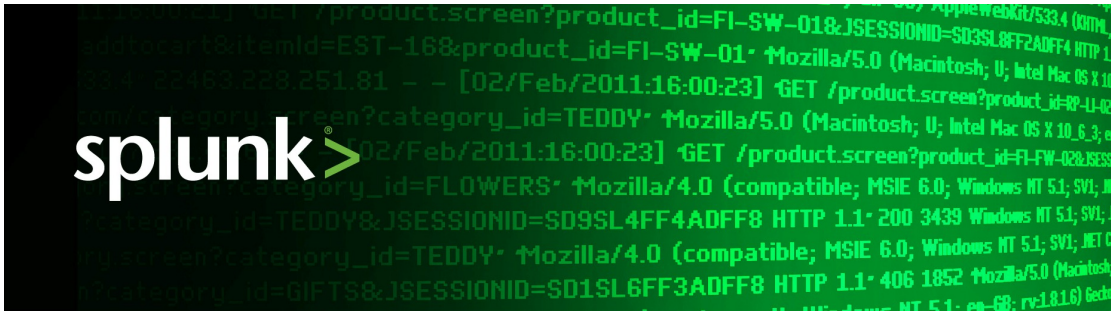


Splunk® Enterprise 6.5.0

数据导入

生成时间：2016 年 9 月 26 日，下午 10:23

Table of Contents

简介

什么数据可以建立索引？

为把您的数据变成可供搜索的数据，请把数据发送至您的 Splunk 部署为其建立索引。在为数据建立索引期间，Splunk Enterprise 会把您的数据转换成一系列包含可搜索字段的**事件**。您可以在 Splunk 为数据建立索引之前和之后，对数据进行精细处理，但通常不需要这样做。为数据创建索引后，就可以开始搜索数据了；或者，您也可以使用数据来创建图表、报表、告警及其他感兴趣的输出。

哪种数据类型？

任何类型。尤其是任何的 IT 流、计算机和历史数据，例如 Windows 事件日志、Web 服务器日志、实时应用程序日志、网络源、系统指标、变更监控、消息队列、归档文件等。

将您的 Splunk 部署指向数据源。并告知一些数据源信息。然后数据源就成为一个数据导入。Splunk Enterprise 会为数据流创建索引，并将其转换为一系列的事件。您可以立刻查看和搜索这些事件。如果所得结果与您的预期不符，您可以调整索引过程，直到它们与您的预期相符。

如果您使用的是 Splunk Enterprise，数据可以与**索引器**位于相同的计算机上（**本地数据**），也可以位于另一台计算机上（**远程数据**）。如果您使用的是 Splunk Cloud，则数据驻留在您的企业网络中，您可以将其发送至您的 Splunk Cloud 部署。通过使用网络源或在生成数据的主机上安装 Splunk **转发器**，您可以将远程数据导入您的 Splunk 部署中。更多有关本地数据与远程数据的信息，请参阅"我的数据位于何处？"

Splunk 提供了多种**应用**和**加载项**，并已为诸如特定于 Windows 或 Linux 的数据源、Cisco 安全数据、Blue Coat 数据等一类的内容预配置了输入。可根据需要在 Splunkbase 中查找合适的应用或加载项。此外，Splunk Entreprise 还附带了许多数据源方法，例如 Web 服务器日志、Java 2 平台、Enterprise Edition (J2EE) 日志或 Windows 性能指标。您可以从 Splunk Web 的**添加数据**页面访问这些方法。如果没有符合您要求的方法和应用，您可以使用通用输入配置功能来指定特定数据源。

更多有关如何配置数据导入的信息，请参阅配置您的输入。

数据源类型

Splunk 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 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 输入进行分类：

- 文件和目录
- 网络事件
- Windows 数据来源
- 其他数据来源

文件和目录

很多数据直接来自文件和目录。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

Splunk Enterprise 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 数据创建索引，但是为了提高可靠性，我们建议您尽可能改用 TCP。

Splunk Enterprise 也可以接收 SNMP 事件（即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，请参阅本手册中的向您的 Splunk 部署发送 SNMP 事件。

Windows 数据来源

Splunk Cloud 和 windows 版本的 Splunk Enterprise 支持广泛的 Windows 特定输入。此外，其 Splunk Web 中还提供了用于定义以下所列 Windows 特定输入类型的页面：

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

注意：要在非 Windows 的 Splunk Enterprise 实例上为 Windows 数据创建索引并进行搜索，您必须首先使用 Windows 实例来收集数据。请参阅有关确定如何监视远程 Windows 数据的注意事项。

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控 Windows 数据。

其他数据源

Splunk 软件也支持其他类型的数据源。例如：

- 先入先出 (FIFO) 队列
- 脚本式输入
  从 API 及其他远程数据接口和消息队列获取数据。
- 模块化输入
  定义自定义输入功能以扩展 Splunk Enterprise 框架。

开始使用数据导入

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。

或者，您可以下载并启用一个应用，如 Splunk App for Microsoft Exchange 或 Splunk IT Service Intelligence。

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”应用或主应用页面，并开始详细浏览您已收集的数据。

- 如需了解更多有关如何配置输入的信息，请参阅配置您的输入。
- 如需了解如何把数据添加至您的 Splunk 部署，请参阅您想如何添加数据？
- 如需了解如何试验添加测试索引，请参阅使用测试索引。
- 如需了解如何添加来源类型，请参阅“设置 Sourcetype 页面”。
- 如需了解何为事件处理及如何予以配置，请参阅 Splunk 软件如何处理您的数据。
- 如需了解如何从您的 Splunk 部署中删除数据，请参阅删除索引数据并重新开始。
- 如需了解如何为您的输入配置默认索引，请参阅在默认索引中指向您的输入。

添加新输入

以下为添加数据的高级程序。

1. 了解您的需求。问自己以下这些问题。
   - 我想要为哪种数据建立索引？请参阅什么数据可以建立索引？
   - 是否有相关应用？请参阅使用应用导入数据。
   - 数据位于何处？位于本地还是远程？请参阅我的数据位于何处？
   - 我是否应该使用转发器来访问远程数据？请参阅使用转发器导入数据。
   - 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”？
2. 创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。
3. 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。
4. 查看您用“搜索”应用添加的测试数据：
   - 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？
   - 数据是否丢失或损坏？
   - 是否获得了最佳结果？
5. 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。
6. 如有必要，删除测试索引中的数据，然后重新开始。
7. 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主索引。

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 软件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”，帮助您针对如何让 Splunk 软件处理您的数据做出决策。请思考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 软件是否无法正确决定时间戳？请参阅时间戳分配如何工作。

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，则所有索引数据均为远程数据。如果您使用的是本地的 Splunk Enterprise 部署，则该问题的答案取决于多个因素，包括：

- 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创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）。如果您的数据位于这样一个资源上，该数据则为本地数据。

本地数据的一些示例包括：

- 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（主机可于开机时获取）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。下面是一些远程资源的示例：

- 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。

例外情况

有时，资源看起来是远程的，但实际不是远程的。我这边举几个例子：

- 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。

使用转发器导入数据

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 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 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。它们代表一个比原始网络源强大得多的解决方案并且它们能进行如下操作：

- 标记元数据（数据来源、来源类型和主机）
- 可配置的缓冲
- 数据压缩
- 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 部署，由该部署为数据建立索引并执行搜索。Splunk 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 部署中，转发器都用作获取数据的主要工具。大型 Splunk 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

以下过程仅供一般性参考。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》*。

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 的任何数据。
3. 为此平台或含数据的主机架构下载 Splunk Enterprise 或通用转发器。
4. 把转发器安装到主机上。
5. 在主机上启用转发并指定一个
6. 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。
7. 确认来自转发器的数据会到达接收索引器。

以下是您可在转发器上配置数据导入的主要方式：

- 在初始部署时指定输入。
- 如为 Windows 转发器，请在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。
- 安装其中包含您所需的输入的应用。
- 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 部署的过程。您可用配置的数据导入找到一个应用。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，例如：

- Splunk IT Service Intelligence
- Splunk App for F5
- Splunk App for Cisco Security
- Splunk App for Websphere Application Server

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？”。特别是，“从哪里获取更多应用和加载项”一文将介绍如何下载和安装应用：

有关如何创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。

配置输入

如需添加新的数据类型到您的 Splunk 部署，请配置一个数据导入。可以通过多种方法来配置数据导入：

- **应用**。Splunk 有多种**应用**，其为多种数据类型提供预先配置输入。有关更多信息，请参阅“使用应用导入数据”。
- **Splunk Web**。可以使用 **Splunk Web** 数据导入页面配置大部分输入。可以从 Splunk 页访问**添加数据**登录页面。另外，您可以在上载或监视文件时预览并调整为文件建立索引的方式。
- **Splunk 命令行界面 (CLI)**。如果您使用的是 Splunk Enterprise，可以用 CLI 来配置绝大多数输入类型。
- **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**配置文件** inputs.conf 中。如果您使用的是 Splunk Enterprise，可以直接编辑该配置文件。一些高级数据导入需求可能需要您对其进行编辑。

另外，如果您将**转发器**配置为将数据从外围计算机发送到中央**索引器**，则可以在安装时指定一些输入。请参阅"使用转发器导入数据"。

使用 Splunk Web

您可以从 Splunk 主页或**设置 > 数据导入**菜单添加数据导入：

- 在 Splunk 主页上，选择**添加数据**。将显示**添加数据**页面，其中提供了大量数据导入类型方法的链接。
- 从系统栏中选择**设置 > 添加数据**。
- 从系统栏**设置**弹出菜单的**数据**部分选择**设置 > 数据导入**。此时将显示"数据导入"页面。您可以在其中查看和管理现有输入以及添加新输入。

添加数据页面上有导入数据的选项。单击图标可转到定义您要上载、监视或转发数据的页面。

- 上载
- 监视器
- 转发。

更多有关如何使用"添加数据"页面的帮助信息，请参阅您想如何添加数据？

应用上下文如何决定 Splunk Enterprise 写入配置文件的位置

您通过 Splunk Web 添加输入时，Splunk Enterprise 会把该输入添加到 `inputs.conf` 的副本。应用上下文，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`。

