下可折叠选项卡进行三种类型的调整：

- **事件换行。**调整 Splunk® Enterprise 将数据拆分成事件的方式。
- **时间戳。**调整 Splunk® Enterprise 确定事件时间戳的方式。
- **高级选项卡。**如果您使用的是 Splunk Enterprise，请编辑 props.conf。

修改事件换行

要修改事件换行参数，请单击**事件换行**。您可以选择以下换行类型：

- **自动。**根据数据内时间戳的位置拆分事件。
- **每行。**把每行视为单个事件。
- **正则表达式**...使用指定的正则表达式将数据拆分成事件。

有关换行的信息，请参阅“配置事件换行”。您可以在搜索中将正则表达式与 rex 搜索命令结合使用，对表达式进行测试。

修改时间戳

如需修改时间戳识别参数，请单击**时间戳**选项卡以将其展开。

您可以从以下提取选项中选择：

- **自动。**自动定位时间戳。
- **当前时间。**使用当前的系统时间。
- **高级。**指定其他高级参数以调整时间戳。

然后，您可以配置其他高级参数：

- **时区。**您想要用于事件的时区。
- **时间戳格式。**代表 Splunk® Enterprise 在数据中搜索时间戳时要使用的时间戳格式的字符串。请参阅“配置时间戳识别”。
- **时间戳前缀。**一个正则表达式，代表显示于时间戳之前的字符。
- **提前量。**Splunk® Enterprise 查找事件或您在“时间戳前缀”中为时间戳指定的正则表达式的字符数。

如果您在“时间戳格式”字段中指定时间戳格式且时间戳并未恰好位于每个事件的开头，则还必须在时间戳前缀字段中指定前缀。否则，Splunk 平台将无法处理格式化指令，而且每个事件都将包含一条无法使用 strptime 的警告。您仍然可以根据 Splunk 平台尝试从问题中恢复的方式而以有效时间戳结束。

有关配置时间戳的信息，请参阅“时间戳分配如何工作”。

进行高级修改

如需修改高级参数，请单击**高级**选项卡。该选项卡显示允许您通过编辑底层 props.conf 文件来指定来源类型属性的选项。

您可以通过指定设置/值对来添加或更改来源类型属性。有关如何设置这些属性的详细信息，请参阅*《管理员手册》*中的“props.conf 配置文件”。

“高级”选项卡显示已选择的来源类型的当前完整属性集：

- 单击**应用**后，由**事件换行**或**时间戳**选项卡中所做更改生成的设置。
- 在您首次预览文件时，自动检测或手动选择的来源类型的任何预先存在的设置。
- 单击**应用设置**后，从**其他设置**文本框应用的设置。

有关如何设置来源类型属性的信息，请参阅*《管理员手册》*中的“props.conf”。还可参阅“时间戳分配如何工作”和“配置事件换行”。

Splunk® Enterprise 如何组合设置

高级选项卡中的设置更改优先级最高。例如，若您使用**时间戳**选项卡改变时间戳设置，且还在**高级选项卡**中进行了互相冲突的时间戳更改，**高级选项卡**更改将优先于您在“时间戳”选项卡中所做的修改。

以下列表显示了 Splunk® Enterprise 组合任意调整与基本默认设置的方式，从优先级最高者开始：

1. 高级选项卡更改
2. 事件换行或时间戳更改
3. 基本来源类型的设置（如果有）
4. 所有来源类型的默认系统设置

如果您在“高级”选项卡中进行更改，然后返回到“事件换行”或“时间戳”选项卡，则这些更改在这些选项卡中将不可见。

查看更改

当您准备好查看更改的效果后，请单击**应用设置**。Splunk Web 将刷新屏幕，以便您可以检查更改对数据产生的影响。

要使用提供的三个调整方法中的任意一个做进一步更改，请单击**应用更改**以查看更改对数据产生的效果。

将修改保存为新来源类型

1. 单击“来源类型”按钮旁边的**另存为**。
2. 在出现的对话框中，为您的新源类型命名、选择它将出现的来源类型类别，以及它使用的应用程序上下文。

字段	描述
名称	新来源类型的名称。
描述	新来源类型的描述。
类别	单击“来源类型”按钮时，来源类型显示为的类别。
应用	新来源类型使用的应用。

3. 请单击**保存**以保存此来源类型并返回到“设置来源类型”页面。

后续步骤

在保存来源类型后，您有以下几个选项：

- （可选）请单击**下一步**以将来源类型应用于您的数据并继续到**输入设置**页面。
- （可选）请单击左尖括号（<）以返回并选择要上载或监视的新文件。
- （可选）单击**添加数据**以返回到“添加数据”向导的起始处。

修改输入设置

前提条件

- 将正确的来源类型应用于您的数据

在您选择数据来源或在上载或监视单个文件时设置您的来源类型后，Splunk® Enterprise 中将显示**修改输入设置**页面。

您可指定数据导入的其他参数，如它的来源类型、应用程序上下文、主机值和来自输入的数据将保存在其中的索引。

配置来源类型

您可通过“来源类型”设置指定要应用到您数据上的来源类型。该设置出现在以下情况中：

- 当您将目录指定为数据来源时。
- 当您将网络输入指定为数据来源时。
- 当您指定由另一个 Splunk 实例转发来的数据来源时。

如果您的数据来源不满足这些标准，则不会显示“来源类型”设置。

指定来源类型

要指定来源类型，请选择以下选项之一：

选项	描述
选择	单击此按钮可将您指定的来源类型应用于数据。
新	单击此按钮可添加新的来源类型。

选择一个现有来源类型

1. 从**选择来源类型**下拉列表中选择最能代表数据来源类型的类别。
2. 从显示的列表中选择来源类型。

添加新的来源类型

1. 请在**来源类型**文本字段中输入新来源类型的名称。
2. 请在**来源类型类别**下拉列表中为来源类型选择一个类别。
3. 请在**来源类型描述**文本字段中输入来源类型的描述。

配置应用上下文

“应用程序上下文”设置确定输入应在其中收集数据的上下文。设置应用程序上下文时，您要确定输入配置要存储到哪个 Splunk 应用中。Splunk 应用在 Splunk 平台上运行，通常用于解决用例。应用程序上下文改进了输入和来源类型定义的可管理性。应用程序上下文根据优先顺序规则加载。请参阅 Splunk Enterprise *《管理员手册》* 中的“配置文件优先顺序”。

单击下拉列表并从中选择您想要的应用程序上下文，然后选择您要此输入在其中运行的应用程序上下文。

配置主机值

Splunk® Enterprise 使用主机标记事件。您可以配置 Splunk 软件决定主机值的方式。通过选择以下可用主机值之一来配置主机值：

主机值	描述
IP	此值使用事件来源的主机的 IP 地址。
DNS	此值使用域名服务 (DNS)。使用 Splunk 软件用 DNS 域名映射决定的主机名称标记事件。
自定义	此值使用您在选择该选项时显示的“主机字段值”文本字段中分配的主机值。

在索引中存储事件

“索引”设置决定该输入的事件应存储于其中的索引。

1. 要使用默认索引，请保持下拉列表选项设置为 `Default`。或者，您也可以单击下拉列表，并选择您希望数据转到其中的索引。
2. （可选）如果您要将数据发送到其中的索引不在列表内，而且您有新建索引的权限，则您可通过单击**新建索引**按钮来新建索引。
3. 在您作出选择之后，单击**下一步**。

在 Splunk Enterprise 中分布来源类型配置

如果您使用 Splunk Web 在 Splunk Cloud 中新建来源类型， Splunk Cloud 将自动管理来源类型配置。但是，如果您使用的是 Splunk Enterprise 且管理的是分布式配置，则您必须根据本主题中介绍的方式分布新的来源类型。

您可使用“设置来源类型”或 Splunk Web 中的来源类型管理页面来新建来源类型，然后您可将这些类型分配给来自特定文件、目录的输入或网络输入。这两个页面都会把新来源类型保存到本地 Splunk Enterprise 实例上的一个 `props.conf` 配置文件。接下来，您可以把这个文件分布到其他 Splunk Enterprise 实例，方便这些实例识别前述新来源类型。

如某分布式环境中有**转发器**获取数据，然后再将数据发送到索引器上，您可以在该环境中使用新来源类型。

如需安装该新来源类型，请遵照下列高级步骤：

1. 将包含来源类型定义的 props.conf 文件发布到 $SPLUNK_HOME/etc/system/local 目标索引器上的目录内；您打算使用自己新建的来源类型，通过目标索引器为数据新建索引。
2. 在将数据发送到这些索引器的转发器上定义输入时，您可以使用该新来源类型。

当转发器将已使用新来源类型标记的数据发送至索引器时，索引器会正确地将其处理成事件。

数据预览 props.conf 文件

当您在“设置 Sourcetype”页面中新建来源类型时，Splunk 软件会把该来源类型定义作为 props.conf 中的一个段落保存在您保存来源类型时所选的应用中。如果您以后新建了其他来源类型，这些来源类型将保存到同一个 props.conf 文件内。

例如，如果您选择了“搜索和报表”应用，此文件将驻留在 $SPLUNK_HOME/etc/apps/search/local/props.conf 中。唯一例外是“系统”应用：如果您在保存来源类型时选择了该应用，此文件将驻留在 $SPLUNK_HOME/etc/system/local. 中。

*请注意：*Splunk Enterprise 实例可能拥有某些配置文件的多个版本，并分散在多个目录中。运行期间，Splunk Enterprise 会按照一组优先顺序规则把配置文件的内容合并在一起。有关配置文件如何运作的背景知识，请参阅“关于配置文件”和“配置文件优先顺序”。

将 props.conf 分布到其他索引器

新建来源类型后，您可以把 props.conf 分布到其他 Splunk Enterprise 实例。之后该实例可为您以新来源类型标记的任何传入数据新建索引。

Splunk 最佳方法是把配置文件置于目标 Splunk Enterprise 实例上该配置文件自己的应用目录内，例如：$SPLUNK_HOME/etc/apps/custom_sourcetype/local/。

如需把配置文件分布到其他 Splunk 实例，您可以使用一个**部署服务器**或其他分布工具。请参阅“更新 Splunk 实例”手册。

注意：Splunk 软件使用 props.conf 中的来源类型定义，把传入的数据分析成事件。为此，您只能将文件分布到执行**分析**的 Splunk Enterprise 实例（即索引器或**重型转发器**）。

在转发器输入中指定新来源类型

转发器（重型转发器除外）中没有 Splunk Web。这意味着，您必须通过 CLI 或 inputs.conf 配置文件来配置转发器输入。当您在该文件中指定输入时，也可指定其来源类型。有关 inputs.conf 的信息，请参阅“配置文件参考”中的 inputs.conf 章节。

1. 如需把某转发器输入标记为新来源类型，请把该来源类型添加到 inputs.conf 中的输入段落内。例如：

   ```
   [tcp://:9995]
   sourcetype = new_network_type
   ```
2. 请确认转发器向其发送数据的所有索引器都有 props.conf 文件的副本，该文件包含 new_network_type 的来源类型定义。当转发器将数据发送到索引器时，它们能够标识新来源类型并正确地设置数据格式。

获取文件和目录的数据

监视文件和目录

要监视 Splunk Cloud 中的文件和目录，几乎在所有情况下都必须使用通用或重型转发器。您在转发器上执行数据收集，然后将数据发送到 Splunk Cloud 实例。
转发器具有三个文件输入处理器：

- **监视器**
- MonitorNoHandle
- 上载

您必须为监视器和 MonitorNoHandle 输入处理器使用转发器，但不需要使用转发器来上载单个文件。您可以使用 Splunk Web 一次将一个文件上载到 Splunk Cloud。

如果使用的是 Splunk Enterprise，则可以直接在 Splunk Enterprise 实例上使用 CLI、Splunk Web 或 input.conf 配置文件来监视文件。您还可以使用通用或重型转发器，如同使用 Splunk Cloud 一样。

您可以使用监视器输入从文件和目录添加几乎所有数据源。但是，您可能希望使用上载输入只监视文件（例如历史数据归档）一次。

在运行 Windows Vista 或 Windows Server 2008 及更高版本的计算机上，您可以使用 MonitorNoHandle 输入来监视 Windows 自动轮换的文件。MonitorNoHandle 输入仅适用于 Windows 计算机。

您可以使用以下方法添加监视器或上载输入：

- 在重型转发器上：请参阅“使用 Splunk Web 监视文件和目录”。
- 在通用或重型转发器上：请参阅“使用 CLI 监视文件和目录”。
- 在通用或重型转发器上：请参阅“使用输入配置文件监视文件和目录”。

您可以使用 CLI 或 inputs.conf 文件添加 MonitorNoHandle 输入。

如果您在重型转发器上使用 Splunk Web 配置文件监视器输入，则可以使用**设置来源类型**页面查看 Splunk 平台如何为文件建立索引。有关详细信息，请参阅**设置来源类型**页面。

监视器处理器如何工作

指定文件或目录的路径时，监视器处理器会获取已写入该文件或目录的所有新数据。使用指定路径的方法，您可以监视实时应用程序日志，例如来自 Web 访问日志、Java 2 Platform Enterprise Edition (J2EE) 或 .NET 应用程序的日志。

转发器会随着新数据的出现监视文件或目录并为其建立索引。您还可以指定已安装目录或共享目录（包括网络文件系统），但前提是转发器可从该目录中读取数据。如果指定的目录包含子目录，只要这些目录可读，监视器进程会以递归的方式检查这些目录中是否有新文件。

您可使用允许列表和排除列表来包含或排除可读取的文件或目录。

如果您禁用或删除监视器输入，则转发器不会停止索引该输入参考的文件。它仅停止再次检查这些文件。要停止所有处理中的数据索引，您必须重新启动转发器。

转发器在重新启动过程中如何处理文件监视

当您重新启动转发器后，其会从重新启动前停止的位置开始继续处理文件。它首先会检查在监视器配置中指定的文件或目录。如果该文件或目录在启动时并不存在，转发器会从上次重新启动时间开始每隔 24 小时检查一次。监视器进程将继续扫描受监视目录的子目录。

监视器数据可能会发生重叠。只要段落名称不同，转发器就会将其视为独立段落，并且与最特定段落匹配的文件将依照段落的设置进行处理。

转发器如何监视归档文件

为了监视归档文件，转发器会在处理之前解压缩归档文件，例如 TAR 或 ZIP 文件。Splunk Enterprise 支持以下几种归档文件类型：

- TAR
- GZ
- BZ2
- TAR.GZ 和 TGZ

- TBZ 和 TBZ2
- ZIP
- Z

如果您向现有归档文件添加新数据，转发器将重新处理整个文件，而不仅限于新数据。这可能会导致产生重复的事件。

转发器如何监视操作系统按计划轮换的文件

监视处理器会检测日志文件轮换，不会处理已重新处理的重命名文件（.tar 和 .gz 归档除外）。请参阅“Splunk 平台如何处理日志文件轮换”。

转发器如何监视不可写的 Windows 文件

Windows 会阻止转发器读取已打开的文件。如果您需要在向文件执行写入期间读取文件，可以使用 monitorNoHandle 输入。

文件监视的限制

转发器无法监视路径超过 1024 个字符（Windows 上为 256 个字符）的文件。

转发器也不会监视具有 .splunk 文件名扩展名的文件，因为具有该扩展名的文件包含 Splunk 元数据。如您需要为具有 .splunk 扩展名的文件建立索引，请使用 add oneshot CLI 命令。

何时使用上载或批处理？

如需为静态文件建立一次性索引，请在 Splunk Cloud 或 Splunk Enterprise 上的 Splunk Web 中选择**上载**。

否则，在转发器上使用 CLI 命令 add oneshot 或 spool 为静态文件建立索引。详细信息请参阅“使用 CLI”。

您可以使用 inputs.conf 文件中的批处理输入类型以一次性加载文件并造成破坏。默认情况下，Splunk 批处理器位于转发器上的 $SPLUNKFORWARDER_HOME/var/spool/splunk 目录中。如果您将某个文件移入此目录，转发器会处理并删除该文件。

何时使用 MonitorNoHandle

在 Windows 向文件执行写入时，该仅适用于 Windows 的输入允许您在 Windows 系统上同时读取文件。您必须使用通用或重型转发器才能使用 Splunk Cloud 的输入。当数据写入文件时，该输入使用内核模式筛选器驱动程序来捕获原始数据。您可以在系统已将其锁定而无法打开进行写入的文件上使用此输入，例如 Windows DNS 服务器日志文件。

使用 MonitorNoHandle 的限制

MonitorNoHandle 输入具有以下限制：

- MonitorNoHandle 仅适用于 Windows Vista 或 Windows Server 2008 及更高版本的操作系统。不适用于早期的 Windows 版本，也不适用于非 Windows 操作系统。
- 使用 MonitorNoHandle 只能监视单个文件。要监视多个文件，您必须为每个文件创建 MonitorNoHandle 输入段落。
- 您无法使用 MonitorNoHandle 监视目录。
- 如果您选择要使用 MonitorNoHandle 监视的文件已经存在，转发器不会为该文件当前的内容建立索引，只会为文件执行写入时传入该文件的新信息建立索引。
- 如果您使用 MonitorNoHandle 监视文件，文件的来源字段是 MonitorNoHandle，而不是文件名。如果您想让来源字段成为文件名，您必须在 inputs.conf 中明确设置该字段。请参阅“使用 inputs.conf 监视文件和目录”。

使用 Splunk Web 监视 Splunk Enterprise 中的文件和目录

您可使用 Splunk Web 从文件和目录添加输入。

转发文件需要其他设置。请参阅以下主题：

- 如果使用是通用转发器，请参阅 Splunk 通用转发器*《转发器手册》*中的“配置通用转发器”。
- 如果使用是重型转发器，请参阅*《转发数据》*手册中的“在 Splunk Enterprise 实例上启用转发”。

转到“新增”页面

您可从 Splunk Web 中的**新增**页面添加输入。

可通过以下两种方式之一访问此页面。

Splunk 设置

1. 单击**设置** > **数据导入**。
2. 单击**文件和目录**。
3. 单击**新建**以添加输入。

Splunk 主页

1. 单击 Splunk 主页中的**添加数据**。
2. 单击**上载**以上载文件、单击**监视**以监视文件或单击**转发**以转发文件。

选择输入来源

1. 如需添加文件或目录输入，请单击Splunk Web 中的**文件和目录**。
2. 在**文件或目录**字段中，输入文件或目录的完整路径。
   要监视已安装在系统上的网络驱动器，请输入 `<myhost>/<mypath>`（适用于 *nix）或 `\\<myhost>\<mypath>`（适用于 Windows）。请确认 Splunk Enterprise 对已安装的驱动器以及您想要监视的文件具有读取访问权限。
3. 选择您要 Splunk Enterprise 监视的文件：
   - 选择**持续监视**以设置持续输入。Splunk Enterprise 为新数据持续监视文件。
   - 选择**索引一次**将服务器上的文件复制到 Splunk Enterprise。
4. 单击**下一步**。
   如果您在**文件或目录**字段中指定目录，则 Splunk Enterprise 会刷新屏幕以显示**允许列表**和**拒绝列表**字段。这些字段允许您输入 Splunk Enterprise 之后用于对包含或排除匹配文件的正则表达式。否则，Splunk Enterprise 将继续前往**设置来源类型**页面，您可在该页面预览 Splunk Enterprise 如何建议对事件建立索引。

有关如何包含和排除数据的更多信息，请参阅“包含或排除特定的传入数据”。

预览您的数据并设置其来源类型

当您添加新文件输入时，Splunk Enterprise 将允许您设置数据的**来源类型**并预览数据在索引后的外观。这将允许您检查数据是否已正确设置格式并作出必要的调整。

有关**设置来源类型**页面的信息，请参阅“将正确的来源类型应用于您的数据”。

如果您跳过数据预览，则会显示**输入设置**页面。

您无法预览目录或归档文件。您也无法使用 `Log to Metrics` 来源类型预览输入。

指定输入设置

您可在**输入设置**页面提供应用程序上下文、默认主机值和索引。所有参数均为可选项。

1. 为此输入选择相应的**应用程序上下文**。
2. 设置**主机**值。

   主机值只是设置生成事件中的**主机**字段。设置此值不会引导 Splunk Enterprise 查找网络中的特定主机。

3. 为此输入设置您希望 Splunk Enterprise 将数据发送到其中的**索引**。如果未定义多个索引且不想使用其中一个索引，请保留“默认”值。
4. 请单击**查看**以查看您已做的所有选择。

查看您的选择

在您提供所有输入设置后，请查看您的选择。Splunk Web 会列出您选定的选项，包括监视器的类型、数据来源、来源类型、应用程序上下文和索引。

1. 查看该设置。
2. 如果它们不符合您的期望，请单击左尖括号（ < ）即可返回到向导中的上一个步骤。否则，请单击**提交**。**成功**页面显示且 Splunk 平台开始索引指定的文件或目录。

使用 CLI 监视 Splunk Enterprise 文件和目录

在 Splunk Enterprise 安装中，您可以使用命令行界面（CLI）监视文件和目录。如需使用 CLI，请从命令提示符或 shell 导航至 `$SPLUNK_HOME/bin/` 目录，并使用目录内的 `splunk` 命令。

CLI 具有内置帮助。键入 `splunk help` 即可访问 CLI 主要帮助。各个命令也各自有帮助页面。键入 `splunk help <command>` 即可访问该帮助。

用于输入配置的 CLI 命令

使用 CLI，以下命令可用于输入配置：

命令	命令语法	操作
add monitor	add monitor [-source] <source> [-parameter value] ...	监视来自 <source> 的输入。
edit monitor	edit monitor [-source] <source> [-parameter value] ...	编辑之前为 <source> 添加的监视器输入。
remove monitor	remove monitor [-source] <source>	移除之前为 <source> 添加的监视器输入。
list monitor	list monitor	列出当前配置的监视器输入。
add oneshot	add oneshot <source> [-parameter value] ...	将来源文件直接复制到 Splunk Enterprise。这将上载文件一次，但 Splunk Enterprise 不会持续监视文件。 您无法使用 oneshot 命令监视远程 Splunk Enterprise 实例上的文件。您无法同时使用命令与递归文件夹或通配符作为数据来源。请提供您所要监视文件的准确来源路径。
spool	spool <source>	使用 sinkhole 目录将来源文件直接复制到 Splunk Enterprise。类似于 add oneshot 命令，不同之处是文件来自于 sinkhole 目录，而不是立即添加。 您无法使用 spool 命令监视远程 Splunk Enterprise 实例上的文件。您无法同时使用命令与递归文件夹或通配符作为数据来源。请提供您所要监视文件的准确来源路径。

用于输入配置的 CLI 参数

可通过设置其他参数更改每种数据导入类型的配置。要设置参数，请使用语法 -parameter value。

每个命令只能设置一个 -hostname、-hostregex 或 -hostsegmentnum。

参数	是否必需？	描述
<source>	是	提供新输入要监视和上载的文件或目录的路径。 此参数可以是值本身。并不需要遵照参数标记。您可以使用 ./splunk monitor <source>，也可以使用 ./splunk monitor -source <source>。
sourcetype	否	为来自输入来源的事件提供 sourcetype 字段值。
index	否	为来自输入来源的事件提供目标索引。
hostname 或 host	否	为来自输入来源的事件提供要设置为主机字段值的主机名。 这些参数在功能上是等效的。
hostregex 或 host_regex	否	提供用于从来源键提取主机字段值的正则表达式。 这些参数在功能上是等效的。
hostsegmentnum 或 host_segment	否	一个整数，用于确定要设置为主机字段值的路径段（以 "/" 分隔）。例如，如果设置为 3，则使用路径的第三个段。 这些参数在功能上是等效的。

rename-source	否	提供要应用于该文件中数据的 source 字段值。
follow-only	否	设置为 "true" 或 "false"。默认为 false。 当设置为“true”时，Splunk Enterprise 会从数据来源末尾开始读取，类似于 tail -f Unix 命令。 此参数不可用于 add oneshot 命令。

示例 1：监视目录中的文件

以下示例显示如何监视 /var/log/ 中的文件。

将 /var/log/ 添加为数据导入：

```
./splunk add monitor /var/log/
```

示例 2：监视 windowsupdate.log

以下示例显示如何监视 Windows Update 日志文件（这是 Windows 记录自动更新的位置）并将数据发送到一个称为 newindex 的索引。

将 C:\Windows\windowsupdate.log 添加为数据导入：

```
splunk add monitor c:\Windows\windowsupdate.log -index newindex
```

示例 3：监视 Internet Information Server (IIS) 日志

本例显示如何监视 Windows IIS 日志的默认位置。

将 C:\windows\system32\LogFiles\W3SVC 添加为数据导入：

```
./splunk add monitor c:\windows\system32\LogFiles\W3SVC
```

示例 4：上载文件

本例显示如何将文件上载到 Splunk Enterprise。Splunk Enterprise 仅使用一次文件。它不会持续监视文件。

直接将 /var/log/applog（在 Unix 上）或 C:\Program Files\AppLog\log.txt（在 Windows 上）上载到 Splunk Enterprise 上时请使用 add oneshot 命令：

Unix	Windows
./splunk add oneshot /var/log/applog	.\splunk add oneshot C:\Program Files\AppLog\log.txt

您也可以使用 spool 命令通过 sinkhole 目录上载文件：

Unix	Windows
./splunk spool /var/log/applog	.\splunk spool C:\Program Files\AppLog\log.txt

两个命令的结果都是相同的。

使用 inputs.conf 监视文件和目录

您可以使用 input.conf 文件通过 Splunk 平台监视文件和目录。input.conf 文件可提供充足的配置选项以设置文件监视输入。如果您使用 Splunk Cloud，则可以使用 Splunk Web 或转发器来配置文件监视输入。

要配置输入，请向 $SPLUNK_HOME/etc/system/local/ 目录中的 input.conf 文件或 $SPLUNK_HOME/etc/apps/ 中您自己的自定义应用程序目录中添加一个**段落**。这些位置位于运行 Splunk Enterprise 或转发器的计算机上。要了解有关 inputs.conf 文件的更多信息，请参阅 Splunk Enterprise*《管理员手册》*中的“inputs.conf”。

可在输入段落中配置多个设置。如果您未指定某个设置的值，Splunk 平台会使用默认设置值。您可以在

$SPLUNK_HOME/etc/system/default/inputs.conf 目录中找到设置的默认值。

有关配置文件的更多信息，请参阅 Splunk Enterprise*管理员手册*中的“关于配置文件”。

配置转发器以将数据发送到 Splunk Cloud

如果要将 Active Directory (AD) 数据发送到 Splunk Cloud，必须在开始编辑转发器上的配置文件之前安装和配置转发器。

1. 将通用转发器安装在您要从中收集 AD 数据的计算机上。
2. 将 Splunk Cloud 通用转发器凭据包安装到计算机中。

使用 inputs.conf 配置文件监视

1. 在运行 Splunk 软件的计算机上，打开 shell 或命令提示符。
2. 将列出的目录更改为 $SPLUNK_HOME/etc/system/local 目录。
3. 如果 inputs.conf 文件不存在，请创建该文件。
4. 打开 inputs.conf 以使用文本编辑器进行编辑。
5. 添加一个引用要监视的文件或目录的段落。例如，要监视 *nix 系统上的 /var/log/messages 文件，请使用以下规范：

   ```
   [monitor:///var/log/messages]
   disabled = 0
   ```

   要监视 Windows 系统上的 C:\Windows\System32\WindowsUpdate.log 文件，请使用以下规范：

   ```
   [monitor://C:\Windows\System32\WindowsUpdate.log]
   disabled = 0
   ```
6. （可选）根据您想要输入执行的操作，添加进一步配置输入的设置。请参阅本主题后续的“配置设置”，或参阅 Splunk Enterprise*《管理员手册》*中的 input.conf 以了解其他设置。

   ```
   [monitor://path/to/file]
   disabled = 0
   setting1 = value
   setting2 = value
   ...
   ```
7. 保存 inputs.conf 文件并将其关闭。
8. 运行以下命令重新启动 Splunk 平台或重新加载配置。如果重新加载配置，Splunk 平台会提示您输入凭据。

   ```
   ./splunk _internal call /services/data/inputs/monitor/_reload -auth
   ```

配置设置

您可以在 monitor 和 batch 输入段落中使用以下设置。

设置	描述	默认
host = <string>	将此段落的主机键设为一个静态初始值。输入处理器在分析和创建索引过程中使用该键设置主机字段，并在搜索过程中使用该字段。Splunk 平台会在 <string> 前面加上 host::。	生成数据的主机的 IP 地址或完全限定域名。
index = <string>	设置用于存储来自此输入的事件的索引。Splunk 平台会在 <string> 前面加上 index::。 有关索引字段的更多信息，请参阅 Splunk Enterprise*《管理索引器和群集》*手册中的“索引如何工作”。	main 索引或设置默认索引的任何值。
sourcetype = <string>	设置来自此输入的事件的来源类型键或字段。此设置显式声明该数据的来源类型，而不是允许 Splunk 平台自动确定。声明来源类型对于可搜索性以及在分析和新建索引期间对此类型数据应用相关格式设置都很重要。 设置 sourcetype 键的初始值。Splunk 平台在分析和创建索引过程中使用该键设置来源类型字段，并在搜索过程中使用该来源类型字段。Splunk 平台会在 <string> 前面加上 sourcetype::。	Splunk 平台会根据数据的各个方面选取一种来源类型。没有默认来源类型。

	有关来源类型的更多信息，请参阅“来源类型为何重要”。	
queue = parsingQueue | indexQueue	指定输入处理器用来存储其所读取事件的位置。设置为 parsingQueue 时，会将 props.conf 文件和其他分析规则应用到您的数据。设置为 indexQueue 时，会将您的数据直接发送到索引。	parsingQueue
_TCP_ROUTING = <tcpout_group_name>,<tcpout_group_name>,...	指定以逗号分隔的 tcpout 组名称列表。您可以通过指定转发器在转发数据时应使用的 tcpout 组，使用该设置以选择性地将您的数据转发到特定索引器。 请在 [tcpout:<tcpout_group_name>] 段落的 outputs.conf 文件中定义 tcpout 组名称。	这些组存在于 outputs.conf 文件中 [tcpout] 段落的 defaultGroup 内。
host_regex = <regular expression>	从每个输入的文件名提取主机的正则表达式。具体来说，Splunk 平台使用正则表达式的第一组作为主机。	如果正则表达式匹配失败，则使用默认的 "host =" 设置。
host_segment = <integer>	把路径的段设为主机，使用 <integer> 决定段。例如，如果 host_segment = 2，则 host 将成为路径的第二段。路径段用正斜杠 (/) 字符进行分隔。	如果数值不是整数或者小于 1，则为默认的 "host =" 设置。

监视语法

监视输入段落可配置 Splunk 平台监视 <path> 中的所有文件（或者，如果 <path> 代表单个文件，则只监视它本身）。您必须依次指定路径前面的输入类型，以便如果路径包含 *nix 计算机中的根目录，就将三个正斜杠添加到路径中。

可在路径中使用通配符。请参阅“使用通配符指定输入路径”。

```
[monitor://<path>]
<setting1> = <val1>
<setting2> = <val2>
...
```

以下是可在定义监视器输入段落时使用的其他属性：

设置	描述	默认
source = <string>	设置来自此输入的事件的来源字段。使用 MonitorNoHandle 输入并想要将来源设置为您正在监视的文件名称时，您可以使用此设置。否则，除非绝对必要，请勿覆盖。请考虑使用来源类型、标记和搜索通配符。输入层提供更准确的字符串来帮助分析和调查问题，并准确记录从中检索数据的文件。 Splunk 平台会在 <string> 前面加上 source::。	输入文件路径（MonitorNoHandle 除外，默认为 MonitorNoHandle）。
crcSalt = <string>	强制 Splunk 平台为具有匹配的循环冗余码校验 (CRC) 的文件建立索引。默认情况下，该软件只对文件的前几行执行 CRC。此行为会阻止对同一文件索引两次，尽管您可能已重新命名了该文件，例如，滚动日志文件。但由于 CRC 只对文件的前几行进行计数，因此不同文件可以具有相同的匹配 CRC。） 如果设置，Splunk 平台会将 string 添加到 CRC。如果设置为 >SOURCE<，Splunk 平台会将完整的来源路径添加到 CRC。添加 >SOURCE< 可确保所监视的每个文件都具有唯一 CRC。 请谨慎使用滚动日志文件的设置。此设置可能会导致日志文件在滚动之后得以重新创建索引。 此设置区分大小写。	N/A
ignoreOlderThan = <time_window>	如果文件的修改时间已超过 <time_window> 阈值，会导致输入停止检查文件的更新。当您监视包含大量历史文件的目录层次结构时，停止文件检查可提高文件跟踪操作的速度。例如，当活动日志文件与不再向其中执行写入操作的旧文件共享目录时。 如果 Splunk 平台首次尝试监视文件时，该文件的修改时间落在 <time_window> 之外，Splunk 平台将不会为该些文件创建索引。	0 （禁用）

	您必须指定 <number><unit>。例如，7d 表示一星期。有效的单位包括 d（天）、h（小时）、m（分钟）和 s（秒）。	
followTail = 0|1	如果设置为 1，监视将从文件末尾（比如 *nix tail -f）开始。此设置仅适用于 Splunk 平台首次尝试监视的文件。此后，Splunk 平台会使用内部文件位置记录跟踪该文件。	0
whitelist = <regular expression>	如果设置，则 Splunk 平台仅监视文件名与指定正则表达式匹配的文件。	N/A
blacklist = <regular expression>	如果设置，则 Splunk 平台不会监视文件名与指定正则表达式匹配的文件。	N/A
alwaysOpenFile = 0 | 1	如果设置为 1，则 Splunk 平台会打开一个文件，以检查其是否已经创建索引。此设置仅对不更新修改时间的文件有用。 此设置用于监控 Windows 上的文件，并且针对 Internet Information Server (IIS) 日志。 请谨慎使用此设置，因为它会增加负载并减慢索引创建速度。	N/A
recursive = true|false	如果设置为 false，Splunk 平台将不会查看其在所监视的目录中发现的子目录。	true
time_before_close = <integer>	在 Splunk 平台可关闭 End-of-file (EOF) 上的文件之前所需的修改时间增量。此设置指示系统不要关闭在过去 <integer> 几秒内更新完毕的文件。	3
followSymlink = true|false	如果设置为 false，Splunk 平台将忽略其在所监视目录中发现的符号链接。	true

MonitorNoHandle 语法

MonitorNoHandle 输入监视文件无需使用 Windows 文件句柄。此输入允许 Splunk 软件读取特殊的 Windows 日志文件，例如 DNS 调试服务器日志。使用此输入时有一些限制：

- MonitorNoHandle 输入段落仅适用于 Windows 系统。
- MonitorNoHandle 输入段落只会监视单个文件。
- 在文件或目录路径中不能使用通配符。
- 无法使用 MonitorNoHandle 段落监视目录。
- MonitorNoHandle 输入段落只会读取写入受监视文件的新数据。它不会引入已经写入该文件的数据。
- 默认情况下，使用 MonitorNoHandle 监视的文件的源元数据设置为 MonitorNoHandle。要指定另一个源，必须使用 inputs.conf 段落中 inputs.conf 文件内的 source 设置进行定义。

有关 MonitorNoHandle 段落的示例，请参阅 MonitorNoHandle，单个 Windows 文件。

批处理语法

使用批处理可为来自某一数据来源的数据设置具有破坏性的一次性输入。

要设置非破坏性的连续输入，请使用 monitor 输入。Splunk 平台会删除其用 batch 输入建立索引的数据。

```
[batch://<path>]
move_policy = sinkhole
<setting1> = <val1>
<setting2> = <val2>
...
```

定义批处理输入时，您必须将 move_policy = sinkhole 设置包含在内。该设置在加载文件时会造成破坏。对于您不想在建立索引后删除的文件，请勿使用批处理输入类型。

要确保 Splunk 平台在您复制某包含新内容的现有文件时为新事件建立索引，请在 props.conf 文件中为该输入来源设置 CHECK_METHOD = modtime 设置。此设置会在时间发生更改时检查文件的修改时间并重新建立索引。Splunk 平台将为整个文件建立索引，这会导致出现重复的事件。有关 props.conf 文件的信息，请参阅 props.conf。

监视输入段落的示例

*单个 *nix 文件*

此示例段落配置 Splunk 平台为单个 /var/log/messages 文件建立索引：

```
[monitor:///var/log/messages]
disabled = 0
sourcetype = unixlog
```

单个 Windows 目录

此 Windows 示例配置 Splunk 平台监视 C:\Windows\Logs 目录以及该目录中的所有文件：

```
[monitor://C:\Windows\Logs]
disabled = 0
```

单个 Windows 目录，文件名中带空格

此 Windows 示例配置 Splunk 平台监视 C:\Program Files\VMWare 目录以及该目录中的所有文件：

```
[monitor://C:\Program Files\VMWare]
disabled = 0
```

多个 Windows 目录

此 Windows 示例告诉 Splunk 平台监视 C:\Windows\Debug 中的所有目录：

```
[monitor://C:\Windows\Debug\*]
disabled = 0
```

*带通配符的多个 *nix 目录*

此示例配置 Splunk 平台监视 /apache/foo/log and /apache/bar/log 等目录：

```
[monitor:///apache/.../log]
```

*一个目录中的多个带通配符的 *nix 目录*

此 *nix 示例配置 Splunk 平台监视一个目录（如 /apache/*.log）中的多个文件：

```
[monitor:///apache/*.log]
```

MonitorNoHandle，单个 Windows 文件

这是 Splunkbase 中 Splunk Add-on for Microsoft Windows 的单个 Windows 文件示例：

```
###### Monitor Inputs for DNS ######
[MonitorNoHandle://$WINDIR\System32\Dns\dns.log]
sourcetype=MSAD:NT6:DNS
disabled=0
```

批处理

此批处理示例加载和删除 system/flight815/ 目录中的所有文件：

```
[batch://system/flight815/*]
move_policy = sinkhole
```

使用通配符指定输入路径

您可以通过编辑 inputs.conf 配置文件手动配置输入。在 Splunk Cloud 中，您可以在收集数据的转发器上编辑此文件。在 Splunk Enterprise 中，您可以在 Splunk Enterprise 实例上编辑此文件。

input.conf 文件中的输入路径规范不使用正则表达式（regexes），而是使用特定于 Splunk 平台的通配符。要指定通配符，必须在 inputs.conf 文件中指定文件和目录监视输入。

当您配置具有通配符的输入路径时，Splunk 平台实例必须对要使用通配符监控的文件的整个路径至少具有读取访问权限。例如，如果要监控路径为 /var/log/server_a/tree_b/directory_c/file.log 的文件，则实例必须具有以下目录的读取权限：

- var
- 对数
- server_a
- tree_b
- directory_c

如果它对路径中所有目录没有读取访问权限，则即使它直接对文件有读取权限，也无法读取文件。

通配符概述

通配符是在搜索文本或者选择多个文件或目录时可替换为一个或多个未指定字符的字符。您可以使用通配符为文件或目录监视输入指定输入路径。有关您可以使用的通配符和示例的说明，请参阅下表：

通配符	描述	等效正则表达式	示例
...	省略号通配符将递归搜索目录以及任何多级别的子目录，以找到匹配项。 如果您指定了文件夹分隔符（例如 //var/log/.../file），该分隔符不会匹配第一个文件夹级别，只会匹配子文件夹。	.*	/foo/.../bar.log 和 /foo/1/bar.log、/foo/2/bar.log、/foo/1/2/bar.log 等匹配。但不匹配 /foo/bar.log 或 /foo/3/notbar.log。 由于单个省略号递归搜索所有文件夹和子文件夹，/foo/.../bar.log 匹配 /foo/.../.../bar.log。
*	星号通配符匹配该特定文件夹路径段中的任意内容。 不同于 ...，"*" 不会递归遍历子文件夹。	[^/]*	/foo/*/bar 匹配以下内容： • /foo/1/bar • /foo/2/bar 但不匹配 • /foo/bar • /foo/1/2/bar /foo/m*r/bar 和 /foo/mr/bar、/foo/mir/bar、/foo/moor/bar 等匹配。 /foo/*.log 和所有扩展名为 .log 的文件匹配，例如 /foo/bar.log。但不匹配 /foo/bar.txt 或 /foo/bar/test.log。 单个句点（.）不是通配符，而是正则表达式，相当于 \.。

如想获取更多具体的匹配项，请把通配符 ... 和 * 结合在一起使用。例如，/foo/.../bar/* 将匹配指定路径下 /bar 目录中的任意文件。

通配符与正则表达式元字符

当确定要监视的一组文件或目录时，Splunk 平台会将监视段落元素拆分成段。段是段落定义中介于目录分隔符字符（/ 或 \）之间的文本块。如果您配置的监视段落中包含同时存在通配符和正则表达式元字符（如 (,), [,] 和 |）的段，这些字符的行为将根据通配符在段落中所处的位置而有所不同。

如果监视段落所包含的具有正则表达式元字符的段位于具有通配符的段之前，元字符将按字面处理，就好像您希望监视文件名或目录名称中包含这些字符的文件或目录。以下示例监视 /var/log/log(a|b).log 文件：

```
[monitor:///var/log/log(a|b).log]
The (a|b) is not treated as a regular expression because no wildcards are present.
```

以下示例监视 /var/log()/ 目录中所有以 log 开头且扩展名为 .log 的文件：

```
[monitor:///var/log()/log*.log]
```

由于所处的段位于通配符之前，() 不会被视为正则表达式。

如果正则表达式元字符在包含通配符的段之中或其之后发生，则 Splunk® Enterprise 会将元字符作为正则表达式处理，并相应地匹配要监视的文件。考虑以下示例：

```
[monitor:///var/log()/log(a|b)*.log]
```

此示例监视 /var/log()/ 目录中所有以 loga 或 logb 开头且扩展名为 .log 的文件。因为下一段中包含通配符，第一组 () 不会被视为正则表达式。第二组 () 则被视为正则表达式，因为它与通配符 * 位于相同的段中。

以下示例监视 /var/ 目录下任意子目录中所有命名为 loga.log 或 logb.log 的文件：

```
[monitor:///var/.../log(a|b).log]
```

Splunk® Enterprise 会将 (a|b) 视为正则表达式，因为前一个段落段中包含通配符 ...。

考虑以下示例：

```
[monitor:///var/.../log[A-Z0-9]*.log]
```

此示例监视 /var/ 目录下任意子目录中所有满足以下条件的文件：

1. 以 log 开头。
2. 包含一个大写字母 (A-Z) 或数字 (0-9)。
3. 包含任何其他字符。
4. 以 .log 结尾。

表达式 [A-Z0-9]* 将被视为正则表达式，因为前一个段落段中包含通配符 ...。

输入示例

要监视 /apache/foo/logs、/apache/bar/logs 和 /apache/bar/1/logs，请创建以下段落：

```
[monitor:///apache/.../logs/*]
```

要监视 /apache/foo/logs、/apache/bar/logs 但不监视 /apache/bar/1/logs 或 /apache/bar/2/logs，请创建以下段落：

```
[monitor:///apache/*/logs]
```

要监视直接位于 /apache/ 下且以 .log 结尾的所有文件，请创建以下段落：

```
[monitor:///apache/*.log]
```

要以递归方式监视 D:\Program Files\Splunk\etc\apps 中的所有日志文件，请创建以下段落：

```
[monitor://D:\Program Files\Splunk\etc\apps\*\...\*.log]
```

监视任意层级子目录下 /apache/ 之下所有以 .log 结尾的文件，请创建以下段落：

```
[monitor:///apache/.../*.log]
```

... 后跟文件夹分隔符将暗示通配符级别文件夹会被排除在外。

```
[monitor:///var/log/.../*.log]
```

跟踪逻辑将成为 ^\/var\/log/.*/[^/]*\.log$

因此，/var/log/subfolder/test.log 将与之匹配，而不符合匹配条件的 /var/log/test.log 将被排除在外。要监视所有文件夹中的所有文件，请进行以下更改：

```
[monitor:///var/log/]

whitelist=\.log$

recursive=true

#true by default
```

通配符和允许列表

Splunk Enterprise 使用标准 Perl 兼容正则表达式（PCRE）语法定义允许列表和拒绝列表。Splunk Cloud 不会以这种方式本机定义允许列表和拒绝列表。

当您在文件输入路径中配置通配符时，Splunk Enterprise 会为该段落创建一个隐式允许列表。无通配符的最长路径将成为监视器段落，而 Splunk Enterprise 将通配符转换为正则表达式。

Splunk Enterprise 将转换后的表达式定位在文件路径的右侧，以便整个路径必须匹配。

例如，在 *nix 中，如果您在 input.conf 配置文件中指定 `[monitor:///foo/bar*.log]` 段落，Splunk Enterprise 会将路径转换为：

```
[monitor:///foo/]
whitelist = bar[^/]*\.log$
```

在 Windows 中，如果您在 input.conf 文件中指定 `[monitor://C:\Windows\foo\bar*.log]` 段落，Splunk Enterprise 会将路径转换为：

```
[monitor://C:\Windows\foo\]
whitelist = bar[^/]*\.log$
```

在 Windows 中，允许列表和拒绝列表规则不支持包含反斜线的正则表达式。使用两个反斜线（\\）即可转义通配符。

包含或排除特定的传入数据

您可以使用“允许列表”和“拒绝列表”规则来确定要在**监视**一个目录或目录集时 Splunk 平台要使用或排除的文件。您也可以把这些设置应用于 `batch` 类型监视输入。当您定义允许列表时，Splunk® Enterprise 仅为您指定的文件建立索引。当您定义拒绝列表时，Splunk 平台会忽略指定的文件并处理其他所有文件。您可以在 inputs.conf 配置文件的 `input` 段落中定义这些过滤器。如果要在 Splunk 平台部署中的所有转发器上应用相同的过滤器，则可以通过设置让部署服务器执行此任务。请参阅*《更新 Splunk Enterprise 实例》*中的“关于部署服务器和转发器管理”。

不需要在配置段落中同时定义允许列表和拒绝列表。这些设置相互独立，但如果文件同时在两个列表中，则拒绝列表过滤器会覆盖允许列表过滤器。如果你同时定义了这两个过滤器，且某文件与二者都匹配，Splunk® Enterprise 不会为该文件创建索引。

列表规则使用正则表达式语法来定义文件名或路径的匹配项。规则必须包含于配置段落内，例如 `[monitor://<path>]`。Splunk 平台会忽略段落中不包括的过滤器列表。当您定义过滤器条目时，您必须使用准确的正则表达式语法。

路由和过滤数据

您可以不将数据输入列入允许列表或拒绝列表，而是过滤特定事件并将其发送到不同的队列或索引。

包含文件

将以下行添加到您的 `monitor` 段落（位于 local/inputs.conf 文件中，该文件为您在其中定义输入的应用上下文）。

```
whitelist = <your_custom regex>
```

例如，要仅监视扩展名为 .log 的文件，请进行以下更改：

```
[monitor:///mnt/logs]
    whitelist = \.log$
```

把多个文件列入允许列表

可以使用 "|"（管道 或 OR）运算符将多个文件列入允许列表的一行中。例如，要包含文件名中带 query.log OR my.log 的文件，请将以下行添加到 input.conf 文件中：

```
whitelist = query\.log$|my\.log$
```

或者，也可以只包含完全匹配的文件。请参阅以下示例：

```
whitelist = /query\.log$|/my\.log$
```

美元符号（ $ ）会将正则表达式定位在行尾处。竖线（ | ）运算符前后没有空格。

排除文件

把下行添加到您的 monitor 段落（位于 /local/inputs.conf 配置文件中，您在该文件中定义输入的应用上下文）。

```
blacklist = <your_custom_regex>
```

如果您为每个想要忽略的文件都创建一个 blacklist 条目，Splunk® Enterprise 只会激活最后一个过滤器。

示例 1：仅排除扩展名为 .txt 的文件

要忽略且不监视含 txt 扩展名的文件，请将以下行添加到 input.conf 文件中：

```
[monitor:///mnt/logs]
    blacklist = \.txt$
```

示例 2：排除扩展名为 .txt 或 .gz 的文件

要忽略且不监视含 txt 扩展名或 .gz 扩展名的所有文件，请将以下行添加到 input.conf 文件中：

```
[monitor:///mnt/logs]
    blacklist = \.(?:txt|gz)$
```

示例 3：排除一整个目录

要忽略监视器输入下的整个目录，请将以下行添加到 input.conf 文件中：

```
[monitor:///mnt/logs]
    blacklist = archive|historical|\.bak$
```

此示例将 Splunk® Enterprise 配置为忽略“归档”或“历史”目录中 /mnt/logs/ 下的所有文件以及扩展名为 *.bak 的所有文件。

示例 4：排除文件名中包含字符串的文件

要忽略名称包含特定字符串的文件，请将以下行添加到 input.conf 文件中：

```
[monitor:///mnt/logs]
   blacklist = 2009022[89]file\.txt$
```

此例忽略 /mnt/logs/ 下的 webserver20090228file.txt 和 webserver20090229file.txt 文件。

示例 5：排除“消息”字段包含特定值的 Windows 事件代码 #4662 事件

要忽略 Message 字段包含值为 Account Name: "example account" 的事件的 Windows 事件代码 #4662 事件，请将以下行添加到 input.conf 文件中：

```
[WinEventLog:Security]
blacklist1 = EventCode = "4662" Message = "Account Name:\s+(example account)"
```

Splunk 平台如何处理日志文件旋转

Splunk 平台会识别操作系统何时旋转其所监视的文件，并且不会再次读取旋转的文件。例如，如果 Splunk 平台正在监视 /var/log/messages，则不会读取 /var/log/messages1。

监视处理器会选取新文件并读取文件的前 256 个字节。然后处理器会对此数据进行哈希处理，以生成开始和结束循环冗余码校验 (CRC)，充当表示文件内容的指纹。Splunk 平台会使用此 CRC 在数据库中查找某个条目，该数据库包含 Splunk 平台之前见过的所有文件的起始 CRC。如果在此数据库中找到匹配项，则查找将返回有关该文件的一些值。最重要的值是 seekAddress，它表示 Splunk 平台已读取到已知文件的字节数，以及 seekCRC，它是该位置数据的指纹。

通过使用此次查找的结果，Splunk 平台可对文件进行分类。

Splunk 平台如何对文件进行分类

Splunk 平台根据 CRC 校验的以下结果对文件进行分类。

CRC 未找到匹配项

如果来自于数据库开始处文件的 CRC 没有匹配项，这表明是一个新文件。然后，Splunk 平台会完成以下步骤：

1. Splunk 平台从文件起始处读取文件数据。
2. Splunk 平台在处理文件时，会使用新 CRC 和 Seek Addresses 更新数据库。

CRC 找到匹配项

如果来自于数据库开始处文件的 CRC 有匹配项，则 Seek Address 位置的内容与文件中该位置存储的 CRC 相匹配，并且该文件的大小比 Splunk 平台存储的 Seek Address 大，该文件之前由 Splunk 平台读取，但自上次读取以来包含新数据。然后，Splunk 平台会完成以下步骤：

1. Splunk 平台打开文件并转到文件内的 seekAddress，这是 Splunk 平台最后完成文件时文件的末尾。
2. Splunk 平台从此处开始读取文件数据。

如果来自于数据库开始处文件的 CRC 有匹配项，但 Seek Address 位置的内容与文件中该位置存储的 CRC 不匹配，则该文件的结果有以下几种可能：

- Splunk Enterprise 读取某些具有相同初始数据的文件，但某些已读取过的内容在位置上发生了修改。
- 该文件是以相同内容开头的不同文件。

由于用于内容跟踪的数据库由 CRC 开始处提供线索，无法为两个不同数据流独立跟踪进度，且需要进一步配置。

使用重复的 CRC 配置文件

因为默认情况下，CRC 启动检查仅针对文件的前 256 个字节，因此非重复文件可能存在重复的开始 CRC，尤其是具有相同标题的文件。要处理这种情况，请进行以下更改：

- 使用 input.conf 配置文件中的 initCrcLength 设置来增加用于 CRC 计算的字符数，使其长度超过文件开头可能存在的任何静态标题。
- 在 inputs.conf 配置文件中为文件配置输入时，请使用 crcSalt 设置。如果将 crcSalt 设置配置为 <SOURCE>，请确保每个文件都具有唯一的 CRC。这样，Splunk 平台假定每个路径名称包含唯一内容。有关配置此设置的其他信息，请参阅“使用 inputs.conf 监视文件和目录”。

请勿将 crcSalt = <SOURCE> 用于操作系统定期旋转的日志文件，或其他会重命名日志文件或将日志文件移动到 Splunk 平台监视的其他位置的方案。该操作会阻止 Splunk 平台在旋转或重命名过程中识别日志文件，从而导致 Splunk 平台多次对数据建立索引。

从网络来源获取数据

从 TCP 和 UDP 端口获取数据

Splunk 平台允许您引入通过网络端口传入的数据。它可以接受来自传输控制协议（TCP）和用户数据报协议（UDP）网络协议的数据。

Splunk® Enterprise 接受来自**重型转发器**或**通用转发器**的此类数据，这些转发器捕获数据并将其发送到 Splunk® Enterprise 实例。出于安全因素考虑，只有当转发器拥有正确的安全套接字层（SSL）证书以连接到实例时，Splunk® Enterprise 才会接受来自该转发器的连接。如果您想从 TCP 或 UDP 数据来源（如 syslog 服务）发送数据，请使用通用转发器侦听该数据来源，并将数据转发至您的 Splunk® Enterprise 部署。

您可以配置转发器以接受任何 TCP 或 UDP 端口上的输入。转发器将使用抵达这些端口的任何数据。您可以使用此方法从网络服务（如 syslog 服务）捕获数据。您还可设置 netcat 服务并将其与网络端口绑定。

网络端口和 Splunk Enterprise

TCP 是以 Splunk Enterprise 数据分发方案为基础的网络协议。使用 TCP 协议从任意远程主机发送数据到您的 Splunk Enterprise 服务器。Splunk Enterprise 可以为来自 syslog-ng 或其他任何通过 TCP 传输的应用程序的远程数据建立索引。

Splunk Enterprise 和通用转发器都支持通过 UDP 进行监视。最佳实践是尽可能使用 TCP 发送网络数据。UDP 并非理想的数据传输协议，原因较多，包括 UDP 无法确保网络封包的交付。

当您监视 TCP 网络端口时，Splunk Enterprise 或通用转发器以该身份运行的用户必须获得授权访问您想要监视的端口。默认情况下，在很多 UNIX 操作系统上，您必须以根用户的身份运行 Splunk Enterprise 才能直接侦听 1024 以下的端口。

请在您使用网络监视输入前确定您的网络设备如何处理外部监视

在您开始使用网络监视器监视网络设备的输出前，请先确认该网络设备与外部网络监视器的交互方式。

如果您将一些网络设备（如 Cisco 自适应安全设备（ASA））配置为记录 TCP 网络活动，且该网络设备无法连接至监视器，则可能会降低设备的性能或停止记录。默认情况下，Cisco ASA 将在遭遇网络拥挤或网络连接问题时停止接受传入的网络连接。

将网络输入添加到转发器并将数据发送到 Splunk Cloud

Splunk Cloud 可以接受仅来自通用或重型转发器的网络数据。在为 Splunk Cloud 收集网络数据之前，您必须具备以下条件：

- 已安装的通用或重型转发器。
- Splunk Cloud 通用转发器凭据包。此凭据包可设置与 Splunk Cloud 实例的转发连接，并确保数据在转发器和 Splunk Cloud 之间安全传输。
- 用于编辑输入和转发配置的文本编辑器。

使用配置文件添加网络输入

在重型转发器或通用转发器上，使用文本编辑器将网络输入的段落添加到 $SPLUNK_HOME/etc/system/local/ 目录中的 input.conf 配置文件、或 Windows 上的 %SPLUNK_HOME%\etc\system\local，或 $SPLUNK_HOME/etc/apps/ 中您自己的自定义应用程序目录中。如您之前未使用过 Splunk 配置文件，请在开始前先参阅 Splunk Enterprise *《管理员手册》* 中的“关于配置文件”。

虽然此过程侧重于配置转发器以将网络数据发送到 {PONYDOCSPRODUCT} 实例，但您无需修改任何 Splunk Enterprise 实例即可执行此操作。

您可以为输入类型配置任意数量的设置和值。如果您未指定某个设置的值，转发器将使用默认值。这些值可以在 Splunk 平台代码中定义，或存在于实例上 $SPLUNK_HOME/etc/system/default/ 目录中的默认配置文件中，或者存在于 Windows 上的 %SPLUNK_HOME%\etc\system\default 中。

以下是配置网络输入的一般过程：

1. 使用文本编辑器打开本节所述其中一个目录下的 inputs.conf 配置文件。
2. 添加代表您要收集的网络数据类型的输入段落。
3. （可选）提供其他设置以配置 Spunk 平台处理数据的方式。
4. 保存该文件并退出文本编辑器。
5. 重新启动转发器或 Splunk Enterprise 实例。

配置 TCP 网络输入

当您配置 TCP 网络输入时，转发器会在该输入上侦听通过 TCP 协议传入的网络数据。

此段落将转发器配置为在指定的 <port> 上侦听由 <remote server> 指定的服务器。如果 <remote server> 为空，则软件将在指定端口上侦听所有连接。

```
[tcp://<remote server>:<port>]
<attrbute1> = <val1>
<attrbute2> = <val2>
...
```

以下设置控制数据在 Splunk 平台上的存储方式：

设置	描述	默认
host = <string>	将此段落的主机字段设为一个静态值。也设置主机键的初始值。Splunk 平台在分析和新建索引过程中使用该键，特别是设置主机字段。它也在搜索时间使用该主机字段。 平台会在 <string> 前面加上 host::。	生成数据的主机的 IP 地址或完全限定域名。
index = <string>	设置 Splunk® Enterprise 从该输入存储事件的索引。Splunk 平台会在 <string> 前面加上 index::。在 Splunk Cloud 中，在配置此设置之前确认此索引存在。	main 或设置默认索引的任何值
sourcetype = <string>	设置来自此输入的事件的来源类型字段。同时声明该数据的来源类型，而不是让 Splunk® Enterprise 决定。这对于可搜索性以及在分析和新建索引期间对此类型数据应用相关格式设置都很重要。 设置 sourcetype 键的初始值。Splunk® Enterprise 在分析和新建索引过程中使用此键，特别是，可在新建索引期间使用此键来设置来源类型字段。Splunk® Enterprise 使用在搜索时间使用的来源类型字段。 Splunk 平台会在 <string> 前面加上 sourcetype::。	Splunk® Enterprise 会根据数据的各个方面选取一种来源类型。没有硬编码的默认值。
source = <string>	设置来自此输入的事件的来源字段。Splunk 平台会在 <string> 前面加上 source::。 除非绝对必要，否则不要覆盖来源键。通过记录从中检索数据的文件，输入层通常提供更准确的字符串来帮助分析和调查问题。在覆盖此值之前，请考虑使用来源类型、标记和搜索通配符。	输入文件路径
indexQueue	指定输入处理器用来存储其所读取事件的位置。 设置为 parsingQueue 时，会将 props.conf 文件和其他分析规则应用到您的数据。设置为 indexQueue 时，会将您的数据直接发送到索引。	分析队列
dns | none	ip 值将主机设置为远程服务器的 IP 地址。 dns 将主机设置为远程服务器的 DNS 项。 none 保留主机的指定值。	ip

配置使用 SSL 加密的 TCP 网络输入

如要从转发器或第三方系统接收已加密且未分析的网络数据，请使用此段落类型。将 <port> 设置为转发器或第三方系统发送未分析的加密数据的端口。

```
[tcp-ssl:<port>]
```

配置 UDP 网络输入

此类型输入段落类似于 TCP 类型，但是，它是在 UDP 网络端口上进行侦听。如果您提供了 <remote server>，指定的端口将仅接受来自该主机的数据。如果您未对 <remote server> 做任何指定，该端口将接受来自任何主机的数据。

```
[udp://<remote server>:<port>]
```

```
<attrbute1> = <val1>
<attrbute2> = <val2>
...
```

以下设置控制 Splunk 平台存储数据的方式：

设置	描述	默认
host = <string>	将此段落的主机字段设为一个静态值。也设置主机键的初始值。Splunk® Enterprise 在分析和新建索引过程中使用该键，特别是设置主机字段。它也在搜索时间使用该主机字段。在 <string> 前面加上 host::。	生成数据的主机的 IP 地址或完全限定域名。
index = <string>	设置 Splunk® Enterprise 从该输入存储事件的索引。在 <string> 前面加上 index::。	main 或设置默认索引的任何值
sourcetype = <string>	设置来自此输入的事件的来源类型字段。也声明该数据的来源类型，而不是允许 Splunk® Enterprise 确定它。这对于可搜索性以及在分析和新建索引期间对此类型数据应用相关格式设置都很重要。 设置 sourcetype 键的初始值。Splunk® Enterprise 在分析和新建索引过程中使用此键，特别是，可在新建索引期间使用此键来设置来源类型字段。它也使用在搜索时间使用的来源类型字段。 在 <string> 前面加上 sourcetype::。	Splunk® Enterprise 会根据数据的各个方面选取一种来源类型。没有硬编码的默认值。
source = <string>	设置来自此输入的事件的来源字段。在 <string> 前面加上 source::。 除非绝对必要，否则不要覆盖来源键。通过记录从中检索数据的文件，输入层通常提供更准确的字符串来帮助分析和调查问题。在覆盖此值之前，请考虑使用来源类型、标记和搜索通配符。	输入文件路径。
indexQueue	设置输入处理器用来存储其所读取事件的位置。设置为 parsingQueue 时，会将 props.conf 文件和其他分析规则应用到您的数据。设置为 indexQueue 时，会将您的数据直接发送到索引。	分析队列
_rcvbuf = <integer>	设置 UDP 端口的接收缓冲区（以字节为单位）。如果值为 0 或负值，则 Splunk® Enterprise 会忽略该值。	值为 1,572,864，除非此值对于操作系统过大。在这种情况下，Splunk® Enterprise 从此默认值持续将值减半，直到缓冲区大小到达可接受级别。
no_priority_stripping = true | false	设置 Splunk Enterprise 处理接收 syslog 数据的方式。 如果您将该设置设置为 true，Splunk® Enterprise 不会从接收的事件中去除 <priority> syslog 字段。 根据此设置的设置方式，Splunk® Enterprise 也会以不同方式设置事件时间戳。当设置为 true 时，Splunk® Enterprise 会优先采用来自数据来源的时间戳。当设置为 false 时，Splunk Enterprise 会为事件分配本地时间。	设置为 false（Splunk® Enterprise 去除 <priority>。）
no_appending_timestamp = true | false	设置 Splunk® Enterprise 将时间戳和主机应用到事件的方式。 如果您将该设置设置为 true，Splunk® Enterprise 不会将时间戳和主机附加到接收的事件。 如果您想将时间戳和主机附加到接收的事件，请勿配置此设置。	false（Splunk® Enterprise 将时间戳和主机附加到事件）

使用 Splunk Web 添加网络输入

您可以使用 Splunk Web 在 Splunk Enterprise 或要配置为将数据发送到 Splunk Cloud 的重型转发器上添加网络输入。Splunk Web 在通用转发器上不可用，并且 Splunk Cloud 无法直接使用 Splunk Web 监视网络输入。

转到“添加数据”页面

您可以通过两种方式进入“添加数据”页面。

要通过 Splunk 设置转到“添加数据”页面，请执行以下步骤：

1. 单击**设置**。
2. 请单击**数据导入**。
3. 请选择 **TCP** 或 **UDP**。
4. 单击**新建本地 TCP** 或**新建本地 UDP** 以添加输入。

要通过 Splunk 主页转到“添加数据”页面，请执行以下步骤：

1. 单击 Splunk 主页中的**添加数据**链接。
2. 请单击**监视**以监视本地计算机上的网络端口或**转发**以从另一个计算机上接收网络数据。

   转发文件需要其他设置。

3. 如果您选择了**转发**，则选择或新建要此输入应用的转发器组。
4. 单击**下一步**。

指定网络输入

1. 单击 **TCP / UDP** 以添加输入。
2. 单击 **TCP** 或 **UDP** 按钮即可选择 TCP 或 UDP 输入。
3. 在**端口**字段中，输入端口号。
4. 在**来源名称覆盖**字段中，如果需要，输入新数据源名称以覆盖默认数据来源值。

   更改“来源名称覆盖”值前，请先咨询 Splunk 支持。

5. 如果是 TCP 网络输入，请决定是希望此端口接受所有主机的连接，还是只接受**仅接受来自以下位置的连接**字段中的一个主机的连接。如果您要输入接受来自一个主机的连接，则输入该主机的主机名或 IP 地址。可以使用通配符指定主机。
6. 单击**下一步**以继续到**输入设置**页面。

指定输入设置

输入设置页面允许您配置来源类型、应用程序上下文、默认主机值和索引。所有这些参数均为可选参数。

1. 设置**来源类型**。这是 Splunk Enterprise 添加到事件中并用来确定处理特性（如时间戳和事件界限）的默认字段。
2. 设置**主机**的值。您有几个选择：
   - 选择 **IP** 将输入处理器设置为使用远程服务器的 IP 地址重写主机。
   - 选择 **DNS** 将主机设置为远程服务器的 DNS 项。
   - 选择**自定义**将主机设置为用户定义的标签。

   有关设置主机值的更多信息，请参阅“关于主机”。

   > host 值只是设置生成事件中的**主机**字段。设置此值不会引导 Splunk 平台查找网络中的特定主机。

3. 对于**索引**，为此输入设置您希望 Splunk Enterprise 将数据发送到其中的索引。如果未定义多个索引来处理不同类型的事件，请保留 default 值。除了用户数据的索引之外，Splunk® Enterprise 还有许多实用工具索引，这些索引也会显示在此下拉框中。
4. 请单击**查看**。

查看您的选择

在输入所有输入设置后，您可查看您的选择。Splunk 平台会列出您选定的选项，包括监视器的类型、数据来源、来源类型、应用程序上下文和索引。

1. 查看该设置。
2. 如果它们不符合您的期望，请单击左尖括号（ < ）即可返回到向导中的上一个步骤。否则，请单击**提交**。

“成功”页面显示且 Splunk 平台开始索引指定的网络输入。

使用 CLI 添加网络输入

您可以在通用或重型转发器上使用 CLI， 配置转发器以将数据发送到 Splunk Cloud。您还可以在 Splunk Enterprise 实例上使用 CLI。如需访问 CLI，请导航至 $SPLUNK_HOME/bin/ 目录（Windows 上为 %SPLUNK_HOME%\bin），并使用 ./splunk 命令。

如果您遇到困难，CLI 内有帮助说明。键入 splunk help 即可访问 CLI 帮助。每个命令也都设有单独的帮助页面，键入 splunk help <command> 即可访问。

以下 CLI 命令可用于网络输入配置：

命令	命令语法	操作
add	add tcp|udp <port> [-parameter value] ...	从 <port> 添加输入。
edit	edit tcp|udp <port> [-parameter value] ...	编辑之前为 <port> 添加的输入。
remove	remove tcp|udp <port>	移除之前添加的数据导入。
list	list tcp|udp [<port>]	列出当前配置的监视器。

<port> 是用于侦听数据的端口号。您运行 Splunk 平台所用的用户身份必须有权访问此端口。

设置以下任何可选参数，即可修改每个输入的配置：

参数	描述
sourcetype	为来自输入来源的事件提供 Sourcetype 字段值。
index	为来自输入来源的事件提供目标索引。
hostname	为来自输入来源的事件提供要设置为主机字段值的主机名。
remotehost	提供只接受来自其数据的 IP 地址。
resolvehost	设置为 true 或 false (T | F)。默认为 false。设置为 true 以使用 DNS 为来自输入来源的事件设置主机字段值。
restrictToHost	提供主机名或 IP 地址以仅接受来自指定主机或 IP 地址的连接。

示例

以下示例显示如何在 *nix 系统上配置 UDP 输入以监视端口 514 并将来源类型设置为 syslog：

```
./splunk add udp 514 -sourcetype syslog
```

以下示例显示如何在 *nix 系统上使用 DNS 名称解析设置 UDP 输入主机值。输入用户名和密码以使用 auth：

```
./splunk edit udp 514 -resolvehost true -auth admin:ch@ng3d
```

更改 TCP 网络输入中的受限制主机

在您新建 TCP 输入时，如果您决定只接受特定主机的连接，则在保存该输入后，您将无法通过 Splunk Web 或 CLI 更改和删除该主机。

要更改或删除某个端口的受限制主机，必须先删除包含旧的受限制主机的输入。之后，您必须添加包含新的受限制主机或没有任何限制的新输入。

UDP 数据包和行合并

Splunk 平台不会将每个 UDP 数据包作为独立事件进行索引。相反，它会对数据流执行事件合并，并且它将没有清晰时间戳的事件合并在一起。

在 props.conf 文件中编辑基础来源类型并将 SHOULD_LINEMERGE 设置设置为 false 即可避免这一问题。这样做可以阻止 Splunk 平台将数据包合并在一起。

Splunk Enterprise 通过 UDP 网络协议处理 syslog 数据的方式

如果您运行 Splunk Cloud，您可以将 Splunk 通用转发器配置为侦听用户数据报协议 (UDP) 网络端口，并将该数据转发到您的 Splunk Cloud 部署。

Splunk Enterprise 索引器可以充当 syslog 服务器，处理符合 syslog 消息传递标准的传入数据流。Splunk Enterprise 还可以充当 syslog 消息发送器。Splunk Cloud 无法发送 syslog 消息，也无法将消息从一台设备移动到另一台设备。

Splunk 平台处理 syslog 输入的方式

当您配置 UDP 网络输入以侦听 Splunk Enterprise 或通用转发器中的 syslog 标准数据流时，通过输入到达的任何 syslog 事件都会收到时间戳和已连接的主机字段。平台在新建索引前会在每个事件前面加上这些字段。当您配置通用转发器以将数据发送到 Splunk Cloud 时，Splunk Cloud 在从通用转发器接收字段时为其建立索引。

Splunk 平台不会以此方式修改传输控制协议（TCP）网络数据包。如果您通过 TCP 发送 syslog 数据，则平台不会从事件中剔除优先级信息。但它会在事件前面加上主机名称和时间戳，除非您对此进行配置。

Splunk Enterprise 处理 syslog 输出的方式

以下部分仅适用于 Splunk Enterprise。Splunk Cloud 和通用转发器都无法将事件转发到另一个 syslog 服务器。

Splunk Enterprise 可将事件转发到其他 syslog 服务器上。此时，它会在事件前面加上优先级信息，这样下游 syslog 服务器就可正确地解读事件。

当该事件到达下游 syslog 服务器时，此计算机会在事件前面加上时间戳、优先级和连接的主机名称（即 Splunk Enterprise 实例）。

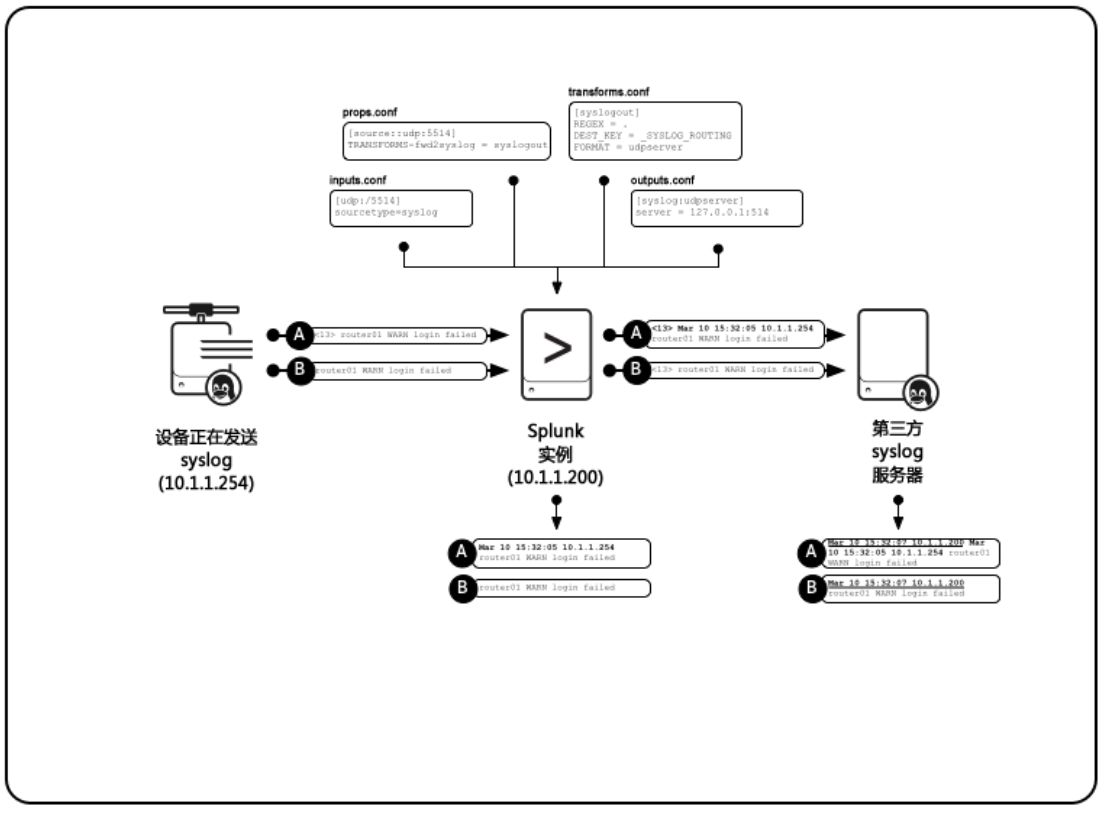
当您将事件转发到 syslog 服务器时，您也可以在事件前面加上时间戳和主机名称。当数据离开 Splunk Enterprise 实例时，您可以将其作为修改数据的一部分。

有关配置路由、筛选和来源类型使用的信息，请参阅 Splunk Enterprise *《转发数据》*手册中的“路由和筛选数据”以及*《管理员手册》*中的“props.conf 规范文件”。

Splunk Enterprise 如何在您对其进行配置以使用 syslog 来源类型时移动 syslog 事件

以下部分仅适用于 Splunk Enterprise。Splunk Cloud 无法将 syslog 事件发送到其他下游 syslog 服务器。

下图显示了 Splunk Enterprise 如何将两条 syslog 消息从一个 syslog 服务器移动到另一个。在本图中，Splunk Enterprise 侦听某个 UDP 网络端口并为传入事件新建索引。另一方面，同一实例将事件转发到第三方 syslog 服务器。



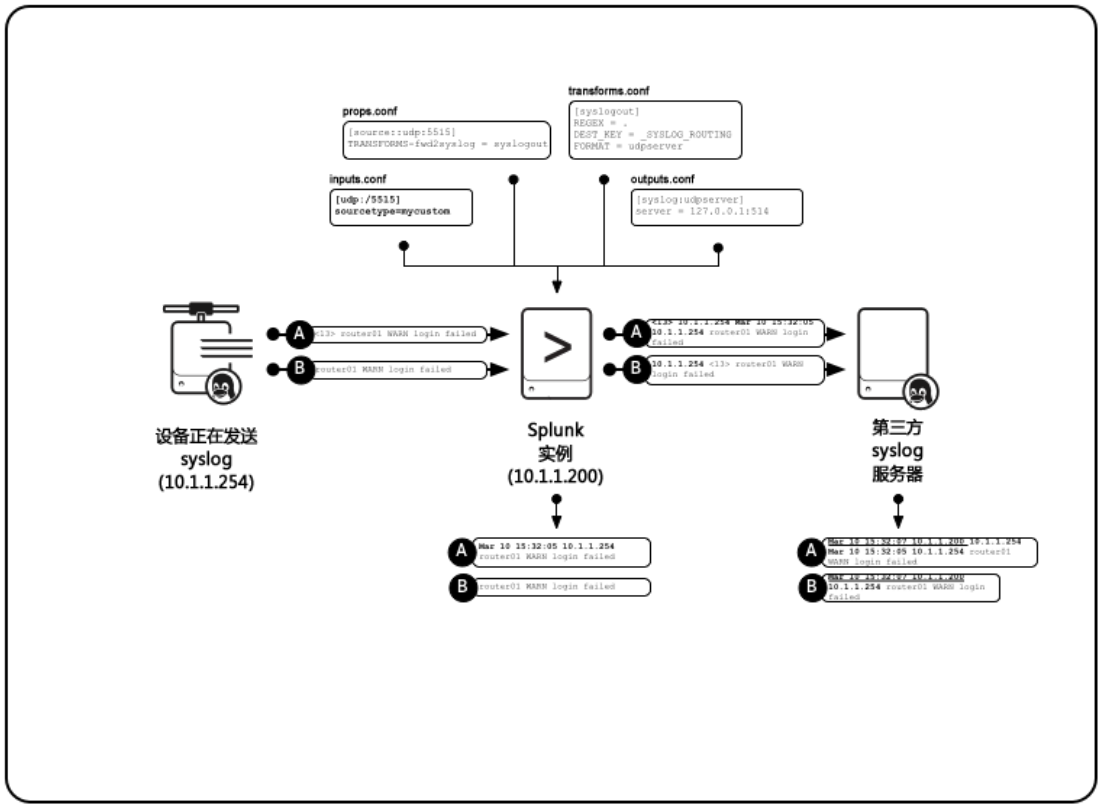
在本图中，消息 A 作为 syslog 事件的来源，消息 B 作为相似事件（其没有与之相关的优先级信息）的来源。一旦接收，Splunk Enterprise 用时间戳和生成事件的主机标记事件。

如果您将实例配置为转发器，则 Splunk Enterprise 在其将事件转发到 syslog 服务器上以前会通过添加在 outputs.conf 中指定的优先级标题来转换事件。一旦它们到达 syslog 服务器，该服务器会在其从 Splunk Enterprise 实例接收到它们时在事件前面加上时间戳和主机数据。

Splunk Enterprise 如何在您配置自定义来源类型时移动 syslog 事件

以下部分仅适用于 Splunk Enterprise。Splunk Cloud 无法移动 syslog 事件。

在本图中，已将 Splunk Enterprise 配置为使用非 syslog 来源类型。初始的消息 A 和消息 B 与第一个示例相同。在本示例中，Splunk Enterprise 在事件前面加上一个原始主机名称或 IP 地址。

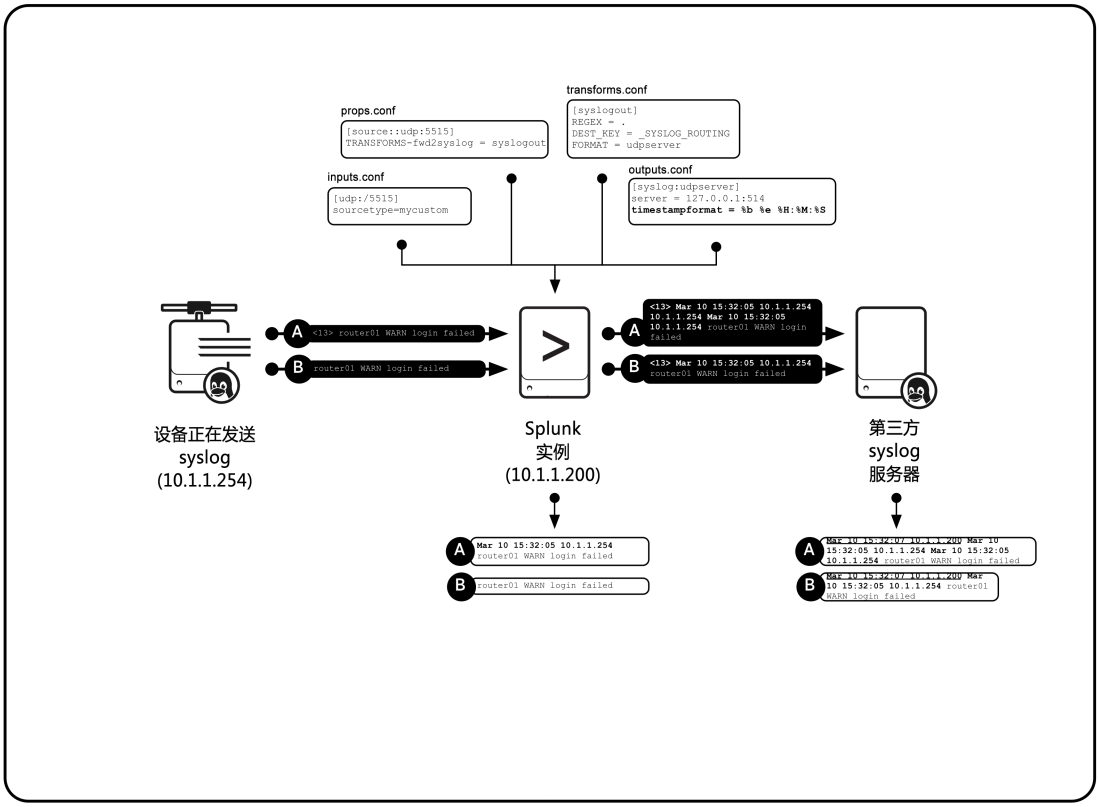


Splunk Enterprise 如何在您为其配置时间戳时移动 syslog 事件

以下部分仅适用于 Splunk Enterprise。Splunk Cloud 无法移动 syslog 事件。

当您转发 syslog 事件时，您也可配置 Splunk Enterprise 以向这些事件中添加时间戳。当您不希望下游服务器添加其自己的时间戳时，您可以为事件添加时间戳。

下表显示了所需的属性并描述了 Splunk Enterprise 处理数据的方式。初始的消息 A 和消息 B 与第一个和第二个示例相同。Splunk Enterprise 在事件前面加上时间戳和原始的主机名或 IP 地址。



使用 Splunk Enterprise 作为 syslog 服务器或消息发送器的注意事项

以下部分仅适用于 Splunk Enterprise。Splunk Cloud 不能用作 syslog 服务器或消息发送器。

尽管可以配置 Splunk Enterprise 直接接收 syslog 事件，不这样进行配置的原因如下所示：

- Splunk 最佳实践包括设置单独计算机，运行 syslog 服务处理 syslog 任务。
- Splunk 平台默认可修改 syslog 数据作为索引过程的一部分。它将时间戳和主机分配给事件。
- 这种情境下，Syslog 数据只流入一个 Splunk Enterprise 实例。在带有多个索引器的部署中，您必须执行其他操作跨索引器分发数据流
- 如果 Splunk Enterprise 因任何原因出现故障，则停机期间到达的任何 syslog 消息可能会丢失，不可恢复。

正常使用时，不要用 Splunk Enterprise 替换 syslog 服务器，除非您没有其他选择。

如果您必须保留原始 syslog 数据（例如，数据保留政策要求访问未经处理的事件），请考虑使用 syslog-ng 等工具，同时将原始数据保存到日志文件，并将这引起事件转发到您的 Splunk Enterprise 实例。这些工具为您提供的优势是，可以在以后需要时为日志文件建立索引。

向您的 Splunk 部署发送 SNMP 事件

简单网络管理协议 (SNMP) 是用于监控网络设备的网络协议。SNMP 数据源包括轮询消息和陷阱。

SNMP 陷阱表示远程代理发送的通知或告警。在典型的网络环境中，中央网络管理系统将收集 SNMP 陷阱。SNMP 轮询需要以下组件：

- 能够接收轮询请求的网络代理设备
- 查询代理以请求特定状态信息的轮询节点

在哪里可以找到 Splunk 平台的 SNMP 支持

Splunk 平台不包括对 SNMP 协议的本机支持。您可以从多个提供 SNMP 支持的 Splunk 应用和工具中进行选择：

- 如果网络管理软件收集的 SNMP 陷阱已写入日志文件，则可以使用转发器监控日志文件并将数据发送到 Splunk 平台。请参阅“使用 inputs.conf 监视文件和目录”。
- 您可以查看 Splunkbase 上可用的应用，以协助您从网络中收集陷阱或轮询 SNMP 数据。请参阅 Splunkbase 上的相关应用。
- 您可以使用 Splunk Stream 通过内置协议支持从 SNMP 消息收集消息统计信息。请参阅《Splunk Stream 安装和配置手