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*《管理员》*手册中的"有关配置文件"。

使用 CLI

如果您使用的是 Splunk Enterprise，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/
```

更多有关 CLI 的信息，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

要想通过 inputs.conf 配置输入，请使用文本编辑器。为每个输入添加**段落**。您可以把该段落添加到 `inputs.conf` 文件，该文件既可以位于 `$SPLUNK_HOME/etc/system/local/` 内，也可以位于自定义应用程序目录内（路径为 `$SPLUNK_HOME/etc/apps/<app name>/local`）。

要配置数据导入，可向其段落添加属性/值对。可在输入段落中设置多个属性。如果您未指定某个属性的值，Splunk Enterprise 会使用该属性的默认值。所有 `inputs.conf` 属性的默认值都位于 `$SPLUNK_HOME/etc/system/default/inputs.conf` 中。

如果您之前未使用过配置文件，请在开始添加输入前参阅"关于配置文件"。

inputs.conf 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输入的主题。例如，要了解如何配置文件输入，请参阅通过 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 `inputs.conf` 规范文件。

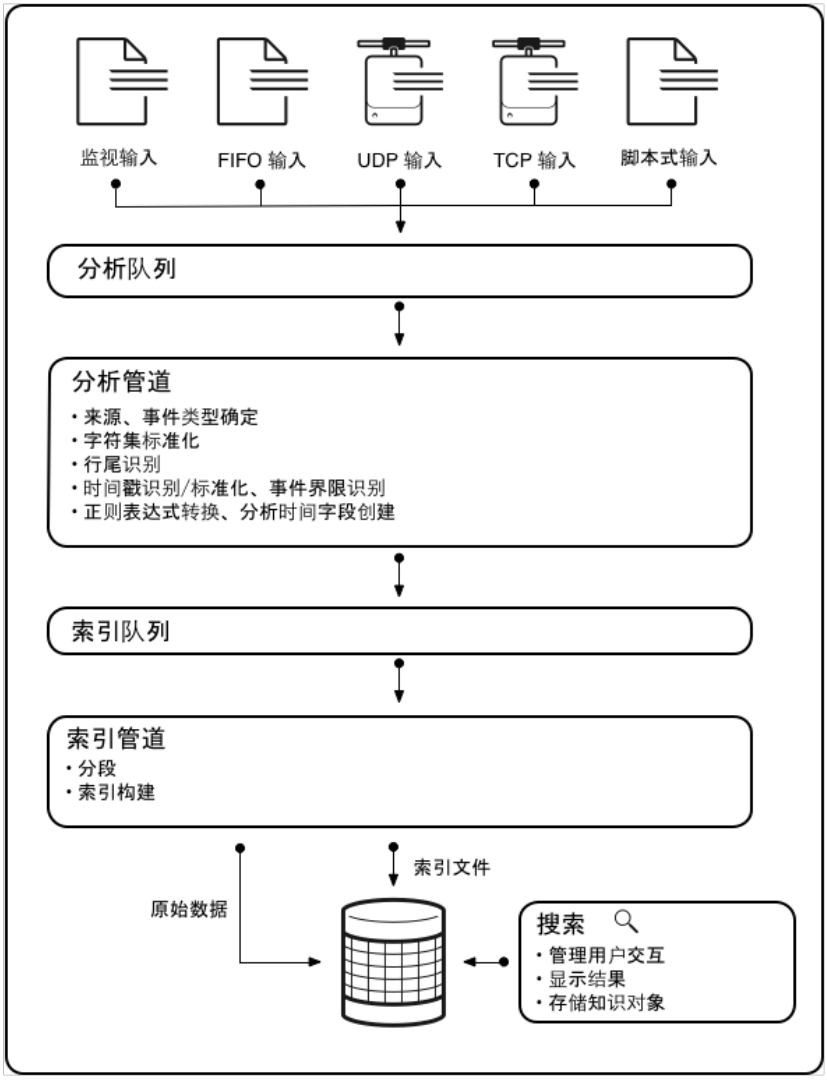
Splunk Enterprise 如何处理您的数据

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

传入的数据通过数据管道，详参*《分布式部署手册》*中的"数据如何通过 Splunk 部署：数据管道"。

下图显示了数据管道中的主要步骤。



事件处理

事件处理将出现在两个阶段：分析和创建索引。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 软件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 软件会转换数据。您可配置其中大部分进程以根据需要进行调整。

Splunk 软件将在分析管道中执行大量操作，包括：

- 为每个事件提取一组默认字段，包括 `host`、`source` 和 `sourcetype`。
- 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 Sourcetype"页面交互修改行尾设置。
- 标识或创建时间戳。Splunk 软件在处理时间戳的同时会识别事件界限。您可使用"设置 Sourcetype"页面交互修改时间戳设置。
- 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的事件数据 （如信用卡号或社会保险号）。
- 根据您的配置，应用自定义元数据到传入事件。

Splunk Enterprise 会在索引管道中执行其他处理，包括：

- 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。
- 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创建。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。其他分析将在接收索引器上进行。

有关事件及其在索引创建过程中所发生情况的更多信息，请参阅本手册中的事件处理概述。

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 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？

把数据添加到您的 Splunk Enterprise 部署的最快方法是使用 Splunk Web。Splunk Web 界面会提供一个

“添加数据”页面

登录 Splunk 部署后将显示主页：

如需添加数据，请单击**添加数据**按钮（在应用列表的右边）。此时将显示“添加数据”页面。（如果您的 Splunk 部署为自助式 Splunk Cloud 部署，请选择**设置**并单击**添加数据**。）

注意：“添加数据”页面在以下情况时不会显示：

- 该实例运行为**搜索头群集**的一部分。请参阅*《分布式搜索》*手册中的“关于搜索头群集化”。
- 该实例为托管式 Splunk Cloud 实例。

使用 Splunk Web 把数据导入您的 Splunk 部署时，有三个选项可供选择：**上载**、**监视**和**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。单击“上载”后，Splunk Web 会跳转到一个启动上载进程的页面。请参阅上载数据。

监视器

“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。单击“监视”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

“转发”选项允许您把数据从**转发器**接收到您的 Splunk 部署中。单击“转发”按钮时，Splunk Web 会将您带到一个从转发器启动数据收集进程的页面。请参阅转发数据。

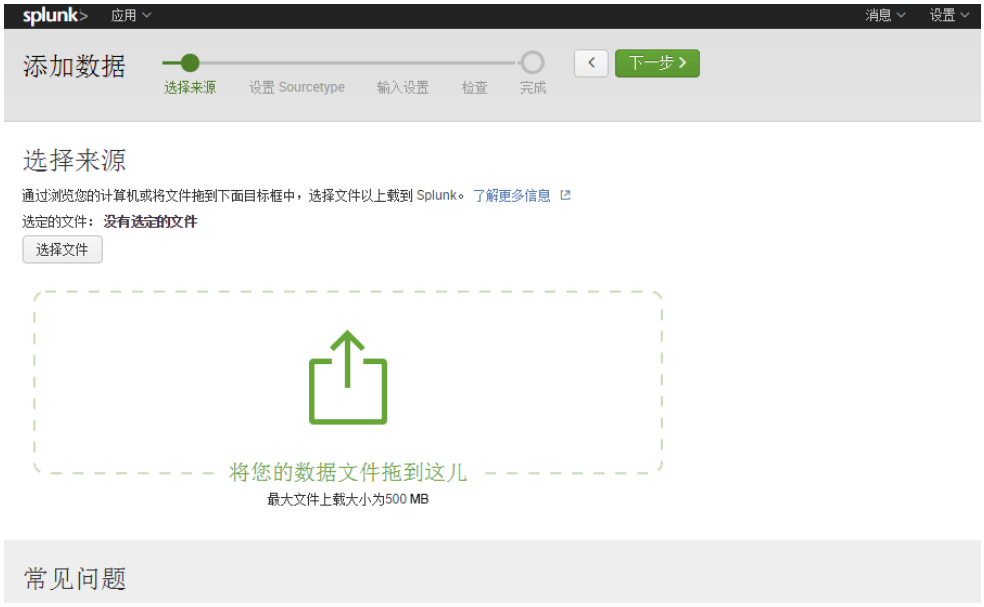
“转发器”选项需要其他配置。仅在单实例 Splunk 环境中使用。

上载数据

“上载”页面允许您指定一个可以直接从计算机上载至 Splunk Enterprise 实例的文件。

注意：从另一个主机导出的 Windows Event Log (.evt) 和 Windows Event Log XML (.evtx) 文件无法使用该上载功能。这是因为上述文件中包含的信息是生成这些文件的主机所特定的。如果不变更格式，其他主机将无法处理这些文件。更多有关处理这类文件时所受限制的信息，请参阅为导出的事件日志（.evt 或 .evtx）文件建立目录。

“上载”页面



您可以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页面上的“将您的数据文件拖到这儿”区域。

或

- 在屏幕左上角，单击**选择文件**并选择您要索引的文件。

Splunk 软件接下来会根据该文件的类型加载并处理该文件。一旦加载完成，您便可单击右上角的绿色**下一步**按钮，以继续到“添加数据”进程中的下一步。

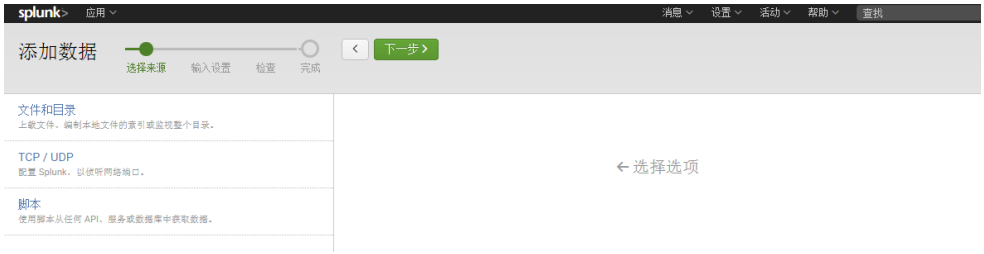
后续步骤

设置 sourcetype

监视数据

您可以使用“监视数据”页面来监视运行 Splunk Enterprise 实例的主机上的文件和网络端口。

“监视”页面



访问“监视”页面时，请选择您希望 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，最后才是您安装在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（Splunk Enterprise 或 Splunk Cloud）和 Splunk Enterprise 实例的运行平台。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 数据源仅在运行 Windows 的主机上可用。

如果无法添加数据导入，您登录的 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。

1. 单击某源即可从左窗格中选定该源。页面更新以您已选择的来源为基础。例如，如果您选择了“文件和目录”，则该页面会更新并出现一字段供您输入文件或目录名称，并介绍 Splunk 软件该如何监视该文件或目录。
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。

后续步骤

设置 sourcetype

转发数据

“转发数据”页面允许您选择已经连接至 Splunk Enterprise 实例的转发器，以将数据配置并发送至该实例。您在“添加数据”页面上单击**转发**按钮的时候，Splunk Web 将加载该页面。

仅在您将 Splunk Enterprise 的单实例用作索引器兼部署服务器时使用该页面；或者仅在您使用自助式 Splunk Cloud 部署并把您的通用转发器配置为部署客户端时使用该页面。如果您有多个执行索引创建的主机，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”。

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。

如需把某一轻型或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”。

如需把某一通用转发器配置为部署客户端，请参阅*《通用转发器》*手册中的“配置通用转发器”。您可以在 Windows 主机上安装转发器期间，把该转发器配置为部署客户端。

“选择转发器”页面

您从“添加数据”页面中选择“转发数据”后，将出现以下页面。

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。

此页面仅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，则该页面会发出警告。

1. 在**选择服务器类**中，单击下列选项中的任意一个。
   - **新建**：创建新的服务器类，或当现有服务器类与您要为其配置输入的转发器组不匹配时。
   - **现有**：使用现有服务器类时。
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。
   注意：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. 如果您选择“选择服务器类”中的**新建**，则为服务器类输入自己能记住的唯一名称。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**即可继续到“设置 Sourcetype”页面。

后续步骤

修改输入设置

“设置 Sourcetype”页面

“设置 Sourcetype”页面允许您通过预览 Splunk 软件如何为数据建立索引改善**事件处理**。请使用此页面确认 Splunk Enterprise 按照您想要其显示的方式为数据创建索引。

在建立索引前预览数据

“设置 Sourcetype”页面在您使用“上载”或“监视”页面将单个文件指定为数据源后会显示。

您可以在“设置 Sourcetypes”页面上对 Splunk 软件为数据建立索引的方式进行调整。您可以交互式地调整和改善索引的建立进程，这样就能在为传入的数据建立索引时，把您的**事件数据**存储为您想要的格式。

将正确的来源类型分配给您的数据

“设置 sourcetype”页面帮助您将正确的**来源类型**应用于您的传入数据。分配给所有传入数据的**默认字段**类型有很多种，来源类型是其中之一，该类型决定了 Splunk 软件在建立索引过程中为数据设置格式的方式。通过将正确的来源类型分配给您的数据，数据的索引版本（**事件数据**）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。

Splunk Enterprise 提供了大量预定义的来源类型，并尝试根据数据的格式把正确的来源类型分配给您的数据。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创建一个新来源类型。

该页面显示了 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。

使用“设置 Sourcetype”页面可进行以下操作：

- 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。
- 应用其他来源类型查看其提供的结果是否更切合您的要求。
- 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。
- 从头创建一个新来源类型。

该页面把所有的新来源类型都保存到 `props.conf` 文件中。您稍后可以把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"分布来源类型配置"。

有关来源类型的信息，请参阅本手册中的"来源类型为何重要"。

请使用"设置 Sourcetypes"页面

当"设置 Sourcetype"页面加载时，Splunk Enterprise 基于您指定的数据选择来源类型。您可接受该推荐或更改它。

下面是"设置 Sourcetype"页面的一个示例：

1. 请参阅页面右侧的预览窗格，以了解 Splunk Enterprise 将如何索引数据。查看事件换行和时间戳。
2. （可选）单击右侧的**查看事件摘要**链接可查看事件摘要。Splunk Web 会在新的窗口中显示该事件摘要。查看"查看事件摘要"。
3. 如果数据以您希望的方式显示，则继续到"步骤 5"。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。"
4. 进行更改后，返回至"步骤 1"以再次预览数据。
5. 若结果符合您的要求，请单击"下一步"以继续到**输入设置**页面。

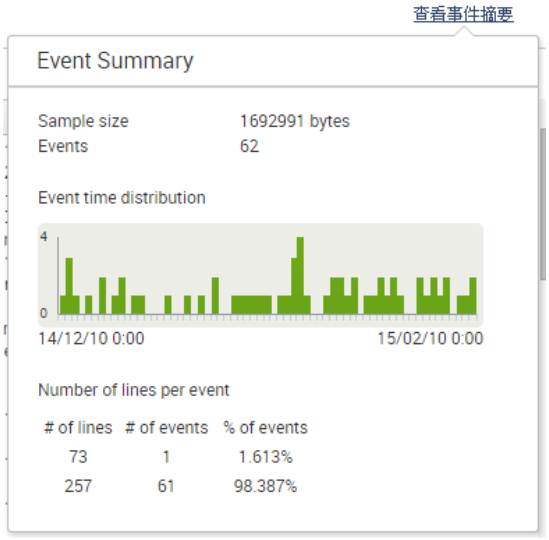
选择一个现有来源类型

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。

注意：Splunk Enterprise 若检测出来源类型，将在 "Sourcetype <sourcetype>" 按钮中显示该来源类型。若无法确定来源类型，则将显示 "Sourcetype: System Defaults"。**输入设置**

1. 单击 **Sourcetype：<sourcetype>** 按钮可用来查看源类型类别列表。在每个类别下存在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在右侧弹出菜单中显示。
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. 请再次查看您的数据。

查看事件摘要



您可以通过单击页面右侧的“查看事件摘要”在数据示例中查看事件摘要。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 软件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，则您可手动调整 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 Sourcetypes”页面左窗格上的**高级**下拉选项卡。当您更改设置时，预览窗格会更新。

注意：仅当 Splunk Enterprise 检测出文件为特定类型或您为文件选择一个特定来源类型时，才会显示一些选项卡。

- 如果 Splunk Enterprise 无法决定如何划分文件，或如果您选择了未定义换行的来源类型时，将显示“事件换行”选项卡。
- 只有当 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，“分隔的设置”选项卡才会出现。

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”。

1. 请单击**事件换行**选项卡。该选项卡将显示**换行类型**按钮，这些按钮会控制 Splunk 软件把文件划分成事件的方式。
   - **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**选项卡。该选项卡扩展以显示提取信息的选项。请从以下选项中任意选择一个：
   - **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、时间戳格式（又称为 strptime() 的特定格式）和包含该时间戳的任何字段。
3. 单击**分隔的设置**选项卡可显示分隔的选项。

   - **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（不包含任何实际数据的行）的正则表达式。
   - **字段名称**：如何决定字段名称：自动方式、基于行数、基于逗号分隔的列表或通过正则表达式。

   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。
4. 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。
   警告：该“高级”选项卡需要您了解 Splunk 功能的高级知识，且此处所做的更改可能会对数据的索引建立造成负面影响。请考虑向 Splunk Professional Services 的成员咨询，以在配置这些选项上获得帮助。

在高级选项卡中更改配置

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**可创建新属性/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。

后续步骤

修改输入设置

准备用于预览的数据

“设置 Sourcetype”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 部署上的文件。尽管它不直接处理网络数据或文件目录，但您能够解决这些限制。

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。*nix 上最常用的工具是 `netcat`。

例如，如果您在 UDP 端口 514 上侦听网络流量，可以使用 `netcat` 指示一些网络数据进入文件中。

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

为了获得最佳结果，请在其逻辑为文件大小达到 2MB 时终止 `netcat` 的 shell 脚本内运行此命令。默认情况下，当您预览文件时，Splunk 软件仅读取前 2MB 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，因为指定任何通配符会导致 Splunk Web 禁用此“设置 Sourcetype”页面。）

文件大小限制

Splunk Web 将显示“设置 Sourcetype”页面中一个文件的前 2MB 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 中的 `max_preview_bytes` 属性取样大量数据。或者，您可以编辑该文件以减少大量相似数据，这样剩余的 2MB 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. 按“设置 Sourcetype”页面中的描述查看事件数据。
2. 修改事件处理设置。
3. 查看更改效果并反复操作，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

如需创建新的来源类型，请使用事件换行和时间戳参数，然后再保存该来源类型。

在“设置 Sourcetype”页面的左侧有可折叠选项卡和链接，您可以执行三种类型的调整：

- **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。
- **高级模式。**如果您使用的是 Splunk Enterprise，请编辑 `props.conf`。

事件换行

要修改事件换行参数，请单击**事件换行**。该栏打开以显示以下按钮：

- **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把数据拆分成事件。

换行相关信息请参阅配置事件换行。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，对表达式进行测试。Splunk 软件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。

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。该选项卡打开以显示以下选项：

对于**提取**，可以选择下列选项之一：

- **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。

“高级”参数包括：

- **时区。**您想要用于事件的时区。
- **时间戳格式。**一个字符串，代表 Splunk Enterprise 在数据中搜索时间戳时要使用的时间戳格式。
- **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。**字符数，为 Splunk Enterprise 应在事件内查找的时间戳（或您在“时间戳前缀”中指定的正则表达式）。

注意：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 Enterprise 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` 的告警。（您仍然可以根据 Splunk Enterprise 尝试从问题中恢复的方式而以有效时间戳结束。）

有关配置时间戳的更多信息，请参阅“配置时间戳”。

高级

如需修改高级参数，请单击**高级**选项卡。该选项卡将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各选项。

您可以通过指定属性/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 props.conf。

“高级”框显示已选择的来源类型的当前完整属性集：

- 由**事件换行**或**时间戳**选项卡中所做更改生成的设置（您单击**应用**后）。
- 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。
- 从**其他设置**文本框应用的设置（您单击**应用设置**后）。

有关如何设置来源类型属性的信息，请参阅“配置”文件参考中的“props.conf”。也可参阅“时间戳分配如何工作”和“事件换行”。

Splunk Enterprise 如何组合设置

高级模式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**高级模式**中进行了互相冲突的时间戳更改，**高级模式**更改将优先于您在“时间戳”选项卡中所做的修改。

下面是 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：

- 高级模式更改
- 事件换行/时间戳更改
- 基本来源类型的设置（如果有）
- 所有来源类型的默认系统设置

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

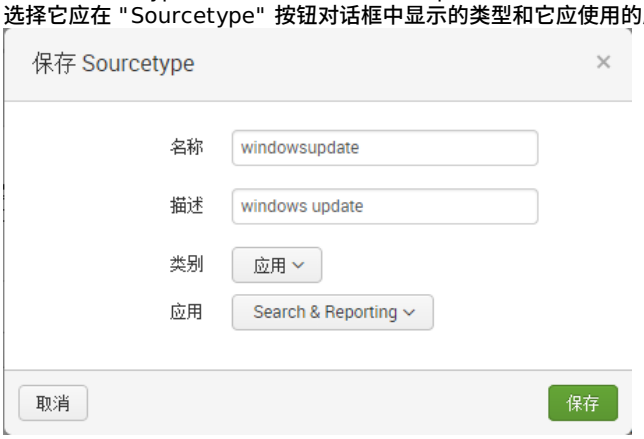
查看更改

当您准备好查看更改的效果后，请选择**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。

要再次使用提供的三个调整方法中的任意一个做进一步更改，请选择**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 “Sourcetype” 按钮旁边的“另存为”。Splunk Web 将显示一个对话框，您可在其中命名新来源类型、选择它应在 "Sourcetype" 按钮对话框中显示的类型和它应使用的应用程序上下文。



2. 输入新来源类型的**名称**。
3. 输入来源类型的**描述**。
4. 当您单击 "Sourcetype" 按钮时，请选择来源类型应在其中显示的**类别**。
5. 请选择新来源类型应该用于的**应用**。
6. 请单击**保存**以保存此来源类型并返回到“设置 Sourcetype”页面。

后续步骤

在保存来源类型后，您有以下几个选项：

1. （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。
2. （可选）单击 "<" 并选择要上载或监视的新文件。
3. （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

在您选择来源（或于上载或监视单个文件时设置您的来源类型后），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、它的应用程序上下文、它的主机值和来自输入的数据应保存在其中的索引。

可进行以下输入设置：

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。该设置将在您执行下列操作时显示：

- 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“来源类型”设置不会显示。

指定来源类型

1. 单击任意一个按钮。
   - **选择**：将您指定的来源类型应用于数据。当您单击“选择”时，会显示一个下拉框。
   - **新建**：添加新的来源类型。当您单击“新建”时，会显示两个文本字段和一个下拉框。

选择一个现有来源类型

1. 从“选择来源类型”下拉框中选择最能代表您想要的来源类型的类别。
2. 从显示的弹出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”下拉框中为来源类型选择一个类别。
3. 请在“来源类型描述”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。应用程序上下文改进了输入和来源类型定义的可管理性。应用上下文根据优先顺序规则加载。请参阅*《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。

- **IP**：使用事件来源的主机的 IP 地址。
- **DNS**：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**：使用您在选择该选项时显示的"主机字段值"文本框中分配的主机值。

索引

"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，请保持下拉列表设置为"默认"。或者，您也可以单击下拉列表，并通过单击列表中的选项来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，而且您有创建索引的权限，则您可通过单击**创建新索引**按钮来创建新的索引。

当您做出选择后，请单击**下一步**以进入"添加数据"过程的最后步骤。

在 Splunk Enterprise 中分布来源类型配置

如果您使用 Splunk Web 在 Splunk Cloud 中创建来源类型， 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创建新的来源类型，然后您可将它们分配给来自特定文件或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；您打算使用自己创建的来源类型，通过目标索引器为数据创建索引。
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"页面中创建来源类型时，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。如果您以后创建了其他来源类型，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

创建好来源类型后，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创建索引。

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

Splunk Enterprise 具有三个文件输入处理器：**监视器**、MonitorNoHandle 和上载。

您可以使用**监视器**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添加一次性输入，例如历史数据归档。

请在运行 Windows Vista 或 Windows Server 2008 及更高版本的主机上使用 MonitorNoHandle 来监视系统自动轮换的文件。MonitorNoHandle 输入仅适用于 Windows 主机。

可使用以下任一方法将输入添加到监视器或上载：

- Splunk Web
- CLI
- inputs.conf

您可以使用 CLI 或 inputs.conf 把输入添加到 MonitorNoHandle。

使用“设置 Sourcetype”页面查看 Splunk 软件如何为文件中的数据建立索引。详细信息请参阅“设置 Sourcetype”页面。

监视器处理器如何工作

指定文件或目录的路径，监视器处理器会获取已写入该文件或目录的所有新数据。您可以此方式来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 应用程序的日志等等。

Splunk Enterprise 会随着新数据的出现监视文件或目录。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 Splunk Enterprise 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**白名单**和**黑名单**包括或排除文件或目录中的数据读取。

如果您禁用或删除监视器输入，则 Splunk Enterprise 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。如需停止正在进行的数据索引建立过程，您必须停止并重启 Splunk 服务器。

Splunk Enterprise 在启动过程中如何处理文件监视

Splunk 服务器将在重启后从停止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，Splunk Enterprise 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，Splunk Enterprise 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。

Splunk Enterprise 如何监视归档文件

归档文件（如 .tar 或 .zip 文件）需先解压缩然后才能建立索引。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，整个文件都将重新建立索引，而不仅限于新数据。这可能会导致产生重复的事件。

Splunk Enterprise 如何监视操作系统按计划轮换的文件

监视进程会检测日志文件轮换，不会处理已为其建立索引的重命名文件（.tar 和 .gz 归档除外）。请参阅 Splunk Enterprise 如何处理日志文件轮换。

Splunk Enterprise 如何监视不可写的 Windows 文件

Windows 会阻止 Splunk Enterprise 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

Splunk Enterprise 无法监视路径超过 1024 个字符的文件。

Splunk Enterprise 也不会监视文件扩展名为 `.splunk` 的文件，因为使用此种扩展名的文件中会包含 Splunk 元数据。如您需要为具有 `.splunk` 扩展名的文件创建索引，请使用 `add oneshot` CLI 命令。

为什么使用上载或批处理？

如需为静态文件创建一次性索引，请在 Splunk Web 中选择**上载**。

您还可以使用 CLI `add oneshot` 或 `spool` 命令来达到同样的目的。详细信息请参阅使用 CLI。

如果您使用的是 Splunk Enterprise，可以用 `batch` 输入类型（位于 `inputs.conf` 中）一次性加载文件并造成破坏。根据默认设置，Splunk 批处理其位于 `$SPLUNK_HOME/var/spool/splunk` 内。如果您将某个文件移入此目录，Splunk Enterprise 会先为该文件建立索引，然后再将文件删除。

注意：有关加载文件归档的最佳做法，请参阅社区 Wiki 上的“如何为不同大小的归档建立索引”。

为什么使用 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。该输入是通过使用内核模式过滤器驱动程序来捕获写入文件的原始数据而实现的。此输入段落适用于已被锁定而无法打开进行写入的文件。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段落，例如 Windows DNS 服务器日志文件。

注意：`MonitorNoHandle` 仅适用于 Windows Vista 或 Windows Server 2008 及更高版本的操作系统。使用 `MonitorNoHandle` 只能监视单个文件。不能监视目录。如果您选择要监视的文件已经存在，Splunk Enterprise 不会为该文件当前的内容建立索引，只会为在文件执行写入时传入该文件的新信息建立索引。

使用 Splunk Web 监视文件和目录

如果您使用的是 Splunk Enterprise，可用 Splunk Web 从文件和目录中添加输入。

转到“新增”页面

您可从 Splunk Web 中的“新增”页面添加输入。

可通过两种方式访问此页面：

- Splunk 主页
- Splunk 设置

Splunk 设置：

1. 单击 Splunk Web 右上角的**设置**。
2. 在“设置”弹出窗口的“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。

Splunk 主页：

1. 请单击 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。

注意：转发文件需要其他设置。请参阅以下主题：

- 如果您使用的是通用转发器，请参阅“配置通用转发器”。
- 如果您使用的是重型和轻型转发器，请参阅“在 Splunk Enterprise 实例上启用转发”。

选择输入来源

1. 如需添加文件或目录输入，请单击**文件和目录**。
2. 在**文件或目录**字段中，指定文件或目录的完整路径。
   要监视共享网络驱动器，请输入以下内容：`<myhost>/<mypath>`（如为 Windows 则输入 `\\<myhost>\<mypath>`）。
   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. 选择您想要 Splunk Enterprise 监视文件的方式：
   - **持续监视**。设置持续输入。Splunk Enterprise 为新数据持续监视文件。
   - **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“白名单”和“黑名单”字段。这些字段允许您指定 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 Sourcetype”页面，您可在该页面预览 Splunk Enterprise 如何建议对事件创建索引。

更多有关如何把数据加入白名单和黑名单的信息，请参阅把特定的传入数据列入白名单或黑名单。

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**并预览它被索引时的外观。这将允许您确保数据已正确设置格式并作出必要的调整。

有关该页面的信息请参阅设置 Sourcetype 页面。

如果您跳过预览数据，则会显示**输入设置**页面。

注意：您无法预览目录或归档文件。

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**名称值。
   注意：主机只是设定生成事件中的**主机**字段。而不是引导 Splunk Enterprise 查找网络中的特定主机。
3. 为此输入设置 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

在您指定所有输入设置后，请查看您的选择。Splunk Web 会列出您选择的选项，包括但不限于监视器的类型、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，单击 < 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 Enterprise 开始索引指定的文件或目录。

使用 CLI 监视文件和目录

可通过 Splunk Enterprise 的命令行界面 (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] ...`	直接把文件 <source> 复制到 Splunk。这将上载文件一次，但 Splunk Enterprise 不会持续监视文件。 您无法对远程 Splunk Enterprise 实例使用 `oneshot` 命令。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。请指定您所要监视文件的准确来源路径。
spool	`spool <source>`	使用 sinkhole 目录把文件 <source> 复制到 Splunk Enterprise。类似于 **add oneshot**，不同之处是文件需要从 sinkhole 目录进行后台打印，而不是立即添加。 您无法对远程 Splunk Enterprise 实例使用 `spool` 命令。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。请指定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。使用以下语法设置参数：`-parameter value`。

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

参数	是否必需？	描述
`<source>`	是	要为新输入监视/上载的文件或目录的路径。 和其他参数不同，该参数的语法可以是数值本身，并不需要遵照参数标记。您既可以使用 "`./splunk monitor <source>`" 也可以使用 "`./splunk monitor -source <source>`"。
`sourcetype`	否	为来自输入来源的事件指定 Sourcetype 字段值。
`index`	否	为来自输入来源的事件指定目标索引。
`hostname` **或** `host`	否	为来自输入来源的事件指定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	指定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	指定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`。

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。Splunk Enterprise 仅使用一次文件。它不会持续监视文件。

直接把 /var/log/applog（在 Windows 上的路径为 C:\Program Files\AppLog\log.txt）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 <code>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

如需向 Splunk Enterprise 配置输入，请把**段落**添加到 $SPLUNK_HOME/etc/system/local/ 中的 inputs.conf 或将其添加到您自己位于 $SPLUNK_HOME/etc/apps/ 中的自定义应用程序目录内。（如需为 Splunk Cloud 配置输入，请使用 Splunk Web。）

可在输入段落中设置多个属性。如果您未指定某个属性的值，Splunk Enterprise 会使用 $SPLUNK_HOME/etc/system/default/inputs.conf 中定义的属性默认值。

更多有关配置文件的信息，请参阅“关于配置文件”。

配置设置

在 monitor 和 batch 输入段落中使用以下属性。

属性	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 Enterprise 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 Enterprise 会在 <string> 前面加上 index::。 更多有关索引字段的信息，请参阅*《管理索引器和群集手册》*中的“索引如何工作”。	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 Sourcetype 键/字段。显式声明该数据的来源类型，而不是允许 Splunk Enterprise 自动确定。这对于可搜索性以及在分析和创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 Enterprise 会在 <string>	Splunk Enterprise 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。

	前面加上 sourcetype::。 更多有关来源类型的信息，请参阅来源类型为何重要。	
queue = parsingQueue | indexQueue	指定输入处理器应该用来存储其所读取事件的位置。设置为 "parsingQueue" 时，会把 props.conf 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，使用该属性以选择性地将您的数据转发到特定索引器。 请在 outputs.conf 中定义 tcpout 组名称，位置就在 [tcpout:<tcpout_group_name>] 段落内。	这些组存在于 outputs.conf 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。特别地，Splunk Enterprise 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 =" 属性
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 '/' 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 =" 属性

监视器语法和示例

监视器输入段落可引导 Splunk Enterprise 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）。您必须依次指定输入类型和路径，以便如果路径包括根目录就将三个斜线放入路径中。

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

以下是可在定义监视器输入段落时使用的附加属性：

属性	描述	默认
source = <string>	设置来自此输入的事件的来源字段。除非绝对必要，否则不要覆盖。请考虑使用来源类型、标记和搜索通配符。通过准确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。 Splunk Enterprise 会在 <string> 前面加上 source::。	输入文件路径
crcSalt = <string>	强制 Splunk Enterprise 使用具有匹配 CRC（循环冗余码校验）的文件。默认情况下，该软件只对文件的前几行执行 CRC 检查。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 只基于文件的前几行，因此不同文件可以具有相同的匹配 CRC，尤其是在它们具有相同标题的情况下。） 如果设置，Splunk Enterprise 会把 string 添加到 CRC。如果设置为 <SOURCE>，Splunk Enterprise 会把完整的来源路径添加到 CRC。这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属性。它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 (modtime) 已超过 <time_window> 阈值，会导致输入停止检查文件的更新。这可提高在监视包含大量历史文件的目录层次结构时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 Enterprise 首次尝试监视文件时，该文件的修改时间落在 <time_window> 之外，Splunk Enterprise 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。这仅适用于 Splunk Enterprise 首次尝试监视的文件。此后，Splunk Enterprise 的内部文件位置记录会跟踪该文件。	0
`whitelist = <regular expression>`	如果设置，则 Splunk Enterprise 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 Enterprise 不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 Enterprise 会打开一个文件，以检查其是否已经创建索引。这仅对不更新修改时间的文件有用。 此属性用于监控 Windows 上的文件，并且主要针对 Internet Information Server (IIS) 日志。 警告：该属性的使用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 Enterprise 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 Enterprise 可关闭 End-of-file (EOF) 上的文件之前所需的修改时间增量。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 Enterprise 将忽略其在所监视目录中发现的符号链接。	true

示例 1.加载 `/apache/foo/logs` 或 `/apache/bar/logs` 等中的任何内容。

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

示例 2.加载 `/apache/` 中以 `.log` 结尾的任何内容。

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

MonitorNoHandle 语法和示例

仅限于 Windows 系统。如使用 `MonitorNoHandle` 段落监视文件则无需使用 Windows 文件句柄。这样，您就可以读取像 Windows DNS 服务器日志文件这样的特殊日志文件。

使用 `MonitorNoHandle` 时，您必须指定有效的文件路径。不能指定目录。如果指定的文件已经存在，则 Splunk Enterprise 不会为该文件中的现有数据创建索引。它只会为系统向该文件写入的新数据创建索引。

您只能使用 inputs.conf 或 CLI 配置 `monitorNoHandle`，不能在 Splunk Web 中对其进行配置。

```
[MonitorNoHandle://<path>]
<attrbute1> = <val1>
<attrbute2> = <val2>
...
```

批处理语法和示例

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**监视器**。请记住，Splunk Enterprise 将在为批处理输入建立索引后**删除**该文件。

```
[batch://<path>]
move_policy = sinkhole
<attrbute1> = <val1>
<attrbute2> = <val2>
...
```

定义批处理输入时，您必须将该属性包含在内，`move_policy = sinkhole`。该设置在加载文件时不会造成任何破坏。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。

示例：此批处理示例加载了 `system/flight815/` 目录中的所有文件，但不递归遍历其中的任何子目录：

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

注意：如需确保在您复制某包含新内容的现有文件时为新事件创建索引，请在 props.conf 中为该来源设置 `CHECK_METHOD = modtime`。这样，当文件发生更改时即会检查文件的修改时间并重新建立索引。请注意，由于将重新为整个文件建立索引，因此这会导致出现重复的事件。

使用通配符指定输入路径

您可以通过编辑 `inputs.conf` 文件手动配置输入。`inputs.conf` 中的输入路径规范不使用正则表达式 (regexes)，而是使用 Splunk 定义的通配符。本主题介绍如何在 `inputs.conf` 中的路径内指定这些通配符。如需指定通配符，您必须使用 `inputs.conf` 来指定文件和目录监视输入。

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。

通配符	描述	Reg. Exp. 等效	示例
`...`	省略号通配符将递归遍历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` 匹配文件 `/foo/1/bar.log`、`/foo/2/bar.log`、`/foo/1/2/bar.log` 等，但不会匹配 `/foo/bar.log` 或 `/foo/3/notbar.log` 由于单个省略号递归遍历所有文件夹和子文件夹，`/foo/.../bar.log` 的匹配方式与 `/foo/.../.../bar.log` 相同。
`*`	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` 匹配文件 `/foo/bar`、`/foo/1/bar`、`/foo/2/bar` 等，但不会匹配 `/foo/1/2/bar`。 `/foo/m*r/bar` 匹配 `/foo/mr/bar`、`/foo/mir/bar`、`/foo/moor/bar` 等。 `/foo/*.log` 匹配所有扩展名为 `.log` 的文件，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 Enterprise 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。如果您指定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。例如：

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

监视 `/var/log/log(a|b).log` 文件。因为不含通配符，`(a|b)` 不会被视为正则表达式。

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

监视 `/var/log()/` 目录中所有以 `log` 开头且扩展名为 `.log` 的文件。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。例如：

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

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。

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

监视 `/var/` 目录下任意子目录中所有命名为 `loga.log` 和 `logb.log` 的文件。Splunk Enterprise 将把 `(a|b)` 视为正则表达式，因为前一个段落段中包含通配符 ' `...`'。

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

监视 `/var/` 目录下任意子目录中所有满足以下条件的文件：

- 以 `log` 开头，
- 包含一个大写字母 (A-Z) 或数字 (0-9)、
- 包含任何其他字符且
- 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

监视 `/apache/foo/logs`、`/apache/bar/logs`、`/apache/bar/1/logs`：

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

监视 `/apache/foo/logs`、`/apache/bar/logs` 但不监视 `/apache/bar/1/logs` 或 `/apache/bar/2/logs`：

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

监视直接位于 `/apache/` 下且以 `.log` 结尾的所有文件：

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

监视任意层级子目录下 `/apache/` 之下所有以 `.log` 结尾的文件：

```
[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，使用：

```
[monitor:///var/log/]
whitelist=\.log$
recurse=true
#true by default
```

通配符与白名单

Splunk Enterprise 使用标准 Perl 兼容正则表达式 (PCRE) 语法定义白名单和黑名单。

在文件输入路径中指定通配符时，Splunk Enterprise 将为该段落创建一个隐式 `whitelist`。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，如果您指定

```
[monitor:///foo/bar*.log]
```

Splunk Enterprise 会将此转换为

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

在 Windows 中，如果您指定

```
[monitor://C:\Windows\foo\bar*.log]
```

Splunk Enterprise 会将此转换为

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

注意：在 Windows 中，`whitelist` 和 `blacklist` 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。

将特定传入数据列入白名单或黑名单

您可以使用**白名单**和**黑名单**规则指定**监视**目录时要获取或排除的文件。您也可以把这些设置应用于 `batch` 输入。当您定义白名单时，Splunk Enterprise 仅为您指定的文件建立索引。当您定义黑名单时，Splunk 软件会忽略指定的文件并处理其他所有文件。请在 `inputs.conf` 中的输入段落内定义白名单和黑名单。如果想把相同的黑名单和白名单应用到所有转发器上，您可以设置一个部署服务器。

不需要在段落中同时定义白名单和黑名单。它们是独立的设置。如果你同时定义了这两个名单，且某文件与二者都匹配，Splunk Enterprise 不会为该文件创建索引，因为 `blacklist` 会覆盖 `whitelist`。

白名单和黑名单规则使用正则表达式语法来定义文件名/路径的匹配项。二者必须包含于配置段落内，例如 `[monitor://<path>]`。Splunk 软件会忽略段落外的白名单和黑名单。当您定义白名单和黑名单条目时，您必须使用准确的正则表达式语法。

如需了解更多有关如何构建正则表达式的信息，请参看 Regular-expressions.info (http://regular-expressions.info) 网站。

发送和过滤数据

您可以不将数据导入列入白名单或黑名单，而是过滤特定事件并将其发送到不同的队列或索引。您还可以使用抓取功能预定义您希望在文件添加到文件系统时为其建立或不建立索引的文件。

白名单（允许）文件

- 把下面的一行添加到您的 `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，只监视扩展名为 `.log` 的文件：