册》。

另请参阅

如果您正在寻找在 Linux 上安装和配置 `snmptrapd` 服务的示例，请查看 Splunk 关于“使用 ITSI 事件分析管理 SNMP 陷阱”的博客文章。

有关将 SNMP 数据源集成到 Splunk Enterprise 的指南，当前的 Splunk 客户可以使用 OnDemand Services 支持服务。请参阅“支持计划”。

获取 Windows 数据

使用 Splunk 平台监视 Windows 数据

您可以将任何类型的 Windows 数据引入 Splunk 平台。例如，您可以为“事件日志”通道、“注册表”或 Active Directory 建立索引。Splunk 软件还提供了一组标准的 Splunk 输入，例如文件和目录、网络监视输入以及脚本式输入。

对于 Splunk Cloud，与许多其他输入类型一样，您必须使用在 Windows 上运行的通用或重型转发器来收集数据并将其发送到您的 Splunk Cloud 实例。Splunk Enterprise 随附适用于多个版本的 Windows 和 Windows Server 的安装程序。如果您运行 Splunk Enterprise，您可以直接在您的 Windows 计算机上安装它或通用转发器。

下列专用输入仅适用于 Windows 安装：

输入	描述	文档
Windows 事件日志	对 Windows 事件日志服务在计算机的任何可用事件日志通道上生成的事件进行监视。您可以在本地 Windows 计算机上收集事件，也可以使用**通用转发器**或 Windows Management Instrumentation (WMI) 远程收集事件。	使用 Splunk Cloud 监视 Windows 事件日志数据
性能监视	请在装有 Splunk Cloud 的 Windows 计算机上收集性能数据，然后告警或报告该数据。性能监视器中提供的任何性能计数器也可用于 Splunk Cloud。您可以通过通用转发器或使用 WMI 本地或远程监视性能。	监视 Windows 性能
通过 WMI 远程监视	Splunk Cloud 可以使用 WMI 通过通用转发器来访问远程计算机上的事件日志和性能数据。	通过 Windows Management Instrumentation (WMI) 监视数据
注册表监视	您可以使用“注册表”监视功能监视本地 Windows 注册表的更改。您可以使用通用转发器从 Windows 计算机收集注册表数据并将数据发送到 Splunk Cloud。	监视 Windows 注册表数据
Active Directory 监视	Splunk Cloud 可以审计任何 Active Directory 更改，包括用户、组、计算机和组策略对象的更改。你可将 Active Directory 数据转发到另一个 Splunk Enterprise 服务器。	监视 Active Directory

将 Windows 数据转发到 Splunk Cloud

监视 Windows 数据的 Splunk Cloud 部署包含以下组件：

- Splunk Cloud 实例，您可以在其中查看 Windows 数据。
- 您要从中收集 Windows 数据的每台 Windows 计算机上的通用转发器。

根据您的 Windows 网络的大小，您可能希望设置一层**中间转发器**来聚合数据并将其发送到您的 Splunk Cloud 实例。如果您想在新建索引之前以任何方式转换此数据，则必须至少使用一个 Splunk Enterprise 重型转发器来执行转换。

Windows 实例上的通用转发器将收集 Windows 数据。然后，使用 Splunk Cloud 通用转发器凭据包将数据发送到 Splunk Cloud，该包处理实例的连接和身份验证。如果您设置了一个中间转发器，该转发器还会使用相同的凭据包来连接到 Splunk Cloud 并在其中进行身份验证。

Splunk Cloud 实例为数据建立索引并使其可供您搜索。您可以安装 Splunk App for Windows Infrastructure，以在预构建的仪表板和报告中查看 Windows 数据。

通用转发器必须以有权访问您要收集的特定 Windows 数据的用户身份运行。有关确定此 Windows 用户的信息，请参阅“选择通用转发器以其身份运行的 Windows 用户”。

将 Windows 数据转发到 Splunk Enterprise

与将 Windows 数据转发到 Splunk Cloud 类似，监视 Windows 数据的 Splunk Enterprise 部署由 Splunk Enterprise 安装和（可选）您要从中收集 Windows 数据的每台 Windows 计算机上的转发器组成。与 Splunk Cloud 部署不同，Splunk Enterprise 可以存在于同一台 Windows 计算机上。

如果您想从另一台 Windows 计算机转发 Windows 数据，可以使用通用转发器，如同您可以并且必须使用 Splunk Cloud 部署一样。

在 Windows 计算机上安装 Splunk Enterprise 的注意事项

当您在 Windows 上安装并部署 Splunk Enterprise 时，请考虑以下事项：

注意事项	描述
验证	要在您网络的远程 Windows 计算机上执行任何操作，Splunk Enterprise 必须以具有能够访问这些计算机凭据的用户身份运行。请在部署前使这些凭据可用。请参阅“确定如何监视远程 Windows 数据的相关注意事项”。
磁盘带宽	Splunk Enterprise 索引器需要占用很多磁盘 I/O 带宽，尤其是在为大量数据新建索引时。请确保您配置所有安装的防毒软件，以避免监视 Splunk Enterprise 目录或进程，因为此类扫描将显著降低性能。
共享的主机	在运行其他服务的主机，如 Exchange、SQL Server 或虚拟机管理程序上安装 Splunk Enterprise 之前，请参阅*《容量规划》*手册中的“Splunk Enterprise 容量规划简介”。

如何将 Windows 数据导入您的 Splunk 部署

您可以使用 Splunk 软件收集以下 Windows 数据：

您可以收集的 Windows 数据	支持文档链接
事件日志	使用 Splunk Cloud 监视 Windows 事件日志数据
文件系统更改	监视 Windows 上的文件系统更改
Active Directory	监视 Active Directory
通过 Windows Management Instrumentation (WMI) 基础设施收集数据	通过 Windows Management Instrumentation (WMI) 监视数据
注册表数据	监视 Windows 注册表数据
性能指标	监视 Windows 性能
主机信息	监视 Windows 主机信息
打印信息	监视 Windows 打印机信息
网络信息	监视 Windows 网络信息

因为只有 Windows 计算机提供这些类型的数据，因此只有 Windows 版的 Splunk 平台可以获取这些数据。其他操作系统无法直接收集 Windows 数据。您可以将 Windows 数据从 Windows 计算机发送到未运行 Windows 的 Splunk 平台实例。如果您使用的是 Splunk Cloud 而且想要监视这些输入，则 Splunk 通用转发器是唯一的选择。

Splunk 平台如何在启动和关机时与 Windows 模块化和脚本化输入互动

在 Splunk 平台中配置脚本化或**模块化** Windows 数据输入时，splunkd 服务会将信号发送到输入以开始收集数据。同样地，您完全关闭 Splunk 平台之后，服务会发送一个不同的信号到输入，指示停止收集数据、清空并退出。

以下表格显示在启动和关闭时 splunkd 服务发送到模块化和脚本化 Windows 输入的信号或控制消息。

过程	信号
启动	CreateProcess
关闭	CTRL_BREAK_EVENT

使用 Splunk Web 收集 Windows 数据

几乎所有 Windows 输入都允许您使用 Splunk Web 界面将数据导入 Splunk Enterprise。MonitorNoHandle 输入是例外，您必须使用配置文件设置该输入。

按照以下步骤在 Splunk Web 中收集 Windows 数据：

1. 登录您的 Splunk 部署。
2. 单击**设置** > **数据导入**。
   数据导入页面显示。
3. 从可用输入列表中，找到要从可用输入列表中添加的 Windows 输入。
4. 在输入的“操作”列中单击**添加新**。
5. 按照您选择的输入类型的说明进行操作。

6. 单击**保存**。大多数情况下，数据集合会立即开始。

使用配置文件收集 Windows 数据

如果您无法使用 Splunk Web 配置 Windows 输入，例如在通用转发器上，您必须使用配置文件。Windows 上的通用转发器安装程序允许您在安装时配置一些 Windows 输入。

很多情况下，配置文件可以对 Splunk Web 提供更多控制。某些输入只可通过这种方式配置。

按照以下步骤使用配置文件收集 Windows 数据：

1. 打开命令提示符或 PowerShell 窗口。
2. 将目录更改为 Splunk 平台实例上的 %SPLUNK_HOME%\etc\system\local 目录。
3. 编辑此目录中的 inputs.conf 配置文件。如果文件不存在，则可能需要创建该文件。
4. 通过定义输入段落、设置和值将输入添加到 inputs.conf 文件。
5. 保存文件并将其关闭。
6. 重新启动 Splunk 平台实例。
   该软件重新加载配置文件并开始收集基于新配置的数据。

有关确定如何监视远程 Windows 数据的注意事项

如果要监视不在本地 Windows 计算机上的 Windows 数据，请考虑以下选项。

Splunk 平台可通过以下两种方式之一收集要建立索引的远程 Windows 数据：

- 从 Splunk 转发器
- 使用 Windows Management Instrumentation (WMI)

如果是 Splunk Cloud 部署，您必须在 Windows 计算机上使用 Splunk 通用转发器才能监视远程 Windows 数据。

使用转发器收集远程 Windows 数据

尽量使用通用转发器来收集远程 Windows 数据。通用转发器具有以下优势：

- 在安装好的计算机上使用的网络和磁盘资源最少。
- 可以将其安装为非特权用户，而 WMI 则设置为需要管理员权限才能访问。
- 如您将通用转发器安装为“本地系统”用户，它将和 WMI 一样具有访问计算机的管理员权限，且从计算机中获取数据时无需验证。
- 通用转发器在大环境中扩展良好，易于安装。使用诸如系统中心配置管理器 (SCCM) 等 Microsoft 部署工具或诸如 Puppet 等第三方分发解决方案即可手动安装通用转发器。

在您安装通用转发器之后，它便会在本地收集信息并将信息发送到 Splunk® Enterprise。您可以在安装期间或之后使用**部署服务器**或手动分发配置更新，以将要收集的数据告知给转发器。您还可将加载项安装到通用转发器中。

使用通用转发器有一些缺点，具体取决于您的网络配置和布局。请参阅本主题中的“Splunk 转发器与 WMI”。

使用 WMI 收集远程 Windows 数据

WMI 框架允许 Splunk 平台从远程 Windows 计算机收集几乎所有类型的数据。在此配置中，Splunk 平台将以您在安装时（或稍后在服务控制面板中）指定的用户身份运行。有关更多信息，请参阅*《安装手册》*中的“选择 Splunk Enterprise 应以其身份运行的 Windows 用户”。

这种配置有以下优点：

- 可为 Splunk 平台赋予指定帐户对网络所具有的相同远程访问权限。
- 允许索引器从整个企业的远程 Windows 计算机收集数据并将该数据放入中央存储库中。
- 非常适用于每个网段至少包含一个索引器的中小型网络。

这种集合方法有一些注意事项。请参阅本主题后面的“通过 WMI 导入数据注意事项”和“Splunk 转发器与 WMI”。

虽然 Active Directory (AD) 监视不使用 WMI，但仍与使用 WMI 的数据导入具有相同的验证注意事项。有关 Splunk 平台如何监视 AD 的信息，请参阅本手册中的“监视 Active Directory”。

通过 WMI 导入数据注意事项

通过 WMI 收集远程 Windows 数据时，请考虑以下注意事项：

- 远程 Windows 数据验证

- 使用受管系统帐户访问 Windows 数据
- 网络和 I/O 使用情况

远程 Windows 数据验证

Windows 需要远程操作验证。如果不了解 Splunk 平台如何通过网络与 Windows 进行交互，则可能会导致不理想的搜索结果或者根本不会获得任何结果。

安装 Splunk 平台时，您可以指定 Splunk 平台以“本地系统”用户或其他用户身份运行。该选项会对安装和数据集合都产生影响。有关更多信息，请参阅*《安装手册》*中的“选择 Splunk Enterprise 应以其身份运行的 Windows 用户”。

您指示 Splunk 平台以其身份运行的用户决定了 Splunk 平台可以从远程计算机上检索的数据类型。要获取您要的数据，您必须为此用户提供相应的权限级别。

大多数情况下，请将 Splunk 平台用户帐户配置为具有您想要收集的数据来源的最小访问权限。这意味着您必须执行以下操作：

- 将用户添加到各种域安全组。
- 根据您需要访问的数据来源修改各种 AD 对象的访问控制列表。

如果您的 AD 域安全策略定期执行密码更改，您还必须执行以下操作：

- 请确保 Splunk 平台的用户密码永不失效，或者您根据密码政策中的定义，在其失效前手动更改密码。
- 在更改密码之后，请重新启动在您网络中的所有主机上以该帐户运行的 Splunk 服务。

另外，您还必须在“本地安全策略”中为此帐户指定“拒绝本地登录”用户权限分配，以防止用户交互登录工作站。与授予域管理员访问权限相比，此方法为您提供更多的控制同时也更安全。

有关更多信息，请参阅本章中涉及到远程访问 Windows 计算机的其他主题。请查看“安全和远程访问权限注意事项”部分，了解如何将 Splunk 平台以其身份运行的用户配置为最小访问权限。

网络和 I/O 使用情况

密切监视网络带宽使用情况，尤其是在 WAN 链路缓慢或较为薄弱的网络中。仅出于这个原因考量，通用转发器是相较于大规模远程数据集合操作而言更好的选择。

还应考虑磁盘带宽问题。无论安装类型如何，防病毒程序驱动程序以及介于 Splunk 平台和操作系统之间的驱动程序应始终配置为忽略 Splunk 平台目录和进程。

Splunk 转发器与 WMI

使用通用转发器从远程 Windows 主机上导入数据。通用转发器提供的数据来源类型最多，提供更多详细数据（例如，在性能监视指标中），可最大程度地减少网络开销，同时减低操作风险和复杂性。另外，在许多情况下，通用转发器比 WMI 更具可扩展性。

在您远程收集数据的情况下（如当企业或安全策略限制代码安装或存在性能或互操作性问题时），您可使用本机 WMI 界面收集事件日志和性能数据。

WMI 和转发器主要有以下几个方面的权衡：

- 性能
- 部署
- 管理

性能

就性能而言，在以下情况下最好选择使用转发器：

方案	注意事项
收集本地事件日志或平面文件。	转发器需要的 CPU 较少，并会预先对数据进行基本压缩，以便减少网络开销。
您希望从某一计算机收集数据而不必担心验证问题。	当您将转发器安装为“本地系统”用户时，转发器将获得访问计算机的管理员权限，让您可以从计算机上收集任何数据。
您希望从繁忙主机（例如，AD 域控制器或始终存在高利用率时段的计算机，如 Exchange、SQL Server、Oracle、VMWare、Hyper-V 或 SharePoint 服务器）收集数据。	在这种情况下考虑使用转发器，原因是 WMI 可能跟不上这些服务生成的数据量。依据设计，WMI 轮询会尽力运行，同时为了防止意外的拒绝服务攻击，Splunk 平台还会对 WMI 调用加以限制。

您担心 CPU 和网络使用。	转发器会尽量少地使用这些资源，而 WMI 会使用更多的 CPU 和网络资源以传输数据。
您担心可扩展性。	通用转发器的扩展性非常好。重型转发器的扩展性不如通用转发器那样好，但是两种类型转发器的扩展性都要比 WMI 好很多。

如果您担心高内存利用率系统上的内存使用情况，WMI 是一个更好的选择。因为转发器提供了更多的轮询选项，并且转发器在收集数据时驻留在本地计算机上，所以转发器所需的内存比 WMI 多。

部署

在以下情况下，最好选择使用转发器部署：

- 您有权控制操作系统的基本版本，这与新建系统映像的情况类似。
- 您有很多数据来源要收集，收集的数据需要进行任意类型的转换时更是如此。

除了少数情况下，您无法在通用转发器到达索引器前使用其处理数据。如果在新建索引之前需要对数据进行任何更改，您必须使用重型转发器。

在以下情况下，最好选择使用 WMI：

- 您在要从中收集数据的计算机上无权控制基本操作系统版本，或者没有域管理员访问权限或本地管理员权限。
- 您希望或只需要从大量主机中收集一组有限的数据（例如，用于使用情况计费的 CPU 数据）。

常见的部署方案是，先使用远程轮询进行测试，然后再于稍后或您大规模部署转发器时，将成功或有用的数据导入添加到您的转发器配置中。

管理

两种机制均提供日志记录和告警功能，方便您了解主机是处于联机状态、脱机状态还是无法连接状态。为防止意外的服务攻击拒绝，Splunk 平台中的 WMI 轮询服务若无法联系主机将降低一段时间内的轮询频率，最终完全停止轮询无法连接的主机。

请不要对经常脱机的计算机（例如笔记本电脑或动态置备的虚拟机）通过 WMI 执行远程轮询。

下表显示了一个数据来源列表，并指出了适用于每种数据来源的数据集合类型：

数据来源	本地转发器	WMI
事件日志	是	是*
性能	是	是
注册表	是	否
Active Directory	是	否
日志文件	是	是**
抓取	是	否

* 如要收集远程事件日志，您必须知道想要集合的事件日志的名称。在本地转发器上，您可以选择收集所有日志，而不考虑名称为何。

** Splunk 平台支持使用 \\SERVERNAME\SHARE 语法进行远程日志文件收集。但是，您必须使用 Common Internet File System (CIFS) 或 Server Message Block (SMB) 作为应用程序层文件访问协议，并且 Splunk 平台必须对共享文件系统和基础文件系统都至少具有读取访问权限。

在 Splunk 平台的非 Windows 实例上搜索 Windows 数据

您可以在非 Windows Splunk 部署上为 Windows 数据建立索引并执行搜索，但您首先必须使用 Splunk 平台的 Windows 实例来获取 Windows 数据。您可以执行此操作，只需将 Splunk 转发器安装到 Windows 计算机上并将其配置为将 Windows 数据转发到 Splunk 平台的非 Windows 实例即可。

您可以通过以下方式之一继续操作：

- 在您想要从中收集数据的所有 Windows 计算机上本地设置转发器。这些转发器可以把 Windows 数据发送到非 Windows 接收实例。
- 在一个单独 Windows 计算机上设置转发器。该转发器可以使用 WMI 从环境中的所有 Windows 计算机收集数据，然后将

合并后的数据转发到 Splunk 平台的非 Windows 接收实例。

监视 Active Directory

Active Directory（AD）数据库（又称 NT Directory Service（NTDS）数据库）是 Windows AD 域或林中用户、计算机、网络、设备和安全对象的中央存储库。您可以使用 Splunk Enterprise 来记录针对 AD 所做的更改，如添加或移除用户、主机，或域控制器（DA）。

如果您使用的是 Splunk Cloud，则必须使用通用转发器从 Windows 域控制器或成员计算机收集 Active Directory 数据并将该数据转发到 Splunk Cloud。在 Splunk Enterprise 上，也可以使用通用转发器，或者可以将 Splunk Enterprise 直接安装到 Windows 计算机上并以这种方式收集 AD 数据。

您可以通过配置 AD 监视来观察 Active Directory 林发生的更改并收集用户和计算机元数据。可以将此功能与动态列表查找功能相结合，使用 AD 中提供的任何信息来修饰或修改事件。请参阅*《知识管理器手册》*中的“关于查找”。

在您通过配置 Splunk Enterprise 来监视您的 Active Directory 后，它会为 AD 架构拍摄基线快照。Splunk Enterpris 使用此快照为要监视的内容建立一个起始点。

AD 监视输入以单独的进程形式运行，该进程名为 `splunk-admon.exe`。Splunk Enterprise 中定义的每个 Active Directory 监视输入都会运行一次该进程。

监视 Active Directory 的原因

如果您要维持 Active Directory 的完整性、安全性和正常运行，那就必须了解它每天的状况。使用 Splunk Enterprise，您可以监视 AD 发生了哪些更改，以及何人在何时进行了这些更改。

您可以将此数据转换为，例如，企业安全合规性或取证报表。还可以使用检索到的入侵告警数据立即进行响应。另外，您可以利用为未来 AD 基础结构规划活动（例如，各域控制器（DC）上操作主服务器角色的分配、AD 副本和全局目录）建立索引时所用的数据新建运行状况报表。

要求

您必须满足下列要求，才能监视 Active Directory 架构：

- Splunk Enterprise 必须在 Windows 上运行。请参阅*《安装手册》*中的“在 Windows 上安装”。
- Splunk Enterprise 必须以域用户身份运行。请参阅*《安装手册》*中的“选择 Splunk Enterprise 应以其身份运行的 Windows 用户”。
- 运行 Splunk Enterprise 的用户一定对您想要监视的所有 AD 对象有读取权限。

监视 Active Directory 的技术注意事项

如想在监视 AD 时获得最佳结果，请注意以下注意事项：

- Splunk 平台的 Windows 版本才有 AD 监视器。Splunk Cloud 无法直接监视 AD。
- 如果无法通过 Splunk Enterprise 的 *nix 版本监视 AD 更改，您可以将 AD 数据从 Splunk Enterprise 的 Windows 版本或通用转发器转发到 *nix 索引器。
- AD 监视进程可以在完整的实例下或任何类型的转发器中运行。
- 监视 AD 更改的主机必须属于您要监视的域或林。
- Splunk Enterprise 以其身份运行的用户也必须属于此域。
- 用户所具有的权限决定了 Splunk 可以监视的 AD 部分。

在安装时决定 Splunk Enterprise 应以何种用户身份运行的相关信息，请参阅*《安装手册》*中的“选择 Splunk Enterprise 应以其身份运行的 Windows 用户”。

AD 监视器如何与 AD 交互

设置 AD 监视输入之后，输入会连接到 AD 域控制器进行验证，如有必要，在收集 AD 架构或更改事件时进行安全 ID（SID）转换。

安装后 AD 监视会使用以下逻辑与 Active Directory 交互：

1. 如果您在定义输入时指定域控制器，那么输入将使用域控制器进行 AD 操作。您可以使用 inputs.conf 中的 `targetDc` 设置或 Splunk Web 中的 `Target domain controller` 字段。
2. 如果您不指定域控制器，那么输入会执行以下操作：
    1. 输入会尝试使用本地系统缓存验证或解决 SID。
    2. 如果监视器无法用该方式验证或解决 SID，它会尝试连接到域控制器（运行输入的计算机曾登录过该控制器）。

3. 如果还是不行，那么输入尝试使用最近的有全局目录副本的 AD 域收集器。
3. 如果您指定的域控制器无效，或无法找到域控制器，那么输入会生成一个错误消息。

AD 监视器不会查找 LDAP 参照

如果 AD 监视器发起轻型目录访问协议 (LDAP) 查询并收到参照，监视器不会查询该参照以完成查询。每个 LDAP 参照代表一个 LDAP 配置问题，您或您的指定管理员必须确定并修复 AD 内的配置问题。

使用 Splunk Web 配置 Active Directory 监视

您可以在 Splunk Web 中或通过编辑配置文件来配置 AD 监视。使用配置文件时，您可以访问更多选项，例如，可以为多个 DC 配置监视器。要通过编辑配置文件来配置 AD 监视，请参阅本主题后面的“使用配置文件配置 Active Directory 监视”部分。

按照以下高级步骤使用 Splunk Web 配置 AD 监视：

1. 转到“添加数据”页面。
2. 选择输入来源。
3. 指定输入设置。
4. 查看您的选择。

转到“添加数据”页面

要通过 Splunk 设置转到“添加数据”页面，请执行以下步骤：

1. 单击 Splunk Web 右上角的**设置**。
2. 请单击**数据导入**。
3. 单击 **Active Directory 监视**。
4. 单击**新建**以添加输入。

要使用 Splunk Web 主页转到“添加数据”页面，请执行以下步骤：

1. 单击 Splunk Web 主页中的**添加数据**。
2. 单击**监视**以监视本地 Windows 计算机的 Active Directory。

选择输入来源

1. 单击 **Active Directory 监视**。
2. 在**集合名称**字段中，为该输入键入一个您可以记住的唯一名称。
3. （可选）在**目标域控制器**字段中输入您要用于监视 AD 的域控制器的主机名或 IP 地址。
4. （可选）在**开始节点**字段中键入您想要输入从其开始监视的 Active Directory 节点。请使用轻型目录访问协议 (LDAP) 格式，例如 `DC=Splunk-Docs,DC=com`。
5. （可选）单击**浏览**按钮来浏览可用的 Active Directory 节点列表，以便浏览可用的 AD 域列表。
6. 如果您想要输入监视在**开始节点**字段中输入的节点的所有子节点，请勾选**监视子树**。
7. 单击**下一步**。

指定输入设置

在“输入设置”页面指定应用程序上下文、默认主机值和索引。所有这些参数均为可选参数。

主机只是设置生成事件中的**主机**字段。它不会指示输入去查找您网络上的特定主机。

1. 为此输入选择相应的**应用程序上下文**。
2. 设置**主机**名称。此设置有多个选项供您选择。有关设置主机值的更多信息，请参阅“关于主机”。
3. 设置 Splunk Enterprise 将数据发送到其中的**索引**。如果未定义多个索引来处理不同类型的事件，请保留**默认**值。除了用户数据的索引之外，Splunk Enterprise 还有很多实用工具索引，这些索引也会显示在此下拉列表中。
4. 请单击**查看**。

查看您的选择

在您指定所有输入设置后，可查看您的选择。Splunk Enterprise 列出所有您所选的选项，包括监视器的类型、数据来源、来源类型、应用程序上下文和索引。

1. 查看该设置。
2. 如果它们不符合您的期望，请单击左尖括号（ < ）即可返回到上一个步骤。否则，请单击**提交**。

成功页面显示且 Splunk Enterprise 开始对 Active Directory 节点建立索引。

使用配置文件配置 Active Directory 监视

您可以在 Splunk Web 中或通过编辑配置文件来配置 AD 监视。使用配置文件时，您可以访问更多选项，例如，可以为多个 DC 配置监视器。

inputs.conf 配置文件控制 Active Directory 监视配置。编辑 %SPLUNK_HOME%\etc\system\local 目录中 inputs.conf 的副本。如果您在默认目录下编辑这些副本，您所做的更改将在升级软件时全部被覆盖。有关配置文件优先顺序的更多信息，请参阅“配置文件优先顺序”。

1. 打开 %SPLUNK_HOME%\etc\system\local\inputs.conf 进行编辑。如果此文件不存在，请新建文件。
2. 添加合适的 AD 监视段落和设置。

默认情况下，当您启用 AD 监视输入时，Splunk Enterprise 会从其可以附加到的第一个域控制器中收集 AD 更改数据。如果可以接受这一点，则无需进行任何其他配置。

inputs.conf 设置

inputs.conf 包含每个 AD 监视输入的一个段落，其中的标题类似如下所示：

```
[admon://<name of stanza>]
```

在每个段落中，您可以指定以下设置：

设置	是否必需？	描述	默认
targetDc	是	您想用于 AD 监视的域控制器的唯一名称。 在以下情况中，请为该设置指定唯一名称： • 您的 AD 非常大，并且您只希望监视特定组织单元 (OU)、子域等的信息。 • 您的特定只读域控制器可用于高安全性环境中的监视目的。 • 您有多个域或林建立了传递信任关系，并且您希望将除运行 Splunk Enterprise 的主机所在树以外的其他某个树设为目标。 • 您希望将多个 AD 监视输入配置为以多个域控制器为目标。例如，跨分布式环境监视 AD 复制。 要以多个 DC 为目标，请为该树中的目标再添加一个 [admon://<uniquename>targetDc] 段落。	n/a
startingNode	否	完全限定轻型目录访问协议 (LDAP) 名称（例如："LDAP://OU=Computers,DC=ad,DC=splunk,DC=com"）用于指定 Splunk Enterprise 在 AD 树中新建索引的起始位置。根据 monitorSubtree 设置的配置情况，该软件将从此处开始向下枚举到子容器。 为获得 AD 数据，startingNode 的值必须落在 Splunk Enterprise 目标 DC 的范围之内。	Splunk Enterprise 可以访问的树中最高根域
monitorSubtree	否	要建立索引的目标 AD 容器的范围。值为 0 时表示只为目标容器建立索引，而不遍历该容器内的子容器。值为 1 时表示枚举其有权访问的所有子容器和域。	1（监视 Splunk Enterprise 具有访问权限的所有域）
baseline	否	输入第一次运行时是否枚举所有现有可用 AD 对象。值为 0 时表示不设置基线。值为 1 时表示设置基线。	1（设置基线）
index	否	要将 AD 监视数据发送到的索引。	default 索引
disabled	否	Splunk Enterprise 是否应运行该输入。值为 0 时表示启用输入；值为 1 则表示禁用输入。	0（启用）

AD 监视配置示例

以下示例显示了如何使用 inputs.conf 文件来监视所需的 AD 网络部分。

请参阅以下示例以从 AD 目录顶部开始建立数据索引：

```
#Gather all AD data that this server can see
```

```
[admon://NearestDC]
targetDc =
startingNode =
```

请参阅以下示例以使用其所在根级别高于目标 OU 的 DC 来进行监视：

```
# Use the pri01.eng.ad.splunk.com domain controller to get all AD metadata for
# the Computers OU in this forest. We want schema data for the entire AD tree, not
# just this node.

[admon://DefaultTargetDc]
targetDc = pri01.eng.ad.splunk.com
startingNode = OU=Computers,DC=eng,DC=ad,DC=splunk,DC=com
```

请参阅以下示例以监视多个域控制器：

```
# Get change data from two domain controllers (pri01 and pri02) in the same AD tree.
# Index both and compare/contrast to ensure AD replication is occurring properly.

[admon://DefaultTargetDc]
targetDc = pri01.eng.ad.splunk.com
startingNode = OU=Computers,DC=eng,DC=ad,DC=splunk,DC=com

[admon://SecondTargetDc]
targetDc = pri02.eng.ad.splunk.com
startingNode = OU=Computers,DC=eng,DC=ad,DC=splunk,DC=com
```

AD 监视输出示例

AD 监视实用工具运行时，该工具会收集 AD 更改事件，供 Splunk 平台稍后为其建立索引。如需在更改事件到达时查看这些事件，请使用“搜索和报表”应用。

Splunk 平台可以为多种类型的 AD 更改事件建立索引。这些传入事件的示例。出于发行目的，这些事件的部分内容已经被遮盖/修改。

更新事件

某一 AD 对象发生更改后，Splunk 平台即会生成此一个更新事件。该软件将把此更改记录为 `admonEventType=Update` 类型。

```
2/1/10
3:17:18.009 PM

02/01/2010 15:17:18.0099
dcName=stuff.splunk.com
admonEventType=Update
Names:
                objectCategory=CN=Computer,CN=Schema,CN=Configuration
                name=stuff2
                displayName=stuff2
                distinguishedName=CN=stuff2,CN=Computers
Object Details:
                sAMAccountType=805306369
                sAMAccountName=stuff2
                logonCount=4216
                accountExpires=9223372036854775807
                objectSid=S-1-5-21-3436176729-1841096389-3700143990-1190
                primaryGroupID=515
                pwdLastSet=06:30:13 pm, Sat 11/27/2010
                lastLogon=06:19:43 am, Sun 11/28/2010
                lastLogoff=0
                badPasswordTime=0
                countryCode=0
                codePage=0
                badPwdCount=0
                userAccountControl=4096
```

```
                objectGUID=blah
                whenChanged=01:02.11 am, Thu 01/28/2010
                whenCreated=05:29.50 pm, Tue 11/25/2008
                objectClass=top|person|organizationalPerson|user|computer
Event Details:
                uSNChanged=2921916
                uSNCreated=1679623
                instanceType=4
Additional Details:
                isCriticalSystemObject=FALSE
                servicePrincipalName=TERMSRV/stuff2|TERMSRV blah
                dNSHostName=stuff2.splunk.com
                operatingSystemServicePack=Service Pack 2
                operatingSystemVersion=6.0 (6002)
                operatingSystem=Windows Vista? Ultimate
localPolicyFlags=0
```

删除事件

某一 AD 对象被标记为删除后，Splunk 平台即会生成一个删除事件。该事件类型与 `admonEventType=Update` 相似，不同点在于此事件类型的末尾包含 `isDeleted=True` 键/值对。

```
2/1/10
3:11:16.095 PM

02/01/2010 15:11:16.0954
dcName=stuff.splunk.com
admonEventType=Update
Names:
                name=SplunkTest
DEL:blah
                distinguishedName=OU=SplunkTest\0ADEL:blah,CN=Deleted Objects
DEL:blah
Object Details:
                objectGUID=blah
                whenChanged=11:31.13 pm, Thu 01/28/2010
                whenCreated=11:27.12 pm, Thu 01/28/2010
                objectClass=top|organizationalUnit
Event Details:
                uSNChanged=2922895
                uSNCreated=2922846
                instanceType=4
Additional Details:
                dSCorePropagationData=20100128233113.0Z|20100128233113.0Z|20100128233113.0Z|16 010108151056.0Z
                lastKnownParent=stuff
                '''isDeleted=TRUE'''
```

同步事件

配置好 AD 监视输入后，Splunk 平台会尝试在启动时捕获 AD 元数据的基线。Splunk 平台将生成 `admonEventType=Sync` 事件类型，该事件类型代表某一 AD 对象的实例及其所有字段值。Splunk 平台会尝试从上次记录的更新序列号 (USN) 捕获所有对象。

重启 Splunk Enterprise 或 `splunk-admon.exe` 进程时，该软件将记录额外的 `sync` 事件。这是正常的。

```
2/1/10
3:11:09.074 PM

02/01/2010 15:11:09.0748
dcName=ftw.ad.splunk.com
admonEventType=Sync
Names:
                name=NTDS Settings
                distinguishedName=CN=NTDS Settings,CN=stuff,CN=Servers,CN=Default-First-Site-Name,CN=Sites,CN=Configuration
```

```
                cn=NTDS Settings
                objectCategory=CN=NTDS-DSA,CN=Schema,CN=Configuration,DC=ad,DC=splunk,DC=com
                fullPath=LDAP://stuff.splunk.com/<GUID=bla bla bla>
                CN=NTDS Settings
Object Details:
                whenCreated=10:15.04 pm, Tue 02/12/2008
                whenChanged=10:23.00 pm, Tue 02/12/2008
                objectGUID=bla bla bla
                objectClass=top|applicationSettings|nTDSDSA
                classPath=nTDSDSA
Event Details:
                instanceType=4
Additional Details:
                systemFlags=33554432
                showInAdvancedViewOnly=TRUE
                serverReferenceBL=CN=stuff,CN=Domain System Volume (SYSVOL share),CN=File Replication Service,CN=System
                options=1
                msDS -hasMasterNCs=DC=ForestDnsZones|DC=DomainDnsZones|CN=Schema,CN=Configuration|CN=Configuration
                msDS-HasInstantiatedNCs=
                msDS-HasDomainNCs=blah
                msDS-Behavior-Version=2
                invocationId=bla bla bla
                hasMasterNCs=CN=Schema,CN=Configuration|CN=Configuration
                dSCorePropagationData=
                dMDLocation=CN=Schema,CN=Configuration
                nTSecurityDescriptor=NT AUTHORITY\Authenticated Users
SchemaName=LDAP://stuff.splunk.com/schema/nTDSDSA
```

Schema 事件

您将 Splunk Enterprise 配置为 AD 监视并重新启动后，Splunk Enterprise 会生成一个 Schema 类型事件：`admonEventType=schema`。此事件显示 Active Directory 结构中每个对象的定义。为每个 AD 对象列出了 `available`、`required` 和 `optional` 字段。如果无法查看所有这些字段，这可能指示 Active Directory 存在问题。

```
02/01/2010 15:11:16.0518
dcName=LDAP://stuff.splunk.com/
admonEventType=schema
className=msExchProtocolCfgSMTPIPAddress
classCN=ms-Exch-Protocol-Cfg-SMTP-IP-Address
instanceType=MandatoryProperties
nTSecurityDescriptor=MandatoryProperties
objectCategory=MandatoryProperties
objectClass=MandatoryProperties
adminDescription=OptionalProperties
adminDisplayName=OptionalProperties
allowedAttributes=OptionalProperties
allowedAttributesEffective=OptionalProperties
allowedChildClasses=OptionalProperties
allowedChildClassesEffective=OptionalProperties
bridgeheadServerListBL=OptionalProperties
canonicalName=OptionalProperties
cn=OptionalProperties
createTimeStamp=OptionalProperties
description=OptionalProperties
directReports=OptionalProperties
displayName=OptionalProperties
displayNamePrintable=OptionalProperties
distinguishedName=OptionalProperties
dSASignature=OptionalProperties
dSCorePropagationData=OptionalProperties
extensionName=OptionalProperties
flags=OptionalProperties
fromEntry=OptionalProperties
frsComputerReferenceBL=OptionalProperties
fRSMemberReferenceBL=OptionalProperties
fSMORoleOwner=OptionalProperties
heuristics=OptionalProperties
isCriticalSystemObject=OptionalProperties
```

```
isDeleted=OptionalProperties
isPrivilegeHolder=OptionalProperties
lastKnownParent=OptionalProperties
legacyExchangeDN=OptionalProperties
managedObjects=OptionalProperties
masteredBy=OptionalProperties
memberOf=OptionalProperties
modifyTimeStamp=OptionalProperties
mS-DS-ConsistencyChildCount=OptionalProperties
mS-DS-ConsistencyGuid=OptionalProperties
msCOM-PartitionSetLink=OptionalProperties
msCOM-UserLink=OptionalProperties
msDFSR-ComputerReferenceBL=OptionalProperties
msDFSR-MemberReferenceBL=OptionalProperties
msDS-Approx-Immed-Subordinates=OptionalProperties
msDs-masteredBy=OptionalProperties
msDS-MembersForAzRoleBL=OptionalProperties
msDS-NCReplCursors=OptionalProperties
msDS-NCReplInboundNeighbors=OptionalProperties
msDS-NCReplOutboundNeighbors=OptionalProperties
msDS-NonMembersBL=OptionalProperties
msDS-ObjectReferenceBL=OptionalProperties
msDS-OperationsForAzRoleBL=OptionalProperties
msDS-OperationsForAzTaskBL=OptionalProperties
msDS-ReplAttributeMetaData=OptionalProperties
msDS-ReplValueMetaData=OptionalProperties
msDS-TasksForAzRoleBL=OptionalProperties
msDS-TasksForAzTaskBL=OptionalProperties
msExchADCGlobalNames=OptionalProperties
msExchALObjectVersion=OptionalProperties
msExchHideFromAddressLists=OptionalProperties
msExchInconsistentState=OptionalProperties
msExchIPAddress=OptionalProperties
msExchTurfList=OptionalProperties
msExchUnmergedAttsPt=OptionalProperties
msExchVersion=OptionalProperties
msSFU30PosixMemberOf=OptionalProperties
name=OptionalProperties
netbootSCPBL=OptionalProperties
nonSecurityMemberBL=OptionalProperties
objectGUID=OptionalProperties
objectVersion=OptionalProperties
otherWellKnownObjects=OptionalProperties
ownerBL=OptionalProperties
partialAttributeDeletionList=OptionalProperties
partialAttributeSet=OptionalProperties
possibleInferiors=OptionalProperties
proxiedObjectName=OptionalProperties
proxyAddresses=OptionalProperties
queryPolicyBL=OptionalProperties
replicatedObjectVersion=OptionalProperties
replicationSignature=OptionalProperties
replPropertyMetaData=OptionalProperties
replUpToDateVector=OptionalProperties
repsFrom=OptionalProperties
repsTo=OptionalProperties
revision=OptionalProperties
sDRightsEffective=OptionalProperties
serverReferenceBL=OptionalProperties
showInAddressBook=OptionalProperties
showInAdvancedViewOnly=OptionalProperties
siteObjectBL=OptionalProperties
structuralObjectClass=OptionalProperties
subRefs=OptionalProperties
subSchemaSubEntry=OptionalProperties
systemFlags=OptionalProperties
unmergedAtts=OptionalProperties
url=OptionalProperties
uSNChanged=OptionalProperties
```

```
uSNCreated=OptionalProperties
uSNDSALastObjRemoved=OptionalProperties
USNIntersite=OptionalProperties
uSNLastObjRem=OptionalProperties
uSNSource=OptionalProperties
wbemPath=OptionalProperties
wellKnownObjects=OptionalProperties
whenChanged=OptionalProperties
whenCreated=OptionalProperties
wWWHomePage=OptionalProperties
```

使用 Splunk Cloud 监视 Windows 事件日志数据

Windows 会在其操作过程中生成日志数据。Windows 事件日志服务可处理此通信的几乎所有方面。它会收集已安装应用程序、服务和系统进程所发布的日志数据，并将日志数据放入事件日志通道中。Microsoft 事件查看器等程序订阅这些日志通道以显示系统上发生的事件。

您可以监视存储在本地计算机上的事件日志通道和文件，也可以从远程计算机收集日志。事件日志监视器将针对您定义的每个事件日志输入运行一次。

要监视 Splunk Cloud 中的 Windows 事件日志通道，请使用 Splunk 通用或重型转发器来收集数据并将其转发到您的 Splunk Cloud 部署。最佳做法是，使用 Splunk Add-on for Windows 来简化将数据导入 Splunk Cloud 的过程。有关使用 Splunk Add-on for Windows 将数据导入 Splunk Cloud 的说明，请参阅*《Splunk Cloud 管理员手册》*中的“将 Windows 数据导入 Splunk Cloud”。

为什么监视事件日志？

Windows 事件日志是 Windows 计算机操作的核心指标。如果您的 Windows 系统出现了问题，则“事件日志”服务已记录这一情况。Splunk 平台的索引、搜索和报告功能使您的日志可以访问。

监视事件日志的要求

活动	要求
监视本地事件日志	• Splunk 通用转发器或 Splunk Enterprise 实例必须在 Windows 上运行。请参阅*《安装手册》*中的“在 Windows 上安装”。 • 为了读取所有的本地事件日志，Splunk 通用转发器或 Splunk Enterprise 实例必须以“本地系统”Windows 用户身份运行。
监视远程事件日志	• 通用转发器或重型转发器必须在您想从中收集事件日志的 Windows 计算机上运行。 • Splunk 通用转发器或重型转发器必须以域或远程用户的身份运行，该域或远程用户具有读取访问远程计算机上 Windows Management Instrumentation (WMI) 的权限。请参阅*《安装手册》*中的“选择 Splunk Enterprise 应以其身份运行的 Windows 用户”。 • 以用户身份运行通用转发器必须具有读取访问您想要收集的事件日志的权限。

从远程计算机收集事件日志数据的安全性和其他注意事项

您可以使用通用转发器、重型转发器或 WMI 从远程计算机收集事件日志数据。最佳做法是，使用通用转发器将远程计算机中的事件日志数据发送到索引器。有关如何安装、配置和使用转发器来收集事件日志数据的信息，请查看*《通用转发器手册》*中的“通用转发器”。如果您无法在要获取数据的计算机上安装转发器，则可以使用 WMI。

要将转发器安装在要收集事件日志数据的远程计算机上，请将转发器作为“本地系统”用户安装在这些计算机上。本地系统用户对本地计算机上的所有数据都具有访问权限，但对远程计算机上的数据没有访问权限。

要使用 WMI 从远程计算机获取事件日志数据，则您必须确保您的网络和 Splunk Enterprise 实例配置正确。不要以“本地系统”用户身份安装 Splunk 软件。用于安装软件的用户决定了 Splunk 软件有权访问的事件日志。若想使用 WMI 正确收集远程数据，必须满足一些要求。有关这些要求的更多信息，请参阅“安全与远程访问注意事项”。

默认情况下，Windows 限制某些事件日志的访问权限，这取决于您所运行的 Windows 版本。例如，默认情况下，只有本地管理员组或全局域管理员组的成员才能读取安全事件日志。

Windows 事件日志监视器如何与 Active Directory 交互

设置 WMI 事件日志监视输入之后，输入会连接到 Active Directory (AD) 域控制器进行验证，如有必要，在开始监视数据前进行任何安全 ID (SID) 转换。

设置后事件日志监视器将使用以下逻辑与 AD 交互：

1. 如果您在定义输入时（利用 inputs.conf 文件中的 evt_dc_name 设置）指定域控制器，那么输入将使用该域控制器进行 AD 操作。
2. 如果您不指定域控制器，那么输入会执行以下操作：
    1. 输入会尝试使用本地系统缓存验证或解决 SID。
    2. 如果监视器无法用该方式验证或解决 SID，它会尝试连接到域控制器（运行输入的计算机曾登录过该控制器）。
    3. 如果还是不行，那么输入尝试使用最近的有全局目录副本的 AD 域收集器。
3. 如果您指定的域控制器无效，或无法找到域控制器，那么输入会生成一个错误消息。

从远程 Windows 计算机收集事件日志

从远程 Windows 计算机上收集数据的两种方法：

- 使用通用转发器
- 使用 WMI

使用通用或重型转发器

您可以在 Windows 计算机上安装一个通用转发器或重型转发器，然后指示该转发器收集事件日志。您可以手动执行此操作，也可以使用部署服务器来管理转发器配置。

1. 通过您想从中收集 Windows 事件日志的 Windows 计算机下载 Splunk Enterprise 或通用转发器软件。
2. 运行通用转发器安装软件包开始安装进程。
3. 根据安装程序的提示配置接收索引器。
4. 安装程序提示您指定输入时，勾选**事件日志**复选框即可启用事件日志输入。
5. 完成安装过程。
6. 在接收索引器上使用 Splunk Web 搜索事件日志数据，如以下示例所示：

```
host=<name of remote Windows machine> sourcetype=Wineventlog
```

安装通用转发器的具体指示，请参阅*《转发器手册》*中的“通过安装程序安装 Windows 通用转发器”。

使用 WMI

如果要使用 WMI 远程收集事件日志，则必须以 Active Directory 域用户身份安装通用或重型转发器。如果选定域用户不是管理员组或域管理员组的成员，您必须将事件日志安全配置为授予域用户对事件日志的访问权限。

要将事件日志安全更改为授予对远程计算机中事件日志的访问权限，您必须满足以下要求：

- 对想从中收集事件日志的计算机具有管理员访问权限。
- 了解安全描述定义语言（SDDL）的工作原理以及如何使用它来分配权限。有关更多信息，请参阅 Microsoft 网站中的 http://msdn.microsoft.com/en-us/library/aa379567(v=VS.85).aspx。
- 决定如何监视您的数据。有关从远程 Windows 计算机收集数据的信息，请参阅“确定如何监视远程 Windows 数据的注意事项”。

您可以使用 wevtutil 实用工具设置事件日志安全性。

1. 将 Splunk Enterprise 实例下载到 Windows 计算机上。
2. 双击安装程序文件开始安装。
3. 安装程序提示您指定用户时，请选择**域用户**。
4. 在下一个安装程序窗格中，输入您希望 Splunk Enterprise 在运行时使用的域用户名和密码。
5. 根据提示完成软件的安装。
6. 软件安装完成后即登入此实例。
7. 使用 Splunk Web 添加远程事件日志输入。请参阅本主题后面的“配置远程事件日志监视”。

一些系统事件日志中显示异常计算机名

在某些 Windows 系统上，您可能会看到一些带有随机生成的计算机名称的事件日志。这是在操作系统安装过程中，用户命名系统之前的那些系统日志记录事件所导致的结果。

该异常情况仅当您通过 WMI 从 Windows 版本远程收集日志时才会发生。

使用 Splunk Web 配置本地事件日志监视

如需获取本地 Windows 事件日志数据，请将 Splunk Enterprise 实例指向事件日志服务。

转到“添加数据”页面

可通过两种方式访问此页面：

- Splunk 设置
- Splunk 主页

通过 Splunk 设置：

1. 单击**设置** > **数据导入**。
2. 单击**本地事件日志集合**。
3. 单击**新建**以添加输入。

通过 Splunk 主页：

1. 单击 Splunk 主页中的**添加数据**链接。
2. 单击**监视**以监视本地 Windows 计算机上的“事件日志”数据或**转发**以转发来自另一个 Windows 计算机的“事件日志”数据。
   Splunk Enterprise 将加载“添加数据 - 选择来源”页面。
3. 如果您选择了**转发**，则选择或新建要此输入应用的转发器组。请参阅本手册中的“转发数据”。
4. 单击**下一步**。

选择输入来源

1. 选择**本地事件日志**
2. 在**选择事件日志**列表中，选择您要此输入监视的“事件日志”通道。
3. 单击要监视的每个“事件日志”通道。
   Splunk Enterprise 将通道从**可用项目**窗口移动到**已选项目**窗口。
4. 要取消选择一个通道，在**可用项目**窗口上单击该通道的名称。
   Splunk Enterprise 将通道从**已选项目**窗口移动到**可用项目**窗口。
5. 要选择或取消选择所有事件日志，单击**添加所有**或**删除所有**链接。

   选择所有通道可导致索引大量数据。

6. 单击**下一步**。

指定输入设置

输入设置页面允许您指定应用程序上下文、默认主机值和索引。所有这些参数均为可选参数。

主机字段只是设置生成事件中的**主机**字段。不会指示 Splunk Enterprise 查找网络中的特定计算机。

1. 为此输入选择相应的**应用程序上下文**。
2. 设置**主机**值。此设置有多个选项供您选择。有关设置主机值的更多信息，请参阅“关于主机”。
3. 设置您希望 Splunk Enterprise 将数据发送到其中的**索引**。如果未定义多个索引来处理不同类型的事件，请保留**默认**值。除了用户数据的索引之外，Splunk Enterprise 还有许多实用工具索引，这些索引也会显示在此下拉框中。
4. 请单击**查看**。

查看您的选择

在指定您的所有输入设置后，您可查看您的选择。Splunk Enterprise 列出所有您所选的选项，包括监视器的类型、数据来源、来源类型、应用程序上下文和索引。

1. 查看该设置。
2. 如果它们不符合您的期望，请单击左尖括号（ < ）即可返回到向导中的上一个步骤。否则，请单击**提交**。

然后，Splunk Enterprise 会显示“成功”页面并开始索引指定的“事件日志”通道。

使用 Splunk Web 配置远程事件日志监视

配置远程事件日志监视的过程几乎和监视本地事件日志的过程相同。

1. 请遵照说明转到“添加数据”页面。请参阅“转到‘添加数据’页面”。
2. 查找并选择**远程事件日志**。
3. 在**事件日志集合名称**字段中，为该输入输入一个可以记住的唯一名称。
4. 在**选择此主机的日志**字段中，输入包含要监视的“事件日志”通道的计算机的主机名或 IP 地址。

   选择所有事件日志通道可导致索引大量数据。

5. 单击**查找日志**按钮即可刷新具有可用“事件日志”通道列表的页面，这些通道位于您输入的计算机上。
6. 单击要监视的每个“事件日志”通道。
   Splunk Enterprise 将通道从**可用项目**窗口移动到**已选项目**窗口。

7. 要取消选择一个通道，在**可用项目**窗口上单击该通道的名称。
   Splunk Enterprise 将通道从**已选项目**窗口移动到**可用项目**窗口。
8. 要选择或取消选择所有事件日志，单击**添加所有**或**删除所有**链接。
9. 在**从其他主机收集同一组日志**字段中，输入包含您先前所选的“事件日志”的其他主机的主机名或 IP 地址。使用逗号分隔多个计算机。
10. 单击绿色**下一步**按钮。
11. 请遵循说明指定输入设置。请参阅“指定输入设置”。
12. 请遵循说明查看您的选择。请参阅“查看您的选择”。

使用 inputs.conf 配置文件配置事件日志监视

在通用或重型转发器上，您可以编辑 input.conf 配置文件以配置 Windows 事件日志监视。

1. 使用 Notepad 或类似的编辑器打开 %SPLUNK_HOME%\etc\system\local\inputs.conf 进行编辑。如果文件不存在，则可能需要创建该文件。
2. 通过添加引用“事件日志”通道的输入段落启用 Windows 时间日志输入。
3. 保存文件并将其关闭。
4. 重新启动 Splunk 平台。

有关使用 inputs.conf 文件配置数据导入的更多信息，请参阅“编辑 inputs.conf”。

为 Windows 事件日志输入指定全局设置

在 inputs.conf 中定义完 Windows 事件日志输入之后，确保您在正确位置明确指定了全局设置。

如果您为 Windows 事件日志输入（如 host、sourcetype 等等）指定了全局设置，您可将这些设置放入以下某个区域：

- 在 [WinEventLog] 全局段落中。此段落等同于其他监视输入的 [default] 段落。例如：

```
[default]
_meta = hf_proxy::meta_test

[WinEventLog]
_meta = hf_proxy::meta_test
host = WIN2K16_DC
index = wineventlog

[WinEventLog://Applications]
disabled = 0
```

- 在您想要监视的事件日志通道的 Windows 事件日志输入段落中。例如：

```
[default]
_meta = hf_proxy::meta_test

[WinEventLog]
host = WIN2K16_DC
index = wineventlog

[WinEventLog://Applications]
disabled = 0
_meta = hf_proxy::meta_test
```

您可以通过查看 %SPLUNK_HOME%\etc\system\default 中的示例或*《管理员手册》*中的规范文件，检查配置文件的默认值。

事件日志监视器配置值

Windows 事件日志 (*.evt) 文件采用二进制格式。您无法像监视普通文本文件一样监视它们。splunkd 服务通过使用适当的 API 读取二进制文件内的数据并为这些数据新建索引，来监视这些文件。

Splunk Enterprise 使用 inputs.conf 中的以下段落来监视默认 Windows 事件日志：

```
# Windows platform specific input processor.
[WinEventLog://Application]
disabled = 0
[WinEventLog://Security]
disabled = 0
[WinEventLog://System]
disabled = 0
```

监视非默认的 Windows 事件日志

您还可以将 Splunk Enterprise 配置为监视非默认 Windows 事件日志。在进行此操作之前，您必须将这些日志导入 Windows 事件查看器。在您导入日志之后，您可以将日志添加到 inputs.conf 的本地副本中，如以下示例所示：

```
[WinEventLog://DNS Server]
disabled = 0
[WinEventLog://Directory Service]
disabled = 0
[WinEventLog://File Replication Service]
disabled = 0
```

使用事件查看器中的“全名”日志属性适当指定复杂的事件日志通道的名称

您可以使用“事件查看器”中的 Full Name 事件日志属性确保自己在 inputs.conf 段落中指定了正确的“事件日志”通道。

例如，若要监视“任务计划程序”应用程序日志 Microsoft-Windows-TaskScheduler-Operational，请执行以下步骤：

1. 打开“事件查看器”。
2. 展开**应用程序和服务日志 > Microsoft > Windows > 任务计划程序**。
3. 右击**操作**，并选择**属性**。
4. 在出现的对话框中复制**全名**字段中的文本。
5. 把该文本附加到 inputs.conf 的 WinEventLog:// 段落中，如以下示例所示：

```
[WinEventLog://Microsoft-Windows-TaskScheduler/Operational]
disabled = 0
```

禁用事件日志段落

要禁用为某事件日志新建索引，请在该事件日志列表之后添加 disabled = 1，该列表位于 %SPLUNK_HOME%\etc\system\local\inputs.conf 中的段落内。

配置用于监视 Windows 事件日志的设置

Splunk 软件使用 inputs.conf 中的下列设置来监视“事件日志”文件：

属性	描述	默认
start_from	如何读取事件。 可接受的值为 oldest（意味着日志读取顺序为从最旧到最新）和 newest（意味着日志读取顺序为从最新到最旧）。 您无法在将该属性设置为 newest 的同时也把 current_only 属性设置为 1。	oldest
current_only	如何为事件新建索引。 可接受值为 1（其中，输入获取在输入首次启动后到达的事件，如 *nix 系统上的 tail -f）或 0（其中，输入首先获取日志中的所有现有事件，然后继续实时监视传入事件）。 您无法在将该属性设置为 newest 的同时也把 current_only 属性设置为 1。	0
checkpointInterval	Windows 事件日志输入保存检查点的频率（以秒为单位）。 检查点会存储已获取事件的 eventID，让 Splunk 软件可以在关机或服务中断后从正确的事件继续执行监视。	0
evt_resolve_ad_ds	Splunk 软件在为 Windows 事件日志通道建立索引时与 Active Directory 交互所用的域控制器。仅在您将 evt_resolve_ad_obj 属性设置为 1 并省略 evt_dc_name 属性时有效。 有效值为 auto（意味着用最近的域控制器来绑定 AD 对象分辨率）或 PDC（意味着将主机位于其内的 AD 站点绑定至主域控制器）。如果您也设置了 evt_dc_name 属性，Splunk 软件将忽略该属性。	auto
	Splunk 软件在为 Windows 事件日志通道建立索引时如何与 Active Directory 交互。有效值为 1（指示将诸如全局唯一标识符 (GUID) 和安全标识符 (SID) 对象的 Active Directory 对象解析为特定 Windows 事件日	

evt_resolve_ad_obj	志通道的权威名称）和 0（指示不要尝试任何解决方案）。 当您将此值设置为 1 时，您可以选择指定要绑定的“域控制器”名称或域的 DNS 名称，Splunk 软件会用该名称解析 AD 对象。如果您未设置此值，Splunk 软件会尝试解析 AD 对象。	0
evt_dc_name	要解析 AD 对象需绑定哪个 Active Directory 域控制器。该名称可以为域控制器的 NetBIOS 名称、域控制器的完全限定 DNS 名称或指定为 $Environment_variable 的环境变量名称。 如果您设置了此属性，则 Splunk 软件会忽略 evt_resolve_ad_ds 属性，该属性控制着 Splunk 软件如何决定把 AD 对象解析绑定至最好的域控制器。 如您指定某环境变量，必须在该环境变量名称前加上美元符号（$）。Splunk 软件把指定的环境变量用作 AD 对象解析要连接的域控制器。例如，若想使用 %LOGONSERVER% 变量，请指定 evt_dc_name = $logonserver。 您可在各个格式前加上两个反斜杠字符。该属性无默认值。	N/A
evt_dns_name	解析 AD 对象所需绑定的域的完全限定 DNS 名称。	N/A
evt_exclude_fields	Windows 事件日志输入在引入 Windows 事件日志数据时要排除的 Windows 事件日志字段的列表。指定此设置后，输入会删除所排除字段的键和值数据。此设置的工作原理与 suppress_* 设置类似，但又有些不同，因为此设置对所有 Windows 事件日志字段均有效，并且排除了您可能包含在允许列表中的字段。发生这种冲突时，实例会记录一个错误。有关可排除的 Windows 事件日志字段的列表，请参阅本主题的后续章节“使用‘白名单’和‘黑名单’创建高级过滤器”。	N/A
suppress_text	是否包含安全事件随附的消息文本。值为 1 时禁止消息文本；值为 0 时保留消息文本。	0
use_old_eventlog_api	是否阅读有事件日志 API 的事件日志事件。 这是一个高级设置。更改前联系 Splunk 支持。 如果设置为 true，输入会使用事件日志 API（而不是 Windows 事件日志 API）从 Windows Server 2008、Windows Vista 和更高的安装版本上的事件日志读取。	false（使用特定于 OS 的 API。）
use_threads	作为对默认写入线程的补充，指定可以新建用来使用允许列表和/拒绝列表正则表达式筛选事件的线程数。 这是一个高级设置。更改前联系 Splunk 支持。 最大线程数为 15。	0
thread_wait_time_msec	发生读取错误后，尝试重新读取事件日志文件的间隔（以毫秒计）。 这是一个高级设置。更改前联系 Splunk 支持。	5000
suppress_checkpoint	保存检查点时，事件日志是否严格遵循 checkpointInterval 设置。 这是一个高级设置。更改前联系 Splunk 支持。 默认情况下，事件日志输入保存零到 checkpointInterval 秒之间的检查点，这取决于传入的事件量。	false
suppress_sourcename	是否将 sourcename 字段从事件中排除。 这是一个高级设置。更改前联系 Splunk 支持。 设置为 true 时，输入会将 sourcename 字段从事件排除，且吞吐量性能（每秒处理的事件数）改善。	false
suppress_keywords	是否将 keywords 字段从事件中排除。 这是一个高级设置。更改前联系 Splunk 支持。 设置为 true 时，输入会将 keywords 字段从事件排除，且吞吐量性能（每秒处理的事件数）改善。	false
suppress_type	是否将 type 字段从事件中排除。 这是一个高级设置。更改前联系 Splunk 支持。 设置为 true 时，输入会将 type 字段从事件排除，且吞吐量性能（每秒处理的事件数）改善。	false

suppress_task	是否将 task 字段从事件中排除。 这是一个高级设置。更改前联系 Splunk 支持。 设置为 true 时，输入会将 task 字段从事件排除，且吞吐量性能（每秒处理的事件数）改善。	false
suppress_opcode	是否将 opcode 字段从事件中排除。 这是一个高级设置。更改前联系 Splunk 支持。 设置为 true 时，输入会将 opcode 字段从事件排除，且吞吐量性能（每秒处理的事件数）改善。	false
whitelist	是否要为与指定文本字符串相匹配的事件建立索引。这是可选属性。 您可指定两种格式中的一种： • 一个或多个“事件日志”事件代码或事件 ID（事件代码/ID 格式）。 • 一个或多个密钥和正则表达式集（高级筛选格式）。 您不可以在单个条目中混合格式。您也不可以在相同段落中混合格式。 Splunk 平台会先处理允许列表，然后再处理拒绝列表。如果没有显示允许列表，Splunk 平台会为所有事件建立索引。 当您使用事件代码/ID 格式： • 对于多个代码/ID，用逗号将列表分隔。 • 对于范围，使用连字符（例如，"0-1000,5000-1000"）。 当使用高级筛选格式时： • 在密钥和代表您的筛选器的正则表达式之间使用 =（例如， whitelist = EventCode=%^1([8-9])$% • 您可在单个高级筛选条目中具有多个密钥/正则表达式集。Splunk 平台从逻辑上连接该集。这意味着，该条目仅在该条目中所有设置为 true 时才有效。 • 通过向 whitelist 属性的末端添加一个数字，每个段落您可以指定多达 10 个白名单，例如 whitelist1...whitelist9。	N/A
blacklist	请不要为匹配指定文本字符串的事件建立索引。这是可选属性。 您可指定两种格式中的一种： • 一个或多个“事件日志”事件代码或事件 ID（事件日志代码/ID 格式）。 • 一个或多个密钥和正则表达式集。（高级筛选格式。） 您不可以在单个条目中混合格式。您也不可以在相同段落中混合格式。 Splunk 平台会先处理允许列表，然后再处理拒绝列表。如果没有显示拒绝列表，Splunk 平台会为所有事件建立索引。 当使用事件日志代码/ID 格式： • 对于多个代码/ID，用逗号将列表分隔。 • 对于范围，使用连字符（例如，0-1000,5000-1000）。 当使用高级筛选格式时： • 在密钥和代表您的筛选器的正则表达式之间使用 =（例如， blacklist = EventCode=%^1([8-9])$% • 您可在单个高级筛选条目中具有多个密钥/正则表达式集。Splunk 平台从逻辑上连接该集。这意味着，该条目仅在该条目中所有设置为 true 时才有效。 • 通过向 blacklist 属性的末端添加一个数字，每个段落您可以指定多达 10 个拒绝列表，例如 blacklist1...blacklist9。	
renderXml	请将事件数据呈现为由“Windows 事件日志”子系统提供的 XML。设置是可选项。 值为 1 或 true 指示要将事件呈现为 XML。值为 0 或 false 指示要将事件呈现为纯文本。 如果您将 renderXml 设为 true，还想新建允许列表或拒绝列表筛选事件数据，您必须使用允许列表或拒绝列表中的 $XmlRegex 特殊密钥。	0（false）
index	此输入将向其发送数据的索引。	默认索引
disabled	输入是否运行。 有效值为 0（指示输入将运行）和 1（指示输入不运行）。	0

使用安全事件日志监视文件更改

您可以监视系统中的文件更改，方法是对一组文件或目录启用安全审计，然后监视安全事件日志通道中的更改事件。事件日志监视输入包括三个属性，您可以在 inputs.conf 中使用它们。例如：

```
[WinEventLog://Security]
disabled = 0
start_from = oldest
current_only = 0
evt_resolve_ad_obj = 1
checkpointInterval = 5
# only index events with these event IDs.
whitelist = 0-2000,3001-10000
# exclude these event IDs from being indexed.
blacklist = 2001-3000
```

要对一组文件或目录启用安全审计，请阅读 MS Technet 中的“如何使用审核安全事件”，网址为 http://technet.microsoft.com/en-us/library/cc727935%28v=ws.10%29.aspx。

您也可以使用 suppress_text 属性来包括或排除安全事件随附的消息文本。

当您在“Windows 事件日志安全”段落中把 suppress_text 设置为 1 时，整个消息文本都不会建立索引，包括任何有关安全事件的上下文信息。如果您需要该上下文信息，请勿在段落中设置 suppress_text。

请参阅以下示例以包含或排除消息文本：

```
[WinEventLog://Security]
disabled = 0
start_from = oldest
current_only = 0
evt_resolve_ad_obj = 1
checkpointInterval = 5
# suppress message text, we only want the event number.
suppress_text = 1
# only index events with these event IDs.
whitelist = 0-2000,2001-10000
# exclude these event IDs from being indexed.
blacklist = 2001-3000
```

要使用特定的域控制器，请设置 evt_dc_name 属性：

```
[WinEventLog://Security]
disabled = 0
start_from = oldest
current_only = 0
evt_resolve_ad_obj = 1
evt_dc_name = boston-dc1.contoso.com
checkpointInterval = 5
# suppress message text, we only want the event number.
suppress_text = 1
# only index events with these event IDs.
whitelist = 0-2000,2001-10000
# exclude these event IDs from being indexed.
blacklist = 2001-3000
```

要使用主域控制器解析 AD 对象，请把 evt_resolve_ad_ds 属性设置为 PDC。否则，它会找到最近的域控制器。

```
[WinEventLog://Security]
disabled = 0
start_from = oldest
current_only = 0
evt_resolve_ad_obj = 1
evt_resolve_ad_ds = PDC
checkpointInterval = 5
# suppress message text, we only want the event number.
```

```
suppress_text = 1
# only index events with these event IDs.
whitelist = 0-2000,2001-10000
# exclude these event IDs from being indexed.
blacklist = 2001-3000
```

使用白名单和黑名单新建高级筛选

除了仅基于事件代码执行筛选，您还可以使用 whitelist 和 blacklist 设置对传入事件执行高级筛选。为此，指定设置中的密钥/正则表达式格式：

```
whitelist = key=<regular expression> [key=<regular expression>] ...
```

在此格式中，key 必须是来自以下表的有效条目：

键	描述
$TimeGenerated	计算机生成事件的时间。Splunk Enterprise 仅将时间字符串生成为事件。
$Timestamp	由“事件日志”服务接收和记录的事件的时间。Splunk Enterprise 仅将时间字符串生成为事件。
$XmlRegex	一个特殊的键，用于将 Splunk Enterprise 配置为过滤 XML 事件。要使用此密钥，将其设为您想要 Splunk Enterprise 筛选的值。您必须配置要对其应用拒绝列表或允许列表的一个输入，以在 XML 中呈现事件。要生成 XML 事件，请在输入段落下指定 renderXml = true 设置。Splunk Enterprise 会合并单个允许列表或拒绝列表中的多个条目。所有过滤器条目必须匹配，过滤器才能触发。
类别	指定事件来源的类别数量。
CategoryString	类别的字符串转换。该转换取决于事件来源。
ComputerName	生成事件的计算机名称。
EventCode	事件的事件 ID 号。对应于事件查看器中的**事件 ID**。
EventType	代表可被记录的五种事件类型中任意一种的数字值：“错误”、“警告”、“信息”、“审计成功”和“审计失败”。仅在运行 Windows Server 2003 及更低版本的计算机或运行 Windows XP 及更低版本的客户端上可用。请参阅 MSDN 上的 Win32_NTLogEvent class (Windows)，网址为 http://msdn.microsoft.com/en-us/library/aa394226(v=vs.85).aspx。
Keywords	用于在事件日志通道内将不同类型的事件进行分类的元素。例如，“安全事件日志”通道具有此元素。
LogName	接收事件的“事件日志”通道名称。对应于事件查看器中的**日志名称**。
消息	事件中消息的文本。
OpCode	事件的安全级别。对应于事件查看器中的 **OpCode**。
RecordNumber	Windows 事件日志记录编号。Windows 计算机上的每个事件获取一个记录编号。该记录编号从 0 开始，并带有系统上生成的第一个事件，且随着每个生成的新事件而增加，直到它达到最大值 4294967295。然后，它滚动回到 0。
Sid	与事件相关联或生成事件的主体（如用户、组、计算机或其他实体）的安全标识符 (SID)。请参阅 MSDN 上的 Win32_UserAccount class，网址为 http://msdn.microsoft.com/en-us/library/windows/desktop/aa394507%28v=vs.85%29.aspx。
SidType	代表与事件相关的 SID 类型的数字值。请参阅 MSDN 上的 Win32_UserAccount class，网址为 http://msdn.microsoft.com/en-us/library/windows/desktop/aa394507%28v=vs.85%29.aspx。
SourceName	生成事件的实体的来源。对应于事件查看器中的**来源**。
TaskCategory	事件的任务类别。事件来源允许您定义类别，以便您可使用“事件查看器”筛选它们（使用 Task Category 字段）。请参阅 MSDN 上的事件类别 (Windows)，网址为 http://msdn.microsoft.com/en-us/library/aa363649%28VS.85%29.aspx。
类型	代表可被记录的五种事件类型中任意一种的数字值：“错误”、“警告”、“信息”、“审计成功”和“审计失败”。仅在运行 Windows Server 2008 或更高版本或 Windows Vista 或更高版本的计算机上可用。请参阅 MSDN 上的 Win32_NTLogEvent class (Windows)，网址为 http://msdn.microsoft.com/en-us/library/aa394226(v=vs.85).aspx。
User	与事件相关的用户。与事件查看器中的**用户**相关联。

<regular expression> 是代表您想要包括（与 whitelist 属性一起使用）或排除（与 blacklist 属性一起使用）的筛选器的任何有效正则表达式。

您可在单个条目行指定多个密钥/正则表达式集。执行此操作时，Splunk Enterprise 从逻辑上连接该设置。这意味着只有满足行中所有集的事件将对包含或排除有效。请参阅以下示例：

```
whitelist = EventCode="^1([0-5])$" Message="^Error"
```

意味着包括 EventCode 在 10 到 15 之间的事件，且包含 Message（以单词 Error 开头）。

您可在每个段落中指定最多 10 个单独的允许列表或拒绝列表条目。为实现此操作，请在单独成行的 whitelist 或 blacklist 条目末尾添加数字：

```
whitelist = key=<regular expression>
whitelist1 = key=<regular expression> key2=<regular expression 2>
whitelist2 = key=<regular expression>
```

您不可以指定具有多个引用相同密钥的密钥/正则表达式集的条目。例如，如果您指定：

```
whitelist = EventCode="^1([0-5])$" EventCode="^2([0-5])$"
```

Splunk Enterprise 会忽略第一组，且仅尝试包含匹配第二组的事件。此种情况下，仅 EventCode 在 20 到 25 之间的事件匹配。EventCode 在 10 到 15 之间的事件则匹配失败。仅条目中的最后组匹配。要解决此问题，指定该段落中的两个单独的条目：

```
whitelist = EventCode="^1([0-5])$"
whitelist1 = EventCode="^2([0-5])$"
```

抑制 Windows 事件日志事件中的字段

有两个选项可通过从 Splunk Platform 实例引入的事件中移除 Windows 事件日志字段的方式来限制数据的引入：

- 使用 inputs.conf 中的 suppress_* 设置从引入的事件中移除某些 Windows 事件日志字段。
- 使用 evt_exclude_fields 设置，您可以通过该设置从 Windows 事件日志事件中移除任何 Windows 事件日志字段。此设置会从该事件中移除所排除的键和值，并且即使允许列表包含此该字段，事件也会被排除。

您可以在 Windows 事件日志监视输入下的 inputs.conf 配置文件中定义这两个设置，例如：

suppress_* 示例（抑制消息文本）

```
[WinEventLog://System]
disabled=0
suppress_text=1
```

evt_exclude_fields 示例

```
[WinEventLog://System]
disabled=0
evt_exclude_fields=EventCode,RecordNumber
```

请参阅本主题前面章节“使用‘白名单’和‘黑名单’创建高级过滤器”中的字段列表。请参阅本主题前面的章节“配置用于监视 Windows 事件日志的设置”，了解有关这些设置的更多信息。

解析事件日志文件中的 Active Directory 对象

若要指定是否为给定的 Windows 事件日志通道解析诸如全局唯一标识符 (GUID) 和安全标识符 (SID) 等 Active Directory 对象，请使用该通道段落的 evt_resolve_ad_obj 属性 (1=enabled, 0=disabled)，该通道段落位于 inputs.conf 的本地副本内。evt_resolve_ad_obj 属性默认为对安全通道启用。

例如：

```
[WinEventLog://Security]
disabled = 0
start_from = oldest
current_only = 0
evt_resolve_ad_obj = 1
checkpointInterval = 5
```

为了解析 AD 对象，Splunk 平台实例应绑定至某域；如需指定该域的域控制器，请使用 evt_dc_name 属性。

evt_dc_name 属性中指定的字符串可以代表域控制器的 NetBIOS 名称或其完全限定域名（FQDN）。可在任一名称类型前面加上两个反斜杠字符（可选）。

以下示例是格式正确的域控制器名称：

- FTW-DC-01
- \\FTW-DC-01
- FTW-DC-01.splunk.com
- \\FTW-DC-01.splunk.com

如需指定要绑定至域的 FQDN，请使用 evt_dns_name 属性。

例如：

```
[WinEventLog://Security]
disabled = 0
start_from = oldest
current_only = 0
evt_resolve_ad_obj = 1
evt_dc_name = ftw-dc-01.splunk.com
evt_dns_name = splunk.com
checkpointInterval = 5
```

当您使用 evt_resolve_ad_obj 和 evt_dc_name 属性时，以下约束适用：

- Splunk 软件会先尝试使用 evt_dc_name 属性中指定的域控制器（DC）解析 SID 和 GUID。如果无法使用此 DC 解析 SID，它会尝试绑定到默认 DC 来执行转换。
- 如果无法联系 DC 来转换 SID，Splunk 软件会尝试使用本地计算机进行转换。
- 如果这些方法都没有用，则 Splunk 会按照在事件捕获时的样子打印 SID。
- Splunk 软件无法转换非 S-1-N-NN-NNNNNNNNNN-NNNNNNNNNN-NNNNNNNNNN-NNNN 格式的 SID。

如果您发现 SID 未正确转换，请审阅 %SPLUNK_HOME%\var\log\splunkd.log 查找可能的问题线索。

指定是从最早的事件还是从最近的事件开始建立索引

使用 start_from 属性可以指定是从最早的事件还是最近的事件开始为事件建立索引。默认情况下，索引建立将从时间最早的数据开始向前建立索引。请勿更改此设置，因为 Splunk 软件使用此方法对 backlog 建立索引后会停止建立索引。

使用 current_only 属性可以指定是否为某给定日志通道内所有的现有事件新建索引。设置为 1 时，仅 Splunk 部署启动时出现的事件才会建立索引。设置为 0 时，所有事件都会建立索引。

例如：

```
[WinEventLog://Application]
disabled = 0
start_from = oldest
current_only = 1
```

在 XML 中显示 Windows Event Log 事件

要使 Splunk Enterprise 在 XML 中生成 Windows Event Log 事件，使用 Windows Event Log 输入段落中的 renderXml 设置：

```
[WinEventLog://System]
 disabled = 0
 renderXml = 1
 evt_resolve_ad_obj = 1
 evt_dns_name = \"SV5DC02\"
```

此输入段落生成如下所示的事件：

```
 <Event xmlns='http://schemas.microsoft.com/win/2004/08/events/event'>
   <System>
     <Provider Name='Service Control Manager' Guid='{555908d1-a6d7-4695-8e1e-26931d2012f4}' EventSourceName='Service Control
Manager'/>
```

```
    <EventID Qualifiers='16384'>7036</EventID>
    <Version>0</Version>
    <Level>4</Level>
    <Task>0</Task>
    <Opcode>0</Opcode>
    <Keywords>0x8080000000000000</Keywords>
    <TimeCreated SystemTime='2014-04-24T18:38:37.868683300Z'/>
    <EventRecordID>412598</EventRecordID>
    <Correlation/>
    <Execution ProcessID='192' ThreadID='210980'/>
    <Channel>System</Channel>
    <Computer>SplunkDoc.splunk-docs.local</Computer>
    <Security/>
  </System>
  <EventData>
    <Data Name='param1'>Application Experience</Data>
    <Data Name='param2'>stopped</Data>
    <Binary>410065004C006F006F006B0075007000530076006300 2F0031000000</Binary>
  </EventData>
</Event>
```

当您指示 Splunk Enterprise 在 XML 中呈现事件时，无论计算机的系统区域为何种语言，XML 事件内的事件密钥均以英语形式呈现。比较在 Windows 服务器的法语版本上生成的以下事件。

标准事件：

```
04/29/2014 02:50:23 PM
LogName=Security
SourceName=Microsoft Windows security auditing.
EventCode=4672
EventType=0
Type=Information
ComputerName=sacreblue
TaskCategory=Ouverture de session sp<ciale
OpCode=Informations
RecordNumber=2746
Keywords=Succès de l'audit
Message=Privilèges spéciaux attribués à la nouvelle ouverture de session.

Sujet :
               ID de sécurité :                 AUTORITE NT\Système
               Nom du compte :                         Système
               Domaine du compte :                   AUTORITE NT
               ID d'ouverture de session :                           0x3e7

Privilèges :                        SeAssignPrimaryTokenPrivilege
                                           SeTcbPrivilege
                                           SeSecurityPrivilege
                                           SeTakeOwnershipPrivilege
                                           SeLoadDriverPrivilege
                                           SeBackupPrivilege
                                           SeRestorePrivilege
                                           SeDebugPrivilege
                                           SeAuditPrivilege
                                           SeSystemEnvironmentP rivilege
                                           SeImpersonatePrivilege
```

XML 事件：

```
<Event xmlns='http://schemas.microsoft.com/win/2004/08/events/event'>
               <System><Provider Name='Microsoft-Windows-Security-Auditing' Guid='{54849625-5478-4994-A5BA-3E3B0328C30D}'/>
                              <EventID>4672</EventID>
                              <Version>0</Version>
                              <Level>0</Level>
                              <Task>12548</Task>
                              <Opcode>0</Opcode>
                              <Keywords>0x8020000000000000</Keywords>
```

```
                        <TimeCreated SystemTime='2014-04-29T22:15:03.280843700Z'/>
                        <EventRecordID>2756</EventRecordID>
                        <Correlation/><Execution ProcessID='540' ThreadID='372'/>
                        <Channel>Security</Channel>
                        <Computer>sacreblue</Computer>
                        <Security/>
                </System>
                <EventData>
                        <Data Name='SubjectUserSid'>AUTORITE NT\Système</Data>
                        <Data Name='SubjectUserName'>Système</Data>
                        <Data Name='SubjectDomainName'>AUTORITE NT</Data>
                        <Data Name='SubjectLogonId'>0x3e7</Data>
                        <Data Name='PrivilegeList'>SeAssignPrimaryTokenPrivilege
                                SeTcbPrivilege
                                SeSecurityPrivilege
                                SeTakeOwnershipPrivilege
                                SeLoadDriverPrivilege
                                SeBackupPrivilege
                                SeRestorePrivilege
                                SeDebugPrivilege
                                SeAuditPrivilege
                                SeSystemEnvironmentP rivilege
                                SeImpersonatePrivilege< /Data>
                </EventData>
</Event>
```

尽管 Data Name 密钥在标准事件中以系统自带语言呈现，但它在 XML 事件中仍以英语呈现。

使用允许列表和拒绝列表筛选基于 XML 的事件

如果您在 XML 中呈现事件，并想要使用允许列表和拒绝列表筛选这些事件，当您构建自己的允许列表或拒绝列表时必须使用特殊密钥 $XmlRegex。

当 Splunk Enterprise 在 XML 呈现的事件中的任何位置找到您用 $XmlRegex 指定的值时，就会触发允许列表或拒绝列表。$XmlRegex 在未使用 renderXml = true 设置显式指定要在 XML 中呈现事件的输入的情况下将不起作用。

$XmlRegex 设置不会搜索键值对。它会配置 Splunk Enterprise，使之预期传入事件以 XML 格式呈现。

以下示例将 whitelist 设置配置为允许 XML 事件。Splunk Enterprise 会为包含 "Error" 一词的所有 XML 事件建立索引。

```
[WinEventLog://System]
 disabled = 0
 renderXml = 1
 evt_resolve_ad_obj = 1
 evt_dns_name = \"SV5DC02\"
 whitelist = $XmlRegex='Error'
```

请参阅本主题前面的“用白名单和黑名单新建高级过滤器”部分了解更多信息和语法。

使用 CLI 配置事件日志监视

您可使用 CLI 配置本地事件日志监视。在您使用 CLI 前，请首先在 inputs.conf 中创建段落条目。请参阅本主题中的“使用 inputs.conf 配置事件日志监视”。

CLI 对远程“事件日志”集合不可用。

要在本地计算机上列出所有配置的“事件日志”通道，请输入以下内容：

```
> splunk list eventlog
```

您还可通过指定特定通道的名称来将其列出，如以下示例所示：

```
> splunk list eventlog <ChannelName>
```

要启用“事件日志”通道，请执行以下操作：

```
> splunk enable eventlog <ChannelName>
```

要禁用通道，请执行以下操作：

```
> splunk disable eventlog <ChannelName>
```

为导出的事件日志文件建立索引

要为导出的 Windows 事件日志（.evt 或 .evtx）文件建立索引，请监视包含导出文件的目录。请参阅“监视文件和目录”。

请不要尝试监视允许写入的 .evt 或 .evtx 文件。Windows 不允许对这些文件进行读取访问。请改用事件日志监视功能。

直接监视 Windows Event Log 文件的限制

直接监视 Windows 事件日志文件具有以下限制：

- 由于 Windows XP 和 Server 2003 系统上的 API 和日志通道处理约束，从这些系统导入的 .evt 文件不包含 Message 字段。这意味着，Message 字段的内容不会显示在您的索引中。
- 如果 .evtx 文件来自运行 Windows Vista 及更高版本或 Windows Server 2008/2008 R2 及更高版本的系统，运行在 Windows XP 和 Windows Server 2003/2003 R2 上的 Splunk Enterprise 无法为这类文件新建索引。
- 运行在 Windows Vista 及更高版本和 Server 2008/2008 R2 及更高版本上的 Splunk Enterprise 可以为 .evt 和 .evtx 文件新建索引。
- 如果您的 .evt 或 .evtx 文件不是来自标准事件日志通道，必须确保该通道所需的任何动态链接库 (DLL) 文件存在于要建立索引的计算机上。
- 如果一个 .evt 或 .evtx 文件位于收集该文件的计算机的主要区域设置和语言，则 Splunk Enterprise 将为该 .evt 或 .evtx 文件新建索引。
- 从其他计算机中导出的文件无法使用“Splunk Web 上载”功能。这是因为上述文件中包含的信息是针对生成这些文件的计算机。如果不变更格式，其他计算机将无法处理这些文件。

如果在一个系统上生成 .evt 或 .evtx 文件，而在另一个系统上监视这些文件，则可能并非每个事件中的所有字段都将像在生成事件的系统上那样进行扩展。这是由 DLL 版本、可用性及 API 的变化造成的。操作系统版本、语言、服务包级别以及安装的第三方 DLL 的差异也可能会产生这种影响。

监视 Windows 上的文件系统更改

Splunk 平台支持通过安全 Windows 事件日志通道监视 Windows 文件系统更改。要监视文件更改，您必须对要监视更改的文件和文件夹启用安全审计，然后使用该事件日志监视器监视安全事件日志通道。这种文件系统更改监视过程替代了已经弃用的文件系统更改监视器输入。

如果您使用的是 Splunk Cloud，而且想要通过“安全事件日志”通道监视 Windows 文件系统更改，请使用 Splunk 通用转发器监视 Windows 计算机上的更改。

要求

您必须满足下列要求，才能监视文件系统更改：

- Splunk 平台必须在 Windows 上运行。请参阅*《安装手册》*中的“在 Windows 上安装”。
- Splunk 平台必须以本地系统用户身份或具有特定安全策略权限的域用户身份运行以读取安全事件日志
- 您必须对希望 Splunk 平台监视其更改的文件或目录启用安全审计

使用安全事件日志监视文件更改

您可以监视系统中的文件更改，方法是对一组文件或目录启用安全审计，然后监视安全事件日志通道中的更改事件。事件日志监视输入包括三个设置，您可以在 inputs.conf 配置文件中使用它们。您无法从 Splunk Web 配置文件系统更改事件的监视。

您可使用安全事件日志和文件系统更改上下文之外的这些设置。该设置列表仅为 inputs.conf 文件可用设置的一个子集。有关其他设置，请参阅“使用 Splunk Cloud 监视 Windows 事件日志数据”。

下表描述了可以在 input.conf 中进行文件监视的配置设置：

设置	描述	默认
whitelist		N/A

	为匹配指定文本字符串的事件建立索引。设置是可选项。 您可指定两种格式中的一种： • 一个或多个“事件日志”事件代码或事件 ID（事件日志代码/ID 格式） • 一个或多个密钥和正则表达式集（高级筛选格式） 您不可以在单个条目中混合格式或在相同段落中混合格式。 Splunk 平台会先处理允许列表，然后再处理拒绝列表。如果没有显示允许列表，Splunk 平台会为所有事件建立索引。 当使用事件代码/ID 格式时，请遵循以下规则： • 对于多个代码/ID，用逗号将列表分隔。 • 对于范围，使用连字符（例如，"0-1000,5000-1000"）。 当使用高级筛选格式时，请遵循以下规则： • 在密钥和代表您的筛选器的正则表达式之间使用 =（例如， `whitelist = EventCode=%^1([8-9])$%` • 您可在单个高级筛选条目中具有多个密钥/正则表达式集。Splunk 平台从逻辑上连接该集。这意味着，该条目仅在该条目中所有设置为 `true` 时才有效。 • 通过向 `whitelist` 设置的末端添加一个数字，每个段落您可以指定多达 10 个允许列表，例如 `whitelist1...whitelist9`。	
blacklist	请不要为匹配指定文本字符串的事件建立索引。设置是可选项。 您可指定两种格式中的一种： • 一个或多个“事件日志”事件代码或事件 ID（事件日志代码/ID 格式） • 一个或多个密钥和正则表达式集（高级筛选格式） 您不可以在单个条目中混合格式或在相同段落中混合格式。 Splunk 平台会先处理允许列表，然后再处理拒绝列表。如果没有显示允许列表，Splunk 平台会为所有事件建立索引。 当使用事件代码/ID 格式时，请遵循以下规则： • 对于多个代码/ID，用逗号将列表分隔。 • 对于范围，使用连字符（例如，"0-1000,5000-1000"）。 当使用高级筛选格式时，请遵循以下规则： • 在密钥和代表您的筛选器的正则表达式之间使用 =（例如， `whitelist = EventCode=%^1([8-9])$%` • 您可在单个高级筛选条目中具有多个密钥/正则表达式集。Splunk 平台从逻辑上连接该集。这意味着，该条目仅在该条目中所有设置为 `true` 时才有效。 • 通过向 `blacklist` 设置的末端添加一个数字，每个段落您可以指定多达 10 个拒绝列表，例如 `blacklist1...blacklist9`。	N/A
suppress_text	是否包含安全事件随附的消息文本。 值为 1 时禁止消息文本。值为 0 时保留该文本。	0

使用白名单和黑名单设置新建高级筛选

除了仅基于事件代码执行筛选，您还可以使用 `whitelist` 和 `blacklist` 设置对传入事件执行高级筛选。为此，指定设置中的密钥/正则表达式格式：

```
whitelist = key=<regular expression> [key=<regular expression] ...
```

在此格式中，`key` 是来自以下列表的有效条目：

键	描述

$TimeGenerated	计算机生成事件的时间。仅将时间字符串生成为事件。
$Timestamp	由“事件日志”服务接收和记录事件的时间。Splunk 平台仅将时间字符串生成为事件。
类别	指定事件来源的类别数量。
CategoryString	类别的字符串转换。该转换取决于事件来源。
ComputerName	生成事件的计算机名称。
EventCode	事件的事件 ID 号。对应于事件查看器中的“事件 ID”。
EventType	代表可被记录的五种事件类型（“错误”、“警告”、“信息”、“审计成功”和“审计失败”）中任意一种的数字值。仅在运行 Windows Server 2003 及更低版本的服务器计算机或运行 Windows XP 及更低版本的客户端上可用。
Keywords	用于在事件日志通道内将不同类型的事件进行分类的元素。例如，“安全事件日志”通道具有此元素。
LogName	接收事件的“事件日志”通道名称。对应于事件查看器中的“日志名称”。
消息	事件中消息的文本。
OpCode	事件的安全级别。对应于事件查看器中的“OpCode”。
RecordNumber	Windows 事件日志记录编号。Windows 服务器上的每个事件获取一个记录编号。该记录编号从 0 开始，并带有系统上生成的第一个事件，且随着每个生成的新事件而增加，直到它达到最大值 4294967295。然后，它滚动回到 0。
Sid	与事件相关联或生成事件的主体（如用户、组、计算机或其他实体）的安全标识符 (SID)。请参阅 MSDN 上的 Win32_UserAccount class，网址为 https://msdn.microsoft.com/en-us/library/windows/desktop/aa394507%28v=vs.85%29.aspx。
SidType	代表与事件相关的 SID 类型的数字值。请参阅 MSDN 上的 Win32_UserAccount class，网址为 https://msdn.microsoft.com/en-us/library/windows/desktop/aa394507%28v=vs.85%29.aspx。
SourceName	生成事件的实体的来源。对应于事件查看器中的“来源”。
TaskCategory	事件的任务类别。事件来源允许您定义类别，以便您可使用“事件查看器”筛选它们（使用“任务类别”字段）。请参阅 MSDN 上的事件类别 (Windows)，网址为 https://msdn.microsoft.com/en-us/library/aa363649%28VS.85%29.aspx。
类型	代表可被记录的五种事件类型（“错误”、“警告”、“信息”、“审计成功”和“审计失败”）中任意一种的数字值。仅在运行 Windows Server 2008 或更高版本的服务器计算机或运行 Windows Vista 或更高版本的客户端上可用。
User	与事件相关的用户。与事件查看器中的“用户”相关联。

<regular expression> 是代表您想要包括（与 whitelist 设置一起使用）或排除（与 blacklist 设置一起使用）的筛选器的任何有效正则表达式。

您可在单个条目行指定多个正则表达式。只有满足行中所有条目的事件才会被包括在内或排除在外。例如，此条目意味着包括 EventCode 在 10 到 15 之间的事件，且包含 Message（以单词 Error 开头）：

```
whitelist = EventCode="^1([0-5])$" Message="^Error"
```

您可在每个段落中指定最多 10 个单独的允许列表或拒绝列表条目。为实现此操作，请在单独成行的 whitelist 或 blacklist 设置条目末尾添加数字：

```
whitelist = key=<regular expression>
whitelist1 = key=<regular expression> key2=<regular expression 2>
whitelist2 = key=<regular expression>
```

您不可以指定具有多个引用相同密钥的正则表达式的条目。

如果您使用多个引用相同密钥的表达式指定一个条目，则 Splunk 平台会忽略第一个正则表达式，且仅尝试包含与第二个正则表达式相匹配的事件。请参阅以下示例：

```
whitelist = EventCode="^1([0-5])$" EventCode="^2([0-5])$"
```

此种情况下，仅 EventCode 在 20 到 25 之间的事件匹配。EventCode 在 10 到 15 之间的事件则匹配失败。仅条目中的最后一个正则表达式匹配。

要解决此问题，指定该段落中的两个单独的条目：

```
whitelist = EventCode="^1([0-5])$"
whitelist1 = EventCode="^2([0-5])$"
```

监视文件系统更改

1. 请确认您有管理员权限。
2. 启用安全审核。针对您运行的 Windows 版本搜索“启用安全审核”。
3. 将 Splunk 平台的事件日志监视器输入配置为监视安全事件日志通道。

有关如何配置“事件日志”监视器输入的说明，请参阅“监视 Windows 事件日志数据”。

文件系统更改监视器示例

以下 inputs.conf 段落显示了如何监视文件系统更改的示例。

此段落收集事件 ID 代码介于 0 到 2000 以及 3001-10000 之间的安全事件。

```
[WinEventLog:Security]
disabled = 0
start_from = oldest
current_only = 0
evt_resolve_ad_obj = 1
checkpointInterval = 5
# only index events with these event IDs.
whitelist = 0-2000,2001-10000
# exclude these event IDs from being indexed.
blacklist = 2001-3000
```

此段落收集事件 ID 代码介于 0 到 2000 以及 3001-10000 之间的安全事件。它还禁止了事件 ID 中的消息文本。

```
[WinEventLog:Security]
disabled = 0
start_from = oldest
current_only = 0
evt_resolve_ad_obj = 1
checkpointInterval = 5
# suppress message text, we only want the event number.
suppress_text = 1
# only index events with these event IDs.
whitelist = 0-2000,2001-10000
# exclude these event IDs from being indexed.
blacklist = 2001-3000
```

通过 Windows Management Instrumentation（WMI）监视数据

Splunk 平台支持使用 Windows Management Instrumentation（WMI）提供程序对远程计算机上的 Windows 性能和事件日志数据进行访问，而无需在这些计算机上安装软件。

要将 WMI 数据导入 Splunk Cloud，您可以在 Windows 计算机上安装通用转发器或重型转发器，并将该转发器配置为使用 WMI 数据输入从其他 Windows 计算机远程收集数据，然后将该数据转发到您的 Splunk Cloud 实例。Splunk Cloud 无法使用 WMI 直接连接到 Windows 计算机，因此通用转发器或重型转发器是唯一的选择。

WMI 数据导入可连接到多个 WMI 提供程序。输入以单独的进程形式在转发器上运行，该进程名为 splunk-wmi.exe。它为**脚本式输入**。

如果可能，当您需要远程收集 Windows 数据时，请直接在要从中收集 Windows 数据的计算机上安装通用转发器，而不是使用 WMI。在许多情况中，WMI 的资源负载量会超过安装通用转发器的资源负载量。如果您要从每个计算机或从非常繁忙的计算机（如域控制器）收集多个事件日志或性能计数器，请使用转发器。请参阅“确定如何监视远程 Windows 数据的相关注意事项”。

您需要什么来监视基于 WMI 的数据？

下面是监视基于 WMI 的数据的最低要求。根据要监视的事件日志或性能计数器，您可能需要其他权限。

有关监视基于 WMI 的数据所需内容的其他详细信息，请参阅本主题中的“安全与远程访问注意事项”。

活动	所需权限
通过 WMI 监视远程事件日志	• Splunk Cloud 必须从在 Windows 上运行的转发器接收数据。 • 如果 Splunk Enterprise 在 Windows 计算机上运行，则可以直接收集 WMI 数据。 • 转发器必须以至少具有 WMI 读取访问权限的 Windows 域用户身份运行。 • 转发器必须以具有想要访问的事件日志的适当访问权限的域用户身份运行。
通过 WMI 监视远程性能监视器计数器	• Splunk Cloud 必须从在 Windows 上运行的转发器接收数据。 • 如果 Splunk Enterprise 在 Windows 计算机上运行，则可以通过 WMI 收集性能数据。 • 转发器必须以至少具有 WMI 读取访问权限的域用户身份运行。 • 转发器必须以具有“Windows 性能数据助手”库适当访问权限的域用户身份运行。

如果您运行 Splunk Enterprise，并且您安装实例的计算机运行 Windows，或者您用于将数据发送到实例的转发器运行 Windows，则您可以通过 WMI 收集数据。

安全与远程访问注意事项

必须正确配置获取 WMI 数据的 Splunk 平台实例和您的 Windows 网络，才能通过 WMI 访问数据。无论为您的数据建立索引的实例是 Splunk Cloud 还是 Splunk Enterprise 实例，在尝试使用该平台通过 WMI 获取数据之前，请查看以下前提条件。

在 Splunk 平台获取基于 WMI 的数据之前：

- 必须使用具有远程网络连接执行权限的 Windows 用户身份安装获取数据的实例。
- 运行实例的 Windows 用户必须是 Active Directory（AD）域或林的成员，而且必须具有查询 WMI 提供程序的相应权限。
- 实例用户还必须是运行实例的计算机上本地管理员组的成员。
- 运行实例的计算机必须能够连接到远程计算机，并且该计算机在连接之后必须能够从远程计算机获取所需的数据。
- Splunk 平台实例与目标计算机必须属于相同 AD 域或林。

Splunk 平台实例运行的用户不必是“域管理员”组的成员，而且出于安全原因，也不应该如此。但是，您必须具有域管理员权限才能为该用户配置访问权限。如果您没有域管理员访问权限，则找到可以为您配置 Splunk 用户访问权限或者可为您提供域管理员权限的人员。

如果您已经将实例作为本地系统 Windows 用户进行安装，则无法通过 WMI 进行远程验证。本地系统用户没有网络中其他计算机的访问权限。无法授予“本地系统”帐户访问另一台计算机的权限。

通过执行以下操作之一，您即可向实例用户授予访问 WMI 提供程序的权限：

- 分配最小权限，如本主题后面的“针对最小访问权限配置 WMI”中所述。建议采用此选项。
- 将用户添加到要轮询的每台 Windows 成员计算机上的本地管理员组。出于安全原因，不建议这样做。
- 将用户添加到“域管理员”全局组。出于安全原因，不建议这样做。

组成员资格和资源访问控制列表（ACL）

为了保持安全完整性，将 Splunk 平台实例运行的 Windows 用户放入域全局组中，并为该组分配 Windows 计算机和资源 ACL 的权限，而不是直接向用户分配该权限。直接向用户分配权限存在安全风险，可能会导致安全审计或未来更改过程出现问题。

针对最小访问权限配置 WMI

如果您为 Splunk 平台配置的运行 Windows 用户身份不是域管理员，则您必须将 WMI 配置为向该用户提供访问权限。按照以下检查表授予所有 Windows 资源（包括 WMI）的最小访问权限。其他信息及分步指导说明，请参阅*《安装》*手册中“为您的 Windows 网络做好 Splunk Enterprise 安装准备”。这些说明也适用于通用转发器。

为方便 Splunk 平台实例使用最小权限方法通过 WMI 收集数据，您必须给 Splunk 平台以其身份运行的用户授予多种级别的访问权限。

- **本地安全策略权限。**Splunk 平台 Windows 用户需要在要为基于 WMI 的数据轮询的每个计算机上定义以下本地安全策略用户权限分配：

    - 从网络访问此计算机
    - 充当操作系统的一部分
    - 作为批处理任务登录
    - 作为服务登录
    - 配置系统性能
    - 替换进程级别标记

- **分布式组件对象模型（DCOM）配置和权限。**必须在要监视的每个 Windows 计算机上都启用 DCOM。另外，Splunk 平台 Windows 用户必须有权访问 DCOM。可通过许多方法执行此操作，但最佳方法是将“分布式 COM 用户”域全局组嵌套到要监视的每个 Windows 计算机上的“分布式 COM 用户”本地组中，然后将 Splunk 平台 Windows 用户添加到“分布式 COM 用户”域全局组中。有关特定详细信息和说明，请在 Microsoft 文档网站上搜索“确保远程 WMI 连接安全”。

- **性能监视器配置和权限。**要使 Splunk 平台能够通过 WMI 访问远程性能对象，Splunk 平台用户必须是“性能日志用户”本地组的成员。实现此目的的最佳方法是，将“性能日志用户”域全局组放入每个 Windows 计算机上的“性能日志用户”本地组中，然后将 Windows 用户分配到“性能日志用户”组。

- **WMI 命名空间安全性。**Splunk 平台访问的 WMI 命名空间必须具备适当的权限。在几乎所有情况下，`Root\CIMV2` 命名空间都可以用于收集 WMI 数据。因为 WMI 没有全局安全性，必须在贵公司的每个 Windows 计算机上手动设置命名空间权限。必须将这些权限分配给根命名空间及其下面的所有子命名空间。在 Microsoft 文档网站上搜索“管理 WMI 安全性”。为 Splunk 用户“根”命名空间内的每个主机启用 WMI 树上的下列权限：
    - 执行方法
    - 启用帐户
    - 远程启用
    - 读取安全

对于使用组策略立即将 WMI 安全设置远程部署到多个计算机，没有标准策略。但是，Microsoft 文档网站上有一篇博客，介绍了如何使用启动脚本来完成此任务。在 Microsoft 文档网站上搜索“使用 GPO（脚本）设置 WMI 命名空间安全”。

- **防火墙配置。**如果您已启用防火墙，则必须将其配置为允许访问 WMI。如果您使用的是 Windows 最新版本中包含的 Windows 防火墙，则例外列表应明确包括 WMI。您必须为原始计算机和目标 Windows 计算机都设置此例外。在 Microsoft 文档网站上搜索“远程连接到 WMI”。

- **用户访问控制（UAC）配置。**如果您运行的是 Windows Vista 和更高版本或 Windows Server 2012 和更高版本系列，UAC 会影响 Windows 分配权限的方式。在 Microsoft 文档网站上搜索“用户帐户控制和 WMI”。

测试 WMI 提供程序的访问权限

在您配置完 WMI 并为 Splunk 平台实例用户设置您的域的访问权限之后，请测试远程计算机的访问权限。

此过程包括临时更改 Splunk 平台实例数据存储目录（`SPLUNK_DB` 所指向的位置）的步骤。在测试 WMI 的访问权限之前，您必须执行此操作。否则可能会导致 WMI 事件缺失。这是因为 `splunk-wmi.exe` 进程每次运行时都会更新 WMI 检查点文件。

如果尝试登录到域控制器，您可能得更改域控制器安全政策，才能为指定用户分配“允许本地登录”政策。

1. 以 Splunk 平台 Windows 用户身份登录到运行 Splunk 平台的 Windows 计算机。
2. 打开命令提示符（单击**开始 > 运行**，然后键入 `cmd`）。
3. 转到 Splunk 平台安装下的 `bin` 子目录（例如 `cd c:\Program Files\Splunk\bin`）。
4. 运行以下命令，确定平台实例当前用来存储其数据的位置：
   ```
   > splunk show datastore-dir
   ```
   请记住平台实例存储其数据的位置。您稍后会重新调用它。
5. 运行以下命令以更改 Splunk 平台用来临时存储其数据的位置：
   ```
   > splunk set datastore-dir %TEMP%
   ```
   在此示例中，数据存储区目录被设置为 TEMP 环境变量指定的目录。如果您希望设置为不同的目录，可以进行更改，但前提是该目录必须已经存在。
6. 重新启动 Splunk 平台实例：
   ```
   > splunk restart
   ```
   Splunk 平台实例重新启动可能需要一点时间。
7. 一旦 Splunk 平台实例完成重启，请测试访问 WMI 提供程序的权限，用远程计算机的名称替换 `<host>`：
   ```
   > splunk cmd splunk-wmi -wql "select * from win32_service" -namespace \\<host>\root\cimv2
   ```
    - 如果您看到有数据流返回但没有错误消息，然后实例能够连接到 WMI 提供程序并成功查询。
    - 如果有错误，则将显示一条包含错误原因的消息。在输出中查找 `error="<msg>"` 字符串，以寻找有关如何更正此问题的线索。

测试 WMI 访问权限之后，请将 Splunk 平台实例重新指向正确的数据库目录，方法是运行以下命令，然后重新启动实例：

```
> splunk set datastore-dir <directory shown from Step 4>
```

配置基于 WMI 的输入

在 Windows 上，Splunk 平台中的所有远程数据集合需通过 WMI 提供程序或转发器发生。

如果您使用的是 Splunk Cloud，必须用转发器收集 WMI 数据并把数据转发到您的 Splunk Cloud 实例。

如果您使用的是 Splunk Enterprise，则可以使用 Splunk Web 或配置文件配置基于 WMI 的输入。使用文件时，有更多的配置选项可供您选择。

使用 Splunk Web 配置基于 WMI 的输入

Splunk Cloud 无法在本机收集 WMI 数据，但您可以在已配置为重型转发器的 Splunk Enterprise 实例上使用 Splunk Web 来收集数据并将数据发送到您的 Splunk Cloud 实例。

要在 Splunk Web 中添加基于 WMI 的输入，请使用“远程时间日志监视”和“远程性能监视”数据导入。有关说明，请参阅以下主题：

- 使用 Splunk Web 配置远程 Windows 性能监视。
- 配置远程 Windows 事件日志监视

使用配置文件配置基于 WMI 的输入

Splunk Cloud 无法在本机收集 WMI 数据，因此如果您不想使用重型转发器，则必须使用通用转发器并修改配置文件。

wmi.conf 配置文件通过 WMI 处理远程数据收集。请参阅此文件以查看基于 WMI 的输入的默认值。如果要更改默认值，请将要更改的设置和值添加到文件的单独版本中，路径为收集 WMI 数据的通用转发器或重型转发器上的 `%SPLUNK_HOME%\etc\system\local\` 目录中。请仅为给定数据导入类型设置您想要更改的设置的值。

请参阅*《管理员手册》*中的 wmi.conf 文件以了解该文件的默认设置和值；有关配置文件的信息,请参阅“关于配置文件”。

wmi.conf 文件包含多个段落：

- 指定全局 WMI 设置的 `[settings]` 段落。
- 一个或多个特定于输入的段落，用于定义如何连接到 WMI 提供程序以从远程 Windows 计算机获取数据。

全局设置

`[settings]` 段落指定全局 WMI 参数。整个段落以及其中的每个设置都是可选的。如果此段落不显示，Splunk 平台会采用系统默认值。

以下设置控制在发生错误的情况下 Splunk 平台如何重新连接到给定 WMI 提供程序。

设置	描述	默认值
initial_backoff	第一次在发生错误之后到尝试重新连接到 WMI 提供程序之前需等待的时间长度（以秒为单位）。如果继续发生连接错误，Splunk 平台将把等待时间延长至两倍，直到达到 max_backoff 中指定的值。	5
max_backoff	调用 max_retries_at_max_backoff 前执行连接尝试之间的等待时间，以秒为单位。	20
max_retries_at_max_backoff	如果连接尝试之间的等待时间达到 max_backoff，则每隔 max_backoff 秒尝试重新连接至提供程序的次数。如果 Splunk 平台仍遇到错误，则它会放弃连接并会在您重新启动之后再次尝试连接到有问题的提供程序。它将继续记录错误，如上面显示的示例。	2
checkpoint_sync_interval	等待状态数据（事件日志检查点）写入磁盘的时间长度（以秒为单位）。	2

特定于输入的设置

特定于输入的段落将 Splunk 平台配置为与 WMI 提供程序连接。这些段落由用于指定 Splunk 平台应收集的数据类型的两个设置之一定义。段落名称可以为任意内容，但通常以 `WMI:` 开头，例如：

```
[WMI:AppAndSys]
```

在 Splunk Web 中配置基于 WMI 的输入时，Splunk 平台会对特定于输入的段落标题使用此命名约定。

您可以在特定于输入的段落中指定两种数据导入类型中的其中一种类型：

数据导入	描述
事件日志	event_log_file 设置将 Splunk 平台配置为从段落定义的来源获取事件日志数据。
Windows 查询语言（WQL）	wql 设置将 Splunk 平台配置为从 WMI 提供程序获取数据。您还必须指定有效的 WQL 语句。当您收集性能数据时，您必须使用该设置。

切勿在一个段落中同时定义这两个设置。只能使用其中的一个属性。否则由该段落定义的输入将不运行。

这两种输入类型通用的设置包括：

设置	描述	默认值
server	要从中获取数据的逗号分隔的 Windows 计算机列表。如果缺少此设置，Splunk 平台会尝试连接到本地计算机。	本地主机
interval	轮询 WMI 提供程序以获取新数据的频率（以秒为单位）。如果此设置不存在或未定义，则段落所定义的输入将不运行。	不适用
disabled	启用还是禁用此设置。设置为 1 时禁用该输入；设置为 0 时启用该输入。	0（启用）

特定于事件日志的参数包括：

设置	描述	默认值
event_log_file	要监视的逗号分隔的事件日志通道列表。	不适用
current_only	是否收集仅在 Splunk 平台运行时发生的事件。如果 Splunk 平台停止时产生了事件，则当 Splunk 平台再次启动时它将不会尝试为这些事件建立索引。请设置为 1 以收集仅在其运行期间发生的事件，请设置为 0 以收集所有事件。	0（收集所有事件）
disable_hostname_normalization	请不要对从 WMI 事件检索的主机名进行规范化。默认情况下，Splunk 平台会将主机名规范化，即针对本地系统标识各种等效的主机名，从而为主机生成单一名称。将此参数设置为 1 时禁用将事件中的主机名规范化；设置为 0 时规范化事件中的主机名	0（规范化 WMI 事件中的主机名）

特定于 WQL 的参数包括：

设置	描述	默认值
wql	有效 WQL 语句。	不适用
namespace	（可选）指定 WMI 提供程序的路径。本地计算机必须能够使用委派的验证连接到远程计算机。如果您未指定远程计算机的路径，Splunk 平台将连接至默认的本地命名空间（\Root\CIMV2）。此默认命名空间是您可能查询的大部分提供程序所驻留的位置。	\\<local server>\Root\CIMV2
current_only	是否需要事件通知查询。有关其他信息，请参阅本主题中的“WQL 查询类型：事件通知与标准”。将此设置设置为 1 时指示 Splunk 平台需要事件通知查询；设置为 0 时需要标准查询。	0（需要标准查询）

WQL 查询类型：事件通知与标准

WQL 段落中的 current_only 设置决定段落收集基于 WMI 的数据时预计使用的查询类型。当您将该设置配置为 1 时，该段落应使用事件通知数据。事件通知数据是指告警您有传入事件的数据。要获取事件通知数据，必须使用事件通知查询。

例如，要了解远程主机何时衍生进程，您必须使用事件通知查询。标准查询没有可通知您发生了事件的实用工具，该查询只能从已经存在的信息返回结果。

反过来，如果您希望了解系统中哪些已在运行的进程是以 "splunk" 单词开头的，则必须使用标准查询。事件通知查询无法告诉您静态和预先存在的信息。

事件通知查询要求为段落定义的 WQL 语句在结构和语法上必须是正确无误的。WQL 格式不正确将会导致段落所定义的输入无法运行。有关具体详细信息和示例，请参阅 wmi.conf 配置文件参考。

WQL 查询段落不更新 WMI 检查点文件

当您通过 WMI 使用 WQL 查询段落收集数据时，Splunk 平台不会更新 WMI 检查点文件，即确定是否已对 WMI 数据建立检索的文件。任何类型的 WQL 查询返回动态数据，因此不可以构建为已生成数据保存检查点的上下文。因此，每次段落运行时，Splunk 平台将其通过 WQL 查询段落收集作为全新数据的 WMI 数据建立索引。此过程可导致对重复事件建立索引，且可能

影响许可量。

如果您需要定期对数据建立索引（如事件日志），则在通用转发器上使用适当的监视器。如果您必须使用 WMI，请使用标准的 WMI 查询类型。

wmi.conf 示例

以下是 wmi.conf 文件的一个示例：

```
[settings]
initial_backoff = 5
max_backoff = 20
max_retries_at_max_backoff = 2
checkpoint_sync_interval = 2

[WMI:AppAndSys]
server = foo, bar
interval = 10
event_log_file = Application, System, Directory Service
disabled = 0

[WMI:LocalSplunkWmiProcess]
interval = 5
wql = select * from Win32_PerfFormattedData_PerfProc_Process where Name = "splunk-wmi"
disabled = 0

# Listen from three event log channels, capturing log events that occur only
# while the Splunk platform runs.  Gather data from three machines.
[WMI:TailApplicationLogs]
interval = 10
event_log_file = Application, Security, System
server = srv1, srv2, srv3
disabled = 0
current_only = 1

# Listen for process-creation events on a remote machine
[WMI:ProcessCreation]
interval = 1
server = remote-machine
wql = select * from __InstanceCreationEvent within 1 where TargetInstance isa 'Win32_Process'
disabled = 0
current_only = 1

# Receive events whenever someone plugs/unplugs a USB device to/from the computer
[WMI:USBChanges]
interval = 1
wql = select * from __InstanceOperationEvent within 1 where TargetInstance ISA 'Win32_PnPEntity' and
TargetInstance.Description='USB Mass Storage Device'
disabled = 0
current_only = 1
```

WMI 数据的字段

为来自输入（基于 WMI）的数据新建索引时，Splunk 平台不仅会设置从中接收数据的原始主机，还会把已接收事件的**来源**设置为 wmi。它将根据以下条件设置传入事件的**来源类型**：

- 对于事件日志数据，Splunk 平台会把来源类型设置为 WinEventLog:<name of log file>。例如，WinEventLog:Application。
- 对于 WQL 数据，Splunk 平台会将来源类型设置为定义输入的段落的名称。例如，对于名为 [WMI:LocalSplunkdProcess] 的段落，Splunk 会把来源类型设置为 WMI:LocalSplunkdProcess。

WMI 和事件转换

WMI 事件不可在索引时间转换。您无法在 Splunk 平台为 WMI 事件建立索引期间修改或提取这些事件。这是因为 WMI 事件是以单一来源形式出现的（脚本式输入），这意味着它们只能作为单一来源进行匹配。

您可以在搜索时间修改和提取 WMI 事件。也可以通过指定 sourcetype [wmi] 在分析时处理基于 WMI 的输入。

有关如何在事件到达 Splunk 平台时转换事件的信息，请参阅本手册中的“关于索引字段提取”。

WMI 输入故障排除

如果 Splunk 平台无法连接到定义的 WMI 提供程序，它会生成如下错误：

```
05-12-2011 02:39:40.632 -0700 ERROR ExecProcessor - message from ""C:\Program Files\Splunk\bin\splunk-wmi.exe"" WMI - Unable to connect to WMI namespace "\\w2k3m1\root\cimv2" (attempt to connect took 42.06 seconds) (error="The RPC server is unavailable." HRESULT=800706BA)
```

发生这种情况时，请尝试以下提示操作：

- 确认轮询实例的计算机可以通过网络访问该计算机。
- 确认您已将实例配置为对 WMI 提供程序的最小访问权限，并已完成所有安全要求。
- 确认 Splunk 平台 Windows 用户有权访问 WMI

如果您在通过 WMI 提供程序接收事件时遇到问题或者您未获得预期的结果，请参阅*《故障排除手册》*中“Splunk 和 WMI 的常见问题”。

监视 Windows 注册表数据

Windows 注册表是 Windows 计算机上的中央配置数据库。几乎所有 Windows 进程和第三方程序都与其进行交互。如果注册表运行不正常，则 Windows 会无法运行。Splunk 平台支持捕获 Windows 注册表设置，并允许您实时监视注册表更改。

当某个程序对配置进行更改时，该程序会将这些更改写入注册表。当该程序再次运行时，它会浏览注册表以读取这些配置。您可以了解 Windows 上的程序和进程添加、更新及删除注册表项的情况。当某个注册表项发生更改时，Splunk 平台会捕获执行更改的进程的名称以及所更改项的整个路径。

如果您使用的是 Splunk Cloud，必须在 Windows 计算机上安装通用转发器从“Windows 注册表”中收集数据并把数据转发到您的 Splunk Cloud 部署。

监视注册表的原因

很多程序和进程始终对它执行读写操作。当某些项目不正常执行时，Microsoft 通常会指示管理员和类似的用户使用 RegEdit 工具直接更改注册表。实时捕获这些编辑以及所有其他更改的功能是了解注册表重要性的第一步。

注册表的运行状况十分重要。Splunk 平台会通知您注册表发生了更改，也还会通知您这些更改是否取得了成功。如果程序和进程无法对注册表执行读写操作，则发生了系统故障。Splunk 平台可以告警您与注册表的交互出现问题，这样您就可以从备份中恢复注册表并保持系统持续运行。

监视注册表的要求

下表列出了监视注册表所需的显式权限。根据要监视的注册表项，您可能还需要其他权限。

活动	所需权限
监视注册表	• Splunk Enterprise 或通用转发器必须在 Windows 上运行。 • Splunk 平台实例必须以本地系统用户的身份运行，或以域用户身份运行且必须具有您想要监视的“注册表”配置单元或项的读取访问权限。

性能注意事项

启用“注册表”监视时，您需指定要监视的“注册表”配置单元：用户配置单元（在 RegEdit 中表示为 HKEY_USERS）或计算机配置单元（表示为 HKEY_LOCAL_MACHINE）。用户配置单元包含 Windows 和程序所需的用户特定配置，而计算机配置单元包含特定于计算机的配置信息，例如服务、驱动程序、对象类及安全描述符的位置。

由于注册表在 Windows 计算机的操作中扮演中心角色，因此同时启用两个注册表路径会导致 Splunk 平台需要监视大量数据。为实现最佳性能，通过配置 inputs.conf 配置文件筛选平台需要建立索引的注册表数据量。

您可以在第一次启动 Splunk 平台以及每次超过指定时间量后重新启动它时捕获 Windows 注册表当前状态的基准快照。此快照使您可以比较注册表在某一时间点时的状态，并可更加轻松地跟踪注册表随时间的更改。

快照进程会占用大量 CPU，可能需要数分钟才能完成。您可以等到将注册表项范围缩小到希望 Splunk 平台监视的特定范围之后，再获取基准快照。

使用配置文件启用注册表事件监视

由于 Windows 注册表几乎一直都在被使用，因此它会生成大量事件。这可能会导致许可授权问题。Splunk 注册表监视功能每天可生成数百 MB 的数据。

Splunk Windows 注册表监视功能使用配置文件 inputs.conf 来确定要在系统上监视的内容。此文件必须驻留在运行注册表监视的计算机上的 %SPLUNK_HOME%\etc\system\local\ 中。

inputs.conf 文件包含您专为改进和筛选希望 Splunk 平台监视的注册表配置单元路径而创建的正则表达式。

inputs.conf 文件中的每个段落代表其定义包含以下属性的特定过滤器：

属性	描述
proc	包含要监视的一个或多个进程的路径的正则表达式。默认值：.* 或所有进程。
hive	包含您要监视的一个或多个条目的配置单元路径的正则表达式。Splunk 支持 Windows 中预定义的根键值映射： • \\REGISTRY\\USER\\ 映射到 HKEY_USERS 或 HKU • \\REGISTRY\\USER_Classes 映射到 HKEY_CLASSES_ROOT 或 HKCR • \\REGISTRY\\MACHINE 映射到 HKEY_LOCAL_MACHINE 或 HKLM • \\REGISTRY\\MACHINE\\SOFTWARE\\Classes 映射到 HKEY_CLASSES_ROOT 或 HKCR • \\REGISTRY\\MACHINE\\SYSTEM\\CurrentControlSet\\Hardware Profiles\\Current 映射到 HKEY_CURRENT_CONFIG 或 HKCC • 由于“注册表”监视器在内核模式中运行，所以没有 HKEY_CURRENT_USER 或 HKCU 的直接映射。使用 \\REGISTRY\\USER\\.* 末尾处的句号和星号是必需的。使用此映射以生成包含登录用户的安全标识符 (SID) 的事件。 • 除使用前面说明的映射之外，您还可以通过使用 \\REGISTRY\\USER\\<SID>（其中 SID 是用户的 SID）来指定您想要监视其“注册表项”的用户。
type	要监视的事件类型的子集。此子集可以是一个或多个 delete, set, create, rename, open, close 或 query。此属性的值必须为 event_types（您在 inputs.conf 中对其进行设置）值的子集。
baseline	是否捕获该特定配置单元路径的基准快照。设置为 1 时代表是；设置为 0 时代表否。
baseline_interval	从获取上一个基准开始必须经过多长时间（以秒为单位）Splunk 平台才能获取另一个启动基准。例如，如果您将 baseline_interval 设为 600，那么当 Splunk 平台启动或重新启动时，如果现有基准超过 600 秒，则会采用现有基准。如果没有基准，那么 Splunk 平台会立即采用基准。如果您也没有将 baseline 设为 1，那么此设置也没有任何影响。默认值为 86,400 秒（1 天）。
disabled	是否启用筛选器。设置为 1 时禁用筛选器；设置为 0 时启用筛选器。

在 Splunk Web 中启用注册表监视

您可以使用 Splunk Web 在 Splunk Enterprise 实例上配置 Windows 注册表监视。

转到“新增”页面

可通过两种方式访问此页面：

- Splunk 主页
- Splunk 设置

通过 Splunk 设置：

1. 单击 Splunk Web 右上角的**设置**。
2. 请单击**数据导入**。
3. 单击**注册表监视**。
4. 单击**新建**以添加输入。

通过 Splunk 主页：

1. 单击 Splunk 主页中的**添加数据**链接。
2. 单击**监视**以监视本地 Windows 计算机的“注册表”数据。

选择输入来源

1. 查找并选择**注册表监视**。
2. 在**集合名称**字段中，为该输入输入一个您可以记住的唯一名称。
3. 在**注册表配置单元**字段中，输入希望 Splunk 平台监视的注册表项的路径。如果您计划监视多个配置单元，则每个配置单元都需要有自己单独的输入。

如果您不确定此路径，请单击**浏览**按钮以选择您希望 Splunk 平台监视的注册表项路径。**注册表配置单元**窗口即会打开并以树视图形式显示注册表。配置单元、项和子项显示为文件夹，值显示为文档图标。HKEY_USERS, HKEY_CURRENT_USER, HKEY_LOCAL_MACHINE, 和 HKEY_CURRENT_CONFIG 配置单元显示为顶级对象。受 HKEY_CLASSES_ROOT 配置单元第一子级中出现的子项数量影响，该配置单元不会显示。要访问 HKEY_CLASSES_ROOT 项目，请选择 HKEY_LOCAL_MACHINE\Software\Classes。

4. 在**注册表配置单元**窗口中，单击所需注册表项的名称。该限定的键名即会显示在窗口底部的**限定名**字段中。
5. 单击**选择**以确认所做选择并关闭窗口。
6. （可选）如果您希望监视起始配置单元下方的子节点，请选择**监视子节点**。

   监视子节点节点决定了 Splunk 平台添加到 inputs.conf 文件中的内容，该文件是 Splunk 平台在您于 Splunk Web 中定义“注册表”监视器输入时新建的。

   如果您使用树视图选择要监视的项或配置单元，并且已选中**监视子节点**，则 Splunk 平台会向所定义输入对应的段落添加一个**正则表达式**。该正则表达式（\\\\?.*）用于筛选未直接引用选定项或其任意子项的事件。

   如果未选中**监视子节点**，则 Splunk 平台会向输入段落添加一个正则表达式，以筛选出未直接引用选定项的事件（包括引用选定项的子项的事件）。

   如果您未使用树视图指定要监视的项，则只有在选中**监视子节点**并且您没有在**注册表配置单元**字段中输入您自己的正则表达式时，Splunk 平台才会添加此正则表达式。

7. 在**事件类型**下方，针对选定注册表配置单元选择希望 Splunk 平台监视的注册表事件类型：

事件类型	描述
Set	如果程序对某个注册表子项执行 SetValue 方法，从而设置了一个值或覆盖现有注册表项中的现有值，Splunk 平台会生成 Set 事件。
Create	如果程序在某个注册表配置单元中执行 CreateSubKey 方法，从而在现有注册表配置单元内新建了一个子项，Splunk 平台会生成 Create 事件。
Delete	当程序执行 DeleteValue 或 DeleteSubKey 方法时，Splunk 平台会生成 Delete 事件。此方法会删除现有特定项的值，或者删除现有配置单元中的项。
Rename	当您在 RegEdit 中重命名某个注册表项或子项时，Splunk 平台会生成 Rename 事件。
Open	当程序对某个注册表子项执行 OpenSubKey 方法时，Splunk 平台会生成 Open 事件，例如，当程序需要注册表中所含的配置信息时的情况。
Close	当程序对某个注册表项执行 Close 方法时，Splunk 平台会生成 Close 事件。这会发生在程序完成读取某个项的内容时，或者您在 RegEdit 中更改某个项的值并退出值输入窗口之后。
Query	当程序对某个注册表子项执行 GetValue 方法时，Splunk 平台会生成 Query 事件。

8. 在**进程路径**字段中输入相应值，您便可指定 Splunk 平台监视哪些进程对注册表的更改。或者，保留 .* 的默认值来监视所有进程。
9. 请指定您是否希望在监视注册表更改之前获取整个注册表的基准快照。要设置基准，请在**基准索引**下方单击**是**。

   基准快照是在获取快照时整个注册表的索引。快照中的注册表保留了原始的索引时间戳。

   </caution>扫描注册表以设置基准索引这一进程会占用大量 CPU 且可能需要一些时间。</caution>
10. 单击**下一步**。

指定输入设置

输入设置页面允许您指定应用程序上下文、默认主机值和索引。所有这些参数均为可选参数。

1. 为此输入选择相应的**应用程序上下文**。
2. 在**主机**字段中，设置主机名称值。

   此字段只是设置生成事件中的**主机**字段。设置此值不会引导 Splunk 平台查找网络中的特定主机。

3. 在**索引**字段中，设置 Splunk 平台应将数据发送到其中的索引。如果未定义多个索引来处理不同类型的事件，请保留“默认”值。除了用户数据的索引之外，Splunk 平台还有很多实用工具索引，这些索引也会显示在此下拉列表框中。
4. 请单击**查看**。

查看您的选择

在您指定所有输入设置后，可查看您的选择。Splunk 平台会列出您选定的所有选项，包括监视器的类型、数据来源、来源类型、应用程序上下文和索引。

1. 查看该设置。
2. 如果这些设置不符合您的期望，单击 < 即可返回到向导中的上一个步骤。
3. 单击**提交**。

然后，Splunk 平台会加载“成功”页面并开始索引指定的“注册表”节点。

查看注册表更改数据

要查看 Splunk 平台已为其新建索引的“注册表”更改数据，请转到“搜索”应用，并搜索来源为 WinRegistry 的事件。例如，当用户登录到某个域时组策略会生成如下所示的事件：

```
3:03:28.505 PM
06/19/2011 15:03:28.505
event_status="(0)The operation completed successfully."
pid=340
process_image="c:\WINDOWS\system32\winlogon.exe"
registry_type="SetValue"
key_path="HKLM\SOFTWARE\Microsoft\Windows\CurrentVersion\Group Policy\History\DCName"
data_type="REG_SZ"
data="\\ftw.ad.splunk.com"
```

每个注册表监视事件包含以下属性：

设置	描述
event_status	尝试更改注册表的结果。此结果将是“(0) 操作成功完成”。否则说明注册表可能存在一些问题，可能需要从备份中进行恢复。
pid	尝试更改注册表的进程的进程 ID。
process_image	尝试更改注册表的进程的名称。
registry_type	process_image 尝试调用的“注册表”操作类型。
key_path	process_image 尝试更改的“注册表”项路径。
data_type	对“注册表”执行更改的 process_image 尝试获取或设置的“注册表”数据类型。
data	对“注册表”执行更改的 process_image 尝试读取或写入的数据。

获取基准快照

当您启用“注册表”监视功能时，您可在下一次 Splunk 平台启动时记录注册表配置单元的基准快照。根据默认设置，此快照涵盖 HKEY_CURRENT_USER 和 HKEY_LOCAL_MACHINE 配置单元。它还新建了何时重新获取快照的时间线。默认情况下，如果基准超过 24 小时，那么 Splunk 平台重启时，会重新获取基准快照。您可以为 inputs.conf 中的所有筛选器自定义该值，只要设置好 baseline_interval 的值（单位为秒）。

当您新建基准快照时，快照会使用注册表数据的索引时间而不是快照新建时间。例如，如果注册表项两年前发生更改，该事件的时间戳将是两年前，而不是基准快照新建时间。

监视 Windows 性能

Splunk® Enterprise 支持实时监视所有 Windows 性能计数器，其中包括对本地和远程性能数据收集的支持。

利用性能监视输入，您可以在 Web 界面中提供“性能监视器”的功能。Splunk 平台使用 Windows 性能数据助手 (PDH) API 在本地 Windows 计算机上实现性能计数器查询。

可用于平台的性能对象、计数器和实例类型取决于计算机上的性能库。Microsoft 和第三方供应商均提供了包含性能计数器的库。有关性能监视的信息，请在 Microsoft 文档网站上搜索“性能计数器”。

要将性能监视器数据导入 Splunk Cloud，您必须在您想从中收集性能指标的 Windows 计算机上运行 Splunk Enterprise 重型转发器或通用转发器，然后将该数据转发到 Splunk Cloud 实例。Splunk Enterprise 的完整实例和通用转发器都支持本地收集性能指标。可通过 Windows Management Instrumentation (WMI) 实现远程性能监视，但这要求 Windows 计算机上的 Splunk 平台实例以具有相应 Active Directory 凭据的用户身份运行。

在 Splunk Enterprise 和通用转发器上，性能监视器输入将以进程的形式运行，该进程称为 splunk-perfmon.exe。此进程将以输入中指定的时间间隔，针对您定义的每个输入运行一次。您可以在 Splunk Web 中或使用配置文件来配置性能监视。

性能监视器输入使用两个文件进行配置。用于配置输入的文件取决于您是要从本地实例还是从远程实例获取性能数据：

- 使用 inputs.conf 配置文件获取本地性能数据。
- 使用 wmi.conf 配置文件从远程计算机获取性能数据。

为什么监视性能指标？

性能监视是 Windows 管理员工具包的一个重要部分。Windows 会生成许多与计算机运行状况相关的数据。正确分析该数据有助于区分计算机是处于正常运行状态还是已经经历了停机。

您需要什么来监视性能计数器

下表列出了监视 Windows 中的性能计数器所需的最低要求。根据要监视的性能对象或计数器，您可能需要其他权限。

有关性能指标监视要求的其他信息，请参阅本主题后面的“安全和远程访问权限注意事项”。

活动	所需权限
监视本地性能指标	* Splunk Cloud 实例必须从转发器接收性能数据。 *转发器必须在 Windows 上运行。请参阅 Splunk Enterprise*《安装手册》*中的“在 Windows 上安装”。 *转发器必须以本地系统 Windows 用户身份运行。Splunk Enterprise*《安装手册》*中的“选择 Splunk Enterprise 应以其身份运行的 Windows 用户”。
通过 WMI 监视其他计算机上的远程性能指标	* Splunk Cloud 实例必须从转发器接收性能数据。 *转发器必须在 Windows 上运行。 *转发器必须以至少具有目标计算机上 WMI 读取访问权限的域或远程用户身份运行。 *转发器必须以具有目标计算机上“性能数据助手”库适当访问权限的域或远程用户身份运行。

安全与远程访问注意事项

在可能的情况下，使用通用转发器将远程计算机中的性能数据发送到 Splunk Cloud 或 Splunk Enterprise 索引器。

Splunk Enterprise 可使用转发器或 WMI 从远程计算机收集数据。

如果您将转发器安装在要收集性能数据的远程 Windows 计算机上，您可以将转发器作为本地系统用户安装在这些计算机上。本地系统用户对本地计算机上的所有数据都具有访问权限，但对远程计算机上的数据没有访问权限。

如果您希望 Splunk Enterprise 使用 WMI 从远程计算机获取性能数据，那么您必须配置 Splunk Enterprise 和您的 Windows 网络。您无法以本地系统用户身份安装 Splunk Enterprise，您选择的用户决定 Splunk Enterprise 可读取的性能监视器对象。

使用有效用户身份安装 Splunk Enterprise 之后，必须先将该用户添加到以下组，然后再启用本地性能监视器输入：

- 性能监视器用户（域组）
- 性能日志用户（域组）

要了解关于 WMI 安全的信息，请参阅“通过 Windows Management Instrumentation (WMI) 监视数据”主题中的“安全与远程访问权限注意事项”。要了解如何使用通用转发器，请参阅 Splunk 通用转发器*《转发器》*手册中的“通用转发器”。

启用本地 Windows 性能监视

在 Splunk Cloud 上，必须从您想要从中收集性能数据的 Windows 计算机转发数据。

在 Splunk Enterprise 中，您可以直接在 Splunk Web 中或使用配置文件来配置本地性能监视。

Splunk Web 是在 Splunk Enterprise 实例上添加性能监视数据导入的首选方法。配置文件中可能会发生拼写错误，并且按照性能监视器 API 定义，精确指定性能监视器对象很重要。有关完整说明，请参阅本主题稍后的“关于在 inputs.conf 中指定性能监视器对象的重要信息”。

使用 Splunk Web 配置本地 Windows 性能监视

您只能在 Splunk Enterprise 实例上使用 Splunk Web 收集 Windows 性能监视指标。

要开始配置 Windows 性能监视指标，请通过 Splunk 设置或 Splunk 主页访问 Splunk Web 中的“新增”页面。

要通过 Splunk 设置连接 Windows 性能监视指标，请执行以下步骤：

1. 单击**设置** > **数据导入**。

2. 单击**本地性能监视**。
3. 单击**新建**以添加输入。
4. 继续执行本主题后面的“选择输入来源”中的步骤。

要通过 Splunk 主页连接 Windows 性能监视指标，请执行以下步骤：

1. 单击 Splunk 主页中的**添加数据**链接。
2. 单击**监视**即可监视来自本地 Windows 计算机的性能数据，或单击**转发**即可接收来自另一台计算机的性能数据。
3. 如果您选择了**转发**，则选择或新建要此输入应用的转发器组。
4. 单击**下一步**。
5. 继续执行本主题后面的“选择输入来源”中的步骤。

选择输入来源

1. 在 Splunk Enterprise 的左窗格中，选择**本地性能监视**。
2. 在**集合名称**字段中，为该输入输入一个您可以记住的唯一名称。
3. 请单击**选择对象**以获取此 Windows 计算机上可用的性能对象列表，然后从该列表中选择您要监视的对象。Splunk Enterprise 显示**选择计数器**和**选择实例**列表框。
4. 在**选择计数器**列表框中，查找您要此输入监视的性能计数器。

   您只能为每个数据导入添加一个性能对象。如果您需要监视多个对象，请为每个对象新建其他数据导入。

5. 单击要监视的每个计数器。Splunk Enterprise 将计数器从**可用计数器**窗口移动到**已选计数器**窗口。
6. （可选）要取消选择一个计数器，在**可用项目**窗口上单击该计数器的名称。
7. （可选）要选择或取消选择所有计数器，请单击**添加所有**或**删除所有**链接。

   选择所有计数器可导致索引大量数据。

8. （可选）在**选择实例**列表框中，通过单击**可用实例**窗口中的实例，选择您要此输入监视的实例。
   选择所有计数器可导致索引大量数据且可能导致许可证违规。
9. 在**轮询间隔**字段中，为输入在轮询尝试之间输入时间（秒）。

   `_Total` 实例是一个特殊实例，许多类型的性能计数器都显示有这个实例。该实例为相同计数器下的所有相关实例的平均值。

10. 单击**下一步**。

指定输入设置

在“输入设置”页面指定应用程序上下文、默认主机值和索引。所有这些参数均为可选参数。

设置此页面上的**主机**只是设置结果事件中的**主机**字段。而不是引导 Splunk Enterprise 查找网络中的特定主机。

1. 在 Splunk Enterprise 中，为**应用程序上下文**字段中的输入选择应用程序上下文。
2. 设置**主机**值。此设置有多个选项供您选择。有关设置主机值的更多信息，请参阅“关于主机”。
3. 设置您希望 Splunk Enterprise 将数据发送到其中的**索引**。如果未定义多个索引来处理不同类型的事件，请保留 `default` 值。
4. 请单击**查看**。

查看您的选择

在您指定输入设置后，请查看您的选择。Splunk Enterprise 列出所有您所选的选项，包括监视器的类型、数据来源、来源类型、应用程序上下文和索引。

1. 查看该设置。
2. 如果它们不符合您的期望，请单击左尖括号（ < ）即可返回到向导中的上一个步骤。否则，请单击**提交**。

然后，Splunk Enterprise 会加载确认页面并开始索引指定的性能指标。有关从文件和目录获取数据的更多信息，请参阅本手册中的“监视 Windows 性能”。

使用配置文件配置本地 Windows 性能监视

inputs.conf 配置文件控制性能监视配置。要使用配置文件设置性能监视，您必须在您想要从中收集性能指标的 Windows 计算机上的 %SPLUNK_HOME%\etc\system\local 中新建或编辑 inputs.conf。如您之前不曾使用过配置文件，请参阅“关于配置文件”。

配置本地 Windows 监视的选项可用于接收转发数据的 Splunk Cloud 实例和 Splunk Enterprise 实例。

`[perfmon://<name>]` 段落在 inputs.conf 中定义性能监视输入。为每个要监视的性能对象指定一个段落。

在每个段落中，您可以指定以下设置：

设置	是否必需？	描述
interval	是	轮询新数据的频率（以秒为单位）。如果该设置不存在，输入将每隔 300 秒（5 分钟）运行一次。
object	是	您希望捕获的性能对象。指定与现有性能监视器对象的名称完全匹配的字符串，或者使用正则表达式来引用多个对象。如果此设置不存在或未定义，则输入无法运行，因为没有默认值。
counters	是	与在 object 设置中指定的对象相关的一个或多个有效性能计数器。使用分号分隔多个计数器。还可以使用星号 (*) 指定某给定 object 下所有可用的计数器。如果此设置不存在或未定义，则输入无法运行，因为没有默认值。
instance	否	与在 counters 设置中指定的性能计数器相关的一个或多个有效实例。使用分号分隔多个实例。请使用星号 (*) 指定所有实例，如果您未在段落中定义该设置，则默认使用星号。
index	否	要将性能计数器数据发送到的索引。如果未定义此设置，则使用默认索引。
disabled	否	是否收集在此输入中定义的性能数据。将此设置设置为 1 时禁用此段落；设置为 0 时启用此段落。如果未定义该设置，则默认为 0。

下表显示了高级选项：

设置	是否必需？	描述
showZeroValue	否	Splunk Enterprise 是否应该收集值为零的事件。 将此设置设置为 1 则收集零值事件，设置为 0 则忽略零值事件。如果不存在，则默认为 0。
samplingInterval	否	Splunk Enterprise 收集性能数据的频率（以毫秒为单位）。 启用高频率性能取样。当您启用高频率性能取样时，Splunk Enterprise 将在每个一定的时间间隔收集一次性能数据，并报告数据平均值以及其他统计信息。默认值为 100 毫秒，必须低于您使用 interval 设置指定的值。
stats	否	Splunk Enterprise 报告的高频率性能取样统计值的分号分隔列表。 允许的值包括 average、min、max、dev 和 count。 默认情况下无设置。
mode	否	启用高性能取样时，此设置控制 Splunk Enterprise 输出事件的方式。 允许的值包括 single、multikv、multiMS 和 multikvMS。 当您启用 multiMS 或 multikvMS 时，Splunk Enterprise 将为其收集的每个性能指标输出两个事件。第一个事件是平均值，第二个事件是统计信息事件。根据您使用的输出模式，统计信息事件包含一种特殊的来源类型：对于 multiMS 模式为perfmonMSStats；对于 multikvMS 模式为 perfmonMKMSStats。 如您未启用高性能取样，multikvMS 输出模式将和 multikv 输出模式一样。 默认值为 single。
useEnglishOnly	否	控制 Splunk Enterprise 在区域为非英语的系统上索引性能指标的方式。具体地说，此设置指示 Splunk Enterprise 在不使用英语的主机上索引性能指标时要使用哪个 Windows 性能监视器 API。 如果设置为 true，则无论系统区域为何种语言，Splunk Enterprise 均会以英语形式收集性能指标。它使用 **PdhAddEnglishCounter()** API 添加计数器字符串。它还禁用正则表达式和与对象和计数器字符串匹配的通配符。 如果设置为 false，Splunk Enterprise 将收集以系统语言呈现的性能指标，并期望您配置以该语言呈现的 object 和 counter 设置。它使用 **PdhAddCounter()** API 添加计数器字符串。您可以使用通配符和正则表达式，但必须指定有效的 object、counters 和 instances 值，这些值特定于操作系统的区域设置。 默认值为 false。

useWinApiProcStats	否	启用后，性能监视器输入中的 useWinApiProcStats 设置使用进程内核模式和用户模式时间来计算进程的 CPU 使用情况。目前，输入使用标准的性能数据助手（PDH）API 计算进程的 CPU 使用情况。 将此设置配置为 true 之后，输入使用核心 Windows API 中的 GetProcessTime() 函数计算进程的 CPU 使用情况，适用于以下性能监视器计数器： • 处理器时间 • 用户时间 • 特权时间 最好的做法是在多核 Windows 计算机上启用 useWinApiProcStats 功能。 确认系统中处理器的数量。可在 Windows 部署中的“系统信息”页面验证内核的总数。 您的系统中的各处理器内核等于最大的处理器性能百分比。因此，对于多核 Windows 部署中的每个核，将 100% 乘以最大的可用处理器性能总百分数。例如，8 核的 Window 环境的最大处理容量是 800%。 通过导航到**数据导入 > 本地性能监视 > 选择系统**，可以在 Splunk Enterprise 部署中验证总处理器容量。 此设置使用的 API 只有英文版。如果您的 Windows 计算机使用的是非英语系统区域，您还必须将 useEnglishOnly 设为 true。 有关计算 Windows 多核计算机上 CPU 使用情况的详细信息，请参阅发行说明中的“性能监视器输入显示多核 Microsoft Windows 计算机上进程的最大使用率百分比”。
formatString	否	控制 Splunk Enterprise 如何格式化性能计数器事件的浮点值输出。 Windows 经常将性能计数器事件打印为浮点值。当未格式化时，则事件以小数点右边的所有有效数字打印。formatString 设置控制着作为每个事件之一部分打印的有效数字的数量。 该设置使用来自 C++ 语言 printf 函数的格式说明符。根据您想要输出事件文本的方式，该功能包括许多种说明符。 当指定格式时，请不要在格式两边使用引号（""）。请仅指定需要以您想要的方式用来格式化字符串的有效字符。 默认值为 %.20g。

无论系统区域为何种语言，均以英语形式收集性能指标

即使 Splunk Enterprise 在其上运行的系统不使用英语语言，您还可以英语形式收集性能指标。

为此，请使用 useEnglishOnly 设置，该设置位于 inputs.conf 中的段落内。无法在 Splunk Web 中配置 useEnglishOnly 设置。

在 inputs.conf 段落中使用 useEnglishOnly 有一些注意事项。请参阅本主题后文中的“注意事项”。

监视性能输入段落的示例

以下是一些显示如何使用 inputs.conf 配置文件监视性能监视对象的示例段落。

```
# Query the PhysicalDisk performance object and gather disk access data for
# all physical drives installed in the system. Store this data in the
# "perfmon" index.
# Note: If the interval setting is set to 0, Splunk resets the interval
# to 1.

[perfmon://LocalPhysicalDisk]
interval = 0
object = PhysicalDisk
counters = Disk Bytes/sec; % Disk Read Time; % Disk Write Time; % Disk Time
instances = *
disabled = 0
index = PerfMon

# Gather SQL statistics for all database instances on this SQL server.
# 'object' setting uses a regular expression "\$.*" to specify SQL
# statistics for all available databases.
[perfmon://SQLServer_SQL_Statistics]
```

```
object = MSSQL\$.*:SQL Statistics
counters = *
instances = *

# Gather information on all counters under the "Process" and "Processor"
# Perfmon objects.
# We use '.*' as a wild card to match the 'Process' and 'Processor' objects.
[perfmon://ProcessandProcessor]
object = Process.*
counters = *
instances = *

# Collect CPU processor usage metrics in English only on a French system.
[perfmon://Processor]
object = Processor
instances = _Total
counters = % Processor Time;% User Time
useEnglishOnly = 1
interval = 30
disabled = 0

# Collect CPU processor usage metrics in the French system's native locale.
# Note that you must specify the counters in the language of that locale.
[perfmon://FrenchProcs]
counters = *
disabled = 0
useEnglishOnly = 0
interval = 30
object = Processeur
instances = *

# Collect CPU processor usage metrics. Format the output to two decimal places only.
[perfmon://Processor]
counters = *
disabled = 0
interval = 30
object = Processor
instances = *
formatString = %.20g
```

关于在 inputs.conf 文件中指定性能监视器对象的重要信息

当您使用 inputs.con 配置文件配置 Windows 性能监视器输入时，必须特别注意确保该文件包含正确的输入语法，否则 Splunk 平台将无法正确为数据创建索引。

指定 perfmon 关键字时全部使用小写形式

在 inputs.conf 中新建性能监视器输入时，您必须对 perfmon 关键字全部使用小写形式。请参阅以下示例：

正确

```
[perfmon://CPUTime]
```

不正确

```
[Perfmon://CPUTime]
[PERFMON://CPUTime]
```

如果此关键字采用大写形式或大小写混合形式，Splunk 平台会警告启动问题，并且指定性能监视器输入将不运行。

指定有效正则表达式以捕获多个性能监视器对象

要在单个性能监视器段落中指定多个对象，您必须使用有效的正则表达式来捕获这些对象。例如，要指定通配符以匹配超出了特定字符数的字符串，请勿使用星号 (*)，而应使用句点后跟星号 （ .*）。如果对象包含美元符号或类似的特殊字符，您可能需要使用反斜杠 （ \ ） 对其进行转义。

如果您不使用正则表达式，则值必须与“性能监视器 API”中的值完全匹配

指定 object、counters 和 instances 设置的值（在 [perfmon://] 段落中）时，请确认这些值与定义于“性能监视器 API”中的值完全匹配，包括大小写，否则输入可能返回错误的数据，甚至完全不返回任何数据。如果输入无法匹配您指定的性能对象、计数器或实例值，则它会将该故障记录到 splunkd.log 文件中。请参阅以下返回失败示例：

```
01-27-2011 21:04:48.681 -0800 ERROR ExecProcessor - message from ""C:\Program Files\Splunk\bin\splunk-perfmon.exe" -noui"
splunk-perfmon - PerfmonHelper::enumObjectByNameEx: PdhEnumObjectItems failed for object - 'USB' with error (0xc0000bb8): The
specified object is not found on the system.
```

使用 Splunk Web 来添加性能监视器数据导入，以确保您正确添加它们。

通过 WMI 启用远程 Windows 性能监视

您可以在 Splunk Web 中或使用配置文件来配置远程性能监视。

通过 WMI 收集性能指标时，您必须将 Splunk 平台实例配置为以具有远程收集性能指标的相应访问权限的 Active Directory (AD) 用户身份运行。在尝试收集这些指标之前，您必须执行此操作。运行 Splunk 平台实例的计算机和 Splunk 从中收集性能数据的计算机都必须驻留在相同 AD 域或林中。

为了防止拒绝服务攻击，已将 WMI 设计为自行限制。Splunk 平台还会减少其随时间进行的 WMI 调用次数，以作为防止这些调用返回错误的预防措施。根据您网络的规模、配置和安全配置文件，最好在要从中收集性能指标的主机上安装本地转发器。请参阅本手册中的“确定如何监视远程 Windows 数据的注意事项”。

基于 WMI 的性能值对比性能监视器值

当通过 WMI 收集远程性能指标时，一些指标返回零值，或者返回的值与性能监视器返回的值不一致。对性能监视器计数器实施 WMI 时的限制导致了此问题。这不是 Splunk 平台的问题，也不是 Splunk 平台如何检索基于 WMI 的数据的问题。

WMI 使用 `Win32_PerfFormattedData_*` 数据类来收集性能指标。有关 Win32 类的更多信息，请参阅 https://docs.microsoft.com/en-us/previous-versions//aa394084(v=vs.85)?redirectedfrom=MSDN。

WMI 将这些类的数据结构定义为 32 位或 64 位无符号整数，具体取决于您运行的 Windows 版本。Windows 性能数据助手 (PDH) API 将性能监视器对象定义为浮点变量。浮点变量意味着，由于四舍五入，您可能会看到基于 WMI 的指标处于异常状态。

例如，如果您在通过 WMI 收集 `Win32_PerfFormattedData_PerfDisk_PhysicalDisk\AvgDiskQueueLength` 指标的同时收集有关“平均磁盘队列长度”性能监视器计数器的数据，即使性能监视器指标返回的值大于零但小于 0.5，基于 WMI 的指标也有可能返回零值。这是因为 WMI 会在显示之前先将值四舍五入。

如果您的性能指标需要额外的粒度，在要从中收集性能数据的每个计算机的通用转发器上配置性能监视输入。然后可以将该数据转发到索引器。与使用基于 WMI 的输入远程收集的数据相比，使用此方法检索的数据更加可靠。

使用 Splunk Web 配置远程 Windows 性能监视

此选项仅在 Splunk Enterprise 上可用，在 Splunk Cloud 实例上不可用。您可以在 Splunk Enterprise 中配置通用转发器并将该数据转发到 Splunk Cloud 实例。

在 Splunk Enterprise 中，通过 Splunk 设置或 Splunk 主页转到 Splunk Web 中的“新增”页面。

要通过 Splunk 设置转到“新增”页面，请执行以下步骤：

1. 单击 Splunk Web 右上角的**设置**。
2. 请单击**数据导入**。
3. 单击**远程性能监视**。
4. 单击**新建**以添加输入。

要通过 Splunk 主页转到“新增”页面，请执行以下步骤：

1. 单击 Splunk 主页中的**添加数据**链接。
2. 单击**监视**以监视来自本地 Windows 计算机的性能数据或**转发**以转发来自另一个 Windows 计算机的性能数据。Splunk Enterprise 将加载**添加数据 - 选择来源**页面。

   转发性能数据需要其他设置。

3. 在左窗格中，查找并选择**本地性能监视**。

选择输入来源

`Win32_PerfFormattedData_*` 类在 Splunk Web 中不显示为可用对象。如果想要监视 `Win32_PerfFormattedData_*` 类，您必须直接把这些类直接添加在 `wmi.conf` 文件中。有关更多信息，请参阅“使用配置文件配置远程 Windows 性能监视”。按以下步骤操作：

1. 在 Splunk Enterprise 的左窗格中，选择“本地性能监视”。

2. 在**集合名称**字段中，为该输入输入一个您可以记住的唯一名称。
3. 在**选择目标主机**字段中，输入您要从其中收集性能数据的 Windows 计算机的主机名或 IP 地址。
4. 单击**查询**以获取在**选择目标主机**字段中指定的 Windows 计算机上可用的性能对象列表。
5. 从**选择类**列表中，选择要监视的对象。Splunk Enterprise 显示**选择计数器**和**选择实例**列表框。
6. 在**选择计数器**列表框中，查找您要此输入监视的性能计数器。

   您只能为每个数据导入添加一个性能对象。这应归于 Microsoft 处理性能监视器对象的方式。许多对象枚举可根据选择动态描述自身的类。这会导致无法根据输入的定义理清哪个性能计数器和实例属于哪个对象。如果您需要监视多个对象，请为每个对象新建其他数据导入。

7. 单击要监视的每个计数器。Splunk Enterprise 将计数器从**可用计数器**窗口移动到**已选计数器**窗口。
8. 要取消选择一个计数器，在**可用项目**窗口上单击该计数器的名称。Splunk Enterprise 将计数器从**已选计数器**窗口移动到**可用计数器**窗口。
9. 要选择或取消选择所有计数器，请单击**添加所有**或**删除所有**链接。

   选择所有计数器可导致索引大量数据（可能比您的许可证所允许的数量大）。

10. 在**选择实例**列表框中，通过单击**可用实例**窗口中的实例，选择您要此输入监视的实例。Splunk Enterprise 将实例移动到**所选实例**窗口。
11. 在**轮询间隔**字段中，为输入在轮询尝试之间输入时间（秒）。

    _Total 实例是一个特殊实例，许多类型的性能计数器都显示有这个实例。该实例为相同计数器下的所有相关实例的平均值。针对该实例收集的数据可能与针对相同计数器下的单个实例收集的数据显著不同。

    例如，在具有两个磁盘的系统上，如果您在 PhysicalDisk 对象下方监视“Disk Bytes/Sec”性能计数器的性能数据，则显示的可用实例包括每个物理磁盘（0 C: 和 1 D:）所对应的一个实例以及 _Total 实例，后者是两个物理磁盘实例的平均值。

12. 单击**下一步**。

指定输入设置

在“输入设置”页面指定应用程序上下文、默认主机值和索引。所有这些参数均为可选参数。

设置**主机**值只是设置生成事件中的**主机**字段。而不是引导 Splunk Enterprise 查找网络中的特定主机。

1. 为此输入选择相应的**应用程序上下文**。
2. 设置**主机**值。此设置有多个选项供您选择。有关设置主机值的更多信息，请参阅“关于主机”。
3. 设置 Splunk Enterprise 应将数据发送到其中的**索引**。如果未定义多个索引来处理不同类型的事件，请保留 default 值。
4. 单击**查看**按钮。

查看您的选择

在您指定输入设置后，请查看您的选择。Splunk Enterprise 列出所有您所选的选项，包括监视器的类型、数据来源、来源类型、应用程序上下文和索引。

1. 查看该设置。
2. 如果它们不符合您的期望，请单击左尖括号（ < ）即可返回到向导中的上一个步骤。否则，请单击**提交**。

然后，Splunk Enterprise 会加载确认页面并开始索引指定的性能指标。有关从文件和目录获取数据的更多信息，请参阅本手册中的“监视 Windows 性能”。

使用配置文件配置远程 Windows 性能监视

wmi.conf 配置文件控制远程性能监视配置。要使用配置文件设置远程性能监视，请新建或编辑 wmi.conf，路径为 %SPLUNK_HOME%\etc\system\local。如您之前不曾使用过配置文件，请在开始前阅读“关于配置文件”中的说明。

对于 Splunk Cloud 实例，在您想从中收集性能数据的计算机上安装通用转发器，并配置该转发器以将数据发送到 Splunk Cloud。在 Splunk Enterprise 实例上，请使用 Splunk Web 新建远程性能监视器输入，除非您对其无访问权限。性能监视器对象、计数器和实例的名称必须完全匹配性能监视器 API 中定义的内容（包括大小写）。Splunk Web 使用 WMI 来获取格式正确的名称，消除键入错误的潜在问题。

对于每个您要监视的远程性能监视器对象，wmi.conf 文件中都包含一个对应的段落。在每个段落中，您可以指定以下设置：

全局设置

	是否		默

设置	必需？	描述	默认
initial_backoff	否	在发生错误的情况下，重试连接到 WMI 提供程序之前需等待的时间（以秒为单位）。如果仍然无法连接到提供程序，连接尝试之间的等待时间将加倍，直到连接成功或等待时间大于或等于 max_backoff 设置。	5
max_backoff	否	尝试重新连接到 WMI 提供程序的最长时间（以秒为单位）。	20
max_retries_at_max_backoff	否	两次重新连接 WMI 提供程序的尝试之间的等待时间达到 max_backoff 秒后，继续尝试重新连接该提供程序的次数。	2
checkpoint_sync_interval	否	等待状态数据刷新到磁盘的时间（以秒为单位）。	2

特定于输入的设置

设置	是否必需？	描述	默认
interval	是	轮询新数据的频率（以秒为单位）。如果此设置不存在，则输入无法运行，因为没有默认值。	N/A
server	否	您希望监视其性能的一个或多个有效主机的逗号分隔的列表。	本地计算机
event_log_file	否	要轮询的一个或多个 Windows 事件日志通道的名称。此设置将 Splunk Enterprise 配置为传入数据采用事件日志格式。 请勿在一段落中使用 event_log_file 设置，前提为该段落已包含有 wql 设置。	N/A
wql	否	有效 Windows 查询语言 (WQL) 语句，用于指定您要远程轮询的性能对象、计数器和实例。此设置指示 Splunk Enterprise 从 WMI 提供程序获取数据。 请勿在一段落中使用 wql 设置，前提为该段落已包含有 event_log_file 设置。	N/A
namespace	否	要查询的 WMI 提供程序所驻留在的命名空间。该设置的值既可以是相对值（例如 Root\CIMV2）也可以是绝对值（例如 \\SERVER\Root\CIMV2），但如果您指定了服务器设置，则必须为相对值。 仅在包含 wql 设置的段落中使用 namespace 设置。	Root\CIMV2
index	否	要将性能计数器数据发送到的所需索引。	default
current_only	否	基于 WMI 的事件集合的特性和交互取决于是定义了 wql 设置还是 event_log_file 设置： • 如已定义了 wql，该设置将指示 Splunk Enterprise 其是否应该预期事件通知查询。设置为 1 时需要事件通知查询；设置为 0 时需要标准查询。 • 如已定义了 event_log_file，将指示 Splunk Enterprise 是否捕获仅在 Splunk Enterprise 运行期间发生的事件。设置为 1 时仅捕获在 Splunk Enterprise 运行期间发生的事件；设置为 0 时从上个检查点开始收集事件（或者，如果不存在检查点，则收集提供的最早事件）。	N/A
disabled	否	指示 Splunk Enterprise 是否收集在此输入中定义的性能数据。设置为 1 时禁用此段落的性能监视；设置为 0 时启用此段落的性能监视。	0

使用 wmi.conf 的示例

以下 wmi.conf 示例收集本地磁盘和内存性能指标，并将它们放入 wmi_perfmon 索引：

```
[settings]
initial_backoff = 5
max_backoff = 20
max_retries_at_max_backoff = 2
checkpoint_sync_interval = 2

# Gather disk and memory performance metrics from the local system every second.
# Store event in the "wmi_perfmon" Splunk index.
```

```
[WMI:LocalPhysicalDisk]
interval = 1
wql = select Name, DiskBytesPerSec, PercentDiskReadTime,PercentDiskWriteTime, PercentDiskTime from \
 Win32_PerfFormattedData_PerfDisk_PhysicalDisk
disabled = 0
index = wmi_perfmon

[WMI:LocalMainMemory]
interval = 10
wql = select CommittedBytes, AvailableBytes, PercentCommittedBytesInUse, Caption from \
 Win32_PerfFormattedData_PerfOS_Memory
disabled = 0
index = wmi_perfmon
```

关于 WQL 查询语句的其他信息

WQL 查询在结构和语法上必须正确无误。如果它们不正确，您可能会得到不符合需要的结果或者根本得不到任何结果。在您编写事件通知查询时（方式是在 WQL 查询驻留的段落中指定 `current_only=1`），您的 WQL 语句必须包含指定此类查询的子句之一：`WITHIN`, `GROUP`, 或 `HAVING`。有关更多信息，请参阅 Microsoft 网站中的 https://docs.microsoft.com/en-us/windows/win32/wmisdk/querying-with-wql?redirectedfrom=MSDN。

Splunk Web 可消除 WQL 语法问题，因为当您使用 Splunk Web 来新建性能监视器输入时，它会生成相应的 WQL 查询。

使用性能监视输入的注意事项

当您使用 Windows 性能监视器输入从 Windows 计算机收集性能监视数据时，请注意以下事项：

在集合性能指标期间内存使用量增加

当您收集有关一些性能对象的数据（例如 `Thread` 对象及其关联计数器）时，您可能会注意到 Splunk Enterprise 部署中的内存使用量增加。这种使用量的增加是正常现象，因为某些性能对象在集合过程中占用的内存比其他对象多。

处理器时间计数器返回的值不会超过 100

由于 Microsoft 使用 `Processor:% Processor Time` 和 `Process:% Processor Time` 计数器计算 CPU 使用量，无论系统中有多少 CPU 或核，这些计数器返回的值都不会超过 100。这种返回设计就是如此。这些计数器从 100% 中去除花在空闲进程上的时间量。

useEnglishOnly 设置的限制

在非英语系统上编辑 inputs.conf 文件以启用性能监视时，`useEnglishOnly` 设置的工作方式存在某些限制。

如果将其设置为 `true`，您无法使用 `object` 和 `counters` 设置的通配符及正则表达式。这些设置必须基于如“性能数据助手”库中定义的有效英语值包含特定条目。您可以指定 `instances` 设置的通配符。下面提供了一个示例：

```
[perfmon://Processor]
object = Processor
instances = _Total
counters = % Processor Time;% User Time
useEnglishOnly = 1
interval = 30
disabled = 0
```

即使系统语言不是英语，`counters` 设置也会包含以英语呈现的值。

如果将其设置为 `false`，您可以使用这些设置的通配符和正则表达式，但必须根据操作系统的语言指定值。以法语形式运行的系统上的段落示例遵循：

```
[perfmon://FrenchProcs]
counters = *
disabled = 0
useEnglishOnly = 0
interval = 30
object = Processeur
instances = *
```

请注意，本示例中的 object 设置已设置为 Processeur，是 Processor 的法语等效值。如果您在此指定英语值，则 Splunk Enterprise 将不会查找性能对象或实例。

使用 useEnglishOnly 设置的其他影响

使用该设置时要考虑其他项目。

- 当您使用 Splunk Web 在非英语操作系统上新建性能监视器输入时，Splunk Web 始终指定 useEnglishOnly = false。
- 此外，您可启用、禁用、复制或删除 Splunk Web 内的这些段落。但是，您无法在 Splunk Web 中编辑它们，除非操作系统的区域与该段落中指定的区域相匹配。
- 您可以使用 Splunk Web 启用、禁用、复制或删除性能监视器段落，只要把 useEnglishOnly 设置设为 true 即可。但是，您无法在 Splunk Web 中编辑它们，除非系统区域为英语。

使用 PowerShell 脚本监视 Windows 数据

PowerShell 是很多 Windows 版本随附的脚本语言。它允许您处理来自命令行界面的 Windows 操作。您可用此脚本语言新建脚本并将那些脚本结果输出为其他脚本的对象。

Splunk 平台支持对通过 PowerShell 脚本接收到的事件的监视。您可使用 PowerShell 输入以运行单个 PowerShell 命令或参考一个 PowerShell 脚本。之后 Splunk 平台为这些命令或脚本作为事件的输出新建索引。

如果您使用的是 Splunk Cloud，而且想监视脚本输出，请使用通用转发器从 Windows 计算机引入数据并把数据发送到您的 Splunk Cloud 部署。

要求

- Splunk Cloud 必须从安装在 Windows 计算机上的通用转发器接收来自 PowerShell 脚本的 Windows 数据。
- Splunk 平台实例必须在 Windows 上运行。请参阅*《安装手册》*中的“在 Windows 上安装”。
- Splunk 平台实例必须配置为以本地系统用户身份运行，这样便可运行所有 PowerShell 脚本。
- 必须在计算机上安装 PowerShell version 3.0 版本或更高版本。必须在计算机上安装 Microsoft .NET 4.5 版本或更高版本。
- 运行 PowerShell 脚本可能有额外的要求，这取决于 Windows 和 PowerShell 的版本。有关详细信息，请参阅 PowerShell 上有关 Microsoft 的文档。

使用配置文件配置输入

1. 编写 PowerShell 命令或脚本以捕获您想要的信息。
2. 在将运行脚本的 Splunk 平台实例上，打开 PowerShell 窗口。
3. 在 %SPLUNK_HOME%\etc\system\local 目录中创建一个 inputs.conf 配置文件。
4. 打开 inputs.conf 文件并进行编辑以启用 Windows PowerShell 输入。
5. 在输入中，使用 script 设置指定脚本的命令或完整路径。
6. （可选）使用 schedule 设置指定命令或脚本将运行的计划。
7. 保存 inputs.conf 文件。
8. 重新启动 Splunk 平台以启用输入。

PowerShell 输入配置值

Splunk 平台使用 inputs.conf 文件中的以下段落来监视由 PowerShell 收集的数据。

设置	描述	默认
script	要运行的 PowerShell 命令或脚本文件。 当您指定一个脚本文件（.ps1）时，请在脚本名称前面加上一个句号和一个空格（. ）。	n/a
schedule	命令或脚本的运行频率。 您可以指定一个数字来指定间隔时间（以秒为单位），也可以使用一个有效的 cron 计划格式。	脚本运行一次
disabled	是否启用输入。 设置为 1 时禁用，设置为 0 时启用。	0（启用）

有关编写 PowerShell 脚本的信息，请参阅本主题后面的“为 PowerShell 输入编写脚本”。

单个命令配置示例

此示例使用 Splunk 软件作为参数安装的主机名称运行 Get-Process 命令行工具和导出到 Select-Object 命令行工具的管道符。每 5 分钟运行命令。

```
    [powershell://Processes-EX1]
    script = Get-Process | Select-Object Handles, NPM, PM, WS, VM, Id, ProcessName,
@{n="SplunkHost";e={$Env:SPLUNK_SERVER_NAME}}
    schedule = */5 * * * *
    sourcetype = Windows:Process
```

脚本配置示例

此示例运行 %SPLUNK_HOME\etc\apps\My-App 目录中的 getprocesses.ps1 脚本。设置这些事件的来源类型，进行 Windows:Process。脚本在周一到周五从上午 9:00 到下午 4:40 每 20 分钟运行一次。

```
    [powershell://Processes-EX2]
    script = . "$SplunkHome\etc\apps\My-App\bin\getprocesses.ps1"
    schedule = */20 * 9-16 * 1-5
    sourcetype = Windows:Process
```

使用 Splunk Web 配置输入

如果您使用的是 Splunk Enterprise，则可以使用 Splunk Web 配置输入。要在通用转发器上配置 PowerShell 输入，请参阅本主题前面的“使用配置文件配置输入”。

按照以下步骤使用 Splunk Web 配置输入：

1. 在 Splunk Web 中选择**设置 > 数据输入**。
2. 请选择 **PowerShell v3 模块化输入**。
3. 单击**新建**。
4. 请在**名称**字段中输入一个输入名称。
5. 请在**命令或脚本路径**字段中输入命令或脚本的路径。
6. 请在**Cron 计划**字段中输入间隔时间或 Cron 计划。
7. 单击**更多设置**复选框选择来源类型、主机和默认索引。
8. 单击**下一步**。

为 PowerShell 输入编写脚本

Splunk 平台提供一个模块化 PowerShell 输入处理程序。PowerShell 处理程序支持 Microsoft PowerShell 3 版本及更高版本。

PowerShell 模块化输入提供单个实例，以及通过 stdin 输入/输出数据流和 XML 流输出提供支持架构、XML 配置的多线程脚本。请参阅*《开发用于 Splunk Web 的视图和应用》*中的“模块化输入概述”。

您可在输入配置文件中定义很多 PowerShell 段落并同时运行它们。您可使用 cron 语法计划每个段落。因为所有脚本都在同一进程中运行，所以脚本共享环境变量（如当前工作目录）。

该输入未在您的 PowerShell 环境中设置一个主机变量。为输入编写脚本时，请勿参考 $host 或使用 Write-Host 或 Out-Host PowerShell 命令行工具。相反，应该使用 Write-Output 或 Write-Error 命令行工具。

基于在架构中定义的公共属性，该输入将所有输出转换为键/值对。

Splunk 平台还包括一个名为 LocalStorage 的 PowerShell 模块，该模块显示三个命令行工具：

- Get-LocalStoragePath
- Export-LocalStorage
- Import-LocalStorage

这些命令行工具使用 Splunk 平台检查点目录并允许您存留脚本计划运行间的数据的键/值对。正常情况下，数据不会从一个调用存留到下一次调用。

指定路径

该输入设置 SplunkHome PowerShell 变量，所以通过编写如下路径即可处理加载项中的脚本：

```
[powershell://MSExchange_Health]
script=. $SplunkHome/etc/apps/TA-Exchange-2010/powershell/health.ps1
```

除 $SplunkHome 外，还有若干其他只读常数变量：

变量	描述
SplunkServerName	为本计算机配置以在事件中使用的名称
SplunkServerUri	Splunk 平台 REST API 地址
SplunkSessionKey	访问 Splunk 平台 REST API 所需的会话秘钥或验证标记
SplunkCheckpointPath	存储持续状态的路径
SplunkServerHost	您想要进行通信的 Splunk 平台实例的名称
SplunkStanzaName	定义该脚本的 inputs.conf 段落的名称

处理 PowerShell 脚本输出

Splunk 平台将您的脚本生成的每个对象视为一个输出并将其转换为一个事件，并以 <event> 和 <data> 标记括起来。Splunk 平台将每个对象的属性转换为键/值对。但是，该值只能是带有引号的字符串，其通过调用 .ToString() 方法进行转换。因此，输出必须简单，且您必须在脚本输出复杂的嵌套对象前将脚本中的这些对象扁平化。