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

把多个文件列入白名单

可以使用 "|" (OR) 运算符将多个文件列入白名单的一行中。例如，把包含 `query.log OR my.log` 的文件名列入白名单：

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

或者，您可以把一个精确的匹配列入白名单。

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

注意：“$” 会将正则表达式定位在行尾处。"|" 运算符前后没有空格。

黑名单（忽略）文件

- 把下面的一行添加到您的 `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）。**

```
blacklist = <your_custom_regex>
```

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

示例：仅把扩展名为 .txt 的文件列入黑名单

仅忽略且不监视扩展名为 `.txt` 的文件：

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

示例 2：把扩展名为 .txt 或 .tgz 的文件列入黑名单

忽略且不监视所有扩展名为 `.txt OR .gz` 的文件（请注意此时要使用 "|"）：

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

示例 3：把一整个目录列入黑名单

如需忽略监视输入下的整个目录：

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

此例指示 Splunk Enterprise 忽略归档或历史目录中 /mnt/logs/ 下的所有文件以及以 *.bak 结尾的所有文件。

示例 4：把文件名中包含字符串的文件列入黑名单

要忽略名称包含特定字符串的文件，您可执行以下操作：

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

此示例将忽略 `webserver20090228file.txt` 和 `webserver20090229file.txt` 文件，这个文件位于 `/mnt/logs/` 下。

Splunk Enterprise 如何处理日志文件轮换

如果 Splunk Enterprise 正在监视的文件（如 `/var/log/messages`）通过操作系统 (`/var/log/messages1`) 滚动，Splunk Enterprise 会识别这一情况，不再读取滚动后的文件。

监视处理器会选取新文件并读取文件的前 256 个字节。处理器然后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)，它充当表示文件内容的指纹。Splunk Enterprise 会使用此 CRC 在数据库中查找某个条目，该数据库包含 Splunk Enterprise 之前见过的所有文件的起始 CRC。成功后，查找返回一些数值和一个 seekCRC。数值中最重要的是 **seekAddress**，即 Splunk Enterprise 已读取到已知文件的字节数。seekCRC 是位置数据的指纹。

通过使用此次查找的结果，Splunk Enterprise 可对文件进行分类。

CRC 检查有三种可能的结果：

- 来自于数据库开始处文件的 CRC 无匹配记录。这指示是一个新文件。Splunk Enterprise 选取该文件，并从文件起始处获取文件内的数据。Splunk Enterprise 在处理文件时，会使用新 CRC 和 Seek Addresses 更新数据库。
- 来自于数据库开始处文件的 CRC 有匹配记录，Seek Address 位置的内容与文件中该位置存储的 CRC 相匹配，且该文件的大小比 Splunk Enterprise 存储的 Seek Address 大。虽然 Splunk Enterprise 之前见过此文件，但是该文件自上次读取之后已经新增了数据。Splunk Enterprise 打开文件，找到 Seek Address--Splunk Enterprise 最后完成文件时则为文件末尾--并从此处开始读取新数据。
- 来自于数据库开始处文件的 CRC 有匹配记录，但 Seek Address 位置的内容与文件中该位置存储的 CRC 不匹配。Splunk Enterprise 之前已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，或它实际上是完全不同的文件，只是以相同内容开始。由于用于内容跟踪的数据库由 CRC 开始处提供线索，没有其他办法为两个不同数据流独立跟踪进度，且需要进一步配置。

因为默认情况下，CRC 启动检查仅针对文件的前 256 个字节进行，因此对于非重复文件可能存在重复的启动 CRC，尤其是具有相同标题的文件。要处理这种情况，您可以：

- 使用 `initCrcLength` 属性（位于 `inputs.conf`）来增加用于 CRC 计算的字符数，使其长度超过您的静态标题。
- 使用 `crcSalt` 属性在 `inputs.conf` 中配置文件，可参阅本手册中"使用 inputs.conf 监视文件和目录"的介绍。把 `crcSalt` 设置为 `<SOURCE>` 时请确保每个文件都具有唯一的 CRC。此设置的效果是，Splunk Enterprise 假定每个路径名称包含唯一内容。

请勿使用含滚动日志文件的 `crcSalt = <SOURCE>`，或其他会重命名日志文件或把日志文件移动到另一个受监视位置的方案。该操作会阻止 Splunk Enterprise 识别滚动日志文件或重命名，从而导致重新对数据建立索引。

从网络来源获取数据

从 TCP 和 UDP 端口获取数据

您可以配置 Splunk Enterprise 以接受任何 TCP 或 UDP 端口上的输入。Splunk Enterprise 将获取抵达这些端口的任何数据。使用此方法从网络服务（如 syslog）捕获数据（默认端口为 UDP 514）。您还可设置 netcat 服务并将其与端口绑定。

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证书时，Splunk Cloud 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）发送数据，请使用 Splunk 通用转发器侦听该数据来源，并把数据转发至您的 Splunk Cloud 部署。

TCP 是以 Splunk Enterprise 数据分发方案为基础的网络协议。如需从任意远程主机发送数据到您的 Splunk Enterprise 服务器，我们建议您使用该协议。Splunk Enterprise 可以为来自 `syslog-ng` 或任何其他通过 TCP 传输的应用程序的远程数据创建索引。

Splunk Enterprise 支持通过 UDP 执行监视，但我们建议您尽量使用 TCP 来发送网络数据。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 以其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。

如您必须使用 UDP 发送网络数据，请参阅 Splunk 社区 Wiki 上的"使用 UDP 连接"以获取相关建议。

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 Splunk Enterprise 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。

如果您在一些网络设备（如思科自适应安全设备，Cisco ASA）上配置 TCP 日志，且该网络设备无法连接至监视器，则可能会导致性能减弱或停止日志，或者是更糟糕的结果。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。

使用 Splunk Web 添加网络输入

使用 Splunk Web 添加来自网络端口的输入：

转到"新增"页面

您可通过两种方式访问此页面：

通过 Splunk 设置：

1. 单击**设置**。

2. 请单击**数据导入**。

3. 请选取 **TCP** 或 **UDP**。

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. 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。

注意：转发文件需要其他设置。

3. 如果您选择了**转发**，则选择或创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.在左窗格中，请单击 **TCP / UDP** 以添加输入。

2.单击 **TCP** 或 **UDP** 按钮即可在 TCP 或 UDP 输入之间进行选择。

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。

注意：更改"源名称覆盖"值前请先咨询 Splunk 帮助团队。

5.如果是 TCP 输入，请指定此端口是应接受所有主机的连接还是只接受**仅接收来自字段的连接**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指定来源类型、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**名称值。您有几个选择：

- **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。
- **DNS。**将主机设置为远程服务器的 DNS 项。
- **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"。

> **注意：主机**只是设定生成事件中的**主机**字段。而不是引导 Splunk Enterprise 查找网络中的特定主机。

3.为此输入设置 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"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

在您指定所有输入设置后，可查看您的选择。Splunk Enterprise 会列出您勾选的选项，包括监视器的类型、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需要，单击 **<** 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网络输入。

使用 CLI 添加网络输入

如需访问 Splunk Enterprise CLI，请导航至 `$SPLUNK_HOME/bin/` 目录，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 主要帮助。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>`	删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 所用的用户身份必须有权访问此端口。

设置以下任何其他参数，即可修改每个输入的配置：

参数	是否必需？	描述
sourcetype	否	为来自输入来源的事件指定 Sourcetype 字段值。
index	否	为来自输入来源的事件指定目标索引。
hostname	否	为来自输入来源的事件指定要设置为主机字段值的主机名。
remotehost	否	指定只接受来自其数据的 IP 地址。
resolvehost	否	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	否	指定此输入只应该接受其连接的主机名或 IP 地址。

示例

- 将 UDP 输入配置为监视端口 514 并将来源类型设置为 "syslog"：

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

- 通过 DNS 设置 UDP 输入的主机值。输入用户名和密码以使用 auth：

```
./splunk edit udp 514 -resolvehost true -auth admin:changeme
```

有关 UDP 最佳使用方法的信息，请参阅社区 Wiki 里“配置 Syslog 输入的最佳方法”。

更改 TCP 网络输入中的受限制主机

在您创建 TCP 输入时，如果您决定只接受特定主机的连接，则一旦保存该输入后，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。

使用 inputs.conf 添加网络输入

如需添加一输入，请为该输入把一个段落添加至位于 $SPLUNK_HOME/etc/system/local/ 内或 $SPLUNK_HOME/etc/apps/ 中您自定义应用程序目录内的 inputs.conf。如您之前未使用过 Splunk 配置文件，请在开始前先参阅*《管理员》*手册中的“关于配置文件”。

您可以在输入类型后面设置任意多个属性和值。如您未指定一个或多个属性的值，Splunk Enterprise 将使用 $SPLUNK_HOME/etc/system/default/ 中预设的默认值（如下文说明）。

配置 TCP 输入

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

指示 Splunk Enterprise 在 <port> 上侦听 <remote server>。如 <remote server> 为空，Splunk Enterprise 将在指定端口上侦听所有连接。

属性	描述	默认
host = <string>	将此段落的主机键/字段设为一个静态值。也设置主机键的初始值。Splunk Enterprise 在分析和创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 应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 Sourcetype 键/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创建索引期间对此类型数据应用相关格式设置都很重要。	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。

	设置 sourcetype 键的初始值。Splunk Enterprise 在分析和创建索引过程中使用此键，特别是，可在创建索引期间使用此键来设置来源类型字段。Splunk Enterprise 使用在搜索时使用的来源类型字段。 在 `<string>` 前面加上 'sourcetype::'。	
`source = <string>`	设置来自此输入的事件的来源键/字段。在 `<string>` 前面加上 'source::'。 **注意：**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
`queue = parsingQueue | indexQueue`	指定输入处理器应该用来存储其所读取事件的位置。 设置为 "parsingQueue" 时，会把 `props.conf` 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
`connection_host = ip | dns | none`	"ip" 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" 将保留主机的指定值。	ip

在 SSL 上配置 TCP 输入

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

如要从转发器或第三方系统接收已加密且未分析的数据，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。

配置 UDP 输入

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

此类型输入段落类似于 TCP 类型，但是，它是在 UDP 端口上进行侦听。

- 如果您指定了 `<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>` - `[udp://<port>]` - 做任何指定，该端口将接受任何主机发送的数据。

属性	描述	默认
`host = <string>`	将此段落的主机键/字段设为一个静态值。也设置主机键的初始值。Splunk Enterprise 在分析和创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 应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 Sourcetype 键/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创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 在分析和创建索引过程中使用此键，特别是，可在创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键/字段。在 `<string>` 前	输入文件路径

	面加上 'source::'。 **注意：**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。	
`queue = parsingQueue | indexQueue`	设置输入处理器应该用来存储其所读取事件的位置。设置为 "parsingQueue" 时，会把 `props.conf` 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 该忽略值。	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。 如果您将该属性设置为 true，Splunk Enterprise 不会从接收的事件中去除 <priority> syslog 字段。 根据此属性的设置方式，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。 如果您将该属性设置为 true，Splunk Enterprise 不会将时间戳和主机附加到接收的事件。 **注意：**如果您想将时间戳和主机附加到接收的事件，请勿设置此属性。	false（Splunk Enterprise 将时间戳和主机附加到事件）

UDP 数据包和行合并

Splunk Enterprise 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` 中编辑基础来源并将 `SHOULD_LINEMERGE` 属性设置为 `false` 即可避免这一问题。这样做可以阻止 Splunk Enterprise 将数据包合并在一起。

问答

有什么问题吗？请访问 Splunk Answers，查看 Splunk 社区有哪些与 UDP 输入、TCP 输入及一般输入相关的问题和解答。

设置并使用 HTTP 事件收集器

HTTP 事件收集器 (HEC) 是一个端点，该端点可让您使用 HTTP 或安全 HTTP (HTTPS) 协议将应用程序事件发送至您的 Splunk 部署。HEC 使用一个基于您生成的令牌的验证模型。之后，您可以使用该令牌配置一个日志库或 HTTP 客户端，这样就可以用特定的格式把数据发送至 HEC。当发送应用程序事件时，该过程不需要转发器。

HEC 专为应用程序开发人员创建，所以一个应用程序只需添加几行代码就可以用来发送数据了。此外，HEC 也是基于令牌的收集器，所以您永远无需对自己应用中或支持文件内的 Splunk Enterprise 凭据进行硬编码。

HEC 作为一个单独的应用运行（称为 `splunk_httpinput`），并将其输入配置存储到 `$SPLUNK_HOME/etc/apps/splunk_httpinput/local` 中。

更多有关在 Splunk Enterprise 上启用 HEC 的信息，请参阅“将数据导入 Splunk 开发门户上的 HTTP 事件收集器中”。

关于事件收集器令牌

令牌是允许日志代理和客户端连接到 HTTP 事件收集器端点的实体。每个令牌都有一个令牌值：一个 32 位的数

字，代理和客户端使用该数字验证他们的 HEC 连接。当它们连接后，它们显示此令牌值。如果 HEC 已有配置好的令牌值且该值已启用，HEC 将接受该连接；之后，代理即可开始以 JavaScript Object Notation (JSON) 格式交付其应用程序事件的有效负载。

HEC 接收事件，Splunk Enterprise 则根据代理用于连接的令牌配置，使用数据来源、来源类型和令牌中指定的索引为 HEC 接收的事件建立索引。如果存在一个转发输出组配置，应用程序事件将在输出组定义它们的时候被转发至其他索引器。

在 Splunk Web 中配置 HTTP 事件收集器

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。如果您使用的是托管式 Splunk Cloud 部署，HEC 必须先由 Splunk 支持启用之后您才能使用。如果是 Splunk Enterprise，请通过 HEC 管理页面中的“全局设置”对话框启用 EC。具体步骤如下：

1. 从系统栏中单击**设置 > 数据导入**。

2. 请单击页面左侧的 **HTTP 事件收集器**。HEC 管理页面加载。

3. 请单击右上角的**全局设置**。

4. 请在**所有令牌**切换按钮中选择**已启用**。

5. 如需为所有 HEC 令牌设置来源类型，请从**默认来源类型**下拉列表中选择一个类别，然后再选择您想要的来源类型。在选择来源类型前，您也可在下拉列表上方的文本字段中键入来源类型的名称。

6. 如需为所有 HEC 令牌设置默认索引，请在**默认索引**下拉列表中选择一个索引。

7. 如需为所有 HEC 令牌设置默认转发输出组，请从**默认输出组**下拉列表中选择一个输出组。

8. 如需使用部署服务器处理 HEC 令牌的配置，请单击**使用部署服务器**复选框。

9. 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. 如需指定 HEC 侦听的端口号，请在 **HTTP 端口号**字段中输入数字。

注意：如需确保日志代理和 HEC 间正常通信，请确认代理上、托管 HEC 的 Splunk 实例上，或者是两者之间都没有防火墙阻止 **HTTP 端口号**字段中指定的端口号。

11. 要保存您的设置，请单击**保存**。对话框会消失，Splunk Web 将保存全局设置并带您返回到 HEC 管理页面。

创建事件收集器令牌

要使用 HTTP 事件收集器，您必须配置至少一个令牌。客户端和代理需要使用令牌才能连接到“事件收集器”以发送数据。

1. 请在“设置”菜单中选择**添加数据**。

2. 选择**监视器**，然后在左窗格中选择 **HTTP 事件收集器**。HEC 端点字段将填充右窗格。

3. 请在**名称**字段中输入令牌的名称，其可说明令牌的用途且您可将其记住。

4. （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。

5. （可选）请在**描述**字段中输入此输入的描述。

6. （可选）请在**输出组**字段中选择一个已有的转发器输出组（在下拉列表中选取）。

> **注意：** 在 `outputs.conf` 中定义输出组。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`。

7. （可选）如您想要启用该令牌的**索引器确认**，单击**启用索引器确认**复选框。

> **注意：** 索引器确认发送自索引器，旨在确认事件是否已完成索引建立。HTTP 事件收集器中的索引器确认不同于*《转发数据手册》*中“防止传输中的数据丢失”内介绍的索引器确认功能。更多有关 HTTP 事件收集器中索引器确认的信息，请参阅“启用索引器确认”。

8. 单击**下一步**。**输入设置**页面将会显示。

9. 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**。请确认端点的所有设置都是您想要的。如果您需要更改设置，单击页面顶部的灰色 **<** 按钮。

11. 如果所有设置都是您想要的，请单击**下一步**。设置成功的页面会加载并显示“事件收集器”生成的令牌值。您可从显示的字段中复制该令牌值并将其粘贴到另一文档中以供后续参考。请参阅“关于事件收集器令牌”。

修改事件收集器令牌

在您创建 HEC 令牌后可对其进行更改。请访问 HEC 管理页面并编辑令牌即可更改令牌的任意特性，包括其名称、描述、默认来源类型、默认索引和输出组。

要更改令牌的属性：

1. 请转到 HEC 管理页面。在**设置**菜单中选择**数据导入**。

2. 请选择 **HTTP 事件收集器**。

3. 请在列表中查找您要更改的令牌。

4. 请在此令牌的**操作**列中单击**编辑**。您也可单击令牌名称的链接。

5. 在**描述**字段中输入更新的文本可编辑令牌的描述。

6. （可选）在**来源**字段中输入文本可更新令牌的来源值。

7. （可选）在**来源类型**下拉列表中可选择一个不同的来源类型。请首先选择一个类别，然后在显示的弹出菜单中选择一个来源类型。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。该索引移动到控制的**已选择的索引**中。

9.（可选）请从**输出组**下拉列表中选择一个不同的输出组。

10.（可选）选择您是否想要为令牌启用索引器确认。

11.单击**保存**。该对话将关闭且 Splunk Web 将带您返回至 HEC 管理页面。

删除事件收集器令牌

如果您不再打算使用某个 HEC 令牌，您也可将其删除。删除某个 HEC 令牌不影响其它 HEC 令牌，也不会禁用 HEC 端点。

> **警告：**您无法撤消此操作。使用此令牌向您的 Splunk 部署发送数据的代理无法再使用此令牌进行验证。您必须生成一个新令牌并更改代理配置以使用新令牌值。

如需删除 HEC 令牌，步骤如下：

1.请转到 HEC 管理页面。请在**设置**菜单中选择**数据导入**。

2.请选择 **HTTP 事件收集器**。

3.请在列表中查找您要删除的令牌。

4.请在此令牌的**操作**列中单击**删除**。

5.请在“删除令牌”对话中单击**删除**。Splunk Enterprise 删除此令牌并带您返回到 HEC 管理页面。

启用和禁用“事件收集器”令牌

您可以从 HEC 管理页面内启用或禁用单个 HEC 令牌。更改某令牌的状态不会更改到其他令牌的状态。要启用或禁用所有令牌，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 令牌的活动状态，步骤如下：

1.请转到 HEC 管理页面。

2.请查找您要切换其状态的令牌。

3.请在此令牌的**操作**列中单击**启用**链接（如果此令牌已启动）或**禁用**链接（如果此令牌未启用）。此令牌状态立即切换且链接基于已更改的令牌状态更改为**启用**或**禁用**。

从开发人员的角度使用 HTTP 事件收集器

开发人员环境内有多个使用“HTTP 事件收集器”的选项。您可以使用我们的 Java、JavaScript (Node.js) 和 .NET

日志库（与常见的日志框架兼容）。也可以使用您喜好的 HTTP 客户端发出 HTTP 请求，并发送您的 JSON 编码事件。

利用操作系统中的 `curl` 指令，通过该指令发出 HTTP 呼叫可以轻松进行测试。

示例：

> **注意：**向端口 8088 发出该 POST 请求，并使用 HTTPS 进行传输。端口和 HTTP 协议的设置配置，可以独立于您的部署中其他任何服务器的设置。

JSON

以下 cURL 语句使用了一个 HTTP 事件收集器令牌示例 (`B5A79AAD-D822-46CC-80D1-819F80D7BFB0`)，并把 `https://localhost` 用作主机名。执行该语句前用您自己的值替换这些值。

JSON 请求

```
curl -k  https://localhost:8088/services/collector/event -H "Authorization: Splunk B5A79AAD-D822-46CC-80D1-
819F80D7BFB0" -d '{"event": "hello world"}'
```

注意：关键“事件”为必填。

JSON 响应

```
{"text": "Success", "code": 0}
```

更多信息

可以在“Splunk 开发人员门户”中找到更多有关开发人员如何使用“HTTP 事件收集器”的内容。使用“HTTP 事件收集器”的完整走查，请参阅“HTTP 事件收集器走查”。

Splunk Enterprise 通过 UDP 处理 syslog 数据的方式

本主题介绍当您使 Splunk Enterprise 在 UDP 网络端口上侦听 syslog 数据时其如何处理其接收的数据。Splunk Enterprise 可用作一个 syslog 服务器或一个 syslog 消息发送器。正常使用时，它不应替代这样的服务器。这是由于 Splunk Enterprise 默认修改 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）。

如果您使用的是 Splunk Cloud，则无法将您的部署配置为 syslog 服务器或 syslog 消息发送器，但您可以对 Splunk 通用转发器进行配置，以便在 UDP 网络端口上执行侦听，并把数据转发到您的 Splunk Cloud 部署。

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），请考虑使用诸如 `syslog-ng` 等工具在将原始数据保存到日志文件的同时把事件转发到您的 Splunk 部署。这样做的一大好处，是您可以等到之后有需要的时候再为日志文件建立索引。

请参阅本主题稍后的图表，该图表说明了 Splunk Enterprise 通过 UDP 处理 syslog 事件的方式。

Splunk Enterprise 处理 syslog 输入的方式

当您配置一个 UDP 网络输入以侦听 Splunk Enterprise 中的 syslog 时，任何通过该输入到达的 syslog 事件会接收一个时间戳和已连接的主机字段。Splunk Enterprise 在创建索引前会在每个事件前面加上这些字段。

在 inputs.conf 中设置 `no_appending_timestamp` 属性即可更改此行为。

如果数据包含 syslog 标题，Splunk Enterprise 会将其剔除，除非您在段落中设置了 `no_priority_stripping` 属性。

Splunk Enterprise 不以此方式修改 TCP 数据包。如果您通过 TCP 发送 syslog 数据，则 Splunk Enterprise 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，除非您对此进行阻止。

Splunk Enterprise 处理 syslog 输出的方式

Splunk Enterprise 也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，此主机会在其前面加上时间戳、优先级和连接的主机名称，皆为 Splunk Enterprise 实例。

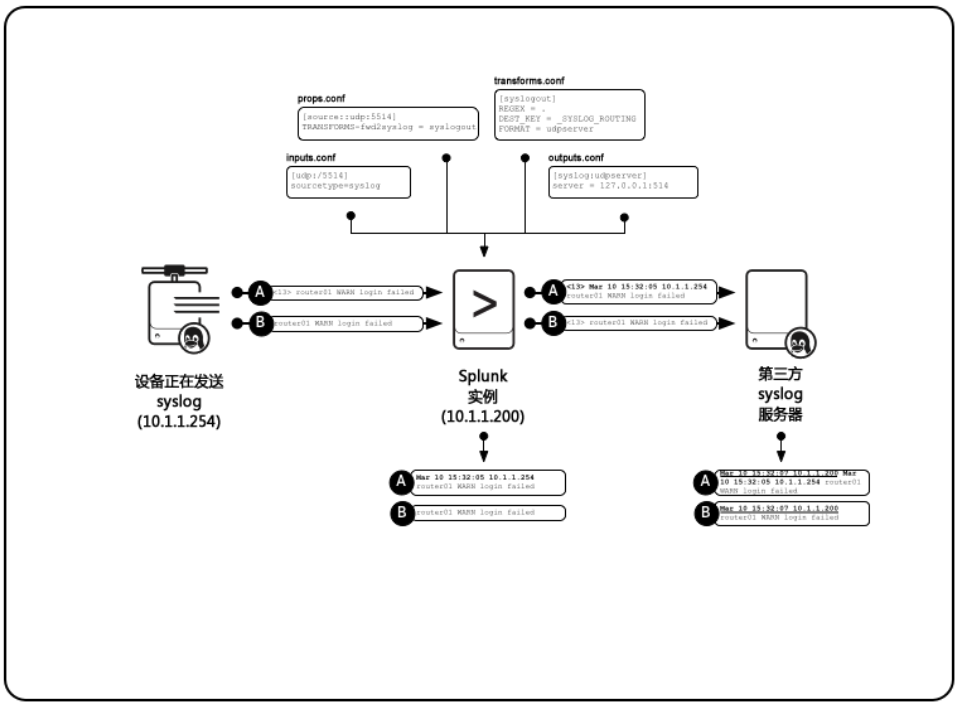
当您将事件转发到 syslog 服务器时，您也可以在事件前面加上时间戳和主机名称。

有关配置路由、过滤和来源类型使用的信息，请参阅*《转发数据》*手册中的“路由和过滤数据”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

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创建索引。另一方面，同一实例将事件转发到第二方、第三方 syslog 服务器。