以下特殊属性名称对 Splunk 平台模块化输入具有重要意义并允许您覆盖 inputs.conf 段落中的默认值：

属性	描述
SplunkIndex	覆盖将要存储输出的索引。
SplunkSource	覆盖输出的 source。
SplunkHost	覆盖输出的 host 名称。
SplunkSourceType	覆盖输出的 sourcetype。
SplunkTime	覆盖 time。如果您未指定时间，则脚本在单个执行中生成的所有对象将获得大致相同的时间戳。这是由于该脚本为输出保存对象，直至其已完成执行，然后会以输出时间标记对象。您必须以 epoch 或 POSIX 时间指定该值，epoch 或 POSIX 时间为一组正整数，代表从 1970 年 1 月 1 日周四 UTC 0:00 起经过的时间（以秒为单位）。

这些属性不会作为对象在键/值输出中显示。

如果您想要设置这些属性并覆盖默认值，请使用含 Select-Object 命令行工具的计算表达式，或使用 Add-Member 命令行工具，来添加 NoteProperty 属性。

处理 PowerShell 脚本输出的注意事项

目前，该输入要求其执行的任何 PowerShell 脚本生成没有任何脚本属性的输出对象。为确保格式正确，通过 Select-Object 命令行工具由管道符传递输出。

目前，在您的管道和运行空间未得以完成前，该输入不会处理脚本的输出。这意味着输入不会处理 ScriptProperty 值。它也意味着您的所有输出基本上具有相同的时间戳，除非您使用 SplunkTime 变量覆盖该时间戳。

写入脚本时，避免长时间运行脚本。请不要编写等待事情发生的脚本，除非一有输出脚本就退出。

监视 Windows 主机信息

您可以使用 Splunk 平台监视本地 Windows 计算机的详细统计信息。

如果您使用的是 Splunk Cloud，则必须使用转发器收集 Windows 主机信息并将其发送至您的 Splunk Cloud 部署。请执行下列高级步骤：

1. 将通用转发器安装在您要从中收集主机信息的 Windows 计算机上。

2. 安装应用以将通用转发器连接到 Splunk Cloud 实例。
3. 配置转发器以收集 Windows 主机信息。

Splunk Enterprise 的完整实例和通用转发器均支持直接本地收集主机信息。在这些实例类型上，Windows 主机监视器输入将作为一个进程运行，该进程称为 splunk-winhostmon.exe。此进程将以 Windows 主机监视输入中指定的时间间隔，针对您定义的每个输入运行一次。在 Splunk Enterprise 上，您可以使用 Splunk Web 配置主机监视，而在通用转发器上，您可以使用 input.conf 配置文件配置输入。

为什么监视主机信息？

您可以监视主机以获得关于您的 Windows 计算机的详细信息。您可以监视系统的更改，例如，软件的安装和删除、服务的启动和停止乃至运行时间。当发生系统故障时，您可以将 Windows 主机监视信息作为取证进程的第一步。借助 Splunk 搜索处理语言，您可以使您的团队能够了解您的 Windows 网络中所有计算机的统计信息。

Splunk 平台可以收集有关 Windows 计算机的以下信息：

常规计算机
: 计算机的牌子和型号、其主机名称及其所属的 Active Directory 域。

操作系统
: 计算机上所安装操作系统和服务包的版本和内部版本号、计算机名称、上次启动时间、安装的内存量和可用内存量以及系统驱动器。

处理器
: 系统中所安装 CPU 的牌子和型号、其速度和版本、处理器和核心数量以及处理器 ID。

磁盘
: 列出系统中所有可用驱动器、（如果可用）其文件系统类型以及总空间量和可用空间量。

网络适配器
: 有关系统中所安装网络适配器的信息，包括制造商、产品名称和 MAC 地址。

服务
: 有关系统中所安装服务的信息，包括名称、显示名称、说明、路径、服务类型、启动模式及状态。

过程
: 有关系统中所运行进程的信息，包括名称、命令行（以及参数）、启动时间以及可执行文件的路径。

要求

要监视主机信息，您必须满足以下要求：

- Splunk Cloud 必须从转发器接收 Windows 主机信息。
- 转发器必须在 Windows 上运行。请参阅*《安装手册》*中的“在 Windows 上安装”。
- 要读取所有本地 Windows 主机信息，转发器必须以本地系统 Windows 用户身份或本地管理员用户身份运行。

安全与远程访问注意事项

默认情况下，通用转发器必须以本地系统用户身份运行才能收集 Windows 主机信息。

在可能的情况下，使用通用转发器将远程计算机中的 Windows 主机信息发送到 Splunk Cloud 或 Splunk Enterprise 索引器。您必须使用通用转发器将 Windows 主机信息发送到 Splunk Cloud。有关如何安装、配置及使用通用转发器收集 Windows 主机数据的信息，请查看《转发器手册》。

如果您选择将转发器安装在要收集 Windows 主机数据的远程计算机上，您可以将转发器作为本地系统用户安装在这些计算机上。本地系统用户对本地计算机上的所有数据都具有访问权限，但对远程计算机上的数据没有访问权限。

如果您以非本地系统用户身份运行 Splunk Enterprise 或通用转发器，则该用户必须具有您要从中收集主机数据的计算机的本地管理员权限和其他权限。请参阅*《安装手册》*中的“选择 Splunk Enterprise 应以其身份运行的 Windows 用户”。

使用 inputs.conf 配置文件配置主机监视

要在 Splunk Cloud 实例上收集 Windows 主机信息，必须在您想从中收集主机信息的 Windows 机器上配置通用转发器。然后，您可以将数据发送到 Splunk Cloud。

您可以通过编辑 inputs.conf 配置主机监视。更多有关如何编辑配置文件的信息，请参阅*《管理员手册》*中的“关于配置文件”。

您还可以直接在 Splunk Enterprise 实例上配置此文件。

要在 input.conf 上配置主机监视，请执行以下步骤：

1. 在您想从中收集 Windows 主机信息计算机上，安装通用转发器。
2. 在转发器上下载并安装 Splunk Cloud 通用转发器凭据包。

3. 在转发器上，使用文本编辑器在 %SPLUNK_HOME%\etc\system\local 中创建一个 input.conf 配置文件并打开进行编辑。
4. 在同一文本编辑器中，打开 %SPLUNK_HOME%\etc\system\default\inputs.conf 并审阅您想要启用的 Windows 事件日志输入。
5. 从 %SPLUNK_HOME%\etc\system\default\inputs.conf 复制您想要启用的 Windows 事件日志输入段落。
6. 将您复制好的段落粘贴到 %SPLUNK_HOME%\etc\system\local\inputs.conf 文件中。
7. 对复制到 %SPLUNK_HOME%\etc\system\local\inputs.conf 文件的段落进行编辑，以收集所需的 Windows 事件日志数据。
8. 保存 %SPLUNK_HOME%\etc\system\local\inputs.conf 文件并将其关闭。
9. 重新启动通用转发器。

在为来自 Windows 主机监视输入的数据新建索引时，Splunk 平台会将所接收事件的**来源**设置为 windows。将传入事件的**来源类型**设置为 WinHostMon。

Windows 主机监视器配置值

Splunk Enterprise 和通用转发器使用 input.conf 配置文件中的以下设置来监视 Windows 主机信息。

设置	是否必需？	描述
interval	是	轮询新数据的频率（以秒为单位）。如果将此间隔设置为负数，则 Splunk 平台将运行一次输入。如果您未定义此设置，则输入将不运行，因为没有默认值。
type	是	要监视的主机信息的类型。可以为 Computer、operatingSystem、processor、disk、networkAdapter、service、process 或 driver 之一。如果此设置不存在，则输入将不运行。
disabled	否	是否运行输入。如果您把该设置设置为 1，平台实例不会运行该输入。

有关示例，请参阅本主题后面的“Windows 主机监视配置示例”。

使用 Splunk Web 配置主机监视

您只能在 Splunk Enterprise 中的 Splunk Web 上配置 Windows 主机信息。按照以下高级步骤通过 Splunk Web 配置主机监视：

1. 转到“添加数据”页面
2. 选择输入来源
3. （可选）指定输入设置
4. 查看您的选择

转到“添加数据”页面

按照以下步骤从“设置”进入“添加数据”页面：

1. 单击**设置** > **数据导入**。
2. 单击**文件和目录**。
3. 单击**新建本地文件和目录**以添加输入。

按照以下步骤从 Splunk Enterprise 主页进入“添加数据”页面：

1. 单击页面上的**添加数据**页面。
2. 单击**监视**以监视来自本地 Windows 计算机的主机信息。

选择输入来源

1. 在左窗格中，查找并选择**本地 Windows 主机监视**。
2. 在**集合名称**字段中，为该输入输入一个唯一、容易记住的名称。
3. 在**事件类型**框中，查找您要此输入监视的主机监视事件类型。
4. 单击要监视的每个类型。
   Splunk Enterprise 将类型从**可用类型**窗口移动到**已选类型**窗口。
5. 要取消选择一个类型，在**已选类型**窗口上单击该类型的名称。
   Splunk Enterprise 将计数器从**已选类型**窗口移动到**可用类型**窗口。
6. （可选）要选择或取消选择所有类型，请单击**添加所有**或**删除所有**链接。

   选择所有类型可索引大量数据，且可能超出您的许可证的数据限制。

7. 在**间隔**字段中，为输入在轮询尝试之间输入时间（秒）。
8. 单击**下一步**。

指定输入设置

转到**输入设置**页面以指定应用程序上下文、默认主机值和索引。所有这些参数均为可选参数。

1. 为此输入选择相应的**应用程序上下文**。
2. 设置**主机**名称。此设置有多个选项供您选择。有关设置主机值的更多信息，请参阅“关于主机”。

   主机只是设置生成事件中的主机字段。而不是配置 Splunk Enterprise 查找网络中的特定主机。

3. 设置要向其发送数据的**索引**。如果未定义多个索引来处理不同类型的事件，请保留**默认**值。除了用户数据的索引之外，Splunk 平台还有多个实用工具索引，这些索引也会显示在此下拉列表中。
4. 请单击**查看**。

查看您的选择

在您指定所有输入设置后，可查看您的选择。Splunk Enterprise 列出所有您所选的选项，包括监视器的类型、数据来源、来源类型、应用程序上下文和索引。

1. 查看该设置。
2. 如果它们不符合您的期望，请单击左尖括号（ < ）即可返回到向导中的上一个步骤。否则，请单击**提交**。

然后，Splunk Enterprise 会加载“成功”页面并开始索引指定的主机信息。

在为来自 Windows 主机监视输入的数据新建索引时，Splunk Enterprise 会把所接收事件的**来源**设置为 `windows`。将传入事件的**来源类型**设置为 `WinHostMon`。

Windows 主机监视配置示例

以下示例说明了如何在 input.conf 中配置 Windows 主机监视。

```
# Queries computer information.
[WinHostMon://computer]
type = Computer
interval = 300

# Queries OS information.
# 'interval' set to a negative number tells Splunk Enterprise to
# run the input once only.
[WinHostMon://os]
type = operatingSystem
interval = -1

# Queries processor information.
[WinHostMon://processor]
type = processor
interval = -1

# Queries hard disk information.
[WinHostMon://disk]
type = disk
interval = -1

# Queries network adapter information.
[WinHostMon://network]
type = networkAdapter
interval = -1

# Queries service information.
# This example runs the input every 5 minutes.
[WinHostMon://service]
type = service
interval = 300

# Queries information on running processes.
# This example runs the input every 5 minutes.
[WinHostMon://process]
```

```
type = process
interval = 300
```

监视 Windows 打印机信息

借助 Splunk 平台，您可以监视本地 Windows 计算机上所有打印机和驱动程序、打印任务和打印机端口的统计信息。您可以收集以下打印系统信息：

- **打印机。**有关打印子系统的信息，例如，所安装打印机的状态以及打印机的添加或删除时间。
- **任务。**有关打印任务的信息，包括打印者、打印内容、任务详细信息以及现有任务的状态。
- **驱动程序。**有关打印驱动程序子系统的信息，包括有关现有打印驱动程序的信息以及打印驱动程序的添加或删除时间。
- **端口。**有关系统中所安装打印机端口的信息，以及这些端口的添加或删除时间。

Splunk 平台的完整实例和通用转发器均支持本地收集打印机子系统信息。如果您使用的是 Splunk Cloud，而且想监视打印机子系统信息，请使用通用转发器来引入信息并把信息发送到您的 Splunk Cloud 部署。

打印机监视器输入将作为一个进程运行，该进程称为 splunk-winprintmon.exe。此进程将以输入中指定的时间间隔，针对您定义的每个输入运行一次。您可以使用 Splunk Web 或 inputs.conf 配置文件配置打印机子系统监视。

监视打印机信息的原因

Windows 打印机监视使您获得有关您的 Windows 打印机子系统的详细信息。您可以监视系统的所有更改，例如，打印机、打印驱动程序和端口的安装及删除、打印任务的开始和完成，同时了解打印者、打印内容和打印时间。当发生打印机故障时，您可以将打印监视信息作为取证进程的第一步。借助 Splunk 搜索处理语言，您可以使您的团队了解您的 Windows 网络中所有打印机的统计信息。

要求

监视主机信息前需要满足以下要求：

- Splunk 平台必须在 Windows 上运行。请参阅*《安装手册》*中的“在 Windows 上安装”。
- 为了读取所有的本地主机信息，Splunk 平台必须以本地系统用户身份运行。

安全与远程访问注意事项

默认情况下，Splunk 平台必须以本地系统用户身份运行才能收集 Windows 打印子系统信息。

请使用通用转发器将远程计算机中的打印机信息发送到索引器。如果您选择将转发器安装在要收集打印机子系统数据的远程计算机上，您可以将转发器作为本地系统用户安装在这些计算机上。本地系统用户对本地计算机上的所有数据都具有访问权限，但对远程计算机上的数据没有访问权限。

如果您以非本地系统用户身份运行 Splunk 平台，该用户必须具有计算机的本地“管理员”权限，以及*《安装手册》*中“选择 Splunk 平台应以其身份运行的 Windows 用户”所详述的其他权限。

使用 Splunk Web 配置打印机信息

按照以下高级步骤在 Splunk Web 中配置打印机信息：

1. 转到“添加数据”页面。
2. 选择输入来源。
3. 指定输入设置。
4. 查看您的选择。

转到“添加数据”页面

选择以下方法之一以进入“添加数据”页面。

要从“设置”下拉列表中添加数据，请执行以下步骤：

1. 单击**设置**。
2. 请单击**数据导入**。
3. 单击**本地 Windows 打印监视**。
4. 单击**新建**以添加输入。

要从 Splunk Web 主页添加数据，请执行以下步骤：

1. 单击**添加数据**。
2. 单击**监视**以监视来自本地 Windows 计算机的打印信息。

3. 在左窗格中，查找并选择**本地 Windows 打印监视**。

选择输入来源

1. 在**集合名称**字段中，为该输入输入一个唯一且容易记住的名称。
2. 在**事件类型**中，查找您要此输入监视的打印监视事件类型。
3. 单击要监视的每个类型一次。
   Splunk 平台将类型从“可用类型”窗口移动到“已选类型”窗口。
4. 要取消选择一个类型，在“已选类型”窗口上单击该类型的名称。
   Splunk 平台将计数器从“已选类型”窗口移动到“可用类型”窗口。
5. （可选）要选择或取消选择所有类型，请单击**添加所有**或**删除所有**。

   选择所有类型可导致索引大量数据（可能比您的许可证所允许的数量大）。

6. 在**基线**控件中，请单击**是**以仅在它启动后运行一次输入。请单击**否**以按**间隔（单位为分钟）** 字段中指定的时间间隔运行输入。
7. 单击**下一步**。

指定输入设置

您可在“输入设置”页面指定应用程序上下文、默认主机值和索引。所有这些参数均为可选参数。

1. 为此输入选择相应的**应用程序上下文**。
2. 设置**主机**名称。此设置有多个选项供您选择。有关设置主机值的更多信息，请参阅“关于主机”。

   主机只是设置生成事件中的主机字段。设置此值不会引导 Splunk 平台查找网络中的特定主机。

3. 设置 Splunk 平台将数据发送到其中的**索引**。如果未定义多个索引来处理不同类型的事件，请保留“默认”值。除了用户数据的索引之外，Splunk 平台还有很多实用工具索引，这些索引也会显示在此下拉列表中。
4. 请单击**查看**。

查看您的选择

在您指定所有输入设置后，可查看您的选择。Splunk 平台会列出您选定的所有选项，包括监视器的类型、数据来源、来源类型、应用程序上下文和索引。

1. 查看该设置。
2. 如果它们不符合您的期望，请单击左尖括号 (<) 即可返回到向导中的上一个步骤。否则，请单击**提交**。

然后，Splunk 平台会加载“成功”页面并开始索引指定的打印信息。

使用 inputs.conf 配置文件配置打印机监视

您可以编辑 inputs.conf 文件来配置打印机监视。请参阅本主题后面的“打印监视配置值和示例”。

1. 打开 Shell 提示符或 PowerShell 窗口。
2. 更改到 %SPLUNK_HOME%\etc\system\local 目录。
3. 使用文本编辑器打开此目录中的 inputs.conf 文件。您可能需要新建该文件。
4. 添加 [WinPrintMon] 配置段落、设置和值以启用 Windows 打印监视输入。
5. 保存文件并将其关闭。
6. 重新启动 Splunk 平台。

有关如何编辑配置文件的信息，请参阅*《管理员手册》*中的“关于配置文件”。

打印监视配置值

Splunk 平台使用 inputs.conf 中的以下设置来监视 Windows 打印机子系统信息：

属性	是否必需？	描述
type	是	要监视的主机信息的类型。可以是 printer、job、driver 或 port。如果此变量不存在，则输入不会运行。
baseline	否	是否生成打印机、任务、驱动程序或端口的现有状态的基准。如您把该属性设置为 1,，Splunk 平台会写入一个基线。Splunk 平台启动时，这可能需要额外的时间和 CPU 资源。
disabled	否	是否运行输入。如果您把该设置设置为 1，Splunk 平台不会运行该输入。

Windows 打印机监视配置示例

以下示例显示如何使用 input.conf 中的 Windows 打印机监视配置设置。

```
# Monitor printers on system.
[WinPrintMon://printer]
type = printer
baseline = 0

# Monitor print jobs.
[WinPrintMon://job]
type = job
baseline = 1

# Monitor printer driver installation and removal.
[WinPrintMon://driver]
type = driver
baseline = 1

# Monitor printer ports.
[WinPrintMon://port]
type = port
baseline = 1
```

用于 Windows 打印监视数据的字段

在为来自 Windows 打印监视输入的数据新建索引时，Splunk 平台会将所接收事件的**来源**设置为 windows。将传入事件的**来源类型**设置为 WinPrintMon。

监视 Windows 网络信息

借助 Splunk Enterprise，您可以监视有关传入或来自 Windows 计算机的网络活动的详细统计信息。在 Splunk Cloud 上，您可以从通用转发器监视 Windows 网络信息，该通用转发器安装在您想从中收集信息的 Windows 计算机上。然后，您可以将该数据转发至 Splunk Cloud。

Splunk 平台可以收集以下网络信息：

- **网络活动。**当 Windows 计算机执行任何类型的网络操作时，Splunk 平台可对其进行监视。
- **地址系列。**网络交易是通过 IPv4 协议还是通过 IPv6 协议进行。
- **数据包类型。**交易中发送的数据包类型，例如 connect 或 transport 数据包）。
- **协议。**网络交易是通过 TCP 协议还是通过 UDP 协议进行。
- **主机。**有关网络交易所涉及主机的信息，包括本地和远程主机、主机用来通信的端口以及所有可用的 DNS 信息。
- **应用程序。**启动了网络交易的应用程序。
- **用户。**启动了网络交易的用户，包括其 ID 和 SID。
- **其他。**有关网络交易的其他信息，包括传输标头大小以及交易是否通过 IPSec 保护。

Windows 版本的 Splunk Enterprise 和通用转发器均支持本地收集网络信息。

网络监视器输入将以进程的形式运行，该进程称为 splunk-netmon.exe。此进程将以输入中选择的时间间隔，针对您定义的每个输入运行一次。您可以使用 Splunk Web 或 inputs.conf 配置文件来配置网络监视。

Windows 网络监视仅在 64 位 Windows 系统上可用。它在 32 位 Windows 系统上不可用。

监视网络信息

Windows 网络监视使您可获得有关您的 Windows 网络活动的详细信息。您可以监视网络中的所有交易，例如，某用户或进程启动了网络连接，或者交易是使用 IPv4 还是 IPv6 地址系列。Splunk® Enterprise 中的网络监视实用工具可以告诉您所涉及的计算机，从而使您能够检测并中断传入或传出的拒绝服务攻击。借助 Splunk 的搜索处理语言，您可以使您的团队大致了解所有 Windows 网络中操作的统计信息。

要求

监视网络信息前需要满足以下要求：

- Splunk Enterprise 或通用转发器必须在 Windows 上运行。请参阅 Splunk Enterprise *《安装手册》*中的“在 Windows 上安装”。

- 计算机上的 Windows 版本必须为以下版本之一：
  - Windows 8.1
  - Windows 10
  - Windows Server 2012 R2
  - Windows Server 2016
  - Windows Server 2019
- Windows 系统必须已经应用所有可用的更新和服务包。如果所有更新未存在于您的 Windows 计算机中，则网络监视输入可能不可用。
- 为读取所有的本地主机信息，Splunk Enterprise 或通用转发器必须以“本地系统”用户或本地管理员帐户的身份运行。

安全与远程访问注意事项

默认情况下，Splunk 平台必须以“本地系统”用户身份运行才能收集 Windows 网络信息。

请使用通用转发器尽可能将远程计算机中的主机信息发送到索引器。如果您选择将转发器安装在要收集 Windows 网络信息的远程计算机上，则您可以将转发器作为“本地系统”用户安装在这些计算机上。本地系统用户对本地计算机上的所有数据都具有访问权限，但对远程计算机上的数据没有访问权限。

如果您以非“本地系统”用户身份运行 Splunk 平台，该用户身份必须具有计算机的本地“管理员”权限，以及*《安装》*手册中“选择 Splunk Enterprise 应以其身份运行的 Windows 用户”所详述的其他显式权限。

使用 inputs.conf 文件配置 Windows 网络监视

要定义某 Windows 网络监视输入，请使用 `[WinNetMon://<name>]` 段落（位于 inputs.conf 文件中）。Splunk 平台使用以下设置来配置 Windows 网络监视器输入。

1. 使用文本编辑器，在要收集 Windows 网络信息的实例上的 %SPLUNK_HOME%\etc\system\local 目录中创建 input.conf 文件。
2. 添加 `[WinNetMon://<name>]` 段落至文件。
3. 根据您希望如何配置 Windows 网络监视，在此任务后面的表中指定一项或多项设置。
4. 保存文件并将其关闭。
5. 重新启动 Splunk 平台。
   立即开始监视 Windows 网络。

下表显示了可以配置以监视 Windows 网络的设置：

设置	描述	默认
disabled = [0|1]	输入是否运行。设置为 1 时禁用该输入；设置为 0 时启用该输入。	0（启用）
index = <string>	此输入将向其发送数据的索引。这是可选属性。	默认索引
remoteAddress = <regular expression>	与网络交易中涉及的远程 IP 地址进行匹配。接受仅代表 IP 地址（而不是主机名）的正则表达式。筛选出远程地址与正则表达式不匹配的事件。通过远程地址与正则表达式匹配的事件传递。 例如：192\.163\..* 匹配 192.163.x.x 范围内的所有 IP 地址。	空字符串（匹配所有内容）
process = <regular expression>	与执行了网络访问的进程或应用程序名称进行匹配。筛选出由与正则表达式不匹配的进程生成的事件。通过由与正则表达式匹配的进程生成的事件传递。	空字符串（匹配所有进程或应用程序）
user = <regular expression>	与执行了网络访问的用户名进行匹配。筛选出由与正则表达式不匹配的用户生成的事件。通过由与正则表达式匹配的用户生成的事件传递。	空字符串（包括所有用户进行的访问）
addressFamily = [ipv4;ipv6]	如果设置此属性，则与网络访问中使用的地址系列进行匹配。接受以分号分隔的值，例如 ipv4;ipv6。	空字符串（包括所有 IP 流量）
packetType = [connect;accept;transport]	与交易中使用的数据包类型进行匹配。接受以分号分隔的值，例如 connect;transport。	空字符串（包括所有数据包类型）
direction = [inbound;outbound]	如果设置此属性，则与网络流量的一般方向进行匹配。Inbound 表示传入监视计算机的流量；outbound 表示离开监视计算机的流量。接受以分号分隔的值，例如	空字符串（包括两个方向）

	inbound;outbound。	
protocol = [tcp;udp]	与指定网络协议进行匹配。 tcp 表示传输控制协议，即网络使用握手和状态来设置交易。udp 表示用户数据报协议，这是一个无状态的即发即弃网络协议。 接受以分号分隔的值，例如 tcp;udp。	空字符串（包括两个协议类型）
readInterval = <integer>	**高级选项。**除非存在输入性能问题，否则请使用默认值。 读取网络监视器筛选器驱动程序的频率（单位为毫秒）。允许调整内核驱动程序的调用频率。频率较高可能会影响网络性能，而频率较低可能会导致事件丢失。最小合法值是 10；最大合法值是 1000。	100
driverBufferSize = <integer>	**高级选项。**除非存在输入性能问题，否则请使用默认值。 它在网络监视器筛选器驱动程序缓冲区中保留的网络数据包数量。控制驱动程序缓存的数据包量。较小的值可能会导致事件丢失，而较大的值可能会增加非分页内存的大小。最小合法值是 128；最大合法值是 8192。	1024
mode = <string>	如何输出每个事件。Splunk 平台可以在 single 或 multikv（键值对）模式中输出每个事件。	single
multikvMaxEventCount = <integer>	**高级选项。**除非存在输入性能问题，否则请使用默认值。 把 mode 设置为 multikv 时，待输出事件的最大量。最小合法值是 10；最大合法值是 500。	100
multikvMaxTimeMs = <integer>	**高级选项。**除非存在输入性能问题，否则请使用默认值。 把 mode 设置为 multikv 时，输出 mulitkv 事件的时间上限（以毫秒为单位）。最小合法值是 100；最大合法值是 5000。	1000

使用 Splunk Web 配置主机监视

您只能使用 Splunk Web 监视 Splunk 平台实例上的 Windows 网络信息。在任何其他的方案中，您必须使用配置文件配置主机监视。

按照以下高级步骤使用 Splunk Web 配置主机监视：

1. 转到“添加数据”页面。
2. 选择输入来源。
3. 指定输入设置。
4. 查看您的选择。

转到“添加数据”页面

选择以下选项之一以进入“添加数据”页面。

要从设置添加数据，请执行以下步骤：

1. 单击 Splunk Web 右上角的**设置**。
2. 请单击**数据导入**。
3. 单击**本地 Windows 网络监视**。
4. 单击**新建**以添加输入。

要从 Splunk Web 主页添加数据，请执行以下步骤：

1. 单击**添加数据**。
2. 单击**监视**以监视来自本地 Windows 计算机的网络信息或**转发**以转发来自另一个 Windows 计算机的网络信息。

   转发网络信息需要其他设置。

   Splunk Web 将显示“添加数据 - 选择来源”页面。
3. 在左窗格中，查找并选择**本地 Windows 网络监视**。

选择输入来源

1. 在**网络监视名称**字段中，为该输入输入一个唯一且容易记住的名称。
2. 在**地址系列**下方，选中您希望 Splunk 平台监视的 IP 地址系列类型。类型为 Ipv4 或 IPv6。
3. 在**数据包类型**下方，选中您希望输入监视的数据包类型。您可以选择**连接**、**接受**或**传输**。
4. 在**方向**下方，选中您希望输入监视的网络方向。选择**入站**（朝向监视主机）或**出站**（远离监视主机）。
5. 在**协议**下方，选中您希望输入监视的网络协议类型。选择 **tcp**（传输控制协议）或 **udp**（用户数据报协议）。
6. 在**远程地址**文本字段中，输入其网络与您希望输入监视的监视主机进行通信的远程主机的主机名或 IP 地址。如果您要监视多个主机，请在此字段中输入正则表达式。
7. 在**进程**文本字段中，输入您希望输入监视其网络通信的进程的部分名称或全名。您可以通过输入正则表达式监视多个进程。
8. 在**用户**文本字段中，输入您希望输入监视其网络通信的用户的部分名称或全名。您可以通过输入正则表达式监视多个用户。
9. 单击**下一步**。

指定输入设置

您可在“输入设置”页面指定应用程序上下文、默认主机值和索引。所有这些参数均为可选参数。

1. 为此输入选择相应的**应用程序上下文**。
2. 设置**主机**名称。此设置有多个选项供您选择。有关设置主机值的更多信息，请参阅“关于主机”。

   主机只是设置生成事件中的 host 字段。设置此值不会引导 Splunk 平台查找网络中的特定主机。

3. 设置 Splunk 平台将数据发送到其中的**索引**。如果未定义多个索引来处理不同类型的事件，请保留**默认**值。除了用户数据的索引之外，Splunk 平台还有很多实用工具索引，这些索引也会显示在此下拉列表中。
4. 请单击**查看**。

查看您的选择

在您指定所有输入设置后，可查看您的选择。Splunk 平台会列出您选定的所有选项，包括监视器的类型、数据来源、来源类型、应用程序上下文和索引。

1. 查看该设置。
2. 如果它们不符合您的期望，请单击左尖括号（<）即可返回到向导中的上一个步骤。否则，请单击**提交**。

加载“成功”页面，并且 Splunk 平台开始索引指定的输入。

用于 Windows 网络监视数据的字段

在为来自 Windows 网络监视输入的数据新建索引时，Splunk 平台将把所接收事件的**来源**设置为 windows。将传入事件的**来源类型**设置为 WinNetMon。

使用 HTTP 事件收集器获取数据

在 Splunk Web 中设置并使用 HTTP 事件收集器

HTTP 事件收集器 (HEC) 可让您通过 HTTP 和安全 HTTP (HTTPS) 协议将数据和应用程序事件发送至 Splunk 部署。HEC 使用基于标记的验证模式。您可以生成标记，然后使用该标记配置一个日志库或 HTTP 客户端，以用特定的格式把数据发送至 HEC。当发送应用程序事件时，该过程不需要 Splunk 转发器。

启用 HEC 之后，您可以在应用程序中使用 HEC 标记将数据发送至 HEC。您无需在您的应用程序或支持的文件中包含 Splunk 凭据即可访问 Splunk 平台实例。

Splunk 软件类型因 HEC 功能而异

HTTP 事件收集器在 Splunk Cloud 和 Splunk Enterprise 上运行。运行方式视您拥有的 Splunk 平台实例类型而定。

HEC 和 Splunk Cloud

您可以在 Splunk Cloud 部署上启用 HEC。以下注意事项适用于在 Splunk Cloud 实例上使用 HEC：

- 如果您需要使用配置文件来配置 HEC 输入，则必须在重型转发器上执行此操作，然后将数据转发至 Splunk Cloud。这是因为 Splunk Cloud 不提供对本地配置文件的访问。
- 您必须向 Splunk 支持提交工单，以便 HEC 能够与 Amazon Web Services (AWS) Kinesis Firehose 一起使用。默认在所有 Splunk Cloud 部署上启用 Standard HEC 且不需要 Splunk 支持工单。
- 您无法更改全局设置。您只能更改您新建的标记的设置。
- 您无法将 HEC 接收的数据转发到另一组 Splunk 索引器，因为 Splunk Cloud 不支持转发输出组。
- 您选择存储 HEC 收到的事件的索引必须已存在。您无法在设置过程中新建索引。
- 目前，索引器确认只适用于 AWS Kinesis Firehose。
- 新建标记后，您可以监视标记进度，因为标记部署在 Splunk Cloud 实例中。

有关如何在 Splunk Cloud 上启用和管理 HEC 的说明，请参阅“在 Splunk Cloud 上配置 HTTP 事件收集器”。

HEC 和 Splunk Enterprise

本地 Splunk Enterprise 平台上的 HEC 功能完整，配置齐全。相比较 Splunk Cloud 上的 HEC，它具备以下其他功能：

- HEC 不仅接受您通过 HTTPS 协议发送的事件，还可以接受您通过 HTTP 协议发送的事件。
- HEC 可以使用可选的转发**输出组**将事件转发到另一个 Splunk 索引器。
- 您可以使用部署服务器在分布式部署中，跨索引器分发 HEC 标记。

有关如何在 Splunk Enterprise 上启用和管理 HEC 的说明，请参阅“在 Splunk Enterprise 上配置 HTTP 事件收集器”。

Splunk 平台如何使用 HTTP 事件收集器标记导入数据

标记是允许日志代理和 HTTP 客户端连接到 HEC 输入的实体。每个标记都有一个唯一值，用 32 字符的全局唯一标识符 (GUID) 表示的 128 位数字。代理和客户端会使用标记验证 HEC 连接。当客户端连接后，它们显示此标记值。如果 HEC 接收到有效的标记，它接受可以以文本或 JavaScript Object Notation (JSON) 格式提供应用程序事件的有效载荷的连接和客户端。

HEC 接收事件，Splunk® Enterprise 依据标记配置为这些事件新建索引。HEC 使用来源、来源类型和标记中指定的索引。如果 Splunk Enterprise 实例上存在转发输出组配置，HEC 将数据转发到该输出组中的索引器。

在 Splunk Cloud 中配置 HTTP 事件收集器

HEC 在 Splunk Cloud 中默认启用 。您可以创建、修改、删除、启用和禁用 HEC 标记。

在 Splunk Cloud 中启用 HTTP 事件收集器

HTTP 事件收集器在 Splunk Cloud 中默认启用 。

在 Splunk Cloud 中新建事件收集器标记

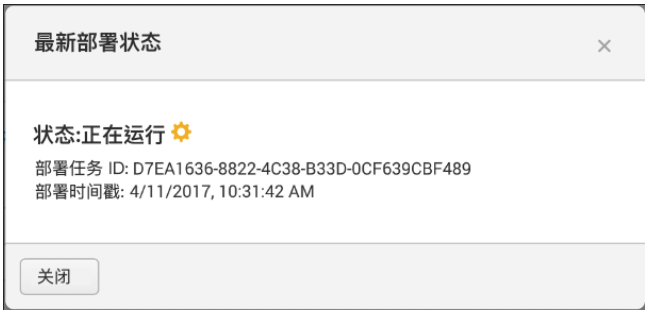
要使用 HEC，您必须配置至少一个标记。Splunk Cloud 在整个部署中分布标记。分布完成后才可以使用标记。

1. 单击**设置** > **添加数据**。
2. 单击**监视**。

3. 单击 **HTTP 事件收集器**。
4. 在**名称**字段，请输入标记的名称。
5. （可选）请在**数据来源名称覆盖**字段中输入数据来源的名称，该名称将会被分配给此端点生成的事件。
6. （可选）请在**描述**字段中输入此输入的描述。
7. （可选）如您想要启用此标记的索引器确认，单击**启用索引器确认**复选框。
8. 单击**下一步**。
9. （可选）编辑来源类型并确认您想将 HEC 事件存储于其中的索引。请参阅“修改输入设置”。
10. 请单击**查看**。
11. 请确认端点的所有设置都是您想要的。
12. 如果所有设置都是您想要的，请单击**提交**。否则，请单击 < 更改。
13. （可选）复制 Splunk Web 显示的标记值并将其粘贴到另一个文档供日后参考。
14. （可选）单击**追踪部署进程**以查看标记是如何部署到其他 Splunk Cloud 部署。当您看到“完成”状态时，您可以用标记发送数据到 HEC。

在 Splunk Cloud 中检查事件收集器标记分布状态

您可以从 HEC 标记页面检查 HEC 标记的分布状态。如果分布正在进行中，那么页面会显示“操作正在进行”和进度条。否则，页面显示“上次部署状态”。



1. 单击**设置 > 数据导入**。
2. 单击 **HTTP 事件收集器**。
3. 单击**操作正在进行**或**上次部署状态**。
4. 查看标记分布的状态。
5. 单击**关闭**。

在 Splunk Cloud 中修改事件收集器标记

在您新建 HEC 标记后可对其进行更改。

1. 单击**设置 > 数据导入**。
2. 单击 **HTTP 事件收集器**。
3. 请在列表中查找您要更改的标记。
4. 请在此标记的**操作**列中单击**编辑**。您也可单击标记名称的链接。
5. （可选）在**描述**字段中输入更新的文本可编辑标记的描述。
6. （可选）在**来源**字段中输入文本可更新标记的来源值。
7. （可选）在**来源类型**下拉列表框中可选择一个不同的来源类型。
    1. 选择一个类别。
    2. 在弹出菜单中出现的来源类型。
    3. （可选）您也可在下拉列表框顶部的文本框中键入来源类型的名称。
8. （可选）在**选择允许的索引**控制的**可用索引**窗格中可选择不同索引。
9. （可选）选择您是否想要为标记启用索引器确认。
10. 单击**保存**。

在 Splunk Cloud 中删除事件收集器标记

您可以删除 HEC 标记。删除某个 HEC 标记不影响其它 HEC 标记，也不会禁用 HEC 端点。

您无法撤消此操作。使用此标记向您的 Splunk 部署发送数据的客户端无法再使用此标记进行验证。您必须生成一个新标记并更改客户端配置以使用新值。

1. 单击**设置 > 数据导入**。
2. 单击 **HTTP 事件收集器**。
3. 请在列表中查找您要删除的标记。
4. 请在此标记的**操作**列中单击**删除**。
5. 请在“删除标记”对话中单击**删除**。

在 Splunk Cloud 中启用和禁用事件收集器标记

您可以在 HEC 管理页面内启用或禁用标记。更改某标记的启用状态不会更改到其他标记的状态。

1. 单击**设置 > 数据导入**。
2. 单击 **HTTP 事件收集器**。
3. 请在此标记的**操作**列中单击**启用**链接（如果此标记未启用）或**禁用**链接（如果此标记已启用）。根据更改的标记状态，此标记状态会切换且链接会更改为**启用**或**禁用**。

在 Splunk Enterprise 中配置 HTTP 事件收集器

您可以在 Splunk Enterprise 中启用 HEC 以及创建、修改、删除、启用和禁用 HEC 标记。

在 Splunk Enterprise 中启用 HTTP 事件收集器

使用事件收集器通过 HTTP 接收事件前，您必须先启用该事件收集器。对 Splunk Enterprise 而言，通过**全局设置**对话框启用 HEC。

1. 单击**设置 > 数据导入**。
2. 单击 **HTTP 事件收集器**。
3. 单击**全局设置**。
4. 请在**所有标记**切换按钮中选择**已启用**。
5. （可选）为所有 HEC 标记选择**默认来源类型**。在选择来源类型前，您也可在下拉列表框上方的文本字段中键入来源类型的名称。
6. （可选）为所有 HEC 标记选择**默认索引**。
7. （可选）为所有 HEC 标记选择**默认输出组**。
8. （可选）如需使用部署服务器处理 HEC 标记的配置，请单击**使用部署服务器**复选框。
9. （可选）如需让 HEC 通过 HTTPS 而非 HTTP 侦听和通信，请单击**启用 SSL** 复选框。
10. （可选）在 **HTTP 端口数**字段中输入一个数字以便 HEC 侦听。

    确认客户端或托管 HEC 的 Splunk 实例都没有防火墙阻止您在“HTTP 端口号”字段中指定的端口号。

11. 单击**保存**。

在 Splunk Enterprise 中新建事件收集器标记

要使用 HEC，您必须配置至少一个标记。

1. 单击**设置 > 添加数据**。
2. 单击**监视**。
3. 单击 **HTTP 事件收集器**。
4. 在**名称**字段，请输入标记的名称。
5. （可选）在**来源名称覆盖**字段，请输入此输入生成的事件的来源名称。
6. （可选）请在**描述**字段中输入此输入的描述。
7. （可选）请在**输出组**字段中选择一个已有的转发器输出组。
8. （可选）如您想要启用此标记的索引器确认，单击**启用索引器确认**复选框。
9. 单击**下一步**。
10. （可选）确认 HEC 事件的来源类型和索引。
11. 请单击**查看**。
12. 请确认端点的所有设置都是您想要的。
13. 如果所有设置都是您想要的，请单击**提交**。否则，请单击 < 更改。
14. （可选）复制 Splunk Web 显示的标记值并将其粘贴到另一个文档供日后参考。

在 Splunk Enterprise 中修改事件收集器标记

在您新建 HEC 标记后可对其进行更改。

1. 单击**设置 > 数据导入**。
2. 单击 **HTTP 事件收集器**。
3. 请在列表中查找您要更改的标记。
4. 请在此标记的**操作**列中单击**编辑**。您也可单击标记名称的链接。
5. （可选）在**描述**字段中输入更新的文本可编辑标记的描述。
6. （可选）在**来源**字段中输入文本可更新标记的来源值。
7. （可选）在**来源类型**下拉列表框中可选择一个不同的来源类型。
    1. 选择一个类别。
    2. 在弹出菜单中出现的来源类型。
    3. （可选）您也可在下拉列表顶部的文本框中键入来源类型的名称。
8. （可选）在**选择允许的索引**控制的**可用索引**窗格中可选择不同索引。

9. （可选）请从**输出组**下拉列表框中选择一个不同的输出组。
10. （可选）选择您是否想要为标记启用索引器确认。
11. 单击**保存**。

在 Splunk Enterprise 中删除事件收集器标记

您可以删除 HEC 标记。删除某个 HEC 标记不影响其它 HEC 标记，也不会禁用 HEC。

您无法撤消此操作。使用此标记向您的 Splunk 部署发送数据的客户端无法再使用此标记进行验证。您必须生成一个新标记并更改客户端配置以使用新标记。

1. 单击**设置** > **数据导入**。
2. 单击 **HTTP 事件收集器**。
3. 请在列表中查找您要删除的标记。
4. 请在此标记的**操作**列中单击**删除**。
5. 请在“删除标记”对话框中单击**删除**。

在 Splunk Enterprise 中启用和禁用事件收集器标记

您可以从 HEC 管理页面内启用或禁用单个 HEC 标记。更改某标记的状态不会更改到其他标记的状态。要启用或禁用所有标记，请使用“全局设置”对话框。请参阅“启用 HTTP 事件收集器”。

如需切换 HEC 标记的活动状态，步骤如下：

1. 单击**设置** > **数据导入**。
2. 单击 **HTTP 事件收集器**。
3. 请在此标记的**操作**列中单击**启用**链接（如果此标记未启用）或**禁用**链接（如果此标记已启用）。此标记状态立即切换且链接基于已更改的标记状态更改为**启用**或**禁用**。

在 Splunk Enterprise 中使用输出组指定索引器组

要索引大量数据，您可能需要多个索引器。仅在 Splunk Enterprise 上，您可以指定索引器组为 HTTP 事件收集器数据建立索引。这些组被称为*输出组*。例如，您可以使用输出组仅索引特定类型的数据或来自特定数据来源的数据。虽然使用输出组将数据路由到特定索引器与 Splunk Enterprise 附带的路由和筛选功能类似，但是输出组允许您按标记指定索引器组。

使用多个索引器组配置输出组时，Splunk Enterprise 会在输出组中的服务器之间平均分布数据。您可在 `outputs.conf` 配置文件中配置输出组。具体来说，对于 HTTP 事件收集器，请编辑位于 `$SPLUNK_HOME/etc/apps/splunk_httpinput/local/` 中（Microsoft Windows 主机上的 `%SPLUNK_HOME%\etc\apps\splunk_httpinput\local\` 中）的 `outputs.conf` 文件。如果此位置无 `local` 目录或 `outputs.conf` 文件，请新建其中一个（或两个）。

HTTP 事件收集器并不是一个应用，但是会将其配置存储在 `$SPLUNK_HOME/etc/apps/splunk_httpinput/` 目录中（Windows 上的 `%SPLUNK_HOME%\etc\apps\splunk_httpinput\` 中），这样可以使用内置应用部署功能轻松部署配置。

发送数据到 HTTP 事件收集器

向 HEC 发送数据时，您必须满足以下所有条件：

- 必须启用 HEC
- 必须至少有一个活跃的 HEC 可用标记
- 必须使用活跃的标记验证 HEC
- 必须为以某种方式传递到 HEC 的数据设置格式。请参阅“为 HTTP 事件收集器设置事件格式”

有几种将数据发送到 HTTP 事件收集器的选项：

- 您可以使用喜好的 HTTP 客户端发出 HTTP 请求，并发送您的 JSON 编码事件。
- 作为开发人员，您可以使用应用程序中的 Java、JavaScript (`node.js`) 和 .NET 日志库将数据发送到 HEC。这些库与常见日志框架兼容。请参阅“Splunk 开发人员门户”中的 Java、JavaScript (Node.js) 和 .NET。

在 Splunk Cloud 中将数据发送到 HTTP 事件收集器

您必须使用 HEC 特定的 URI 发送数据。

Splunk Cloud 免费试用版中 HEC URI 的标准格式如下所示：

```
<protocol>://inputs.<host>:<port>/<endpoint>
```

Splunk Cloud 中 HEC URI 的标准格式如下所示：

```
<protocol>://http-inputs-<host>:<port>/<endpoint>
```

Google Cloud 上 Splunk Cloud 中 HEC URI 的标准格式如下所示：

```
<protocol>://http-inputs.<host>:<port>/<endpoint>
```

其中：

- <protocol> 是 http 或 https
- 您必须在 <host> 之前添加 http-inputs-
- <host> 是运行 HEC 的 Splunk Cloud 实例
- <port> 是 HEC 端口号
  - Splunk Cloud 免费试用版上为 8088
  - Splunk Cloud 实例上默认为 443
- <endpoint> 是您想使用的 HEC 端点。大多情况下，您可以使用 JavaScript Object Notation (JSON) 格式事件的 /services/collector 端点或原始事件的 services/collector/raw 端点。

如果您在发送数据时未在 Splunk Cloud 主机名前面加上这些前缀，则数据无法到达 HEC。

在 Splunk Enterprise 中将数据发送到 HTTP 事件收集器

您可以将数据发送到 HEC 的特定统一资源标识符 (URI)。

Splunk Enterprise 中 HEC URI 的标准格式如下所示：

```
<protocol>://<host>:<port>/<endpoint>
```

其中：

- <protocol> 是 http 或 https
- <host> 是运行 HEC 的 Splunk 实例
- <port> 是 HEC 端口号，默认号码为 8088，但是您可在 HEC 全局设置中更改
- <endpoint> 是您想使用的 HEC 端点。大多情况下，您可以使用 JavaScript Object Notation (JSON) 格式事件的 /services/collector 端点或原始事件的 services/collector/raw 端点。

使用 HTTP 请求发送数据到 HEC 的示例

以下示例在端口 8088 向 HEC 发送 HTTP POST 请求，并使用 HTTPS 传输。此示例使用 curl 命令以生成请求，但是您可以使用命令行或更符合您的需求的其他工具。

您可以在 Splunk Enterprise 或 Splunk Cloud 部署中配置不同于其他 HEC 实例设置的网络端口和 HTTP 协议设置。

以下 cURL 命令使用了一个 HTTP 事件收集器标记示例 (B5A79AAD-D822-46CC-80D1-819F80D7BFB0)，并把 https://hec.example.com 用作主机名。运行此命令前用您自己的值替换这些值。

JSON 请求和响应

当您发出 JSON 请求以将数据发送到 HEC 时，您必须在命令中指定“事件”键。

```
curl -k  https://hec.example.com:8088/services/collector/event -H "Authorization: Splunk B5A79AAD-D822-46CC-80D1-819F80D7BFB0"
-d '{"event": "hello world"}'
{"text": "Success", "code": 0}
```

关于 HEC 的更多信息

- 有关 HEC 标记的信息，请参阅“Splunk 平台如何使用 HTTP 事件收集器标记导入数据”中的“get_data_in”。
- 关于定义转发输出组的信息，请参阅“使用 outputs.conf 配置转发器”。您也可以在 Splunk Web 中设置转发，Splunk Web 会生成默认的输出组，名为 default-autolb-group。
- 有关索引器确认的信息，请参阅“HTTP 事件收集器索引器确认”。HTTP 事件收集器中的索引器确认与索引器确认和索引器群集中介绍的索引器确认功能不同。

关于开发人员的 HEC 的更多信息。

- 关于使用 HTTP 事件收集器的开发人员内容，请参阅《Splunk 开发人员门户》中的“Splunk HTTP 事件收集器简介”，或“HTTP 事件收集器示例”。

使用配置文件设置并使用 HTTP 事件收集器

HTTP 事件收集器 (HEC) 会将其设置存储在两个配置文件（inputs.conf 和 outputs.conf）中的 Splunk Enterprise 实例上。这些文件无法在 Splunk Cloud 实例上访问，且必须通过 Splunk Web 管理 Splunk Cloud 实例上的配置。

用配置文件配置 HEC 输入和配置其他数据导入略有不同。在许多情况下，您可以编辑 $SPLUNK_HOME/etc/system/local 目录中的文件。对于 HEC，编辑 $SPLUNK_HOME/etc/apps/splunk_httpinput/local/ 目录中的文件。

无论 Splunk Enterprise 实例拥有多少 inputs.conf 文件，以及这些文件驻留在哪里，Splunk Enterprise 都会使用位置优先规则组合所有设置。请参阅*《管理员手册》*中的“配置文件优先顺序”。

要使用配置文件设置 HEC，请执行以下步骤：

1. 在 $SPLUNK_HOME/etc/apps/splunk_httpinput 目录中，如果 local 目录不存在，则新建该目录。
2. 更改到 $SPLUNK_HOME/etc/apps/splunk_httpinput/local 目录。
3. 如果 inputs.conf 文件不存在，请新建此文件。
4. 使用文本编辑器打开 inputs.conf 进行编辑。
5. 指定全局设置和标记设置（如本主题后续“标记相关设置”所述）。

   HEC 标记必须是全局唯一标识符 (GUID)。

6. 保存文件并将其关闭。
7. 重新启动 Splunk Enterprise 使更改生效。

标记相关设置

HEC 会存储 inputs.conf 配置文件中标记管理相关的设置。

您可以指定设置是全局应用于所有标记还是仅应用于特定标记：

- [http] 段落包含应用于所有标记的全局设置。
- [token_name] 段落可应用于单个标记，其中 token_name 表示用户分配的标记名称。此处指定的设置会覆盖 [http] 段落中指定的设置。

Inputs.conf 文件包含有关各设置的基本说明信息。

全局设置

[http] 段落包含应用于所有标记的全局设置。

参数	描述
dedicatedIoThreads	HTTP 事件收集器服务器上分配器线程数量。默认值为 2。除非 Splunk 支持人员要求，否则不应更改此设置。此参数值不得超过 Splunk Enterprise 服务器上物理 CPU 核心的数量。
disabled	是否禁用标记。1 表示 true；0 表示 false。默认值为 1。如果在 [http] 段落中设置为 1，则此参数禁用所有标记。
enableSSL	HTTP 事件收集器服务器协议是 HTTP 还是 HTTPS。1 表示启用 HTTPS；0 表示启用 HTTP。默认值为 1。HTTP 事件收集器和 Splunk Enterprise 实例共享 SSL 设置，且无法拥有 enableSSL 设置，该设置和 Splunk Enterprise 实例上的设置不同。
index	全局默认索引。可以在单个标记段落中设置之后，或者在事件数据的标题包含设为不同值的 index 参数时覆盖此参数。您可以使用 indexes 参数，以每个标记为基础限制此参数的允许值集。
maxSockets	Splunk Enterprise 接受同时连接的 HTTP 事件收集器的数量（以整数表示）。您可以限制此数量来限制资源使用情况。如果设置为 0，Splunk Enterprise 会自动设为主机上最多允许打开文件的三分之一。如果此数字小于 50，将会设为 50。如果此数字大于 400000，将会设为 400000。如果设为负数，则没有限制。默认为 0。
maxThreads	活动 HTTP 交易可以使用线程数量（以整数表示）。您可以限制此数量来限制资源使用情况。如果设置为 0，Splunk Enterprise 会自动将限制设为主机上最多可使用线程的三分之一。如果此数字小于 20，将会设为 20。如果此数字大于 150000，将会设为 150000。如果 maxSockets 不是负数，且 maxThreads 大于 maxSockets，那么 Splunk Enterprise 会将 maxThreads 设为和 maxSockets 相同。如果设为负数，则没有限制。默认为 0。
outputgroup	全局默认输出组。输出组是一组由 Splunk Enterprise 管理员设置用于索引数据的索引器。如果没有指定输出组，事件数据将进入本地索引器。如果给定输出组无效，则数据会遭到丢弃，错误信息将记录在 splunkd.log 中。有关指定输出组的更多

	信息，请参阅本主题后面的“输出组相关设置”。
port	HTTP 事件收集器服务器端口。默认值为 8088。此端口号不能再使用中。
sourcetype	全局默认来源类型。可以在单个标记段落中设置之后，或者通过将标题中包含 sourcetype 参数的事件数据设为不同值来覆盖此参数。
useDeploymentServer	是否使用部署服务器。如果设为 1（true），写入 serverclass.conf 文件中 repositoryLocation 属性指定的位置。默认为 0，并写入 $SPLUNK_HOME/etc/apps。

每个标记设置

http://<token_name> 段落可应用于单个标记，其中 <token_name> 表示用户分配的标记名称。此处指定的设置会覆盖 [http] 段落中指定的设置。

参数	描述
connection_host	标记的默认主机类型。该参数可以设置为以下任意一个文本值： • dns 表示主机值是发送数据的系统 IP 地址的反向 DNS 条目。 • ip 表示主机值是发送数据的系统 IP 地址。 • none 将主机值设为 HTTP 主机标题中指定的连接主机。通常是 Splunk 平台实例主机名。
disabled	是否禁用标记。1 表示 true；0 表示 false。默认值为 0。
index	标记的默认索引。可通过将标题中包含 index 参数的事件数据设为不同值覆盖此参数。
indexes	可用于索引数据的可用索引列表。
persistentQueueSize	**保留队列**的最大大小。此参数值在表 <integer>[KB|MB|GB] 中。默认值为 0，表示没有保留队列。保留队列可以通过将输入队列中的数据保存到磁盘，防止传输数据丢失。设置后，persistentQueueSize 参数的值必须大于 queueSize 参数的值。有关保留队列的更多信息，请参阅“使用保留队列有助于防止数据丢”。
source	标记的默认来源。可通过将标题中包含 source 参数的事件数据设为不同值来覆盖此参数。
queueSize	内存中输入队列的最大大小。此参数值采用 <integer>[KB|MB|GB] 格式。默认值为 500KB。
sourcetype	标记的默认来源类型。可通过将标题中包含 sourcetype 参数的事件数据设为不同值来覆盖此参数。
标记	HTTP 事件收集器标记。标记必须是唯一的 GUID。

输出组相关设置

应用于转发和负载均衡的设置存储在 outputs.conf 中，包括用于指定 HTTP 事件收集器输出组的设置。这些设置和 Splunk Enterprise 管理员用于在索引器间管理转发和负载均衡的设置一样。

在 [tcpout] 段落中指定全局设置，并在 [tcpout:<target_group>] 段落中指定每个输出组设置。

outputs.conf 文件包含有关各设置的基本说明信息。有关更多信息，请参阅*《转发数据》*手册中的“关于转发和接收”以及 Splunk 通用转发器*《转发器手册》*中的“使用 outputs.conf 配置转发器”。

全局设置

[tcpout] 段落定义将数据转发到哪些输出组。

参数	描述
defaultGroup	以 <target_group>, <target_group>, ... 形式命名的一个或多个目标输出组名称列表，以逗号隔开。输出组名称之后在 [tcpout:<target_group>] 段落中的 outputs.conf 文件内指定。将发送数据到指定组。

每个输出组设置

[tcpout:target_group] 段落定义由 <target_group> 指示的目标输出组的配置。您可以拥有任意数量的目标组。如果指定了多个目标组，则转发器会将数据克隆到各目标组。

参数	描述

blockWarnThreshold	输出组管道发送失败计数线程。默认值为 100。符合线程后，失败消息将在 Splunk Web 中显示为横幅。要有效禁用此警告，将此值设为非常大的值，例如 2000000。
server	<server name>]:<port>, [<ip>|<server name>]:<port>, ...。对于上述各系统，您必须包括端口号、IP 地址或服务器名称。

通过 CLI 设置和使用 HTTP 事件收集器

您可以使用 Splunk 命令行界面 (CLI) 的 http-event-collector 参数及其选项管理 Splunk Enterprise 服务器上的 HTTP 事件收集器 (HEC)。

无法通过 CLI 在 Splunk Cloud 实例上使用 HEC。如果您有 Splunk Cloud 实例，请登录该实例并改为从 Splunk Web 管理 HEC。

有关 CLI 的更多信息，请参阅 Splunk Enterprise *《管理员手册》* 中的以下主题：

- 关于 CLI
- 使用 CLI 来管理远程 Splunk 服务器

CLI 语法

当您通过 CLI 管理 HEC 时，有两种语法可供使用：

- 适用于所有其他 HEC 操作（例如新建、删除和显示标记等）的语法
- 用于将数据发送至 HEC 的语法

以下语法适用于所有操作（将数据发送至 HEC 除外）：

```
splunk http-event-collector <command><token-name> [<option2>] [<-parameter1><value1>] [<-parameter2><value2>] <data>
```

所有 HTTP 事件收集器命令（send 除外）假设命令名称之后的第一个选项是标记的名称。此外，create 命令假设第二个选项是标记的描述（用引号引起来）。

使用以下语法将数据发送至 HEC：

```
splunk http-event-collector send -uri <uri_value> -name <token-name><data>
```

如果您想要将 CLI 命令应用于全局配置，请勿包括 -name <token-name> 参数。例如，以下语法会启用 HTTP 事件收集器：

```
splunk http-event-collector enable -uri <uri_value><data>
```

支持的 CLI 命令

Splunk Enterprise 支持以下特定于 HTTP 事件收集器的 CLI 命令：

命令	描述
create	新建标记。
delete	删除标记。
list	显示所有可用标记。
update	更改标记属性。
enable	启用标记。
disable	禁用标记。
help	显示帮助。
send	将数据发送至端点。

支持的 CLI 参数

HEC 支持以下 CLI 参数。您必须在 CLI 参数后随即紧跟其值。用引号将包含空格的任意值引起来。

参数	描述
-uri	Splunk 服务器的统一资源标识符（URI）采用的格式：scheme://host:port。作为设置此参数的替换方案，您还可以设置 $SPLUNK_URI 环境变量。要使用的端口号必须为 Splunk 服务器的管理端口（默认为 8089），而不是 HTTP 事件收集器端口（默认为 8088）。
-auth	Splunk 服务器用户身份验证格式为：username:password。如果此参数缺失，将提示您输入用户名和密码。
-name	标记的名称。
-disabled	是否禁用标记。1 表示 true；0 表示 false。
-description	标记描述。
-indexes	标记接受的索引列表。
-index	标记默认索引。Splunk Enterprise 会将此值分配给尚未设置索引值的数据。
-source	标记的默认来源值。Splunk Enterprise 会将此值分配给尚未设置来源值的数据。
-sourcetype	标记的默认来源类型值。Splunk Enterprise 会将此值分配给尚未设置来源类型值的数据。
-outputgroup	标记的默认输出组值。输出组是一组由 Splunk 软件管理员设置用于索引数据的索引器。Splunk Enterprise 会将此值分配给尚未设置输出组值的数据。
-port	HTTP 事件收集器服务器端口。默认值为 8088，但是您可以使用此参数更改。
-enable-ssl	HTTP 事件收集器服务器协议是 HTTP 还是 HTTPS。1 表示 HTTPS；0 表示 HTTP。
-dedicated-io-threads	HTTP 事件收集器服务器上分配器线程数量。默认值为 2。除非 Splunk 支持人员要求，否则不应更改此设置。此参数值不得超过 Splunk Enterprise 服务器上物理 CPU 核心的数量。
-output-format	输出格式。txt 表示文本；json 表示 JSON。默认值为 txt。

示例 CLI 语法

以下示例 CLI 条目可新建名为 new-token 的标记、将其分配给指定的 URI、为其添加描述、将其设为 disabled，并指示应保存到 **log** 索引中的 HTTP 事件收集器数据。

```
splunk http-event-collector create new-token -uri https://localhost:8089 -description "this is a new token" -disabled 1 -index log
```

以下示例 CLI 条目启用名为 myapp 的标记、将其分配给指定的 URI，并设置用户身份验证：

```
splunk http-event-collector enable -name myapp -uri https://localhost:8089 -auth admin:changeme
```

以下示例 CLI 条目使用指定的标记和 URI 将数据发送至 HTTP 事件收集器。

```
splunk http-event-collector send -uri https://localhost:8089 -token new-token {"this is the data to send"}
```

使用 cURL 管理 HTTP 事件收集器标记、事件和服务

您可以使用 cURL Web 数据传输应用程序，通过表现层状态转换（REST）API 管理 Splunk® Enterprise 实例上 HTTP 事件收集器（HEC）的标记、事件和服务。使用 REST API，您可以无缝管理 HEC 对象，而无需使用 Splunk Web 或 CLI。

使用 cURL 管理 HTTP 事件收集器

所有 HEC 标记操作都可以使用 cURL 进行 REST。Splunk® Enterprise 将这些标记存储在以下 REST API 端点，假设您的 Splunk® Enterprise 服务器管理地址如下：

```
https://localhost:8089/servicesNS/admin/splunk_httpinput/data/inputs/http/
```

使用 cURL 列出现有的 HTTP 事件收集器标记

您可以使用 cURL 列出 Splunk® Enterprise HEC 中的现有标记。例如，以下示例 cURL 命令通过用户管理员列出了 https://localhost:8089 上 Splunk® Enterprise 实例中存在的标记：

```
curl -k -u admin:changeme https://localhost:8089/servicesNS/admin/splunk_httpinput/data/inputs/http
```

使用 cURL 新建 HTTP 事件收集器标记

要使用 cURL 新建标记，请使用名称属性。例如，以下示例 CLI 命令通过用户管理员在 https://localhost:8089 的 Splunk® Enterprise 实例上新建了名为 mytoken 的标记：

```
curl -k -u admin:changeme https://localhost:8089/servicesNS/admin/splunk_httpinput/data/inputs/http -d name=mytoken
```

使用 cURL 编辑 HTTP 事件收集器标记

您可以使用 cURL 更新任何标记属性（名称或值除外）。例如，以下示例 cURL 命令通过用户管理员在 https://localhost:8089 的 Splunk® Enterprise 实例上更新了 mytoken 标记的说明：

```
curl -k -u admin:changeme https://localhost:8089/servicesNS/admin/splunk_httpinput/data/inputs/http/mytoken  -d description=abc
```

您可以更新以下任何参数：

参数	描述
disabled	是否禁用标记。1 表示 true；0 表示 false。
描述	标记描述。
indexes	标记接受的索引列表。
index	标记的默认索引。Splunk® Enterprise 会将此值分配给尚未设置索引值的数据。
source	标记的默认来源值。Splunk® Enterprise 会将此值分配给尚未设置来源值的数据。
sourcetype	标记的默认来源类型值。Splunk® Enterprise 会将此值分配给尚未设置来源类型值的数据。
outputgroup	标记的默认输出组值。输出组是一组由 Splunk 软件管理员设置用于索引数据的索引器。Splunk® Enterprise 会将此值分配给尚未设置输出组值的数据。
port	HTTP 事件收集器服务器端口。默认值为 8088，但是您可以使用此参数更改。
enableSSL	HTTP 事件收集器服务器协议是 HTTP 还是 HTTPS。1 表示 HTTPS；0 表示 HTTP。
dedicatedIoThreads	HTTP 事件收集器服务器上分配器线程数量。默认值为 2。除非 Splunk 支持人员要求，否则不应更改此设置。此参数值不得超过 Splunk® Enterprise实例上物理 CPU 核心的数量。
useACK	索引事件之后返回确认。设为 1 启用。

使用 cURL 启用或禁用 HTTP 事件收集器标记

您可以使用 cURL 启用或禁用标记。更改某标记的状态不会更改到其他标记的状态。要启用或禁用标记，请使用 HTTP POST 命令、标记名称、启用或禁用端点。例如，以下命令通过用户管理员在 https://localhost:8089 的 Splunk® Enterprise 实例上禁用名为 mytoken 的标记：

```
curl -k -X "POST" -u admin:changeme https://localhost:8089/servicesNS/admin/splunk_httpinput/data/inputs/http/mytoken/disable
```

类似地，以下示例通过用户管理员在 https://localhost:8089 的 Splunk® Enterprise 实例上启用名为 mytoken 的标记：

```
curl -k -X "POST" -u admin:changeme https://localhost:8089/servicesNS/admin/splunk_httpinput/data/inputs/http/mytoken/enable
```

使用 cURL 启用或禁用 HTTP 事件收集器

您可以使用 cURL 批量更改所有标记启用或禁用 HTTP 事件收集器本身。使用启用或禁用端点时不要指定标记名称。要启用或禁用 HTTP 事件收集器，请使用 HTTP POST 命令、启用或禁用端点。例如，以下示例通过用户管理员在

https://localhost:8089 的 Splunk® Enterprise 实例上禁用 HTTP 事件收集器：

```
curl -k -X "POST" -u admin:changeme https://localhost:8089/servicesNS/admin/splunk_httpinput/data/inputs/http/http/disable
```

使用 cURL 删除 HTTP 事件收集器标记

要使用 cURL 删除标记，请使用 HTTP DELETE 命令和标记名称。例如，以下示例 CLI 命令通过用户管理员从 https://localhost:8089 的 Splunk® Enterprise 实例中删除名为 mytoken 的标记：

```
curl -k -X "DELETE" -u admin:changeme https://localhost:8089/servicesNS/admin/splunk_httpinput/data/inputs/http/mytoken
```

使用 cURL 管理 HEC 事件和服务

以下命令显示您可以将事件发送到 HEC 服务和管理该服务的方式。此列表不是完整的列表，但是可以让您了解 HEC 可以完成哪些事情。

将事件发送到 HEC

以下示例说明了基本的 HEC 使用情况。包括含有端口和端点的 Splunk® Enterprise 实例地址、验证标记，以及根据 HEC 事件数据格式规格设定格式的事件数据和元数据。

```
curl -k "https://http-inputs-mysplunkserver.splunkcloud.com:8088/services/collector" \
    -H "Authorization: Splunk CF179AE4-3C99-45F5-A7CC-3284AA91CF67" \
    -d '{"event": "Hello, world!", "sourcetype": "manual"}'
```

使用基本验证将事件发送到 HEC

这个示例演示了基本验证，可用于替代 HTTP 验证。要使用基本验证，请将请求中用冒号分隔的用户各和密码对作为 `-u` 参数提交，其中我们将任意字符串用作用户名并将标记用作 `<password>`：`<user>:<password>`.

```
# Basic auth
curl -k -u "x:CF179AE4-3C99-45F5-A7CC-3284AA91CF67" "https://http-inputs-
mysplunkserver.splunkcloud.com:8088/services/collector/event" \
    -d '{"sourcetype": "mysourcetype", "event": "Hello, world!"}'
```

在一个请求中将多个事件发送到 HEC

以下示例演示了在一个请求中发送多个事件。尽管您可以在单个请求中发送多个事件，但是您不能跨多个请求拆分一个事件。

```
curl -k "https://http-inputs-mysplunkserver.splunkcloud.com:8088/services/collector" \
    -H "Authorization: Splunk CF179AE4-3C99-45F5-A7CC-3284AA91CF67" \
    -d '{"event": "Pony 1 has left the barn"}{"event": "Pony 2 has left the barn"}{"event": "Pony 3 has left the barn",
"nested": {"key1": "value1"}}'
```

将原始文本发送到 HEC

以下示例演示了将原始文本发送到 HEC。请注意原始端点，以及通道标识符和来源类型说明的使用，后两者都是使用 URL 查询参数完成。

```
curl -k "https://http-inputs-mysplunkserver.splunkcloud.com:8088/services/collector/raw?channel=00872DC6-AC83-4EDE-8AFE -
8413C3825C4C&sourcetype=mydata" -H "Authorization: Splunk CF179AE4-3C99-45F5-A7CC-3284AA91CF67" -d '1, 2, 3... Hello, world!'
```

将原始批处理事件发送到 HEC

以下示例演示了如何将原始批处理事件发送到 HEC。在这种情况下，命令会跨多个日志发送 splunkd。该命令指示索引器必须将这些事件分配给 `splunkd_access` 的来源类型，并指定必须将这些事件发送到 `main` 索引。

```
# HEC Raw batching
curl -k "https://http-inputs-mysplunkserver.splunkcloud.com:8088/services/collector/raw?channel=00872DC6-AC83-4EDE-8AFE -
8413C3825C4C&sourcetype=splunkd_access&index=main" \
    -H "Authorization: Splunk CF179AE4-3C99-45F5-A7CC-3284AA91CF67" \
    -d '127.0.0.1 - admin [28/Sep/2016:09:05:26.875 -0700] "GET /servicesNS/admin/launcher/data/ui/views?count=-1 HTTP/1.0" 200
```

```
126721 - - - 6ms
127.0.0.1 - admin [28/Sep/2016:09:05:26.917 -0700] "GET /servicesNS/admin/launcher/data/ui/nav/default HTTP/1.0" 200 4367 - - -
6ms
127.0.0.1 - admin [28/Sep/2016:09:05:26.941 -0700] "GET /services/apps/local?search=disabled%3Dfalse&count=-1 HTTP/1.0" 200
31930 - - - 4ms
127.0.0.1 - admin [28/Sep/2016:09:05:26.954 -0700] "GET /services/apps/local?search=disabled%3Dfalse&count=-1 HTTP/1.0" 200
31930 - - - 3ms
127.0.0.1 - admin [28/Sep/2016:09:05:26.968 -0700] "GET /servicesNS/admin/launcher/data/ui/views?digest=1&count=-1 HTTP/1.0"
200 58672 - - - 5ms'
```

启用索引器确认将事件发送到 HEC

以下示例演示了如何通过启用索引器确认将事件发送到 HEC。此示例和基本示例之间唯一的区别在于是否包括通道标识符。索引器确认也可以和原始数据结合使用。

```
# Indexer ack
curl -k "https://http-inputs-mycompany,splunkcloud.com:8088/services/collector?channel=00872DC6-AC83-4EDE-8AFE -8413C3825C4C" \
    -H "Authorization: Splunk CF179AE4-3C99-45F5-A7CC-3284AA91CF67" \
    -d '{"event": "Hello, world!", "sourcetype": "manual"}'
```

检查 HEC 索引器确认状态

以下示例演示了如何检查之前的 HEC 请求的索引状态。该示例会发送请求到 **ack** 端点，包括 `acks` 键，该键被设为要查询状态的三种确认标识符（ackID）。

```
# Check ack status
curl -k "https://http-inputs-mysplunkserver.splunkcloud.com:8088/services/collector/ack?channel=00872DC6-AC83-4EDE-8AFE -
8413C3825C4C" \
    -H "Authorization: Splunk CF179AE4-3C99-45F5-A7CC-3284AA91CF67" \
    -d '{"acks": [1,3,4]}'
```

从发送到 HEC 的事件中提取 JASON 字段

以下示例演示了如何指示 Splunk® Enterprise 从发送到 HEC 的事件中提取 JSON 字段。

```
# Extracting JSON fields
curl -k "https://http-inputs.mysplunkserver.splunkcloud.com:8088/services/collector" \
    -H "Authorization: Splunk CF179AE4-3C99-45F5-A7CC-3284AA91CF67" \
    -d '{"sourcetype": "_json", "event": {"a": "value1", "b": ["value1_1", "value1_2"]}}'
```

从发送到 HEC 的事件中提取显式 JASON 字段

以下示例和之前的示例类似，但是明确指定了 JSON 字段。

```
# Explicit JSON fields
curl -k "https://mysplunkserver.example.com:8088/services/collector/event" \
    -H "Authorization: Splunk CF179AE4-3C99-45F5-A7CC-3284AA91CF67" \
    -d '{"event": "Hello, world!", "sourcetype": "cool-fields", "fields": {"device": "macbook", "users": ["joe", "bob"]}}'
```

关于 HTTP 事件收集器索引器确认

HTTP 事件收集器（HEC）仅支持 Splunk Enterprise 中的**索引器确认**。Splunk Cloud 不支持 HEC 中的索引器确认。

虽然用途相似，名称相同，但是 HEC 中的索引器确认和用于转发的索引器确认功能不一样。有关转发的索引器确认的信息，请参阅*《Splunk Enterprise 转发数据》*手册中的“防止传输中的数据丢失”。Splunk Cloud 也不支持基于转发的索引器确认。

为什么选择索引器确认

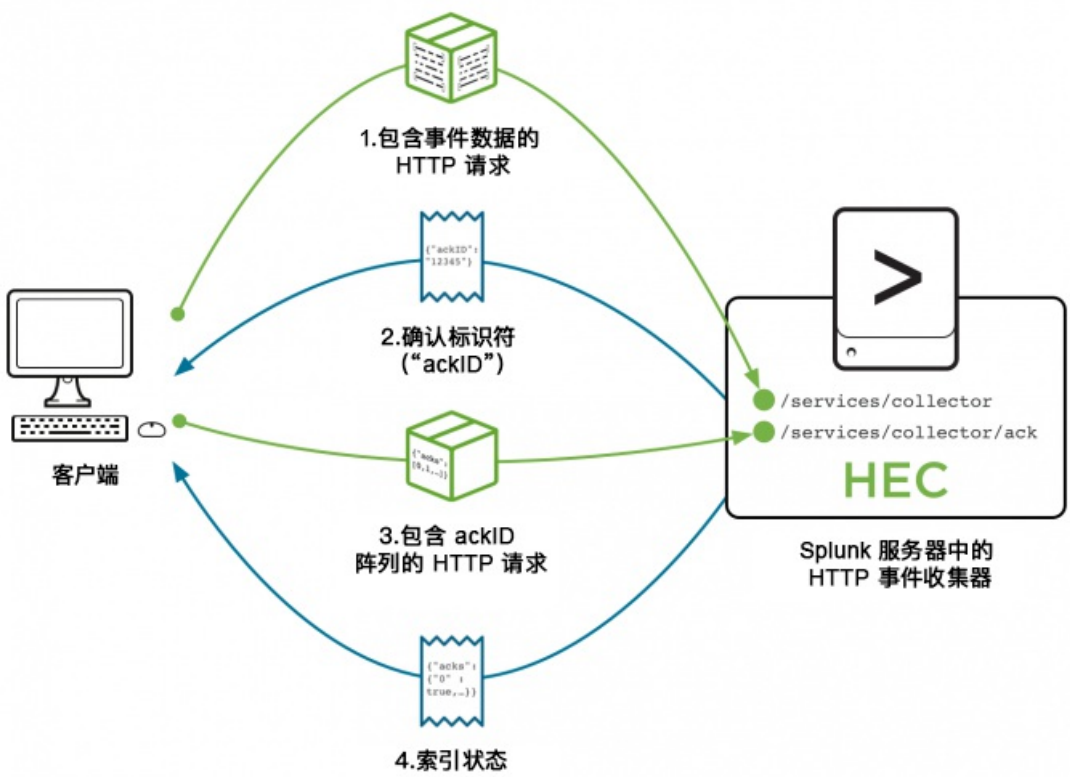
默认情况下，如果 HEC 成功接收事件，则会立即向数据的发送者发送 HTTP 状态 200 代码。但是，这仅仅表明该事件数据看起来有效，并且 HEC 在事件数据进入处理管道之前发送状态消息。处理期间，由于中断或系统故障，有几个地方的事件可能在 Splunk Enterprise 建立索引前丢失。尽管 HEC 已采取措施防止数据丢失，但是要完全防止这种情况是不可能的，特别是在网络故障或硬件崩溃的情况下。这时需要启用索引器确认。

索引器确认如何工作

您可以按标记启用索引器确认。索引器确认过程和以下包装追踪情景类似：

运输公司在运送包裹时会发出追踪号码。运输公司在交付最终号码后更新追踪号码的状态，届时，您方便时可使用追踪号码检索状态，检查包裹是否成功抵达。

下图从上到下按顺序说明了索引器确认过程。后面段落中用编号指代每个步骤：



每次客户端使用启用索引器确认的标记将请求发送到 HEC 端点 (1) 时；HEC 会将确认标识符返回到客户端 (2)。响应正文是一个带确认标识符的 JavaScript Object Notation (JSON) 对象，如下所示：

```
{"ackID":"2"}
```

然后，客户端可以使用标识符查询 HEC，确认已为请求中与标识符相对应的所有事件建立索引 (3)。客户端将查询发送到特定端点 (/services/collecto

```
{"acks":[0,1,2]}
```

接下来，HEC 以状态信息响应客户端 (4)。回复正文包含客户端查询的每个请求的状态。true 状态表示已用所需要的复制因子复制与 ackID 对应的事件

```
{"acks": {"0": true, "1": false, "2": true}}
```

因为 false 状态可表示任意问题数量，因此只能在合理预期请求可能处于传输期的时间范围内查询 ackID。客户端检索到 ackID true 状态后，HEC 会

在 Splunk Enterprise 中为 HEC 启用索引器确认

您可以在 Splunk Web 中或 input.conf 配置文件中启用索引器确认。

使用 Splunk Web 为 HEC 启用索引器确认

您在 Splunk Web 中新建 HEC 标记后，请勾选第一个屏幕上标记为**启用索引器确**认的复选框，然后继续标记新建流程。

启用索引器确认

有关在 Splunk Web 中新建 HEC 标记的信息，请参阅“在 Splunk Web 中设置和使用 HTTP 事件收集器”。

使用 input.conf 配置文件启用索引器确认

您可以通过编辑 HEC 应用中的 input.conf 配置文件，为 Splunk Enterprise 上的现有标记启用索引器确认。

1. 打开 inputs.conf 文件，路径如下：
   - 在 *nix 中：$SPLUNK_HOME/etc/apps/splunk_httpinput/local/inputs.conf
   - 在 Windows 中：%SPLUNK_HOME%\etc\apps\splunk_httpinput\local\inputs.conf
2. 在与您想要启用索引器确认的标记对应的段落中，添加以下行： useACK=true
3. 保存并关闭文件。

关于通道和发送数据

启用索引器确认后将事件发送到 HEC 和未启用此设置发送事件类似。有一个至关重要的区别：当您启用索引器确认时，您必须在发送事件时指定一个*通道*。HEC 中引入了通道的概念，主要用于防止快速客户端妨碍慢速客户端的性能。当您为每个客户端分配一个通道后，因为 Splunk Enterprise 上通道处理在 HTTP 请求中向 HEC 发送数据和请求确认是否为请求中包含的事件建立索引时，必须包括一个匹配通道标识符。如果没有，您将收到错误消息“数据通道

```
curl https://mysplunk.com/services/collector -H "X-Splunk-Request-Channel: FE0ECFAD-13D5-401B-847D-77833BD77131" -H "Authorizati
```

或者，X-Splunk-Request-Channel 标题字段可作为 URL 查询参数发送，如下所示：

```
curl https://mysplunk.com/services/collector?channel=FE0ECFAD-13D5-401B-847D-77833BD77131 -H "Authorization: Splunk BD274822-96A
```

索引器确认也可以和原始 JSON 数据结合使用。在此情况下，要在请求中使用的端点是 /services/collector/raw。请参阅“为 HTTP 事件收集器设置事通道设计使您可以将唯一通道分配给向 HEC 发送数据的每个客户端。每个通道都有一个通道标识符 (ID)，它必须是全局唯一标识符 (GUID)，可可以随机

索引状态的查询

为标记启用索引器确认后，客户端使用该标记向 HEC 发送的每个请求会向发送者返回以下包含在简单 JSON 对象中的确认标识符 (ackID)，其中 <int>

```
{"ackID":"<int>"}
```

要验证索引器是否已为请求中包含的事件建立索引，请查询以下端点，其中 <host> 和 <port> 分别表示 Splunk 平台实例的主机名和端口号：

```
https://<host>:<port>/services/collector/ack
{"acks":[0,1,2]}
```

以下是一个示例 cURL 语句，用于向 Splunk Enterprise 查询标识符为“0”、“1”、“2” 和 “3” 的请求中所包含事件的索引状态：

```
curl -k https://<host>:<port>/services/collector/ack?channel=FE0ECFAD-13D5-401B-847D-77833BD77131 -H "Authorizat
```

此查询中需要填写数据通道 ID (?channel=FE0ECFAD-13D5-401B-847D-77833BD77131) 和验证标题 ("Authorization: Splunk BD274822-96AA-4DA6

回复正文包含查询其状态的每个请求的状态。以下示例响应表示已成功为 ackID 为 "0" 和 "2" 的请求建议索引，但没有成功为 ackID 为 "1" 和 "

```
{"acks": {"0": true, "1": false, "2": true, "3": false}}
```

通道限制和索引状态到期

Splunk Enterprise 会在内存中缓存确认 ID 及其相应的状态信息。要防止服务器内存不足和阻止恶意或不正常的客户端，引入了几种新的限制设置。

要防止通道过载并防止新建过多的通道，在 limits.conf 配置文件的 http_input 段落中引入了几种新设置：

设置	值类型	默认值	描述
max_number_of_acked_requests_pending_query_per_ack_channel	整数	1000000	指定了每个通道中等待查询的 ackID 的最大数量及其对应的状态信息。如果客户端在启用索引器确认的情况下做出很多请求，此设置可防止客户端通道充满 ackID 和状态信息，且防止客户端接收服务器繁忙错误。
max_number_of_ack_channel	整数	1000000	指定客户端可以为此 Splunk 服务器实例获取的最大通道数。如果单个客户端尝试获取的通道数多于此限制，则请求失败，显示服务器繁忙错误。此设置用于防止客户端获取过多通道。
max_number_of_acked_requests_pending_query	整数	10000000	指定了所有通道中 ackID 的最大数量及其对应的状态信息。

要防止达到限制，Splunk Enterprise 可清除闲置了一段时间的通道，释放这些通道占用的内存。您可以使用以下设置进行此操作，该设置可在 inputs.conf 配置文件中进行全局（[http] 段落）级别设置：

- 在 *nix 中：$SPLUNK_HOME/etc/apps/splunk_httpinput/local/inputs.conf
- 在 Windows 中：%SPLUNK_HOME%\etc\apps\splunk_httpinput\local\inputs.conf

参数	值类型	默认值	描述
ackIdleCleanup	布尔值	false	设为 true 后，此参数会导致服务器删除闲置了一定秒数（如 maxIdleTime 设置中所设）的通道。
maxIdleTime	整数	600	指定了删除通道前通道可闲置的最大秒数。

索引器确认客户端行为

为确保将数据成功引入到 Splunk 平台中，请配置客户端，让客户端可以对 HEC 端点返回的响应代码采取行动。如果客户端无法根据生成的响应代码采取行动，则可能会发生数据丢失。有关更多信息，请参阅“可能的错误代码”。

遵循这些准则确保连接到 HEC 的客户端不会出现任何恶意行为或最终达到本主题前面所述的限制。索引器确认客户端必须：

- 新建自身的 GUID 以用作通道标识符。
- 仅使用该通道发送请求。
- 保存请求后 HEC 返回的每个确认标识符（ackID）。
- 定期（例如，每 10 秒）连续轮询 /services/collector/ack REST 端点，确认及时收到确认状态。由于 Splunk Enterprise 会在客户端检索状态信息后将其删除，这会释放服务器内存。
- 重新发送 HEC 在一定时间（例如，5 分钟）内未发送确认的任何事件数据。可以肯定，到此时间时，事件数据已丢失。重新发送事件数据后，最好的做法是在表示可能是重复数据的事件中添加其他额外数据。可能之前已为事件建立了索引，但状态由于清理通道而过期，或者可能已清除 HEC 状态缓存。