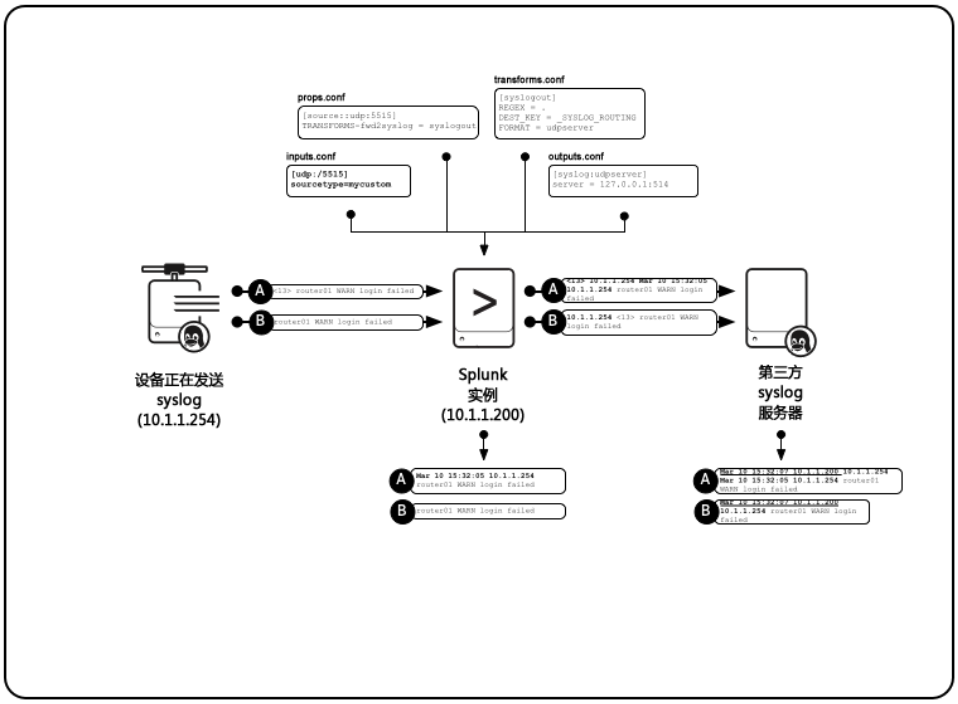
在本图中，消息 A 作为 syslog 事件的来源，消息 B 作为相似事件（其没有与之相关的优先级信息）的来源。一旦接收，Splunk Enterprise 用时间戳和生成事件的主机标记事件。

如果您已把此实例配置为转发器，Splunk Enterprise 会先通过添加优先级标题（您在 `outputs.conf` 中所指定的）转换事件，然后再将事件转发到 syslog 服务器。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

在本图中，已将 Splunk Enterprise 配置为使用非 syslog 来源类型。

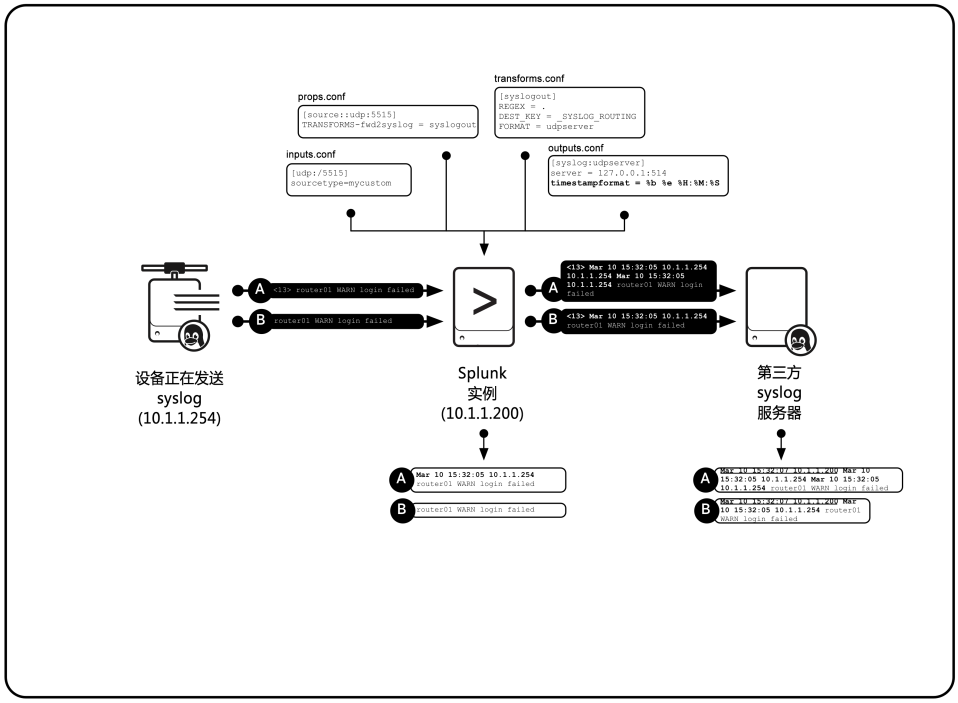
初始的消息 A 和消息 B 与第一个示例相同。在本示例中，Splunk Enterprise 在事件前面加上一个原始主机名称或 IP 地址。



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

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，您可以为事件创建时间戳。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。



向您的 Splunk 部署发送 SNMP 事件

简单网络管理协议 (SNMP) 陷阱是由远程设备发出的告警。本主题将介绍如何向 Splunk 部署发送 SNMP 陷阱。

注意：本主题中显示的过程（针对 *nix 和 Windows）仅用作示例。向 Splunk 部署发送 SNMP 陷阱的方法有很多种。例如，您可以不使用 Net-SNMP，而使用 Snare 或 SNMPGate 等其他工具把 SNMP 陷阱写入您可以监视的文件。

如何为 SNMP 陷阱创建索引

如果是 Splunk Enterprise，为 SNMP 陷阱建立索引的最有效方法是把陷阱写入 Splunk Enterprise 服务器上的某个文件，然后再对 Splunk Enterprise 进行配置，从而达到监视该文件的目的。如果您使用的是 Splunk Cloud，可以把数据写入受 Splunk 通用转发器监视的某个文件。

如需通过配置 Splunk Enterprise 来获取 SNMP 陷阱数据，步骤如下：

1.请配置远程设备以直接将它们的陷阱发送到 Splunk Enterprise 实例 IP 地址。SNMP 陷阱的默认端口为 `udp:162`。

2.请将 SNMP 陷阱写入 Splunk Enterprise 实例上的一个文件，如"将 SNMP 陷阱写入 Splunk Enterprise 服务器上的一个文件"中所述。

3.请将 Splunk 配置为监视该文件，如"监视文件和目录"中所述。

注意：本主题不包括 SNMP 轮询，这是查询远程设备的一种方式。

将 SNMP 陷阱写入 Splunk Enterprise 实例上的一个文件

使用您喜爱的 SNMP 软件将 SNMP 陷阱写入文件。有关可用 SNMP 软件的信息，请访问 SNMP 门户 (http://www.snmplink.org) 网站。

*对于 *nix*

您可以在 *nix 上使用 Net-SNMP 项目 `snmptrapd` 二进制把 SNMP 陷阱写入文件。

在系统上安装 `snmptrapd` 之前，请参阅随附 *nix 分发的 `snmptrapd` 版本的本地文档。您也可以参阅 `snmptrapd` 的手册页面。

最简单的配置如下：

```
# snmptrapd -Lf /var/log/snmp-traps
```

注意：`snmptrapd` 5.3 及更高版本不会自动接受并记录通知（即使未提供显式配置亦如此），而会对所有传入通知进行访问控制检查。如您运行 `snmptrapd` 时访问控制设置不当，该应用程序将不会处理所述陷阱。指定以下内容可避免发生这种情况：

```
# snmptrapd -Lf /var/log/snmp-traps --disableAuthorization=yes
```

如需查看 `snmptrapd` 版本，请从命令提示符运行 `snmptrapd --version`。

SNMP 疑难解答

如果无法向您的 Splunk 部署发送 SNMP 陷阱，请考虑以下方法：

- UDP 端口 162 是有特权的网络端口。如您需要使用该端口，则必须以根用户的身份运行 `snmptrapd`。
- 测试时，您可以使用 `-f` 标记将 `snmptrapd` 保持在前台。
- 您可以使用 `-Lo` 标记代替 `-Lf` 以记录到标准输出。
- 您可以使用 `snmptrapd` 命令生成示例陷阱，如下所示：

```
# snmptrap -v2c -c public localhost 1 1
```

对于 Windows

要将 SNMP 陷阱记录到 Windows 上的一个文件：

1.从 NET-SNMP 网站下载并安装适用于 Windows 的 `NET-SNMP` 最新版本。

注意：禁止将 OpenSSL 库安装在系统上，因为其与 NET-SNMP 相冲突。

2.把 `snmptrapd` 注册为一项服务，此时须使用 `NET-SNMP` 安装中包含的脚本。

3.编辑 `C:\usr\etc\snmp\snmptrapd.conf`：

```
snmpTrapdAddr [System IP]:162
authCommunity log [community string]
```

4. 默认的日志位置为 `C:\usr\log\snmptrapd.log`

使用管理信息库 (MIB)

管理信息库 (MIB) 提供 SNMP 陷阱所报告的数字对象 ID (OID) 与可读文本之间的映射。尽管 `snmptrapd` 完全可以在没有任何 MIB 文件的情况下工作，但结果无法以完全相同的方式显示。

接收 SNMP 陷阱的设备的供应商可以提供特定 MIB。例如，可以使用在线 Cisco SNMP Object Navigator 定位所有 Cisco 设备 MIB。

要添加新 MIB 文件：

1. 下载 MIB 文件并将其复制到 MIB 搜索目录。在 Net-SNMP 的 *nix 版本上，默认位置为 `/usr/local/share/snmp/mibs`。通过将 `-m` 参数提供给 `snmptrapd`，您可以设置不同的目录。

2. 指示 `snmptrapd` 加载 MIB，具体做法为：将冒号分隔列表传递给 `-m` 参数。

注意：

- 如果您为 `-m` 参数中的参数添加前导 '+' 字符，`snmptrapd` 将同时加载 MIB 和默认列表，而不是覆盖该列表。
- 特殊关键字 `ALL` 指示 `snmptrapd` 加载 MIB 目录中的所有 MIB 模块。

例如，要加载 MIB 目录中的所有 MIB 模块：

```
snmptrapd -m +ALL
```

获取 Windows 数据

使用 Splunk Enterprise 监视 Windows 数据

Splunk 软件可以为许多不同类型的 Windows 数据建立索引。该数据几乎可以是任何内容："事件日志"通道、"注册表"或 Active Directory。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。

下列专用输入仅适用于 Splunk Enterprise 的 Windows 安装。如果您使用的是 Splunk Cloud 而且想要监视这些输入，请使用 Splunk 通用转发器。

- **Windows 事件日志。**对计算机的任何可用事件日志通道上由 Windows 事件日志服务生成的事件进行监视。您可以在本地计算机上收集事件，也可以使用通用转发器或 Windows Management Instrumentation (WMI) 远程收集事件。
- **性能监视。**请在装有 Splunk Enterprise 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 Enterprise。您可以通过通用转发器或 WMI 本地或远程监视性能。
- **通过 WMI 远程监视。**Splunk Enterprise 可以使用 WMI 来访问远程计算机上的事件日志和性能数据。
- **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。可使用通用转发器从远程计算机收集注册表数据。
- **Active Directory 监视。**Splunk Enterprise 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。

Splunk App for Windows Infrastructure

Splunk App for Windows Infrastructure 可为 Windows 服务器和桌面管理提供数据导入、搜索、报表、告警和仪表板。您可以从一个位置对 Windows 操作系统进行监视、管理及故障排除。此应用包括 CPU、磁盘 I/O、内存、事件日志、配置及用户数据的输入，以及用于为 Windows 事件日志创建索引的基于 Web 的设置 UI。

在 Windows 上部署 Splunk Enterprise 的初始注意事项

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

- **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"有关确定如何监视远程 Windows 数据的注意事项。"
- **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创建索引时。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”。

从任何 Windows 服务器上收集数据的最有效方式，是在您想要从中收集数据的主机上安装通用转发器。通用转发器将使用有限的资源。某些情况下，例如“注册表”监视，您必须使用转发器，因为无法通过 WMI 收集“注册表”数据。

如何将 Windows 数据导入您的 Splunk 部署

您可以收集下列 Windows 统计数据：

- Windows 事件日志
- 文件系统更改
- Active Directory
- Windows Management Instrumentation (WMI) 基础设施上的数据
- 注册表数据
- 性能指标
- 主机信息
- 打印信息
- 网络信息

您只需在 Windows 主机上就可以收集所有这些类型的数据。其他操作系统无法直接收集 Windows 数据。您可以把 Windows 数据从 Windows 主机转发到未运行 Windows 的 Splunk Enterprise 实例。如果您使用的是 Splunk Cloud 而且想要监视这些输入，请使用 Splunk 通用转发器。

使用 Splunk Web 收集 Windows 数据

几乎所有 Windows 输入都允许您使用 Splunk Web 界面收集 Windows 数据。`MonitorNoHandle` 输入是一个例外，您必须使用配置文件才能设置该输入。

1.登入您的 Splunk 部署。

2.请单击右上角的**设置**，然后再单击**数据导入**。**数据导入**页面显示。

3.通过单击输入的“操作”列中的**新建**，您可查找您想要在可用输入列表中添加的输入。

4.有关您选择的输入类型，请遵循后续页面中的说明。

5.单击**保存**。

在大多数情况下，数据收集也马上开始。

使用配置文件收集 Windows 数据

如您无法使用 Splunk Web 创建和启用数据导入，例如当您使用通用转发器收集数据时，您必须使用配置文件。许多情况下，使用配置文件比使用 Splunk Web 提供更多的控制和可配置性。某些输入只可通过使用配置文件配置。

注意：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。

1.从命令提示符或 PowerShell 窗口转到 `%SPLUNK_HOME%\etc\system\local` 目录。

2.在此目录中编辑 `inputs.conf`。您可能需要创建该文件。

3.通过定义输入段落把输入添加到 `inputs.conf` 文件。

4.保存文件并将其关闭。

5.重新启动 Splunk 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

本主题将介绍监视远程 Windows 数据时的注意事项。

远程 Windows 数据概述

Splunk Enterprise 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
- 通过 Windows Management Instrumentation (WMI)

如果是 Splunk Cloud 部署，您必须使用 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 或 IBM BigFix 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 部署。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“转发器与通过 WMI 远程收集”部分。

使用 WMI 收集远程 Windows 数据

Windows Management Instrumentation (WMI) 框架允许 Splunk Enterprise 从远程 Windows 计算机收集几乎所有类型的数据。在此配置中，Splunk Enterprise 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。

此配置：

- 可为 Splunk Enterprise 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

这种收集方法有一些注意事项。请参阅本主题中的转发器与 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 Enterprise 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

当从 WMI 收集远程 Windows 数据时，请考虑以下内容：

远程 Windows 数据验证

Windows 需要远程操作验证。如果不了解 Splunk Enterprise 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。本部分提供有关收集远程 Windows 数据安全性的指南。

安装 Splunk Enterprise 时，您可以指定它以“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。

您指示 Splunk Enterprise 以其身份运行的用户决定了 Splunk Enterprise 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 Enterprise 用户帐户配置为具有您想要收集的数据来源的“最小权限”访问权限。这需要：

- 将用户添加到各种域安全组。
- 根据您需要访问的数据来源更改各种 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：

- 请确保 Splunk Enterprise 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，您应该在“本地安全策略”中为 Splunk Enterprise 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。与分发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。

本手册中与远程访问 Windows 计算机相关的各个“数据导入”主题将包含有关如何针对最小权限访问配置 Splunk Enterprise 的运行用户身份的其他信息和建议。请查看这些页面中的“安全与远程访问注意事项”部分。

使用受管系统帐户访问 Windows 数据

您可以在最新的 Windows Server 版本上使用受管服务帐户 (MSA) 来解决密码失效问题。请参阅*《安装》*手册中“Windows Server 2008 和 Windows 7 上的受管服务帐户”。

网络和 I/O 使用情况注意事项

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收集操作而言更好的选择。

还应考虑磁盘带宽问题。防毒扫描仪驱动程序以及介于 Splunk Enterprise 和操作系统之间的驱动程序应始终配置为忽略 Splunk Enterprise 目录和进程，无论安装类型为何。

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。

以下是在 WMI 与转发器之间权衡时的主要方面：

- 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：

- 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。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 Enterprise 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，最好选择使用部署：

- 您有权控制操作系统的基本版本，这与创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。

注意：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。如果在创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：

- 您无权控制基本操作系统版本，或者您在要从中收集数据的计算机上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 Enterprise 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。

此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收集类型。

数据来源与收集方法

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，您必须知道想要收集的事件日志的名称。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 Enterprise 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 CIFS（Common Internet File System 或 Server Message Block）作为应用层文件访问协议，并且 Splunk Enterprise 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 Enterprise 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 Enterprise 的 Windows 实例来获取 Windows 数据。您可以执行此操作，只需将 Splunk Enterprise 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 Enterprise 的非 Windows 实例即可。

然后通过以下两种方式进行操作：

- 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 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 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 Splunk 通用转发器才能收集 Active Directory 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。

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。

为什么监视 Active Directory？

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。Splunk Enterprise 允许您监视 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创建运行状况报表。

您需要什么来监视 Active Directory？

下表列出了监视 Active Directory 架构所需的权限。

活动	所需权限
监视 Active Directory 架构	* Splunk Enterprise 必须在 Windows 上运行 * Splunk Enterprise 必须以域用户身份运行 * Splunk Enterprise 运行时的用户身份必须对所有要监视的 AD 对象具有读取访问权限

监视 Active Directory 的技术注意事项

如想在监视 AD 时获得最佳结果，请阅读并了解下列事项：

- Splunk Enterprise 的 Windows 版本才有 AD 监视器。

- 如果无法通过 Splunk Enterprise 的 *nix 版本监视 AD 更改，您可以把 AD 数据从 Splunk Enterprise 的 Windows 版本转发到 *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 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 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 应以其身份运行的用户”。

AD 监视器不会查找 LDAP 参照

如果 AD 监视器发起 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应确定并修复 AD 内的配置问题。

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，有更多的选项可供使用，例如，可以为多个 DC 配置监视器。

使用 Splunk Web 配置 AD 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 **Active Directory 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 Active Directory。

选择输入来源

1. 在左窗格中选择 **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）您可以单击**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

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。

注意：主机只是设定生成事件中的**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“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，单击 **<** 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 Active Directory 节点。

使用配置文件配置 AD 监视

`inputs.conf` 配置文件控制着 Active Directory 监视配置。编辑 `inputs.conf` 的副本，该文件位于 `%SPLUNK_HOME%\etc\system\local` 目录中。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，请参阅本手册中的“配置文件优先顺序”。

1. 打开 `%SPLUNK_HOME%\etc\system\local\inputs.conf` 进行编辑。如果不存在，您可能需要创建此文件。
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，其标题如下所示：

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

在每个段落中，您可以指定：

属性	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，请为该属性指定唯一名称： • 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 树中创建索引的起始位置。根据 `monitorSubtree` 属性的配置情况，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。	“默认”索引。
`disabled`	否	Splunk 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

下列示例将示范如何使用 `inputs.conf` 来监视您所需的 AD 网络部分。

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 =
```

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
```

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

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

Splunk 软件可以为多种类型的 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 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 软件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|16010108151056.0Z
                lastKnownParent=stuff
                '''isDeleted=TRUE'''
```

同步事件

配置好 AD 监视输入后，Splunk 软件会尝试在启动时捕获 AD 元数据的基线。Splunk 软件将生成 `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 软件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“同步”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 对象列出了可用字段、必填字段和可选字段。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

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，并将这些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。

Splunk Enterprise 可以监视存储在本地计算机上的事件日志通道和文件，并可从远程计算机上收集日志。该事件日志监视器将作为输入处理器在 `splunkd` 服务内运行。它将针对您在 Splunk Enterprise 中定义的每个事件日志输入运行一次。如果您使用的是 Splunk Cloud，而且想监视事件日志通道，请使用 Splunk 通用转发器来收集数据并把数据发送到您的 Splunk Cloud 部署。

为什么监视事件日志？

Windows 事件日志是 Windows 主机操作的核心指标 - 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 Enterprise 的索引、搜索和报告功能使您的日志可以访问。

您需要什么来监视事件日志？

活动：	所需权限：
监视本地事件日志	Splunk Enterprise 必须在 Windows上运行 Splunk Enterprise 必须以“本地系统”用户身份运行才能读取所有本地事件日志
监视远程事件日志	通用转发器必须在您想从其中收集事件日志的 Windows 主机上运行。 或 Splunk Enterprise 必须在 Windows 上运行 且 Splunk Enterprise 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目标主机上 Windows Management Instrumentation (WMI) 的权限 AND Splunk Enterprise 以其身份运行的用户必须具有读取访问您想要的事件日志的权限

安全与远程访问注意事项

Splunk Enterprise 可使用 WMI 或转发器从远程计算机收集事件日志数据。Splunk 建议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”。

要将转发器安装在要收集事件日志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。

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 实例配置正确。您无法将 Splunk 平台安装为“本地系统”用户，而您安装的用户将决定 Splunk 软件会看到的事件日志。若想使用 WMI 正确收集远程数据，必须满足一些要求。更多有关这些要求的信息，请参阅本手册监视基于 WMI 的数据主题中的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。特别是，默认情况下本地管理员组或全局域管理员组的成员只能读取安全事件日志。

从远程 Windows 计算机收集事件日志

从远程 Windows 主机上收集数据的方法有很多种：

使用通用转发器

您可以在 Windows 主机上安装一个通用转发器，然后指示该转发器来收集事件日志。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。

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

1. 通过您想从中收集 Windows 事件日志的 Windows 主机从 Splunk 下载通用转发器软件。
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。搜索字符串示例如下：

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

使用 WMI

如果您选择使用 WMI 收集事件日志，则必须使用 Active Directory 域用户身份安装 Splunk Enterprise。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：

- 对想从中收集事件日志的服务器具有管理员访问权限。
- 了解安全描述定义语言 (SDDL)（外部链接）的工作原理以及如何使用它来分配权限。

如果您运行的是 Windows Vista、Windows 7、Windows Server 2008/2008 R2 或 Server 2012 R2，请使用 `wevtutil` 实用工具来设置事件日志安全性。

有关从远程 Windows 计算机收集数据的信息，请参阅有关确定如何监视远程 Windows 数据的注意事项。

1. 把 Splunk Enterprise 实例下载到 Windows 主机上。
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域用户名和密码，Splunk Enterprise 在运行时会使用这组域用户名和密码。
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. 根据配置远程事件日志监视中的介绍，使用 Splunk Web 添加远程事件日志输入。

一些系统事件日志中显示异常主机名

在 Windows Vista 和 Server 2008 R2 系统上，您可能会看到一些主机名称随机生成的事件日志。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 从上述 Windows 版本远程收集日志时才会发生。

使用 Splunk Web 配置事件日志监视

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

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**本地事件日志集合**。

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**，则选择或创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.在左窗格中，请选择**本地事件日志**

2.在**选择事件日志**列表框中，请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。**重要提示：**选择所有通道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. 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. 设置**主机**名称值。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"。

注意：主机只是设定生成事件中的**主机**字段。而不是引导 Splunk Enterprise 查找网络中的特定主机。

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"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，单击 **<** 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"事件日志"通道。

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. 请遵循转到"新增"页面"中的说明转到"新增"页面。

2. 在左窗格中，查找并选择**远程事件日志**。

3. 在**事件日志集合名称**字段中，为该输入输入一个您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。

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主机上。

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。

警告：勾选所有通道后，系统将为大量数据创建索引，可能会超过您的 Splunk 许可证所支持的数量。

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。使用逗号分隔多个主机。

10. 单击绿色**下一步**按钮。

11. 请遵循说明以指定输入设置，如"指定输入设置"中所述。

12. 请遵循说明以查看您的选择，如"查看您的选择"中所述。

使用 inputs.conf 配置事件日志监视

编辑 `inputs.conf` 以配置事件日志监视。有关使用 `inputs.conf` 配置数据导入的信息，请参阅本手册中"配置您的输入"。

注意：您始终可以通过在 `%SPLUNK_HOME%\etc\system\default` 中或《管理员手册》规范文件中查找示例来审阅某一配置文件的默认值。

通过编辑 `inputs.conf` 启用事件日志输入：

1. 使用 Notepad 或类似的编辑器打开 `%SPLUNK_HOME%\etc\system\local\inputs.conf` 进行编辑。如果不存在，您可能需要创建此文件。

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 Enterprise。

下一部分将介绍用于事件日志监视的可用配置值。

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创建索引，来监视这些文件。

Splunk Enterprise 使用位于 `inputs.conf` 内的下列段落来监视默认的 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。如下所示，您可以在导入日志后把这些日志添加到 `inputs.conf` 的本地副本中:

```
[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"全名事件日志"属性确保自己在某 `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：

1.启动"事件查看器"。

2.展开 `Applications and Services Logs > Microsoft > Windows > TaskScheduler`。

3.右击 `Operational`，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 `WinEventLog://` 段落：

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

禁用事件日志段落

要禁用为某事件日志创建索引，请在该事件日志列表下方添加 `disabled = 1`，该列表位于 `%SPLUNK_HOME%\etc\system\local\inputs.conf` 中的段落内。

Splunk 软件使用 inputs.conf 中的下列属性来监视"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`	如何为事件建立索引。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）。	0