通过分布式部署扩展 HTTP 事件收集器

您可以将 HTTP 事件收集器（HEC）用作分布式 Splunk 平台部署的一部分。Splunk 软件会像处理任何其他输入一样处理 HEC 数据。

可以在 Splunk Cloud 部署上配置 HEC。当您配置 Splunk Cloud 分布式部署时，该部署中的索引器会自动配置 HEC。

如果要在 Splunk Enterprise 分布式部署上调整 HEC，请在继续之前熟悉分布式 Splunk Enterprise 部署。请参阅以下相关链接：

- 有关分布式部署的更多信息，请参阅*《分布式部署手册》*中的“使用 Splunk Enterprise 组件调整您的部署规模”，以及*《容量规划手册》*中的“Splunk Enterprise 部署的各个组件”。
- 有关部署服务器的更多信息，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”。

放置 HEC 的地方

HEC 可放在重型转发器或索引器上。和任何数据输入配置一样，当数据进入系统时，HEC 必须驻留在组件中。最佳做法是将 HEC 放在重型转发器上。**通用转发器**不支持 HEC。和任何其他数据输入配置一样，必要时，您也可以直接将 HEC 放在群集或非群集索引器上。

部署服务器分发任何包含 HEC 配置的应用。此配置包含以下信息：

- HTTP 事件收集器默认值（端口、SSL、来源类型、索引）
- SSL 设置
- HTTP 事件收集器标记

各 HEC 输入条目必须包含标记的有效通用唯一标识符 (UUID)。HEC 将其配置存储在 $SPLUNK_HOME/etc/apps/splunk_httpinput/ 目录或 Windows 上的 %SPLUNK_HOME%\etc\apps\splunk_httpinput\。

将 HEC 放置并分发到重型转发器上

在使用**转发器**的分布式 Splunk 平台部署上使用 HEC。使用部署服务器将 HEC 配置分发到 Splunk 平台部署中的重型转发器上。

如果您打算通过部署服务器分发 HEC 配置，请将 useDeploymentServer 选项（位于部署服务器上 inputs.conf 的 [http] 段落中）设置为 1。将此选项设为 1 并在部署服务器上做出基于 UI 的 HEC 更改之后，这些更改将直接放在 $SPLUNK_HOME/etc/deployment-apps/splunk_httpinput/ 文件夹中，而不是 $SPLUNK_HOME/etc/apps/splunk_httpinput/ 中。有关更多信息，请参阅“inputs.conf”。

请参阅*《转发数据》*手册中的“部署重型转发器”了解更多。

将 HEC 放置并分发到非群集索引器上

从以下两种方式中选择一种使用带非群集索引器的 HEC：

- 通过使用部署服务器将配置分发到 HEC，将 HEC 放置在转发给索引器的重型转发器上。
- 通过使用部署服务器将配置分发到 HEC，直接将 HEC 放在索引器上。

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”了解更多。

有关配置部署客户端的更多信息，请参阅*《更新 Splunk Enterprise 实例》*手册中的“配置部署客户端”。

将 HEC 放置并分发到索引器群集对等节点上

在使用**索引器群集**的分布式 Splunk 平台部署上使用 HEC。

使用配置软件包方法将 HEC 配置分发到对等节点。使用 HEC 端口号、首选协议（HTTP 或 HTTPS）、SSL 设置和 HTTP 事件收集器标记，将部署转发器上的 HEC 与部署索引器群集的对等节点相连。在运行群集管理器节点的 Splunk Enterprise 实例上集中管理标记。

请参阅*《管理索引器和索引器群集》*手册中的“使用转发器在索引器群集中获取数据”和“更新通用对等节点配置和应用”。

示例 serverclass.conf 文件

服务器类是一组可以作为一个单元进行管理的部署客户端。将您希望在 HEC 部署中使用的部署客户端分配给普通的服务器类。之后，在您将 HEC 设置分发给部署客户端之后，只有该服务器类的成员才能接收配置设置。以下示例 serverclass.conf 文件为 HEC 定义了服务器类 FWD2Local。

```
[global]
whitelist.0=*
restartSplunkd=true
stateOnClient = enabled

[serverClass:FWD2Local]
whitelist.0=*
```

```
[serverClass:FWD2Local:app:splunk_httpinput]
```

关于部署 HEC 设置，您可以将 HEC 视为名为 `splunk_httpinput` 的应用。在段落内，您可以设置客户端筛选属性和几个非筛选属性。

关于可用的客户端筛选属性的更多信息，请参阅*《更新 Splunk Enterprise 实例手册》*中的“通过 serverclass.conf 定义筛选器”。要了解有关可用非筛选属性的更多信息，请参阅*《更新 Splunk Enterprise 实例手册》*中的“使用 serverclass.conf 定义服务器类”。

另请参阅

有关分布式部署的更多信息，请参阅*《分布式部署手册》*中的“Splunk Enterprise 分布式部署概述”一章，以及*《容量规划手册》*中的“Splunk Enterprise 部署的各个组件”。

有关分布式部署，包括高级配置选项和常规示例的更多信息，请参阅《更新 Splunk 实例手册》。

为 HTTP 事件收集器设置事件格式

HTTP 事件收集器 (HEC) 通过一系列 HTTP 请求接收来自客户端的事件。各请求可包含一个 HEC 标记、通道标识符标题、事件元数据或事件数据，具体取决于您的事件是原始格式还是依据 JavaScript Object Notation (JSON) 的标准格式。您可以在 Splunk Cloud 和 Splunk Enterprise 中为 HEC 设置事件格式。

要了解有关 HEC、其工作方式以及如何设置 HEC 的更多信息，请参阅“在 Splunk Web 中设置并使用 HTTP 事件收集器”。

HEC 标记

在 HTTP 事件收集器可接受数据进行索引前，您必须验证运行收集器的 Splunk Cloud 或 Splunk Enterprise 实例。您可以使用新建 HEC 输入时生成的标记进行验证。当您在 Splunk 服务器上使用标记管理端点生成标记时，会生成全局唯一标识符 (GUID) 格式的标记。有关更多信息，请参阅“使用 cURL 管理 HTTP 事件收集器标记、事件和服务”。使用标记管理端点可以保证标记是唯一的。

您可以使用以下几种方式验证实例：通过 HTTP 验证、基本验证或查询字符串。

HTTP 验证

将标记放到个 HTTP 请求的验证标题中，如下所示：

```
"Authorization: Splunk <hec_token>"
```

以下示例显示了用于与 HEC 通信的典型 `curl` 命令上下文中的 HTTP 验证：

```
curl -k -H "Authorization: Splunk 12345678-1234-1234-1234-1234567890AB"
https://mysplunkserver.example.com:8088/services/collector/event -d '{"sourcetype": "my_sample_data", "event": "http auth
ftw!"}'
```

基本验证

要在 HEC 中执行基本验证，请在 `-u` 后面的请求中包含用冒号隔开的用户密码对，插入 HEC 标记作为 `<password>`: `"<user>:<password>"`。`<user>` 可以是任意字符串。例如：

```
-u "x:<hec_token>"
```

以下示例显示了 `curl` 命令上下文中的基本验证：

```
curl -k -u "x:12345678-1234-1234-1234-1234567890AB" https://mysplunkserver.example.com:8088/services/collector/event -d
'{"sourcetype": "my_sample_data", "event": "basic auth ftw!"}'
```

查询字符串

您可以将 HEC 标记指定为您在查询中为 HEC 指定的 URL 中的查询字符串。例如：

```
?token=<hec_token>
```

以下示例显示了 `curl` 命令上下文中的查询字符串验证：

```
curl -k https://mysplunkserver.example.com:8088/services/collector/event?token=12345678-1234-1234-1234-1234567890AB -d
'{"sourcetype": "my_sample_data", "event": "query string ftw!"}'
```

您还必须为每个标记启用查询字符串验证。通过 Splunk 支持打开案例以编辑 $SPLUNK_HOME/etc/apps/splunk_httpinput/local/inputs.conf 文件中的文件。此文件中按名称显示标记，形式为 http://<token_name>。

在您想要启用查询字符串验证的每个标记段落内，让 Splunk 支持添加或更改以下设置：

```
allowQueryStringAuth = true
```

在 Splunk Enterprise 上，您可以直接在实例上进行这些配置。重新启动实例后更改生效。

通道标识符标题

如果对 HEC 的请求包括原始事件，您必须在该请求中包括 X-Splunk-Request-Channel 标题字段。您必须将标题字段设为唯一通道标识符 (GUID)。有关更多信息，请参阅“关于通道和发送数据”。以下示例显示了构成有效请求的 cURL 命令：

```
curl https://http-inputs-<customer>.splunkcloud.com/services/collector/raw  -H "X-Splunk-Request-Channel: FE0ECFAD-13D5-401B-847D-77833BD77131" -H "Authorization: Splunk BD274822-96AA-4DA6-90EC-18940FB2414C" -d '<raw data string>' -v
```

或者，您可以将 X-Splunk-Request-Channel 标题字段设置为 URL 查询参数：

```
curl https://http-inputs-<customer>.splunkcloud.com/services/collector/raw?channel=FE0ECFAD-13D5-401B-847D -77833BD77131 -H "Authorization: Splunk BD274822-96AA-4DA6-90EC-18940FB2414C" -d '<raw data string>' -v
```

如果与 HTTP 事件收集器进行身份验证所使用的标记启用了索引器确认，您还必须在索引器状态查询中包含通道标识符。请参阅“启用索引器确认”了解更多信息。

事件元数据

下表显示了可包括在事件元数据中的键。这些键均是可选择的。任意未包括在事件中的键-值对将设为为 Splunk 平台实例上标记定义的值。

键	描述
"time"	事件时间。默认时间格式为 UNIX 时间格式，格式为 <sec>.<ms>。例如，1433188255.500 表示 epoch 后 1433188255 秒和 500 毫秒，或者 GMT 2015 年 6 月 1 日星期一下午 7:50:55。
"host"	要分配给事件数据的 host 值。此键通常是您发送数据所使用的客户端主机名。
"source"	要分配给事件数据的 source 值。例如，如果您正在从您正在开发的应用发送数据，则将此键设为应用名称。
"sourcetype"	要分配给事件数据的 sourcetype 值。
"index"	用来为事件数据建立索引的索引名称。如果标记已设置了 indexes 参数，您在此处指定的索引必须在允许索引列表内。
"fields"	fields 键不适用于原始数据。此键指定包含将在索引时间进行定义的显式自定义字段的 JSON 对象。包含 "fields" 属性的请求必须发送到 /collector/event 端点，否则不会对其进行索引。更多信息，请参阅“索引字段提取”。

有了原始事件，您可使用查询字符串在全局级别、标记层和请求层配置元数据。有关更多信息，请参阅“关于事件收集器标记”和“从 CLI 使用 HEC”。在请求中指定的元数据将应用于从请求提取的所有事件。

事件数据

事件数据可以分配给 HTTP 请求的 JSON 对象中的 "event" 键，或者可以是原始文本。JSON 事件包中的 "event" 键级别和元数据键一样。在"event" 键值的大括号内，数据可以是任何格式字符串、数字、其他 JSON 对象等等。

您可以通过将多个事件和请求组合在一起，将多个事件打到一个事件包中。通过批量处理事件，您可以指定请求的任意事件元数据都将应用于请求所包含的所有事件。如果您需要索引大量数据，批量处理事件可极大提高速度。

示例 1：事件元数据作为包含在 JSON 中的字符串

以下示例是格式正确的事件元数据和事件数据，作为包含在 JSON 对象中的字符串：

```
{
    "time": 1426279439, // epoch time
    "host": "localhost",
    "source": "random-data-generator",
    "sourcetype": "my_sample_data",
    "index": "main",
```

```
    "event":  "Hello world!"
}
```

示例 2：JSON 对象作为事件数据

以下示例是 JSON 对象作为格式正确的事件中的事件数据：

```
{
    "time": 1437522387,
    "host": "dataserver992.example.com",
    "source": "testapp",
    "event": {
        "message": "Something happened",
        "severity": "INFO"
    }
}
```

示例 3：批量处理数据

以下是批量处理数据的示例。HTTP 事件收集器的批处理协议涉及逐个堆叠的事件对象，而不是 JSON 数组。这些事件即使只包含 "event" 和 "time" 键却也仍然有效：

```
{
  "event":"event 1",
  "time": 1447828325
}

{
  "event":"event 2",
  "time": 1447828326
}
```

示例 4：带有授权标题的 cURL 语句

以下示例是一个简单的“hello world!”cURL 语句，其中包括授权标题、目的地端点和简单的事件数据。请注意：请求会进入 /services/collector/event 端点，其中所有 JSON 格式的事件请求必须进入：

```
curl -k -H "Authorization: Splunk 12345678-1234-1234-1234-1234567890AB" https://localhost:8088/services/collector/event -d
'{"event":"hello world"}'
```

示例 5：发送原始数据的 cURL 语句

以下示例 cURL 语句演示了发送原始事件数据。请注意：发送原始事件数据时必须添加通道 ID。另外，请求会进入 /services/collector/raw 端点，这也是所有原始事件请求必须进入的端点：

```
curl -k http://localhost:8088/services/collector/raw -H 'Authorization: Splunk B5A79AAD-D822-46CC-80D1-819F80D7BFB0' -H 'x-
splunk-request-channel: 18654C68-B28B-4450-9CF0-6E7645CA60CA' -d 'hello world'
```

示例 6：cURL 语句作为 URL 参数

或者，此示例 cURL 语句会将通道 ID 传送为 URL 参数：

```
curl -k http://localhost:8088/services/collector/raw?channel=18654C68-B28B-4450-9CF0-6E7645CA60CA -H 'Authorization: Splunk
B5A79AAD-D822-46CC-80D1-819F80D7BFB0'  -d 'hello world'
```

关于如何使用 cURL 设置事件数据格式并发送数据的其他示例，请参阅“使用 cURL 管理 HTTP 事件收集器标记、事件和服务”。

事件分析

HTTP 事件收集器端点会从 HTTP 请求提取事件，并在发送到索引器之前对这些事件进行分析。由于事件数据格式是预先决定的，Splunk 平台能够快速分析数据并为数据建立索引。和其他导入数据的方法相比，这种更快的分析可以提高数据吞吐量，并缩短事件处理时间。

您可在 props.conf 配置文件中配置提取规则。要了解更多信息，请参阅“配置基于规则的来源类型识别”。

原始事件分析

原始事件分析在 Splunk Cloud 的当前版本和 Splunk Enterprise 6.4.0 及更高版本中可用。

HTTP 事件收集器可分析原始文本并提取一个或更多事件。HEC 期望 HTTP 请求包含一个或多个包含有效换行规则的事件。HEC 接受请求后，会将其事件传入会提取字段（如时间戳）的管道。HEC 会使用基于时间戳的换行策略，但是您可以在 props.conf 配置文件中设置来源类型覆盖此策略。请参阅“配置基于规则的来源类型识别”。

您必须在单个 HTTP 请求中包含事件。无法跨多个请求。

要容纳原始事件，请使用 services/collector/raw 端点。

此端点需要补充 X-Splunk-Request-Channel 标题字段，您必须将该字段设为唯一的通道标识符（GUID）。有关更多信息，请参阅“关于通道和发送数据”。每个包含原始事件的 HTTP 请求必须包含通道标识符。以下为一个 cURL 语句示例，构成一个有效请求：

```
curl https://http-inputs-<customer>.splunkcloud.com/services/collector/raw  -H "X-Splunk-Request-Channel: FE0ECFAD-13D5-401B-847D-77833BD77131" -H "Authorization: Splunk BD274822-96AA-4DA6-90EC-18940FB2414C" -d '<raw data string>' -v
```

或者，X-Splunk-Request-Channel 标题字段可作为 URL 查询参数发送：

```
curl https://http-inputs-<customer>.splunkcloud.com/services/collector/raw?channel=FE0ECFAD-13D5-401B-847D -77833BD77131 -H "Authorization: Splunk BD274822-96AA-4DA6-90EC-18940FB2414C" -d '<raw data string>' -v
```

如果与 HTTP 事件收集器进行身份验证所使用的标记启用了索引器确认，您还必须在索引器状态查询中包含通道标识符。请参阅“启用索引器确认”了解更多信息。

有了原始事件，您可使用查询字符串在全局级别、标记层和请求层配置元数据。有关更多信息，请参阅“关于事件收集器标记”和“从 CLI 使用 HEC”。在请求中指定的元数据将应用于从请求提取的所有事件。

来源类型层已启用时间戳提取规则以提取时间戳。可以识别最常见的时间戳格式，例如 "current-time" 键，但是如果无法提取任何时间戳，则会根据当前时间分配一个。关于其他元数据，您可以在 props.conf 文件中配置提取规则。有关更多信息，请参阅“配置基于规则的来源类型识别”。

有关 services/collector/raw 端点 cURL 请求的更多示例，请参阅 Splunk Enterprise *《REST API 参考手册》*中的“输入端点示例”。

有关通道的更多信息，请参阅“启用索引器确认”中的*关于通道和发送数据*。

用 HTTP 事件收集器自动进行索引字段提取

Splunk® Enterprise 为数据创建索引时，它将数据流分析为一系列事件。在此过程中，它会将许多字段添加到事件数据中。这些字段包括 Splunk 软件自动添加的默认字段和您指定的任何自定义字段。向事件添加字段的过程称为字段提取。有两类字段提取：搜索-时间字段提取和索引字段提取。索引时将索引字段并入索引，并成为事件数据的一部分。

索引字段提取无法和发送到 **services/collector/raw** 端点的数据结合使用。有关更多信息，请参阅 Splunk Enterprise *《REST API 参考手册》*中的“services/collector/raw”。

表单 HEC 请求触发索引字段提取

您可以以两种方式触发 JavaScript Object Notation（JSON）字段的索引提取：作为主要 event 数据的一部分或和 event 数据分离，但仍与事件关联。

在 event 属性中使用嵌套 JSON

将 event 属性（在发送给 HEC 的 JSON 顶层）分配给 JSON 对象，该对象包含要索引作为键值对的自定义字段。例如，以下 event 属性（从 HTTP 请求发送到 HEC）指定了两种自定义字段： club 和 wins。

```
"event": {"club":"glee", "wins":["regionals","nationals"]}
```

在此示例中，wins 属性为多值 JSON 数组。wins 字段被分配数组中的两个值。

在与 event 属性相同的等级上，您还必须包括 sourcetype 属性，并将其设为已启用索引提取的来源类型。您可以在 props.conf 配置文件中使用将INDEXED_EXTRACTIONS 设置配置为 JSON 的任何来源类型，包括内置来源类型（如 _json）。请参阅以下示例：

```
"sourcetype":"_json"
```

以下为将事件发送到 Splunk® Enterprise 实例上的 HEC 的 cURL 命令示例。在此情况中，事件数据中包含两个将在索引时间提取的自定义字段：

```
# Extracting JSON fields
curl -k https://mysplunkserver.example.com:8088/services/collector -H "Authorization: Splunk 12345678-1234-1234-1234-1234567890AB" -d '{"sourcetype": "_json", "event": {"club":"glee", "wins":["regionals","nationals"]}}'
```

在 JSON 顶层添加 fields 属性

包括发送到 HEC 的 JSON 顶层的 `fields` 属性，即与 `event` 属性等级相同。添加此属性指定了与主要 `event` 数据分离的显式自定义字段。如果您不想将自定义字段包含在事件数据中，但是想要用一些其他信息（如数据位置）来注释数据，则此方法有用。通常来说，使用此方法也比嵌套 JSON 方法更快。

请注意：您必须将包含 `fields` 属性的 HEC 请求发送到 **`/collector/event`** 端点。否则，不会对其进行索引。

将 `fields` 属性分配给 JSON 对象，该对象包含作为键值对索引的自定义字段。例如，以下 `fields` 属性（从 HTTP 请求发送到 Splunk 平台实例）指定了两种自定义字段：`club` 和 `wins`。

```
"fields": {"club":"glee", "wins":["regionals","nationals"]}
```

在此示例中，`wins` 属性设置为多值 JSON 数组。`wins` 字段被分配数组中的两个值。

在与 `event` 和 `fields` 属性相同的等级上，您还必须包括 `sourcetype` 属性，将其设为已启用索引提取的来源类型。您可以在 `props.conf` 文件中使用将 `INDEXED_EXTRACTIONS` 设置配置为 `JSON` 的任何来源类型，包括内置来源类型（如 `_json`）。请参阅以下示例：

```
"sourcetype":"_json"
```

以下为将事件发送到 Splunk® Enterprise 实例上的 HEC 的 cURL 命令示例。在此情况中，事件数据中包含两个将在索引时间提取的自定义字段：

```
# Explicit JSON fields
curl -k https://mysplunkserver.example.com:8088/services/collector/event -H "Authorization: Splunk 12345678-1234-1234-1234-1234567890AB" -d '{"event": "Hello, McKinley High!", "sourcetype": "_json", "fields": {"club":"glee", "wins":["regionals","nationals"]}}'
```

只有字符串可用作字段值。

搜索索引-提取字段

索引数据后，您可以使用双冒号（ `::` ）索引提取表示法搜索此事件，显示如下：

```
sourcetype=_json club::glee
```

关于使用提取字段检索事件的更多信息，请参阅 Splunk Enterprise *《搜索手册》* 中的“使用字段检索事件”。

将指标发送到指标索引

如果您收集了指标数据，可使用 HTTP 事件收集器（HEC）直接将该数据发送到指标索引。

Splunk Enterprise *《指标》* 手册中的以下主题提供了将数据发送到指标索引的最新说明：

- 从 collectd 导入指标
- 从其他来源导入指标

另请参阅

- Splunk Enterprise *《指标》* 手册中的“指标概述”
- Splunk Enterprise *《管理索引器和索引器群集》* 手册中的“创建自定义索引”
- *《REST API 参考手册》* 中的“/collector 和 /collector/raw”

HTTP 事件收集器 REST API 端点

您可以在 Splunk Enterprise《**REST API 参考手册**》中找到完整的 HTTP 事件收集器 REST API 端点参考。为方便起见，每个端点都已链接至此处。

REST API 端点	描述
data/inputs/http	访问或更新 HTTP 事件收集器全局配置标记和应用程序标记。
data/inputs/http/{name}	管理 {name} HTTP 事件收集器标记。HTTP，如 data/inputs/http/http 中所示，表示全局配置。
data/inputs/http/{name}/disable	禁用 {name} HTTP 事件收集器标记。
data/inputs/http/{name}/enable	启用 {name} HTTP 事件收集器标记。
services/collector	使用 Splunk 平台 JavaScript Object Notation（JSON）事件协议将事件发送到 HTTP 事件收集器。
services/collector/ack	查询事件索引状态。
services/collector/event	当 /event URL 中的 auto_extract_timestamp 参数设置为 true 时，使用 Splunk 平台 JSON 事件协议将带有时间戳的事件发送到 HTTP 事件收集器。 • 时间戳示例如下 2017-01-02 00:00:00 • 如果事件的 JSON 信封中有时间戳，则 Splunk 会首先接受该时间戳。 • 如果事件的 JSON 信封中没有时间戳，则合并管道将从事件中提取时间戳。 • 如果 /event URL 中使用 time=xxx，那么会禁用 auto_extract_timestamp。
services/collector/event/1.0	此端点的工作方式和 **services/collector/event** 端点一样，但为了方便未来实现可伸缩性而引入协议版本。
services/collector/health	检查 HTTP 事件收集器的运行状况。Splunk Cloud 和 Splunk Enterprise 版本 6.6.0 及更高版本支持此端点。
services/collector/mint	将 Splunk MINT 格式的数据发布到 HTTP 事件收集器。
services/collector/mint/1.0	此端点的工作方式和 **receivers/token/mint** 端点一样，但为了方便未来实现可伸缩性而引入协议版本。
services/collector/raw	将原始数据直接发送到 HTTP 事件收集器。
services/collector/raw/1.0	此端点的工作方式和 **services/collector/raw** 端点一样，但为了方便未来实现可伸缩性而引入协议版本。
services/collector/s2s	该端点通过 HTTP 从 Splunk 通用转发器接收 Splunk TCP 数据。与 Splunk Enterprise 8.1.0 和更高版本兼容。

HTTP 事件收集器示例

HTTP 事件收集器（HEC）输入具有无数的使用案例。以下示例说明了如何使用 HEC 为数据流建立索引。示例还显示了必须将数据发送到 HEC 输入的方式。您可以使用这些示例对如何将您自己的数据发送到 Splunk Cloud 或 Splunk Enterprise 中的 HEC 进行建模。

此页面上的示例使用 curl 命令。通常，示例命令使用如下参数：

参数	描述
-d	使用此参数向 HEC 提供事件。您可以将原始文本或 JSON 格式的文本发送到 HEC。
-u	使用此参数指定用户。使用基本验证时需要此参数。
-H	使用此参数指定标头。无论您是使用 HTTP 验证还是基本验证，您都必须提供标头以将事件提交给 HEC。标头是您包含 HEC 标记的方式。

-k 参数不安全，请勿用它来检查安全证书。请勿在生产环境中或需要安全性的地方使用此参数。

向 HEC 提交事件不需要使用 curl 命令。您可以使用任何与 HTTP 和 REST 规范兼容的工具或应用程序。

示例 1：基本示例

此示例说明了基本的 HEC 使用情况。包括 Splunk 平台实例地址、端口和 REST 端点，以及验证标记、事件数据和元数据。此示例根据 HEC 事件数据格式规格设定格式。

```
curl -k "https://mysplunkserver.example.com:8088/services/collector" \
    -H "Authorization: Splunk CF179AE4-3C99-45F5-A7CC-3284AA91CF67" \
    -d '{"event": "Hello, world!", "sourcetype": "manual"}'
```

示例 2：将多个事件发送到 HEC

此示例演示了如何在一个请求中发送多个事件。尽管您可以在一个请求中发送多个事件，但是您不能跨多个请求拆分一个事件。

```
curl -k "https://mysplunkserver.example.com:8088/services/collector" \
    -H "Authorization: Splunk CF179AE4-3C99-45F5-A7CC-3284AA91CF67" \
    -d '{"event": "Pony 1 has left the barn"}{"event": "Pony 2 has left the barn"}{"event": "Pony 3 has left the barn",
"nested": {"key1": "value1"}}'
```

示例 3：将原始文本发送到 HEC

此示例演示了将原始文本发送到 HEC。要发送原始文本，您必须使用**原始**端点以及通道标识符和来源类型规范。您可以使用 URL 查询参数提交这两项设置。

```
curl -k "https://mysplunkserver.example.com:8088/services/collector/raw?channel=00872DC6-AC83-4EDE-8AFE -
8413C3825C4C&sourcetype=mydata" -H "Authorization: Splunk CF179AE4-3C99-45F5-A7CC-3284AA91CF67" -d '1, 2, 3... Hello, world!'
```

示例 4：将多个原始文本事件发送到 HEC

以下示例演示了如何将原始批处理事件发送到 HEC。在这种情况下，命令会跨多个日志发送 splunkd。该命令指示索引器为这些事件分配 `splunkd_access` 的来源类型，并指定它们将进入主索引。

```
curl -k "https://mysplunkserver.example.com:8088/services/collector/raw?channel=00872DC6-AC83-4EDE-8AFE -
8413C3825C4C&sourcetype=splunkd_access&index=main" \
    -H "Authorization: Splunk CF179AE4-3C99-45F5-A7CC-3284AA91CF67" \
    -d '127.0.0.1 - admin [28/Sep/2016:09:05:26.875 -0700] "GET /servicesNS/admin/launcher/data/ui/views?count=-1 HTTP/1.0" 200
126721 - - - 6ms
127.0.0.1 - admin [28/Sep/2016:09:05:26.917 -0700] "GET /servicesNS/admin/launcher/data/ui/nav/default HTTP/1.0" 200 4367 - - -
6ms
127.0.0.1 - admin [28/Sep/2016:09:05:26.941 -0700] "GET /services/apps/local?search=disabled%3Dfalse&count=-1 HTTP/1.0" 200
31930 - - - 4ms
127.0.0.1 - admin [28/Sep/2016:09:05:26.954 -0700] "GET /services/apps/local?search=disabled%3Dfalse&count=-1 HTTP/1.0" 200
31930 - - - 3ms
127.0.0.1 - admin [28/Sep/2016:09:05:26.968 -0700] "GET /servicesNS/admin/launcher/data/ui/views?digest=1&count=-1 HTTP/1.0"
200 58672 - - - 5ms'
```

示例 5：HEC 事件数据的索引器确认

此示例演示如何通过传入 HEC 数据的索引器确认将事件发送到 HEC。此示例和基本示例之间的区别在于是否包括通道标识符。索引器确认也可以和原始数据结合使用。

```
curl -k "https://mysplunkserver.example.com:8088/services/collector?channel=00872DC6-AC83-4EDE-8AFE-8413C3825C4C" \
    -H "Authorization: Splunk CF179AE4-3C99-45F5-A7CC-3284AA91CF67" \
    -d '{"event": "Hello, world!", "sourcetype": "manual"}'
```

示例 6：检查索引器确认状态

此示例演示了如何检查之前的 HEC 请求的索引状态。该命令会发送请求到**确认** REST 端点，并包括 `acks` 键，该键设为要查看其状态的三个确认标识符。

```
curl -k "https://mysplunkserver.example.com:8088/services/collector/ack?channel=00872DC6-AC83-4EDE-8AFE-8413C3825C4C" \
    -H "Authorization: Splunk CF179AE4-3C99-45F5-A7CC-3284AA91CF67" \
    -d '{"acks": [1,3,4]}'
```

示例 7：提取 JSON 字段

此示例演示了如何指示 Splunk 平台从发送到 HEC 的事件中提取 JSON 字段。

```
curl -k "https://mysplunkserver.example.com:8088/services/collector" \
    -H "Authorization: Splunk CF179AE4-3C99-45F5-A7CC-3284AA91CF67" \
    -d '{"sourcetype": "_json", "event": {"a": "value1", "b": ["value1_1", "value1_2"]}}'
```

示例 8：显式 JSON 字段

此示例和之前的示例类似，但是明确指定了 JSON 字段。

```
curl -k "https://mysplunkserver.example.com:8088/services/collector/event" \
    -H "Authorization: Splunk CF179AE4-3C99-45F5-A7CC-3284AA91CF67" \
    -d '{"event": "Hello, world!", "sourcetype": "cool-fields", "fields": {"device": "macbook", "users": ["joe", "bob"]}}'
```

示例 9：基本验证

此示例演示了基本验证，可用于替代之前示例中使用的 HTTP 验证。要使用基本验证，请使用 `-u` 参数在请求中包括用冒号隔开的用户/密码对。您可以对 `<user>` 字符串使用任何内容，标记是 `<password>`。

```
curl -k -u "x:CF179AE4-3C99-45F5-A7CC-3284AA91CF67" "https://mysplunkserver.example.com:8088/services/collector/event" \
    -d '{"sourcetype": "mysourcetype", "event": "Hello, world!"}'
```

HTTP 事件收集器故障排除

您可以通过查看错误日志对 HTTP 事件收集器（HEC）进行故障排除。您还可以使用配置文件设置日志记录，使用监视控制台中包含的仪表板调查实例性能，以及检测其他缩放问题。

日志

HTTP 事件收集器将其自身相关数据保存到日志文件。您可以使用 Splunk Cloud 或 Splunk Enterprise 搜索这些使用指标以浏览系统范围内、每个标记、每个来源类型等使用趋势，以及评估 HEC 性能。只要启用了 HEC，就会记录指标。默认禁用 HEC，这样不会记录数据，除非您启用。

您还可以在 Splunk Enterprise 的 splunkd.log 日志文件中查看 HEC 错误日志。请参阅*《故障排除手册》*中的“启用调试日志记录”，了解如何在 Splunk Enterprise 实例上启用调试。

日志文件位置和管理

Splunk Enterprise 将 HTTP 事件收集器指标写入 $SPLUNK_HOME/var/log/introspection/splunk/http_event_collector_metrics.log 文件。

当注销并重新登录 Splunk Cloud 或启动 Splunk Enterprise 实例时，Splunk 平台会创建一个新的 http_event_collector_metrics.log 文件。任何具有该名称的现有文件都将被重命名。

您可以在 limits.conf 配置文件中配置 HTTP 事件收集器指标的记录频率。默认频率为 60 秒。HEC 会继续记录系统级别的指标，即使没有数据导入活动也如此。如果没有活动，每 24 小时约可以产生 200 千字节（KB）指标日志数据。指标日志文件最大大小为 25 百万字节（MB）。如果日志文件达到该限制，则 Splunk 平台会重命名日志文件并新建一个文件。一次最多可存储五个指标日志文件。

Props.conf 配置文件定义了读取和索引指标日志文件的参数。

搜索 HTTP 事件收集器指标数据

Splunk 平台将 HEC 指标数据放入 `_introspection` 索引。要使用 Splunk 平台搜索累计的 HEC 指标，请使用以下搜索命令：

```
index="_introspection" token
```

指标日志数据格式

Splunk 平台将 HEC 指标数据以 JSON 格式记录到日志中。这表明日志既易于读取，也可以和其他 Splunk Cloud 或 Splunk Enterprise 日志格式保持一致。单个条目由输入摘要指标（`series = http_event_collector`）和每个标记指标（`series =`

http_event_collector_token）组成，如以下示例所示：

```
{
   "datetime":"09-01-2016 19:21:19.014 -0700",
   "log_level":"INFO",
   "component":"HttpEventCollector",
   "data":{
      "series":"http_event_collector",
      "transport":"http",
      "format":"json",
      "total_bytes_received":0,
      "total_bytes_indexed":0,
      "num_of_requests":0,
      "num_of_events":0,
      "num_of_errors":0,
      "num_of_parser_errors":0,
      "num_of_auth_failures":0,
      "num_of_requests_to_disabled_token":0,
      "num_of_requests_to_incorrect_url":0,
      "num_of_requests_in_mint_format":0,
      "num_of_ack_requests":0,
      "num_of_requests_acked":0,
      "num_of_requests_waiting_ack":0
   }
}

{
   "datetime":"08-22-2016 12:38:04.854 -0700",
   "log_level":"INFO",
   "component":"HttpEventCollector",
   "data":{
      "token_name":"test",
      "series":"http_event_collector_token",
      "transport":"http",
      "format":"json",
      "total_bytes_received":57000,
      "total_bytes_indexed":44000,
      "num_of_requests":1000,
      "num_of_events":1000,
      "num_of_errors":0,
      "num_of_parser_errors":0,
      "num_of_requests_to_disabled_token":0,
      "num_of_requests_in_mint_format":0
   }
}
```

HEC 摘要指标

即使没有输入活动，Splunk 平台也会累计系统范围摘要指标。通过 `"series":"http_event_collector"` 识别这些指标。

有关 HEC 摘要指标的字段说明，请参阅下表：

字段	描述	值
组件	HTTP 事件收集器指标数据标识符。	HttpEventCollector
data:format	HTTP 事件收集器数据格式。	json
data:num_of_auth_failures	由于标记无效导致失败的验证总数。	无符号的整数
data:num_of_errors	每个标记错误总数，包括以下选项： • 数据格式不正确 • 没有授权 • 授权错误 • 连接问题	无符号的整数
data:num_of_events	HTTP 事件收集器端点接收的每个标记事件总数。	无符号的整数

data:num_of_parser_errors	由于事件数据格式不正确产生的每个标记分析错误的总数。	无符号的整数
data:num_of_requests	HTTP 事件收集器端点接收的有效的每个标记单个 HTTP 或 HTTPS 请求总数。每个请求可有一个或多个数据事件。	无符号的整数
data:num_of_requests_to_incorrect_url	对错误的 URL 的请求总数。	无符号的整数
data:num_of_requests_in_mint_format	来自 Splunk MINT 的请求总数。	无符号的整数
data:num_of_requests_to_disabled_token	每个标记请求禁用标记的请求总数。	无符号的整数
data:series	指标数据类型。	http_event_collector
data:total_bytes_indexed	发送到索引器的每个标记数据总数。	无符号的整数
data:total_bytes_received	通过调用 receive/token 端点接收的每个标记数据总数。	无符号的整数
data:transport	HTTP 事件收集器数据的数据传输协议。	http
datetime	与数据关联的日期和时间。采用以下格式： MM-DD-YYYY HH:MM:SS.SSS +/-GMTDELTA	字符串
log_level	记录严重性级别。	信息

每个标记指标

和系统范围摘要指标相反，仅当启用 HEC 时，Splunk 平台才累计每个标记指标。通过 `"series":"http_event_collector_token"` 识别这些指标。

limits.conf 配置文件中的 `[http_input]` 段落可定义记录间隔，以及为这些指标记录的标记最大数量。

有关每个标记指标的字段说明，请参阅下表：

字段	描述	值
组件	HTTP 事件收集器指标数据标识符。	HttpEventCollector
data:format	HTTP 事件收集器数据格式。始终以 JSON 格式记录指标。	json
data:num_of_errors	错误数量，包括以下内容： • 数据格式不正确 • 没有授权 • 授权错误 • 连接问题	无符号的整数
data:num_of_events	HTTP 事件收集器端点接收的事件数量。	无符号的整数
data:num_of_parser_errors	由于事件数据格式不正确产生的分析错误数量。	无符号的整数
data:num_of_requests	HTTP 事件收集器端点接收的有效的单个 HTTP 或 HTTPS 请求数量。每个请求可有一个或多个数据事件。	无符号的整数
data:num_of_requests_in_mint_format	来自 Splunk MINT 的请求总数。	无符号的整数
data:num_of_requests_to_disabled_token	禁用标记的请求数量。	无符号的整数
data:series	指标数据类型。	http_event_collector_token
data:token_name	标记名称。	字符串
data:total_bytes_indexed	发送到索引器的数据总数。	无符号的整数
data:total_bytes_received	通过调用 receive/token 端点接收的数据总数。	无符号的整数
data:transport	HTTP 事件收集器数据的数据传输协议。	http
datetime	与数据关联的日期和时间。采用以下格式： MM-DD-YYYY HH:MM:SS.SSS +/-GMTDELTA	字符串

log_level	记录严重性级别。	信息

使用配置文件记录

limits.conf 和 props.conf 文件控制指标数据记录和索引行为。

limits.conf

$SPLUNK_HOME/etc/system/default/limits.conf 文件中的 [http_input] 段落控制 HTTP 事件收集器指标数据记录。

有关所有 HTTP 事件收集器相关参数以及那些与指标不相关参数的信息，请参阅 Splunk Enterprise*《管理员手册》*中关于 limits.conf 的 [http_input] 段落文档。

Limits.conf 采用以下参数：

参数	默认值	描述
max_number_of_tokens	10000	表示 HTTP 事件收集器指标报告的标记最大数的无符号整数。
metrics_report_interval	60	表示 HTTP 事件收集器指标报表间隔秒数的无符号整数。

props.conf

$SPLUNK_HOME/etc/system/default/props.conf 文件中的 [http_event_collector_metrics] 段落控制读取和索引 HTTP 事件收集器日志文件。

请参阅以下示例：

```
[source::.../http_event_collector_metrics.log(.\d+)?]
sourcetype = http_event_collector_metrics

...

[http_event_collector_metrics]
SHOULD_LINEMERGE = false
TIMESTAMP_FIELDS = datetime
TIME_FORMAT = %m-%d-%Y %H:%M:%S.%l %z
INDEXED_EXTRACTIONS = json
KV_MODE = none
JSON_TRIM_BRACES_IN_ARRAY_NAMES = true
```

Props.conf 采用以下参数：

参数	默认	描述
SHOULD_LINEMERGE	false	指定每行事件的布局。设置为 true 允许同一行中有多个事件。设置为 false 将多个事件放到多个单独的行中。
TIMESTAMP_FIELDS	datetime	日志条目时间字段名称。
TIME_FORMAT	%m-%d- %Y %H:%M:%S.%l %z	日志条目时间字段格式。
INDEXED_EXTRACTIONS	json	指标日志格式。始终以 JSON 格式记录指标。
KV_MODE	无	键-值数据标识符。设置为 none 表示无键-值数据。始终没有指标记录。
JSON_TRIM_BRACES_IN_ARRAY_NAMES	true	是否从 JSON 数组名称中修剪大括号字符。

可能的错误代码

以下状态代码对所有 HTTP 事件收集器端点有特殊含义：

状态代码	HTTP 状态代码 ID	HTTP 状态代码	状态消息
0	200	确认	Success

1	403	禁止	Token disabled
2	401	未授权	Token is required
3	401	未授权	Invalid authorization
4	403	禁止	Invalid token
5	400	错误请求	No data
6	400	错误请求	Invalid data format
7	400	错误请求	Incorrect index
8	500	内部错误	Internal server error
9	503	服务不可用	Server is busy
10	400	错误请求	Data channel is missing
11	400	错误请求	Invalid data channel
12	400	错误请求	Event field is required
13	400	错误请求	Event field cannot be blank
14	400	错误请求	ACK is disabled
15	400	错误请求	Error in handling indexed fields
16	400	错误请求	Query string authorization is not enabled

为确保将数据成功引入到 Splunk 平台中，请配置客户端，让客户端可以对 HEC 端点返回的响应代码采取行动。如果客户端无法根据生成的响应代码采取行动，则可能会发生数据丢失。

使用监视控制台调查实例性能

监视控制台为 HEC 提供预构建仪表板，这样您可以用来调查实例性能。有关更多信息，请参阅下面的主题：

- 对于 Splunk Cloud，请参阅*《Splunk Cloud 管理员手册》*中的“监视您的 Splunk Cloud 部署”一章。
- 对于 Splunk Enterprise，请参阅*《监视 Splunk Enterprise》*手册中的“关于监视控制台”一章。

监视控制台可提供预构建仪表板以监视 HTTP 事件收集器。请参阅索引：输入：“HTTP 事件收集器”（*《监视 Splunk Enterprise》*手册中）。

检测缩放问题

如果您出现性能减速问题，或者想要提高 HTTP 事件收集器部署速度，以下因素可影响性能。

HTTP 和 HTTPS

通过 HTTP 发送数据和通过 HTTPS 发送数据相比，性能有明显提升。

批处理

如果您将多个事件批处理到单个请求中，可加快数据传输速度。因为请求元数据应用于请求中的所有事件，所以整体发送的数据减少。关于如何打包事件数据的更多信息，请参阅“为 HTTP 事件收集器设置事件格式”。

HTTP 处于活动状态

在连接上设置处于活动状态可提高性能。只要发送数据的客户端支持 HTTP 1.1 并设为支持 HTTP 永久连接，就可以通过处于活动状态优化性能。

保留队列

保留队列通过将输入队列中的数据存储到光盘来降低性能。有更多信息，请参阅“使用保留队列帮助防止数据丢失”。

获取其他类型数据的位置

监视先进先出（FIFO）队列

您可以通过编辑 Splunk Enterprise 实例上的 input.conf 配置文件来配置先进先出（FIFO）输入。如果您使用的是 Splunk Cloud，请使用重型转发器读取、索引和转发 FIFO 队列。Splunk Web 不支持 FIFO 输入的定义。

通过 FIFO 队列发送的数据不会保留在计算机内存中，这对于数据源而言并不是一种可靠的方法。要保证数据完整性，请使用**监视**输入。有关**监视**输入的更多信息，请参阅“监视文件和目录”。

将 FIFO 输入添加到 inputs.conf 中

如您之前未使用过配置文件，请在开始前先参阅 Splunk Enterprise *《管理员手册》* 中的“关于配置文件”。

要添加 FIFO 输入，请编辑 input.conf 文件并添加 FIFO 输入段落。请向 $SPLUNK_HOME/etc/system/local/ 目录中的 input.conf 文件或 $SPLUNK_HOME/etc/apps/ 中您自己的自定义应用程序目录添加段落。如果文件不存在，则可能需要创建该文件。

该输入段落将对 Splunk Enterprise 进行配置，以便从指定路径下的 FIFO 队列进行读取：

```
[fifo://<path>]
<setting1> = <val1>
<setting2> = <val2>
...
```

可以在 FIFO 输入段落中使用以下设置：

设置	描述	默认
host = <string>	将此段落的主机键或字段设为一个静态值。在 <string> 前面加上 host::。 此设置可设置主机键的初始值。分析和建立索引期间会使用该键来设置主机字段。它也在搜索时间使用该主机字段。	生成数据的主机的 IP 地址或完全限定域名。
index = <string>	存储来自此输入的事件的索引。在 <string> 前面加上 index::。	main 索引或已设置为默认索引的任何值。
sourcetype = <string>	来自此输入的事件的来源类型键或字段。此设置显式声明该数据的来源类型，而不是使其自动确定。声明来源类型对于可搜索性以及在分析和新建索引期间对此类型数据应用相关格式设置都很重要。 此设置可设置来源类型键的初始值。分析和建立索引期间会使用该键来设置来源类型字段。它还是在搜索时间使用的来源类型字段。 • 在 <string> 前面加上 sourcetype::。 • 有关来源类型的更多信息，请参阅“来源类型为何重要”。	Splunk 软件会根据数据的各个方面选取一种来源类型。没有硬编码的默认值。
source = <string>	设置来自此输入的事件的来源键或字段。在 <string> 前面加上 source::。 除非绝对必要，否则不要覆盖数据来源字段。输入层提供更准确的字符串来帮助分析和调查问题，并准确记录从中检索数据的文件。在覆盖此值之前，请考虑使用来源类型、标记和搜索通配符。	输入文件路径。
queue = [parsingQueue|indexQueue]	输入处理器用来存储其所读取事件的位置。 设置为 parsingQueue 时，会将 props.conf 文件和其他分析规则应用到您的数据。设置为 indexQueue 时，会将您的数据直接发送到索引。	默认为 parsingQueue。

监视对文件系统的更改

已弃用此功能。

Splunk Enterprise 5.0 版本中已弃用此功能。这意味着：尽管当前版本的 Splunk 平台继续保留此功能，但可能会从未来版本中删除。作为替代方法，您可以：

- 了解如何监视 Windows 系统上的文件系统更改。
- 使用 *nix 系统上的 auditd 守护程序和来自守护程序的监视器输出。

有关所有弃用功能的列表，请参阅*《发行说明》*中的“弃用功能”主题。

Splunk 平台**文件系统更改监视器**跟踪您的文件系统中发生的更改。监视器会对您指定的目录进行监视，当该目录发生任何更改时，它将生成一个事件。在系统中编辑、删除或添加任何文件时，此监视器都可以检测到。它能检测到任何文件的更改，包括非 Splunk 平台特定文件的文件。

例如，您可以配置文件系统更改监视器以监视 /etc/sysconfig/ 目录，并在系统配置发生更改时立即向您发出告警。

要了解如何使用内置的 Microsoft 审计工具监视 Windows 上的文件系统更改，请参阅“监视文件系统更改”。

文件系统更改监视器只能在 Splunk 平台的本地版本中运行。如果您使用 Splunk Cloud，则必须使用通用或重型转发器将文件系统更改数据发送到 Splunk Cloud 实例。

文件系统更改监视器如何工作

文件系统更改监视器使用以下属性检测 *nix 文件系统上的更改：

- 修改日期/时间
- 组 ID
- 用户 ID
- 文件模式（读取/写入属性等）
- 文件内容的 Secure Hash Algorithm-256 (SHA256) 哈希（可选）

可以配置文件系统更改监视器的以下功能：

- 使用正则表达式的允许列表
  - 指定要检查的文件（无论是什么文件）
- 使用正则表达式的拒绝列表
  - 指定要跳过的文件
- 目录递归
  - 包括符号链接遍历
  - 扫描多个目录，每个目录使用各自的轮询频率
- 加密签名
  - 新建文件系统更改的分布式审计线索
- 发生添加/更改事件时为整个文件新建索引
  - 用于发送整个文件和/或哈希的大小截止点
- 由 Splunk 平台新建索引并且可搜索的所有更改事件

只要 $SPLUNK_HOME/etc/ 目录的内容有任何进程更改、删除或添加，文件系统更改监视器就会生成**审计事件**。本地 Splunk 实例首次启动时，会为 $SPLUNK_HOME/etc/ 目录及其所有子目录中的每个文件生成审计事件。之后，配置中的任何更改（不考虑来源）都会为受影响的文件生成一个审计事件。

文件系统更改监视器根据您配置文件系统监视输入的方式将数据发送到各种索引。如果已为输入配置 signedaudit 设置，则实例会将文件系统更改发送到**审计索引**。如果尚未配置 signedaudit，则实例会将事件写入**主**索引，除非您指定另一个索引。

文件系统更改监视器仅跟踪发生的更改。文件系统更改监视器不会跟踪执行更改的帐户用户名。对于用户级监视，请考虑使用内置操作系统审计工具，这种工具可以访问上述信息。

不要将文件系统更改监视器配置为监视您的根文件系统。如果启用目录递归，这样做可能会耗用大量资源。

配置文件系统更改监视器的前提条件

- 如果您使用 Splunk cloud，则必须配置通用或重型转发器以连接到您的 Splunk Cloud 实例，以便将文件系统更改数据发送到该实例。
- 必须使用配置文件才能配置文件系统更改监视器。

配置文件系统更改监视器

要使用文件系统更改监视器来监视任何目录，请添加或编辑 [fschange] 段落至 inputs.conf 配置文件（路径为要从中收集文件系统更改信息的计算机上的 $SPLUNK_HOME/etc/system/local/ 目录，或您自己位于 $SPLUNK_HOME/etc/apps/ 的自定义应用程序目录）。有关配置文件的一般信息，请参阅《管理员》手册中的“关于配置文件”。

只要您更改了 inputs.conf 文件中的 [fschange] 段落就必须重启 Splunk 平台。

1. 在要向 Splunk Cloud 发送数据的 Splunk 平台实例上，使用文本编辑器打开 input.conf 配置文件。
2. 添加 [fschange:<directory>] 段落以指定 Splunk 平台应监视其更改的文件或目录。
3. 保存 inputs.conf 文件并将其关闭。
4. 重新启动 Splunk 平台实例。文件系统更改监视立即开始。

文件系统更改监视器语法

以下是 [fschange] 段落的语法：

```
[fschange:<directory or file to monitor>]
<setting1> = <val1>
<setting2> = <val2>
...
```

Splunk 平台监视目录及其子目录中发生的所有添加、更新和删除。任何更改都会生成一个事件，Splunk平台会对其建立索引。<directory or file to monitor> 默认为 $SPLUNK_HOME/etc/。

文件系统更改监视器的配置文件设置

文件系统更改监视器的所有设置都是可选的。以下是可用设置的列表：

设置	描述	默认
index=<indexname>	Splunk 平台存储它生成的所有事件的索引。	**main**，除非您启用了审计事件签名，否则为 **_audit**。
recurse=<true | false>	是否递归 [fschange] 中指定的目录内的所有目录。设置为 true 可递归所有子目录，设置为 false 则仅指定当前目录。	true
followLinks=<true | false>	文件系统更改监视器是否会跟踪符号链接。设置为 true 可跟踪符号链接，设置为 false 则不可跟踪符号链接。	false 如果在具有指向同一目录中文件的符号链接的目录上设置 followLinks，则可能会发生文件系统循环。
pollPeriod=N	每 N 秒检查一次此目录是否发生了更改。	3,600 秒 如果对受监视目录中的文件进行更改，则文件系统审计事件可能需要 1 到 3600 秒的时间进行生成并在审计搜索中变为可用。
hashMaxSize=N	为小于或等于 N（字节）的每个文件计算 SHA1 哈希。 此哈希可用作检测文件/目录更改的另一种方法。	-1，splunk平台不会计算用于更改检测的哈希）
signedaudit=<true | false>	发送经过加密签名的添加、更新或删除事件。 设置为 true 以在 _audit 索引中生成事件。如您正在配置 index 设置，请设置为 false。将 signedaudit 设置为 true 时，确保已在 audit.conf 配置文件中启用了审计。	false
fullEvent=<true | false>	如果检测到添加或更新更改，则发送完整事件。 受 sendEventMaxSize 属性进一步限制。	false

sendEventMaxSize=N	仅当事件大小小于或等于 N 字节时才发送完整事件。 这对已索引的文件数据的大小进行了限制。	-1（无限制）
sourcetype = <string>	设置来自此输入的事件的来源类型。 把 "sourcetype::" 加到 <string> 的前面。	audittrail （如果 signedaudit=true）或 fs_notification（如果 signedaudit=false）
filesPerDelay = <integer>	插入 delayInMills 指定的延迟时间，之前须先处理 <integer> 文件。此设置会对文件系统监视加以限制，使其不会消耗太多的 CPU。	n/a
delayInMills = <integer>	处理好每个 <integer> 文件（按照 filesPerDelay 中的指定）后使用的延迟时间（以毫秒为单位）。此设置用于对文件系统监视加以限制，使其不会消耗太多的 CPU。	
filters=<filter1>,<filter2>,...<filterN>	对于在监视器轮询周期内找到的每个文件或目录，都会按从左到右的顺序应用上面的每个筛选器。	n/a

为文件系统更改监视器定义筛选器

如需使用 filters 属性定义要使用的筛选器，请按下列步骤添加 [filter...] 段落：

```
[filter:blacklist:backups]
regex1 = .*bak
regex2 = .*bk
[filter:whitelist:code]
regex1 = .*\.c
regex2 = .*\.h

[fschange:/etc]
filters = backups,code
```

以下列表描述了 Splunk 平台如何处理 fschange 允许列表和拒绝列表逻辑：

- 各个事件将遍历筛选器列表，直到其找到第一个匹配。
- 如果与事件匹配的第一个筛选器是允许列表，则 Splunk 平台将为该事件新建索引。
- 如果与事件匹配的第一个筛选器是拒绝列表，则筛选器将阻止事件被索引。
- 如果事件到达链的末尾时未找到任何匹配，则 Splunk 平台将为该事件新建索引。

如果希望默认情况下事件未显式匹配允许列表时 Splunk 平台不为其新建索引，请在链的末尾添加将与所有剩余事件匹配的拒绝列表。

例如：

```
...
filters = <filter1>, <filter2>, ... terminal-blacklist

[filter:blacklist:terminal-blacklist]
regex1 = .?
```

如果您排除某个目录（包括在一系列允许列表之后添加末尾拒绝列表这种方式），则 Splunk 平台会排除其所有子文件夹和文件，因为它们不会通过任何允许列表。要适应这种情况，请在筛选器中的拒绝列表项之前显式包含所有所需的文件夹和子文件夹。

显式包含和末尾排除的示例

此配置将监视指定目录中扩展名为 .config、.xml、.properties 和 .log 的文件并忽略所有其他文件。

在此示例中，文件系统更改监视器可以排除目录。如果出现这种情况，则 Splunk 平台也会排除其所有子文件夹和文件。只对指定目录中的文件进行监视。

```
[filter:whitelist:configs]
regex1 = .*\.config
regex2 = .*\.xml
```

```
regex3 = .*\.properties
regex4 = .*\.log

[filter:blacklist:terminal-denylist]
regex1 = .?

[fschange:/var/apache]
index = sample
recurse = true
followLinks = false
signedaudit = false
fullEvent = true
sendEventMaxSize = 1048576
delayInMills = 1000
filters = configs,terminal-denylist
```

将文件系统更改监视器与通用转发器一起使用

要从通用转发器转发文件系统更改监视器事件，您必须使用 `signedaudit = false` 和 `index=_audit` 设置配置 `exchange` 段落。

```
[fschange:<directory or file to monitor>]
signedaudit = false
index=_audit
```

Splunk 平台将利用该解决方法为文件系统更改监视器事件新建 **_audit** 索引，其中来源类型为 `fs_notification`，来源为 `fschangemonitor`，而并非来源和来源类型为 `audittrail`。

通过脚本式输入从 API 及其他远程数据接口获取数据

Splunk Cloud 通过通用或重型转发器，可接受来自您提供的脚本的事件。

脚本式输入与 `ipconfig`、`iostat`、`netstat`、`top` 等 Windows 和 *nix 命令行工具结合使用非常有用。您可以使用脚本式输入从 API 和其他远程数据接口和消息队列中获取数据。之后，您可以对此数据使用 `vmstat` 和 `iostat` 一类的工具来生成指标和状态数据。在 Windows 平台上，可以使用中介 Windows 批处理文件（.bat）或 PowerShell（.ps1）文件来启用基于文本的脚本，例如，Perl 和 Python 脚本。

您可以从 Splunk Enterprise 上的 Splunk Web 中的“设置”菜单，或者通过编辑通用或重型转发器上的 input.conf 配置文件配置脚本式输入。

脚本式输入启动某个脚本时，该脚本将继承 Splunk Enterprise 或通用转发器环境。唯一可能引发脚本和脚本输出生成问题的环境变量是库路径（在 Linux、Solaris 和 FreeBSD 上通常称为 `LD_LIBRARY_PATH`）。在 Splunk 平台实例上使用脚本式输入时，请清除任何可能影响脚本操作的环境变量。

Splunk Enterprise 和通用转发器会将输入式脚本发送到 `stderr` I/O 通道的所有消息记录到 `splunkd.log`。

前提条件

要添加脚本式输入，必须先编写一个输入。如需了解如何编写脚本式输入，请参阅*《开发用于 Splunk Web 的视图和应用》*手册中的“构建脚本式输入”一章。

在 Splunk Web 中添加脚本式输入

在 Splunk Enterprise 或重型转发器实例上，按照以下高级步骤在 Splunk Web 中添加脚本式输入：

1. 转到“新增”页面。
2. 选择输入来源。
3. （可选）指定输入设置。
4. 查看您的选择。

转到“添加数据”页面

要使用“设置”转到“添加数据”页面，请执行以下步骤：

1. 在 Splunk Web 中单击**设置**。
2. 请单击**数据导入**。
3. 单击**脚本**。
4. 单击**新建**以添加输入。

要使用 Splunk Web 主页转到“添加数据”页面，请执行以下步骤：

1. 在 Splunk Web 中，单击**添加数据**。
2. 单击**监视**以监视本地计算机上的脚本或，单击**转发**以从远程计算机上的脚本转发数据。
   Splunk Web 将显示**添加数据 - 选择来源**页面。
3. 在左窗格中，选择**脚本**。

从脚本式输入中转发数据需要其他设置。

选择输入来源

1. 在**脚本路径**下拉列表中，选择脚本驻留的路径。
   Splunk Web 将更新该页面以包括**脚本名称**下拉列表。
2. 在**脚本名称**下拉列表中，选择要运行的脚本。
   Splunk Web 将更新该页面以用脚本名称填充**命令**字段。
3. 在**命令**字段中，添加任何所需的参数以调用该脚本。
4. 在**间隔**字段中，输入 Splunk Enterprise 在调用该脚本前等待的时间量（单位为秒）。
5. （可选）在**来源名称覆盖**字段中，如果需要，输入新的来源名称以覆盖默认来源值。
6. 单击**下一步**。

指定输入设置

您可在**输入设置**页面指定应用程序上下文、默认主机值和索引。所有这些参数均为可选参数。有关设置主机值的更多信息，请参阅“关于主机”。

设置此页面上的**主机**只是设置结果事件中的**主机**字段。而不是引导 Splunk Enterprise 查找网络中的特定主机。

1. 选择此脚本的来源类型。单击**选择**以从本地计算机上的可用来源类型列表中选取，或单击**手动**以输入来源类型名称。
2. 为此输入选择相应的**应用程序上下文**。
3. 设置**主机**名称。此设置有多个选项供您选择。
4. 设置 Splunk Enterprise 将数据发送到其中的**索引**。如果未定义多个索引来处理不同类型的事件，请保留**默认**值。除了用户数据的索引之外，Splunk Enterprise 还有多个实用工具索引，这些索引也会显示在此下拉列表中。
5. 请单击**查看**。

查看您的选择

在您指定所有输入设置后，可查看您的选择。Splunk Web 会列出您选定的所有选项，包括监视器的类型、数据来源、来源类型、应用程序上下文和索引。

1. 查看该设置。
2. 如果它们不符合您的期望，请单击左尖括号（ < ）即可返回到向导中的上一个步骤。否则，请单击**提交**。

Splunk Web 显示“成功”页面。

使用 inputs.conf 配置文件添加脚本式输入

在 inputs.conf 中添加脚本式输入，操作方法为：在该文件中添加一个 [script] 段落。您可以在 Splunk Enterprise 或通用转发器上执行此操作，然后将该信息转发到 Splunk Cloud。

语法

[script] 段落的语法如下所示，其中 $SCRIPT 是脚本位置的完整路径：

```
[script://$SCRIPT]
<attrbute1> = <val1>
<attrbute2> = <val2>
...
```

$SCRIPT 也可以是以 .path 后缀结尾的文件路径。该特殊的后缀允许您使用段落来指向另一个存在于主机文件系统中任意地方的命令或脚本。请参阅“使用 .path 后缀引用外部脚本”。您在段落中引用的文件必须遵循以下部分“脚本式输入的脚本放在哪里”中介绍的位置限制。

脚本式输入的脚本放在哪里

您在 $SCRIPT 中引用的脚本只能驻留在主机文件系统上以下位置中的任意一个：

- $SPLUNK_HOME/etc/system/bin
- $SPLUNK_HOME/etc/apps/<your_app>/bin
- $SPLUNK_HOME/bin/scripts

最佳方法是将您的脚本放入 bin/ 目录，该目录离在主机文件系统上调用您的脚本的 inputs.conf 文件最近。例如，如果您要配置 $SPLUNK_HOME/etc/system/local/inputs.conf，请将您的脚本放在 $SPLUNK_HOME/etc/system/bin/ 中。如果您要使用 $SPLUNK_HOME/etc/apps/$APPLICATION/ 中的应用程序，请将您的脚本放在 $SPLUNK_HOME/etc/apps/$APPLICATION/bin/ 中。

属性

所有属性均为可选项。以下是可用属性的列表：

属性	描述	默认
interval = <number>|<cron schedule>	运行指定命令的频率。可以指定代表秒数的整数值，也可以指定一个有效的 cron 计划。 指定 cron schedule 后，该脚本不会在启动时运行，而会在 cron 计划所定义的时间运行。 Splunk Enterprise 会针对每个实例保留脚本的一次调用。时间间隔基于脚本何时完成。如果您将脚本配置为每 10 分钟运行一次，且该脚本需要 20 分钟才能运行完成，下一次运行将会在第一次运行之后 30 分钟进行。 要获得稳定的数据流，请输入 1 或一个小于脚本时间间隔的值。要获得单次数据流，请输入 -1。把 interval 设置为 -1 会导致每次启动时都运行该脚本。	60 秒
index = <string>	存储来自此输入的事件的索引。Splunk Enterprise 会在 <string> 前面加上 index::。 更多有关索引字段的信息，请参阅*《管理索引器和索引器群集》*手册中的“索引如何工作”。	main，或已设置为默认索引的值。
sourcetype = <string>	设置来自此输入的事件的 sourcetype 字段。在 <string> 前面加上 sourcetype::。 显式声明该数据的来源类型，而不是使其自动确定。这对于可搜索性以及在分析和新建索引期间对此类型数据应用相关格式设置都很重要。 设置 sourcetype 键的初始值。Splunk Enterprise 在分析和新建索引过程中使用此键，特别是，可在新建索引期间使用此键来设置 sourcetype 字段。它也在搜索时间使用 sourcetype 字段。	没有硬编码的默认值。Splunk Enterprise 会根据数据的各个方面选取一种来源类型。
source = <string>	设置来自此输入的事件的 source 键。 除非绝对必要，否则不要覆盖 source 键。通常，输入层将提供更准确的字符串来帮助分析和调查问题，同时准确记录从中检索数据的文件。在覆盖此值之前，请考虑使用来源类型、标记和搜索通配符。 Splunk Enterprise 会在 <string> 前面加上 source::。	输入文件路径
disabled = <true | false>	输入是否运行。如果您想要禁用该输入，请设置为 true。	false

持续运行脚本

如果要持续运行脚本，可以将脚本编写为永不退出并设置一个较短的时间间隔。这样有助于确保出现问题时脚本可以重新启

动。Splunk Enterprise 会追踪衍生的脚本，并在退出时将其关闭。

使用包装脚本

最好针对将命令与参数结合使用的脚本式输入编写一个包装脚本。某些情况下，命令可以包含脚本式输入验证您在 Splunk Web 中所输入文本时会进行转义的特殊字符。对先前配置的输入进行的更新将无法保存。

在验证文本时，Splunk Enterprise 会对未含于路径中的字符进行转义，如等号 (=) 和分号 (;)。例如，以下脚本式输入在您于 Splunk Web 中对其进行编辑时未正确保存，因为脚本式输入把参数中的等号 (=) 转义成了 myUtil.py 实用工具：

```
[script://$SPLUNK_HOME/etc/apps/myApp/bin/myUtil.py file=my_datacsv]
disabled = false
```

为了避免这个问题，可以编写一个包含脚本式输入的包装脚本，或者在脚本式输入段落名称中使用特殊的 .path 参数。有关编写包装脚本的信息，请参阅*开发用于 Splunk Web 的视图和应用*手册中的“脚本式输入概述”。

通过直接编辑 inputs.conf 更新脚本式输入可以避免验证。

使用 .path 后缀引用外部脚本

除了编写包装脚本之外，您也可以对脚本式输入进行配置，以引用位于主机文件系统中任意位置的脚本或可执行文件。

您引用的脚本可以包含调用您所需脚本或可执行文件的单独行。您可以使用此文件来呼叫位于 Splunk Enterprise 环境之外的运行时间环境。例如，若您在同一台主机上既安装了随附 Python 的 Splunk Enterprise 又单独安装了 Python，您可以用 .path 方法来引用主机上安装的第二个 Python。

按照以下步骤使用 .path 后缀引用外部脚本：

1. 使用 Splunk Web 或编辑 inputs.conf 并用以 .path 结尾的脚本名称指定一个脚本式输入段落。例如：

   ```
   [script://myfile.path]
   disabled = 0
   ```
2. 根据“脚本式输入的脚本放在哪里”中的介绍，把您在该段落中引用的文件放入适当的目录。
3. 通过编辑此文件来指定您需要的脚本或可执行文件。例如：
   ```
   /path/to/myscript -arg1 arg -arg2 arg
   ```

含 Inputs.conf 的脚本式输入的示例

以下示例使用 input.conf 配置各种脚本式输入。

Unix top 命令

本示例介绍将 UNIX top 命令用作数据导入来源：

1. 新建一个应用程序目录。本示例使用 scripts/。

   ```
   $ mkdir $SPLUNK_HOME/etc/apps/scripts
   ```
2. 创建一个 bin/ 目录。所有脚本都必须从您的应用程序目录内的 bin/ 目录中运行。

   ```
   $ mkdir $SPLUNK_HOME/etc/apps/scripts/bin
   ```
3. 在 bin/ 目录中创建一个脚本。本示例使用一个较小的 shell 脚本 top.sh。

   ```
   $ #!/bin/sh
   top -bn 1  # linux only - different OSes have different parameters
   ```
4. 将该脚本变为可执行文件。

   ```
   chmod +x $SPLUNK_HOME/etc/apps/scripts/bin/top.sh
   ```
5. 在 shell 中运行该脚本，由此来测试该脚本是否有效。

   ```
   $SPLUNK_HOME/etc/apps/scripts/bin/top.sh
   ```
   该脚本会发送一个 top 输出。
6. 将脚本条目添加到 $SPLUNK_HOME/etc/apps/scripts/local/ 中的 input.conf。

   ```
   [script:///opt/splunk/etc/apps/scripts/bin/top.sh]
   interval = 5                    # run every 5 seconds
   sourcetype = top                # set sourcetype to top
   source = script://./bin/top.sh   # set source to name of script
   ```
7. 根据默认设置，Splunk Enterprise 会将单个 top 条目拆分成多个事件，因此您可能需要修改 props.conf 来解决此问

题。如有必要，编辑 props.conf 并将服务器配置为只在输出中不存在的内容前面断开。
例如，将以下内容添加到 $SPLUNK_HOME/etc/apps/scripts/default/props.conf 中将强制所有行成为单个事件：

```
[top]
BREAK_ONLY_BEFORE = <stuff>
```

由于 top 输出中没有时间戳，您必须指示 Splunk Enterprise 使用当前的时间。在 props.conf 中设置以下参数：

```
DATETIME_CONFIG = CURRENT
```

使用 .path 段落引用外部脚本

下面的示例使用特殊的 .path 段落设置来引用 Python 的外部构建，由此在您的主机上运行脚本。

1. 编辑 inputs.conf。

```
[script://loglogs.path]
disabled = 0
```

2. 在 $SPLUNK_HOME/etc/system/bin 中放置或创建 loglogs.path。
3. 通过编辑 loglogs.path 来引用 Python 的外部版本。

```
/usr/bin/python logit.py --source /opt/files/my_files --target /opt/files/my_files/processed --logfile
/opt/src/my_sources/logfiles
```

将 interval 属性设置为 cron 计划

在前面的示例中，您还可以通过指定字符串将 interval 属性设置为 cron 计划。

例如，以下字符串表示脚本在每小时开始时运行，一小时运行一次：

```
0 * * * *
```

以下字符串表示脚本在周一到周五，从上午 9 点到下午 5 点每 15 分钟运行一次。

```
*/15 9-17 * * 1-5
```

以下字符串表示脚本在每个偶数月份（二月、四月、六月，以此类推）的第一天，从午夜到上午 7 点以及下午 8 点到午夜，在每小时的 15 分、35 分和 55 分运行。

```
15,35,55 0-6,20-23 1 */2 *
```

更多有关设置 cron 计划的信息，请参阅 Crontab 网站上 https://crontab.org 中的 CRONTAB(5)。

使用 Journald 输入获取数据

Journald 输入是一种模块化输入，用于收集 64 位 Intel x86（x86-64）芯片架构计算机上 Linux Journald 系统日志记录过程生成的日志。Journald 进程运行于特定计算机，这些计算机要运行使用 systemd 作为系统管理服务的 Linux 操作系统版本。systemd-journald.service 将日志条目写入 Journald 数据库，并使用 **journalctl** 实用程序读取。

Splunk 平台部署会从**通用转发器**接受 journald 数据，通用转发器会捕获数据并将数据发送给 Splunk 平台部署的其他部分。

Journald 读取器的工作方式

Journald 输入会从 inputs.conf 文件中定义的段落获取其配置。每个定义的段落都包含定义 Splunk 通用转发器收集哪些数据的设置。已定义的段落的数量决定了运行 **journalctl** 进程的数量。例如，如果您定义五个段落，则 **journalctl** 实用程序会运行五次。如果配置了五个相同的已定义段落，则 **journalctl** 实用程序会向 Splunk 软件报告五次相同的数据。

您在 journald 段落中定义的每个设置对数据来说都是一个过滤器。如果未使用参数对其进行启用和配置，则它会从最早的条目开始引入日志的全部内容。要优化数据引入，请定义过滤器。Journald 配置支持数据过滤器的说明性和限制性声明。如果传递了一个或多个 FIELD=VALUE 匹配参数，则会对输出进行相应的过滤。例如，_SYSTEMD_UNIT=httpd.service 指结构化日志条目的组件。

版本和字段兼容性

journald 输入收集器使用 systemd 实用程序 journalctl。旧版本 journalctl（版本 236 和更低版本）不支持 journalctl-include-fields 配置参数。

要查看 journactl 的版本，请在命令行界面（CLI）中输入以下行：

```
journalctl --version
```

Journalctl 版本	支持的字段
236 及更高版本	journalctl-include-fields、journalctl-exclude-fields、journalctl-filter、journalctl-grep、 journalctl-user-unit
236 及更低版本	journalctl-exclude-fields、journalctl-filter、journalctl-grep、 journalctl-user-unit

配置 Journald 输入

在 Splunk Enterprise x86_64 芯片组 Linux 通用转发器上配置 Journald 输入，以将 Journald 日志引入到 Splunk 平台部署中。使用 .conf 文件进行输入配置。

1. 在您的 Splunk 通用转发器上，导航到 splunkforwarder/etc/apps/journald_input/default/。
2. 复制 inputs.conf 文件。
3. 导航到 splunkforwarder/etc/apps/journald_input/local/。
4. 粘贴 inputs.conf 文件的副本。
5. 使用文本编辑器打开 inputs.conf 文件。
6. 导航到 journald 段落，并使用已包含或排除所需数据字段的参数来配置此段落。请参阅以下示例：
   ```
   [journald://my-stanza]
   # The following fields are included by default, and required: MESSAGE, _REALTIME_TIMESTAMP, _CURSOR
   journalctl-include-fields = PRIORITY,CMD,EXE
   journalctl-exclude-fields =
   journalctl-filter =_SYSTEMD_UNIT=my.service,_PID=232+_SYSTEMD_UNIT=sshd
   journalctl-grep =^WARN.*disk,.*errno=\d+\S+restarting
   journalctl-user-unit =unit1,unit2
   ```
   有关参数的完整列表，请参阅本主题中的**参考**部分。
7. 确认您的字段配置是以下划线开头。Journald 输入段落配置要求使用下划线
8. 保存更改。
9. 重新启动 Splunk 通用转发器。
10. （可选）使用部署服务器将设置更改推送到 Splunk 平台部署中的其他转发器。有关更多信息，请参阅*《更新 Splunk Enterprise 实例》*手册中的“使用转发器管理管理应用”主题。

参考

有关可用参数的完整列表，请参阅下表：

参数	描述
journalctl-include-fields = <string>	过滤要发送到 Splunk 平台实例的字段。
journalctl-exclude-fields = <string>	过滤哪些字段不发送给 Splunk 平台实例。
journalctl-filter = <string>	使用 journalctl 中的**匹配**概念的过滤器。例如，_SYSTEMD_UNIT=avahi-daemon.service _PID=28097 + _SYSTEMD_UNIT=dbus.service 将显示 PID 为 28097 的来自 Avahi 服务进程的所有消息，以及来自 D-Bus 服务的所有消息。
journalctl-unit = <string>	显示指定 systemd 单元的消息。
journalctl-identifier = <string>	显示指定 syslog 标识符 SYSLOG_IDENTIFIER 的消息。
journalctl-priority =	按消息优先级或优先级范围过滤输出。

<string>	
journalctl-boot = <string>	来自特定引导的消息。
journalctl-facility = <string>	通过 syslog 工具过滤输出。
journalctl-grep = <string>	将输出过滤到 MESSAGE= 字段与指定的正则表达式匹配的条目。
journalctl-user-unit = <string>	显示指定用户会话单元的消息。
journalctl-dmesg = <string>	仅显示内核消息。
journalctl-quiet = <string>	抑制所有信息类消息。
journalctl-freetext = <string>	保留以备将来使用。

对 Journald 输入进行故障排除

Journald 输入的最佳做法

- 检查您的 splunkd.log 文件中的错误。
- 从简单的配置开始，然后再构建更复杂的东西。
- 有关配置的更多信息，请参阅此产品的规格文件。

验证数据

要验证您的 Journald 输入在收集什么，请在命令行界面（CLI）中输入以下内容：

ps aux | grep journalctl。

您需要在 inputs.conf 中定义至少一个段落才能收集数据。

数据重复

配置 Journald 时，每个段落都会运行一个日志读取器。重复的段落定义将导致数据重复。

数据收集效率

使用 include-fields 比使用 exclude-fields 更有效，因为 include-fields 的过滤发生在 Journald 级别，而不是 Splunk 通用转发器级别。

字段不可用

查看您的 journalctl 版本。较早的版本不包含 include-fields 和 grep。在部署服务器方案中，有些计算机使用较新的 journalctl，而有些计算机使用较旧的版本。Splunk 部署服务器会将相同的配置推送给新旧计算机，但旧计算机无法收集数据。

针对字段选项缺少进行故障排除

Journald 输入使用外部开发的实用程序才能发挥作用。因此，较旧版本的 journalctl 可能无法使用所有的配置选项。CentOS 7 过滤配置中可能缺少以下选项：

您可以借助 journalct-include-fields 选项声明要检索的 Journald 字段。如果没有此选项，您可能只能指定不想包含的字段。例如，如果您只想检索 MESSAGE 字段，则必须列出您不想 Splunk 平台实例引入的每个字段。此选项可能会生成多个冗长的数据字段。

使用 grep 选项，您可以在数据到达索引器之前应用正则表达式过滤器。

Journald 日志在 Splunk 通用转发器上首次启动时无法打开游标文件

在通用转发器上首次启动 Journald 模块化输入时，如果您收到以下 Journald 日志错误：

```
09-04-2020 03:50:49.221 +0000 ERROR ExecProcessor - message from "/home/ec2-user/splunkforwarder/bin/splunkd journald-modinput
'$@'" journald-modinput - unable to open cursor file /home/ec2-
user/splunkforwarder/var/lib/splunk/modinputs/journald/s1.checkpoint for reading (No such file or directory)
09-04-2020 03:50:49.221 +0000 INFO  ExecProcessor - message from "/home/ec2-user/splunkforwarder/bin/splunkd journald-modinput
'$@'"  journald-modinput -  starting (mode = Data Collection) Config -  serverHost=ip-172-31-49-163.us-west-2.compute.internal
serverUri=https://127.0.0.1:8089 checkpointDir=/home/ec2-user/splunkforwarder/var/lib/splunk/modinputs/journald
stanza=journald://s1 PropertiesMap: {host -> '$decideOnStartup' index -> 'default' interval -> '30' journalctl-exclude-fields -
> '__MONOTONIC_TIMESTAMP,__SOURCE_REALTIME_TIMESTAMP' journalctl-include-fields ->
'PRIORITY,_SYSTEMD_UNIT,_SYSTEMD_CGROUP,_TRANSPORT,_PID,_UID,_MACHINE_ID,_GID,_COMM,_EXE' journalctl-quiet -> 'true'}
```

游标文件（Splunk 软件用此文件来跟踪从日志中引入的数据）可能第一次无法使用。它不会影响 Journald 输入功能。数据引入正常。它会生成此错误日志响应。

Journald ID 缺少检查点文件错误

如果您首先运行 journald 输入，将收到以下错误：

```
10-03-2020 18:29:35.241 -0600 ERROR ExecProcessor - message from "/opt/splunkforwarder/bin/splunkd journald-modinput '$@'"
journald-modinput - unable to open cursor file /opt/splunkforwarder/var/lib/splunk/modinputs/journald/all.checkpoint for
reading (No such file or directory)
```

没有其他要执行的操作。这是一个检查点文件，其中包含 Splunk 软件已读取的最后一条记录的 journald ID。如果您需要重新开始读取，Splunk 软件会在此错误中的 ID 后重新启动。

journald 读取器启动时，首先会检查 Splunk 软件是否需要在上次停止后继续读取。该进程第一次运行时，没有创建检查点文件，因此读取器将无法找到检查点文件，并将错误写入日志。

配置事件处理

事件处理概述

Splunk 平台为事件建立索引，事件是驻留在计算机数据中的活动记录。事件提供生成机器数据的系统的相关信息。术语**事件数据**指 Splunk 平台索引的内容。

下面是一个示例事件：

```
172.26.34.223 - - [01/Jul/2017:12:05:27 -0700] "GET /trade/app?action=logout HTTP/1.1" 200 2953
```

在为事件建立索引时，Splunk 软件会执行以下任务：

任务	链接
配置字符集编码	配置字符集编码
配置多行事件的换行	配置事件换行
标识事件时间戳，并在无时间戳的情况下将时间戳应用于事件	配置事件时间戳
提取一组有用的标准字段，如 host、source 和 sourcetype	关于默认字段
将事件分段	关于事件分段
将元数据动态分配给事件（若已指定）	动态分配默认字段
使数据匿名（若已指定）	使数据匿名

有关索引处理的概述，请参阅*《管理索引器和索引器群集》*手册中的“索引概述”一章。

配置字符集编码

您可为您的数据源配置**字符集编码**。Splunk 软件具有内置的字符集规范，为**部署**的国际化提供支持。Splunk 软件支持多种语言，包括某些不使用 8 位通用编码字符集转换格式 (UTF-8) 编码的语言。

默认情况下，Splunk 软件会尝试将 UTF-8 编码应用于您的数据来源。如果某数据来源未使用 UTF-8 编码或为非 ASCII 文件，Splunk 软件会尝试将该来源中的数据转换为 UTF-8 编码，除非您在 props.conf 文件中设置 CHARSET 键时指定了要使用的字符集。

您可以通过在大多数 *nix 系统上使用 `iconv -l` 命令来检索有效字符编码规范的列表。Windows 上的 `iconv` 端口可用。

支持的字符集

Splunk 软件支持的字符集极为广泛，包括以下关键字符集：

- UTF-8
- UTF-16LE
- Latin-1
- BIG5
- SHIFT-JIS `

下表显示了常见支持的字符集及其对应的语言的简短列表。

语言	代码
阿拉伯语	CP1256
阿拉伯语	ISO-8859-6
亚美尼亚语	ARMSCII-8
白俄罗斯语	CP1251
保加利亚语	ISO-8859-5

捷克斯洛伐克语	ISO-8859-2
格鲁吉亚语	Georgian-Academy
希腊语	ISO-8859-7
希伯来语	ISO-8859-8
日语	EUC-JP
日语	SHIFT-JIS
韩语	EUC-KR
俄语	CP1251
俄语	ISO-8859-5
俄语	KOI8-R
斯洛伐克语	CP1250
斯洛文尼亚语	ISO-8859-2
泰语	TIS-620
乌克兰语	KOI8-U
越南语	VISCII

有关更多受支持的字符集，请参阅本主题后面的“支持字符集的综合列表”部分。

手动指定字符集

要手动指定字符集，您需要编辑 props.conf 文件。如果您使用的是 Splunk Cloud 部署，请在转发器上编辑 props.conf 文件。如果您使用的是 Splunk Enterprise 部署，则可以在 Splunk Enterprise 部署中编辑此文件。

要手动指定一个应用到输入上的字符集，请在 props.conf 文件中设置 CHARSET 键：

```
[spec]
CHARSET=<string>
```

例如，若您的主机生成的数据为希腊文且该主机使用 ISO-8859-7 编码，请在 props.conf 文件中为该主机设置 CHARSET=ISO-8859-7。在本示例中，该主机称为“GreekSource”：

```
[host::GreekSource]
CHARSET=ISO-8859-7
```

Splunk 软件仅分析具有 UTF-8 映射的字符编码。

自动指定字符集

Splunk 软件可以使用其字符集编码算法自动检测语言和适当的字符集。

如需通过配置 Splunk 软件使其自动为特定输入检测适当的语言和字符集编码，请在 props.conf 文件中为该输入设置 CHARSET=AUTO。如果您使用的是 Splunk Cloud 部署，则可以在转发器上编辑此文件。如果您使用的是 Splunk Enterprise 部署，则可以在 Splunk Enterprise 部署中编辑此文件。

例如，若想为主机“my-foreign-docs”自动检测字符集编码，请在 props.conf 文件中为该主机设置 CHARSET=AUTO：

```
[host::my-foreign-docs]
CHARSET=AUTO
```

使 Splunk Enterprise 识别字符集

如果您使用的是 Splunk Cloud 而且想把一个字符集编码添加到您的 Splunk 部署，请提交 Splunk 支持工单。如果您使用的是 Splunk Enterprise 部署，则可以训练 Splunk 软件识别字符集。

您可以通过将示例文件添加到以下路径并重新启动 Splunk Enterprise 来训练 Splunk Enterprise 识别字符集：

```
$SPLUNK_HOME/etc/ngram-models/_<language>-<encoding>.txt
```

例如，如果您要使用 "vulcan-ISO-12345" 字符集，请将规范文件复制到以下路径：

```
/SPLUNK_HOME/etc/ngram-models/_vulcan-ISO-12345.txt
```

将示例文件添加到指定路径后，Splunk 软件即可识别使用了新字符集的数据来源，并自动在索引时将这些来源转换成 UTF-8 格式。

支持字符集的综合列表

前面“支持的字符集”部分中所述的常见字符集是 CHARSET 属性可支持的一个小子集。Splunk 软件也支持字符集的长列表。在 Splunk 平台支持的字符集中，它还支持其别名，该列表与 *nix iconv 实用工具支持的列表一样。

Splunk 软件在匹配 CHARSET 时忽略标点符号和大小写。例如，utf-8、UTF-8 和 utf8 都视为相同。

以下列表显示了所有支持的字符集，括号中是别名：

- utf-8（CESU-8、ANSI_X3.4-1968、ANSI_X3.4-1986、ASCII、CP367、IBM367、ISO-IR-6、ISO646-US ISO_646.IRV:1991、US、US-ASCII、CSASCII）
- utf-16le（UCS-2LE、UNICODELITTLE）
- utf-16be（ISO-10646-UCS-2、UCS-2、CSUNICODE、UCS-2BE、UNICODE-1-1、UNICODEBIG、CSUNICODE11、UTF-16）
- utf-32le（UCS-4LE）
- utf-32be（ISO-10646-UCS-4、UCS-4、CSUCS4、UCS-4BE、UTF-32）
- utf-7（UNICODE-1-1-UTF-7、CSUNICODE11UTF7）
- c99（java）
- utf-ebcdic
- latin-1（CP819、IBM819、ISO-8859-1、ISO-IR-100、ISO_8859-1:1987、L1、CSISOLATIN1）
- latin-2（ISO-8859-2、ISO-IR-101、ISO_8859-2:1987、L2、CSISOLATIN2）
- latin-3（ISO-8859-3、ISO-IR-109、ISO_8859-3:1988、L3、CSISOLATIN3）
- latin-4（ISO-8859-4、ISO-IR-110、ISO_8859-4:1988、L4、CSISOLATIN4）
- latin-5（ISO-8859-9、ISO-IR-148、ISO_8859-9:1989、L5、CSISOLATIN5）
- latin-6（ISO-8859-10、ISO-IR-157、ISO_8859-10:1992、L6、CSISOLATIN6）
- latin-7（ISO-8859-13、ISO-IR-179、L7）
- latin-8（ISO-8859-14、ISO-CELTIC、ISO-IR-199、ISO_8859-14:1998、L8）
- latin-9（ISO-8859-15、ISO-IR-203、ISO_8859-15:1998）
- latin-10（ISO-8859-16、ISO-IR-226、ISO_8859-16:2001、L10、LATIN10）
- ISO-8859-5（CYRILLIC、ISO-IR-144、ISO_8859-5:1988、CSISOLATINCYRILLIC）
- ISO-8859-6（ARABIC、ASMO-708、ECMA-114、ISO-IR-127、ISO_8859-6:1987、CSISOLATINARABIC、MACARABIC）
- ISO-8859-7（ECMA-118、ELOT_928、GREEK、GREEK8、ISO-IR-126、ISO_8859-7:1987、ISO_8859-7:2003、CSISOLATINGREEK）
- ISO-8859-8（HEBREW、ISO-8859-8、ISO-IR-138、ISO8859-8、ISO_8859-8:1988、CSISOLATINHEBREW）
- ISO-8859-11
- roman-8（HP-ROMAN8、R8、CSHPROMAN8）
- KOI8-R（CSKOI8R）
- KOI8-U
- KOI8-T
- GEORGIAN-ACADEMY
- GEORGIAN-PS
- ARMSCII-8
- MACINTOSH（MAC、MACROMAN、CSMACINTOSH）

  这些 MAC* 字符集适用于 MacOS 9。macOS X 等更高版本，使用 unicode。

- MACGREEK
- MACCYRILLIC
- MACUKRAINE
- MACCENTRALEUROPE
- MACTURKISH
- MACCROATIAN
- MACICELAND
- MACROMANIA
- MACHEBREW
- MACTHAI
- NEXTSTEP

- CP850（850、IBM850、CSPC850MULTILINGUAL）
- CP862（862、IBM862、CSPC862LATINHEBREW）
- CP866（866、IBM866、CSIBM866）
- CP874（WINDOWS-874）
- CP932
- CP936（MS936、WINDOWS-936）
- CP949（UHC）
- CP950
- CP1250（MS-EE、WINDOWS-1250）
- CP1251（MS-CYRL、WINDOWS-1251）
- CP1252（MS-ANSI、WINDOWS-1252）
- CP1253（MS-GREEK、WINDOWS-1253）
- CP1254（MS-TURK、WINDOWS-1254）
- CP1255（MS-HEBR、WINDOWS-1255）
- CP1256（MS-ARAB、WINDOWS-1256）
- CP1257（WINBALTRIM、WINDOWS-1257）
- CP1258（WINDOWS-1258）
- CP1361（JOHAB）
- BIG-5（BIG-FIVE、CN-BIG5、CSBIG5）
- BIG5-HKSCS(BIG5-HKSCS:2001)
- CN-GB（EUC-CN、EUCCN、GB2312、CSGB2312）
- EUC-JP（EXTENDED_UNIX_CODE_PACKED_FORMAT_FOR_JAPANESE、CSEUCPKDFMTJAPANESE）
- EUC-KR（CSEUCKR）
- EUC-TW（CSEUCTW）
- GB18030
- GBK
- GB_1988-80（ISO-IR-57、ISO646-CN、CSISO57GB1988、CN）
- HZ（HZ-GB-2312）
- GB_2312-80（CHINESE、ISO-IR-58、CSISO58GB231280）
- SHIFT-JIS（MS_KANJI、SJIS、CSSHIFTJIS）
- ISO-IR-87（JIS0208 JIS_C6226-1983、JIS_X0208 JIS_X0208-1983、JIS_X0208-1990、X0208、CSISO87JISX0208、ISO-IR-159、JIS_X0212、JIS_X0212-1990、JIS_X0212.1990-0、X0212、CSISO159JISX02121990）
- ISO-IR-14（ISO646-JP、JIS_C6220-1969-RO、JP、CSISO14JISC6220RO）
- JISX0201-1976（JIS_X0201、X0201、CSHALFWIDTHKATAKANA）
- ISO-IR-149（KOREAN、KSC_5601、KS_C_5601-1987、KS_C_5601-1989、CSKSC56011987）
- VISCII（VISCII1.1-1、CSVISCII）
- ISO-IR-166（TIS-620、TIS620-0、TIS620.2529-1、TIS620.2533-0、TIS620.2533-1）
- UCS-2-INTERNAL、UCS-2-SWAPPED、UCS-4-INTERNAL、UCS-4-SWAPPED

配置事件换行

一些事件由多行组成。默认情况下，Splunk 平台能正确处理大多数的多行事件。如果有 Splunk 平台未正确处理的多行事件，您可以对平台进行配置，更改其换行行为。

如果您使用的是 Splunk Cloud，则必须在需要配置事件换行的地方转发任何数据，因为在 Splunk Web 界面中无法配置事件换行。您必须使用已配置为将数据发送到 Splunk Cloud 实例的重型转发器，以将传入数据分成多行，然后根据需要将它们合并为事件。

如果您使用的是 Splunk Enterprise，则可以在对传入数据流建立索引的任何实例上配置设置并遵循本主题中的过程。

Splunk 平台如何确定事件界限

Splunk 平台用两个阶段来确定事件界限：

1. 换行，即使用 LINE_BREAKER 设置将传入的数据流拆分成多个单独的行。默认情况下，LINE_BREAKER 值是新行和回车的任意序列。在正则表达式格式中，这表示为以下字符串：`([\r\n]+)`。您通常不需要调整此设置，但在必要时，您必须在将数据发送到 Splunk Cloud 或 Splunk Enterprise 索引器的转发器上的 props.conf 配置文件中进行配置。LINE_BREAKER 设置需要正则表达式格式的值。
2. 行合并，即使用 SHOULD_LINEMERGE 设置将之前分隔的行合并成事件。默认情况下，Splunk 平台会执行行合并，SHOULD_LINEMERGE 的值为 true。您通常不需要调整此设置，但在必要时，您必须在将数据发送到 Splunk Cloud 的转发器上的 props.conf 配置文件中配置此设置。如果您通过将 SHOULD_LINEMERGE 属性设置为 false 将 Splunk 平台配置为不执行行合并，那么平台会根据 LINE_BREAKER 设置确定的内容将传入数据拆分为行。

对于 Splunk 平台，换行相对有效，而行合并则相对较慢。使用 LINE_BREAKER 设置可以生成您想要在换行阶段取得的结果。如果您的大量数据均包含多行事件，这将有用。

还有其他配置设置可帮助您将传入数据流拆分成事件，例如换行。

如何配置事件界限

许多事件日志都采用严格的一个事件一行的格式，但有些也不采用此格式。Splunk 平台通常可以识别事件边界，但是如果事件边界识别没有发生或发生错误，您可以在 props.conf 配置文件中的设置自定义规则以建立事件边界。

配置事件界限的要求

在尝试为事件配置事件边界之前，请确认已具备以下条件：

- 了解正则表达式。LINE_BREAKER 设置使用正则表达式确定事件的边界。
- 以下其中之一，具体取决于您使用的是 Splunk Cloud 还是 Splunk Enterprise：
  - 已配置为将数据发送到 Splunk Cloud 实例的重型转发器。您可以下载 Splunk Cloud 实例随附的 Splunk Cloud 通用转发器凭据包，并将其安装在 Splunk 重型转发器上。
  - Splunk Enterprise 索引器或重型转发器（如果您使用的是 Splunk Enterprise）。
- 表示要在其中配置自定义换行的数据流的文件。

编辑 props.conf 配置文件以配置多行事件

1. 检查要建立索引的文件以确定其事件格式。
2. 在文件中，查找事件中要设置为事件开头或结尾的模式。
3. 使用文本编辑器，在您配置为将数据发送到 Splunk Cloud 的转发器上，编辑 $SPLUNK_HOME/etc/system/local/props.conf 配置文件。
4. 在 props.conf 配置文件中，添加必要的换行和行合并设置，以配置转发器对传入的数据流执行正确的换行。
5. 保存文件并将其关闭。
6. 重新启动转发器来提交更改。

可使用两种方法来处理多行事件：

- 拆分数据流并将其重组成事件。
- 使用 LINE_BREAKER 设置将数据流直接拆分为实际事件。

拆分数据流并将其重组成事件

此方法通常可以简化配置过程，因为它为您授予对可用于定义行合并规则的若干设置的访问权限。

您必须在指定用于将数据发送到 Splunk Cloud 实例的重型转发器上执行这些步骤。

1. 在将数据发送到 Splunk Cloud 实例的转发器上，使用文本编辑器打开 $SPLUNK_HOME/etc/system/local/props.conf 进行编辑。
2. 在此文件中，请在 props.conf 配置文件中指定一个段落，它代表了您想要拆分并将其重组成事件的数据流。
3. 在此段落中使用将数据流拆分成多行的正则表达式配置 LINE_BREAKER 设置。
4. 将 SHOULD_LINEMERGE 设置添加到该段落，并将其值设置为 true。
5. 配置其他行合并设置，例如 BREAK_ONLY_BEFORE 和其他设置，即可指定转发器如何将各行重组成事件。有关行合并设置的更多信息，请参阅本主题后面的“仅当 SHOULD_LINEMERGE 设置为 true 时应用的属性”。

如果您的数据符合默认的 LINE_BREAKER 值（任意数量的新行和回车），则无需更改 LINE_BREAKER 设置。如果不符合，设置 SHOULD_LINEMERGE=true 并使用行合并设置重组数据。

使用 LINE_BREAKER 设置将数据流直接拆分为实际事件

使用 LINE_BREAKER 设置定义事件边界可能会增加您的索引速度，但是使用可能会更难。如果您发现索引缓慢且您的大量数据均包含多行事件，则此方法有显著的改善作用。

1. 请在 props.conf 中指定一个段落，它代表了您想要直接拆分为事件的数据流。
2. 在此段落下，使用正则表达式配置 LINE_BREAKER 设置，表达式应和您想要用来将原始数据流拆分成事件的边界匹配。
3. 添加 SHOULD_LINEMERGE 设置，并将其配置为 false。

换行常规设置

下表列出 props.conf 文件中影响换行的设置。

属性	描述	默认
TRUNCATE = <non-negative integer>	更改默认的最大行长度（以字节为单位）。尽管此设置为字节度量，但当此属性以其他方式获得多字节字符的中间字符时，Splunk 平台仍会向下舍入行长度。	10,000

	如果希望永不截断，请设置为 0。然而，非常长的行通常表示垃圾数据。	
LINE_BREAKER = <regular expression>	正则表达式会在任何行合并发生前确定 Splunk 平台如何将原始文本流拆分为初始事件。此设置取决于 SHOULD_LINEMERGE 设置，稍后将进行介绍。 正则表达式必须包含捕获组，用于定义已标识的匹配子组件的一对括号。 无论表达式匹配到的位置在哪里，Splunk 平台会将第一个捕获组的开头视为前一个事件的结尾，而将第一个捕获组的结尾视为下一个事件的开头。 该平台会丢弃第一个捕获组的内容。该内容不会出现于任何事件中，因为平台认为此文本位于两行之间。 使用 LINE_BREAKER 设置分隔多行事件时，处理速度将显著提升（与使用 SHOULD_LINEMERGE 将单个行重组成多行事件相反）。如果您的大部分数据均包含多行事件，请考虑使用此方法。 请参阅 props.conf 规范文件了解如何配合使用 LINE_BREAKER 和分支表达式的信息及其他信息。	([\r\n]+) Splunk 平台会将每行的数据拆分成一个事件，并由任意数量的回车（\r）或新行（\n）字符分隔。
LINE_BREAKER_LOOKBEHIND = <integer>	前一个原始数据块中有剩余数据时，LINE_BREAKER_LOOKBEHIND 会指示 Splunk 平台应用了 LINE_BREAKER 正则表达式的原始数据块（与下一个数据块相连）结尾之前的字符数。如果您处理极大或多行事件，则可能需要增加此值，而非采用默认值。	100
SHOULD_LINEMERGE = [true|false]	设置为 true 时，Splunk 平台会将多个输入行组合成单个事件，配置则以下节中介绍的设置为基础。	true

仅当 SHOULD_LINEMERGE 设置为 true 时应用的属性

将 SHOULD_LINEMERGE 设置为默认值 true 时，请使用下列额外的设置定义换行行为。

属性	描述	默认
BREAK_ONLY_BEFORE_DATE = [true|false]	设置为 true 时，Splunk 平台会在遇到含日期的新行时新建的事件。	true 如果您将 DATETIME_CONFIG 设置配置为 CURRENT 或 NONE，此属性没有意义，因为在这些情况下，Splunk 平台无法识别时间戳。
BREAK_ONLY_BEFORE = <regular expression>	设置之后，Splunk 平台会在遇到与正则表达式匹配的新行时新建事件。	空字符串
MUST_BREAK_AFTER = <regular expression>	设置之后，只要正则表达式与当前的行匹配，Splunk 平台就会为下一个输入行新建事件。如果匹配到另一个规则，平台仍可能会在当前行之前中断。	空字符串
MUST_NOT_BREAK_AFTER = <regular expression>	设置之后，只要正则表达式与当前行匹配，Splunk 平台就不会在匹配到 MUST_BREAK_AFTER 表达式之前拆分当前行之后的各行。	空字符串
MUST_NOT_BREAK_BEFORE = <regular expression>	设置之后，只要正则表达式与当前行匹配，Splunk 平台就不会拆分当前行之前的最后一个事件。	空字符串
MAX_EVENTS = <integer>	指定 Splunk 平台添加到任意事件的输入行的最大数。读取指定行数后，软件进行事件换行。	256 行

配置事件换行的示例

指定事件分隔符

以下示例将 Splunk 平台配置为将仅由数字组成的任何行标识为其来源类型设置为 my_custom_sourcetype 的任何数据的新事件的开始。

```
[my_custom_sourcetype]
BREAK_ONLY_BEFORE = ^\d+\s*$
```

将多行合并为单个事件

下列日志事件包含作为同一请求一部分的多个行。请求之前的区分符是 "Path"。

```
{{"2006-09-21, 02:57:11.58", 122, 11, "Path=/LoginUser
Query=CrmId=ClientABC&ContentItemId=TotalAccess&SessionId=3A1785URH117BEA&Ticket=646A1DA4STF896EE&
amp;SessionTime=25368&ReturnUrl=http://www.clientabc.com, Method=GET, IP=209.51.249.195, Content=", ""}}


{{"2006-09-21, 02:57:11.60", 122, 15, "UserData:<User CrmId="clientabc"
UserId="p12345678"><EntitlementList></EntitlementList></User>", ""}}
{{"2006-09-21, 02:57:11.60", 122, 15, "New Cookie:
SessionId=3A1785URH117BEA&Ticket=646A1DA4STF896EE&CrmId=clientabc&UserId=p12345678&AccountId=& amp;AgentHost=man&AgentId=man,
MANUser: Version=1&Name=&Debit=&Credit=&AccessTime=&BillDay=&Status=&Language=&Country=
&Email=&EmailNotify=&Pin=&PinPayment=&PinAmount=&PinPG=&PinPGRate=&PinMenu=&", ""}}
```

若想为该多行事件正确地新建索引，请在您的配置中使用 `Path` 区分符。将以下内容添加到您的 `$SPLUNK_HOME/etc/system/local/props.conf` 文件中。

```
[source::source-to-break]
SHOULD_LINEMERGE = True
BREAK_ONLY_BEFORE = Path=
```

此代码将 Splunk 平台配置为合并该事件的各行，且仅在术语 `Path=` 之前中断。

多行事件换行和分段限制

Splunk 平台会对巨大事件应用换行和分段限制：

限制	描述
MAX_EVENTS 行中的事件	如果平台中的多行事件超出 MAX_EVENTS 中指定的行数，在该限制下进行事件换行，将 BREAK_ONLY_BEFORE_DATE 设置为 false（如果为 true），然后降低任一 MUST_NOT_BREAK_BEFORE 或 MUST_NOT_BREAK_AFTER 规则。这会导致事件不会像您期望的那样换行。要解决问题，您可以提高 MAX_EVENTS 设置，但是您可以通过将 SHOULD_LINEMERGE 设置更改为 false，并使用 LINE_BREAKER 设置指定事件边界获得更好的结果。
长度超过 10,000 个字节的行。	Splunk 平台使用 LINE_BREAKER 和 TRUNCATE 设置将 10KB 以上的事件评估并拆分为多行，每行 10KB。它会将 meta::truncated 字段附加到所有截断行的末尾处。如果您还将 SHOULD_LINEMERGE 配置为 true，则平台使用 props.conf 规则评估所有其他事件数据，直到创建完整事件为止。
超过 100,000 个字节的事件的分段	Splunk Web 会在搜索结果中显示一个事件的前 100,000 个字节。不过，极长行的前 100,000 个字节之后的段仍可搜索。
超过 1,000 个段的事件的分段	Splunk Web 会在搜索结果中显示一个事件的前 1,000 个段，该事件的各段由空格分隔开而且会在鼠标悬停在其上时会突出显示。它将事件的剩余部分显示为无交互格式的原始文本。

配置事件时间戳

您可以控制 Splunk 平台如何处理 Splunk Cloud 和 Splunk Enterprise 中的时间戳。

检查以下事件示例：

```
172.26.34.223 - - [01/Jul/2017:12:05:27 -0700] "GET /trade/app?action=logout HTTP/1.1" 200 2953
```

事件中的时间信息 `[01/Jul/2017:12:05:27 -0700]` 是 **时间戳**。

Splunk 平台使用时间戳按时间关联事件、在 Splunk Web 中新建直方图，并为搜索设置时间范围。大部分事件都包含时间戳；如果事件不包含时间戳信息，Splunk 平台会尝试在索引时间为事件分配时间戳值。

大多数情况下，Splunk 平台都能正确地提取时间戳，但在某些情况下，您可能需要配置时间戳处理。例如，处理分布式部署时，您可能需要重新配置时间戳识别和格式。

有关如何配置时间戳的具体说明，请参阅本册手中的“配置时间戳”一章。

配置索引字段提取

Splunk 软件可在索引时间提取各种字段。您可以配置和修改软件执行此字段提取的方式。

Splunk 软件在索引时间可以提取下列字段：

- 默认字段
- 自定义字段
- 文件标头字段

Splunk 软件始终为每个事件提取一组默认字段。您可以将其配置为提取自定义字段，对于某些数据，还可以配置为提取文件标头字段。

有关索引字段提取的更多信息，请参阅“配置索引字段提取”一章。

使数据匿名

您可能需要将您索引至 Splunk Cloud 的数据中的敏感个人信息（例如信用卡或社保号）匿名或以掩码显示。您可以使事件中的部分机密字段匿名，以保护隐私，同时提供足够的剩余数据以跟踪事件。

要使用 Splunk Cloud 使数据匿名，您必须将 Splunk Enterprise 实例配置为重型转发器，并在将数据发送至 Splunk Cloud 之前使用该实例使传入数据匿名。

有两种可以使用重型转发器使数据匿名的方式：

- **使用 SEDCMD 设置。**此设置存在于您在重型转发器上配置的 props.conf 配置文件中。像使用 sed *nix 脚本一样完成替换。此方法更简单，配置所需的时间更少，而且比正则表达式转换的速度更快。但是，可以调用 SEDCMD 设置的次数及其可用的功能是有限的。有关此方法的说明，请参阅“使用 sed 脚本使数据匿名”。
- **使用正则表达式（regex）转换。**此方法配置时间较长，但是初次配置之后易于修改。还可以更灵活地将此方法分配给多个数据导入。有关此方法的说明，请参阅“使用正则表达式转换使数据匿名”。

上述两个选项在 Splunk Enterprise 中也可用，可在其中完成重型转发器或索引器的配置。

使数据匿名的前提条件

您必须选择一组要匿名的事件，才能使数据匿名。

- 首先，选择要匿名的事件
- 然后，您可以：
  - 使用 props.conf 配置文件通过 sed 脚本使事件匿名
  - 使用 props.conf 和 transforms.conf 配置文件通过正则表达式转换使事件匿名

选择要匿名的事件

您可以根据数据是来自特定来源还是主机，或者数据是否带有特定来源类型标记来使事件数据匿名。您必须在 props.conf 配置文件中指定选择数据的方法。您在 props.conf 文件中指定的段落名称将决定 Splunk 平台会如何选择和处理事件进行匿名。

请参阅以下段落规范：

- [host::<host>] 段落匹配包含指定主机的事件
- [source::<source>] 段落匹配带有特定来源的事件
- [<sourcetype>] 段落匹配带有特定来源类型的事件
  - 作为最佳实践，您必须随后在 inputs.conf 文件中指定来源类型才能使此段落类型生效。此选项是一个 Splunk 最佳实践。

用 sed 脚本替换事件中的字符串

您可以使用 sed 脚本和 SEDCMD 方法替换字符串或替代字符。有关 sed 样式替换，请参阅以下语法：

```
SEDCMD-<class> = s/<regex>/<replacement>/flags
```

SEDCMD 设置具有以下组件：

- regex 是用 Perl 编程语言编写的正则表达式。它代表您要替换的内容。
- replacement 是要替换任何正则表达式匹配的字符串。
- flags 可以是字母 g 以替换所有匹配，也可以是数字以替换指定匹配。

用 sed 脚本替换事件中的字符

有关 sed 字符替换，请参阅以下语法：

```
SEDCMD-<class> = y/<string1>/<string2>/
```

该语法会把 string1 中出现的所有字符替代为 string2 中的字符。

使用具有 transforms.conf 的正则表达式转换使数据匿名

transforms.conf 配置文件中的每个段落会定义转换类别，您可为指定的来源类型、来源或主机从 props.conf 文件中引用。

转换具有若干设置和变量，允许您指定变更内容和位置，但是以下设置最为重要：

- REGEX 设置会指定将指向您想要匿名的事件中的字符串的正则表达式
- FORMAT 设置会指定掩码值
- $1 变量代表正则表达式前面的事件文本，而该正则表达式表示您想要进行掩码的事件中的字符串
- $2 变量代表正则表达式后面的事件文本
- DEST_KEY = _raw 将 FORMAT 中的值写入日志中的原始值。此操作将使事件匿名。

该正则表达式处理器不处理多行事件。通过将字符串 (?m) 放在 transforms.conf 文件中的正则表达式之前，即可将该事件指定为多行事件，这样事件就可以跨多个行。

使用 sed 脚本使数据匿名

您可以使用 sed 脚本替换或替代事件中的字符串，来实现数据匿名。

Sed 是 *nix 实用工具，可根据您使用的命令或您为实用工具提供的参数读取文件和修改输入。很多 *nix 用户使用实用工具快速转换传入数据，因为实用工具的多功能性。您可以在 props.conf 文件中使用类似 sed 的语法以通过脚本将 Splunk 平台中的数据掩码。

以下是如何掩码文件的示例。

假设您有一个包含社会保险号和信用卡号的名为 accounts.log 的日志文件：

```
...
ss=123456789, cc=1234-5678-9012-3456
ss=123456790, cc=2234-5678-9012-3457
ss=123456791, cc=3234-5678-9012-3458
ss=123456792, cc=4234-5678-9012-3459
...
```

您想要以掩码显示字段，这样该字段为：

```
...
ss=XXXXX6789, cc=XXXX-XXXX-XXXX-3456
ss=XXXXX6790, cc=XXXX-XXXX-XXXX-3457
ss=XXXXX6791, cc=XXXX-XXXX-XXXX-3458
ss=XXXXX6792, cc=XXXX-XXXX-XXXX-3459
...
```

您可以使用 inputs.conf 和 props.conf 配置文件更改 Splunk 平台访问时来自 accounts.log 文件的数据。这些配置文件位于 Splunk Cloud 重型转发器或 Splunk Enterprise 索引器上的 $SPLUNK_HOME/etc/system/local/ 目录中。

使用 sed 脚本使数据匿名的前提条件

您必须满足以下要求才能使用 sed 脚本使数据匿名：

- 拥有您想要匿名的数据
- 了解正则表达式如何工作
- 具有指向您要匿名的数据所在位置的 inputs.conf 配置文件
- 具有引用使数据匿名的 sed 脚本的 props.conf 配置文件

配置 inputs.conf 文件以使用 sed 脚本

在此示例中，您可以新建来源类型 SSN-CC-Anon，并将其分配给 accounts.log 文件的数据导入。您新建的转换会用此来源类型了解要转换哪些数据。虽然有其他选项可使用 SEDCMD 转换日志文件中的传入数据，但是最佳实践是新建来源类型，然后为 props.conf 文件中的来源类型分配转换。

1. 在运行重型转发器的计算机上，如果 inputs.conf 文件尚不存在，请在 $SPLUNK_HOME/etc/system/local 目录中创建该文件。
2. 使用文本编辑器打开 $SPLUNK_HOME/etc/system/local/inputs.conf。
3. 添加以下段落以引用 accounts.log 文件并将来源类型分配至 accounts.log 数据。

```
[monitor:///opt/appserver/logs/accounts.log]
sourcetype = SSN-CC-Anon
```

4. 保存文件并将其关闭。

在 props.conf 中定义 sed 脚本

在此示例中，props.conf 会使用 SEDCMD 设置直接执行转换。

SEDCMD 词干后面的 -Anon 子句可以是可帮助您识别转换脚本如何操作的任何字符串。该子句必须存在，因为该子句和 SEDCMD 词干构成了脚本的类别名称。等号（=）后面的文本是调用转换的正则表达式。

1. 在运行重型转发器的计算机上，如果 props.conf 文件尚不存在，请在 $SPLUNK_HOME/etc/system/local 目录中创建该文件。
2. 使用文本编辑器打开 $SPLUNK_HOME/etc/system/local/props.conf。
3. 添加以下段落以引用您在 inputs.conf 文件中创建的来源类型以进行掩码转换。

```
[SSN-CC-Anon]
SEDCMD-Anon = s/ss=\d{5}(\d{4})/ss=xxxxx\1/g s/cc=(\d{4}-){3}(\d{4})/cc=xxxx-xxxx-xxxx-\2/g
```

4. 保存文件并将其关闭。
5. 重新启动重型转发器。

使用正则表达式转换使数据匿名

您可通过新建**转换**以掩码显示数据。转换会接收传入数据并根据您提供的配置更改数据。在此情况下，转换使用遮盖真实的敏感数据的字符替换部分数据，同时保留原始的数据格式。

假设您有一个名为 MyAppServer.log 的应用程序服务器日志文件，包含事件如下所示：

```
"2006-09-21, 02:57:11.58",  122, 11, "Path=/LoginUser Query=CrmId=ClientABC&
ContentItemId=TotalAccess&SessionId=3A1785URH117BEA&Ticket=646A1DA4STF896EE&
SessionTime=25368&ReturnUrl=http://www.clientabc.com, Method=GET,IP=209.51.249.195,
Content=", ""
"2006-09-21, 02:57:11.60",  122, 15, "UserData:<User CrmId="clientabc"
UserId="p12345678"><EntitlementList></EntitlementList></User>", ""
"2006-09-21, 02:57:11.60",  122, 15, "New Cookie: SessionId=3A1785URH117BEA&
Ticket=646A1DA4STF896EE&CrmId=clientabcUserId=p12345678&AccountId=&AgentHost=man&
AgentId=man, MANUser: Version=1&Name=&Debit=&Credit=&AccessTime=&BillDay=&Status=
&Language=&Country=&Email=&EmailNotify=&Pin=&PinPayment=&PinAmount=&PinPG=
&PinPGRate=&PinMenu=&", ""
```

您想要更改数据，这样 sessionID 和 Ticket 字段将以掩码显示，事件显示如下所示：

```
"2006-09-21, 02:57:11.58",  122, 11, "Path=/LoginUser Query=CrmId=ClientABC&
ContentItemId=TotalAccess&SessionId=##########7BEA&Ticket=##########96EE&
SessionTime=25368&ReturnUrl=http://www.clientabc.com, Method=GET,IP=209.51.249.195,
Content=", ""
```

您可以使用 inputs.conf、props.conf 和 transforms.conf 文件更改 Splunk 平台访问时来自 MyAppServer.log 文件的数据。所有这些配置文件位于 Splunk Cloud 重型转发器或 Splunk Enterprise 索引器上的 $SPLUNK_HOME/etc/system/local/ 目录中。

使用正则表达式转换使数据匿名的前提条件

要以掩码显示敏感数据，您必须满足以下要求：

- 拥有您想要匿名的数据

- 了解正则表达式如何工作
- 具有指向此数据所在位置的 inputs.conf 配置文件
- 具有以掩码显示数据的 transforms.conf 配置文件
- 具有 props.conf 配置文件，该文件针对您想要掩码以显示的数据引用了 transforms.conf 文件

配置 inputs.conf

在此示例中，您可新建 MyAppServer-Anon 来源类型。您新建的转换会用此来源类型了解要转换哪些数据。您可以从其他选项中选择要转换的数据。

按照以下步骤为本示例配置 input.conf 文件：

1. 在运行重型转发器的计算机上，如果 inputs.conf 文件尚不存在，请在 $SPLUNK_HOME/etc/system/local 目录中创建该文件。
2. 使用文本编辑器打开 $SPLUNK_HOME/etc/system/local/inputs.conf。
3. 添加以下段落以引用 MyAppServer.log 文件并将来源类型分配至 MyAppServer.log 数据。

   ```
   [monitor:///opt/MyAppServer/logs/MyAppServer.log]
   sourcetype = MyAppServer-Anon
   ```
4. 保存文件并将其关闭。

配置 transforms.conf 文件

Splunk 平台使用 transforms.conf 文件执行数据的转换。按照以下步骤为本示例配置 transforms.conf 文件：

1. 在运行重型转发器的计算机上，如果 transforms.conf 文件尚不存在，请在 $SPLUNK_HOME/etc/system/local 目录中创建该文件。
2. 使用文本编辑器打开 $SPLUNK_HOME/etc/system/local/transforms.conf。
3. 添加以下文本定义使 sessionID 字段匿名的转换，这样会只显示字段中的最后四个字符：

   ```
   [session-anonymizer]
   REGEX = (?m)^(.*)SessionId=\w+(\w{4}[&"].*)$
   FORMAT = $1SessionId=########$2
   DEST_KEY = _raw
   ```
4. 直接将以下文本添加到 session-anonymizer 段落下方，以定义用于 Ticket 字段的转换，类似 sessionID 字段：

   ```
   [ticket-anonymizer]
   REGEX = (?m)^(.*)Ticket=\w+(\w{4}&.*)$
   FORMAT = $1Ticket=########$2
   DEST_KEY = _raw
   ```
5. 保存文件并将其关闭。

配置 props.conf 配置文件

Props.conf 指定用于使数据匿名的转换。该文件会引用您在 transforms.conf 文件中定义的一个或多个转换类别。

本示例中，session-anonymizer 和 ticket-anonymizer 是您在 transforms.conf 文件中定义的转换类别名称。

按照以下步骤为本示例配置 props.conf 文件：

1. 在运行重型转发器的计算机上，如果 props.conf 文件尚不存在，请在 $SPLUNK_HOME/etc/system/local 目录中创建该文件。
2. 使用文本编辑器打开 $SPLUNK_HOME/etc/system/local/props.conf。
3. 添加以下段落以引用您在 transforms.conf 文件中创建的转换以进行掩码转换。

   ```
   [MyAppServer-Anon]
   TRANSFORMS-anonymize = session-anonymizer, ticket-anonymizer
   ```
4. 保存文件并将其关闭。
5. 重新启动重型转发器。

使用 sed/SEDCMD 脚本的替换示例

假设您有一个想为其新建索引的文件 abc.log，而且想用大写字母“A”、“B”和“C”替代事件中的所有小写字母“a”、“b”或“c”。

将以下段落和设置添加至 props.conf 文件：

```
[source::.../abc.log]
SEDCMD-abc = y/abc/ABC/
```

Splunk 平台已使用“A”、“B”和“C”分别替代每个“a”、“b”和“c”。搜索 source="*/abc.log" 时，您的数据中不会显示小写字母 "a"、"b" 和 "c"。

匿名数据的注意事项

使数据匿名可能会有以下注意事项。

使用 sed 脚本使数据匿名的限制

如果您使用 SEDCMD 方法使数据匿名，则以下限制使用：

- SEDCMD 脚本在索引时间内仅适用于 _raw 字段。使用正则表达式转换，您可以将更改应用于其他字段。
- 您不能对单个 props.conf 文件中的同一主机、数据来源或来源类型使用多个 SEDCMD 类型转换。

使用正则表达式转换使数据匿名的限制

如果您使用正则表达式转换使数据匿名，则包括您定义转换并将其设为大于最大预期事件的数字时的 LOOKAHEAD 设置。否则，匿名可能会失败。

Splunk Cloud 和 Splunk Enterprise 索引器不会分析结构化数据

您向 Splunk Cloud 或 Splunk Enterprise 索引器转发结构化数据时，即使您已经在该索引器上配置了 props.conf 文件（使用 INDEXED_EXTRACTIONS 设置进行配置），该平台也不会分析该数据。转发数据会跳过索引器上的下列处理队列，这就排除了对于该数据的分析：

- parsing
- aggregation
- typing

转发器必须先分析数据，然后再将该数据发送到 Splunk Cloud 或 Splunk Enterprise 索引器。要实现该目标，您必须在发送数据的转发器上设置 props.conf 文件。这包括配置 INDEXED_EXTRACTIONS 设置，以及其他任何分析、筛选、匿名及路由规则。

通用转发器只能分析结构化数据。请参阅“转发从结构化数据文件中提取的数据”。

配置时间戳

时间戳分配如何工作

时间戳处理是**事件处理**中的一个关键步骤。Splunk 软件通过以下方式使用**时间戳**：

- 按时间关联**事件**
- 在 Splunk Web 中创建时间线直方图
- 为搜索设置时间范围

Splunk 软件在**索引时间**为事件添加时间戳。软件使用它在原始事件数据中找到的信息自动分配时间戳值。如果事件不包含显式时间戳，Splunk 软件会尝试通过其他方法为其分配一个时间戳值。对于某些数据，您可能需要帮助 Splunk 软件学习如何识别时间戳。您可以通过配置时间戳提取来完成此操作。要配置时间戳提取，请参阅本主题后面的“配置时间戳提取”部分。

Splunk 软件使用协调世界时 (UTC) 时间格式将时间戳值存储于 `_time` 字段中。

如果您使用的是 Splunk Cloud 并且需要修改时间戳提取，请使用重型转发器来引入数据并将数据发送到您的 Splunk Cloud 实例。您可以在重型转发器上配置时间戳提取。如果使用的是 Splunk Enterprise 并且需要修改时间戳提取，则可以修改 Splunk Enterprise 实例上的配置文件。

有关事件处理的更多信息，请参阅“配置事件处理”一章。

Splunk 软件如何分配时间戳

Splunk 软件使用下列优先顺序规则为事件分配时间戳：

1. 该软件使用显式 `TIME_FORMAT`（如果有）在事件自身中查找时间或日期。您可以在 `props.conf` 文件中配置 `TIME_FORMAT` 属性。
2. 如果未配置数据的 `TIME_FORMAT`，Splunk 软件会尝试在事件自身中自动识别时间或日期。Splunk 软件使用事件的来源类型（其中包括 `TIME_FORMAT` 信息）试图查找时间戳。
3. 如果事件有时间和日期，但没有年份，Splunk 软件会确定年份，并从该日期构建时间戳。请参阅“Splunk 软件如何确定没有年份的时间戳”。
4. 如果数据来源中的所有事件都没有日期，Splunk 软件会尝试在来源名称或文件名称中查找日期。文件名中未标识时间，因此事件必须具有时间，即使它们没有日期。
5. 就文件来源而言，如果无法在文件名称中识别出日期，Splunk 软件会使用文件的修改时间。
6. 还有最后一种方法，Splunk 软件会在为每个事件建立索引时把时间戳设置为当前的系统时间。

Splunk 软件只能从数据来源中提取日期，而不能提取时间。如果需要从来源中提取时间，请使用转换配置。请参阅“在索引时间新建自定义字段”。

Splunk 软件如何确定没有年份的时间戳

如果 Splunk 软件发现事件中的时间戳没有包含年份元素，软件会使用以下逻辑确定年份：

1. 使用上次分析的事件日期或当前时钟时间来标识当前日期。
2. 然后以该日期所在的年份为基准，并通过多个测试确定该年份：
    1. 如果新事件的日期是 12 月 31 日，而当前日期是 1 月 1 日，则基准年减一年。
    2. 如果新事件的日期是 1 月 1 日，而当前日期是 12 月 31 日，则基准年加一年。
    3. 如果新事件的日期是 2 月 29 日，则软件会确定当前年份是否为闰年。
    4. 如果当前年份是闰年，则以该年为基准年。如果不是，则使用前一个闰年。
3. 如果以上测试都无法成功确定基准年，软件则会使用以下步骤来确定年份：
    1. 计算从 1 月 1 日起的天数来确定新事件在一年中的日期。
    2. 如果有上一个事件的日期信息，并且该事件在一年中的天数大于新事件在一年中的天数加上 4 天，则基准年加一年。
    3. 如果没有上一个事件的日期信息，而新事件在一年中的天数大于当前日期在当年的天数加上 2 天，则基准年减一年。
4. 软件随后将基准年分配给事件的时间戳。此时间戳仍然必须通过时间范围检查，才会生效。

示例 1

如果 Splunk 软件在 2017 年 5 月 26 日遇到一个日期为 `26 Jun` 的新事件，而且软件无法确定上一个事件的年份，则按照以下步骤确定事件时间戳：

1. 由于软件无法确定上一个事件的年份，所以它将基准年设为 `2017`，因为这是当前日期的年份。
2. 该软件会检查日期是 12 月 31 日还是 1 月 1 日。12 月 31 日和 1 月 1 日测试失败，因此基准年仍然为 `2017`。
3. 该软件会检查日期是否出现在闰年。闰年测试失败，因为此日期不是 2 月 29 日。基准年仍然为 `2017`。
4. Splunk 软件计算 6 月 26 日在当年的天数为 `177` 天。

5. 由于软件无法确定上一个事件的年份，所以将这个数字中加了 2，即 179。
6. 然后，软件将 179 与当前日期在当年的天数（2017 年 5 月 26 日，即第 147 天）相比较。
7. 由于 179 大于 147，所以软件将年份从 2017 减为 2016。
8. 软件随后构建此新的时间戳：26 Jun 2016。
9. 如果新时间戳在设置的时间范围内，软件会为事件添加此时间戳。

示例 2

如果 Splunk 软件在 2017 年 5 月 26 日遇到一个日期为 10 Apr的新事件，且软件确定之前事件的年份为 2017，则按照以下步骤确定事件时间戳：

1. 由于软件确定了上一个事情的年份，所以会将 2017 设为基准年。
2. 该软件会检查日期是 12 月 31 日还是 1 月 1 日。12 月 31 日和 1 月 1 日测试失败，因此基准年仍然为 2017。
3. 该软件会检查日期是否出现在闰年。闰年测试失败，因为此日期不是 2 月 29 日。基准年仍然为 2017。
4. Splunk 软件计算 4 月 10 日在当年的天数为 100 天。
5. 由于有上一个事件的年份信息，所以软件会将此数字加上 4，即 104。
6. 然后，软件将 104 与当前日期在当年的天数（2017 年 5 月 26 日，即第 147 天）相比较。
7. 由于 104 小于 147，所以软件将年份从 2017 增加为 2018。
8. 软件随后构建此新的时间戳：10 Apr 2018。
9. 默认情况下，这个新时间戳是无效的，因为它超出了默认的 MAX_DAYS_HENCE 设置，该设置将有效时间戳限制为将来 2 天。软件会使用当前日期 26 May 2017 作为时间戳，并将此时间戳应用到事件中。

配置时间戳提取

大多数事件不需要特殊时间戳处理。Splunk 软件会自动为事件识别和提取时间戳。对于某些数据来源和分布式部署，您可能需要配置时间戳的提取方式，以确保时间戳的格式正确。

可通过两种方式配置时间戳提取：

- 使用 Splunk Web 中的**设置来源类型**页面交互调整示例数据上的时间戳。如果您对结果满意，请将更改保存为新来源类型，然后将此来源类型应用于您的数据导入。请参阅“将正确的来源类型分配给您的数据”。
- 请直接编辑 props.conf 配置文件。请参阅“配置时间戳识别”。

您还可以出于以下目的配置时间戳提取：

- 应用时区偏移。请参阅“指定时间戳的时区”。
- 从具有多个时间戳的事件中提取正确的时间戳。请参阅“为具有多个时间戳的事件配置时间戳分配”。
- 提高索引性能。请参阅“调整时间戳识别以获得更佳的索引性能”。

从新输入添加数据时的注意事项

如果您要为来自新输入的数据建立索引，然后发现需要调整时间戳提取过程，则在进行配置更改之后，您必须为此数据重新建立索引。

请考虑预览您的数据以防止需要为数据重新建立索引。或者，您也可以在生产 Splunk Cloud 环境的单独索引中测试新的数据导入。如果您使用的是 Splunk Enterprise，则可以先创建单独的测试 Splunk 部署，然后再将数据添加到您的生产实例。创建单独的测试部署允许您删除并重新建立索引，直到获得预期的结果。

配置时间戳识别

大多数事件不需要特殊时间戳处理。Splunk 平台可以正确识别和提取时间戳。但是，对于某些来源和分布式部署，您可能需要配置 Splunk 平台如何提取时间戳，以确保时间戳的格式正确。

您可以通过以下方式配置时间戳提取：

- 使用 Splunk Web 中的“设置来源类型”页面交互调整示例数据上的时间戳。如果您对结果满意，可将更改保存为新来源类型，然后将此来源类型应用于数据导入。仅当您将文件直接上载到实例时，此选项才在 Splunk Cloud 上可用。请参阅“将正确的来源类型分配给您的数据”。
- 使用 props.conf 配置文件。在以下条件下，此选项在 Splunk Cloud 和 Splunk Enterprise 上均可用：
  - 如果您使用的是 Splunk Cloud 并且需要修改时间戳提取，请使用重型转发器来引入数据并将数据发送到您的 Splunk Cloud 实例。重型转发器允许您指定提取时间戳的配置。您不能直接使用通用转发器或 Splunk Cloud 进行这些配置。
  - 在 Splunk Enterprise 实例上，如果您需要修改时间戳提取，请在索引器上指定配置。在必须转发数据的情况下，您必须配置重型转发器才能处理这些更改。

时间戳处理器

在 Splunk Enterprise 实例上，默认情况下您可以在 $SPLUNK_HOME/etc/datetime.xml 中找到时间戳处理器。正常情况下，您无需编辑此文件，除非您要处理不寻常的自定义时间戳。您无法在 Splunk Cloud 实例上编辑此文件，因为您没有访问 Splunk Cloud 文件系统的权限。

如果您需要配置时间戳识别，则可以如本主题所述，通过编辑 props.conf 配置文件中的时间戳设置来进行更改。

如果您有自定义的时间戳且无法通过配置 props.conf 文件进行处理，请用 DATETIME_CONFIG 设置替代您自己的时间戳处理器。此设置指定 Splunk 平台进行时间戳处理应使用的文件。

在 props.conf 配置文件中编辑时间戳属性

您可以在重型转发器上编辑时间戳属性以确保 Splunk Cloud 查看并使用正确的时间戳，或者您可以直接在 Splunk Enterprise 实例上编辑时间戳。

如需配置 Splunk 平台如何识别时间戳，请编辑 props.conf 配置文件。使用此文件，您可以控制 Splunk 平台如何处理它在传入事件中看到的时间戳。请参阅*《管理员手册》*中的“props.conf 规范文件”。

props.conf 文件有几个用于时间戳处理的设置。例如，可以使用 TIME_FORMAT 设置为基于 strptime() 字符串到时间结构转换函数的时间戳指定格式。

还可以配置在事件中的哪个位置查找时间戳、事件使用的时区，或如何处理不同货币的时间戳。

按照以下步骤配置时间戳识别：

1. 对于从转发器接收数据的 Splunk Cloud 实例或 Splunk Enterprise 实例，安装新的 Splunk Enterprise 实例并将其配置为重型转发器。
2. 如果您使用的是 Splunk Cloud，请在转发器上安装 Splunk Cloud 通用转发器凭据包。
3. 在转发器上，使用文本编辑器编辑 $SPLUNK_HOME/etc/system/local/ 目录或 $SPLUNK_HOME/etc/apps/ 中您自己的自定义应用程序目录中的 props.conf 文件。
4. 指定适合您的需求、适合您的传入数据并符合语法的时间戳识别设置。有关更多信息，请参阅本主题中的“语法概述”。
5. 保存更改并重新启动转发器。
   转发器开始根据您提供的新配置处理时间戳。

如您之前不曾使用过配置文件，请参阅*《管理员手册》*中的“关于配置文件”。

语法概述

当您配置时间戳识别时，您的配置文件及其设置遵循适当的语法非常重要。语法错误可能会导致错误的时间戳和事件引入。

请参阅以下语法概述：

```
[<spec>]
DATETIME_CONFIG = <filename relative to the $SPLUNK_HOME directory>
TIME_PREFIX = <regular expression>
MAX_TIMESTAMP_LOOKAHEAD = <integer>
TIME_FORMAT = <strptime-style format>
TZ = <POSIX time zone string>
MAX_DAYS_AGO = <integer>
MAX_DAYS_HENCE = <integer>
MAX_DIFF_SECS_AGO = <integer>
MAX_DIFF_SECS_HENCE = <integer>
```

在此语法中，<spec>可以具有以下任何含义：

- <sourcetype>，即事件的来源类型。
- host::<host>，其中 <host> 是事件的主机值。
- source::<source>，其中 <source> 是事件的来源值。

如果一事件中包含的数据匹配到 <spec> 的值，段落中指定的时间戳规则将适用于该事件。为处理不同的 <spec> 值，您可以拥有多个段落。

时间戳有效性设置和对事件的影响

默认情况下，Splunk® Enterprise 会对所有事件建立索引，除非您通过其他方法特别筛选掉了一些事件。

如果您在一个段落中设置 MAX_DAYS_AGO、MAX_DAYS_HENCE、MAX_DIFF_SECS_AGO 或 MAX_DIFF_SECS_HENCE 设置，而且事件的时间戳落

在这些设置的参数之外，那么，Splunk® Enterprise 会使用以下算法来确定时间戳：

- Splunk 软件使用前一个事件的时间戳来分配当前事件的时间戳。
- 如果前一个事件的时间戳无法确定，Splunk 软件会使用当前的索引时间为事件分配时间戳。

Splunk® Enterprise 不会因为事件落在这些设置的参数之外而将其丢弃。

时间戳设置

可使用以下时间戳配置设置：

设置	描述	默认
DATETIME_CONFIG = <filename relative to $SPLUNK_HOME>	为配置 Splunk 时间戳处理器，请指定一个要使用的文件。 正常情况下，您无需创建自己的时间戳处理器文件或修改默认 datetime.xml 文件。其他 props.conf 设置通常就足够了。然而，如果您的数据采用自定义时间戳格式，则您可能需要自己的 datetime.xml 文件版本。 设置 DATETIME_CONFIG = NONE 可以防止时间戳处理器运行。时间戳处理关闭时，Splunk® Enterprise 不在事件的文本中查找时间戳；相反，它使用事件的 time of receipt，即事件通过其输入到达的时间。如果是基于文件的输入，事件时间戳将取自输入文件的修改时间。 设置 DATETIME_CONFIG = CURRENT 可以在 Splunk® Enterprise 为其新建索引时将当前的系统时间分配给所有事件。 CURRENT 和 NONE 作为DATETIME_CONFIG 的值显式禁用时间戳识别，因此默认的事件界限检测（BREAK_ONLY_BEFORE_DATE = true）可能无法如预期般运行。运用这些设置时，请使用 SHOULD_LINEMERGE、BREAK_ONLY_* , MUST_BREAK_* 或这两个设置来控制事件合并。	$SPLUNK_HOME/etc/datetime.xml
TIME_PREFIX = <regular expression>	设置之后，Splunk® Enterprise 会先使用指定的正则表达式查找匹配，然后再尝试提取时间戳。该时间戳算法仅在第一个正则表达式匹配项结尾之后的那个事件文本中查找时间戳。 使用确切指向事件时间戳前面的正则表达式。例如，若时间戳在您的事件中位于abc123 短语之后，则将 TIME_PREFIX 设置为 abc123。 如果在事件文本中找不到 TIME_PREFIX，则无法提取时间戳。	空字符串
MAX_TIMESTAMP_LOOKAHEAD = <integer>	指定 Splunk 平台应在事件中查找时间戳的深度或字符数。 此约束从 TIME_PREFIX 所定位的位置处开始应用。 例如，如果 TIME_PREFIX 所定位的位置是事件中的 11 个字符处且 MAX_TIMESTAMP_LOOKAHEAD 设置为 10，则时间戳提取将被限定为字符 11 至 20。 如果设置为 0 或 -1，将会禁用时间戳识别的长度约束。但是，这会对性能产生负面影响，导致输入行的长度有所缩放，或者为事件拆分重新定义 LINE_BREAKER 时，事件大小有所缩放。	128 个字符
TIME_FORMAT = <strptime-style format>	为提取时间戳，请指定一个 strptime() 格式的字符串。	空字符串

	strptime() 是用于指定时间格式的 Unix 标准。有关更多信息，请参阅“增强的 strptime() 支持”部分。 TIME_FORMAT 在 TIME_PREFIX 后（如无 TIME_PREFIX 设置则正好在事件开头）开始读取。如您使用 TIME_PREFIX，其必须与时间戳开始前的所有字符（包括时间戳开始前的字符）相匹配。如您未设置 TIME_PREFIX 但设置了 TIME_FORMAT，则时间戳必须出现在每个事件的开头。否则，Splunk 平台将无法处理格式化指令，而且每个事件都将包含一条无法使用 strptime 的警告。 为了获得最佳结果，请使用 <strptime-style format> 描述具体的日期和时间。 如果 <strptime-style format> 中包含小时组件却不包含分钟组件，TIME_FORMAT 将忽略小时组件。它将该格式视为异常且认为精度是仅日期。	
TZ = <timezone_identifier>	事件的时区按以下方式决定： • 如果事件在原始文本中包含时区（如 UTC 或 -08:00），请直接使用。 • 否则，如果 TZ 被设置为有效的时区字符串，则使用它。使用 TZ 时区数据库中的值指定时区设置。请参阅 Wikipedia 上的 [1]。 • 如果到达索引器的事件源自转发器，且该索引器和转发器均为 6.0 或更高版本，则请使用该转发器提供的时区。 • 否则，请使用运行 splunkd 的系统的时区。 更多详情和示例，请参阅“指定时间戳的时区”。	空字符串
TZ_ALIAS = <key=value>[,<key=value>]...	针对如何解释从事件中提取到的时区字符串提供管理员层级的控制。例如，EST 可以指东部（美国）标准时间或东部（澳大利亚）标准时间。有许多其他具有多种扩展的三个字母时区首字母缩写。 如果这些值的默认映射按预期运行，则无需使用 TZ_ALIAS。例如，EST 默认映射到东部美国。 对 TZ 值毫无影响。仅影响来自事件文本的时区字符串，无论该事件文本是任何已配置好的 TIME_FORMAT 或基于模式的猜测回退。 该设置是以逗号分隔的 key=value 对列表。 键根据事件时区指示符的文本进行匹配，值是将时间戳映射到 UTC/GMT 时要使用的时区指示符。 该值是表示所需偏移的另一个 TZ 指示符。 例如： TZ_ALIAS = EST=GMT+10:00 请参阅*《管理员手册》*中的 props.conf 规范文件，了解更多示例。	未设置
MAX_DAYS_AGO = <integer>	指定已提取日期可以有效的最大过去天数（从当前日期算起）。 例如，若 MAX_DAYS_AGO = 10，Splunk 平台将忽略距离当前日期 10 天以上的过去日期，改而使用前一个事件的时间戳；如果无法确定前一个事件的时间戳，Splunk 平台将使用事件当前的索引时间。 最大可设置过去天数为 10951。	2,000 天 如果您的数据已超过 2,000 天，请增加此设置。
MAX_DAYS_HENCE = <integer>	指定已提取日期可以有效的最大未来天数（从当前日期算	2 天

	起）。 例如，若 MAX_DAYS_HENCE = 3，Splunk 平台将忽略距离当前日期 3 天以上的未来日期，并使用前一个事件的时间戳；如果无法确定前一个事件的时间戳，平台将使用事件当前的索引时间。 窗口较小时出现误报的可能性较小。更改此设置时请谨慎。 如果您的服务器日期设置错误或者时区提前了一天，请将此值至少设置为 3。允许最长未来一天的时间戳提取。 最大可设置天数为 10950。	
MAX_DIFF_SECS_AGO = <integer>	如果事件时间戳和前一个时间戳相比早了 <integer> 秒，仅当该事件时间戳的时间格式与数据来源内大多数时间戳相同时，Splunk 平台才会接受它。 如果您的时间戳毫无顺序，请考虑增加此值。 最大可设置秒数为 2147483646。	3,600 秒（1 小时）
MAX_DIFF_SECS_HENCE = <integer>	如果事件时间戳和前一个时间戳相比晚了 <integer> 秒，仅当该事件时间戳的时间格式与数据来源内大多数时间戳相同时，Splunk 平台才会接受它。 如果您的时间戳毫无顺序，或者如果您具有写入少于一周的日志，请考虑增加此值。 最大可设置秒数为 2147483646。	604800 秒（1 周）

增强的 strptime() 支持

使用 props.conf 文件中的 TIME_FORMAT 设置来配置时间戳分析。此设置采用 strptime() 格式字符串来提取时间戳。

Splunk 平台实施支持其他格式的 Unix strptime() 增强版本，从而允许微秒、毫秒、任何时间宽度格式及一些其他时间格式与其兼容。包括以下其他格式：

格式	描述
%N	适用于 GNU 日期-时间纳秒。通过提供宽度来指定亚秒分析：%3N = 毫秒，%6N = 微秒，%9N = 纳秒。亚秒级分析取决于设置 ADD_EXTRA_TIME_FIELDS。请参阅*《管理员手册》*中的 props.conf。
%Q, %q	适用于毫秒，微秒则适用于 Apache Tomcat。如果宽度已指定， %Q 和 %q 可以设置任何时间精度的格式。
%l	适用于采用 12 小时制格式的小时。如果 %l 出现在 %S 或 %s 之后（如 %H:%M:%S.%l），它将具有毫秒的 log4cpp 含义。
%+	适用于标准 Unix 日期格式时间戳。
%v	适用于 BSD 和 OSX 标准日期格式。
%Z	时区缩写。如果没有时区信息，则没有任何内容。
%z, %:z, %::z	国际组织标准 (ISO) 时区偏移指示器 8601 格式。比如：PST 为 -0800、GMT 为 +0000、或时区无法确定的情况下为无。如果时间戳偏移中包含小时和分钟，例如 -08:00，则使用 %:z。如果时间戳偏移中包含小时、分钟和秒，例如 -08:00:00，则使用 %::z。
%o	适用于 AIX 时间戳支持。 %o 用作 %Y 的别名。
%p	AM 或 PM 的区域设置等效值。可能没有等效值。
%s	Epoch（10 位）。

以文字圆点和亚秒指示符（如 %Q、 %q、 %N）结尾的 strptime 表达式将终止圆点和转换指示符视为可选项。如果文本中缺少 .subseconds 部分，仍可提取时间戳。

strptime() 格式表达式示例

以下是一些日期格式的示例，使用 strptime() 表达式来处理这些日期格式：

日期格式	表达式
1998/12/31	%Y-%m-%d
98-12-31	%y-%m-%d
1998 years, 312 days	%Y years, %j days
Jan 24, 2003	%b %d, %Y
January 24, 2003	%B %d, %Y
1397477611.862	%s.%3N

Splunk 平台目前无法识别时间戳中的非英文月份名称。如果您具有向日志文件写入非英文月份名称的应用程序，请将该应用程序重新配置为使用数字月份（如果可能）。

示例

您的数据可能包含易于识别的时间戳，如以下示例所示：

```
...FOR: 04/24/07 PAGE 01...
```

要提取此时间戳，请在 props.conf 文件中添加以下段落：

```
[host::foo]
TIME_PREFIX = FOR:
TIME_FORMAT = %m/%d/%y
```

您的午夜还有一个包含时区信息的时间戳：

```
⋯Valid_Until=Thu Dec 31 17:59:59 GMT-06:00 2020
```

要提取此时间戳，请在 props.conf 文件中添加以下段落：

```
[host::bar]
TIME_PREFIX = Valid_Until=
TIME_FORMAT = %a %b %d %H:%M:%S %Z%:z %Y
```

您的数据中可能包含其他被分析为时间戳的信息。以下面的例子为例：

```
...1989/12/31 16:00:00 Wed May 23 15:40:21 2007...
```

Splunk 平台将该日期提取为无用的 Dec 31, 1989。此例中，请配置 props.conf 文件以从 host::foo 内的事件中提取正确的时间戳：

```
[host::foo]
TIME_PREFIX = \d{4}/\d{2}/\d{2} \d{2}:\d{2}:\d{2} \w+\s
TIME_FORMAT = %b %d %H:%M:%S %Y
```

此配置假设 host::foo 中的所有时间戳格式均相同。为避免可能的时间戳错误，尽量将您的 props.conf 段落配置得具体一点。

更多有关从包含多个时间戳的事件中提取正确时间戳的信息，请参阅“为含多个时间戳的事件配置时间戳分配”。

针对特定需求配置时间戳

您可以使用本主题中介绍的设置来配置时间戳提取处理器以满足某些专门用途。例如，您可以为以下需求配置时间戳：

- 应用时区偏移。请参阅“指定时间戳的时区”。
- 从具有多个时间戳的事件获取正确的时间戳。请参阅“为具有多个时间戳的事件配置时间戳分配”。
- 提高索引性能。请参阅“调整时间戳识别以获得更佳的索引性能”。

配置时间戳如何出现在搜索结果中

您可以使用浏览器区域设置来配置 Splunk Web 在搜索结果中显示时间戳的方式。有关设置浏览器区域设置的信息，请参阅*《管理员手册》*中的“配置用户语言和区域设置”。

重新配置时间戳如何出现在原始数据中

尽管 Splunk 平台使用浏览器区域来配置时间戳在搜索结果中的显示方式，但原始数据仍保留其原始格式。您可能希望对此进行更改，以便数据格式在原始数据和搜索结果中都为标准化格式。使用 props.conf 和 transforms.conf 文件完成此操作。

例如，假设原始事件中的时间戳数据与以下类似：

```
06/07/2011 10:26:11 PM
```

但您希望它与其在搜索结果中的显示方式相对应：

```
07/06/2011 10:26:11 PM
```

您可以使用 props.conf 和 transforms.conf 文件来转换原始事件中的时间戳。

在 transforms.conf 文件中，添加以下段落：

```
[resortdate]
REGEX = ^(\d{2})\/(\d{2})\/(\d{4})\s([^/]+)
FORMAT = $2/$1/$3 $4
DEST_KEY = _raw
```

在 props.conf 中添加本段落，其中 <spec> 符合您的数据：

```
[<spec>]
TRANSFORMS-sortdate = resortdate
```

为具有多个时间戳的事件配置时间戳分配

如果一个事件中包含多个时间戳，您可以指定事件在建立索引期间使用的时间戳。为包含 syslog 主机链数据的事件建立索引时，配置时间戳尤其有用。

通过编辑 props.conf 配置文件配置位置时间戳提取。虽然在 Splunk Web 中配置时间戳提取的能力有限，但您可以通过在重型转发器上使用配置文件来获得最佳结果。有关此配置文件的更多信息，请参阅“props.conf”。

在 Splunk Cloud 实例上，您必须在配置转发器以将数据发送到 Splunk Cloud 实例后在重型转发器上配置时间戳。有关针对时间戳编辑 props.conf 文件的一般信息，请参阅“配置时间戳识别”。

如果您使用的是 Splunk Enterprise 而且需要修改时间戳提取，请在您的索引器计算机上执行配置；或者，如果您想要转发数据，请使用重型转发器并在重型转发器运行的计算机上执行配置。

配置位置时间戳提取

如需指定您想要提取的时间戳的位置，请执行以下步骤：

1. 将 TIME_PREFIX 和 MAX_TIMESTAMP_LOOKAHEAD 设置添加到 props.conf 文件的段落中。
   通过设置 TIME_PREFIX 的正则表达式，您可以指定一个指示时间戳查找起始点的字符模式。
2. 设置 MAX_TIMESTAMP_LOOKAHEAD 设置的值即可指定在事件中查找时间戳的深度（位于 TIME_PREFIX 位置之后）。
   通过约束提前时间量，您可以提高确定和提取时间戳的准确性和性能。

设置好 TIME_PREFIX 设置后，Splunk 平台会先扫描事件文本寻找其正则表达式的匹配项，然后再尝试提取时间戳。该时间戳算法在第一个正则表达式匹配项结尾之后的那个文本中查找时间戳。例如，如果将 TIME_PREFIX 设置设为 abc123，仅在第一个 abc123 后出现的文本才会用于时间戳提取。

TIME_PREFIX 也会设置 MAX_TIMESTAMP_LOOKAHEAD 设置的起始点。提前量始于 TIME_PREFIX 正则表达式中文本的匹配部分之后。例如，如果 TIME_PREFIX 通过事件的前 11 个字符与文本匹配，且您想要提取的时间戳始终位于接下来的 30 个字符内，您可以设置 MAX_TIMESTAMP_LOOKAHEAD=30。时间戳提取将被限制在以字符 12 开始、以字符 41 结束的文本范围内。

示例

检查以下示例事件：

```
1989/12/31 16:00:00 Wed May 23 15:40:21 2007 ERROR UserManager - Exception thrown
Ignoring unsupported search for eventtype: /doc sourcetype="access_combined"
NOT eventtypetag=bot
```

要将时间戳识别为时间信息 May 23 15:40:21 2007 的第二个字符串，请按如下所示配置 props.conf 文件：

```
[source::/Applications/splunk/var/spool/splunk]
TIME_PREFIX = \d{4}\/\d{2}\/\d{2} \d{2}:\d{2}:\d{2} \w+\s
MAX_TIMESTAMP_LOOKAHEAD = 21
```

此配置指示 Splunk 平台定位与第一个时间戳构建相匹配的事件，但若在接下来的 21 个字符内（从 MAX_TIMESTAMP_LOOKAHEAD 设置中获得的数字）出现另一个时间戳，则 Splunk 平台会忽略第一个时间戳。Splunk 平台将找到第二个时间戳，因为它始终出现在此 21 个字符的限制内。

设置 MAX_TIMESTAMP_LOOKAHEAD 的值，仅根据您的时间戳提取需求查找事件，可以优化时间戳提取速度。本示例中的 MAX_TIMESTAMP_LOOKAHEAD 已优化，因此仅查找事件中位于正则表达式值后的 21 个字符。

用 datetime.xml 配置高级时间戳识别

Splunk 平台使用 datetime.xml 文件从建立索引的事件中提取日期和时间戳。文件包含说明 Splunk 平台如何从原始事件数据中进行提取的正则表达式。

在几乎所有情况中，您不需要修改 datetime.xml。

如果要在 Splunk Cloud 实例上修改此文件，请提交支持工单。无法手动修改 Splunk Cloud 实例上的文件。相反，如有必要，请在包含要发送到 Splunk Cloud 的数据的计算机上的通用转发器上配置文件。

当您用 props.conf 配置时间戳识别时，Splunk Enterprise 使用 datetime.xml 配置其时间戳处理器并针对来源、来源类型或主机从事件中提取时间戳。如果软件无法处理事件数据中的时间戳，您可以通过自定义 datetime.xml 版本训练 Splunk Enterprise 提取时间戳。

Datetime.xml 文件的结构

datetime.xml 文件由以下类型的代码块组成：

- 定义时间戳单个元素的代码块
- 使用文件中已定义的其他元素的代码块
- 提取模式代码块

每个定义代码块都有一个或多个 <text> 定义，其中包含 Splunk Enterprise 用于提取时间戳元素的正则表达式。

定义时间戳单个元素的代码块

这些单个元素可以包含年、月、日、小时和分钟等信息。以下示例代码块定义 Splunk Enterprise 用于提取事件数据中文本值月元素（例如，Jan、Mar）的正则表达式：

```
<define name="_litmonth"  extract="litmonth">
<text><![CDATA[(?<![\d\w])(jan|\x{3127}\x{6708}|feb|\x{4E8C}\x{6708}|mar|\x{4E09}\x{6708}|apr|\x{56DB}\x{6
708}|may|\x{4E94}\x{6708}|jun|\x{516D}\x{6708}|jul|\x{4E03}\x{6708}|aug|\x{516B}\x{6708}|sep|\x{4E5D}\x{6708}|
oct|\x{5341}\x{6708}|nov|\x{5341}\x{3127}\x{6708}|dec|\x{5341}\x{4E8C}\x{6708})[a-z,\.;]*]]></text>
</define>
```

使用文件中已定义的其他元素的代码块

以下示例代码块定义提取小时、分钟、秒、亚秒、某段时间和时区的 _time 元素：

```
<define name="_time" extract="hour, minute, second, subsecond, ampm, zone">
<text><![CDATA[(?<!\d)]]></text>
        <use name="_hour"/>
        <text><![CDATA[:]]></text>
        <use name="_minute"/>
        <text><![CDATA[:]]></text>
```

```
        <use name="_second"/>
        <text><![CDATA[(?:(?: \d{4})?[:,\.](\d+))? {0,2}]]></text>
        <use name="_ampm"/>
        <text><![CDATA[ {0,2}]]></text>
        <use name="_zone"/>
        <text><![CDATA[(?!:\d)]]></text>
</define>
```

提取模式代码块

提取模式代码块定义尝试从传入事件数据提取时间和日期的顺序。timePatterns 块定义 Splunk Enterprise 尝试提取时间戳的顺序，datePatterns 块定义提取日期的方式。

新建或修改自定义 datetime.xml 文件

在几乎所有情况中，您不需要修改 datetime.xml。而应配置 props.conf 配置文件提取时间戳。有关说明，请参阅“配置时间戳识别”的“在 props.conf 配置文件中编辑时间戳属性”部分。

如果 Splunk Enterprise 没有正确使用 props.conf 文件提取日期和时间，您可能需要修改或用自定义版本替换 datetime.xml。您可以使用 splunk train CLI 命令获取时间戳数据示例并生成您可以用于创建提取时间戳的自定义 datetime.xml 的代码。

splunk train CLI 命令已弃用，但是仍可用于帮助您根据您的示例时间戳数据为 datetime.xml 创建模式。

在您通过 splunk train 创建模式文件之后，您可以复制默认的 datetime.xml 文件并将修改添加到该文件中，或者您可创建仅包含自定义时间戳定义的新的 datetime.xml。

不要直接修改 $SPLUNK_HOME/etc/datetime.xml。每次升级时，Splunk Enterprise 都会覆盖此文件，对于自定义源类型以及实例上的所有其他源类型来说，由更改导致文件中出现任何错误都可能会造成严重的、持久的数据引入问题。如果您想要更改默认文件，将副本保存到 $SPLUNK_HOME/etc/system/local 中并在该路径下更改。

要创建或修改自定义 datetime.xml，请执行下列高级步骤：

1. 创建示例时间戳模式文件。
2. 针对文件运行 splunk 培训 CLI 命令。
3. 使用输出创建自定义 datetime.xml 文件。
4. 参考您的时间戳配置中的自定义 datetime.xml 文件。

创建示例时间戳模式文件

1. 通过提示窗口或 PowerShell 窗口，创建文本文件。
2. 将时间戳数据示例粘贴到此文件中。
3. 保存文件并将其关闭。
4. 改为 $SPLUNK_HOME/bin 目录。

针对文件运行 splunk 培训 CLI 命令

1. 改为 $SPLUNK_HOME/bin 目录：

   ```
   cd $SPLUNK_HOME/bin
   ```
2. 运行 splunk train CLI 命令：

   ```
   ./splunk train dates
   ```
3. 在软件询问您希望执行的操作之后，输入 L、l 或 learn 执行“learn”操作。
4. 输入包含时间戳示例的文件路径。
   Splunk Enterprise 显示示例的第一行，并提示您输入表示时间戳的值：

   ```
   ------------------------------------
   Interactively learning date formats.
   ------------------------------------
   INSTRUCTIONS: If a sample line does not have a timestamp, hit Enter.
   If it does have a timestamp, enter the timestamp separated by commas
   in this order: month, day, year, hour, minute, second, ampm, timezone.
   Use a comma as a placeholder for missing values.  For example, for a
   sample line like this "[Jan/1/08 11:56:45 GMT] login", the input
   should be: "Jan, 1, 08, 11, 56, 45, , GMT" (note missing AM/PM).
   Spaces are optional.
   SAMPLE LINE 1:
   ```

```
        Tue Jul 10 21:23:06 PDT 2007 Received Trade 330 with detail user: user3456 date: date: 10Jul200721:
        23:06 action: sell 3583 MNAG @ 42
-------------------------------------------------------------------------------
Enter timestamp values as: month, day, year, hour, minute, second, ampm, timezone.
   >
```

5. 输入月、日、年、小时、分钟、秒、时间段（上午/下午）和时区的值。
   如果值足够，Splunk 软件将显示以下消息，以表示其已记住了此模式：

```
Learned pattern.
-------------------------------------------------------------------------------
If you are satisfied that the timestamps formats have been learned, hit control-c.
-------------------------------------------------------------------------------
```

6. 如果 Splunk Enterprise 已正确学习时间戳格式，按 Ctrl+C。
   Splunk 软件会显示与以下类似的文本：

```
Patterns Learned.
It is highly recommended that you make changes to a copy of the default datetime.xml file.
For example, copy "/Applications/splunk/etc/datetime.xml" to "/Applications/splunk/etc/system/local/datetime.xml", and
work with that file.
In that custom file, add the below timestamp definitions, and add the pattern names
to timePatterns and datePatterns list.
For more details, see http://www.splunk.com/doc/latest/admin/TrainTimestampRecognition
-------------------------------------------------------------------------------
<define name="mycustom_date" extract="day,litmonth,year,">
<text><![CDATA[:\d+\s\w+\s(\d+)\s(\w+)\s(\d+)]]></text>
</define>
<define name="mycustom_time" extract="hour,minute,second,ampm,">
<text><![CDATA[(\d+):(\d+):(\d+)\s(\w+)]]></text>
</define>
-----------------------------------------------------
What operation do you want to perform? (default=learn)
-----------------------------------------------------
                          Enter choice: [Learn]/Test/Quit >
```

7. 查看输出中的模式定义。如果时间戳示例的定义符合您的要求，通过键入 Q、q 或 quit 退出 splunk train 会话。或者，重新键入 L、l 或 learn 尝试重新进行培训。

使用输出创建自定义 datetime.xml 文件

在您成功培训 Splunk Enterprise 了解您的自定义时间戳之后，您必须将 splunk train 生成的定义添加到 datetime.xml 的自定义版本中。

您可以使用以下选项创建此文件：

- 将时间戳定义添加到现有的 datetime.xml 中。这是首选方法。
- 创建新的 datetime.xml 文件，其中仅包含您自定义的时间戳定义。当数据源类型是非常严格的格式，但是 Splunk Enterprise 却错误地选择了更宽广的默认格式时，这个选项更好。

请勿编辑 $SPLUNK_HOME/etc/datetime.xml。始终复制此文件并将自定义时间戳样式添加到副本中。

1. 复制 $SPLUNK_HOME 目录中的 datetime.xml。

```
cd $SPLUNK_HOME/etc
cp datetime.xml system/local/
```

2. 打开 $SPLUNK_HOME/etc/system/local/datetime.xml 进行编辑。
3. 在单独的一行中，将 splunk train 命令生成的，以 define name 开头的代码块复制到文件中。此代码块可以放在 <datetime> 和 <timePatterns> 条目之间的任何位置。
4. 在 <timePatterns> 块中，添加引用您之前在 datetime.xml 文件中添加的定义行的行。
5. 在 <datePatterns> 代码块中，添加您在上一步添加的相同行。
6. 保存自定义 datetime.xml 文件并将其关闭。

请参阅本主题后面介绍的“自定义 datetime.xml 配置的示例”查看示例。

参考您的时间戳配置中的自定义 datetime.xml 文件

构建自定义 datetime.xml 文件后，您可以在 props.conf 中引用此文件，以提取自定义时间戳。您可以针对任何主机、源或源类型设置自定义时间戳提取模式。

1. 在 $SPLUNK_HOME/etc/system/local 中，如果 props.conf 不存在，则新建该文件。
2. 打开 $SPLUNK_HOME/etc/system/local 中的 props.conf 进行编辑。
3. 如果不存在，为需要自定义时间戳提取的主机、源或源类型添加段落。
4. 在此段落中，添加指向自定义 datetime.xml（相对于 $SPLUNK_HOME 目录）的 DATETIME_CONFIG 设置。例如：

```
[mysourcetype]
DATETIME_CONFIG = /etc/system/local/datetime.xml
MAX_TIMESTAMP_LOOKAHEAD = 128
MAX_DAYS_AGO = 28
```

5. 必要时，为需要自定义提取的其他主机、源或源类型重复之前的步骤。
6. 保存 props.conf 并将其关闭。
7. 重新启动 Splunk 平台。
8. 确认为与包含自定义时间戳提取模式的主机、源或源类型匹配的事件正确提取了时间戳。

自定义 datetime.xml 配置的示例

以下代码块是如何正确配置自定义 datetime.xml 文件的示例。

例如，假设 splunk train 命令生成了以下代码：

```
<define name="mycustom_date" extract="day,litmonth,year,">
<text><![CDATA[:\d+\s\w+\s(\d+)\s(\w+)\s(\d+)]]></text>
</define>
<define name="mycustom_time" extract="hour,minute,second,ampm,">
<text><![CDATA[(\d+):(\d+):(\d+)\s(\w+)]]></text>
</define>
```

请参阅“示例 1a”和“示例 1b”部分，了解从该代码继续进行的方法。请参阅“示例 2”部分了解下一个步骤。

示例 1a：修改现有的 datatime.xml

继续前面的示例，然后您可以将这些定义块添加到您之前复制的 $SPLUNK_HOME/etc/system/local 中的现有 datetime.xml 中：

```
<datetime>

<define name="mycustom_date" extract="day,litmonth,year,">
<text><![CDATA[:\d+\s\w+\s(\d+)\s(\w+)\s(\d+)]]></text>
</define>

<define name="mycustom_time" extract="hour,minute,second,ampm,">
<text><![CDATA[(\d+):(\d+):(\d+)\s(\w+)]]></text>
</define>

<... existing configurations removed for clarity ...>

<timePatterns>
      <use name="_time"/>
      <use name="_hmtime"/>
      <use name="_hmtime"/>
      <use name="_dottime"/>
      <use name="_combdatetime"/>
      <use name="_utcepoch"/>
      <use name="_combdatetime2"/>
      <use name="mycustom_time"/>
</timePatterns>

<datePatterns>
<use name="_usdate1"/>
      <use name="_usdate2"/>
      <use name="_isodate"/>
      <use name="_eurodate1"/>
      <use name="_eurodate2"/>
      <use name="_bareurlitdate"/>
      <use name="_orddate"/>
      <use name=" combdatetime"/>
```

```
      <use name="_masheddate"/>
      <use name="_masheddate2"/>
      <use name="_combdatetime2"/>
      <use name="mycustom_date"/>
</datePatterns>

</datetime>
```

示例 1b：只包含时间戳配置的新的 datetime.xml

除了“示例 1a”部分中显示的示例之外，您还可以在 $SPLUNK_HOME/etc/system/local 中新建 datetime.xml 文件，如下所示：

```
<datetime>

<define name="mycustom_date" extract="day,litmonth,year,">
<text><![CDATA[:\d+\s\w+\s(\d+)\s(\w+)\s(\d+)]]></text>
</define>

<define name="mycustom_time" extract="hour,minute,second,ampm,">
<text><![CDATA[(\d+):(\d+):(\d+)\s(\w+)]]></text>
</define>

 <timePatterns>
       <use name="mycustom_date"/>
 </timePatterns>

 <datePatterns>
       <use name="mycustom_time"/>
 </datePatterns>

 </datetime>
```

示例 2：为您的自定义源类型引用 props.conf 中的新 datetime.xml

完成前面的示例后，您可以引用 props.conf 中您的源类型配置中的自定义 datetime.xml 文件，如下所示：

```
$SPLUNK_HOME/etc/system/local/props.conf


[my_custom_sourcetype]
DATETIME_CONFIG=/etc/system/local/datetime.xml
SHOULD_LINEMERGE=false
NO_BINARY_CHECK=true
```

指定时间戳的时区

如果您从不同的时区为数据新建索引，可以利用时区偏移来检查各时区在搜索时是否关联正确。可以根据事件的主机、来源或来源类型配置时区。

如需修改时间戳提取，您的 Splunk Cloud 架构必须包含重型转发器，并且您必须编辑重型转发器上的 props.conf 文件。请在重型转发器运行的计算机上执行配置。

如果您更改了主机的时区设置，则必须重启 Splunk Enterprise 或转发器，软件才能检测到所做的更改。

有关在 props.conf 中编辑时间戳的一般信息，请参阅“配置时间戳识别”。

如果您使用的是 Splunk Enterprise 而且需要修改时间戳提取，请在您的索引器计算机上执行配置；或者，如果您想要转发数据，请使用重型转发器并在重型转发器运行的计算机上执行配置。

Splunk 软件如何确定时区

为了确定要分配给时间戳的时区，Splunk 软件将使用以下逻辑：

- 若有，请使用原始事件数据中指定的时区（例如 PST、-0800）。

- 如果事件与段落指定的主机、数据来源或来源类型相匹配，请使用 TZ 属性（设置于 props.conf 中）。
- 如果转发器和接收索引器均为 6.0 或更高版本，请使用转发器提供的时区。
- 使用为事件建立索引的主机的时区。

在 props.conf 中指定时区

要配置时区设置，请编辑在 $FORWARDER_HOME/etc/system/local/ 内或 $FORWARDER_HOME/etc/apps/ 中您自己的自定义应用程序目录内的 props.conf 文件。有关配置文件的一般信息，请参阅 Splunk Enterprise *《管理员手册》* 中的“关于配置文件”。

通过将 TZ 属性添加到 props.conf 文件中的适当段落来配置时间区。TZ 属性将识别区域信息 TZ ID。请在主机、来源或来源类型的段落内，把 TZ 属性设置为所需时区的 TZ ID。确保您输入的事件的时区是来自该主机、来源或来源类型的时区。

要查看所有时区 TZ ID 的列表，请参阅 https://en.wikipedia.org/wiki/List_of_tz_database_time_zones。

您没有在 Splunk 平台中为索引器配置时区，但在基本操作系统中为其配置。只要索引器主机系统上的时间设置正确，就能正确地计算事件时区的偏移。

在 props.conf 中指定时区的示例

以下是一些示例，示范如何在 props.conf 中指定时区。

第一个示例中，事件从纽约市（采用美国/东部时区）和加利福尼亚山景城（采用美国/太平洋时区）进入转发器。要正确地处理这两组事件的时间戳，您必须将转发器的 props.conf 的时区设置为分别指定为美国/东部和美国/太平洋。

只要事件来自其名称与正则表达式 nyc.* 相匹配的主机，第一个示例就将该事件的时区设置为美国/东部：

```
[host::nyc*]
TZ = US/Eastern
```

第二个示例则将任何来自 /mnt/ca/... 路径中各来源的事件的时区设置为美国/太平洋：

```
[source::/mnt/ca/...]
TZ = US/Pacific
```

Zoneinfo（TZ）数据库时区值

Zoneinfo 数据库是时区值的公共维护数据库。TZ 数据库的位置和内容取决于操作系统。以下列表显示了几个常用操作系统的 TZ 数据库的位置。

- Splunk 软件的 UNIX 版本依赖于您正在上面运行 UNIX 分布所附带的 TZ 数据库。大多数 UNIX 分布都将该数据库存储在 /usr/share/zoneinfo 目录中。
- Splunk 软件的 Solaris 版本将 TZ 信息存储在 /usr/share/lib/zoneinfo 目录中。
- Splunk 软件的 Windows 版本随附一个 TZ 数据库的副本。

有关 tz 数据库时区的列表，请参阅“tz 数据库时区列表”。

映射从事件数据提取的时区字符串

使用 TZ_ALIAS 属性（位于 props.conf 中）更改 Splunk 解释出现于事件数据中的时区首字母缩写字符串的方式。例如，“EST”默认指东部（美国）标准时间，但您的事件数据可能使用此值改为指定东部（澳大利亚）标准时间。要将“EST”的含义更改为后者，请使用以下语法设置属性：

```
TZ_ALIAS = EST=GMT+10:00
```

然后，如果 Splunk 软件在事件数据中遇到 "EST"，就会将其解释为 "GMT+10:00"，而非默认的 "GMT- 5:00"。

如本例所示，您可以将时区字符串映射到现有字符串并加上其偏移值。也可以将一个 TZ 字符串直接映射到另一个 TZ 字符串。

映射时区字符串时，请确保同时处理时区的冬夏版本。例如，如果您映射东部标准时间，则还必须映射东部夏令时。测试您的软件以查看其产生什么时区字符串。

TZ_ALIAS 的语法为：

```
TZ_ALIAS = <key=value>[,<key=value>]...
```

有关编辑 props.conf 文件的更多信息（包括示例），请参阅 Splunk Enterprise*《管理员手册》*的“配置文件参考”章节中的“props.conf 规范”。

为用户的搜索结果设置时区

添加或编辑用户时，您可以设置一个用户时区。此用户的搜索结果将出现在指定时区中。然而，此设置并不更改实际事件数据，因为这些数据的时区已在索引时间内确定。有关设置此值的信息，请参阅*《确保 Splunk 平台安全》*手册中的“使用 Splunk Web 配置用户”。

调整时间戳识别以获得更佳的索引性能

要加快索引创建速度，可以使用 props.conf 配置文件调整时间戳处理器在事件中的查找深度。甚至可以完全关闭时间戳处理器。

如果您使用的是 Splunk Cloud 而且需要修改时间戳提取，请使用重型转发器并在重型转发器运行的计算机上执行配置。

如果您使用的是 Splunk Enterprise 并且需要修改时间戳提取，请在您的索引器计算机上执行配置。如果您想要转发数据，请使用重型转发器并在重型转发器运行的计算机上执行配置。如果您使用的是 Splunk Cloud 而且需要修改时间戳提取，请使用重型转发器并在重型转发器运行的计算机上执行配置。有关针对时间戳编辑 props.conf 配置文件的信息，请参阅“配置时间戳识别”。