	您无法在将该属性设置为 `1` 的同时也把 `start_from` 属性设置为 `newest`。	
`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。	5
`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`（意味着选择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`
`evt_resolve_ad_obj`	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”名称和/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 = $logonserver`。 您可在各个格式前加上两个反斜杠字符。该属性无默认值。	N/A
`evt_dns_name`	解析 AD 对象所需绑定的域的完全限定 DNS 名称。	N/A
`suppress_text`	是否包含安全事件随附的消息文本。值为 1 时禁止消息文本；值为 0 时保留消息文本。	0
`whitelist`	是否要为与指定文本字符串相匹配的事件建立索引。这是可选属性。 您可指定两种格式中的一种： • 一个或多个“事件日志”事件代码或事件 ID（事件代码/ID 格式）。 • 一个或多个密钥和正则表达式集（高级过滤格式）。 您不可以在单个条目中混合格式。您也不可以在相同段落中混合格式。 Splunk 软件会先处理白名单，然后再处理黑名单。如果没有任何白名单，Splunk 软件将为所有事件建立索引。 当您使用事件代码/ID 格式： • 对于多个代码/ID，用逗号将列表分隔。 • 对于范围，使用连字符（例如，"0-1000,5000-1000"）。 当使用高级过滤格式时： • 在密钥和代表您的过滤器的正则表达式之间使用 '='（例如，“白名单 = EventCode=%^1([8-9])$%” • 您可在单个高级过滤条目中具有多个密钥/正则表达式集。Splunk Enterprise 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。（高级过滤格式。） 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 软件会先处理白名单，然后再处理黑名单。如果没有任何黑名单，Splunk 软件将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"）。 当使用高级过滤格式时： • 在密钥和代表您的过滤器的正则表达式之间使用 '='（例如，“黑名单 = EventCode=%^1([8-9])$%” • 您可在单个高级过滤条目中具有多个密钥/正则表达式集。Splunk 软件从逻辑上连接该集。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 个黑名单，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。这是可选属性。 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。	0 (false)
`index`	此输入应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（指示输入应该运行）和 1（指示输入不应该运行）。	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，方法是对一组文件和/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` 中使用这三个属性。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 上的“如何使用审核安全事件”(http://technet.microsoft.com/zh-cn/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。

注意：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。该消息文本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`。

```
[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 '**whitelist**' 和 '**blacklist**' 创建高级过滤*

除了仅基于事件代码执行过滤，您还可以使用 `whitelist` 和 `blacklist` 属性对传入事件执行高级过滤。为此，指定属性中的密钥/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：

关键	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”服务接收和记录的事件的时间。Splunk Enterprise 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 及更早版本的主机或运行 Windows XP 及更早版本的客户端上可用。请参阅 MSDN 上的 "Win32_NTLogEvent class (Windows)" (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”。
Message	事件中消息的文本。
OpCode	事件的安全级别（事件查看器中的 "OpCode"）。
RecordNumber	Windows 事件日志记录编号。Windows 主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" (http://msdn.microsoft.com/en-us/library/windows/desktop/aa394507%28v=vs.85%29.aspx)。
SidType	代表与事件相关联的 SID 类型的数字值。请参阅 MSDN 上的 "Win32_UserAccount class" (http://msdn.microsoft.com/en-us/library/windows/desktop/aa394507%28v=vs.85%29.aspx)。
SourceName	生成事件的实体的数据来源（事件查看器中的“数据来源”）
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”过滤它们（使用“任务类别”字段）。请参阅 MSDN 上的事件类别 (Windows) (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（“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 或更新版本的主机或运行 Windows Vista 或更新版本的客户端上可用。请参阅 MSDN 上的 "Win32_NTLogEvent class (Windows)" (http://msdn.microsoft.com/en-us/library/aa394226(v=vs.85).aspx)。
User	与事件相关联的用户。与事件查看器中的“用户”相关联。

和 `<regular expression>` 是任何代表您想要包括（通过 `whitelist` 属性使用时）或排除（通过 `blacklist` 属性使用时）的过滤器的有效正则表达式。

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。例如，此条目：

```
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 个单独的白名单或黑名单条目。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>
```

注意：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])$"
whitelist1 = EventCode="^2([0-5])$"
```

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 (GUIDs) 和安全标识符 (SIDs) 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 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
```

evt_dc_name 及 evt_resolve_ad_obj 属性的使用限制

当您使用 evt_resolve_ad_obj 和 evt_dc_name 属性时：

- 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 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创建索引。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 中显示事件

要使 Splunk Enterprise 在 XML 中生成事件，请使用 renderXml 属性：

```
[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>410065004C006F006F006B007500700053007600630002F0031000000</Binary>
  </EventData>
</Event>
```

当您指示 Splunk Enterprise 在 XML 中呈现事件时，无论主机系统区域为何种语言，XML 事件内的事件密钥均以英语呈现。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 spéciale
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
                                           SeSystemEnvironmentPrivilege
                                           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
                                        SeSystemEnvironmentPrivilege
                                        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 CLI 配置事件日志监视

您可使用 CLI 配置本地事件日志监视。使用 CLI 之前请先在 `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"。

注意：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：

```
> splunk list eventlog
```

您还可通过指定特定通道的名称来将其列出：

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

要启用"事件日志"通道：

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

要禁用通道：

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

为导出的事件日志（.evt 或 .evtx）文件建立索引

要为导出的 Windows 事件日志文件建立索引，请根据监视文件和目录的说明来监视包含导出文件的目录。

警告：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。

约束

- 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"消息"字段。这意味着，"消息"字段的内容不会显示在您的 Splunk 索引中。
- 如果 .evtx 文件导出自运行 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 无法为这类文件创建索引。
- 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 文件创建索引。
- 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 .evt 或 .evtx 文件位于收集该文件的计算机的主要区域设置/语言，则 Splunk Enterprise 将为该 .evt 或 .evtx 文件创建索引。
- 从其他主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生成这些文件的主机所特定的。如果不变更格式，其他主机将无法处理这些文件。

注意：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 等的差异也可能会产生这种影响。

问答

有什么问题吗？请访问 Splunk Answers，查看 Splunk 社区有哪些与 Windows 事件日志相关的问题和答案。

监视文件系统更改

Splunk Enterprise 支持通过安全事件日志通道监视 Windows 文件系统更改。要启用对文件和目录更改的监视，请先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。

您需要什么来监视文件系统更改？

活动：	所需权限：
监视文件系统更改	• Splunk Enterprise 必须在 Windows 上运行，且 • Splunk Enterprise 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，且 • 您必须对希望 Splunk Enterprise 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，方法是对一组文件和/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` 中使用这三个属性。

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属性。该属性列表还仅为 inputs.conf 可用属性的一个子集。其他属性请阅读本手册中的监视 Windows 事件日志数据。

属性	描述	默认
`whitelist`	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（高级过滤格式）。 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 Enterprise 首先处理白名单，然后再处理黑名单。如果没有显示白名单，Splunk Enterprise 会为所有事件建立索引。 当使用事件代码/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"）。 当使用高级过滤格式时： • 在密钥和代表您的过滤器的正则表达式之间使用 '='（例如，“白名单 = EventCode=%^1([8-9])$%” • 您可在单个高级过滤条目中具有多个密钥/正则表达式集。Splunk Enterprise 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（高级过滤格式）。 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 Enterprise 首先处理白名单，然后再处理黑名单。如果没有显示白名单，Splunk Enterprise 会为所有事件建立索引。	

	当使用事件代码/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"）。 当使用高级过滤格式时： • 在密钥和代表您的过滤器的正则表达式之间使用 '='（例如，“白名单 = EventCode=%^1([8-9])$%” • 您可在单个高级过滤条目中具有多个密钥/正则表达式集。Splunk Enterprise 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 个黑名单，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 `whitelist` 和 `blacklist` 创建高级过滤

除了仅基于事件代码执行过滤，您还可以使用 `whitelist` 和 `blacklist` 属性对传入事件执行高级过滤。为此，指定属性中的密钥/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：

关键	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”服务接收和记录的事件的时间。Splunk Enterprise 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 及更早版本的服务器计算机或运行 Windows XP 及更早版本的客户端上可用。请参阅 MSDN 上的 Win32_NTLogEvent class (Windows) (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”。
Message	事件中消息的文本。
OpCode	事件的安全级别（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 (http://msdn.microsoft.com/en-us/library/windows/desktop/aa394507%28v=vs.85%29.aspx)
SidType	代表与事件相关联的 SID 类型的数字值。请参阅 MSDN 上的 Win32_UserAccount class (http://msdn.microsoft.com/en-us/library/windows/desktop/aa394507%28v=vs.85%29.aspx)
SourceName	生成事件的实体的数据来源（事件查看器中的“数据来源”）
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”过滤它们（使用“任务类别”字段）。请参阅 MSDN 上的事件类别 (Windows) (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（“错误”、“警告”、“信息”、“审计成功”和“审计失败”）中的

	一种事件类型的数字值。仅在运行 Windows Server 2008 或之后版本的服务器计算机或运行 Windows Vista 或之后版本的客户端上可用。请参阅 MSDN 上的 Win32_NTLogEvent class (Windows) (http://msdn.microsoft.com/en-us/library/aa394226(v=vs.85).aspx)。
User	与事件相关联的用户。与事件查看器中的“用户”相关联。

`<regular expression>` 是任何代表您想要包括（通过 `whitelist` 属性使用时）或排除（通过 `blacklist` 属性使用时）的过滤器的有效正则表达式。

要了解有关正则表达式以及如何使用正则表达式的更多信息，请访问 Regularexpressions.info (http://www.regular-expressions.info) 网站。

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：

```
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 个单独的白名单或黑名单条目。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>
```

注意：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。例如，如果您指定：

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

Splunk 软件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。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

您可监视一组文件或目录的文件系统更改。

1.请确认您有管理员权限。

2.按照 MS Technet 上的“如何使用审计安全事件”(http://technet.microsoft.com/zh-cn/library/cc727935%28v=ws.10%29.aspx) 启用安全审计。

3.将 Splunk Enterprise 的事件日志监视器输入配置为监视安全事件日志通道。

注意：有关如何配置“事件日志”监视器输入的说明，请参阅本手册中的监视 Windows 事件日志数据。

文件系统更改监视器示例

以下是 `inputs.conf` 段落，显示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 Enterprise 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无代理式访问。您可以从环境中的所有 Windows 计算机提取事件日志，而无需在这些计算机上进行任何安装。

如果可能，请使用**通用转发器**而非 WMI 从远程计算机上收集数据。在许多情况中，WMI 的资源负载量会超过 Splunk 通用转发器。如果您要从每个主机或从非常繁忙的主机（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。请参阅本手册中的有关确定如何监视远程 Windows 数据的注意事项。如果您使用的是 Splunk Cloud，必须用通用转发器从 WMI 提供程序收集数据并把数据转发到您的 Splunk Cloud 部署。

基于 WMI 的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运行，该进程名为 `splunk-wmi.exe`。它为**脚本式输入**。

您需要什么来监视基于 WMI 的数据？

下面是监视基于 WMI 的数据的基本最低要求。根据要监视的日志或性能计数器，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，请参阅本主题后面的“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 Enterprise 必须在 Windows 上运行 * Splunk Enterprise 必须以至少具有 WMI 的读取访问权限的域用户身份运行 * Splunk Enterprise 必须以具有所需事件日志的相应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 Enterprise 必须在 Windows 上运行 * Splunk Enterprise 必须以至少具有 WMI 读取访问权限的域用户身份运行 * Splunk Enterprise 必须以具有“性能数据助手”库适当访问权限的域用户身份运行。

安全与远程访问注意事项

必须正确配置 Splunk Enterprise 和您的 Windows 网络才能实现 WMI 数据访问。在尝试使用 Splunk Enterprise 获取 WMI 数据前，查看以下前提条件。

在 Splunk Enterprise 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用户身份安装 Splunk。
- 运行 Splunk Enterprise 的用户必须是 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。
- Splunk 用户还必须是运行 Splunk Enterprise 的计算机上本地管理员组的成员。
- 运行 Splunk Enterprise 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 Enterprise 实例与目标计算机必须属于相同 AD 域或林。

Splunk Enterprise 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 Splunk Enterprise 作为本地系统用户进行安装，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”帐户访问另一台主机的权限。

通过执行以下操作之一，您即可向 Splunk 用户授予访问 WMI 提供程序的权限：

- 将它添加到要轮询的每个成员主机上的本地管理员组（出于安全原因，不建议这样做）。
- 将该用户添加到域管理员全局组（出于安全原因，不建议这样做）。
- 按照如下详细说明分配最小权限（推荐方式）。

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 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 Enterprise 配置的运行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。仅授予所有 Windows 资源（包括 WMI）的最小访问权限。要授予此类型的访问权限，请按照以下检查表进行操作。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"。

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

要在域范围内部署这些用户权限分配，请使用**域安全政策** (`dompol.msc`) Microsoft 管理控制台 (MMC) 管理单元。部署完毕后，成员主机将在下一个 AD 复制周期期间继承网络上的这些权限分配。请重新启动这些计算机上的 Splunk Enterprise 实例以使更改生效。

要将该访问权限特定扩展到域控制器，请使用**域控制器安全政策** (`dcpol.msc`) 管理单元分配权限。

- **本地安全策略权限。**Splunk 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
- 替换进程级别令牌

- **分布式组件对象模型 (DCOM) 配置和权限。**必须在要监视的每个计算机上都启用 DCOM。另外，必须为 Splunk Enterprise 用户分配访问 DCOM 的权限。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计算机上的"分布式 COM 用户"本地组中，然后将 Splunk Enterprise 用户添加到"分布式 COM 用户"域全局组中。有关授予 Splunk Enterprise DCOM 的用户访问权限的高级选项，请参阅 MSDN 上的"确保远程 WMI 连接安全"(http://msdn.microsoft.com/en-us/library/aa393266(VS.85).aspx)。

- **性能监视器配置和权限。**要使 Splunk Enterprise 能够通过 WMI 访问远程性能对象，Splunk Enterprise 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"域全局组嵌套到每个成员主机上的"性能日志用户"本地组，然后向全局组分配用户。

- **WMI 命名空间安全性。**Splunk Enterprise 访问的 WMI 命名空间（最常使用 `Root\CIMV2`）必须具备适当的权限。因为没有全局 WMI 安全性，必须在贵公司的每个主机上手动设置这些权限。使用 WMI 安全性 MMC 管理单元 (`wmimgmt.msc`) 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

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。请参阅 Microsoft TechNet 上的"管理 WMI 安全性" (https://technet.microsoft.com/en-us/library/cc731011.aspx)。

注意：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，没有标准实用工具。但是，MSDN 博客上的"通过 GPO 设置 WMI 命名空间安全"(http://blogs.msdn.com/spatdsg/archive/2007/11/21/set-wmi-namespace-security-via-gpo-script.aspx) 提供了一些说明指导，介绍如何创建一个可置于组策略对象 (GPO) 中的启动脚本；一旦应用到所需主机，GPO 将设置命名空间安全。之后您可以在域范围内部署此 GPO 或将其部署到一个或多个组织单元 (OU)。

- **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计算机都设置此例外。更多详细信息请参阅 MSDN 上的 "Connecting to WMI Remotely Starting with Vista" (http://msdn.microsoft.com/en-us/library/aa822854(VS.85).aspx)。

- **用户访问控制 (UAC) 配置。**如果您运行的是 Windows Vista、Windows 7、Windows 8.1 或 Windows Server 2008 或 2012 系列，UAC 会影响 Windows 分配权限的方式。请参阅 MSDN 上的"用户帐户控制和 WMI"(http://msdn.microsoft.com/en-us/library/aa826699(v=vs.85).aspx)。

测试 WMI 提供程序的访问权限

在您配置完 WMI 并为 Splunk 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 Enterprise 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 用户身份登录到运行 Splunk Enterprise 的计算机。

2.打开命令提示符（单击**开始 -> 运行**，然后键入 cmd）。

3.转到 Splunk Enterprise 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 Splunk Enterprise 当前用来存储其数据的位置：

```
> splunk show datastore-dir
```

注意：请记住 Splunk Enterprise 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 Enterprise 用来临时存储其数据的位置：

```
> splunk set datastore-dir %TEMP%
```

注意：在此示例中，数据存储区目录被设置为在 TEMP 环境变量中指定的当前目录。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 Enterprise：

```
> splunk restart
```

注意：Splunk Enterprise 重新启动可能需要一点时间。

7.一旦 Splunk Enterprise 完成重启，请测试访问 WMI 提供程序的权限，用远程主机的名称替换 <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 Splunk Enterprise 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 Enterprise 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 Splunk Enterprise：

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

配置基于 WMI 的输入

在 Windows 上，Splunk Enterprise 中的所有远程数据收集需通过 WMI 提供程序或转发器发生。请参阅本手册中的有关确定如何监视远程 Windows 数据的注意事项。

您可以在 Splunk Web 中或通过编辑配置文件来配置基于 WMI 的输入。使用配置文件时，有更多的选项可供您选择。

使用 Splunk Web 配置基于 WMI 的输入

要添加基于 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。请参阅使用 Splunk Web 配置远程 Windows 性能监视。也请参阅配置远程 Windows 事件日志监视

使用配置文件配置基于 WMI 的输入

wmi.conf 处理远程数据收集配置。请参阅此文件以查看基于 WMI 的输入的默认值。如您想要更改默认值，请编辑 wmi.conf 文件的副本，该文件位于 %SPLUNK_HOME%\etc\system\local\ 内。请仅为给定数据导入类型设置您想要更改的属性的值。请参阅*《管理员》*手册中的“关于配置文件”。

wmi.conf 包含若干段落：

- 指定全局 WMI 参数的 [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参数都是可选的。如果此段落不显示，Splunk Enterprise 会采用系统默认值。

如果 Splunk Enterprise 无法连接到定义的 WMI 提供程序，它会在 splunkd.log 中生成一个错误：

```
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)
```

以下属性控制在发生错误的情况下 Splunk Enterprise 如何重新连接到给定 WMI 提供程序。

属性	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 Enterprise 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 Enterprise 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

特定于输入的段落指示 Splunk Enterprise 如何连接到 WMI 提供程序。这些段落由用于指定 Splunk Enterprise 应收集的数据类型的两个属性之一定义。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 Enterprise 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：

- **事件日志。** event_log_file 属性指示 Splunk Enterprise 从段落中定义的来源获取事件日志数据。
- **Windows 查询语言 (WQL)。** wql 属性指示 Splunk Enterprise 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，您必须使用该属性。

切勿在一个段落中同时定义这两个属性。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属性包括：

属性	描述	默认值
server	要从中获取数据的逗号分隔的主机列表。如果缺少此属性，Splunk Enterprise 会假定您希望连接到本地计算机。	本地主机
interval	指示 Splunk Enterprise 轮询新数据的频率（以秒为单位）。如果此属性不存在或未定义，则段落所定义的输入将不运行。	N/A
disabled	指示 Splunk Enterprise 是否启用此属性。请将此参数设置为 1 以禁用输入，请将此参数设置为 0 以启用输入。	0（启用）

特定于事件日志的参数包括：

属性	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。	N/A
current_only	是否收集仅在其运行时发生的事件。如果 Splunk Enterprise 停止时产生了事件，则当 Splunk Enterprise 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 Enterprise 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：

属性	描述	默认值

wql	有效 WQL 语句。	N/A
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 Enterprise 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。Microsoft 提供了一个包含适用于 Windows XP 和更高版本 Windows 的命名空间的列表 (http://msdn.microsoft.com/en-us/library/aa394084(VS.85).aspx)。	\\<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”。将此属性设置为 1 时指示 Splunk Enterprise 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 属性决定段落收集基于 WMI 的数据时预计使用的查询类型。当您将该属性设置为 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 Enterprise 不会更新 WMI 检查点文件 - 确定是否已对 WMI 数据建立检索的文件。设计就是如此 - 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。这意味着，每次段落运行时，Splunk Enterprise 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。这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 Splunk Enterprise 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）的数据创建索引时，Splunk Enterprise 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 Enterprise 会把来源类型设置为 `WinEventLog:<name of log file>`。例如，`WinEventLog:Application`。
- 对于 WQL 数据，Splunk Enterprise 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 Enterprise 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 Enterprise 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 Enterprise 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。之后，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 Enterprise 会捕获执行更改的进程的名称以及所更改项的整个路径。

Windows“注册表”输入监视器作为一个进程运行，称为 `splunk-regmon.exe`。

如果您使用的是 Splunk Cloud，必须用通用转发器从“Windows 注册表”中收集数据并把数据转发到您的 Splunk Cloud 部署。

为什么监视注册表？

注册表可能是 Windows 操作最常用但最少被了解的组件。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 Enterprise 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 Enterprise 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。

您需要什么来监视“注册表”？

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 必须在 Windows 上运行 且 * Splunk Enterprise 必须以本地系统用户身份运行 或

	* Splunk Enterprise 必须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`）和/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 Enterprise 需要监视大量数据。为实现最佳性能，可以通过配置 `inputs.conf` 过滤 Splunk Enterprise 为其创建索引的“注册表”数据量。

同样，您可以在第一次启动 Splunk Enterprise 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。

快照进程会占用一些 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 Enterprise 监视的特定范围之后，再获取基准快照。

在 Splunk Web 中启用注册表监视

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. 在左窗格中，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 Enterprise 监视的注册表项的路径。

4. （可选）如果您不确定此路径，请单击**浏览**按钮以选择您希望 Splunk Enterprise 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`。

5. 在**注册表配置单元**窗口中，单击所需注册表项的名称以选择此项。

该限定的键名即会显示在窗口底部的**限定名**字段中。

6. 单击**选择**以确认所做选择并关闭窗口。

7. （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**。

注意：监视子节点节点决定了 Splunk Enterprise 添加到 `inputs.conf` 文件中的内容，该文件是 Splunk Enterprise 在您于 Splunk Web 中定义“注册表”监视器输入时创建的。

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 Enterprise 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 (`\\\\?.*`) 用于过滤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 Enterprise 会向输入段落添加一个正则表达式，以过滤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所需项，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 Enterprise 才会添加此正则表达式。

8.在**事件类型**下方，针对选定注册表配置单元选择希望 Splunk Enterprise 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 Enterprise 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创建了一个新的子项，Splunk Enterprise 会生成 Create 事件。
Delete	当程序执行 DeleteValue 或 DeleteSubKey 方法时，Splunk Enterprise 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 Enterprise 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 Enterprise 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 Enterprise 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 Enterprise 会生成 Query 事件。

9.在**进程路径**字段中输入相应值，您便可指定 Splunk Enterprise 监视哪些进程对注册表的更改。或者，保留 `C:\.*` 的默认值来监视所有进程。

10.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**。

注意：基准快照是在获取快照时整个注册表的索引。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。

11.单击绿色**下一步**按钮。

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.设置**主机**名称值。此设置有多个选项供您选择。更多有关设置主机值的信息，请参阅关于主机。

> **注意：主机**只是设定生成事件中的**主机**字段。而不是引导 Splunk Enterprise 查找网络中的特定主机。

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“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，单击 **<** 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“注册表”节点。

查看注册表更改数据

要查看 Splunk Enterprise 已为其创建索引的“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。

属性	描述
event_status	尝试更改注册表的结果。这应始终为 "(0) The operation completed successfully."。否则说明注册表可能存在一些问题，最终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。

筛选传入注册表事件

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。此文件需驻留在运行注册表监视的服务器上的 $SPLUNK_HOME\etc\system\local\ 中。

inputs.conf 包含您专为优化和过滤想要 Splunk 监视的“注册表”配置单元而创建的特定正则表达式。

inputs.conf 中每个段落都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。
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 or 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\\.*（请注意末尾处的句号和星号）生成包含登录用户安全标识符 (SID) 的事件。 • 或者，您也可以通过使用 \\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。可以是一个或多个 delete, set, create, rename, open, close 或 query。此处的值必须为 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	Splunk Enterprise 在重新获取快照之前需要关闭的时间（以秒为单位）。默认值为 86,400 秒（1 天）。
disabled	是否启用过滤器。设置为 1 时禁用过滤器；设置为 0 时启用过滤器。

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 Enterprise 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。它还确定了重新获取快照的时间线：默认情况下，如果自上个检查点之后 Splunk Enterprise 的停止时间超过了 24 小时，它会重新获取基准快照。您可以为 inputs.conf 中的所有过滤器自定义该值，只要设置好 baseline_interval 的值（单位为秒）。

监视 Windows 性能

Splunk Enterprise 不但支持实时监视所有 Windows 性能计数器，而且还支持本地和远程收集性能数据。

利用 Splunk Enterprise 的性能监视实用工具，您可以在 Web 界面中实现“性能监视器”的功能。Splunk Enterprise 使用性能数据助手 (PDH) API 在本地计算机上实现性能计数器查询。

可用于 Splunk Enterprise 的性能对象、计数器和实例类型取决于系统上安装的性能库。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，请参阅 MSDN 上的 "Performance Counters" (http://msdn.microsoft.com/en-us/library/aa373083%28v=VS.85%29.aspx)。

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 WMI (Windows Management Instrumentation) 实现远程性能监视，但这要求 Splunk Enterprise 以具有相应 Active Directory 凭据的用户身份运行。如果您使用的是 Splunk Cloud，而且想监视 Windows 性能指标，必须使用 Splunk 通用转发器来收集数据并把数据发送到您的 Splunk Cloud 部署。

性能监视器输入将作为一个进程运行，称为 `splunk-perfmon.exe`。它将按照在输入中指定的时间间隔，针对每个定义的输入运行一次。您可以使用 Splunk Web、 `inputs.conf`（用于本地性能数据）或 `wmi.conf`（用于来自远程计算机的性能数据）来配置性能监视。

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 会生成许多与系统运行状况相关的数据。正确分析该数据有助于区分系统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权限。根据要监视的性能对象或计数器，您可能需要其他权限。

有关监视性能指标所需内容的其他信息，请阅读本主题后面的“安全与远程访问注意事项”。

活动	所需权限
监视本地性能指标	* Splunk Enterprise 必须在 Windows 上运行。 * Splunk Enterprise 必须以“本地系统”用户身份运行。
通过 WMI 监视其他计算机上的远程性能指标	* Splunk Enterprise 必须在 Windows 上运行。 * Splunk Enterprise 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。 * Splunk Enterprise 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

Splunk Enterprise 可使用转发器或 WMI 从远程计算机收集数据。Splunk 建议使用通用转发器将远程计算机中的性能数据发送到索引器。请参阅《通用转发器手册》。

如果您将转发器安装在要收集性能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。

如果您希望 Splunk Enterprise 使用 WMI 从远程计算机获取性能数据，则您必须配置 Splunk Enterprise 和您的网络。不能将 Splunk Enterprise 作为“本地系统”用户进行安装，您选择的用户身份决定了 Splunk Enterprise 将会看到的内容。请参阅本手册“监视 WMI 数据”主题中的安全与远程访问注意事项。

使用有效用户身份安装 Splunk Enterprise 之后，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）

启用本地 Windows 性能监视

您可以在 Splunk Web 中或使用配置文件来配置本地性能监视。

Splunk Web 是添加性能监视数据导入的首选方法。这是因为使用配置文件可能会发生键入错误，而且，完全按照性能监视器 API 的定义指定性能监视器对象，这一点非常重要。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

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**本地性能监视**。

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**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**，则选择或创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.在左窗格中，查找并选择**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。

注意：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，请为每个对象创建其他数据导入。

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。

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.要取消选择一个计数器，在“可用项目”窗口上单击该通道的名称。Splunk Enterprise 将计数器从“已选计数器”窗口移动到“可用计数器”窗口。

7.要选择或取消选择所有计数器，单击“添加所有”或“删除所有”链接。

警告：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。

8.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。

注意："_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，在安装了两个磁盘的系统上，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- 且 "_Total" 实例是两个物理磁盘实例的平均值。

9.在**轮询间隔**字段中，为输入在轮询尝试之间输入时间（秒）。

10.单击绿色**下一步**按钮。

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.设置**主机**名称值。此设置有多个选项供您选择。更多有关设置主机值的信息，请参阅关于主机。

> **注意：主机**只是设定生成事件中的**主机**字段。而不是引导 Splunk Enterprise 查找网络中的特定主机。

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“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，单击 **<** 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性能指标。更多有关从文件和目录获取数据的信息，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` 控制着性能监视数据。要使用配置文件设置性能监视，请创建或编辑 `inputs.conf`，路径为 `%SPLUNK_HOME%\etc\system\local`。如您之前不曾使用过配置文件，请参阅"关于配置文件"。

`[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，您可以指定以下属性。

属性	是否必需？	描述
`interval`	是	轮询新数据的频率（以秒为单位）。如果该属性不存在，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（包括大小写）的字符串，或者使用正则表达式来引用多个对象。如果此属性不存在或未定义，则输入将不运行，因为没有默认值。
`counters`	是	与在 `object` 中指定的对象相关联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。如果此属性不存在或未定义，则输入将不会运行，因为没有默认值。
`instances`	否	与在 `counters` 中指定的性能计数器相关联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属性，则默认使用星号。
`index`	否	要将性能计数器数据发送到的索引。如果不存在，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。设置为 1 时禁用此段落；设置为 0 时启用此段落。如果不存在，则默认为 0（启用）。
`showZeroValue`	否	**高级选项。** Splunk Enterprise 是否应该收集值为零的事件。 设置为 1 则收集零值事件，设置为 0 则忽略零值事件。如果未出现，默认设置为 0（忽略零值事件）。
`samplingInterval`	否	**高级选项。** Splunk 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 (ms)，必须低于您使用 `interval` 属性指定的值。
`stats`	否	**高级选项。** 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（禁用）。
`mode`	否	**高级选项。** 启用高性能取样时，此属性控制 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（`perfmonMSStats` 为 `multiMS` 的输出模式，`perfmonMKMSStats` 则为 `multikvMS` 的输出模式），统计信息事件有特殊的 sourcetype。 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	**高级选项。**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，它指示 Splunk 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。它还会禁用与 object 和 counter 属性匹配的正则表达式及通配符。 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 属性。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。
formatString	否	**高级选项。**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 属性控制着作为每个事件之一部分打印的有效数字的数量。 该属性使用来自 C++ 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。可在 cplusplus.com 上的 "printf - C++ reference"(http://www.cplusplus.com/reference/cstdio/printf/) 中找到有示例的参考。 当指定格式时，请不要使用引号 ("")。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 属性，该属性位于 inputs.conf 中的段落内。无法在 Splunk Web 中配置 useEnglishOnly。

注意：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

以下是一些向您显示如何使用 inputs.conf 监视性能监视对象的段落的示例。