调整时间戳提前量

时间戳提前量确定时间戳处理器在事件中查找时间戳的字符数。调整 MAX_TIMESTAMP_LOOKAHEAD 设置可调整时间戳处理器的查找深度。

时间戳处理器查看事件的默认字符数为 128。您可以将 MAX_TIMESTAMP_LOOKAHEAD 设置设为较低的值以加速索引。如果时间戳始终出现在事件的第一部分，则执行此操作。

本示例在来自数据来源 foo 的事件的头 20 个字符内查找时间戳：

```
[source::foo]
MAX_TIMESTAMP_LOOKAHEAD = 20
...
```

禁用时间戳处理器

可以彻底关闭时间戳处理器以提高索引性能。通过将 DATETIME_CONFIG 设置配置为 NONE，可以关闭匹配到特定主机、来源或来源类型的事件的时间戳处理。当 DATETIME_CONFIG=NONE 时，Splunk 软件不会在事件文本中查找时间戳。相反，Splunk 软件会使用事件的接收时间，或事件通过其输入获得接收的时间。对于基于文件的输入（如 monitor）而言，时间戳来自输入文件的修改时间。

您也可以通过将 DATETIME_CONFIG 设置设为 CURRENT 来提升索引性能，这样做可以在建立索引时将当前的系统时间分配给所有事件。

本示例将关闭来自 foo 来源的事件的时间戳提取：

```
[source::foo]
DATETIME_CONFIG = NONE
...
```

CURRENT 和 NONE 都禁用时间戳识别，因此默认的事件界限检测（BREAK_ONLY_BEFORE_DATE = true）可能无法如预期般运行。使用这些设置时，请指定 SHOULD_LINEMERGE 或 BREAK_ONLY_* 及 MUST_BREAK_* 设置以控制事件合并。

配置索引字段提取

关于索引字段提取

为数据建立索引时，Splunk 软件会把数据流分析为一系列**事件**。在此过程中，软件会将许多**字段**添加到**事件数据**中。这些字段包括 Splunk 软件自动添加的**默认字段**和您指定的任何自定义字段。

向事件添加字段的过程称为**字段提取**。字段提取有两种类型：

- **索引字段的提取**，当字段存储在索引中并成为事件数据的一部分时发生。索引字段有两种类型：
  - Splunk 软件自动添加到每个事件的默认字段。有关更多详情，请参阅“关于默认字段”。
  - 自定义字段，由您指定。有关更多详情，请参阅“在索引时间新建自定义字段”。
- **搜索时间字段提取**，在搜索数据时进行。Splunk 软件会在编制搜索结果时新建那些字段，而且不会把它们存储在索引中。有关此类型字段提取的信息，请参阅《知识管理器手册》中的“关于字段”。

当使用这些字段时，应考虑大多数计算机数据不具有结构或具有不断更改的结构。如果是非结构化数据，为将灵活性最大化，请使用搜索时间字段提取。您定义搜索时间字段提取后对其进行修改也很容易。

其他数据类型可能展示出更加固定的结构，或者结构可能已在数据或文件的事件中定义。您可以对 Splunk 软件进行配置，以便读取此类文件（如逗号分隔值 (CSV) 文件、制表符分隔值 (TSV) 文件、管道符分隔值文件和 JavaScript 对象符号 (JSON) 数据来源）的结构，并在索引时映射字段。要了解这些进程如何运作，请参阅“使用结构化数据从文件中提取字段”。

关于默认字段（host、source、sourcetype 等）

为数据建立索引时，Splunk 软件使用许多字段标记每个事件。这些字段成为索引**事件数据**的一部分。自动添加的字段称为**默认字段**。

默认字段有许多用途：

- 默认字段 `index` 识别事件处于其中的索引。
- 默认字段 `linecount` 描述事件包含的行数。
- 默认字段 `timestamp` 指定事件发生的时间。

为了正确新建事件，Splunk 软件在为数据新建索引时会使用某些字段的值，尤其是 `sourcetype`。新建数据索引之后，您可以在搜索中使用默认字段。

默认字段的完整列表遵循以下：

字段类型	字段列表	描述
内部字段	`_raw, _time, _indextime, _cd`	这些字段中包含 Splunk 软件用于其内部进程的信息。
基本默认字段	`host, index, linecount, punct, source, sourcetype, splunk_server, timestamp`	这些字段提供有关事件的基本信息，例如事件来源、事件包含的数据种类、事件所属的索引、事件所包含的行数以及事件发生时间。
默认日期时间字段	`date_hour, date_mday, date_minute, date_month, date_second, date_wday, date_year, date_zone`	这些字段为事件时间戳提供了额外的可搜索的粒度。 **注意：**只有包含时间戳信息（由各自的系统生成）的事件才会包含 `date_*` 字段。如果某事件包含 `date_*` 字段，则此字段表示直接来自该事件本身的时间/日期值。如果您在索引或输入时指定了时区转换或更改了时间/日期值（例如，将时间戳设置为索引或输入时间），那么这些字段将不再表示事件本身的时间/日期值。

有关搜索角度方面默认字段的信息，请参阅《知识管理器手册》中的“使用默认字段”。

您也可以为索引内的包含事项指定附加的自定义字段。请参阅本章中的“在索引时间新建自定义字段”。

本主题侧重于三个关键默认字段：

- **host**
- **source**
- **sourcetype**

定义 host、source 和 sourcetype

Host、source 和 sourcetype 字段定义如下：

- **host** - 事件的主机值通常是事件来源的网络主机的主机名、IP 地址或完全限定的域名。主机值允许您找到源于特定设备的数据。有关主机的更多信息，请参阅“关于主机”。
- **source** - 事件的来源是事件来源的文件、流或其他输入的名称。若是在文件和目录处受监视的数据，来源的值为完整路径，如 /archive/server1/var/log/messages.0 或 /var/log/。对于基于网络的数据来源，source 的值为协议和端口，例如 UDP:514。
- **sourcetype** - 事件的来源类型是指作为事件来源的数据导入的格式，如 access_combined 或 cisco_syslog。来源类型决定为数据设置格式的方式。有关来源类型的更多信息，请参阅“来源类型为何重要”。

Source 与 sourcetype

数据来源和来源类型都是默认字段，但它们截然不同且易于混淆。

- **Source** 是特定事件来源的文件、流或其他输入的名称。
- **Sourcetype** 决定了 Splunk 软件根据数据性质把传入的数据流处理为单个事件的方式。

来源类型相同的事件也有可能源自不同的来源，例如，若您监视 source=/var/log/messages 并从 udp:514 处接收直接 syslog 输入。如果您搜索 sourcetype=linux_syslog，来自这两个数据来源的事件都将被返回。

在哪些条件下应覆盖 host 和 sourcetype 分配？

Splunk 软件大多数时候都能自动识别正确*且*有用的主机和 sourcetype 值。但是，确实会出现要求您干预此过程并提供覆盖值的情况。

覆盖 host 分配

下列情况下，您可能想要更改默认的 host 分配：

- 批量加载最初由其他主机生成的归档数据，并想让这些事件具备此主机值。
- 从不同的主机转发数据。（除非您另有指定，否则转发器将分配其主机名。）
- 您正在集中式日志服务器环境下工作，这意味着接收自此服务器的所有数据都将具有相同的主机，即使其源于其他位置也是如此。

有关主机的详细信息，请参阅“配置主机值”一章。

覆盖 sourcetype 分配

下列情况下，您可能想要更改默认的 sourcetype 分配：

- Splunk 软件无法正确地为数据自动设置格式，从而导致诸如时间戳出错或事件换行等问题。
- 您希望将来源类型应用于具有特定输入的特定事件，例如源自一组离散主机的事件，或者甚至与特定 IP 地址或 userid 关联的事件。

您还可以执行一些步骤来扩展 Splunk 软件自动识别的来源类型范围，或者通过这些步骤来简单地重命名来源类型。

动态分配默认字段

此功能允许您在 Splunk 软件获取默认字段（也称为“元数据”）时将它们动态地分配给各事件。使用此功能动态指定传入数据的来源类型、主机或来源。此功能主要用于脚本数据 -- **脚本式输入**或由脚本处理的现有文件。

请勿同时使用动态元数据分配和文件监视（tail）输入。有关文件输入的更多信息，请参阅本手册中的“监视文件和目录”。

注意：模块化输入功能已经取代了此 ***SPLUNK*** 标头功能。如果 host、source 和 sourcetype 需要动态生成的值，请考虑编写一个模块化输入。

要使用此功能，请将单个动态输入标头附加到文件中并指定要将值分配到的元数据字段。可用的元数据字段为 sourcetype、

host 和 source。

可以使用此方法分配元数据，而不是编辑 inputs.conf、props.conf 和 transforms.conf。

配置单个输入文件

要将此功能用于现有输入文件，请编辑文件（手动或使用脚本）以添加单个输入标头：

```
***SPLUNK*** <metadata field>=<string><metadata field>=<string> ...
```

1. 把 <metadata field>=<string> 设置为有效的元数据/值对。可以指定多个对。例如，sourcetype=log4j host=swan。
2. 将单个标头添加到文件中的任意位置处。标头之后的所有数据都将附加有您分配的属性和值，直至到达文件的结尾。
3. 将您的文件添加至 $SPLUNK_HOME/var/spool/splunk 或其他任何受 Splunk 监视的目录。

使用脚本进行配置

更为常见的情况是编写脚本，以将输入标头动态添加到传入数据流中。您的脚本也可以根据输入文件的内容动态设置标头。

在索引时间新建自定义字段

通常，您应试图在**搜索时间**内提取字段。但是，有时您可能需要添加到会在**索引时间**应用于事件的自定义**索引字段**集。

例如，某些搜索时间字段提取可能显著影响了搜索性能。例如，若您通常使用诸如 foo!=bar 或 NOT foo=bar 等表达式搜索大型事件集合，且 foo 字段的值几乎*始终*是 bar，也会发生这种情况。

如果搜索时间提取的字段的值时常位于字段之外，那么您也可能希望添加索引字段。例如，若您通常仅搜索 foo=1，但 1 出现在很多*不*具 foo=1 的事件中，您可以把 foo 添加到 Splunk 在索引时间提取的字段列表。

有关创建自定义字段提取的更多信息，请参阅《知识管理器手册》中的“关于字段”。

如果您使用的是 Splunk Cloud 而且想要定义索引时间字段提取，请打开一个 Splunk 支持问题。

除非绝对必要，否则切勿将自定义字段添加到 Splunk 软件在索引时间自动提取并为其建立索引的默认字段组。该组字段包括 timestamp、punct、host、source 和 sourcetype 等字段。添加此字段列表会对索引性能和搜索时间产生负面影响，因为每个索引字段都会增加可搜索索引的大小。索引字段的灵活性也较低 -- 无论何时更改字段集，都必须为整个数据集重建索引。

请参阅“关于默认字段”。

定义其他索引字段

通过编辑 props.conf、transforms.conf 和 fields.conf 定义其他索引字段。

在 $SPLUNK_HOME/etc/system/local/ 内或 $SPLUNK_HOME/etc/apps/ 中您自己的自定义应用程序目录内编辑这些文件。更多有关配置文件的一般信息，请参阅《管理员手册》中的“关于配置文件”。

在分布式环境中将配置更改放置在何处

如果您具有**分布式搜索**部署，处理在搜索节点（索引器）和搜索头之间进行拆分。您必须部署如下更改：

- 把 props.conf 和 transforms.conf 更改部署到每个搜索节点。
- 把 fields.conf 更改部署到搜索头。

如果您正在搜索节点前采用重型转发器，则 props 和 transforms 处理在转发器（而不是在搜索节点）上发生。因此，您必须将 props 和 transforms 更改部署给转发器，而不是搜索节点。

有关 Splunk Enterprise 分布式组件的详细信息，请参阅*《分布式部署手册》*中的“使用 Splunk Enterprise 组件调整部署规模”。

有关配置设置的放置位置详情，请参阅*《管理员手册》*中的“配置参数和数据管道”。

字段名称语法限制

您可以根据以下说明分配字段名称：

- 有效的字段名称字符包括 a-z、A-Z、0-9 或 _。

- 字段名称不能以 0-9 或 _ 开头。Splunk 为其内部变量保留前导下划线。
- 避免分配和任何默认字段名称匹配的字段名称。
- 不要分配包含国际字符的字段名称。

在 transforms.conf 中添加新字段的正则表达式段落

在 transforms.conf 中定义索引时间字段转换时，请遵照以下格式（注意：其中某些属性，如 LOOKAHEAD 和 DEST_KEY，仅在某些使用案例中才需要）：

```
[<unique_transform_stanza_name>]
REGEX = <regular_expression>
FORMAT = <your_custom_field_name>::$1
WRITE_META = [true|false]
DEST_KEY = <KEY>
DEFAULT_VALUE = <string>
SOURCE_KEY = <KEY>
REPEAT_MATCH = [true|false]
LOOKAHEAD = <integer>
```

请注意以下事项：

- <unique_stanza_name> 对所有转换来说都是必须的，REGEX 亦如此。
- REGEX 是处理数据以提取字段的正则表达式。
  - REGEX 中的名称捕获组直接被提取到字段，意味着您不需要为简单的字段提取案例指定 FORMAT。
  - 若 REGEX 可以同时提取字段名称*和*它相应的值，您可以利用以下特殊捕获群组跳过在 FORMAT 属性中指定映射：

    ```
    _KEY_<string>, _VAL_<string>
    ```

- 例如，以下两项是等效的：

  使用 FORMAT：

  ```
  REGEX = ([a-z]+)=([a-z]+)
  FORMAT = $1::$2
  ```

  不使用 FORMAT：

  ```
  REGEX = (?<_KEY_1>[a-z]+)=(?<_VAL_1>[a-z]+)
  ```

- FORMAT 为可选。用于指定您正在提取的字段-值对的格式，包括您想要添加的任何字段名称或值。您不需要指定 FORMAT，前提是您得有包含名称捕获组的简单 REGEX。
- FORMAT 行为根据在搜索时间还是索引时间内提取而不同。
  - 对于索引时间转换，您可以使用 $n 来指定每个 REGEX 匹配的输出（如 $1、$2 等）。
  - 若 REGEX 不含 n 组，则匹配失败。
  - FORMAT 默认为 <unique_transform_stanza_name>::$1。
  - 特殊标识符 $0 代表 DEST_KEY 中的内容（执行 REGEX 之前，且在索引时间字段提取 DEST_KEY 为 _meta 的情况下）。有关更多信息，请参阅下面的“Splunk 如何构建索引字段”。
  - 对于索引时间字段提取，您可以多种方式设置 FORMAT。既可以是如下所示的 <field-name>::<field-value> 设置：

    ```
    FORMAT = field1::$1 field2::$2 （其中 REGEX 为捕获的组 "field1" 和 "field2" 提取字段值）
    ```

  或者：

  ```
  FORMAT = $1::$2 （其中 REGEX 既提取字段名称也提取字段值）
  ```

  不过，您还可以设置新建连接字段的索引时间字段提取：

  ```
  FORMAT = ipaddress::$1.$2.$3.$4
  ```

  使用 FORMAT 新建连接字段时，$ 是唯一的特殊字符，了解这一点十分重要。**仅**在其后跟数字且**仅**此数字适用于现有捕获组时，才将其视为正则表达式捕获组的前缀。

  因此，若您的正则表达式只含一个捕获组且其值为 bar，则

  ```
  FORMAT = foo$1 将生成 foobar
  FORMAT = foo$bar 将生成 foo$bar
  FORMAT = foo$1234 将生成 foo$1234
  FORMAT = foo$1\$2 将生成 foobar\$2
  ```

- WRITE_META = true 将提取的字段名称和字段值写入 Splunk 在其内存储索引字段的 _meta 中。此属性设置对所有索引时间字段提取而言都是必须的，DEST_KEY = _meta 字段除外（请参阅下文中有关 DEST_KEY 的介绍）。
  - 更多有关 _meta 及其在索引字段新建中所扮演角色的信息，请参阅下文中的“Splunk 如何构建索引字段”。

- DEST_KEY 对索引时间字段提取而言是必须的，其中 WRITE_META = false 或未进行设置。该字段指定了 Splunk 发送 REGEX 结果的位置。
  - 对于索引时间搜索，DEST_KEY = _meta，这是 Splunk 存储索引字段的位置。其他可能的 KEY 值请参阅本手册中的 transforms.conf 页面。
  - 有关 _meta 及其在索引字段新建中所扮演角色的更多信息，请参阅下文中的“Splunk 如何构建索引字段”。
  - 使用 DEST_KEY = _meta 时，您还应把 $0 添加到 FORMAT 属性的开头。 $0 代表 DEST_KEY 值（Splunk 执行 REGEX之前，即 _meta 之前）。
  - **注意：** $0 值绝不可能派生*自* REGEX。

- DEFAULT_VALUE 为可选。该属性的值将写入 DEST_KEY，除非 REGEX 失败。
  - 默认为空。

- SOURCE_KEY 为可选。您用它来识别其值需应用到 REGEX 的某 KEY。
  - 根据默认设置，SOURCE_KEY = _raw，意味着其将应用到全体所有事件。
  - 通常与 REPEAT_MATCH 结合使用。
  - 其他可能的 KEY 值请参阅本手册中的 transforms.conf 页面。

- REPEAT_MATCH 为可选。将其设置为 true 即可多次运行 REGEX，运行位置为 SOURCE_KEY。
  - REPEAT_MATCH 从上一匹配的停止位置开始并继续，直至不再找到任何匹配。适用于每个事件的预期字段-值匹配数未知的情况。
  - 默认为 false。

- LOOKAHEAD 为可选。用于指定要在事件中搜索的字符数。
  - 默认为 4096。如果有些事件的行长度超过 4096 个字符，您可能想要增加 LOOKAHEAD 值。
  - 尤其是，如果您需要匹配的文本超过此字符数的限度，则您需要增加此值。
  - 但是，要注意，当扫描更大的文本段时，复杂的正则表达式可具有非常高的成本。当使用多个贪婪（greedy）分支或提前量/推后量时，速度会成二次方下降或下降更快。

注意：有关正则表达式语法和用法的入门，请参阅 Regular-Expressions.info。您可以在搜索中将正则表达式与 rex 搜索命令结合使用，以对表达式进行测试。Splunk 还有一个非常有用的第三方工具列表，可用于编写和测试正则表达式。

注意：正则表达式中捕获组必须识别遵循字段名称语法限制的字段名称。它们只能包含 ASCII 字符（a-z、A-Z、0-9 或_）。国际字符无效。

将新字段链接到 props.conf

要 props.conf，请添加下列各行：

```
[<spec>]
TRANSFORMS-<class> = <unique_stanza_name>
```

请注意以下事项：

- <spec> 可以为：
  - <sourcetype>，事件的 sourcetype。
  - host::<host>，其中 <host> 为事件的主机。
  - source::<source>，其中 <source> 为事件的来源。
  - **注意：**设置 <spec> 时您可以使用正则表达式类型的语法。此外，来源和来源类型段落在匹配时区分大小写，主机段落则不区分。更多信息请参阅 props.conf 规范文件。
- <class> 是唯一的文字字符串，用于标识所提取字段（键）的命名空间。**注意：** <class> 值*不*需要遵循字段名称语法限制（请参阅上文）。您可以使用 a-z、A-Z 和 0-9 以外的其他字符，也可以使用空格。
- <unique_stanza_name> 是 transforms.conf 中的段落名称。

注意：对于索引时间字段提取，props.conf 使用 TRANSFORMS-<class> 而非用于配置搜索时间字段提取的 EXTRACT-<class>。

针对新字段将条目添加到 fields.conf 中

Splunk 平台使用 fields.conf 中的配置来确定哪些自定义字段提取应被视为**索引字段**。

如果新的索引字段来自一个完全或部分由非结构化数据组成的源类型，则可以为每个自定义索引字段创建一个单独的配置。这些配置的段落名称即为字段名称。

如果您要添加字段且这些字段来自完全由结构化数据（如 JSON 格式的数据）组成的源类型，则可以设计使用通配符表达式匹配索引字段集的源类型范围配置。请参阅“使用结构化数据从文件中提取字段”。

针对新的索引字段，把一个条目添加到 fields.conf：

```
[<your_custom_field_name>]
INDEXED=true
```

- <your_custom_field_name> 是您添加至 transforms.conf 中的唯一段落中设置的自定义字段名称。
- 设置 INDEXED=true 即说明已为该字段新建索引。

若在搜索时间提取了某一同名字段，则必须设置该字段的 INDEXED=false。此外，如果事件字段值的提取时间并非索引时间，而*是*搜索时间，则还必须设置 INDEXED_VALUE=false。

例如，假设您在索引时间执行简单的 <field>::1234 提取。这项操作可以运行，但若您同时执行一个基于诸如 A(\d+)B 等正则表达式的搜索时间字段提取，就会出现问题，因为在此情况下，字符串 A1234B 会为 1234 字段生成一个值。该值将在搜索时间找到 1234 事件，导致 Splunk 无法在索引时间使用 <field>::1234 提取定位这些事件。

重新启动 Splunk 使更改生效

直到所有受影响的组件都关闭并重启 Splunk 之后，针对 props.conf 和 transforms.conf 等配置文件所做的更改才会生效。

Splunk 如何构建索引字段

Splunk 通过向 _meta 写入内容来构建索引字段。其工作方式如下：

- _meta 由 transforms.conf （其中包含 DEST_KEY = _meta 或 WRITE_META = true）中的所有匹配转换进行修改。
- 所有匹配转换都可以覆盖 _meta，因此，您可以使用 WRITE_META = true 来附加 _meta。
  - 如您未使用 WRITE_META，则把您的 FORMAT 设置为以 $0 开头。
- _meta 在分析期间彻底构建完成后，Splunk 将以下列方式解释文本：
  - 文本拆分成多个单元；每个单元都以空格分隔。
  - 引号（" "）将字符归组成更大的单元，而不考虑空格。
  - 引号紧前面的反斜杠（ \ ）禁用引号的归组属性。
  - 反斜杠前面的反斜杠禁用此反斜杠。
  - 包含双冒号（::）的文本单元变成提取字段。双冒号左侧的文本成为字段名称，右侧文本则成为值。

具有正则表达式提取值且包含引号的索引字段通常不起作用，反斜杠可能还会导致问题。在搜索时间内提取的字段没有这些限制。

下面是涉及引号和反斜杠以禁用引号和反斜杠的一组索引时间提取的示例：

```
WRITE_META = true
FORMAT = field1::value field2::"value 2" field3::"a field with a \" quotation mark" field4::"a field which
ends with a backslash\\"
```

当 Splunk 新建字段名称时

请牢记：Splunk 新建字段名称之后，会应用字段名称语法限制这些字段名称。

1. 所有非 a-z、A-Z 和 0-9 字符均替换为下划线（_）。

2. 所有前导下划线都会被删除。在 Splunk 中，前导下划线预留用于**内部字段**。

索引时间字段提取示例

下面是索引时间字段提取的配置文件设置的一组示例。

定义新索引字段

此基础示例将新建一个名为 err_code 的索引字段。

transforms.conf

在 transforms.conf 中添加：

```
[netscreen-error]
REGEX =  device_id=\[\w+\](?<err_code>[^:]+)
FORMAT = err_code::"$1"
```

```
WRITE_META = true
```

此段落包含 device_id=（后跟带括号的单词）和以冒号结尾的文本字符串。事件的来源类型为 testlog。

注释：

- FORMAT = 行包含下列值：
  - err_code:: 是字段的名称。
  - $1 指写入到索引中的新字段。它是 REGEX 提取到的值。
- WRITE_META = true 是一个将 FORMAT 的内容写入索引的说明。

props.conf

将下列各行添加到 props.conf：

```
[testlog]
TRANSFORMS-netscreen = netscreen-error
```

fields.conf

将下列各行添加到 fields.conf：

```
[err_code]
INDEXED=true
```

重新启动 Splunk，使配置文件更改生效。

使用一个正则表达式定义两个新索引字段

本示例新建了两个索引字段，分别称为 username 和 login_result。

transforms.conf

在 transforms.conf 中添加：

```
[ftpd-login]
REGEX = Attempt to login by user: (.*): login (.*)\.
FORMAT = username::"$1" login_result::"$2"
WRITE_META = true
```

此段落用于查找文字文本 Attempt to login by user: 并提取一个用户名，该用户名之后依次为冒号、结果和句号。行可能类似如下所示：

```
2008-10-30 14:15:21 mightyhost awesomeftpd INFO Attempt to login by user: root: login FAILED.
```

props.conf

将下列各行添加到 props.conf：

```
[ftpd-log]
TRANSFORMS-login = ftpd-login
```

fields.conf

将下列各行添加到 fields.conf：

```
[username]
INDEXED=true

[login_result]
INDEXED=true
```

重新启动 Splunk，使配置文件更改生效。

在索引时间内连接来自事件段的字段值

本示例将向您展示如何借助 FORMAT 选项，使用索引时间转换来提取并合并一事件的单独段，以新建单一字段。

假设您有以下事件：

```
20100126 08:48:49 781 PACKET 078FCFD0 UDP Rcv 127.0.0.0 8226 R Q [0084 A NOERROR] A (4)www(8)google(3)com(0)
```

现在，您需要执行的操作是提取 (4)www(8)google(3)com(0)，将其作为 dns_requestor 字段的值。不过，您不需要那些无用的括号和数字，您只需要看起来像 www.google.com 的内容。如何实现此目的？

transforms.conf

首先，要在 transforms.conf（名为 dnsRequest）中设置转换：

```
[dnsRequest]
REGEX = UDP[^\(]+\(\d\)(\w+)\(\d\)(\w+)\(\d\)(\w+)
FORMAT = dns_requestor::$1.$2.$3
```

此转换将定义一个名为 dns_requestor 的自定义字段，并使用转换的 REGEX 提取 dns_requestor 值对的三个段，然后再使用 FORMAT 排序这些段，各段间使用句点分隔，看起来就像正常的 URL。

注意：这种将事件段连接成完整字段值的方法*仅*可用于索引时间提取；搜索时间提取有一些实际限制会妨碍此方法。如果发现自己必须以此方式使用 FORMAT，您必须新建索引字段才能达成目标。

props.conf

下一步就是在 props.conf 中定义一个字段提取，该 props.conf 文件引用了 dnsRequest 转换，并将该转换应用到来自 server1 来源类型的事件：

```
[server1]
TRANSFORMS-dnsExtract = dnsRequest
```

fields.conf

最后，您需要在 fields.conf 中输入以下段落：

```
[dns_requestor]
INDEXED = true
```

重新启动 Splunk，使配置文件更改生效。

使用结构化数据从文件中提取字段

许多结构化数据文件（如逗号分隔值（CSV）文件和 Internet 信息服务器（IIS）Web 服务器日志）都在文件标头中具有可在建立索引期间被提取为字段的信息。您可以对 Splunk Enterprise 和 Splunk 通用转发器进行配置，以把这些值自动提取入可供搜索的字段内。例如，CSV 文件以一个标头行开始，其中包含后续行中各值的列标头，例如：

```
host,status,message,"start date"
srv1.splunk.com,error,"No space left on device",2013-06-10T06:35:00
srv2.splunk.com,ok,-,2013-06-11T06:00:00
```

索引字段提取功能支持的输入类型

该功能支持下列几种输入类型：

- 仅限基于文件的输入（例如监视文件、目录或归档。）
- 使用 oneshot 输入类型的输入（或通过 Splunk Web 中的“上载”功能。）

该功能不支持模块化输入、网络输入或任何其他输入类型。

更多有关来源类型和时间戳的信息

- 有关如何在导入结构化数据文件时设置来源类型的信息，请参阅“设置来源类型”页面。
- 有关如何在预览索引结果时调整时间戳的信息，请参阅“调整时间戳和事件换行”。
- 更多有关配置文件的一般信息，请参阅《管理员》手册中的“关于配置文件”。

使用 Splunk Web 从结构化数据文件中提取字段

当您上载或监视机构化数据文件时，Splunk Web 会加载“设置来源类型”页面。此页面允许您预览 Splunk Web 如何为数据建立索引。请参阅“设置来源类型”页面。

1. 从 Splunk Web “添加数据”页面中选择**上载**或**监视**作为您想要添加数据的方法。
2. 指定您想要软件监视的结构化数据文件。Splunk Web 会加载“设置来源类型”页面。它基于其对数据的解读为数据设置来源类型。例如，若您上载一个 CSV 文件，该文件将把来源类型设置为 csv。
3. 审阅页面右侧预览窗格中的事件。根据当前的来源类型为事件设置格式。
4. 如已正确设置事件格式，请单击“下一步”以转到“修改输入设置”页面。否则，请通过修改时间戳、事件换行和分隔的设置来配置事件格式，直到预览的事件符合您的要求。
5. 如果您不想将设置另存为新的来源类型，请返回步骤 4。否则请单击**另存为**按钮，把设置保存为新的来源类型。
6. 请在显示的对话框中键入新来源类型的名称和描述。
7. 请为来源类型选择类别，通过从“类别”下拉列表中选择您所需的类别。
8. 请选择新来源类型要应用到的应用程序上下文，通过从“应用”下拉列表中的条目中进行选择。
9. 请单击“保存”以保存来源类型。
10. 请返回步骤 4 以进入“修改输入设置”页面。

有许多列的结构化数据文件可能无法显示 Splunk Search 中的所有提取字段

如果索引有许多列的结构化数据文件（比如，有 300 列的 CSV 文件），您以后可能遇到以下问题，搜索应用不会返回或显示该文件的所有字段。Splunk 软件已正确索引所有字段后，因为对 Splunk 软件搜索时间如何提取字段的配置设置而出现异常。

在 Splunk 软件显示 Splunk Web 中的字段之前，必须先通过执行搜索时间字段提取来提取这些字段。默认情况下，搜索时间可自动提取的字段数的限制为 100。可以在 $SPLUNK_HOME/etc/system/local 中编辑 limits.conf 文件，将此数字设置得更高，并更改 limit 设置为高于结构化数据文件中列数的数字。

```
[kv]
limit = 300
```

如果您需要处理许多大型 CSV 文件，您可能想要将设置配置为能反映出您期望结构化数据文件有的最大行数相同的列数。

使用配置文件启用自动基于标头的字段提取

您也可使用 inputs.conf 和 props.conf 的组合以从结构化数据文件中提取字段。在 $SPLUNK_HOME/etc/system/local/ 内或 $SPLUNK_HOME/etc/apps/<app_name>/local 中您自己的自定义应用程序目录内编辑这些文件。Inputs.conf 将指定您要监视的文件和文件中所包含的、要应用到事件上的来源类型，props.conf 则将定义来源类型本身。如果使用的是 Splunk Enterprise，您可以在索引器计算机上或运行 Splunk 通用转发器的计算机上编辑设置。必须重启 Splunk Enterprise，对 inputs.conf 和 props.conf 所做的更改才能生效。如果您使用的是 Splunk Cloud 而且想要从结构化数据中配置字段的提取，请使用 Splunk 通用转发器。

结构化数据的 props.conf 属性

如需为包含标头的文件配置字段提取，请在 props.conf 中修改下列属性：有关 props.conf 中的其他属性，请查看 props.conf 规范文件。

属性	描述	默认
INDEXED_EXTRACTIONS = <CSV|W3C|TSV|PSV|JSON|HEC>	指定文件类型以及 Splunk Enterprise 应对文件使用的提取和/或分析方法。 **注意：**若设置 INDEXED_EXTRACTIONS=JSON，请确保您并未同时设置同一来源类型的 KV_MODE = json，因为二者同时设置会导致重复提取 JSON 字段，亦即索引时间提取一次，搜索时间再提取一次。	n/a（未设置）
PREAMBLE_REGEX	一些文件包含序言行。该属性包含一个正则表达式，Splunk 软件使用该表达式忽略任何匹配行。	n/a
FIELD_HEADER_REGEX	指定前缀标头行模式的正则表达式。Splunk 软件会把第一个匹配行分析成标头字段。请注意，实际标头在匹配模式之后开始，该匹配模式不包括在解析的标头字段中。可以在此属性中指定特殊字符。	n/a
FIELD_DELIMITER	指定监视的文件或来源中使用哪些字符分隔字段。可以在此属性中指定特殊字符。	n/a

FIELD_QUOTE	指定指定的文件或来源中用于引号的字符。可以在此属性中指定特殊字符。	n/a
HEADER_FIELD_ACCEPTABLE_SPECIAL_CHARACTERS	指定可以显示在标题字段的特殊字符。如果没有设置，Splunk 软件会用下划线替换既不是字母数字也不是空格的所有字符。如果配置此设置，如果特殊字符和您指定的字符匹配，Splunk 软件不会替换标题字段名称中的特殊字符。例如，如果您将此设置配置为 .，在引入 CSV 期间，此设置不会用下划线替换该字符。	n/a
HEADER_FIELD_DELIMITER	指定标头行中使用哪些字符分隔字段名称。可以在此属性中指定特殊字符。如果未指定 HEADER_FIELD_DELIMITER，则 FIELD_DELIMITER 应用于标头行。	n/a
HEADER_FIELD_QUOTE	指定标头行中使用哪些字符为字段名称两侧加上引号。可以在此属性中指定特殊字符。如果未指定 HEADER_FIELD_QUOTE，则 FIELD_QUOTE 应用于标头行。	n/a
HEADER_FIELD_LINE_NUMBER	指定包含标头字段的文件中的行数。如果设置为 0，Splunk 则尝试自动在文件内查找标头字段。	0
TIMESTAMP_FIELDS = field1,field2,...,fieldn	一些 CSV 和结构化文件的时间戳包含事件中的多个字段，这些字段以分隔符分隔。此属性指示 Splunk 软件指定所有以逗号分隔形式组成时间戳的此类字段。	Splunk Enterprise 尝试自动提取事件的时间戳。
FIELD_NAMES	一些 CSV 和结构化文件的标头可能会缺失。此属性将指定标头字段名称。	n/a
MISSING_VALUE_REGEX	如果在结构化数据文件中查找到与指定的正则表达式相匹配的数据，Splunk 软件会把此行中字段的值视为空值。	n/a

特殊字符或值可用于某些属性

您可在某些属性中使用特殊字符或值，如空格、垂直和水平制表符以及换页等。下表列出了这些字符：

特殊值	Props.conf 表示
换页	\f
空格	space 或 ' '
水平制表符	\t 或 tab
垂直制表符	\v
空格	whitespace
无	none 或 \0
文件分隔符	fs 或 \034
组分隔符	gs 或 \035
记录分隔符	rs 或 \036
单位分隔符	us 或 \037

您可仅为下列属性使用这些特殊字符：

- FIELD_DELIMITER
- FIELD_HEADER_REGEX
- FIELD_QUOTE

编辑配置文件以新建和引用来源类型

要新建并引用新来源类型以提取具有标头的文件：

1. 使用文本编辑器在适当位置打开 props.conf 文件，详情请参阅本主题稍早的“使用配置文件启用自动基于标头的字段提取”所述。如果 props.conf 文件不存在，您必须新建一个。
2. 定义新 Sourcetype，方法是新建一个告诉 Splunk Enterprise 如何使用上面介绍的属性提取文件标头和结构化文件数据的段落。可以在文件中定义任意数量的段落，因而也可以定义任意数量的 Sourcetype。例如：

```
[HeaderFieldsWithFewEmptyFieldNamesWithSpaceDelim]
```

```
FIELD_DELIMITER=,
HEADER_FIELD_DELIMITER=\s
FIELD_QUOTE="
```

3. 保存 props.conf 文件并将其关闭。
4. 如果 inputs.conf 文件不存在，请在相同的目录中新建一个。
5. 打开文件进行编辑。
6. 添加一个表示 Splunk Enterprise 要从中提取标头和结构化数据的文件的段落。可以为要从中提取标头和结构化数据的文件或目录添加任意数量的段落。例如：

```
[monitor:///opt/test/data/StructuredData/HeaderFieldsWithFewEmptyFieldNamesWithSpaceDelim.csv]
sourcetype=HeaderFieldsWithFewEmptyFieldNamesWithSpaceDelim
```

7. 保存 inputs.conf 文件并将其关闭。
8. 重启 Splunk Enterprise 或通用转发器，更改才会生效。

按源类型对索引化的结构化数据字段进行范围划分，以提高搜索性能

从具有固定语义模式（如 JSON）的结构化数据格式中提取字段，往往会产生一组名称相同的字段，这是由于这些格式采用的是分层字段命名系统。当您更新 fields.conf 以将这些字段集添加到自定义**索引字段**集中时，可以使用源类型范围的通配符表达式来覆盖一个配置中的所有自定义索引字段集。

此方法不但可以让您便捷地将提取的结构化数据字段组配置为自定义索引字段，还能让您在搜索这些字段时享有优势。对于使用相同自定义索引字段的搜索，相较于字段分别在 fields.conf（使用 [<field name>] 段落）中配置的搜索相比，字段在 fields.conf 中使用源类型范围通配符表达式配置的搜索其完成速度会更快。

有关 fields.conf 如何用于从非结构化数据中识别单个索引时间字段的更多信息，请参阅“在索引时间新建自定义字段”。

fields.conf 中源类型范围自定义索引字段配置的格式

这是源类型范围自定义索引字段配置的格式。它使用通配符表达式来定位已从结构化数据中提取的一组字段。

```
[sourcetype::<sourcetype>::<wildcard_expression>]
INDEXED=true
```

<sourcetype> 必须仅包含结构化数据，例如 JSON 格式的数据。不要将源类型范围的通配符表达式应用于包含非结构化数据的源类型。

<wildcard_expression> 应该定位一组提取的结构化数据字段。与通配符表达式匹配的所有字段名称都以指定的源类型为范围。

您可以使用通配符表达式来分配字段名称，如下所示：

- 有效的字段名称字符包括 **a-z**、**A-Z**、**0-9**、.、: 和 _。
- 字段名称不能以 **0-9** 或 _ 开头。Splunk 为其内部变量保留前导下划线。
- 不要分配包含国际字符的字段名称。
- 不要分配和任何**默认字段名称**匹配的字段名称。
- 不要创建名称与在搜索时提取的字段名称冲突的索引字段。

设置 INDEXED=true，以表示已为配置匹配的字段集建立索引。

源类型范围的索引字段配置要求在 limits.conf 中设置 indexed_fields_expansion=t。

源类型范围的通配符表达式示例

此配置会为从 splunk_resource_usage 源类型提取的所有以字符串 data. 开头的字段建立索引。

```
[sourcetype::splunk_resource_usage::data.*]
INDEXED=True
```

使用此配置时，如果您仅为 [data.search_props.delta_scan_count] 设置了 fields.conf 配置，以下搜索的运行速度更快。

```
index=_introspection data.search_props.delta_scan_count=0
```

转发从结构化数据文件中提取的字段

把从结构化数据文件中提取到的字段转发到重型转发器或通用转发器上。

要转发从结构化数据文件中提取的字段：

1. 对监视文件的 Splunk 实例进行配置，以把数据转发到重型转发器或通用转发器上。
2. 对接收实例进行配置：
3. 为了正确处理数据的事件换行和时间戳，请在监视实例上配置 props.conf 和 inputs.conf。您可以通过以下两种方法完成此操作。

    - 要使用 Splunk Web，请遵循本主题稍早的“使用 Splunk Web 从结构化数据文件中提取字段”一文所述。
    - 要使用配置文件，请遵循本主题稍早的“编辑配置文件以新建和引用 sourcetype”一文中所述。

4. （可选）如您需要在为数据新建索引前以任何方式转换该数据，请编辑 transforms.conf。
5. 重启接收实例。
6. 重启监视实例。
7. 在接收实例上使用“搜索”应用确认字段已从结构化数据文件中正确地提取出，并正确地建立了索引。

从结构化数据文件中提取字段的注意事项

Splunk 软件不会分析已转发到索引器上的结构化数据

您向索引器转发结构化数据时，即便您已经在该索引器上配置了 props.conf（使用 INDEXED_EXTRACTIONS 进行配置），Splunk 软件也不会在数据达到索引器时分析数据。转发的数据会在索引器上跳过下列管道，因此在该索引器上排除了这些数据的任何分析：

- parsing
- merging
- typing

转发数据在到达索引器时必须已经过分析。

必须在转发器上配置已转发的结构化数据的字段提取设置

如果要把您从结构化数据文件中提取到的字段转发到另一个 Splunk 实例，您必须配置 props.conf 设置，该设置在发送数据的转发器上定义字段提取。这包括 INDEXED_EXTRACTIONS 的配置，以及其他任何分析、筛选、匿名及路由规则。由于已转发的数据必须到达已解析的索引器，因此在为数据建立索引的实例上执行这些操作将无效。

当您使用 Splunk Web 修改事件换行和时间戳设置时，它将把您做的所有更改记录为 props.conf 的一个段落。您可在“设置来源类型”页面上的“高级”选项卡中找到这些设置。

使用“高级”选项卡中的“复制到剪切板”链接，把针对 props.conf 所做的更改复制到系统剪切板。之后，您可以将该段落粘贴到文本编辑器内的 props.conf 中，该文本编辑器位于监视和转发类似文件的 Splunk 实例上。

仅为包含数据的标头字段建立索引

从结构化数据文件中提取标头字段时，Splunk 软件仅提取至少一个行中存在数据的那些字段。如果标头字段的所有行中均无数据，Splunk 软件将跳过此字段（亦即，不为该字段建立索引）。例如，采用下列 csv 文件：

```
header1,header2,header3,header4,header5
one,1,won,,111
two,2,too,,222
three,3,thri,,333
four,4,fore,,444
five,5,faiv,,555
```

读取该文件时，Splunk 软件注意到 header4 列中的各行均为空白，因此不会为该列中的标头字段或任何行建立索引。这意味着无法在索引中搜索 header4 及行中的任何数据。

但是，如果 header4 字段包含带空字符串的行（例如 “”），该字段和下面的所有行都会建立索引。

在允许标题字段使用特殊字符时要小心

HEADER_FIELD_ACCEPTABLE_SPECIAL_CHARACTERS 设置旨在管理列标题有 . 或 : 这样的字符的情况。如果您没有使用此设置，在引入期间，Splunk 软件会用下划线替换这些字符。

不支持在文件中间重命名标头字段

一些软件（如 Internet Information Server）支持在文件中间重命名标头字段。但 Splunk 软件无法识别这类更改。如果您尝试为在文件中间重命名了标头字段的文件建立索引，Splunk 软件不会为重命名的标头字段建立索引。

示例配置和数据文件

下面是示例 inputs.conf 和 props.conf，可使您了解如何使用文件标头提取属性。

如需在本地提取数据，请编辑 inputs.conf 和 props.conf 来为结构化数据文件定义输入和 Sourcetype，并使用上面介绍的属性指示 Splunk 软件如何处理文件。要将此数据转发到另一 Splunk 实例，请在转发实例上编辑 inputs.conf 和 props.conf，在接收实例上编辑 props.conf。

Inputs.conf

```
[monitor:///opt/test/data/StructuredData/CSVWithFewHeaderFieldsWithoutAnyValues.csv]
sourcetype=CSVWithFewHeaderFieldsWithoutAnyValues

[monitor:///opt/test/data/StructuredData/VeryLargeCSVFile.csv]
sourcetype=VeryLargeCSVFile

[monitor:///opt/test/data/StructuredData/UselessLongHeaderToBeIgnored.log]
sourcetype=UselessLongHeaderToBeIgnored

[monitor:///opt/test/data/StructuredData/HeaderFieldsWithFewEmptyFieldNamesWithSpaceDelim.csv]
sourcetype=HeaderFieldsWithFewEmptyFieldNamesWithSpaceDelim

[monitor:///opt/test/data/FieldHeaderRegex.log]
sourcetype=ExtractCorrectHeaders
```

Props.conf

```
[CSVWithFewHeaderFieldsWithoutAnyValues]
FIELD_DELIMITER=,

[VeryLargeCSVFile]
FIELD_DELIMITER=,

[UselessLongHeaderToBeIgnored]
HEADER_FIELD_LINE_NUMBER=35
TIMESTAMP_FIELDS=Date,Time,TimeZone
FIELD_DELIMITER=\s
FIELD_QUOTE="

[HeaderFieldsWithFewEmptyFieldNamesWithSpaceDelim]
FIELD_DELIMITER=,
HEADER_FIELD_DELIMITER=\s
FIELD_QUOTE="

[ExtractCorrectHeaders]
FIELD_HEADER_REGEX=Ignore_This_Stuff:\s(.*)
FIELD_DELIMITER=,
```

示例文件

下面是在上述 inputs.conf 和 props.conf 示例中引用的文件的片段，可使您了解文件的结构。

您可能需要向右滚动一点才能看到所有内容。

CSVWithFewHeaderFieldsWithoutAnyValues.csv

```
vqmcallhistoryid,serialnumber,vqmavgjbenvdelay,vqmavgjbenvnegdelta,vqmavgjbenvposdelta,vqmbitrate,vqmburstcoun
t,vqmburstlengthavgms,vqmburstlengthavgpkts,vqmburstlostrateavg,vqmburstrvaluelq,vqmccmid,vqmcalldurationms,vq
mcallid,vqmcallstart,vqmdegradationdelay,vqmdegradationdiscard,vqmdegradationecholevel,vqmdegradationloss,vqmd
egradationnoiselevel,vqmdegradationrecency,vqmdegradationsignallevel,vqmdegradationvocoder,vqmdelayavgmsec,vqm
delaycurrentmsec,vqmdelaydecreasecount,vqmdelayincreasecount,vqmdelaymaxmsec,vqmdelayminmsec,vqmdestifc,vqmdes
tintname,vqmdiscardrateavg,vqmdiscards,vqmdscp,vqmduplicatepkts,vqmearlypeakjitterms,vqmearlypkts,vqmearlythre
sholdms,vqmearlythresholdpc,vqmearlytotalcount,vqmearlyunderthresholdcount,vqmexcessburst,vqmexcessgap,vqmexte
```

```
rnalrvaluecqin,vqmexternalrvaluecqout,vqmexternalrvaluelqin,vqmexternalrvaluelqout,vqmfrom,vqmgapcount,vqmgapl
engthavgms,vqmgaplengthavgpkts,vqmgaplostrateavg,vqmgaprvalue,vqmjbmaxdelay,vqmjbmindelay,vqmjbnomdelay,vqmjbt
ype,vqmjbresetcount,vqmlatepeakjitterms,vqmlatepkts,vqmlatethresholdms,vqmlatethresholdpc,vqmlatetotalcount,vq
mlateunderthreshold,vqmlocaldelayaveragems,vqmlocaldelaymaximumms,vqmlocaldelayminimumms,vqmloss,vqmlossrateav
g,vqmmaxjbenvnegdelta,vqmmaxjbenvposdelta,vqmmeanpdvabsmaxavg,vqmmeanpdvavg,vqmmeanpdvmaxavg,vqmmeanpdvtrue,vq
mminjbenvnegdelta,vqmminjbenvposdelta,vqmmoscq,vqmmoscqfixed,vqmmoslq,vqmmoslqfixed,vqmmosnominal,vqmmosnomina
lfixed,vqmmospq,vqmmospqfixed,vqmnetworklossrateavg,vqmonewaydelayaverage,vqmonewaydelayinstant,vqmonewaydelay
maximum,vqmoriginationdelayaverage,vqmoriginationdelayinstant,vqmoriginationdelaymaximum,vqmoutoforder,vqmover
rundiscardpkts,vqmppdvms,vqmpdvaveragems,vqmpdvmaximumms,vqmpktsrcvd,vqmrvaluecq,vqmrvalueg107,vqmrvaluelq,vqm
rvaluenominal,vqmreliabilityindex,vqmresynccount,vqmrtdelayaverage,vqmrtdelayinstant,vqmrtdelaymaximum,vqmrtpd
esturi,vqmrtpdestinationip,vqmrtpdestinationport,vqmrtpssrc,vqmrtpsourceip,vqmrtpsourceport,vqmrtpsourceuri,vq
msourceintname,vqmsrcifc,vqmstreamquality,vqmterminatedelayaverage,vqmterminatedelayinstant,vqmterminatedelaym
aximum,vqmthroughputindex,vqmto,vqmunderrundiscardpkts,vqmvocoderclass,vqmvocodertype,created,modified
99152,CFG0730084,-3,-2,356,64000,1,280,14,14.29,36,3499,201000,BW163736844290611-173170743@10.253.143.13,2011-06-29
12:37:37.292,0,4.68,1.43,0.19,0,0,0,0,52,60,15,17,60,10,0,Loopback,0.48,48,46,0,30,1334,10,99.55,10008,9962,0,
0,,,,,6096147095,2,100590,5029,0.48,87,200,10,50,1,625,487.5,8767,50,99.58,93,50,,518,,2,0.5, -
60,975,488,179,192,999.3,0,0,4.07,,4.12,,4.2,,4.03,,0.02,63,76,76,,,,43,0,6.8,0,520,10054,87,87,89,93,9,79,12
,12,12,6096147089,10.255.43.12,10010,706222942,10.253.136.231,25110,6096147095,eth
0/1,2,0,54,80,80,18500,6096147089,48,1,0,2011-06-29 12:41:47.303,2011-06-29 12:41:47.303
99154,CFG0730084,-3,-1,251,64000,4,195,9,20.52,28,3494,359000,BW163502270290611594566299@10.253.143.13,2011-06-29
12:35:02.324,0,2.88,1.11,3.44,0,0,0,0,40,40,26,24,50,10,0,Loopback,0.31,54,46,0,31,2455,10,99.8,17769,17732,0,
0,,,,,6096147095,5,71556,3577,0.62,87,200,10,50,1,1120,496.5,15437,50,99.73,123,74,,529,,65,0.67, -
62,993,496.5,126,139,3404.7,0,0,4.04,,4.07,,4.2,,3.94,,0.36,58,64,69,,,,49,0,286,0,529,17839,86,86,87,93,9,13
7,8,8,8,6096147089,10.255.43.12,10000,536353626,10.253.136.231,25096,6096147095,eth
0/1,2,0,48,60,70,30400,6096147089,54,1,0,2011-06-29 12:41:47.342,2011-06-29 12:41:47.342
```

VeryLargeCSVFile.csv

```
IncidntNum,Category,Descript,DayOfWeek,Date,Time,PdDistrict,Resolution,Location,X,Y
030203898,FRAUD,"FORGERY, CREDIT CARD",Tuesday,02/18/2003,16:30,NORTHERN,NONE,2800 Block of VAN NESS AV,-
122.424612993055,37.8014488257836
000038261,WARRANTS,WARRANT ARREST,Thursday,04/17/2003,22:45,NORTHERN,"ARREST, BOOKED",POLK ST / SUTTER ST,-
122.420120319211,37.7877570602182
030203901,LARCENY/THEFT,GRAND THEFT PICKPOCKET,Tuesday,02/18/2003,16:05,NORTHERN,NONE,VAN NESS AV / MCALLISTER ST,-
122.42025048261,37.7800745746105
030203923,DRUG/NARCOTIC,SALE OF BASE/ROCK COCAINE,Tuesday,02/18/2003,17:00,BAYVIEW,"ARREST, BOOKED",1600 Block of KIRKWOOD AV,-
122.390718076188,37.7385560584619
030203923,OTHER OFFENSES,CONSPIRACY,Tuesday,02/18/2003,17:00,BAYVIEW,"ARREST, BOOKED",1600 Block of KIRKWOOD AV,-
122.390718076188,37.7385560584619
030203923,OTHER OFFENSES,PROBATION VIOLATION,Tuesday,02/18/2003,17:00,BAYVIEW,"ARREST, BOOKED",1600 Block of KIRKWOOD AV,-
122.390718076188,37.7385560584619
```

UselessLongHeaderToBeIgnored.log

```
************ Start Display Current Environment ************
WebSphere Platform 6.1 [ND 6.1.0.21 cf210844.13]  running with process name sammys_cell_A\fsgwws189Node_A\sammys_A_c01_s189_m06
and process id 17904
Detailed IFix information: ID: 6.1.0-WS-WASSDK-AixPPC32-FP0000021  BuildVrsn: null  Desc: Software Developer Kit 6.1.0.21
ID: 6.1.0-WS-WAS-AixPPC32-FP0000021  BuildVrsn: null  Desc: WebSphere Application Server 6.1.0.21
ID: 6.1.0-WS-WASSDK-AixPPC32-FP0000019  BuildVrsn: null  Desc: Software Developer Kit 6.1.0.19
ID: 6.1.0-WS-WAS-AixPPC32-FP0000019  BuildVrsn: null  Desc: WebSphere Application Server 6.1.0.19
ID: sdk.FP61021  BuildVrsn: null  Desc: WebSphere Application Server 6.1.0.21
ID: sdk.FP61019  BuildVrsn: null  Desc: WebSphere Application Server 6.1.0.19
ID: was.embed.common.FP61021  BuildVrsn: null  Desc: WebSphere Application Server 6.1.0.21
ID: was.embed.FP61021  BuildVrsn: null  Desc: WebSphere Application Server 6.1.0.21
```

HeaderFieldsWithFewEmptyFieldNamesWithSpaceDelim.csv

```
"Field 1"  "Field 3" "Field 4"  "Field 6"
Value11,Value12,Value13,Value14,Value15,Value16
Value21,Value22,Value23,Value24,Value25
Value31,Value32,Value33,Value34,Value35, Value36
```

FieldHeaderRegex.log

```
Garbage
Garbage
Garbage
Ignore_This_Stuff: Actual_Header1 Actual_Header2
```

问答

有什么问题吗？请访问 Splunk Answers，查看 Splunk 社区有哪些与提取字段相关的问题和答案。

用引入时 eval 处理事件

引入时 eval 是一种**转换**类型，可评估索引时间的表达式。引入时 eval 可提供搜索时 eval 提供的大部分功能。主要不同在于引入时 eval 是在索引之前处理事件数据，评估所产生的新字段和值会发送到索引器。

更多有关搜索时 eval 表达式的信息，请参阅*搜索手册*中的“使用 eval 命令和函数”。

您可以使用引入时 eval 表达式新建字段，在传入数据上执行多种操作，包括数学、统计和加密函数。请参阅*搜索参考*中的评估函数。

为什么使用引入时 eval？

引入时 eval 可提供单独使用正则表达式很难或不可能实现的引入时转换（如使指标数据规范化）的备选方法。请参阅*《指标》*手册中的“以 log-to-metrics 对话为目标的示例”。

通过引入时 eval ，您可以设置引入时查找，这使您可以在引入数据时以及在建立索引之前使用查找字段来丰富数据。如果您对几乎所有事件都执行了某些查找，则可能需要将它们设置为引入时查找。请参阅“通过引入时查找减少查找开销”。

引入时 eval 还会允许您直接控制索引时字段。例如，您可使用引入时 eval 精确控制索引时字段如何存储在 Splunk Enterprise 索引的**原始数据日志**中。有关更多信息，请参阅*管理索引器和索引器群集*中的“索引器如何存储索引”。

引入时 eval 语法和使用情况

引入时 eval 会对搜索时 | eval 命令采用类似格式。有关更多信息，请参阅*《搜索参考》*中的 eval。

transforms.conf 中的引入时 eval 段落包含一个 INGEST_EVAL 表达式。例如：

```
[eval1]
INGEST_EVAL= field3=length (_raw) *2
```

您还可以链接多个逗号隔开的 INGEST_EVAL 表达式，例如：

```
[eval2]
INGEST_EVAL= field4=_time, field5=field4+1
```

有关 INGEST_EVAL 的详细使用信息和示例，请参阅 transforms.conf。

使用 props.conf 中 EVAL-fieldname 设置的搜索时计算字段不可用。

使用引入时 eval 进行的索引前数据处理可能会影响性能。

配置引入时 eval 转换

您可以将包含转换段落的 transforms.conf 文件与引用该文件的 props.conf 文件结合使用，用配置其他索引时转换的相同方法配置基于 eval 的转换。您还必须在搜索头上配置 fields.conf 文件，以启用最近索引的 eval 字段搜索。

要用引入时 eval 处理事件数据，请配置以下文件：

配置 transforms.conf

要配置用于引入时 eval 的 transforms.conf，请遵循以下步骤：

1. 新建一个 transforms.conf 文件（在 $SPLUNK_HOME/etc/system/local 目录中）。
2. 添加指定 INGEST_EVAL 表达式的引入时 eval 段落。例如，以下 INGEST_EVAL 表达式会新建一个名为 eval_user 的新字段，并用 username 字段中的值的小写形式填充字段：

```
[myeval]
INGEST_EVAL = eval_user=lower(username)
```

配置 props.conf

要配置用于引入时 eval 的 props.conf，请遵循以下步骤：

1. 新建一个 props.conf（在 $SPLUNK_HOME/etc/system/local 目录中）。
2. 添加指定您想要处理的数据（如 <my_sourcetype>）的段落，并引用 transforms.conf 中的引入时 eval 段落。例如：

```
[my_sourcetype]
TRANSFORMS = myeval
```

引入评估转换要求 props.conf 中有源类型段落

您可以按照任何顺序结合使用 props.conf 中的基于 eval 的转换和基于 regex 的转换。您列出转换的顺序决定转换相对于 tranforms.conf 中其他段落的运行时间。例如，TRANSFORMS = eval1,regex1,eval2,regex2 以该特定顺序运行四种不同的 transforms.conf 段落。

配置 fields.conf

要配置 fields.conf 启用引入时 eval 字段搜索，请执行以下操作：

1. 在搜索头上，在$SPLUNK_HOME/etc/system/local 目录中新建 fields.conf 文件。
2. 添加引用 INGEST_EVAL 表达式新建的索引字段的段落，如下所示：

```
[eval_user]
INDEXED = True
```

有关如何配置索引时转换的更多信息，请参阅“定义其他索引字段”。

示例

关于 eval 表达式的基本和扩展示例，请参阅*搜索引用*中的 eval。

通过引入时查找减少查找开销

如果您通常会把特定查找应用于所有传入事件，可以考虑在引入时使用引入时查找对它们进行处理。要做到这一点，您可以配置一个使用 lookup() 评估函数的引入时 eval，在将查找表中的值添加到索引之前先把这些值添加到事件中。

引入时查找的前提条件

本部分介绍在尝试配置引入时查找之前应该了解的事项。

lookup() eval 函数

引入时查找依赖于 eval 的lookup() 函数。

lookup() 函数以与 lookup 命令类似的方式执行 CSV 查找，两者的不同之处在于，前者针对 eval 利用 JSON 函数（如 json_object），以 JSON 对象的形式应用输出字段和值。lookup() 函数的输出是一个 JSON 对象。

有关 lookup() 函数和 JSON 函数的更多信息，请参阅*《搜索参考》*中的以下主题：

- 对比和条件函数
- JSON 函数

Filesystem 访问

您必须具有 Splunk 文件系统访问权限才能设置引入时 eval 表达式。无法在应用程序或用户的上下文中配置引入时 eval 表达式。

- 请参阅 Splunk Enterprise*《管理员手册》*中的“如何编辑配置文件”了解具体步骤。
- 您可以有几个具有相同名称的配置文件，分散在默认目录、本地目录和应用目录中。请参阅 Splunk Enterprise*《管理员手册》*中“在何处放置（或查找）已修改的配置文件”。
- 如果您的 Splunk 部署使用分布式搜索，请使用配置软件包 **deployer** 将 CSV 查找文件和配置推送到群集成员和对等节点。请参阅*《分布式搜索》*中的“使用 deployer 分布应用和配置更改”。

不要更改或复制默认目录中的配置文件。默认目录中的文件必须保持原样并位于其原始位置。更改本地目录中的文件。

引入时 eval 表达式

引入时查找是一种引入时 eval 表达式。您可以使用引入时 eval 表达式新建字段，在传入数据上执行多种操作，包括数学、统计和加密函数。有关概述，请参阅“用引入时 eval 处理事件”。

在 $SPLUNK_HOME/etc/system/local 中使用 INGEST_EVAL 设置在 transforms.conf 中配置引入时 eval 表达式。

CSV 查找文件和定义

引入时查找是 CSV 文件查找，因此使用 CSV 文件作为其查找表。引入时查找预期 CSV 查找文件存储于 $SPLUNK_HOME/etc/system/lookups。如果有 Splunk 的单个实例，则可以手动将文件加载到此位置。如果您有分布式搜索环境，则可以使用 deployer 来更新对等节点上的文件。

您可以选择指定 CSV 查找定义而不是 CSV 查找文件。CSV 查找定义包括对 CSV 查找文件的引用。CSV 查找定义还可以包括过滤器、字段和值匹配规则及其他设置。如果您指定 CSV 查找定义，则必须在 $SPLUNK_HOME/etc/system/local 的 transforms.conf 段落中对它进行配置。请参阅*《知识管理器手册》*中的“配置 CSV l查找”。

引入时查找语法和使用情况

引入时查找的语法与 lookup 命令的语法相似，并且与自动搜索时查找的配置语法相似。请参阅*《知识管理器手册》*中的“设为自动查找。

引入时查找通过 INGEST_EVAL 表达式运行 lookup() 函数。其语法如下所示：

```
[lookup1]
INGEST_EVAL= <string>=lookup("<lookup_table>", json_object("<input_field>",<match_field>,...),
json_array("<output_field>",...))
```

如果为 <lookup_table> 提供的第一个带引号的字符串缺少“.csv”文件描述符，则 Splunk 软件会假定它是 CSV 查找定义的名称。

如果您希望与定义关联的各种设置应用于引入时查找，请指定查找定义。定义可以包括过滤器、过滤器、字段和值匹配规则等。

lookup() 函数可以使用多个 <input_field>/<match_field> 对来标识事件，并且可以将多个 <output_field> 值应用于这些事件。以下是带有多个输入、匹配项和输出的有效 lookup() 语法的示例。

```
[lookup1]
INGEST_EVAL= <string>=lookup("<lookup_table>", json_object("<input_field1>", <match_field1>, "<input_field2>", <match_field2>),
json_array("<output_field1>", "<output_field2>", "<output_field3>")
```

您可以设置 INGEST_EVAL 表达式，这些表达式将 lookup() 函数嵌套在另一个 eval 函数中。此示例使用 json_extract 函数从由 lookup() 函数生成的 JSON 对象中提取字段值：

```
[lookup-extract]
INGEST_EVAL= status_detail=json_extract(lookup("http_status.csv", json_object("status", status),
json_array("status_description")), "status_description")
```

这会导致在引入时向事件中添加字段值对，例如 status_detail=Created 和 status_detail=Not Found，具体取决于这些事件中 status 字段的值。

Limits.conf 设置

两个 limits.conf 设置使您可以在整个部署范围内控制引入时查找的使用情况。

ingest_max_memtable_bytes

ingest_max_memtable_bytes 设置为用于引入时查找处理的 CSV 查找表的大小设置了上限。这样可以确保引入时查找文件永远不会占用过多的内存空间。默认情况下，当与 INGEST_EVAL 一起使用时，大于 10mb 的查找表就不能用于 lookup() eval 函数。当引入时查找使用的查找表的大小超过 ingest_max_memtable_bytes 设置时，任何依赖于它的查找都会失败。重新启动 Splunk 时，splunkd.log 中会显示指示其失败的错误消息。

ingest_max_memtable_bytes 设置有意与 max_memtable_bytes 分开，后者是搜索时查找的类似设置。它将设置为较低的值，以便索

引管道在内存和 CPU 使用方面不会受到太大影响。

ingest_lookup_refresh_period_secs

如果您有基于 CSV 查找表的引入时查找，而且该查找表经常更改，则可以调整 `ingest_lookup_refresh_period_secs` 设置以确保捕获这些更改。默认情况下，此设置可确保引入时会与 `lookup()` 函数一起使用的内存中查找表每 60 秒刷新一次。

配置主机值

关于主机

事件的**主机**字段值就是事件源自其中的物理设备的名称。由于主机字段值是一个**默认字段**，也就是说 Splunk® Enterprise 会向其建立索引的每个事件分配一个主机，因此可以使用该字段搜索特定主机已经生成的所有事件。

主机值通常是事件所源自的网络计算机的主机名、IP 地址或完全限定的域名。

Splunk Cloud 和 Splunk Enterprise 都在索引时间分配主机名，但您可以直接在 Splunk Enterprise 实例上配置主机分配，您必须在 Splunk Cloud 的通用或重型转发器上执行此配置。

Splunk® Enterprise 如何确定主机值

Splunk® Enterprise 会按照以下顺序检查设置并使用遇到的第一个主机设置，由此来为每个事件分配主机值：

1. 您在 transforms.conf 配置文件中指定的事件特定的任何主机分配。对于 Splunk Cloud，您必须使用重型转发器通过事件分配主机名。
2. （如有）新建了事件的输入的默认主机值。
3. 最初引入数据的索引器或转发器的默认主机值。

默认主机值

如果没有为来源指定任何其他主机规则，Splunk® Enterprise 将为主机字段分配一个默认值，该值应用于从任何输入进入实例的所有数据。默认主机值是最初引入数据的索引器或转发器的主机名或 IP 地址。当 Splunk Enterprise 或（在 Splunk Cloud 的情况下）重型转发器在发生事件的服务器上运行时，行为属于正常情况，无需进行手动干预。

更多信息请参阅“设置 Splunk 平台实例的默认主机”。

文件或目录输入的默认主机

如果您在中央日志归档中运行 Splunk Cloud、Splunk Enterprise，或使用了从您环境中其他计算机转发来的文件，您可能需要覆盖来自特定输入的事件的默认主机分配。

有两种方法可用于向接收自特定输入的数据分配主机值：您既可以为来自特定输入的所有数据定义一个静态主机值，也可以让 Splunk 平台为数据来源的部分路径或文件名动态分配主机值。采用将每个主机的日志归档分隔为不同子目录的目录结构时，后一种方法非常有用。

更多信息请参阅“设置文件或目录输入的默认主机”。

事件特定的分配

一些情况下需要通过检查事件数据来分配主机值。例如，如果中央日志主机向 Splunk Enterprise 部署发送事件，则可能有多个计算机向该主日志服务器提供数据。要确保每个事件都具有其来源服务器的主机值，您需要使用事件的数据来确定主机值。

更多信息请参阅“基于事件数据设置主机值”。

处理分配不当的主机值

如果事件数据使用了错误的主机值标记，您可以通过多种方法解决或处理此问题。有关最常见情况的修复，请参阅“建立索引后更改主机值”。

标记主机值

您可以标记主机值来辅助稳健搜索的执行。标记允许您将各组主机群集到有用的可搜索类别中。

详情请参阅 Splunk Enterprise *《知识管理器手册》* 中的“关于标记和别名”。

设置 Splunk 平台实例的默认主机

事件主机值是网络中事件源自其中的物理设备的 IP 地址、主机名或完全限定的域名。由于 Splunk 软件为事件建立索引，而且在索引时间为这些事件都分配了一个 host 值，因此主机值搜索可以让您轻松地找到源自特定设备的数据。

您无法更改 Splunk Cloud 上的默认主机名。相反，您可以根据输入、来源和来源类型分配主机名。从特定设备查找数据仅在

Splunk Enterprise 上可用。

默认主机分配

如果您尚未针对某数据来源指定其他主机规则，则事件的默认主机值为计算机的主机名或 IP 地址，该计算机运行引入事件数据的 Splunk 平台实例。如果事件源自于 Splunk 平台实例本身，则该主机分配正确，无需更改任何设置。但是，如果您的数据都转发自不同主机或者要批量加载归档数据，最好更改该数据的默认主机值。

要设置主机字段的默认值，可以使用 Splunk Web 或编辑 inputs.conf 配置文件。

使用 Splunk Web 设置默认主机值

按照以下步骤将为传入 Splunk 实例的所有事件设置主机字段的默认值。您可以覆盖各个来源或事件的值。

1. 在 Splunk Web 中单击**设置 > 服务器设置**。
2. 在“设置”页面上单击**常规设置**。
3. 在“常规设置”页面上，向下滚动至**索引设置**部分并更改**默认主机名**。
4. 保存更改。

使用 inputs.conf 设置默认主机值

默认主机分配于安装期间在 inputs.conf 配置文件中设置。您可以通过编辑 $SPLUNK_HOME/etc/system/local/ 目录或 $SPLUNK_HOME/etc/apps/ 中您自己的自定义应用程序目录中的该文件来修改主机值。有关更多信息，请参阅“inputs.conf”。

主机分配设置出现在文件的 [default] 段落中。

以下是 inputs.conf 文件中默认主机分配的格式：

```
[default]
host = <string>
```

将 <string> 值设置为您选择的默认主机值。 <string> 默认为生成数据的主机的 IP 地址或域名。请勿在 <string> 值上加引号。例如，输入 host=foo，而不是 host="foo"。

编辑 inputs.conf 文件之后，Splunk 平台实例必须重新启动，这样您所做的更改才能生效。

根据默认设置，host 设置被配置为变量 $decideOnStartup，意味着该设置被设置为运行 splunkd 的计算机的主机名称。守护程序每次启动时都将重新解释该值。

覆盖接收自特定输入的数据的默认主机值

如果对中央日志归档运行 Splunk Enterprise，或者在环境中使用其他主机转发而来的文件，则可能需要覆盖来自特定输入的事件的默认主机分配。

有两种方法可用于向接收自特定输入的数据分配主机值：您既可以为来自特定输入的所有数据定义一个静态主机值，也可以为数据来源的部分路径或文件名动态分配主机值。采用将每个主机的日志归档分隔为不同子目录的目录结构时，后一种方法非常有用。

更多信息请参阅“设置文件或目录输入的默认主机”。

使用事件数据覆盖默认主机值

一些情况下需要通过检查事件数据来分配主机值。例如，若您用中央日志主机向您的 Splunk 平台部署发送事件，您可能有多个主机服务器向该主日志服务器提供数据。要确保每个事件都具有其来源服务器的主机值，您需要使用事件的数据来确定主机值。

更多信息请参阅“基于事件数据设置主机值”。

设置文件或目录输入的默认主机

您可以为通用转发器和 Splunk Enterprise 上特定文件或目录输入的所有数据设置主机值。您既可以静态设置主机，也可以动态设置主机。

在 Splunk Cloud 上，您必须使用通用转发器来分配主机值，作为收集要发送到 Splunk Cloud 的数据的一部分。您不能在

Splunk Web 中配置主机名。

若您静态设置主机值，Splunk 平台会为从指定文件或目录输入接收的每个事件分配相同的主机。

若您动态设置主机值，Splunk 平台将使用正则表达式或数据来源完整目录路径的一个段从数据来源输入中提取相同的主机名。

您也可以根据事件来源或来源类型值以及其他各类信息把主机值分配给来自特定文件或目录输入的事件。请参阅“根据事件数据设置主机值”。

目前，您无法启用网络（TCP 和 UDP）或**脚本式输入**的默认主机值的设置。

静态设置默认主机值

此方法将把单个默认主机值应用到特定文件或目录输入生成的每个事件。

静态主机值分配仅影响某个输入生成的新事件。不能向已经索引的数据分配默认主机值。相反，您必须用主机值标记现有事件。请参阅*《知识管理器手册》*中的“在‘设置’中定义和管理标记”。

编辑 inputs.conf 配置文件以静态设置默认主机

如需为受监视的文件或目录输入指定主机值，您可以编辑 inputs.conf 配置文件。编辑 inputs.conf 文件时，请在定义输入的段落中设置 host 设置。如果您使用 Splunk Cloud，请在运行通用转发器的计算机上配置该设置。

```
[monitor://<path>]
host = <your_host>
```

编辑 $SPLUNK_HOME/etc/system/local/ 或 $SPLUNK_HOME/etc/apps/ 中您自己的自定义 Splunk 应用程序目录中的 input.conf 文件。

有关配置文件的一般详细信息，请参阅*《管理员手册》*中的“关于配置文件”。有关输入和输入类型的更多信息，请参阅“什么数据可以建立索引？”。

静态主机值分配示例

本示例涵盖任何来自 /var/log/httpd 的事件。任何来自此输入的事件将接收到一个 host 值，该值属于 webhead-1。

```
[monitor:///var/log/httpd]
host = webhead-1
```

使用 Splunk Web 静态设置默认主机

在 Splunk Enterprise 中，您可以在添加或编辑文件或目录输入时随时定义这类输入的主机。

要在新建输入时设置默认主机，请参阅本主题后面的“设置新输入的默认主机”。

要在现有输入上静态设置默认主机，请执行以下步骤：

1. 在 Splunk Web 中单击**设置 > 数据导入**。
2. 单击**文件和目录**。
3. 在**文件和目录**页面单击现有输入的名称即可更新该输入。
4. 在**主机**部分中，从**设置主机**下拉列表中选择**常量值**。
5. 在**主机字段值**字段中输入输入的静态主机值。
6. 单击**保存**。

设置新输入的默认主机

创建输入时，您必须遵循不同的过程来设置默认主机。

1. 在 Splunk Web 中单击**设置 > 数据导入**。
2. 单击**文件和目录**。
3. 在**文件和目录**页面单击**新建**即可添加输入。
4. 指定您想要监视的文件或目录，并指定任意允许列表或拒绝列表。
5. 单击**下一步**。
6. （可选）设置新输入的来源类型。
   如您指定了目录，“设置来源类型”页面将不会出现。
7. 单击**下一步**。
8. 在**主机**部分里的**输入设置**页面单击**常量值**。

9. 在**主机字段值**字段输入该输入的主机名称。
10. 单击**审阅**以继续到“审阅”页面。
11. 单击**提交**来新建输入。

动态设置默认主机值

此方法从来源输入路径段或从正则表达式提取文件或目录输入的主机值。例如，如果要索引归档的目录且目录中每个文件的名称都包含相关的主机信息，则可以提取此信息并将其分配给主机字段。

您可以在搜索中将正则表达式与 rex 搜索命令结合使用，以对表达式进行测试。

使用 inputs.conf 文件动态设置默认主机

配置 inputs.conf 可让您设置动态主机提取规则。有关配置文件的一般详细信息，请参阅*《管理员手册》*中的“关于配置文件”。

使用 host_regex 属性设置事件主机

1. 编辑 $SPLUNK_HOME/etc/system/local/ 或 $SPLUNK_HOME/etc/apps/ 中您自己的自定义应用程序目录中的 input.conf。
2. 使用 host_regex 属性把主机字段覆盖为通过正则表达式提取到的值。

```
[monitor://<path>]
host_regex = <your_regular_expression>
```

3. 保存 inputs.conf 文件。
4. 重新启动 Splunk 平台实例。

正则表达式会从每项输入的文件名称中提取 host 值。输入会把正则表达式的第一个捕获组用作主机。若正则表达式匹配失败，输入会把默认的 host 属性设置为主机。

使用 host_segment 属性设置事件主机

host_segment 值使用从数据来源路径段中提取到的值覆盖主机字段。

1. 编辑 $SPLUNK_HOME/etc/system/local/ 或 $SPLUNK_HOME/etc/apps/ 中您自己的自定义应用程序目录中的 input.conf。
2. 将 host_segment 属性添加到段落，通过这种方法用从数据来源路径段中提取到的值覆盖主机字段。例如，若数据来源的路径为 /var/log/<host server name> 且您想要使主机服务器名称或第三个段成为主机值，请按以下步骤设置 host_segment：

```
[monitor://var/log/]
host_segment = 3
```

3. 保存 inputs.conf 文件。
4. 重新启动 Splunk 平台实例。

动态主机分配示例

本示例中，正则表达式将为所有来自 /var/log/foo.log 的事件分配一个主机值 foo：

```
[monitor://var/log]
host_regex = /var/log/(\w+)
```

本示例将主机值分配给 apache/logs 路径中的第三段：

```
[monitor://apache/logs/]
host_segment = 3
```

使用 Splunk Web 动态设置默认主机

1. 单击**设置 > 数据导入**。
2. 单击**文件和目录**。
3. 在**文件和目录**页面单击现有输入的名称即可更新该输入。
4. 在**主机**部分中，从**设置主机**下拉列表中选择以下两个选项。

提取方法	步骤
使用正则表达式提取主机名	1. 选择**路径里的正则表达式**。 2. 在**正则表达式**字段中输入要提取的主机的正则表达式。
从数据来源路径中的段提取主机名	1. 选择**路径里的段**。 2. 在**段编号**字段中输入段编号。例如，如果来源路径为 /var/log/<host server name> 而您想要使主机服务器名称成为主机值，请输入 3 即可提取第三段。

5. 单击**保存**。

动态设置新输入的默认主机

创建输入时，您必须遵循不同的过程来动态设置默认主机。

1. 单击**设置 > 数据导入**。
2. 单击**文件和目录**。
3. 在**文件和目录**页面单击**新建**即可添加输入。
4. 指定您想要监视的文件或目录，并指定任意允许列表或拒绝列表。
5. 单击**下一步**。
6. （可选）设置新输入的来源类型。
   如您指定了目录，“设置来源类型”页面将不会出现。
7. 单击**下一步**。
8. 在**主机**部分里的**输入设置**页面上单击**路径正则表达式**或**路径里的段**。
9. 如您选择了**路径正则表达式**，请输入从**正则表达式**字段数据来源路径中提取主机名时使用的正则表达式。否则，请输入决定**段编号**字段中主机名时使用的数据来源路径段的编号。
10. 单击**审阅**以继续到“审阅”页面。
11. 单击**提交**来新建输入。

设置 host_segment 属性来提取主机名的注意事项

使用 inputs.conf 段落中的 host_segment 属性时须遵循一些注意事项：

- 不能在同一个段落中同时指定 host_regex 和 host_segment 属性。
- 在同一个段落中同时指定 host_segment 和 source 属性后，host_segment 属性的行为将会更改：
  - 若您为数据来源指定的值包含正斜线（/），则将根据您在 host_segment 中指定的段编号提取主机值。
  - 若 source 不含正斜线（/），或您指定的 host_segment 值大于 source 中可用的段数量，那么 Splunk 平台将无法提取主机值，只能使用提取数据的主机的名称。

以下示例显示了当 source 不包含正斜线或您指定的 host_segment 值大于 source 中可用的段数时会发生的情况：

主机名	来源路径	Inputs.conf 配置	结果主机值
server01	/mnt/logs/server01	[monitor:///mnt/logs/] host_segment = 3	server01
server01	/mnt/logs/server01	[monitor:///mnt/logs/server01] source = /mnt/logs/server01 host_segment = 3	server01
server02	/mnt/logs/server02	[monitor:///mnt/logs/server02] source = serverlogs host_segment = 3	server02

基于事件数据设置主机值

通过对 Splunk 平台进行配置，您可以根据事件中的数据将主机名分配到您的事件。您可以使用事件数据覆盖 Splunk 平台的默认分配，方法是为事件数据提供正则表达式并配置两个配置文件，以确定平台何时覆盖事件的主机名。

在 Splunk Cloud 上，您必须配置重型转发器才能执行主机名分配，然后将该数据转发到您的 Splunk Cloud 实例。您必须执行此额外步骤，因为您无法直接编辑 Splunk Cloud 实例上的配置文件。

在 Splunk Enterprise 上，您可以在索引器或重型转发器上编辑配置文件。在任何情况下，您都不能使用通用转发器，因为通用转发器无法转换数据，除非在某些有限的情况下。

有关正则表达式语法和用法的入门，请参阅网站 http://www.regular-expressions.info/reference.html。您可以在搜索中将正则表达式与 rex 搜索命令结合使用，以对表达式进行测试。

使用配置文件覆盖事件中的主机名默认字段

在引入数据时，Splunk 平台会使用默认字段标记事件数据。为 Splunk 平台建立索引的事件创建主机名覆盖需要您根据其中一些默认字段在收集数据的 Splunk 平台实例上编辑两个配置文件。

关于更新 transforms.conf 文件

第一个文件 transforms.conf 通过使用正则表达式来配置主机名覆盖，以确定实例何时覆盖或转换主机名默认字段。您可以通过确定事件数据中触发转换的确切内容来提供正则表达式，然后将该正则表达式提供给 transforms.conf 文件。您将此正则表达式作为文件中的一个段落提供，当传入的事件数据与您指定的正则表达式匹配时，Splunk 平台将触发覆盖。

关于更新 props.conf 文件

第二个文件 props.conf 确定可以应用主机名覆盖的默认字段。这些字段在文件中显示为一个段落，指定默认字段，Splunk 平台可在其中修改传入事件的主机名字段。

您可以将主机名覆盖应用于以下默认字段：

- 来源，使用 source::<source> 关键字
- 来源类型，使用 sourcetype=<sourcetype> 关键字
- 主机名，使用 host::<host> 关键字

当您在 props.conf 文件中指定这些默认字段之一时，会发生主机名覆盖。在 Splunk 平台可以覆盖主机名之前，必须发生以下事件：

- 传入事件数据中的主机、来源或来源类型必须与您在 props.conf 文件中指定的内容相匹配，以激活 transforms.conf 文件中的主机名覆盖转换配置。
- 事件数据必须与您为要触发的主机名覆盖转换设置的正则表达式匹配。

创建主机名覆盖

1. 查看您的事件数据以确定表示您希望 Splunk 平台何时执行主机名覆盖的字符串。此字符串将成为您稍后在过程中提供的正则表达式。请参阅本主题后续的示例。
2. 查看本节中的“配置具有主机名覆盖转换的 transforms.conf 段落”和“配置 props.conf 段落以引用主机名覆盖转换”部分，以了解主机名覆盖的段落语法如何工作。
3. 在要覆盖主机名的重型转发器上，打开文本编辑器。
4. 使用该编辑器，打开 $SPLUNK_HOME/etc/system/local/transforms.conf 文件进行编辑。
5. 向此文件添加一个段落，表示 Splunk 平台何时执行主机名覆盖。
6. 保存 transforms.conf 文件并将其关闭。
7. 打开文件 $SPLUNK_HOME/etc/system/local/props.conf 进行编辑。
8. 向此文件添加一个段落，表示要应用主机名覆盖的默认字段。
9. 保存 props.conf 文件并将其关闭。
10. 重新启动重型转发器。

在 Splunk Enterprise 上，您可以在引入数据的实例或向实例发送数据的重型转发器上执行此过程。

更多有关配置文件的一般信息，请参阅 Splunk Enterprise *《管理员手册》* 中的“关于配置文件”。

使用主机名覆盖转换配置 transforms.conf 段落

transforms.conf 文件控制 Splunk 平台在何处以及如何转换传入的事件数据。

transforms.conf 中的主机名覆盖转换段落使用以下语法：

```
[<unique_stanza_name>]
REGEX = <your_regex>
FORMAT = host::$1
DEST_KEY = MetaData:Host
```

本段落有几点事项需要注意：

- 使用语法的 <unique_stanza_name> 部分来引用 props.conf 配置文件中的转换。最佳做法是让它反映其涉及到主机值。
- <your_regex> 是正则表达式，用于标识要在事件中提取主机值的位置并将该值指定为该事件的默认字段。
- FORMAT = host::$1 将 REGEX 值写入 host:: 字段。

配置 props.conf 段落以引用主机名覆盖转换

props.conf 文件引用了执行转换的 transforms.conf 文件中的段落：

```
[<spec>]
TRANSFORMS-<class> = <unique_stanza_name>
```

本段落有几点事项需要注意：

- <spec> 可以是以下任何值：
  - <sourcetype>，即事件的来源类型。
  - host::<host>，其中 <host> 是事件的主机值。
  - source::<source>，其中 <source> 是事件的来源值。
- <class> 是您想要为转换指定的唯一标识符。
- <unique_stanza_name> 是在 transforms.conf 中创建的段落的名称。

主机名默认字段覆盖示例

假设从 houseness.log 文件的以下事件集开始。您希望 Splunk 平台将每个事件的主机默认字段设置为在事件中找到的主机名。主机位于日志文件中每行的第三个位置。在本示例中，它为 fflanda。

```
41602046:53 accepted fflanda
41602050:29 accepted rhallen
41602052:17 accepted fflanda
```

首先，在 transforms.conf 文件中新建一个段落，并提供提取主机值的正则表达式：

```
[houseness]
DEST_KEY = MetaData:Host
REGEX = \s(\w*)$
FORMAT = host::$1
```

接下来，在 props.conf 配置文件的段落中引用 transforms.conf 段落。请参阅以下示例：

```
[source::.../houseness.log]
TRANSFORMS-rhallen=houseness
SHOULD_LINEMERGE = false
```