```
# 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 attribute 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' attribute 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 中指定性能监视器对象的重要信息

指定 perfmon 关键字时全部使用小写形式

在 inputs.conf 中创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，例如：

正确	不正确
[perfmon://CPUTime]	[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 Enterprise 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一定字符数量的字符串，请勿使用 *，而应使用 .*。如果对象包含美元符号或类似特殊字符，您可能需要使用反斜线 (\) 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 属性的值（在 [perfmon://] 段落中）时，请确保这些值与定义于“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。如果输入与您指定的性能对象、计数器或实例值匹配失败，该输入将把此故障记录到 splunkd.log。例如：

```
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 Enterprise 配置为以具有远程收集性能指标的相应访问权限的 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 的计算机和 Splunk 从中收集性能数据的计算机必须驻留在相同 AD 域或林中。

注意：为了防止拒绝服务攻击，已将 WMI 设计为自行限制。Splunk Enterprise 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。请参阅本手册中的有关确定如何监视远程 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。这是由于 WMI 实现中的性能监视器计数器限制造成的，而与 Splunk Enterprise 或其如何检索基于 WMI 的数据无关。

WMI 使用 `Win32_PerfFormattedData_*` 类来收集性能指标。有关特定类的更多信息，请参阅 MSDN 上的 "Win32 Classes" (http://msdn.microsoft.com/en-us/library/aa394084%28v=vs.85%29.aspx)。

WMI 将这些类的数据结构定义为 32 位或 64 位无符号整数，具体取决于您所运行的 Windows 版本。同时，性能监视器对象将被定义为浮点变量。这意味着，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，最好在要从中收集性能数据的每个计算机上配置通用转发器的性能监视输入。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

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**远程性能监视**。

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。

注意：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

1.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。

2.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。

3.单击“查询”按钮以获取在“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。

注意：`Win32_PerfFormattedData_*` 类不会在 Splunk Web 中显示为可用对象。如果想要监视 `Win32_PerfFormattedData_*` 类，您必须直接把这些类添加在 `wmi.conf` 中。

4.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。

注意：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，请为每个对象创建其他数据导入。

5.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。

6.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。

7.要取消选择一个计数器，在“可用项目”窗口上单击该通道的名称。Splunk Enterprise 将计数器从“已选计数

器”窗口移动到“可用计数器”窗口。

8.要选择或取消选择所有计数器，单击“添加所有”或“删除所有”链接。**重要提示：**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）。

9.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。

注意："_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，在安装了两个磁盘的主机上，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- 且 "_Total" 实例是两个物理磁盘实例的平均值。

10.在**轮询间隔**字段中，为输入在轮询尝试之间输入时间（秒）。

11.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.设置**主机**名称值。此设置有多个选项供您选择。更多有关设置主机值的信息，请参阅关于主机。

> **注意：主机**只是设定生成事件中的**主机**字段。而不是引导 Splunk Enterprise 查找网络中的特定主机。

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“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.单击绿色**查看**按钮。

查看您的选择

在指定所有输入设置后，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，单击 **<** 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性能指标。

更多有关从远程计算机获取性能监视数据的信息，请参阅本手册中的监视 WMI 数据。

使用配置文件配置远程 Windows 性能监视

可通过 wmi.conf 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，请创建和/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。

请使用 Splunk Web 创建远程性能监视器输入，除非您对其无访问权限。这是因为性能监视器对象、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。

对于您想要监视的所有远程性能监视器对象，`wmi.conf` 中都包含一个对应的段落。在每个段落中，您可以指定以下内容。

全局设置

属性	是否必需？	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` 属性。	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

属性	是否必需？	描述	默认
interval	是	轮询新数据的频率（以秒为单位）。如果此属性不存在，则输入将不运行，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。该属性告知 Splunk Enterprise 传入数据采用事件日志格式。 请勿在一段落中使用 event_log_file 属性，前提为该段落已包含有 wql 属性。	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。此属性指示 Splunk Enterprise 从 WMI 提供程序获取数据。 请勿在一段落中使用 wql 属性，前提为该段落已包含有 event_log_file 属性。	N/A
namespace	否	要查询的 WMI 提供程序所驻留在的命名空间。该属性的值既可以是相对值 (Root\CIMV2) 也可以是绝对值 (\\SERVER\Root\CIMV2)，但如果您指定了 server 属性，则必须为相对值。 仅使用某段落中的 namespace 属性，该段落中包含 wql 属性。	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。 • 如已定义了 wql，该属性将指示 Splunk Enterprise 其是否应该预期事件通知查询。设置为 1 时指示 Splunk 需要事件通知查询；设置为 0 时需要标准查询。有关 WQL 和事件通知查询的其他要求，请参阅下文。 • 如已定义了 event_log_file，将指示 Splunk 是否仅捕获在 Splunk 运行期间发生的事件。设置为 1 时指示 Splunk 仅捕获在 Splunk 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 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

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。尤其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（`WITHIN`, `GROUP`, 和/或 `HAVING`）的子句之一。有关其他信息，请参阅以下 MSDN 文章："使用 WQL 查询"。

Splunk Web 可消除 WQL 语法问题，因为当您使用 Splunk Web 来创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

在收集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"线程"对象及其关联计数器）时，您可能会注意到 Splunk 中的内存使用量增加。这是正常现象，因为某些性能对象在收集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。这是故意将这些计数器设计为从 100% 中去除花在空闲进程上的时间量。

在非英语安装上，useEnglishOnly 属性有使用限制

在非英语系统上编辑 inputs.conf 以启用性能监视时，`useEnglishOnly` 属性的工作方式存在某些限制。

如果把属性设置为 `true`，您无法使用 `object` 和 `counters` 属性的通配符及正则表达式。这些属性必须基于如"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` 属性的通配符。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` 属性也会包含以英语呈现的值。

如果把属性设置为 `false`，您可以使用这些属性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` 属性已设置为 `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 属性的其他影响

使用该属性时要考虑其他项目。

- 当您使用 Splunk Web 在非英语操作系统上创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` 属性设置为 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。您可用此脚本语言创建脚本并将那些脚本结果输出为其他脚本的对象。

Splunk Enterprise 支持对通过 PowerShell 脚本接收到的事件的监控。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 Enterprise 为这些命令或脚本作为事件的输出创建索引。

如果您使用的是 Splunk Cloud，而且想监视脚本输出，请使用通用转发器来收集数据并把数据发送到您的 Splunk Cloud 部署。

您需要什么来使用 PowerShell 脚本监视数据？

活动	所需权限
使用 PowerShell 脚本监视数据	Splunk Enterprise 必须在 Windows 上运行。 Splunk Enterprise 必须以本地系统用户身份运行，这样便可运行所有 PowerShell 脚本。 必须在主机上安装 PowerShell 3.0 版本或更高版本。 必须在主机上安装 Microsoft .NET 4.5 版本或更高版本。

使用配置文件配置输入

要使用 `inputs.conf` 配置文件在 Splunk Enterprise 中配置某 PowerShell 输入：

1.编写 PowerShell 命令或脚本以捕获您想要的信息。

2.把 `inputs.conf` 从 `%SPLUNK_HOME%\etc\system\default` 复制到 `etc\system\local`。

3.打开此文件并进行编辑以启用 Windows PowerShell 输入。

4.重新启动 Splunk Enterprise。

PowerShell 输入配置值

Splunk 使用 `inputs.conf` 中的以下段落来监视 PowerShell 收集的数据。

属性	描述	默认
`script`	要执行的 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`	命令或脚本应执行的频率是多少。 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）

以下是关于如何配置输入的一些示例：

单个命令示例：

```
    [powershell://Processes-EX1]
    script = Get-Process | Select-Object Handles, NPM, PM, WS, VM, Id, ProcessName,
@{n="SplunkHost";e={$Env:SPLUNK_SERVER_NAME}}
    schedule = 0 */5 * ? * *
    sourcetype = Windows:Process
```

脚本示例：

```
[powershell://Processes-EX2]
script = . "$SplunkHome\etc\apps\My-App\bin\getprocesses.ps1"
schedule = 0 */5 * ? * *
sourcetype = Windows:Process
```

更多有关编写脚本的指南，请参阅为 PowerShell 输入编写脚本。

使用 Splunk Web 配置输入

要使用 Splunk Web 配置 PowerShell 输入：

1. 从系统栏选择**设置 > 数据导入**。

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。

通过单击“更多设置”复选框，您也可选择来源类型、主机和默认索引。

7. 单击**下一步**。“成功”页面会加载。

为 PowerShell 输入编写脚本

架构

Splunk Enterprise 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` 流和 XML 流输出提供支持架构、XML 配置的多线程脚本。

您可定义很多 PowerShell 段落并同时运行它们。您可通过 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）。

注意：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，该输入自动将所有输出转换为键/值对。

Splunk Enterprise 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 Enterprise 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` 变量，所以通过编写如下路径即可轻松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：

- `SplunkHome` - 您在其中安装 Splunk Enterprise 的目录（有助于附加 `/etc/apps/` 路径至）
- `SplunkServerName` - 为在事件中使用而为本计算机配置的名称
- `SplunkServerUri` - Splunk Enterprise REST API 地址
- `SplunkSessionKey` - 访问 Splunk Enterprise REST API 所需的会话密钥（验证令牌）
- `SplunkCheckpointPath` - 存储持续状态的路径
- `SplunkServerHost` - 您想要与其通信的 Splunk Enterprise 实例的名称
- `SplunkStanzaName` - 定义该脚本的 inputs.conf 段落的名称

处理输出

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，该输入不处理您的脚本的输出。这意味着输入不处理 `ScriptProperty` 值，也意味着您应避免长时间运行的脚本。您不应该编写等待事情发生的脚本，除非每次有输出您就退出。它也意味着您的所有输出实质上具有相同的时间戳，除非您根据下面的建议，使用 `SplunkTime` 变量覆盖该时间戳。

Splunk Enterprise 把您的脚本生成的每个对象均视为一个输出，并将其转换为用 <event> 和 标记括起来的事件。Splunk Enterprise 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，该输出必须简单，您应该在它们输出前展平脚本中的所有复杂嵌套对象。

有几个特殊属性名称，它们对 Splunk Enterprise 模块输入具有重要意义并允许您覆盖 inputs.conf 段落中的默认值。它们是：

- `SplunkIndex` - 覆盖其内将存储输入的索引
- `SplunkSource` - 覆盖输出的“来源”
- `SplunkHost` - 覆盖输出的“主机”名称
- `SplunkSourceType` - 覆盖输出的“来源类型”
- `SplunkTime` - 覆盖“事件”。如果您未指定此，则您的脚本在单个执行中生成的所有对象将大致获得相同时间戳。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。

监视 Windows 主机信息

Splunk Enterprise 支持监视本地 Windows 计算机的详细统计信息。它可以收集有关 Windows 主机的以下信息：

- **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的版本和内部版本号，以及所有服务包；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。
- **进程。**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。

Splunk Enterprise 的完整实例和通用转发器均支持本地收集主机信息。如果您使用的是 Splunk Cloud，而且想监视主机信息，请使用通用转发器来收集数据并把数据发送到您的 Splunk Cloud 部署。

主机监视器输入将作为一个进程运行，该进程称为 `splunk-winhostmon.exe`。此进程将按照在输入中指定的时间间隔，针对每个定义的输入运行一次。您可以使用 Splunk Web 或 `inputs.conf` 配置主机监视。

为什么监视主机信息？

Windows 主机监视使您获得关于您的 Windows 主机的详细信息。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 Enterprise 的搜索语言，您可以使您的团队能够大致了解您的 Windows 网络中所有计算机的统计信息。

监视主机信息需要什么？

活动	所需权限
监视主机信息	* Splunk Enterprise 必须在 Windows 上运行。 * 为读取所有的本地主机信息，Splunk 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 Enterprise 必须以本地系统用户身份运行来收集 Windows 主机信息。

Splunk 建议使用通用转发器将远程计算机中的主机信息发送到索引器。有关如何安装、配置及使用转发器收集 Windows 主机数据的信息，请查看《转发数据手册》。

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，则该用户身份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。该用户身份同时也必须拥有其他权限，详情请参阅*《安装》*手册中的“选择 Splunk Enterprise 应以其身份运行的 Windows 用户”。

使用 Splunk Web 配置主机监视

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**文件和目录**。

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，为该输入输入一个您可以记住的唯一名称。

3.在**事件类型**列表框中，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。**注意：**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）。

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

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.设置**主机**名称值。此设置有多个选项供您选择。更多有关设置主机值的信息，请参阅关于主机。

> **注意：主机**只是设定生成事件中的**主机**字段。而不是引导 Splunk Enterprise 查找网络中的特定主机。

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“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，单击 **<** 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。

使用 inputs.conf 配置主机监视

您可以通过编辑 `inputs.conf` 来配置主机监视。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"。

1. 创建一个 `inputs.conf`（路径为 `%SPLUNK_HOME%\etc\system\local`）并打开进行编辑。

2. 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。

3. 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。

4. 把您复制好的段落粘贴到 `%SPLUNK_HOME%\etc\system\local\inputs.conf` 中。

5. 对该段落进行编辑，以收集所需的 Windows 事件日志数据。

6. 保存 `%SPLUNK_HOME%\etc\system\local\inputs.conf` 并关闭。

7. 重新启动 Splunk Enterprise。

Windows 主机监视器配置值

Splunk Enterprise 使用 `inputs.conf` 中的下列属性监视 Windows 主机信息。

属性	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 Enterprise 将运行一次输入。如果您未定义此属性，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。如果此属性不存在，则输入将不运行。
`disabled`	否	是否运行输入。如果您把该属性设置为 `1`，Splunk Enterprise 不会运行该输入。

Windows 主机监视配置示例

以下是一些有关如何使用 `inputs.conf` 中的 Windows 主机监视配置属性的示例。

```
# 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 ever 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
```

用于 Windows 主机监视数据的字段

在为来自 Windows 主机监视输入的数据创建索引时，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`。

问答

有什么问题吗？请访问 Splunk Answers，查看 Splunk 社区有哪些与 Windows 主机信息相关的问题和答案。

监视 Windows 打印机信息

Splunk Enterprise 支持监视本地 Windows 主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。它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、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 Enterprise 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获取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` 配置打印机子系统监视。

为什么监视打印机信息？

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 Enterprise 的搜索语言，您可以使您的团队大致了解您的 Windows 网络中所有打印机的统计信息。

监视打印机信息需要什么？

活动	所需权限
监视主机信息	* Splunk Enterprise 必须在 Windows 上运行。 * 为了读取所有的本地主机信息，Splunk Enterprise 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 Enterprise 必须以本地系统用户身份运行来收集 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 Enterprise，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”所详述的其他权限。

使用 Splunk Web 配置打印机信息

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**本地 Windows 打印监视**。

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，为该输入输入一个您可以记住的唯一名称。

2.在**事件类型**列表框中，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。Splunk Enterprise 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 Enterprise 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”链接。**重要提示：**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**单选按钮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.单击绿色**下一步**按钮。

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.设置**主机**名称值。此设置有多个选项供您选择。更多有关设置主机值的信息，请参阅关于主机。

注意：主机只是设定生成事件中的**主机**字段。而不是引导 Splunk Enterprise 查找网络中的特定主机。

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“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，单击 **<** 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打印信息。

使用 inputs.conf 配置主机监视

您可以通过编辑 `inputs.conf` 来配置主机监视。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”。

1.把 inputs.conf 从 `%SPLUNK_HOME%\etc\system\default` 复制到 `etc\system\local`。

2.使用“资源管理器”或 `ATTRIB` 命令移除此文件的“只读”标记。

3.打开此文件，然后进行编辑以启用 Windows 打印监视输入。

4.重新启动 Splunk。

打印监视配置值

Splunk Enterprise 使用 inputs.conf 中的下列属性来监视 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。可以为 printer, job, driver 或 port 之一。如果此变量不存在，则输入将不运行。
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 Enterprise 会写入一个基线。Splunk Enterprise 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。如果您把该属性设置为 1，Splunk Enterprise 不会运行该输入。

Windows 主机监视配置示例

以下是一些有关如何使用 inputs.conf 中的 Windows 主机监视配置属性的示例。

```
# 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创建索引时，Splunk Enterprise 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 WinPrintMon。

问答

有什么问题吗？请访问 Splunk Answers，查看 Splunk 社区有哪些与 Windows 打印监视相关的问题和答案。

监视 Windows 网络信息

Splunk Enterprise 支持监视有关传入或来自 Windows 主机的网络活动的详细统计信息。它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 Enterprise 可对其进行监视。
- **地址系列。**网络事务是通过 IPv4 协议还是通过 IPv6 协议进行。
- **数据包类型。**事务中发送的数据包类型（例如 'connect' 或 'transport' 数据包）。
- **协议。**网络事务是通过 TCP 协议还是通过 UDP 协议进行。
- **主机。**有关网络事务所涉及主机的信息，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。**启动了网络事务的应用程序。
- **用户。**启动了网络事务的用户，包括其 ID 和 SID。
- **其他。**有关网络事务的其他信息，包括传输标头大小以及事务是否通过 IPSec 保护。

Splunk Enterprise 的完整实例和通用转发器均支持本地收集网络信息。如果您使用的是 Splunk Cloud，而且想监视网络信息，请使用通用转发器来收集数据并把数据发送到您的 Splunk Cloud 部署。

网络监视器输入将以进程的形式运行，该进程称为 splunk-netmon.exe。此进程将按照在输入中指定的时间间隔，针对每个定义的输入运行一次。您可以使用 Splunk Web 或 inputs.conf 配置网络监视。

Splunk Enterprise 中的 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。

为什么监视网络信息？

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 Enterprise 的搜索语言，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。

监视网络信息需要什么？

活动	要求

监视网络信息	• Splunk 必须在 Windows 上运行。
	• 计算机上的 Windows 版本必须为以下版本之一：
	  ◦ Windows Vista。
	  ◦ Windows 7。
	  ◦ Windows 8。
	  ◦ Windows 8.1。
	  ◦ Windows Server 2008。
	  ◦ Windows Server 2008 R2 或
	  ◦ Windows Server 2012 R2。
	• Windows 系统必须已经应用所有可用的更新和服务包，其中包括运行 Windows Vista、Windows 7、Windows Server 2008 和 Windows Server 2008 R2 的计算机上的 Kernel-Mode Driver Framework 版本 1.11 更新。
	• Splunk 必须以“本地系统”用户身份或使用本地管理员帐户运行以读取所有本地主机信息。

安全与远程访问注意事项

默认情况下，Splunk Enterprise 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 Enterprise，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 Splunk Web 配置主机监视

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**本地 Windows 网络监视**。

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。Splunk Web 将显示“添加数据 - 选择数据来源”页面。

注意：转发网络信息需要其他设置。

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，为该输入输入一个您可以记住的唯一名称。

2.在**地址系列**下方，选中您希望 Splunk Enterprise 监视的 IP 地址系列类型（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（**连接**、**接受**或**传输**中的任何一项）。

4.在**方向**下方，选中您希望输入监视的网络方向（**入站**（朝向监视主机）或**出站**（远离监视主机）之一）。

5.在**协议**下方，选中您希望输入监视的网络协议类型（**tcp**（传输控制协议）或 **udp**（用户数据报协议）之一）。

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。

注意：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。

注意：和远程地址一样，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。

注意：与远程地址和进程实体一样，您可以通过在此字段中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

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.设置**主机**名称值。此设置有多个选项供您选择。更多有关设置主机值的信息，请参阅关于主机。

注意：主机只是设定生成事件中的**主机**字段。而不是引导 Splunk Enterprise 查找网络中的特定主机。

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“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，单击 **<** 即可返回到向导中的上一个步骤。否则，单击绿色**提交**按钮。

然后，Splunk Enterprise 会加载“成功”页面并开始索引指定的打印信息。

使用 inputs.conf 配置网络监视

您可以通过编辑 `inputs.conf` 来配置网络监视。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”。

1.把 inputs.conf 从 `%SPLUNK_HOME%\etc\system\default` 复制到 `etc\system\local`。

2.使用“资源管理器”或 `ATTRIB` 命令移除此文件的“只读”标记。

3.打开此文件，然后进行编辑以启用 Windows 网络监视输入。

4.重新启动 Splunk。

下一部分介绍用于主机监视的特定配置值。

Windows 主机监视器配置值

要定义某 Windows 网络监视输入，请使用 `[WinNetMon://<name>]` 段落（位于 `inputs.conf` 中）。Splunk Enterprise 使用以下属性来配置 Windows 网络监视器输入。

属性	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>	此输入应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。过滤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。过滤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。过滤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 "inbound;outbound"。	（空字符串 - 包括两个方向。）
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”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。 读取网络监视器过滤器驱动程序的频率（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。 它应在网络监视器过滤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 Enterprise 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	1000

	大合法值是 5000。	

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创建索引时，Splunk Enterprise 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`。

确认完全修补您的 Windows 计算机

如果您在 Windows Vista、Windows 7、Windows Server 2008 或 Windows Server 2008 R2 计算机上运行网络监视输入时遇到问题，则请确认您已使用所有可用修补程序（包括作为“知识库”文章 2685811 一部分的 Kernel-Mode Driver Framework 版本 1.11 更新 (http://support.microsoft.com/kb/2685811)）来更新该计算机。如果此更新未存在于您的系统中，则网络监视输入可能不可用。