本示例段落包含传统的设置 SHOULD_LINEMERGE = false，可在每个换行符处划分事件。

稍后事件将在搜索结果中显示如下：

建立索引后更改主机值

在建立索引后，您可能会注意到一些事件的主机值有误。例如，您可能直接在 Splunk 平台实例上将 Web 代理日志收集到目录中，并且添加该目录作为输入而忘记覆盖主机字段值，则将导致主机值与 Splunk 平台实例主机的值相同。

如果发生此类情况，可以按照从容易到复杂选择以下方案。您可以使用 Splunk Cloud 实例执行所有这些操作：

- 请删除并重新为数据建立索引。请参阅 Splunk Enterprise *《管理索引器和索引器群集》* 手册中的“删除索引和索引的数据”。

- 使用搜索删除主机值不正确的特定事件，然后重新索引这些事件。请参阅 Splunk Enterprise*《管理索引器和索引器群集》*手册中的“删除整个索引”。
- 标记不正确的主机值，然后使用标记进行搜索。请参阅 Splunk Enterprise*《管理索引器和索引器群集》*手册中的“在‘搜索’中为字段-值对设置标记”。
- 请设置一个逗号分隔值（CSV）查找以查找主机，在查找文件中将其映射到新的字段名称，然后在搜索中使用新名称。请参阅 Splunk Enterprise*《知识管理器手册》*中的“查找配置简介”。
- 为主机字段创建别名到新字段（例如 `temp_host`），设置一个 CSV 查找以使用名称 `temp_host` 查找正确的主机名称，然后再借助该查找，用新的查找值覆盖原始的 `host`（使用 `OUTPUT` 选项定义该查找）。请参阅 Splunk Enterprise*《知识管理器手册》*中的“在 Splunk Web 中创建字段别名”和“查找配置简介”。

在这些选项中，删除和重新建立索引能够为您带来最佳性能，而且也是最容易的操作。如果您无法删除数据并为其重新建立索引，则最后一个选项提供了最快的替代方案。

有关覆盖主机字段值的更多信息，请参阅“覆盖主机字段的值”。

配置来源类型

来源类型为何重要

来源类型是 Splunk 平台分配给所有传入数据的众多**默认字段**之一。来源类型会告知平台您所获数据的类型，方便平台在建立索引期间智能地为数据设置格式。来源类型还可以用于数据分类，进而简化搜索。

来源类型将决定如何为传入的数据设置格式

因为来源类型控制着 Splunk 平台如何为传入的数据设置格式，所以把正确的来源类型分配给您的数据很重要。由此一来，数据的索引版本（**事件数据**）将符合您的要求，包含适当的**时间戳**和**事件**分隔符。这有助于后期数据的搜索变得更容易。

Splunk 软件提供了大量预定义的来源类型。获取数据时，Splunk 平台通常会自动选择正确的来源类型。如果您的数据为专用数据，则可能需要手动选择其他预定义来源类型。如果您的数据非常罕见，则您可能需要新建一个包含自定义事件处理设置的全新来源类型。如果您的数据源包含异类数据，则可能需要根据每个事件而不是每个来源分配来源类型。

与任何其他字段相似，索引数据后，您同样可以使用来源类型字段搜索事件数据。由于来源类型是分类数据的主要方法，因此您将在搜索中频繁使用来源类型。

常见来源类型

任何常见数据导入格式都可以是来源类型。大多数来源类型都是日志格式。例如，Splunk 平台会自动识别的常见来源类型包括：

来源类型	描述
access_combined	对于 NCSA 组合日志格式 HTTP Web 服务器日志。
apache_error	对于标准的 Apache Web 服务器错误日志。
cisco_syslog	对于通过 Cisco 网络设备（包括 PIX 防火墙、路由器和 ACS）生成的标准 syslog，通常使用远程 syslog 到中央日志主机。
websphere_core	WebSphere 的核心文件导出。

如需查看完整的预定义来源类型列表，请参阅本手册中的“预置来源类型列表”。

配置来源类型

您可以对来源类型执行两种基本的配置类型：

- 为传入数据显式分配来源类型。
- 重新建立来源类型，或者通过修改现有来源类型来新建来源类型。

分配来源类型

大多数情况下，Splunk 平台会为您的数据决定最佳来源类型，并自动将最佳来源类型分配给传入的事件。然而，在某些情况下，您可能需要为数据显式分配来源类型。通常在定义数据导入时执行此分配。有关如何改进来源类型分配的详细信息，请参阅以下主题：

- 覆盖自动来源类型分配
- 基于每个事件覆盖来源类型
- 配置基于规则的来源类型识别
- 新建来源类型
- 重命名来源类型

有关 Splunk 平台如何分配源类型的更多信息，请参阅“Splunk 平台如何分配来源类型”。

新建来源类型

如果现有来源类型都不符合您的数据需求，则您可以新建一个类型。

Splunk Web 允许用户调整来源类型设置，使其更好地符合数据。实际上是一个可视化的来源类型编辑器。请参阅“使用‘设置来源类型’页面”。

如果您使用的是 Splunk Cloud，请使用 Splunk Web 定义来源类型。如果您使用的是 Splunk Enterprise，则可以使用

Splunk Web 或通过编辑 props.conf 配置文件并在那里添加来源类型段落来新建来源类型。请参阅“新建来源类型”。

预览数据以测试和修改来源类型

Splunk Web 允许您查看把来源类型应用到输入的效果。这样您便可以预览生成的事件而无需将事件真正提交给索引。您也可以编辑时间戳和事件换行设置，然后将修改结果另存为新的来源类型。有关数据预览功能如何用作来源类型编辑器的信息，请参阅“使用‘设置来源类型’页面”。

搜索来源类型

sourcetype 是来源类型搜索字段的名称。您可以使用 sourcetype 字段从任意来源类型中查找相似的数据类型。例如，即使 WebLogic 正从多个域或主机（以 Splunk 术语表示）中执行记录，您仍可以通过搜索 sourcetype=weblogic_stdout 查找所有 WebLogic 服务器事件。

Splunk 平台如何分配来源类型

Splunk 平台采用各种方法在索引时间为事件数据分配来源类型。Splunk Cloud 和 Splunk Enterprise 以相同的方式执行这些方法。不同之处在于，在 Splunk Cloud 上，您只能使用 Splunk Web 更改源类型配置。如果不向 Splunk 支持提交工单，则无法访问任何配置文件。

Splunk 平台处理事件数据时，将按照定义好的优先顺序逐步执行这些方法。从 inputs.conf 和 props.conf 配置文件中静态配置的来源类型配置开始，接着是基于规则的来源类型关联，之后是使用自动来源类型识别和自动来源类型学习等方法。这些方法可以在 Splunk 平台将来源类型值自动分配给其他事件的同时，让您配置 Splunk 平台将来源类型值应用到特定事件类型的方法。

以下列表显示 Splunk 平台如何为数据导入决定来源类型。Splunk 平台先从第一个方法开始，然后再根据需要往下执行其他方法，直到能够确定来源类型为止。

- 基于数据导入的显式来源类型规范
- 基于数据源的显式来源类型规范
- 基于规则的来源类型识别
- 自动来源类型匹配
- 延迟的基于规则的来源类型关联
- 自动来源类型学习

基于数据导入的显式来源类型规范

若为数据导入找到了显式来源类型，Splunk 平台会停在此处。

如果您使用 Splunk Cloud，则在 Splunk Web 中进行配置。如果您使用 Splunk Enterprise，您可以在 Splunk Web 或在 inputs.conf 配置文件中进行配置。以下是配置 input.conf 文件以将来源类型分配给文件输入的语法：

```
[monitor://<path>]
sourcetype=<sourcetype>
```

在 Splunk Web 中定义输入时也可以分配来源类型。有关文件输入执行此操作的相关信息，请参阅本手册中的“使用 Splunk Web 监视 Splunk Enterprise 中的文件和目录”。此过程与网络或其他类型的输入的过程相似。

更多信息请参阅“指定输入的来源类型”。

基于数据源的显式来源类型规范

若为特定数据来源找到了显式来源类型，Splunk 平台会停在此处。

如果您使用 Splunk Enterprise，或者您想使用重型转发器将此数据转发到 Splunk Cloud，则可以使用以下语法在 props.conf 配置文件中进行配置：

```
[source::<source>]
sourcetype=<sourcetype>
```

更多信息请参阅“指定来源的来源类型”。

基于规则的来源类型识别

Splunk 平台接下来会为来源类型查找您已经新建的规则。

如果您使用 Splunk Enterprise，则可以在 props.conf 文件中创建来源类型分类规则：

```
[rule::<rule_name>]
sourcetype=<sourcetype>
MORE_THAN_[0-100] = <regex>
LESS_THAN_[0-100] = <regex>
```

有关设置来源类型识别规则的信息，请参阅“配置基于规则的来源类型识别”。

自动来源类型匹配

Splunk 平台接下来会尝试使用自动来源类型识别来匹配看上去相似的文件并为这些文件分配来源类型。

Splunk 平台会在任意文件或网络输入流的前几千行中计算模式的签名。这些签名标识了重复单词模式、标点符号模式、行长度等内容。计算签名时，Splunk 平台会将签名与它提供给已知“预置”来源类型的一组签名进行对比。如果确定匹配，则将该来源类型分配给数据。

有关 Splunk 平台默认可识别的来源类型列表，请参阅“预置来源类型列表”。

延迟的基于规则的来源类型关联

如果到目前为止仍尚未确定来源类型，Splunk 平台会查寻延迟的规则。

运行方式和基于规则的关联相似。如果您使用 Splunk Enterprise，则可以在 props.conf 文件中创建 delayedrule:: 段落。这种获取全部来源类型的方法很有用，以免 Splunk Enterprise 在使用智能匹配时错过任何来源类型。

延迟规则关联的一个范例就是基于规则的步骤中使用 rule:: 定义的相当特定的来源类型的通用版本。例如，您可以使用 rule:: 获取具有特定 syslog 来源类型（如“sendmail syslog”或“cisco syslog”）的事件数据，然后再通过 delayedrule:: 将通用的 syslog 来源类型应用到剩余的 syslog 事件数据。

语法如下文所示：

```
[delayedrule::$RULE_NAME]
sourcetype=$SOURCETYPE
MORE_THAN_[0-100] = $REGEX
LESS_THAN_[0-100] = $REGEX
```

有关设置或移除来源类型识别延迟规则的更多信息，请参阅“配置基于规则的来源类型识别”。

自动来源类型学习

如果无法使用前述方法为事件分配来源类型，Splunk 平台会为事件签名新建一个来源类型。Splunk 平台将学习到的模式信息存储在 sourcetypes.conf 配置文件中。

覆盖自动来源类型分配

Splunk® Enterprise 会尝试向数据自动分配来源类型。您可以指定要分配的来源类型。您也可以对 Splunk 平台进行配置，这样它就可以根据数据导入或数据来源分配来源类型。

Splunk 平台为数据分配来源类型时会采用优先顺序规则，有关该规则的详细信息，请参阅“来源类型为何重要”中的“Splunk 平台如何分配来源类型”。

只能在监视输入的文件和目录中，或者您上传的文件中覆盖。您无法覆盖网络输入上的来源类型。此外，覆盖只会影响在设置覆盖之后到达的新数据。要更正已经索引的事件的来源类型，可转而为来源类型新建一个标记。请参阅*《知识管理器手册》*中的“在‘搜索’中为字段-值对设置标记”。

可以根据数据的输入和来源为数据指定来源类型。

指定输入的来源类型

您可以为来自特定输入（例如 /var/log/）的数据分配来源类型。如果您使用的是 Splunk Cloud，请使用 Splunk Web 定义来源类型。如果使用的是 Splunk Enterprise，您可以在 Splunk Web 中或通过编辑 inputs.conf 配置文件来定义来源类型。

请注意，通过输入分配来源类型不是很细致。当您通过输入指定源类型时，Splunk 平台会为所有来自输入的数据分配相同的来源类型，即使其中某些数据来自不同的来源或主机亦如此。若想避开来源类型自动分配改而采取更具针对性的方法，您可以按照“指定来源的来源类型”部分中所述，根据数据的来源为其分配来源类型。

使用 Splunk Web

定义数据导入之后，您可以设置一个来源类型值应用到来自该输入的所有传入数据。您可以从列表中选择一个来源类型，也可以输入自己的来源类型值。

要为输入选择一个来源类型，请更改您要添加的数据导入类型的来源类型设置。例如，对于文件输入，请完成以下步骤：

1. 单击 Splunk Web 右上角的**设置**。
2. 在“设置”下拉列表的**数据**部分中，单击**数据导入**。
3. 单击**文件和目录**。
4. 单击**新建**以添加输入。
5. 在**添加数据**页面，请浏览或输入您想要监视的文件的名称，然后单击**下一步**。
6. 在**设置来源类型**页面，请单击**来源类型**下拉列表并从预置来源类型列表中选择。请参阅“预置来源类型列表”。Splunk Web 会更新此页面，显示数据接收到新来源类型时的外观。
7. 如果您要更改来源类型，请使用**事件换行**、**时间戳**和**高级**选项卡以修改设置并更新数据预览。请参阅本手册中的“将正确的来源类型分配给您的数据”。
8. 如果您要将来源类型保存为其他名称，请单击**另存为...**打开**保存来源类型**对话框以保存新的来源类型。否则，请转到步骤 10。
9. 如果选择保存来源类型，请输入名称、描述、类别和来源类型应适用的应用。请参阅“将修改另存为新来源类型”。
10. 请单击**下一步**为数据设置来源类型并转到“输入设置”页面。请参阅“修改输入设置”。

Splunk® Enterprise 此时会将选定的来源类型分配给针对该输入索引的所有事件。

使用 inputs.conf 配置文件

在 Splunk Enterprise 实例的 inputs.conf 配置文件中配置输入时，可以为该输入指定来源类型。在 Splunk Cloud 实例上，您可以在包含要收集并转发到 Splunk Cloud 实例的数据的计算机上配置通用转发器。

编辑 $SPLUNK_HOME/etc/system/local/ 或 $SPLUNK_HOME/etc/apps/ 中您自己的自定义应用程序目录中的 input.conf 文件。有关配置文件的一般信息，请参阅*《管理员手册》*中的“关于配置文件”。

若要指定来源类型，请在输入段落中加入 sourcetype 属性。例如：

```
[tcp://:9995]
connection_host=dns
sourcetype=log4j
source=tcp:9995
```

本例中将来自端口 9995 上 TCP 输入的任何事件的来源类型设置为 log4j。

请勿在属性值两侧加引号。例如，正确的格式是 sourcetype=log4j，而不是 sourcetype="log4j"。

指定来源的来源类型

使用 props.conf 文件覆盖自动来源类型匹配并向来自特定来源的所有数据显式分配单一来源类型。

编辑 $SPLUNK_HOME/etc/system/local/ 或 $SPLUNK_HOME/etc/apps/ 中您自己的自定义应用程序目录中的 props.conf。有关配置文件的一般信息，请参阅*《管理员手册》*中的“关于配置文件”。

如果要覆盖来源类型，则必须在配置输入的转发器上的 props.conf 中配置该设置。

若要覆盖来源类型分配，请将您的来源的一个段落添加到 props.conf。在该段落中，为实现灵活性，可根据需要使用正则表达式 (regex) 语法标识来源路径。然后通过加入一个 sourcetype 属性来指定来源类型。例如：

```
[source::.../var/log/anaconda.log(.\d+)?]
sourcetype=anaconda
```

只要事件源自其内的任何来源中包含字符串 /var/log/anaconda.log 且该字符串后接有任意数量的数字字符，本示例都会把来源类型设置为 anaconda。

您的段落来源路径正则表达式（例如 [source::.../web/....log]）必须尽可能具体。避免使用以 ... 结尾的正则表达式。例如，请避免如下做法：

```
[source::/home/fflanda/...]
sourcetype=mytype
```

这种格式非常危险。它将指示 Splunk 平台以 mytype 文件而不是 GZIP 文件形式处理 /home/fflanda 中的任何 GZIP 文件。

相反，使用以下格式编写：

```
[source::/home/fflanda/....log(.\d+)?]
sourcetype=mytype
```

配置基于规则的来源类型识别

您可以使用基于规则的来源类型识别扩大 Splunk 软件可以识别的来源类型范围。

您必须使用配置文件配置基于规则的来源类型识别。如果您使用的是 Splunk Cloud，则在将数据发送到 Splunk Cloud 之前，您必须使用重型转发器配置来源类型识别。您可能还需要提交支持工单以确认您的 Splunk Cloud 部署识别源类型规则。

如果您使用的是 Splunk Enterprise，则可以在 props.conf 配置文件中创建一个 rule:: 段落，以将特定来源类型和一组限定条件关联起来。获取数据时，Splunk 平台会将指定的来源类型分配给符合规则限定条件的文件输入。

基于规则的来源类型识别如何工作

可以在 props.conf 文件中创建两种类型的规则：规则和延迟规则。二者之间的区别是来源键入过程期间 Splunk 平台对二者执行检查的时间不同。处理每组传入数据的过程中，Splunk 平台会使用多种方法确定来源类型：

- Splunk 平台检查显式来源类型定义后，平台会查看所有 rule:: 段落（定义于 props.conf 文件中），并尝试根据这些段落内指定的分类规则将来源类型和数据进行匹配。
- 若无法使用可用的 rule:: 段落找到匹配的来源类型，Splunk 平台将尝试使用来源类型自动匹配，识别与其过去学习到的来源类型相似的模式。
- 如果该方法失败，Splunk 平台接下来将检查所有 delayedrule:: 段落（位于 props.conf 文件中），并试着使用这些段落中的规则，把数据和来源类型进行匹配。

Splunk 平台为数据分配来源类型时会采用优先顺序规则，有关该规则的详细信息，请参阅“Splunk 平台如何分配来源类型”。

您可以配置 Splunk 平台，这样 rule:: 段落就将包含特定来源类型的分类规则，而 delayedrule:: 段落则将包含通用来源类型的分类规则。通过这种方式，Splunk 平台会将一般来源类型应用于大量不符合更为专用的来源类型条件的事件。

例如，您可以使用 rule:: 段落获取具有特定 syslog 来源类型（如 sendmail_syslog 或 cisco_syslog）的数据，然后再配置一个 delayedrule:: 段落，以把通用 syslog 来源类型应用到任何剩下的 syslog 数据。

在 props.conf 中创建来源键入规则

要设置来源键入规则，请编辑 $SPLUNK_HOME/etc/system/local/ 或 $SPLUNK_HOME/etc/apps/ 中您自己的自定义应用程序目录中的 props.conf 配置文件。有关配置文件的一般信息，请参阅*《管理员手册》*中的“关于配置文件”。

按照以下步骤创建规则：

1. 在 props.conf 中，添加 rule:: 或 delayedrule:: 段落。在段落标题中提供该规则的名称。

   ```
   [rule::<rule_name>] OR [delayedrule::<rule_name>]
   ```
2. 在段落正文中声明来源类型。

   ```
   [rule::<rule_name>] OR [delayedrule::<rule_name>]
   sourcetype=<source_type>
   ```
3. 在声明来源类型之后，在 MORE_THAN 和 LESS_THAN 属性中设置一个数字值，与必须包含正则表达式指定的字符串的输入行百分比相对应。例如，MORE_THAN_80 表示至少 80% 的行必须包含关联的表达式。 LESS_THAN_20 表示少于 20% 的行可以包含关联的表达式。

   您可以在规则中设置任意数量的 MORE_THAN 和 LESS_THAN 条件。仅当数据符合规则中所有语句的限定时，才会将规则的来源类型分配给数据文件。例如，可以定义一个规则，仅在多于 60% 的行符合一个正则表达式且少于 20% 的行符合另一个正则表达式时，才将特定来源类型分配给文件输入。

   按照以下语法定义来源类型分配规则：

   ```
   [rule::<rule_name>] OR [delayedrule::<rule_name>]
   sourcetype=<source_type>
   MORE_THAN_[0-99] = <regex>
   LESS_THAN_[1-100] = <regex>
   ```

尽管命名中包含“more than”，MORE_THAN_ 属性实际上表示“大于或等于”。类似地，LESS_THAN_ 属性表示“少于或等于”。

您可以在搜索中将正则表达式与 rex 搜索命令结合使用，以对表达式进行测试。

示例

以下示例显示了配置来源类型规则的方法。

Postfix syslog 文件

```
# postfix_syslog sourcetype rule
[rule::postfix_syslog]
sourcetype = postfix_syslog
# If 80% of lines match this regex, then it must be this type
MORE_THAN_80=^\w{3} +\d+ \d\d:\d\d:\d\d .* postfix(/\w+)?\[\d+\]:
```

可分隔文本的延迟规则

```
# breaks text on ascii art and blank lines if more than 10% of lines have
# ascii art or blank lines, and less than 10% have timestamps
[delayedrule::breakable_text]
sourcetype = breakable_text
MORE_THAN_10 = (^(?:---|===|\*\*\*|___|=+=))|^\s*$
LESS_THAN_10 = [: ][012]?[0-9]:[0-5][0-9]
```

预置来源类型列表

Splunk 软件随附内置或预置的**来源类型**，把传入的数据分析成事件。

Splunk 平台可以自动识别多数的预置来源类型并将其自动分配给传入的数据。您还可以手动分配 Splunk 平台无法自动识别的预置来源类型。要手动分配来源类型，请参阅“覆盖自动来源类型分配”。

对于重型或通用转发器，您还可以从 input.conf 配置文件配置来源类型。如果您使用 Splunk Enterprise，则可以从 Splunk Web 或从 input.conf 文件分配来源类型。

如果预置的来源类型与您的数据相匹配，请使用预置的来源类型，因为 Splunk 平台已经知道如何正确地为预置来源类型建立索引。如果您的数据不符合任何预置来源类型，您可以新建自己的来源类型。请参阅“新建来源类型”。即使没有自定义属性，Splunk 平台实际上也可以为任何格式的数据建立索引。

自动识别的来源类型

下表显示了自动识别的来源类型：

来源类型	来源	示例
access_combined	NCSA 组合 格式 http web 服务器日志。可通过 Apache 或其他 web 服务器生成。	10.1.1.43 - webdev [08/Aug/2005:13:18:16 -0700] "GET / HTTP/1.0" 200 0442 "-" "check_http/1.10 (nagios-plugins 1.4)"
access_combined_wcookie	NCSA 组合格式 http web 服务器日志。可通过 Apache 或其他 web 服务器生成，在末尾添加 cookie 字段。	"66.249.66.102.1124471045570513" 59.92.110.121 - - [19/Aug/2005:10:04:07 -0700] "GET /themes/splunk_com/images/logo_splunk.png HTTP/1.1" 200 994 "http://www.splunk.org/index.php/docs" "Mozilla/5.0 (X11; U; Linux i686; en-US; rv:1.7.8) Gecko/20050524 Fedora/1.0.4-4 Firefox/1.0.4" "61.3.110.148.1124404439914689"
access_common	NCSA 普通格式 http web 服务器日志。可通过 Apache 或其他 web 服务器生成。	10.1.1.140 - - [16/May/2005:15:01:52 -0700] "GET /themes/ComBeta/images/bullet.png HTTP/1.1" 404 304
apache_error	标准 Apache web 服务器错误日志	[Sun Aug 7 12:17:35 2005] [error] [client 10.1.1.015] File does not exist: /home/reba/public_html/images/bullet_image.gif

asterisk_cdr	标准 Asterisk IP PBX 呼叫详细记录	"","5106435249","1234","default","""James Jesse""<5106435249>","SIP/5249-1ce3","","VoiceMail","u1234","2005-05-26 15:19:25","2005-05-26 15:19:25","2005-05-26 15:19:42",17,17,"ANSWERED","DOCUMENTATION"
asterisk_event	标准 Asterisk 事件日志（管理事件）	Aug 24 14:08:05 asterisk[14287]: Manager 'randy' logged on from 127.0.0.1
asterisk_messages	标准 Asterisk 消息日志（错误和警告）	Aug 24 14:48:27 WARNING[14287]: Channel 'Zap/1-1' sent into invalid extension 's' in context 'default', but no invalid handler
asterisk_queue	标准 Asterisk 队列日志	1124909007|NONE|NONE|NONE|CONFIGRELOAD|
cisco_syslog	所有 Cisco 网络设备生成的标准 Cisco syslog，包括 PIX 防火墙、路由器、ACS 等。通常经由远程 syslog 到中央日志主机。	Sep 14 10:51:11 stage-test.splunk.com Aug 24 2005 00:08:49: %PIX-2-106001: Inbound TCP connection denied from IP_addr/port to IP_addr/port flags TCP_flags on interface int_name Inbound TCP connection denied from 144.1.10.222/9876 to 10.0.253.252/6161 flags SYN on interface outside
db2_diag	标准 IBM DB2 数据库管理和错误日志	2005-07-01-14.08.15.304000-420 I27231H328 LEVEL: Event PID : 2120 TID : 4760 PROC : db2fmp.exe INSTANCE: DB2 NODE : 000 FUNCTION: DB2 UDB, Automatic Table Maintenance, db2HmonEvalStats, probe:900 STOP : Automatic Runstats: evaluation has finished on database TRADEDB
exim_main	Exim MTA mainlog	2005-08-19 09:02:43 1E69KN-0001u6-8E => support-notifications@splunk.com R=send_to_relay T=remote_smtp H=mail.int.splunk.com [10.2.1.10]
exim_reject	Exim 拒绝日志	2005-08-08 12:24:57 SMTP protocol violation: synchronization error (input sent without waiting for greeting): rejected connection from H=gate.int.splunk.com [10.2.1.254]
linux_messages_syslog	标准 Linux syslog，位于大多数平台上的 /var/log/messages	Aug 19 10:04:28 db1 sshd(pam_unix)[15979]: session opened for user root by (uid=0)
linux_secure	Red Hat、Debian 和均衡分布 Linux 身份验证日志	Aug 18 16:19:27 db1 sshd[29330]: Accepted publickey for root from ::ffff:10.2.1.5 port 40892 ssh2
log4j	由任何 J2EE 服务器使用 log4j 生成的 Log4j 标准输出	2005-03-07 16:44:03,110 53223013 [PoolThread-0] INFO [STDOUT] got some property...
mysqld_error	标准 MySQL 错误日志	050818 16:19:29 InnoDB: Started; log sequence number 0 43644 /usr/libexec/mysqld: ready for connections. Version: '4.1.10a-log' socket: '/var/lib/mysql/mysql.sock' port: 3306 Source distribution
mysqld	标准 MySQL 查询日志，转换为文本后也与 MySQL 的二进制日志相匹配	53 Query SELECT xar_dd_itemid, xar_dd_propid, xar_dd_value FROM xar_dynamic_data WHERE xar_dd_propid IN (27) AND xar_dd_itemid = 2
postfix_syslog	通过 *nix syslog 工具报告的标准 Postfix MTA 日志	Mar 1 00:01:43 avas postfix/smtpd[1822]: 0141A61A83: client=host76-117.pool80180.interbusiness.it[80.180.117.76]
sendmail_syslog	通过 *nix syslog 工具报告的标准 Sendmail MTA 日志	Aug 6 04:03:32 nmrjl00 sendmail[5200]: q64F01Vr001110: to=root, ctladdr=root (0/0), delay=00:00:01, xdelay=00:00:00, mailer=relay, min=00026, relay=[101.0.0.1] [101.0.0.1], dsn=2.0.0, stat=Sent (v00F3HmX004301 Message accepted for delivery)
sugarcrm_log4php	使用 log4php 实用工具报告的标准 Sugarcrm 活动日志	Fri Aug 5 12:39:55 2005,244 [28666] FATAL layout_utils - Unable to load the application list language file for the selected language(en_us) or the default language(en_us)
weblogic_stdout	采用标准本机 BEA 格式的 Weblogic 服	####<Sep 26, 2005 7:27:24 PM MDT><Warning><WebLogicServer><bea03><asiAdminServer><ListenThread.Default><<WLS

weblogic_stdout	格式的 WebLogic 服务器日志	Kernel>><><BEA-000372><HostName: 0.0.0.0, maps to multiple IP addresses:169.254.25.129,169.254.193.219>
websphere_activity	Websphere 活动日志，通常也称为服务日志	------------------------------------ ComponentId: Application Server ProcessId: 2580 ThreadId: 0000001c ThreadName: Non-deferrable Alarm : 3 SourceId: com.ibm.ws.channel.framework.impl. WSChannelFrameworkImpl ClassName: MethodName: Manufacturer: IBM Product: WebSphere Version: Platform 6.0 [BASE 6.0.1.0 o0510.18] ServerName: nd6Cell01\was1Node01\TradeServer1 TimeStamp: 2005-07-01 13:04:55.187000000 UnitOfWork: Severity: 3 Category: AUDIT PrimaryMessage: CHFW0020I: The Transport Channel Service has stopped the Chain labeled SOAPAcceptorChain2 ExtendedMessage: ----------
websphere_core	Websphere 的核心文件导出	NULL-- 0SECTION TITLE subcomponent dump routine NULL=============================== 1TISIGINFO signal 0 received 1TIDATETIME Date: 2005/08/02 at 10:19:24 1TIFILENAME Javacore filename: /kmbcc/javacore95014.1122945564.txt NULL ------------------------------------ --------- 0SECTION XHPI subcomponent dump routine NULL =============================== 1XHTIME Tue Aug 2 10:19:24 20051XHSIGRECV SIGNONE received at 0x0 in <unknown>. Processing terminated. 1XHFULLVERSION J2RE 1.3.1 IBM AIX build ca131-20031105 NULL
websphere_trlog_syserr	采用 IBM 本机 trlog 格式的标准 Websphere 系统错误日志	[7/1/05 13:41:00:516 PDT] 000003ae SystemErr R at com.ibm.ws.http.channel. inbound.impl.HttpICLReadCallback.complete (HttpICLReadCallback.java(Compiled Code)) (truncated)
websphere_trlog_sysout	采用 IBM 本机 trlog 格式的标准 Websphere 系统输出日志。与 Resin 和 Jboss 的 log4j 服务器日志相似。作为系统错误日志的示例格式，但包含更低的严重性和信息事件。	[7/1/05 13:44:28:172 PDT] 0000082d SystemOut O Fri Jul 01 13:44:28 PDT 2005 TradeStreamerMDB: 100 Trade stock prices updated: Current Statistics Total update Quote Price message count = 4400 Time to receive stock update alerts messages (in seconds): min: -0.013 max: 527.347 avg: 1.0365270454545454 The current price update is: Update Stock price for s:393 old price = 15.47 new price = 21.50
windows_snare_syslog	通过第三方 Intersect Alliance Snare 代理报告给 *nix 服务器上远程 syslog 的标准 windows 事件日志	0050818050818 Sep 14 10:49:46 stage-test.splunk.com Windows_Host MSWinEventLog 0 Security 3030 Day Aug 24 00:16:29 2005 560 Security admin4 User Success Audit Test_Host Object Open: Object Server: Security Object Type: File Object Name: C:\Directory\secrets1.doc New Handle ID: 1220 Operation ID: {0,117792} Process ID: 924 Primary User Name: admin4 Primary Domain: FLAME Primary Logon ID: (0x0,0x8F9F) Client User Name: - Client Domain: - Client Logon ID: - Accesses SYNCHRONIZE ReadData (or ListDirectory) Privileges -Sep

特殊来源类型

下表显示了特殊的来源类型：

来源类型	来源	示例
known_binary	文件名与通常作为二进制文件而不是日志文件的模式相匹配	MP3 文件、图像、.rdf 文件、.dat 文件和其他明显的非文本文件

预置来源类型

下表显示了预置来源类型，包括自动识别的来源类型和不自动识别的来源类型：

类别	来源类型
应用程序服务器	log4j、log4php、weblogic_stdout、websphere_activity、websphere_core、websphere_trlog、catalina、ruby_on_rails
数据库	db2_diag、mysqld、mysqld_error、mysqld_bin、mysql_slow
电子邮件	exim_main、exim_reject、postfix_syslog、sendmail_syslog、procmail

操作系统	linux_messages_syslog、linux_secure、linux_audit、linux_bootlog、anaconda、anaconda_syslog、osx_asl、osx_crashreporter、osx_crash_log、osx_install、osx_secure、osx_daily、osx_weekly、osx_monthly、osx_window_server、windows_snare_syslog、dmesg、ftp、ssl_error、syslog、sar、rpmpkgs
Metrics	collectd_http、metrics_csv、statsd
网络	novell_groupwise、tcp
打印机	cups_access、cups_error、spooler
路由器和防火墙	cisco_cdr、cisco:asa、cisco_syslog、clavister
VoIP	asterisk_cdr、asterisk_event、asterisk_messages、asterisk_queue
Web 服务器	access_combined、access_combined_wcookie、access_common、apache_error、iis*
Splunk 软件	splunk_com_php_error、splunkd、splunkd_crash_log、splunkd_misc、splunkd_stderr、splunk-blocksignature、splunk_directory_monitor、splunk_directory_monitor_misc、splunk_search_history、splunkd_remote_searches、splunkd_access、splunkd_ui_access、splunk_web_access、splunk_web_service、splunkd_conf*、django_access、django_service、django_error、splunk_help、mongod
非日志文件	csv*、psv*、tsv*、_json*、json_no_timestamp、fs_notification、exchange*、generic_single_line
其他	snort、splunk_disk_objects*、splunk_resource_usage*、kvstore*

带星号 (*) 标记的来源类型使用 INDEXED_EXTRACTIONS 属性，该属性将 props.conf 中的其他属性设置为特定默认值，且要求特殊处理以转发至另一个 Splunk 平台实例。请参阅“转发从结构化数据文件中提取的字段”。

了解来源类型配置

若想了解 Splunk 平台在为给定的来源类型建立索引时使用了哪些配置信息，您可以使用 btool 实用工具显示转发器上的属性。如果您使用的是 Splunk Enterprise，则可以在 Splunk Enterprise 实例上执行此操作。

更多有关使用 btool 的信息，请参阅*《故障排除手册》*中的“使用 btool 排除配置故障”。

以下示例向您展示如何列出 tcp 来源类型的配置：

```
$ ./splunk btool props list tcp
[tcp]
BREAK_ONLY_BEFORE = (=\+)+
BREAK_ONLY_BEFORE_DATE = True
CHARSET = UTF-8
DATETIME_CONFIG = /etc/datetime.xml
KV_MODE = none
LEARN_SOURCETYPE = true
MAX_DAYS_AGO = 2000
MAX_DAYS_HENCE = 2
MAX_DIFF_SECS_AGO = 3600
MAX_DIFF_SECS_HENCE = 604800
MAX_EVENTS = 256
MAX_TIMESTAMP_LOOKAHEAD = 128
MUST_BREAK_AFTER =
MUST_NOT_BREAK_AFTER =
MUST_NOT_BREAK_BEFORE =
REPORT-tcp = tcpdump-endpoints, colon-kv
SEGMENTATION = inner
SEGMENTATION-all = full
SEGMENTATION-inner = inner
SEGMENTATION-outer = foo
SEGMENTATION-raw = none
SEGMENTATION-standard = standard
SHOULD_LINEMERGE = True
TRANSFORMS =
TRANSFORMS-baindex = banner-index
TRANSFORMS-dlindex = download-index
TRUNCATE = 10000
maxDist = 100
pulldown_type = true
```

基于每个事件覆盖来源类型

您可以在 Splunk 平台上基于每个事件覆盖来源类型，方法是使用重型转发器将事件分配给新的来源类型并将这些事件发送到 Splunk Cloud。在 Splunk Enterprise 上，您可以直接在实例本身上覆盖来源类型。

此来源类型分配发生在分析时，在平台完成其初始来源类型分配之后。有关该过程的更多信息，请参阅“来源类型为何重要”中的“Splunk 平台如何分配来源类型”。

由于此类覆盖在分析时发生，因此，覆盖仅在索引器或重型转发器上有效。在通用转发器上或直接在 Splunk Cloud 上则无效。若想了解在输入、分析、新建索引过程中不同时刻可使用哪些配置，请参阅*《管理员手册》*中的“配置参数和数据管道”。

要逐一配置事件覆盖，请结合使用 transforms.conf 和 props.conf 配置文件来指定必须使用新来源类型的事件以及事件必须使用的来源类型。

若想了解如何为来自特定输入或采用特定来源的事件数据配置基础来源类型覆盖，请参阅“覆盖来源类型自动分配”。

配置

要逐一配置事件覆盖，需要新建两个段落，一个位于 transforms.conf 文件中，另一个位于 props.conf 文件中。编辑 $SPLUNK_HOME/etc/system/local/ 目录或 $SPLUNK_HOME/etc/apps/ 中您自己的自定义应用程序目录中的这些文件。更多有关配置文件的一般信息，请参阅*《管理员手册》*中的“关于配置文件”。

编辑 transforms.conf 文件

1. 打开 $SPLUNK_HOME/etc/system/local/transforms.conf 文件进行编辑。
2. 在 transforms.conf 中新建一个遵循以下语法的段落：

```
[<unique_stanza_name>]
REGEX = <your_regex>
FORMAT = sourcetype::<your_custom_sourcetype_value>
DEST_KEY = MetaData:Sourcetype
```

3. 保存文件并将其关闭。

在此文件中，设置的含义如下：

- <unique_stanza_name> 意味着其涉及到的来源类型。稍后将在 props.conf 段落中使用此名称。
- <your_regex> 是一个正则表达式，标识了要应用自定义来源类型的事件（例如，采用特定主机名或其他字段值的事件）。
- <your_custom_sourcetype_value> 是要应用到 <your_regex> 选择的事件的来源类型。

您可以在搜索中将正则表达式与 rex 搜索命令结合使用，以对表达式进行测试。请参阅*《搜索参考》*中的“rex”。

编辑 props.conf 文件

1. 打开 $SPLUNK_HOME/etc/system/local/props.conf。
2. 在 props.conf 文件中新建一个段落来引用您在 transforms.conf 文件中指定的段落：

```
[<spec>]
TRANSFORMS-<class> = <unique_stanza_name>
```

请参阅下表以了解本段落中每个占位符变量的含义：

占位符变量	描述
spec	可以设置为以下选项： ◦ <sourcetype>，或事件的来源类型 ◦ host::<host>，其中 <host> 是事件的主机值 ◦ source::<source>，其中 <source> 是事件的来源值
<class>	任何您想要为转换指定的唯一标识符
<unique_stanza_name>	在 transforms.conf 中创建的段落的名称

3. 保存文件并将其关闭。
4. 重新启动 Splunk 平台实例。

示例：为来自单个输入但不同主机的事件分配来源类型

假设有一个共享 UDP 输入 UDP514。Splunk 平台实例通过此输入索引来自多个主机的大量数据。已确定您需要将一个名为 my_log 的特定来源类型应用到来源于三个特定主机（host1、host2 和 host3）且通过 UDP514 输入到达实例的数据。

您可以使用 Splunk 软件为 syslog 事件提取主机字段时常用的正则表达式开始操作。您可以在 $SPLUNK_HOME/etc/system/default/transforms.conf 中找到它：

```
[syslog-host]
REGEX = :\d\d\s+(?:\d+\s+|(?:user|daemon|local.?)\.\w+\s+)*\[?(\w[\w\.\-]{2,})\]?\s
FORMAT = host::$1
DEST_KEY = MetaData:Host
```

您可以修改此正则表达式，这样就可以把匹配范围缩小到来自目标主机名的事件。在此示例中，主机名为 host1、host2 和 host3：

```
REGEX = :\d\d\s+(?:\d+\s+|(?:user|daemon|local.?)\.\w+\s+)*\[?(host1|host2|host3)[\w\.\-]*\]?\s
```

现在，您可以在转换中使用修改过的正则表达式，该转换会把 my_log 来源类型应用到来自那三个主机的事件：

```
[set_sourcetype_my_log_for_some_hosts]
REGEX = :\d\d\s+(?:\d+\s+|(?:user|daemon|local.?)\.\w+\s+)*\[?(host1|host2|host3)[\w\.\-]*\]?\s
FORMAT = sourcetype::my_log
DEST_KEY = MetaData:Sourcetype
```

然后可以在 props.conf 段落中指定该转换，标识事件的特定输入：

```
[source::udp:514]
TRANSFORMS-changesourcetype = set_sourcetype_my_log_for_some_hosts
```

新建来源类型

您可以通过以下几种方式在 Splunk 平台上新建来源类型：

- 使用 Splunk Web 中的“设置来源类型”页面作为添加数据的一部分。
- 按照“添加来源类型”中的说明，在“来源类型”管理页面中新建一个来源类型。
- 编辑 props.conf 配置文件。除非您在通用转发器上定义来源类型并将其发送到 Splunk Cloud，否则此选项在 Splunk Cloud 上不可用。

尽管通过编辑驻留在转发器上的配置文件也可以配置单个转发器，实现新建来源类型的目的；但是，新建来源类型的最佳做法是使用 Splunk Web，该方法可以确保 Splunk 平台部署上的来源类型前后一致。

设置来源类型作为在 Splunk Web 中新建数据导入的一部分

Splunk Web 中的“设置来源类型”页面允许您查看为数据应用来源类型的效果。还允许您根据需要调整来源类型设置。您可以将更改另存为新的来源类型，然后将其分配给数据导入。

此页面允许您对**时间戳**和**事件**分隔符执行大多数常见的调整操作。对于其他修改，则可以直接编辑基础的 props.conf 文件。更改设置时，您可以立即查看更改如何影响事件数据。

该页面仅当您指定或上传单个文件时显示。当您指定任何其他类型的数据来源时，不会显示该页面。

要了解有关“设置来源类型”页面以及如何为数据分配来源类型的更多信息，请参阅“将正确的来源类型分配给您的数据”。

您还可使用“来源类型”管理页面新建来源类型。请参阅“添加来源类型”。

编辑 props.conf 配置文件以新建来源类型

如果您使用的是 Splunk Enterprise，则可以通过编辑 props.conf 配置文件和添加新来源类型段落来新建来源类型。有关 props.conf 文件的详细信息，请参阅 Splunk Enterprise*《管理员手册》*中的 props.conf 规范。有关配置文件的一般信息，请参阅 Splunk Enterprise*《管理员手册》*中的“关于配置文件”。

以下条目是 props.conf 文件中一个条目的示例。该条目先定义 access_combined 来源类型，然后再把该来源类型分配给与指定来源匹配的文件。您可通过使用正则表达式在数据来源中配置多个文件或目录。

```
[access_combined]
pulldown_type = true
```

```
maxDist = 28
MAX_TIMESTAMP_LOOKAHEAD = 128
REPORT-access = access-extractions
SHOULD_LINEMERGE = False
TIME_PREFIX = \[
category = Web
description = National Center for Supercomputing Applications (NCSA) combined fo
rmat HTTP web server logs (can be generated by apache or other web servers)

[source::/opt/weblogs/apache.log]
sourcetype = access_combined
```

要编辑 props.conf 文件，请执行以下步骤：

1. 在要新建来源类型的计算机上，如果 $SPLUNK_HOME/etc/system/local/props.conf 文件尚不存在，请新建该文件。

   您可能需要创建本地目录。如果您使用某个应用，请转到 $SPLUNK_HOME/etc/apps 目录中的应用。

2. 使用文本编辑器打开 $SPLUNK_HOME/etc/system/local 目录中的 props.conf 文件。
3. 为新的来源类型添加一个段落并指定 Splunk 软件处理来源类型时将使用的设置。

   ```
   [my_sourcetype]
   setting1 = value
   setting2 = value
   ```

   有关设置列表，请参阅 Splunk Enterprise*《管理员手册》*中的 props.conf 规范。

4. （可选）若您知道来源类型将被应用于其中的文件的名称，请在 [source::<source>] 段落中指定这些文件：

   ```
   [my_sourcetype]
   setting1 = value
   setting2 = value
   <br>
   [source::.../my/logfile.log]
   sourcetype = my_sourcetype
   ```

5. 保存 props.conf 文件。
6. 重新启动 Splunk Enterprise。新来源类型在重新启动完成后即生效。

指定事件换行和时间戳

当您新建来源类型时，需要指定一些重要设置：

- 事件换行：要了解如何使用 props.conf 文件指定事件换行，请参阅“配置事件换行”。
- 时间戳：要了解如何使用 props.conf 文件指定时间戳，请参阅“配置时间戳识别”以及本手册中“配置时间戳”一章中的其他主题。

您也可以为事件换行和时间戳配置许多其他设置。有关更多信息，请参阅 Splunk Enterprise*《管理员手册》*中的 props.conf 规范。

管理来源类型

在“来源类型”页面上新建、编辑和删除来源类型。要访问 Splunk Web 中的“来源类型”页面，请转至**设置 > 来源类型**。尽管此页面和“设置来源类型”页面名称相似，但这些页面提供不同的功能。

“来源类型”页面显示已在 Splunk® Enterprise 实例上配置的所有来源类型。页面上会显示您的部署提供的默认来源类型和您添加的所有来源类型。

对来源类型进行排序

默认情况下，“来源类型”页面按字母顺序对来源类型进行排序。您可通过单击“名称”、“类别”和“应用”列的标题栏更改页面排序的方式。

每个标题栏（除了“操作”）都用作一次切换。单击一次将按升序排序，再次单击将按降序排序。

筛选来源类型

您可对您在“来源类型”页面上查看的来源类型数目进行筛选。

按共性筛选

要只查看最常见的来源类型，请单击页面顶部**只显示常用**复选框。常用来源类型是最常见的来源类型。他们有 `pulldown_type "1"` 的来源类型字段值。未选择**只显示常用**对话框时，页面显示在实例上定义的所有来源类型。

按类别筛选

要仅查看属于某个类别的来源类型，请单击**类别**下拉列表并选择您想要的类别。只有属于该类别的来源类型才会显示。要再次查看所有来源类型，请从“类别”下拉列表中选择**全部**。

按应用程序上下文筛选

要仅查看属于某个应用程序上下文中的来源类型，请单击**应用**下拉列表并选择该来源类型所应用到的应用程序上下文。仅应用到该应用程序上下文的来源类型会显示。要再次查看所有来源类型，请从“应用”下拉列表中选择**全部**。

按字符串筛选

要仅查看名称中包含某个特定字符串的来源类型，请在“应用”下拉列表旁的**筛选器**文本框中键入该字符串，然后按 **Enter** 键。仅名称或描述与您已在“筛选器”框中键入的内容相匹配的来源类型会显示。要再次查看所有来源类型，请单击“筛选器”文本框右侧的 **x** 按钮。

默认情况下，“来源类型”管理页面最多在一个页面上显示 20 个来源类型。如果您想要查看更多或更少来源类型，请单击页面右侧的**每页面 20** 并选择您想要查看的来源类型数目。选项为一个页面上有 10 个、20 个、50 个或 100 个列表。

修改来源类型

要修改来源类型，请在列表中单击其名称，或在**操作**列单击**编辑**。随即显示“编辑来源类型”页面。

编辑来源类型对话框允许您更改来源类型的配置。您可以更改以下字段：

字段	描述
描述	来源类型的描述。
目标应用	来源类型应用到的应用程序上下文。 您无法更改 Splunk 软件附带的来源类型的应用目标。
类别	来源类型作为其成员的类别。单击该按钮从类别列表中选择所需的类别。当您保存时，来源类型会在您选择的类别中显示。 **Log to Metrics** 来源类型类别是专门用于将日志事件转换为指标数据点。请参阅本主题后面的“关于 Log to Metrics 来源类型类别”。
索引提取	在索引时间使用结构化数据从文件中提取字段的格式。请为已建立索引的提取选择一个最能代表文件内容的类型： • none：文件不包含结构化数据。 • json：文件是 JavaScript Object Notation (JSON) 格式。 • csv：文件有逗号分隔值。 • tsv：文件有制表符分隔值。 • psv：文件有管道符 (|) 分隔值。 • w3c：文件符合万维网联盟 (W3C) 日志格式。 • field_extraction：此文件包含非结构化事件，其中带有 <field>=<value> 格式的字段。
时间戳	为源自来源文件的事件确定时间戳的方式。请参阅本主题后面的“时间戳”部分。
高级	显示该来源类型的所有配置（按键/值格式）。请参阅本主题后面的“高级”部分。

时间戳

对话框的“时间戳”部分控制着 Splunk 软件为源自来源文件的事件确定时间戳的方式。您可以选择以下选项：

- 自动：使用默认的逻辑从事件中提取时间戳。
- 当前时间：使用当前的系统时间。
- 高级：使用高级内部逻辑从事件中提取时间戳。

当您选择“时间戳提取”部分中的**高级**时，以下高级配置可用：

- 时区：指定要分配给时间戳的时区。
- 时间戳格式：指定来源事件中的时间戳格式。可用格式来自 `strptime()` 编程功能的属性。

例如，假设来源文件包含的日志具有此格式的时间戳：

```
6 Jun 2015 18:35:05
```

则可在**时间戳格式**字段中指定以下内容：

```
%d %b %Y %H:%M:%S
```

考虑以下示例：

```
Tue Jun 4 2:55:18 PM 2015
```

如果来源文件包含的日志具有该格式的时间戳，则在**时间戳格式**字段中指定以下内容：

```
%a %b %d %I:%M:%S %p %Y
```

对于您可以用来定义时间戳格式的字符串列表，请参阅 die.net Linux 站点上 https://linux.die.net/man/3/strptime 中的 strptime(3)。

时间戳前缀：一个正则表达式，代表在时间戳之前显示的字符。当 Splunk® Enterprise 在事件中查看该字符集时，它期望时间戳在这之后发生。

提前量：指定在事件中查找时间戳时会查看的字符数上限。如果您在“时间戳前缀”字段中指定了一个正则表达式，则指定的字符数未超过正则表达式为时间戳所代表的字符串。

高级

“高级”部分向您显示该来源类型的所有配置（按键/值格式）。这代表了定义该来源类型的 props.conf 配置文件中的内容。您可直接编辑每个设置或添加与删除设置。

请慎重使用“高级”部分。在此添加或更改值会导致数据的索引建立过程出错。

要删除设置，请单击每个设置右侧的 x。要添加条目，请单击对话框底部的**新建设置**以显示字段的键/值对。请在**名称**字段中输入键名称并在**值**字段中输入键值。

关于 Log to Metrics 来源类型类别

Log to Metrics 来源类型类别是用于在引入时将日志事件转换为指标数据点。如果您选择此来源类型类别，将显示**指标**选项卡。关于此来源类型和“指标”选项卡中的字段的更多信息，请参阅*指标*中的“将事件日志转换为指标数据点”。

添加来源类型

如需新建来源类型，请执行以下步骤：

1. 请单击屏幕右上方的**新建来源类型**按钮。
   新建来源类型对话框打开。
2. 按照相同的步骤修改现有来源类型以添加来源类型。有关您可以配置的字段的信息，请参阅本主题前面的“修改来源类型”。
3. 当您完成配置来源类型时，请单击**保存**。

删除来源类型

您只能删除您新建的或应用随附的来源类型。无法删除内置来源类型。

删除来源类型会造成重大影响，尤其是已经在使用中的来源类型。

在继续之前考虑以下后果：

- 删除来源类型后，建立的数据索引可能不正确。使数据可按您后续想要的方式进行搜索会产生很大效果。许多应用和加载项使用来源类型查找数据，在缺失的来源类型下得以索引的数据是那些应用和加载项不会查看的数据。

- 该来源类型使用的任何配置（例如，字段提取、索引时间筛选和时间戳格式）都已永久性丢失。
- 您无法撤销删除来源类型。在这种情况下，仅可从备份中恢复定义了来源类型的 props.conf 文件，或手动重新创建来源类型。

要删除来源类型，请执行以下步骤：

1. 在要删除的来源类型的“操作”列中单击**删除**。您无法删除内置的来源类型，只能删除您新建的或应用随附的来源类型。将出现一个对话框，询问您是否确认删除。
2. 如果您确定要删除该来源类型，请单击**删除**。

该操作将关闭对话框，而且 Splunk Web 会带您返回至“来源类型”管理页面。

搜索时间重命名来源类型

您可能会遇到要重命名来源类型的情况。例如，假设您将输入意外分配给错误的来源类型。或者意识到在搜索时间实际上两个名称不同的来源类型需要按照相同的方式处理。

如果使用的是 Splunk Enterprise，您可以添加 rename 设置（位于 props.conf 配置文件中）在搜索时间为事件分配新的来源类型。如果您需要对其进行搜索，Splunk Enterprise 会将原始的来源类型移动至名为 _sourcetype 的单独字段。

在 Splunk Cloud 上，您必须打开支持工单才能重命名来源类型，因为在 Splunk Cloud 实例上无法编辑 props.conf 文件，并且无法使用重型转发器，因为重命名来源类型仅适用于您已创建索引的数据。

索引的事件仍包含原始来源类型名称。来源类型的重命名操作只在搜索时间发生。此外，重命名来源类型只会进行重命名。而不会解决因开始分配了错误的来源类型而导致的任何事件数据索引格式问题。

要重命名来源类型，请将 rename 设置添加到 props.conf 文件中的来源类型段落：

```
rename = <string>
```

一个来源类型名称只能包含从 a 到 z 的字母、从 0 到 9 的数字和冒号（ : ）字符以及下划线（ _ ）字符。

例如，假设您对应用程序服务器使用来源类型 cheese_shop。然后意外将大量数据索引为来源类型 whoops。您可以在 props.conf 文件中使用以下段落，将 whoops 重命名为 cheese_shop：

```
[whoops]
rename=cheese_shop
```

现在，搜索 cheese_shop 时将返回所有 whoops 事件以及具有 cheese_shop 来源类型的所有事件：

```
sourcetype=cheese_shop
```

如果需要单独列出 whoops 事件，您可以在搜索中使用 _sourcetype：

```
_sourcetype=whoops
```

来自重命名的来源类型的数据仅使用目标来源类型的搜索时间配置（本例中为 cheese_shop）。Splunk 平台将忽略原始来源类型的所有字段提取。

管理事件分段

关于事件分段

事件分段功能可以在**索引**时间和**搜索**时间将事件划分成可搜索的**段**。段可以分类为主要段或次要段。次要段在主要段内进行拆分。例如，IP 地址 192.0.2.223 是一个主要段。但是，该主要段可以拆分成次要段，例如 192 或 0 以及 192.0.2 之类的次要段组。

可以定义事件分段的详细程度。这非常重要，因为索引时间分段影响索引和搜索速度、存储大小以及使用键盘缓冲功能的能力，其中 Splunk Web 提供与键入搜索栏的文本相匹配的项目。另一方面，搜索时间分段影响搜索速度以及通过从 Splunk Web 的显示结果选择项目来新建搜索的能力。

有关索引时间和搜索时间区别的更多信息，请参阅*《管理索引器和群集》*手册中的“索引时间对比搜索时间”。

可以按照“设置事件数据的分段”所述，在 props.conf 中将分段分配给特定类别的事件。

如果使用的是 Splunk Cloud，则在重型转发器计算机上配置索引时间分段。若要配置搜索时间分段，则必须提交支持工单。

如果您使用的是 Splunk Enterprise，则既可以在索引器或重型转发器计算机上配置索引时间分段，也可以在搜索头上配置搜索时间分段。

事件分段的类型

分段有三种主要类型或级别，您可以在索引或搜索时间内对其进行配置。您还可以禁用分段。位于 $SPLUNK_HOME/etc/system/default 中的 segmenters.conf 配置文件定义所有可用的分段类型。根据默认设置，索引时间分段被设置为 indexing 类型，此类型结合了内部分段和外部分段。搜索时间分段设置为完全分段。

内部分段

内部分段将事件拆分成可能的最小次要段。例如，若诸如 192.0.2.223 等 IP 地址进行内部分段，会被拆分成 192、0、2 和 223。在索引时间内设置内部分段会加快索引和搜索速度并减少磁盘使用量。然而，它对键盘缓冲功能有所限制，以便用户只能在次要段级别进行预先输入。

外部分段

外部分段与内部分段相反。在外部分段的情况下，Splunk 平台仅为主要段建立索引。例如，若 IP 地址 192.0.2.223 以 192.0.2.223 的形式列入索引，意味着您无法搜索短语的各个单独部分。不过，您仍可以使用通配符搜索短语的各部分。例如，若搜索 192.0*，您将获得任何 IP 地址以 192.0 开头的事件。此外，外部分段将禁用单击搜索结果不同段的能力，例如同一个 IP 地址的 192.0 段。外部分段往往或多或少地比完全分段高效些，而内部分段往往高效得多。

完全分段

完全分段是内部和外部分段的组合。完全分段的情况下，系统在为 IP 地址新建索引时会包含一个主要段和多个次要段，包括诸如 192.0 和 192.0.2 等次要段组合。这是效率最低的索引选项，但在搜索方面的功能却最强。

无分段

最节省空间的分段设置是完全禁用分段。但是，这对搜索有巨大影响。通过禁用分段，您将把搜索限制在索引字段内，如时间、数据来源、主机和来源类型。搜索关键字将不返回任何结果。必须通过搜索命令传送搜索以进一步限制结果。请参阅*《搜索参考》*中的“搜索”。此设置仅在不需要任何高级搜索功能的情况下使用。

配置分段类型

segmenters.conf 定义分段类型。如有必要，可以定义自定义分段类型。

有关默认提供的分段类型的信息，请参阅 $SPLUNK_HOME/etc/system/default 中的 segmenters.conf 文件。

不要修改默认文件。如果要更改现有的分段段落或完全创建新段落，可以在 $SPLUNK_HOME/etc/system/local/ 目录中的文件或 $SPLUNK_HOME/etc/apps/ 中的自定义应用目录中指定要更改的设置。

设置特定主机、来源或来源类型的分段类型

可以配置要应用于特定主机、来源或来源类型的索引时间和搜索时间分段。如果您定期运行涉及特定来源类型的搜索，则可以使用此功能改善这些搜索的性能。类似地，若您经常为大量的 syslog 事件新建索引，可以利用此功能来减少这些事件占用的总体磁盘空间。

有关如何将分段类型应用于特定事件类别的信息，请参阅“设置事件数据的分段”。

设置事件数据的分段

默认情况下，Splunk 软件会在索引期间将事件分段，这样才能执行灵活性最大的搜索操作。有许多可用的分段类型，必要时，您可以新建其他类型。您采用的分段类型影响索引速度、搜索速度以及索引占用的磁盘空间量。要了解分段以及各类型分段之间权衡的更多信息，请参阅“关于分段”。

Splunk 软件也可以在搜索时间内将事件分段。可以按照“在 Splunk Web 中设置搜索时间分段”所述，在 Splunk Web 中设置搜索时间分段。

如果您知道想要如何搜索或处理来自特定主机、来源或来源类型的事件，则可以为此特定类型事件配置索引时间分段。也可以为特定类型的事件配置搜索时间分段选项。

在 props.conf 中指定分段

通过将分段类型分配给 props.conf 中的适当段落，为特定主机、来源或来源类型的事件指定分段。在段落中，可以分配已在 segmenters.conf 中定义的分段类型（或“规则”）。这些可以是预定义类型（如内部、外部或完全），也可以是您定义的自定义类型。有关定义自定义类型的更多信息，请阅读“配置分段类型”。

在 props.conf 中配置的使用这些类型的属性取决于您在配置的是索引时间分段还是搜索时间分段：

- 若是索引时间分段，请使用 SEGMENTATION 属性。
- 若是搜索时间分段，则使用 SEGMENTATION-<segment selection> 属性。

可以在段落中定义这两个属性之一或全部。

将您的段落添加至 $SPLUNK_HOME/etc/system/local/props.conf。

索引时间分段

SEGMENTATION 属性决定着索引时间使用的分段类型。语法如下：

```
[<spec>]
SEGMENTATION = <seg_rule>
```

[<spec>] 可以为：

- <sourcetype>：事件数据的来源类型。
- host::<host>：事件数据的主机值。
- source::<source>：事件数据的来源。

SEGMENTATION = <seg_rule>

- 这将为 [<spec>] 事件指定索引时间使用的分段类型。
- <seg_rule>
  - 定义的分段类型或“规则”的位置为 segmenters.conf
  - 常见设置有 inner、outer、none 和 full，但默认文件中也包含其他预定义的分段规则。
  - 根据“配置分段类型”中的说明，通过编辑 $SPLUNK_HOME/etc/system/local/segmenters.conf 新建您自己的自定义规则。

搜索时间分段

SEGMENTATION-<segment_selection> 属性有助于决定搜索时间使用的分段类型。语法如下：

```
[<spec>]
SEGMENTATION-<segment_selection> = <seg_rule>
```

[<spec>] 可以为：

- <sourcetype>：事件数据的来源类型。
- host::<host>：事件数据的主机值。
- source::<source>：事件数据的来源。

SEGMENTATION-<segment_selection> = <seg_rule>

- 这为 [<spec>] 事件在 Splunk Web 中指定搜索时间使用的分段类型。

- <segment_selection> 可以是以下各项中的其中一项：full、inner、outer 或 raw。
  - 这四个值是在**结果显示选项**面板（从 Splunk Web 中搜索结果正上方的**选项**直接调用）的**事件分段**下拉框中显示的一组选项。
  - 请注意，这些值只是一组可用的 Splunk Web 下拉选项。可以使用此属性指定选项调用的实际分段类型，这可能与下拉选项本身的名称不同。例如，您甚至可以定义“内部”下拉选项来调用“外部”分段类型，而非您可能希望的那样。
  - 如“在 Splunk Web 中设置搜索时间分段”中所述，通过把下拉选项映射到 <seg_rule>，用户稍后可以在查看搜索结果时指定该选项，由此来设置搜索时间分段。

- <seg_rule>
  - 定义的分段类型或“规则”的位置为 segmenters.conf
  - 常见设置有 inner、outer、none 和 full，但默认文件中也包含其他预定义的分段规则。
  - 根据“配置分段类型”中的说明，通过编辑 $SPLUNK_HOME/etc/system/local/segmenters.conf 新建您自己的自定义规则。

示例

本示例同时为 syslog 事件设置索引时间和搜索时间分段规则。

把以下内容添加到 [syslog] 来源类型段落（位于 props.conf 内）：

```
[syslog]
SEGMENTATION = inner
SEGMENTATION-full= inner
```

本段落把所有来源类型为 syslog 的事件的索引时间分段更改为内部分段。这还会导致 Splunk Web 中的 full 单选按钮调用相同事件的内部分段。

注意：必须重新启动 Splunk Enterprise，这样才能将更改应用于搜索时间分段。必须为数据重新建立索引，以将索引时间分段更改应用于现有数据。

在 Splunk Web 中设置搜索时间事件分段

Splunk Web 允许您为搜索结果设置分段。虽然设置分段与索引时间分段无关，但 Splunk Web 搜索时间分段影响浏览器交互，而且会加速搜索结果。

要设置搜索结果分段，请执行以下步骤：

1. 执行搜索并查看结果。
2. 单击返回的事件集上方的**格式**。
3. 在**单击选择**下拉菜单中，从可用选项中选择：**完整**、**内部**或**外部**。

可以按照“设置事件数据的分段”所述，配置这些下拉菜单选项的含义。

改善数据导入过程

使用测试索引测试输入

将新输入添加到 Splunk® Enterprise 实例上的生产索引之前，最好先通过将输入添加到测试索引来对其进行测试。验证接收的数据正确且生成的事件格式可用后，可以将输入配置为生产索引。您可以继续在不同时段以此方法测试新输入。如果您发现起始输入并非所需，则可以一直处理测试索引，直至对结果满意。

您还可以预览 Splunk® Enterprise 对数据建立索引到测试索引中的方式。在预览期间，可交互调整事件处理设置。有关详细信息，请参阅“将正确的来源类型分配给您的数据”。

使用测试索引

1. 使用 Splunk Web，或者如果您有 Splunk Enterprise，也可以使用 CLI 来新建测试索引；编辑 indexes.conf 配置文件亦可新建测试索引。
2. 配置数据导入时，将事件发送到测试索引。可以在 Splunk Web 中执行此操作。对于每个输入，执行以下步骤：
    1. 从**添加数据**页面配置输入时，请选中**更多设置**选项。它会显示几个新字段，包括一个名为**索引**的字段。
    2. 从**索引**下拉列表中，选择您的测试索引。该数据输入的所有事件现在都将转到此索引。
    3. 对要发送到测试索引的每个数据输入都重复此过程。
3. 搜索时，在搜索命令中指定测试索引。默认情况下，Splunk® Enterprise 对 **main** 索引进行搜索。使用 index= 命令，如下所示：

```
index=test_index
```

若要搜索事件的测试索引，而该些事件又来自您新建的输入，请将字段边栏的时间范围更改为“实时 > 所有时间（实时）”。生成的实时搜索将显示写入到此索引的所有事件，而不管其提取的时间戳值如何。若您正在为进入您的索引的历史数据新建索引，而您的索引不会显示在“上一个小时”或“实时 > 30 分钟窗口”搜索的搜索结果内，此结果尤其有用。

要了解如何通过 Splunk Web 新建和使用自定义索引，请参阅 Splunk Enterprise *《管理索引器和索引器群集》*手册中的“新建自定义索引”。也可以在 inputs.conf 配置文件中配置输入时指定索引。

删除索引数据并重新开始

若想清除测试索引并在 Splunk Enterprise 上重新开始，请使用 CLI clean 命令，如 Splunk Enterprise *《管理索引器和索引器群集》*手册中的“删除索引和索引的数据”所述。

配置输入以使用默认索引

对测试结果满意并准备真正开始新建索引时，编辑数据导入以使其将数据发送至默认的 **main** 索引，而非测试索引。对于已经设置的每个数据输入，按照设置测试索引的相反步骤进行操作：

1. 返回到最初配置输入的位置。例如，如果您从 Splunk Web 中的**添加数据**页面配置输入，则返回到此输入的配置屏幕：
    1. 选择**系统 > 系统配置 > 数据导入**。
    2. 选择输入的数据类型以查看此类型所有已配置输入的列表。
    3. 选择要编辑的特定数据输入。选择输入会将您带到可对其进行编辑的屏幕。
    4. 选择**显示高级设置**选项。转到名为**索引**的字段。
    5. 在**索引**下拉列表中，选择 **main** 索引。该数据输入的所有事件现在都将转到此索引。

若您改用 inputs.conf 文件配置输入，可以直接在该文件中更改索引，如 Splunk Enterprise *《管理索引器和索引器群集》*手册中的“创建自定义索引”所述。

现在，您进行搜索时，无需再在搜索命令中指定索引。默认情况下，Splunk 软件会搜索 **main** 索引。

使用保留队列帮助防止数据丢失

保留队列使您可以将输入队列中的数据存储到磁盘。在 Splunk Cloud 部署中，如果您配置为将数据发送到 Splunk Cloud 实例的**转发器**备份，保留队列可以帮助防止数据丢失。在 Splunk Enterprise 部署中，保留队列适用于转发器或**索引器**。无法直接在单个 Splunk Cloud 实例中配置保留队列。

默认情况下，转发器和索引器有 500 KB 的内存中输入队列。如果输入流的运行速率高于转发器或索引器能够处理的速率，当转发器上的输入队列达到最大限制时会出现不希望的后果。在您通过 UDP 协议发送网络数据的情况下，该数据会在队列中减少且会丢失。对于其他类型的数据输入，可以备份生成数据的应用程序。

通过实施保留队列，可以帮助防止此数据删除或丢失的发生。使用保留队列时，在内存中队列已满后，转发器或索引器便会将输入流写入到磁盘上的文件中。然后，它处理来自内存中和磁盘队列的数据，直至达到能够再次直接从数据流开始处理的时间

点。

虽然在处理过程有备份的情况下，保留队列可以预防数据丢失，但如果转发器或索引器崩溃，您仍可能会丢失数据。例如，转发器会将一些输入数据保留在内存队列中以及保留在保留队列文件中。如果转发器崩溃，内存中数据将丢失。同样，在分析或索引管道中但尚未写入到磁盘的数据在出现崩溃时也会丢失。

何时可以使用保留队列？

保留队列可用于某些类型的输入，但并非全部。一般来讲，保留队列可用于短暂特性的输入（例如网络输入），但不可用于有其自己保留形式的输入（例如监视文件）。

保留队列可用于以下输入类型：

- 使用 TCP 协议的网络输入
- 使用 UDP 协议的网络输入
- 先进先出（FIFO）输入
- 脚本式输入
- Windows 事件日志输入
- HTTP 事件收集器标记

保留队列不可用于以下输入类型：

- 监视输入
- 批处理输入
- 文件系统更改监视器
- 来自 Splunk 转发器的 splunktcp 输入

配置保留队列

使用 inputs.conf 配置文件配置保留队列。您可以在您配置为将数据发送到 Splunk Cloud 的通用转发器上配置保留队列。您还可以在 Splunk Enterprise 索引器上配置保留队列。直接在将数据发送到索引器的索引器或转发器上使用相同的过程。

输入不共享队列。可以在特定输入的段落中配置保留队列。

1. 在将数据转发到 Splunk Cloud 的计算机上，使用文本编辑器打开 $SPLUNK_HOME/etc/system/local/inputs.conf 文件进行编辑。
2. 找到或添加要启用保留排队的输入段落。
3. 在该输入段落中指定以下设置：

   ```
   persistentQueueSize = <integer>(KB|MB|GB|TB)
   ```
4. 保存文件并将其关闭
5. 重新启动转发器。

有关 inputs.conf 文件的更多信息，请参阅 Splunk Enterprise*《管理员手册》*中的“inputs.conf”。

配置保留队列示例

下面是为 TCP 网络输入指定 100 MB 保留队列的示例：

```
[tcp://9994]
persistentQueueSize=100MB
```

下面是为 Windows 事件日志输入指定 1GB 保留队列的示例：

```
[WinEventLog]
persistentQueueSize=1GB
```

Windows 事件日志监视器仅接受默认 Windows 事件日志段落的保留队列配置。无法为特定事件日志通道配置保留队列。可以为特定的 Windows 主机监视输入配置保留队列。

保留队列位置

保留队列具有硬编码位置，该位置因输入类型而异。

对于网络输入，保留队列位于 $SPLUNK_HOME/var/run/splunk/[tcpin|udpin]/pq__<port>。

在文件名中放置两个下划线：pq__<port></port>，非 pq_<port></port>。

请参阅以下示例：

- TCP 端口 2012 的保留队列为 $SPLUNK_HOME/var/run/splunk/tcpin/pq__2012。
- UDP 端口 2012 的保留队列为 $SPLUNK_HOME/var/run/splunk/udpin/pq__2012。

对于 FIFO 输入，保留队列驻留在 $SPLUNK_HOME/var/run/splunk/fifoin/<encoded path> 中。

对于脚本式输入，保留队列驻留在 $SPLUNK_HOME/var/run/splunk/exec/<encoded path> 中。Inputs.conf 文件中的 FIFO 脚本式输入段落派生出 <encoded path>。

输入过程故障排除

以下是在 Splunk® Enterprise 上的数据导入过程中排除故障时可以采取的一些初步措施。

决定您为何找不到预期事件的原因

将某输入添加到您的 Splunk® Enterprise 部署时，该输入相对地添加到了您所处的应用中。一些应用会将输入数据写入到指定的索引中。如果您确定数据就在 Splunk® Enterprise 部署中却找不到，请确认您查找的索引是否正确。请参阅*《搜索手册》*中的“从索引中检索事件”。对于当前使用的角色，您可能需要把一些索引添加到默认索引列表中。

- 有关角色的详细信息，请参阅*《确保 Splunk 平台安全》*手册中的“添加和编辑角色”。
- 更多有关数据导入问题故障排除的信息，请参阅本主题的其他部分或查看*《故障排除手册》*中的“我无法找到我的数据！”

如果您使用的是 Splunk Enterprise 而且通过编辑 inputs.conf 配置文件添加了输入，Splunk Enterprise 可能无法立即识别添加的输入。自上次重启时开始，Splunk Enterprise 每隔 24 小时查找一次输入。所以如果您添加了一个新的段落来监视目录或文件，Splunk Enterprise 最晚会在 24 个小时之后开始为该目录或文件中的内容建立索引。要确保 Splunk Enterprise 立即识别您的输入并为其建立索引，请通过 Splunk Web 或 CLI 添加该输入，或在编辑 inputs.conf 文件后重启 Splunk 服务。

尾文件故障排除

您可以使用 **FileStatus** 表现层状态转换（REST）端点获取尾文件的状态。例如：

```
curl https://serverhost:8089/services/admin/inputstatus/TailingProcessor:FileStatus
```

您也可以监视 **fishbucket**，这是一个子目录，Splunk 软件使用该子目录追踪文件内容的索引建立进度。在 Splunk Enterprise 部署中，fishbucket 会驻留在 $SPLUNK_DB/fishbucket/splunk_private_db。如果是 Splunk Cloud 部署，您无法对该子目录进行物理访问。

要监视该 fishbucket，请使用 REST 端点。有关其他信息，请查看《REST API 参考手册》。

解决 Splunk Enterprise 上的引入阻塞问题

有时，由于未知原因，Splunk Enterprise 数据引入会变慢。导致速度变慢的其中一个可能因素是 Splunk Enterprise 索引器上可用的非活动输入通道数。

输入通道说明

索引器必须跟踪其处理的每个唯一数据流的状态。例如，索引器从一组尾文件中引入了一些数据，在为这些数据换行时，索引器会以一种无法预测的顺序从这些文件接收数据。各种文件的不同部分可以相互交错。索引器会通过一种名为输入通道的数据结构来跟踪每个文件的状态，以防止这种交错导致一个文件的换行干扰另一文件的换行。

输入通道存储了多种信息，包括以下信息：

- 断行器的状态
- 聚合器的状态
- punct 状态
- props.conf 中用于输入的设置

索引器遇到的每个来源、来源类型或主机流都有一个唯一的输入通道。

非活动输入通道说明

出于性能和内存使用率的原因，索引器不会永远保持输入通道。如果某个通道在一段时间未使用，例如，特定来源、来源类型

和主机元组的数据在一段时间后仍未出现，则该通道就可以由其他流重用。Splunk Enterprise 有多种设置可以控制非活动通道的回收行为。您可以在 limits.conf 配置文件中配置这些设置。

例如，假设索引器刚遇到一个新的流。因此，它需要一个输入通道以便在引入该流时可以保存该流的状态。此时，它必须决定是创建一个新的输入通道（但会耗用更多的内存），还是重用一个非活动通道（但如果该非活动通道变为活动状态时，则会导致性能下降）。

在决定是否使用非活动输入通道时，索引器会遵循以下决策过程：

1. 如果非活动通道数小于或等于 limits.conf 配置文件中为 lowater_inactive 设置所设的值，则创建一个输入通道。
2. 如果非活动通道数大于 limits.conf 文件中为 max_inactive 设置所设的值，或者最旧的非活动通道的未使用时间大于为 inactive_eligibility_age_seconds 所设的值，执行以下其中一个操作：
   - 回收最旧的非活动输入通道。
   - 创建一个输入通道。

换一种说法：

- 如果非活动通道数当前小于lowater_inactive 值，则索引器会创建一个新的输入通道。
- 如果非活动通道数当前大于max_inactive 值，则索引器会回收一个非活动输入通道。
- 如果非活动通道数大于 lowater_inactive 值却小于 max_inactive 值，则索引器会回收最旧的非活动输入通道，只要该通道的未使用时间大于 inactive_eligibility_age_seconds 秒。否则，它会创建一个新的输入通道。

现在，max_inactive 设置的设置值为 auto。这会将索引器配置为，根据运行实例的计算机中可用的内存量来调整 max_inactive 设置。

配置手动或自动非活动输入通道限制

您可以调整索引器可保留的最大非活动输入通道数。手动增加此数字会增加索引器所用的内存量。如果这个数字较低，就表示索引器占用的内存较少，但索引器需要创建的新输入通道数会增加，这可能会大幅降低索引器在处理传入数据时的性能，具体取决于索引器遇到的源、源类型和主机的数量。每个非活动输入通道占用大约 5KB 的内存。

1. 在要调整非活动输入通道限制的索引器上，打开 shell 提示符、命令提示符或文本编辑器。
2. 打开文件 $SPLUNK_HOME/etc/system/local/limits.conf 进行编辑。
3. 在该文件中找到 [input_channels] 段落。如果文件中没有这个段落，需创建一个。
4. 在 [input_channels] 段落下，添加以下行：

   ```
   max_inactive = <positive integer>
   ```

   如果您希望索引器自动管理非活动通道的数量，请将该行更改为

   ```
   max_inactive = auto
   ```
5. 保存文件并将其关闭。
6. 重新启动 Splunk Enterprise 实例以应用更改。

有关 max_inactive、lowater_inactive 和 inactive_eligibility_age_seconds 设置的更多信息，请参阅 limits.conf 规格文件。

找不到转发的数据？

确保转发器运行正常且对索引器可见。您可以使用分布式管理控制台 (DMC) 排除 Splunk 拓扑结构故障，并找到任何转发器故障的根源问题。有关详细信息，请参阅*监视 Splunk Enterprise*。

解决数据质量问题

当您将数据导入 Splunk 平台时，您可以解决以下事件处理和数据质量问题：

- 不正确的换行
- 不正确的事件换行
- 不正确的时间戳提取

换行问题

以下症状表明可能存在换行问题：

- 事件比您预期的少且事件较大，尤其是单行事件。
- “监视控制台**数据质量**仪表板”显示换行问题。
- 您可能在 Splunk Web 数据导入工作流或在 splunkd.log 文件中看到以下错误消息：“Truncating line because limit

of 10000 bytes has been exceeded（截断行，因为已超过 10000 字节的限制）”。

诊断

要确认您的 Splunk 平台是否有换行问题，请执行以下一项或多项故障排除步骤：

- 访问“监视控制台**数据质量**仪表板”。检查仪表板表格是否存在换行问题。请参阅*《监视 Splunk Enterprise》*中的“关于监视控制台”。
- 在 splunkd.log 文件中查找消息，如下所示：

```
12-12-2016 13:45:48.709 -0800 WARN LineBreakingProcessor - Truncating line because limit of 10000 bytes has been exceeded with a line length >= 301367
```

- 搜索事件。出现多个事件组合或单一事件拆分为多个事件则表明有换行问题。

解决方案

要解决换行问题，请在 Splunk Web 中完成以下步骤：

1. 单击**设置 > 添加数据**。
2. 单击**上载**以通过上载文件进行测试，或单击**监视**以重新进行监视输入。
3. 选择有数据示例的文件。
4. 单击**下一步**。
5. 在**设置来源类型**页面，使用左侧面板中的选项进行设置，直到您的示例数据正确分成事件。要配置 LINE_BREAKER 或 TRUNCATE，请单击**高级**。
6. 完成数据导入工作流或记录正确的设置，然后用他们纠正现有的导入配置。

使用“设置来源类型”页面中的选项进行设置时，LINE_BREAKER 设置可能不正确。LINE_BREAKER 设置必须有一个捕获组，且该组必须和事件匹配。

比如，您可能有一个不匹配的 LINE_BREAKER 值。在 splunkd.log 文件中查找“Truncating line because limit of 10000 bytes has been exceeded”的消息或在 Splunk Web 中查找以下消息：

如果您找到此消息，进行以下操作：

1. 确认 LINE_BREAKER 已按照您的预期正确配置为将您的数据分割为多行。
2. 确认您在 LINE_BREAKER 设置中指定的字符串存在于您的数据中。
3. 如果 LINE_BREAKER 配置正确但行很长，或如果您正在使用 LINE_BREAKER 作为定义事件的唯一方法，绕过索引管道后的合并行操作，确认 TRUNCATE 设置足够大，能够包含 LINE_BREAKER 拆分的整个数据片段。
   TRUNCATE 的默认值为 10,000。如果您的事件大于 TRUNCATE 值，您可能想要增加 TRUNCATE 值。考虑到性能和内存使用情况，请不要设置 TRUNCATE 为无限制。

如果您没有指定捕获组，LINE_BREAKER 会被忽略。

有关更多信息，请参阅“配置事件换行”。

事件换行或聚合问题

事件换行问题可能与 BREAK_ONLY_BEFORE_DATE 和 MAX_EVENTS 设置及任意具有关键字 BREAK 的 props.conf 配置文件设置相关。

如果您看到以下指示器，则可能存在聚合问题：

- “监视控制台**数据质量**仪表板”中显示的聚合问题。

- Splunk Web 数据导入工作流中的错误。
- 计算事件。如果事件缺失且较大，尤其是单行事件，可能会有事件换行问题

诊断

要确认 Splunk 平台是否有事件换行问题，请进行以下一项或多项故障排除步骤：

- 查看“监视控制台**数据质量**仪表板”。
- 搜索由多个事件组合而成的事件。
- 检查 splunkd.log 中的下列消息：
  ```
  12-07-2016 09:32:32.876 -0500 WARN AggregatorMiningProcessor - Breaking event because limit of 256 has been exceeded
  12-07-2016 09:32:32.876 -0500 WARN AggregatorMiningProcessor - Changing breaking behavior for event stream because
  MAX_EVENTS (256) was exceeded without a single event break. Will set BREAK_ONLY_BEFORE_DATE to False, and unset any
  MUST_NOT_BREAK_BEFORE or MUST_NOT_BREAK_AFTER rules. Typically this will amount to treating this data as single-line only.
  ```

解决方案

对于事件换行，确定是否因以下某个原因而发生这种问题：

- 您的事件已被正确识别，但是事件太大而超出限制。MAX_EVENTS 定义事件中的最大行数。
- 您的事件未被正确识别。

如果您的事件超出 MAX_EVENTS 设置的限制，您可提高限制。注意该大型事件对索引性能、搜索性能和资源使用方法而言不是最优的。大型事件搜索成本较高。两个限制的上限值将产生的字符数为每行 10,000 个字符（如 TRUNCATE 所定义）乘以 256 行（按 MAX_EVENTS 设置）。两个限制的组合是非常大的事件。

如果是由未被正确识别的事件导致（由这种原因造成的可能性更大），那么 Splunk 平台的事件换行不正常。检查以下内容：

- 事件换行策略。默认策略在计划日期之前换行，所以如果 Splunk 平台不提取时间戳，就不会进行事件换行。要诊断和解决此问题，请调查时间戳提取。请参阅“时间戳分配如何工作”。
- 您的事件换行正则表达式。

有关更多信息，请参阅以下主题：

- 调整时间戳识别以获得更佳的索引性能
- 配置事件换行

时间戳问题

时间戳问题可能与 props.conf 配置文件中的以下设置有关：

- DATETIME_CONFIG
- TIME_PREFIX
- TIME_FORMAT
- MAX_TIMESTAMP_LOOKAHEAD
- TZ

如果您看到以下指示器，则可能存在时间戳分析问题：

- 时间戳分析问题在“监视控制台**数据质量**仪表板”中显示。
- Splunk Web 数据导入工作流中出现错误。
- 计算事件。如果您缺失事件，且有大型事件，尤其是单行事件，分析可能是个问题。
- 时区分配不正确。
- 由 Splunk 平台分配的 _time 值和原始数据中的时间不匹配。

诊断

要确认您是否有时间戳问题，请进行以下一项或多项操作：

- 访问“监视控制台**数据质量**仪表板”。检查表中的时间戳分析问题。根据多种回退分配时间戳。有关更多详情，请参阅“时间戳分配如何工作”。对于大多数回退，即使其中一个成功分配时间戳，您仍然会在“监视控制台”仪表板中获得一个问题。
- 搜索由多个事件组合而成的事件。
- 查看 splunkd.log 文件的消息，如下所示：

```
12-09-2016 00:45:29.956 -0800 WARN  DateParserVerbose - Failed to parse timestamp. Defaulting to timestamp of previous event
(Fri Dec  9 00:45:27 2016). Context: source::/disk2/sh-demo/splunk/var/log/splunk/entity.log|host::svdev-sh-demo|entity-
too_small|682235
12-08-2016 12:33:56.025 -0500 WARN AggregatorMiningProcessor - Too many events (100K) with the same timestamp: incrementing
```

```
timestamps 1 second(s) into the future to insure retrievability
```

使用相同的时间戳索引所有事件，使搜索该时间范围无效。

解决方案

要解决时间戳问题，请完成以下步骤：

- 确保每个事件有完整的时间戳，包括年份、完整日期、完整时间和时区。
- 请参阅“配置时间戳识别”以查看更多可能的解析步骤。