问答

有什么问题吗？请访问 Splunk Answers，查看 Splunk 社区有哪些与 Windows 网络监视相关的问题和答案。

获取其他类型数据的位置

监视先进先出 (FIFO) 队列

本主题介绍如何通过编辑 `inputs.conf` 文件配置“先进先出 (FIFO)”输入，该文件位于您安装 Splunk 通用转发器的 Splunk Enterprise 主机或计算机上。（Splunk Web 当前不支持 FIFO 输入的定义。）如果您使用的是 Splunk Cloud，请使用通用转发器读取 FIFO 队列。

注意：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。

将 FIFO 输入添加到 inputs.conf 中

要添加 FIFO 输入，请向位于 `$SPLUNK_HOME/etc/system/local/` 或您自己的自定义应用程序目录（路径为 `$SPLUNK_HOME/etc/apps/`）中的 inputs.conf 添加一个段落。

如果您之前未使用过配置文件，请在开始之前先阅读*《管理员》*手册中的“关于配置文件”。

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>]
<attribute1> = <val1>
<attribute2> = <val2>
...
```

可以在 FIFO 段落中使用以下属性：

属性	描述	默认
`host = <string>`	将此段落的主机键/字段设为一个静态值。在 `<string>` 前面加上 'host::'。 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` 或您已设置为默认索引的任何内容。
`sourcetype = <string>`	来自此输入的事件的 sourcetype 键/字段。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创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。分析和建立索引期间会使用该键来设置来源类型字段。它还是在搜索时使用的来源类型字段。 • 在 `<string>` 前面加上 'sourcetype::'。 • 更多有关来源类型的信息，请参阅本手册中的来源类型为何重要。	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/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]	输入处理器应该用来存储其所读取事件的位置。 设置为 "parsingQueue" 时，会把 props.conf 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 6.x 版本的 Splunk 软件继续保留此功能，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 Enterprise **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。此监视器是完全可配置的，系统上的任何文件被编辑、删除或添加时，都可以检测（而不仅仅是检测特定于 Splunk 的文件）。

例如，您可以指示文件系统更改监视器去监视 /etc/sysconfig/，并在系统配置发生更改时立即向您发出告警。

要监视 Windows 上的文件系统更改，请参阅本手册中的监视文件系统更改，了解如何使用 Microsoft 自带的审计工具。

文件系统更改监视器如何工作

文件系统更改监视器利用以下内容检测更改：

- 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 SHA256 哈希（可选）

可以配置文件系统更改监视器的以下功能：

- 使用正则表达式的白名单
  - 指定要检查的文件（无论是什么文件）
- 使用正则表达式的黑名单
  - 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
  - 创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创建索引
  - 用于发送整个文件和/或哈希的大小截止点
- 由 Splunk Enterprise 创建索引并且可搜索的所有更改事件

根据默认设置，只要 $SPLUNK_HOME/etc/ 的内容有所更改、删除或添加，文件系统更改监视器就会生成**审计事件**。Splunk Enterprise 首次启动时，会为 $SPLUNK_HOME/etc/ 目录和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。如您已配置好 signedaudit=true，Splunk Enterprise 会为文件系统更改创建**审计索引** (index=_audit)。如果 signedaudit 未启用，根据默认设置，Splunk Enterprise 会把事件写入**主**索引，除非您指定了另一个索引。

文件系统更改监视器不会跟踪执行更改的帐户用户名，仅跟踪发生的更改。对于用户级监视，请考虑使用自带的操作系统审计工具，这种工具可以访问上述信息。

警告：不要将文件系统更改监视器配置为监视您的根文件系统。如果启用了目录递归，则这样做非常危险并且非常

耗时。

配置文件系统更改监视器

在 inputs.conf 中配置文件系统更改监视器。不支持在 Splunk Web 中配置文件系统更改监视器。只要您更改了 `[fschange]` 段落就必须重启 Splunk Enterprise。

1. 打开 `inputs.conf`。

2. 添加 `[fschange:<directory>]` 段落以指定 Splunk Enterprise 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 Enterprise。文件系统更改监视立即开始。

如果您想要与**转发**结合使用该功能，请遵循这些指导原则：

- 要将事件发送给远程索引器，请使用**重型转发器**。
- 如果您无法使用重型转发器，请遵照使用通用转发器中的配置说明。

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`（路径为 `$SPLUNK_HOME/etc/system/local/`）或您自己位于 `$SPLUNK_HOME/etc/apps/` 的自定义应用程序目录中。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”。

语法

以下是 `[fschange]` 段落的语法：

```
[fschange:<directory or file to monitor>]
<attribute1> = <val1>
<attribute2> = <val2>
...
```

请注意以下事项：

- Splunk Enterprise 监视目录及其子目录中发生的所有添加/更新/删除。
- 任何更改都会生成一个被 Splunk 索引的事件。
- `<directory or file to monitor>` 默认为 `$SPLUNK_HOME/etc/`。

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
`index=<indexname>`	存储所有生成事件的索引。	`main`（除非您已经启用了审计事件签名）。
`recurse=<true | false>`	是否递归 <code[fschange]</code> 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 **警告：**如果您设置 `followLinks` 时不够谨慎，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 如果您执行更改，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	-1（不使用哈希进行更改检测）。

	法。	
signedaudit=<true | false>	发送经过加密签名的添加/更新/删除事件。 设置为 true 以在 _audit 索引中生成事件。如您正在设置 index 属性，请设置为 false。**注意：**把 signedaudit 设置为 true 时，确保已在 audit.conf 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（若 signedaudit=true）或 fs_notification（若 signedaudit=false）
filesPerDelay = <integer>	插入 delayInMills 指定的延迟时间，之前须先处理 <integer> 文件。这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）。这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过滤器。有关定义过滤器的信息，请参阅下一部分。	n/a

定义过滤器

如需使用 filters 属性定义要使用的过滤器，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 Enterprise 如何处理 fschange “白名单”和“黑名单”逻辑：

- 各个事件将遍历过滤器列表，直到其找到第一个匹配。
- 如果与事件匹配的第一个过滤器是白名单，则 Splunk Enterprise 将为该事件创建索引。
- 如果与事件匹配的第一个过滤器是黑名单，则过滤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 Enterprise 将为该事件创建索引。这表示有一个内置的“全部通过”隐式过滤器。

如果希望默认情况下事件未显式匹配白名单时 Splunk Enterprise 不为其创建索引，请在链的末尾添加将与所有剩余事件匹配的黑名单。

例如：

```
...
filters = <filter1>, <filter2>, ... terminal-blacklist

[filter:blacklist:terminal-blacklist]
regex1 = .?
```

如果您将某个目录永远列入黑名单（包括在一系列白名单之后添加末尾黑名单这种方式），则 Splunk Enterprise 将其所有子文件夹和文件都列入黑名单，因为它们不会通过任何白名单。要实现此目的，请在过滤器中的黑名单项目之前将所有所需文件夹和子文件夹显式加入白名单。

显式白名单和末尾黑名单实例

该配置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。

在本例中，目录可能已被加入黑名单。如果出现这种情况，则 Splunk Enterprise 将其**所有**子文件夹和文件也加入黑名单。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-blacklist]
regex1 = .?

[fschange:/var/apache]
index = sample
recurse = true
followLinks = false
signedaudit = false
fullEvent = true
sendEventMaxSize = 1048576
delayInMills = 1000
filters = configs,terminal-blacklist
```

与通用转发器结合使用

要从通用转发器转发文件系统更改监视器事件，您必须设置 `signedaudit = false` 和 `index=_audit`。

```
[fschange:<directory or file to monitor>]
signedaudit = false
index=_audit
```

Splunk Enterprise 将利用该解决方法为文件系统更改监视器事件创建 `_audit` 索引，其中 `sourcetype` 设置为 `fs_notification`，`source` 则设置为 `fschangemonitor`，而并非默认值 `audittrail` （`sourcetype` 和 `source` 均如此）。

通过脚本式输入从 API 及其他远程数据接口获取数据

Splunk Enterprise 可接受来自您提供的脚本的事件。脚本式输入可与诸如 `ipconfig`、`iostat`、`netstat`、`top` 等 indows 和 *nix 命令行工具结合使用。您可以使用脚本式输入从应用程序接口 (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。您可以使用中介 Windows 批处理 (`.bat`) 或 PowerShell (`.ps1`) 文件在 Windows 平台上启用基于文本的脚本，如 Perl 和 Python 脚本。

本主题介绍如何添加您已编写好的脚本式输入。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”。

您可以在“设置”菜单中或通过编辑 inputs.conf 来配置脚本式输入。

脚本式输入启动某个脚本时，该脚本将继承 Splunk Enterprise 环境。请清除任何可能影响脚本操作的环境变量。唯一可能引发问题的环境变量是库路径（在 Linux、Solaris 和 FreeBSD 上通常称为 `LD_LIBRARY_PATH`）。

Splunk Enterprise 会把输入式脚本发送到 `stderr` I/O 通道的所有消息记录到 `splunkd.log`。

在 Splunk Web 中添加脚本式输入

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**脚本**。
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**以监视本地计算机上的脚本或**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 - 选择数据来源”页面。
3. 在左窗格中，查找并选择**脚本**。

注意：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一个新的“脚本名称”下拉列表。
2. 在**脚本名称**下拉列表中，输入要运行的脚本。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 在调用该脚本前应等待的时间量（单位为秒）。
5. （可选）在**来源名称覆盖**字段中，如果需要，输入新数据源名称以覆盖默认数据来源值。
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。有关设置主机值的更多信息，请参阅“关于主机”。

在此页面上设置**主机**只能设置结果事件中的**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。您可选择**选择**以从本地计算机上的可用来源类型列表中选取，或“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**名称值。此设置有多个选项供您选择。
4. 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“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。
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，单击 **<** 即可返回到向导中的上一个步骤。否则，请单击**提交**。

Splunk Web 会显示“成功”页面并开始为指定的 Active Directory 节点建立索引。

使用 inputs.conf 添加脚本式输入

在 `inputs.conf` 中添加输入式脚本，操作方法为：添加一个 `[script]` 段落。

语法

以下为 `[script]` 段落的语法：

```
[script://$SCRIPT]
<attrbute1> = <val1>
<attrbute2> = <val2>
...
```

- `$SCRIPT` 是脚本位置的完整路径。
- `$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 [Getdatafromscriptedinputs#Use_the_.path_suffix_to_reference_external_scripts|Use the .path suffix to reference external scripts]。您在段落中引用的文件必须遵守本主题中“脚本式输入的脚本放在哪里”内介绍的位置限制。

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

最佳方法是把您的脚本放入 bin/ 目录，该目录离在主机文件系统上呼叫您的脚本的 inputs.conf 最近。例如，如果您要配置 $SPLUNK_HOME/etc/system/local/inputs.conf，请把您的脚本放入 $SPLUNK_HOME/etc/system/bin/。如要使用 $SPLUNK_HOME/etc/apps/$APPLICATION/ 中的应用程序，请把您的脚本放入 $SPLUNK_HOME/etc/apps/$APPLICATION/bin/。

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>	执行指定命令的频率。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执行，而会在 cron 计划所定义的时间执行。 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。如果您把脚本配置为每 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。 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::。 有关索引字段的更多信息，请参阅《管理索引器和群集》手册中的“索引如何工作”。	main 或您已设置为默认索引的任何内容。
sourcetype = <string>	设置来自此输入的事件的 Sourcetype 键/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创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 将此键用于分析/创建索引，特别是，可在创建索引期间设置来源类型字段。它也在搜索时使用来源类型字段。	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键/字段。 • **注意：**除非绝对必要，否则不要覆盖来源键。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。这会导致对先前配置的输入进行的更新无法保存。

Splunk Enterprise 会在验证文本时对不应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呼叫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。

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。

   ```
   [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。

   ```
   /path/to/myscript -arg1 arg -arg2 arg
   ```

含 Inputs.conf 的脚本式输入的示例

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. 所有脚本都应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. 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。

   ```
   $SPLUNK_HOME/etc/apps/scripts/bin/top.sh
   ```
   该脚本应发送一个 `top` 输出。
6. 把脚本条目添加至 `inputs.conf`，路径为 `$SPLUNK_HOME/etc/apps/scripts/local/`。

   ```
   [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
   ```

   注意：您可能需要修改 props.conf：

   - 根据默认设置，Splunk Enterprise 会把单个 `top` 条目拆分成多个事件。
   - 解决此问题最简单的方法是告诉此服务器只在输出中不存在的内容前面断开。

   例如，把以下内容添加至 `$SPLUNK_HOME/etc/apps/scripts/default/props.conf` 会强制所有行变成单个事件：

   ```
   [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。使用 `props.conf` 并设置以下内容：

```
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. 把 `loglogs.path` 放入 `$SPLUNK_HOME/etc/system/bin`，或在该路径下创建一个。
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

在上述示例中，您还可以把 `interval` 属性设置为 "cron" 计划，只需指定如下的字符串即可：

`0 * * * *`：表示在每小时开始时运行，一小时运行一次。

`*/15 9-17 * * 1-5`：表示周一到周五，从上午 9 点到下午 5 点每 15 分钟运行一次。

`15,35,55 0-6,20-23 1 */2 *`：表示在每个偶数月份（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。

更多有关设置 cron 计划的信息，请参阅 Crontab 网站上的 CRONTAB(5)。

通过抓取找到要监控的更多数据来源

已弃用此功能。
Splunk Enterprise 6.0 版本中已弃用此功能。这意味着，尽管这将继续运行，但可能在未来版本删除。作为替代方法，您可以搜索文件和目录以手动监视。 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。

使用 `crawl` 搜索命令来搜索您的文件系统或网络，以寻找新的数据源并将其添加到您的 Splunk Enterprise 索引。

请通过编辑 crawl.conf 来更改默认的抓取器设置。您可以在运行 `crawl` 时覆盖抓取器默认值。

`crawl` 会生成抓取活动日志，并将其存储在 `$SPLUNK_HOME/var/log/splunk/crawl.log` 中。

更改抓取器默认值

编辑 `$SPLUNK_HOME/etc/system/local/crawl.conf` 即可更改默认的抓取器配置设置。您分别定义文件和网络抓取器，单独在其各自的段落中进行。

语法

`crawl.conf` 包含两个段落：`[files]` 和 `[network]`，分别定义文件和网络抓取器的默认值。

有关这两个段落的可定义属性及其默认值的信息，请阅读 crawl.conf 规范文件。

示例

此为 `crawl.conf` 文件的一个示例，包含为文件和网络抓取器定义的设置：

```
[files]
bad_directories_list= bin, sbin, boot, mnt, proc, tmp, temp, home, mail, .thumbnails, cache, old
bad_extensions_list= mp3, mpg, jpeg, jpg,  m4, mcp, mid
bad_file_matches_list= *example*, *makefile, core.*
```

```
packed_extensions_list= gz, tgz, tar, zip
collapse_threshold= 10
days_sizek_pairs_list= 3-0,7-1000, 30-10000
big_dir_filecount= 100
index=main
max_badfiles_per_dir=100

[network]
host = myserver
subnet = 24
```

配置事件处理

事件处理概述

事件是驻留在日志文件中的活动及计算机数据的记录。事件是 Splunk 软件建立索引的主要内容。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 索引的内容。

下面是一个示例事件：

```
172.26.34.223 - - [01/Jul/2005:12:05:27 -0700] "GET /trade/app?action=logout HTTP/1.1" 200 2953
```

在为事件建立索引时，Splunk 软件会：

- 配置字符集编码。
- 配置多行事件的换行。
- 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）。
- 提取一组有用的标准字段，如 `host`、`source` 和 `sourcetype`。
- 将事件分段。
- 将元数据动态分配给事件（若已指定）。
- 使数据匿名（若已指定）。

索引处理的概述请参阅《管理索引器和群集手册》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 软件会尝试把该来源中的数据转换为 UTF-8 编码，除非您设置 `CHARSET` 键（在 `props.conf` 内）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，包括如下主要类别：

- UTF-8
- UTF-16LE
- Latin-1
- BIG5
- SHIFT-JIS

有关详尽列表，请参阅本主题结尾处的“支持字符集的综合列表”。

以下为 Splunk 软件支持的主要字符集和它们对应的语言：

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

手动指定字符集

要手动指定一个应用到输入上的字符集，请设置 `CHARSET` 键（在 `props.conf` 中）：

```
[spec]
CHARSET=<string>
```

例如，若您的主机生成的数据为希腊文（在本示例中称为 "GreekSource"）且该主机使用 ISO-8859-7 编码，请为该主机设置 `CHARSET=ISO-8859-7`（在 `props.conf` 中）：

```
[host::GreekSource]
CHARSET=ISO-8859-7
```

注意：Splunk 软件仅分析具有 UTF-8 映射的字符编码。某些 EUC-JP 字符不具有已映射的 UTF-8 编码。

自动指定字符集

Splunk 软件可以使用其复杂的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 软件使其为特定输入检测适当的语言和字符集编码，请为该输入设置 `CHARSET=AUTO`（设置位置位于 `props.conf` 内）。例如，若想为主机 “my-foreign-docs” 自动检测字符集编码，请为该主机设置 `CHARSET=AUTO`（设置位置位于 `props.conf` 中）：

```
[host::my-foreign-docs]
CHARSET=AUTO
```

训练 Splunk 软件使其识别字符集

如果您想要使用 Splunk 软件无法识别的字符集编码，请向以下路径添加示例文件并重启 Splunk Enterprise，由此对其进行训练使其识别目标字符集：

```
$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，并自动在索引时间把这些来源转换成成 UTF-8 格式。

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，请向 Splunk 支持提交问题。

支持字符集的综合列表

之前所述的常见字符集是字符集属性可支持的一个小子集。Splunk 软件也支持字符集和别名的长列表，该列表与 *nix `iconv` 实用工具支持的列表一样。

注意：Splunk 软件在匹配字符集时忽略标点符号和大小写，因此，会将 "utf-8"、"UTF-8" 和 "utf8" 都视为相同。

以下是完整列表，括号中是别名：

- utf-8（也称为 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（也称为 UCS-2LE、UNICODELITTLE）
- utf-16be（也称为 ISO-10646-UCS-2、UCS-2、CSUNICODE、UCS-2BE、UNICODE-1-1、UNICODEBIG、CSUNICODE11、UTF-16）
- utf-32le（也称为 UCS-4LE）
- utf-32be（也称为 ISO-10646-UCS-4、UCS-4、CSUCS4、UCS-4BE、UTF-32）
- utf-7（也称为 UNICODE-1-1-UTF-7、CSUNICODE11UTF7）
- c99（也称为 java）
- utf-ebcdic
- latin-1（也称为 CP819、IBM819、ISO-8859-1、ISO-IR-100、ISO_8859-1:1987、L1、CSISOLATIN1）
- latin-2（也称为 ISO-8859-2、ISO-IR-101、ISO_8859-2:1987、L2、CSISOLATIN2）
- latin-3（也称为 ISO-8859-3、ISO-IR-109、ISO_8859-3:1988、L3、CSISOLATIN3）
- latin-4（也称为 ISO-8859-4、ISO-IR-110、ISO_8859-4:1988、L4、CSISOLATIN4）
- latin-5（也称为 ISO-8859-9、ISO-IR-148、ISO_8859-9:1989、L5、CSISOLATIN5）
- latin-6（也称为 ISO-8859-10、ISO-IR-157、ISO_8859-10:1992、L6、CSISOLATIN6）
- latin-7（也称为 ISO-8859-13、ISO-IR-179、L7）
- latin-8（也称为 ISO-8859-14、ISO-CELTIC、ISO-IR-199、ISO_8859-14:1998、L8）
- latin-9（也称为 ISO-8859-15、ISO-IR-203、ISO_8859-15:1998）
- latin-10（也称为 ISO-8859-16、ISO-IR-226、ISO_8859-16:2001、L10、LATIN10）
- ISO-8859-5（也称为 CYRILLIC、ISO-IR-144、ISO_8859-5:1988、CSISOLATINCYRILLIC）
- ISO-8859-6（也称为 ARABIC、ASMO-708、ECMA-114、ISO-IR-127、ISO_8859-6:1987、CSISOLATINARABIC、MACARABIC）
- ISO-8859-7（也称为 ECMA-118、ELOT_928、GREEK、GREEK8、ISO-IR-126、ISO_8859-7:1987、ISO_8859-7:2003、CSISOLATINGREEK）
- ISO-8859-8（也称为 HEBREW、ISO-8859-8、ISO-IR-138、ISO8859-8、ISO_8859-8:1988、CSISOLATINHEBREW）
- ISO-8859-11
- roman-8（也称为 HP-ROMAN8、R8、CSHPROMAN8）
- KOI8-R（也称为 CSKOI8R）
- KOI8-U
- KOI8-T
- GEORGIAN-ACADEMY
- GEORGIAN-PS
- ARMSCII-8
- MACINTOSH（也称为 MAC、MACROMAN、CSMACINTOSH）[注意：这些 MAC* 字符集适用于 MacOS 9；OS/X 使用 unicode]
- MACGREEK
- MACCYRILLIC
- MACUKRAINE
- MACCENTRALEUROPE
- MACTURKISH
- MACCROATIAN
- MACICELAND
- MACROMANIA
- MACHEBREW
- MACTHAI
- NEXTSTEP
- CP850（也称为 850、IBM850、CSPC850MULTILINGUAL）
- CP862（也称为 862、IBM862、CSPC862LATINHEBREW）
- CP866（也称为 866、IBM866、CSIBM866）
- CP874（也称为 WINDOWS-874）
- CP932
- CP936（也称为 MS936、WINDOWS-936）
- CP949（也称为 UHC）
- CP950
- CP1250（也称为 MS-EE、WINDOWS-1250）
- CP1251（也称为 MS-CYRL、WINDOWS-1251）
- CP1252（也称为 MS-ANSI、WINDOWS-1252）

- CP1253（也称为 MS-GREEK、WINDOWS-1253）
- CP1254（也称为 MS-TURK、WINDOWS-1254）
- CP1255（也称为 MS-HEBR、WINDOWS-1255）
- CP1256（也称为 MS-ARAB、WINDOWS-1256）
- CP1257（也称为 WINBALTRIM、WINDOWS-1257）
- CP1258（也称为 WINDOWS-1258）
- CP1361（也称为 JOHAB）
- BIG-5（也称为 BIG-FIVE、CN-BIG5、CSBIG5）
- BIG5-HKSCS（也称为 BIG5-HKSCS:2001）
- CN-GB（也称为 EUC-CN、EUCCN、GB2312、CSGB2312）
- EUC-JP（也称为 EXTENDED_UNIX_CODE_PACKED_FORMAT_FOR_JAPANESE、CSEUCPKDFMTJAPANESE）
- EUC-KR（也称为 CSEUCKR）
- EUC-TW（也称为 CSEUCTW）
- GB18030
- GBK
- GB_1988-80（也称为 ISO-IR-57、ISO646-CN、CSISO57GB1988、CN）
- HZ（也称为 HZ-GB-2312）
- GB_2312-80（也称为 CHINESE、ISO-IR-58、CSISO58GB231280）
- SHIFT-JIS（也称为 MS_KANJI、SJIS、CSSHIFTJIS）
- ISO-IR-87（也称为 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（也称为 ISO646-JP、JIS_C6220-1969-RO、JP、CSISO14JISC6220RO）
- JISX0201-1976（也称为 JIS_X0201、X0201、CSHALFWIDTHKATAKANA）
- ISO-IR-149（也称为 KOREAN、KSC_5601、KS_C_5601-1987、KS_C_5601-1989、CSKSC56011987）
- VISCII（也称为 VISCII1.1-1、CSVISCII）
- ISO-IR-166（也称为 TIS-620、TIS620-0、TIS620.2529-1、TIS620.2533-0、TIS620.2533-1）

配置事件换行

一些事件由多行组成。默认情况下，Splunk 软件能正确处理大多数的多行事件。如果有 Splunk 软件未正确处理的多行事件，您可以对软件进行配置，更改其换行行为。

Splunk 软件如何确定事件界限

Splunk 软件用两个步骤来确定事件界限：

1.换行，即使用 `LINE_BREAKER` 属性正则表达式值把传入的字节流拆分成多个单独的行。根据默认设置，`LINE_BREAKER` 是新行和回车的任意序列（即 `([\r\n]+)`）。

2.行合并，只有在您把 `SHOULD_LINEMERGE` 属性设置为 "true"（默认设置）的时候才会发生。此步骤使用所有其他行合并设置（如 `BREAK_ONLY_BEFORE`, `BREAK_ONLY_BEFORE_DATE`, `MUST_BREAK_AFTER`, 等）来合并之前拆分成多行的时间。

如果第二步未运行（因为您把 `SHOULD_LINEMERGE` 属性设置为 "false"），各事件只是 `LINE_BREAKER` 属性所决定的单个行。第一步相对有效，第二步则相对较慢。适当使用 `LINE_BREAKER` 正则表达式可以生成您想要在第一步取得的结果。如果您的大量数据均包含多行事件，这将有用。

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 软件通常能够识别事件界限。然而，如果事件界限识别运行不正常，您可以在 props.conf 中设置自定义规则。

若要配置多行事件，

1.首先请检查事件格式。确定事件中要设置为事件开头或结尾的模式。

2.然后再编辑 `$SPLUNK_HOME/etc/system/local/props.conf` 并设置必要的属性，以配置您的数据。

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

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属性的访问权限。

1.请在 `props.conf` 中指定代表您想要拆分或重组成事件的数据流的段落。

2.在此段落中使用 `LINE_BREAKER` 属性把数据流拆分成多行。

3.把 `SHOULD_LINEMERGE` 属性设置为 true。

4.设置您的行合并属性（`BREAK_ONLY_BEFORE` 等），即可指示 Splunk 软件把各行重组成事件。

如果您的数据符合默认的 LINE_BREAKER 设置（任意数量的新行和回车），则无需改变 LINE_BREAKER。如果不符合，只要设置 SHOULD_LINEMERGE=true 并使用行合并属性就能重组各行。

使用 LINE_BREAKER 功能将数据流直接拆分为实际事件

这可能会加快索引速度，但处理起来稍微困难一些。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 指定代表您想要直接将其拆分成事件的数据流的段落。

2. 配合使用 LINE_BREAKER 属性和 SHOULD_LINEMERGE=false。

换行常规属性

这些是影响换行的 props.conf 属性：

属性	描述	默认
TRUNCATE = <non-negative integer>	请更改默认的最大行长度（以字节为单位）。尽管此属性为字节量度，但当此属性以其他方式获得多字节字符的中间字符时，Splunk 仍会向下舍入行长度。 希望永不截断时设置为 0（然而，非常长的行通常表示垃圾数据）。	10,000 个字节
LINE_BREAKER = <regular expression>	正则表达式会在任何行合并发生前（如下文所述，如已由 SHOULD_LINEMERGE 属性指定）决定如何将原始文本流拆分为初始事件。 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）。 无论匹配到的位置在哪里，Splunk 软件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。 Splunk 软件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 Splunk 软件认为此文本位于两行之间。 使用 LINE_BREAKER 分隔多行事件时，处理速度将显著提升（与使用 SHOULD_LINEMERGE 把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 软件会把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) 字符进行分隔。
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 软件应用了 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 个字符
SHOULD_LINEMERGE = [true|false]	设置为 true 时，Splunk 软件会把多个输入行组合成单个事件，配置则以下节中介绍的属性为基础。	true

仅在 SHOULD_LINEMERGE 设置为 true 时应用的属性

设置 SHOULD_LINEMERGE=true（默认）时，请使用下列属性定义换行行为：

属性	描述	默认
BREAK_ONLY_BEFORE_DATE = [true|false]	设置为 true 时，Splunk 软件会在遇到含日期的新行时创建新的事件。	true **注意：**如把 DATETIME_CONFIG 设置为 CURRENT 或 NONE，该属性将毫无意义，因为在这些情况下，Splunk 软件不会识别时间戳。

BREAK_ONLY_BEFORE = <regular expression>	设置之后，Splunk 软件会在遇到与正则表达式匹配的新行时创建新的事件。	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 软件就会为下一个输入行创建新的事件。如果匹配到另一个规则，Splunk 软件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 软件就不会在匹配到 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 软件就不会拆分当前行之前的最后一个事件。	空字符串
MAX_EVENTS = <integer>	指定要添加到任一事件的输入行的最大数量。 • Splunk 软件会在读取到指定的行数后中断。	256 行

配置事件换行的示例

指定事件分隔符

```
[my_custom_sourcetype]
BREAK_ONLY_BEFORE = ^\d+\s*$
```

假设任何仅由数字组成的行是任何其来源类型设置为 my_custom_sourcetype 的数据的新事件的开始。

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"。对于本示例，假设所有这些行均需显示为单个事件条目。

```
{{"2006-09-21, 02:57:11.58", 122, 11, "Path=/LoginUser
Query=CrmId=ClientABC&ContentItemId=TotalAccess&SessionId=3A1785URH117BEA&Ticket=646A1DA4STF896EE&SessionTime=25368&R
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=&AgentHost=man&AgentId=m
MANUser:
Version=1&Name=&Debit=&Credit=&AccessTime=&BillDay=&Status=&Language=&Country=&Email=&EmailNotify=&Pin=&PinPayment=&P
""}}
```

若想为该多行事件正确地创建索引，请在您的配置中使用 Path 区分符。请把以下内容添加到您的 $SPLUNK_HOME/etc/system/local/props.conf：

```
[source::source-to-break]
SHOULD_LINEMERGE = True
BREAK_ONLY_BEFORE = Path=
```

此代码将指示 Splunk 软件合并该事件的各行，且仅在术语 Path= 之前中断。

多行事件换行和分段限制

Splunk 软件会对巨大事件应用换行和分段限制：

- **超过 10,000 个字节的行。**为超过 10,000 个字节的行建立索引时，Splunk 软件会将这些行拆分为多个 10,000 字节的行。它将把字段 meta::truncated 附加到每个截断部分的结尾。它仍将这些行归组到单个事件中。
- **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。

问答

有什么问题吗？请访问 Splunk Answers 以查看在 Splunk 社区中对于换行有哪些相关的问题和解答。

配置事件时间戳

本主题介绍 Splunk 软件如何处理时间戳。

检查以下示例事件：

```
172.26.34.223 - - [01/Jul/2005:12:05:27 -0700] "GET /trade/app?action=logout HTTP/1.1" 200 2953
```

事件中的时间信息：

```
[01/Jul/2005:12:05:27 -0700]
```

是一个**时间戳**。

Splunk 软件使用时间戳按时间关联事件、在 Splunk Web 中创建直方图，并为搜索设置时间范围。大部分事件都包含时间戳；在事件不包含时间戳信息的情况下，Splunk 软件会尝试在索引时间为事件分配时间戳值。

大多数情况下，Splunk 软件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，处理一些来源或分布式部署时，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信息，请参阅本册手中的"配置时间戳"一章。

配置索引字段提取

本主题介绍 Splunk 软件会在索引时间提取哪些字段，并带领您进入新的一章介绍如何配置提取。

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。

更多有关索引字段提取的信息，请参阅本手册中的配置索引字段提取一章。

使数据匿名

本主题介绍如何对您传入 Splunk 部署的数据（如信用卡号和社会保险号）进行匿名处理。

建立日志事件的索引时，您可能希望以掩码显示敏感个人数据。可能不希望出现在索引中的数据的两个示例是信用卡号和社会保险号。本主题描述如何以掩码显示机密字段部分以保护隐私，同时提供足够的剩余数据以跟踪事件。

您有两种方法可以对数据进行匿名处理：

- 通过正则表达式 (regex) 转换。
- 通过一个 `sed` 脚本。

如果您正在运行 Splunk Enterprise 而且想对数据进行匿名处理，请根据本主题中的介绍配置您的索引器或重型转发器。如果您正在向 Splunk Cloud 转发数据而且想对数据进行匿名处理，请使用重型转发器并根据本主题中的介绍对其进行配置。

使用正则表达式转换使数据匿名

您可以配置 `transforms.conf`，通过正则表达式以掩码显示数据。

本示例以掩码显示应用程序服务器日志中 `SessionId` 和 `Ticket number` 字段最后四个字符以外的所有字符。

下面是一个所需输出的示例：

```
SessionId=############7BEA&Ticket=############96EE
```

示例输入：

```
"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=&", ""
```

要以掩码显示数据，请修改 `props.conf` 和 `transforms.conf` 文件（位于您的 `$SPLUNK_HOME/etc/system/local/` 目录中）。

配置 props.conf

1.编辑 `$SPLUNK_HOME/etc/system/local/props.conf` 并添加以下段落：

```
[<spec>]
TRANSFORMS-anonymize = session-anonymizer, ticket-anonymizer
```

在本段落中，`<spec>` 必须是以下内容之一：

- `<sourcetype>`，事件的来源类型。
- `host::<host>` 其中 `<host>` 是事件的主机。
- `source::<source>` 其中 `<source>` 是事件的来源。

本示例中，`session-anonymizer` 和 `ticket-anonymizer` 为任意的 TRANSFORMS 类名称；您在对应的 `transforms.conf` 文件内的段落中定义了这些类名称的操作。使用您在 `transforms.conf` 中创建的类名称。

配置 transforms.conf

2.在 `$SPLUNK_HOME/etc/system/local/transforms.conf` 中添加您的 TRANSFORMS：

```
[session-anonymizer]
REGEX = (?m)^(.*)SessionId=\w+(\w{4}[&"].*)$
FORMAT = $1SessionId=########$2
DEST_KEY = _raw
[ticket-anonymizer]
REGEX = (?m)^(.*)Ticket=\w+(\w{4}&.*)$
FORMAT = $1Ticket=########$2
DEST_KEY = _raw
```

在本次转换中：

- `REGEX` 应指定正则表达式，其指向您想要匿名的事件中的字符串。`FORMAT` 指定掩码值。
- `$1` 全是正则表达式的前导文本，`$2` 全是正则表达式之后的事件文本。
- `DEST_KEY = _raw` 指定将 `FORMAT` 中的值写入到日志中的原始值，从而修改事件。

注意：该正则表达式处理器不处理多行事件。通过把 `(?m)` 放在 `transforms.conf` 中的正则表达式之前，就可以把该事件指定为多行事件，这是一种解决方法。

通过 `sed` 脚本匿名数据

您也可以使用 `sed` 脚本替换或替代事件中的字符串，来实现数据匿名。

大多数 UNIX 用户都熟悉 `sed`，这是一个 Unix 实用工具，可以根据命令列表的指定读取文件并修改输入。Splunk Enterprise 让您可以使用类 `sed` 语法（在 `props.conf` 中）使您的数据匿名。

在 props.conf 中定义 sed 脚本

1.编辑或创建 `props.conf` 的副本（在 `$SPLUNK_HOME/etc/system/local` 中）。

创建一个 `props.conf` 段落，其使用 `SEDCMD` 来指示 `sed script`：

```
[<spec>]
SEDCMD-<class> = <sed script>
```

在本段落中，`<spec>` 必须是以下内容之一：

- `<sourcetype>`，事件的来源类型。
- `host::<host>` 其中 `<host>` 是事件的主机。
- `source::<source>` 其中 `<source>` 是事件的来源。

`sed script` 在索引时间内仅适用于 `_raw` 字段。支持 `sed` 命令的以下子集：

-
  - 替换 (`s`)
  - 字符替代 (`y`)。

2. 更改 `props.conf` 后重启 Splunk 实例即可启用配置。

使用正则表达式匹配替换字符串

`sed` 替换的语法为：

```
SEDCMD-<class> = s/<regex>/<replacement>/flags
```

在本段落中：

- `regex` 是一个 PERL 正则表达式。
- `replacement` 是替换该正则表达式匹配的字符串。它使用 "\n" 代表后参考，其中 `n` 为个位数。
- `flags` 既可以是替换所有匹配的 "g"，也可以是替换特定匹配的数字。

示例

在下例中，您要为包含社会保险号和信用卡号的数据建立索引。在索引时间内，您希望以掩码显示这些值，以便仅后四位数显示在您的事件中。`props.conf` 段落可能类似如下所示：

```
[source::.../accounts.log]
SEDCMD-accounts = s/ssn=\d{5}(\d{4})/ssn=xxxxx\1/g s/cc=(\d{4}-){3}(\d{4})/cc=xxxx-xxxx-xxxx-\2/g
```

在您的帐户事件中，“社会保险”号显示为 `ssn=xxxxx6789`，信用卡号显示为 `cc=xxxx-xxxx-xxxx-1234`。

替代字符

`sed` 字符替代的语法为：

```
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。

示例

您有一个想为其创建索引的文件 `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"。请把以下内容添加到您的 `props.conf`：

```
[source::.../abc.log]
SEDCMD-abc = y/abc/ABC/
```

搜索 `source="*/abc.log"` 的时候，您在数据中应该找不到小写字母 "a"、"b" 和 "c"。Splunk Enterprise 已使用 "A"、"B" 和 "C" 分别替代每个 "a"、"b" 和 "c"。

匿名数据的注意事项

Splunk 索引器不会分析结构化数据

您向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`（使用 `INDEXED_EXTRACTIONS` 进行配置），索引器也不会分析该数据。转发数据跳过索引器上的下列队列，这就排除了索引器上对于该数据的任何解析：

- `parsing`
- `aggregation`
- `typing`

转发数据在到达索引器时必须已经过分析。要实现该目标，您还必须在发送数据的转发器上设置 `props.conf`。这包括 `INDEXED_EXTRACTIONS` 的配置，以及其他任何分析、过滤、匿名及路由规则。

通用转发器能够单独地为结构化数据执行这些任务。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

Splunk 软件使用**时间戳**执行下列操作：

- 按时间相关联**事件**。
- 在 Splunk Web 中创建时间线直方图。
- 为搜索设置时间范围。

在**索引时间**把时间戳添加到事件。它通常使用原始事件数据中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，它可能需要您帮忙告知如何识别时间戳。

时间戳值存储在 `_time` 字段中（采用 UTC 时间格式）。

时间戳处理是**事件处理**中的关键步骤之一。有关事件处理的详细信息，请参阅本手册中的"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. 它使用显式 `TIME_FORMAT`（如果有）在事件自身中查找时间或日期。配置 `TIME_FORMAT` 属性（在 `props.conf` 中）。

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 Enterprise 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. 如果事件没有时间或日期，Splunk 软件会使用同一个数据来源中最新的前一个事件的时间戳。

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。文件名中不标识当天的时间。（这要求事件具有时间，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。

注意：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。

配置时间戳

大多数事件不需要任何特殊时间戳处理。Splunk 软件会自动为事件识别和提取时间戳。但是，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 Sourcetype"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"设置 Sourcetype"页面。
- 直接编辑 props.conf。请参阅配置时间戳识别。

您还可以对时间戳提取进行配置，以便：

- 应用时区偏移。
- 从具有多个时间戳的事件获取正确的时间戳。
- 改进索引性能。

从新输入添加数据时的注意事项

如果您从新输入为某些数据建立索引，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重新建立索引。

或者，您也可以先在测试版 Splunk 部署（或 Splunk 生产实例上的另一个索引）中测试新的数据导入，然后再把数据添加到您的生产实例。按此方式，您可删除并重新建立索引，直到您获得您想要的结果。

配置时间戳识别

大多数事件不需要特殊时间戳处理。Splunk 软件可以正确识别和提取事件的时间戳。但是，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 Sourcetype"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。请参阅"设置 Sourcetype"页面。
- 请直接编辑 props.conf，如本主题中所介绍。

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。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。

时间戳处理器

默认情况下，时间戳处理器会驻留在 `$SPLUNK_HOME/etc/datetime.xml` 中。正常情况下，您无需改动此文件，除非您要

处理不寻常的自定义时间戳。如果您需要以某种方式配置时间戳识别，可以根据本主题下文中的描述，通过设置 props.conf 时间戳进行必要的更改。

如果您有自定义的时间戳且无法通过配置 props.conf 进行处理，请用 DATETIME_CONFIG 属性替代您自己的时间戳处理器。此属性会指定 Splunk 软件进行时间戳处理时使用的文件。

在 props.conf 中编辑时间戳属性

如需配置 Splunk 软件如何识别时间戳，请编辑 props.conf。有许多属性都与时间戳相关。特别是，您可以通过使用 TIME_FORMAT 属性来指定时间戳的 strptime() 格式，由此来决定 Splunk 软件识别时间戳的方式。

您也可以设置与时间戳相关的其他属性。包括指定在事件中的哪个位置查找时间戳、要使用的时区，或如何处理不同货币的时间戳。

编辑 props.conf 文件，该文件位于 $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。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"。

要设置时间戳识别，请在 props.conf 中配置一个或以上的时间戳属性。有关这些属性和其他属性的详细信息，请参阅 props.conf 规范文件。

语法概述

时间戳属性的语法概述遵循以下：

```
[<spec>]
DATETIME_CONFIG = <filename relative to $SPLUNK_HOME>
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>
```

在本语法中，<spec> 可以是：

- <sourcetype>，事件的来源类型。
- host::<host>，其中 <host> 为事件的主机值。
- source::<source>，其中 <source> 为事件的来源值。

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。

时间戳有效性属性和对事件的影响

默认情况下，所有事件都会建立索引，除非您通过其他方法特别过滤掉了一些事件。

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 属性，而且事件的时间戳落在这些属性的参数之外，那么，Splunk 软件会使用以下算法执行下列操作 ：

- 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。

事件不会因为落在这些属性的参数之外而遭丢弃。

时间戳属性

您可以使用 props.conf 来设置下列时间戳属性：

属性	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，您既不需要创建自己的时间戳处理器文件，也不需要修改默认的 datetime.xml 文件。本主题中介绍的其他 props.conf 属性通常可以调整时间戳识别功能，来满足您的需求。然而，如果您的数据采用自定义时间戳格式，则可能需要替换您自己的此文件版本。	$SPLUNK_HOME/etc/datetime.xml

	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 软件不会在事件的文本中查找时间戳；相反，它会使用事件的“接收时间”，即事件通过其输入获得接收的时间。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 可以在创建索引时把当前的系统时间分配给所有事件。 **注意：** CURRENT 和 NONE 都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 和/或 BREAK_ONLY_* , MUST_BREAK_* 设置来控制事件合并。	
TIME_PREFIX = <regular expression>	设置之后，Splunk 软件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。 应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，您应该把 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 软件应在事件中查找时间戳的深度（多少个字符）。 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，将会有效地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）。	150 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。 strptime() 为指定时间格式的 Unix 标准。更多信息请参阅下文中的增强的 strptime() 支持。 TIME_FORMAT 在 TIME_PREFIX 后（如无 TIME_PREFIX 属性则直接在事件开头）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，时间戳必须出现在每个事件的开头，否则 Splunk 软件将无法处理格式化指令，所有事件也会因为无法使用 strptime 而包含一个告警。（您仍有可能获得有效的时间戳，具体取决于 Splunk 软件尝试从问题中恢复的方式。） 为了获得最佳结果，<strptime-style format> 应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。	空字符串
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。	空字符串

	• 否则，如果 TZ 被设置为有效的时区字符串，则使用它。使用 zoneinfo TZ 数据库中的值指定时区设置。 • 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，请参阅本手册中的“指定时间戳的时区”。	
`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字母时区首字母缩写。 如果这些值的传统 Splunk 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 (key) 根据事件时区指示符的文本进行匹配，值 (value) 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。 示例：`TZ_ALIAS = EST=GMT+10:00`（请参阅“配置文件参考”中的 props.conf 示例文件，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 软件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 软件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 **注意：**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（从当前日期算起）。 例如，若 `MAX_DAYS_HENCE = 3`，软件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，软件将使用事件当前的索引时间。 **注意：**窗口较小时出现误报的可能性较小。更改此属性时请谨慎。 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。	2 天
`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 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来源内大多数时间戳相同时，Splunk 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 TIME_FORMAT 属性（位于 props.conf 内）来配置时间戳分析。此属性采用 strptime() 格式字符串来提取时间戳。

Splunk 软件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。其他格式包括：

%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。
%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 在 %s 上出现在 %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什么都没有）。
%z, %:z, %::z	ISO-8601 格式的数字时区偏移（例如，-0800 表示 PST 或 +0000 表示 GMT；或者，如果无法确定时区则什么都没有）。请使用 %:z，前提是时间戳偏移中包含小时和分钟（例如 -08:00）；也可以使用 %::z，前提是时间戳偏移中包含小时、分钟和秒钟（例如 -08:00:00）。
%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。（注意：可能无任何内容。）
%s	Epoch（10 位）

注意：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：

1998-12-31	%Y-%m-%d
98-12-31	%y-%m-%d
1998 years, 312 days	%Y 年， %j 天
Jan 24, 2003	%b %d, %Y
January 24, 2003	%B %d, %Y
1397477611.862	%s.%3N

注意：Splunk 软件目前还无法识别时间戳中的非英语月份名称。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，例如：

```
...FOR: 04/24/07 PAGE 01...
```

要提取此时间戳，请在 props.conf 中添加本段落：

```
[host::foo]
TIME_PREFIX = FOR:
TIME_FORMAT = %m/%d/%y
```

包括时区信息的其他示例：

```
...Valid_Until=Thu Dec 31 17:59:59 GMT-06:00 2020
```

要提取该时间戳，请将该内容添加至 `props.conf`：

```
[host::bar]
TIME_PREFIX = Valid_Until=
TIME_FORMAT = %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，例如：

```
...1989/12/31 16:00:00 Wed May 23 15:40:21 2007...
```

Splunk 软件将该日期提取为无用的 Dec 31, 1989。此例中，请配置 `props.conf` 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，尽量把您的 `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属性来配置时间戳提取处理器以满足某些专门用途，例如：

- 应用时区偏移。
- 从具有多个时间戳的事件获取正确的时间戳。
- 改进索引性能。

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“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 软件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` 执行此操作。以下是一个示例：

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
```

但您希望它与以下类似（以便与其在搜索结果中的外观相对应）：

```
07/06/2011 10:26:11 PM
```

本示例简要显示了如何使用 `props.conf` 和 `transforms.conf` 来转换原始事件中的时间戳。

在 `transforms.conf` 中添加本段落：

```
[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
```

问答

有什么问题吗？请访问 Splunk Answers，查看 Splunk 社区有哪些与时间戳识别和配置相关的问题和答案。

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，这尤其有用。

通过编辑 props.conf 配置位置时间戳提取。有关编辑时间戳 `props.conf` 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。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。

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，请把 `TIME_PREFIX` 和 `MAX_TIMESTAMP_LOOKAHEAD` 属性添加到 `props.conf` 段落。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。设置 `MAX_TIMESTAMP_LOOKAHEAD` 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）。通过约束提前量 (lookahead)，您可以提高准确度和性能。

设置好 `TIME_PREFIX` 后，Splunk 软件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。该时间戳算法仅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。因此，如果把 `TIME_PREFIX` 设置为 `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` 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

假设您有类似如下所示的一个事件：

```
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
```

要将时间信息 `May 23 15:40:21 2007` 的第二个字符串识别为时间戳，请按如下所示配置 `props.conf`：

```
[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 软件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` 属性中获得的数字）出现另一个时间戳，则 Splunk 软件会*忽略*第一个时间戳。Splunk 软件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。

注意：设置 `MAX_TIMESTAMP_LOOKAHEAD` 的值，仅根据您的时间戳提取需求查找事件，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。

指定时间戳的时区

如果您从不同的时区为数据创建索引，可以利用时区偏移来确保各时区在搜索时关联正确。可以根据事件的主机、来源或来源类型配置时区。

在 props.conf 中配置时区。有关编辑时间戳 `props.conf` 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。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。

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。

注意：如果您使用的是 Splunk Enterprise 而且您更改了主机计算机的时区设置，您必须重启 Splunk Enterprise 才能检测到所做的更改。

在 props.conf 中指定时区

要配置时区设置，请编辑在 `$SPLUNK_HOME/etc/system/local/` 内或 `$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”。

通过把 `TZ` 属性添加到 `props.conf` 中的适当段落来配置时间区。`TZ` 属性将识别 zoneinfo TZ ID。（请参阅 zoneinfo (TZ) 数据库中的所有时区 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。这应是来自此主机、来源或 Sourcetype 的事件所处的时区。

您没有在 Splunk Enterprise 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）进入索引器。要正确处理这两组事件的时间戳，该索引器的 `props.conf` 需要将时区分别指定为美国/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
```

第二个示例则把任何来自 `/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

Zoneinfo 数据库是时区值的公共维护数据库。

- 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以下目录中：`/usr/share/zoneinfo`。
- Splunk 软件的 Solaris 版本将 TZ 信息存储在以下目录中：`/usr/share/lib/zoneinfo`。
- Splunk 软件的 Windows 版本随附一个 TZ 数据库的副本。

所有获准可用的 `TZ` 值请参阅 zoneinfo (TZ) 数据库。

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，请按如下所示设置属性：

```
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偏移值。也可以只将一个 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，如果映射 EST 的同时还映射 EDT - 具体视您的区域设置对而定。测试您的软件以查看其产生什么时区字符串。

可以指定多个映射。`TZ_ALIAS` 的语法为：

```
TZ_ALIAS = <key=value>[,<key=value>]...
```

有关更多信息（包括示例），请参阅《配置文件参考》中的 props.conf 规范和示例文件。

为用户的搜索结果设置时区

使用 Splunk 验证添加或编辑用户时，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 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

如需加快索引的建立速度，您可以使用 props.conf 调整 Splunk 时间戳处理器在事件中的查找深度，或者甚至可以完全关闭时间戳处理器。

有关编辑时间戳 `props.conf` 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。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。

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深度（多少个字符）。设置 `MAX_TIMESTAMP_LOOKAHEAD` 属性可调整时间戳处理器的查找深度。

时间戳处理器在事件中查找的默认字符数为 150。您可以将 `MAX_TIMESTAMP_LOOKAHEAD` 设置为更低值以加速索引。如果时间戳始终出现在事件的第一部分，则应特别执行此操作。

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。通过把 `DATETIME_CONFIG` 属性设置为 `NONE`，可以关闭匹配到指定主机、来源或 sourcetype 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"，即事件通过其输入获得接收的时间。对于基于文件的输入（如监视器）而言，这意味着时间戳来自输入文件的修改时间。

您也可以通过把 `DATETIME_CONFIG` 设置为 `CURRENT` 来提升索引性能，这样做可以在建立索引的时候把当前的系统时间分配给所有事件。

示例：

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
...
```

注意： `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，它将许多**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：

- **索引字段提取**，这在本主题的开头已简要介绍，是构成本章的基础。这些字段存储在索引中，并成为事件数据的一部分。
- **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创建那些字段，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"。

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**。请参阅本章中的关于默认字段。
- **自定义字段**，由您指定。请参阅本手册中的在索引时间创建自定义字段。

注意：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。如果是这类数据，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，以便在索引时间读取这些文件类型（如逗号分隔值文件 (CSV)、制表符分隔值文件 (TSV)、管道符分隔值文件和 JavaScript 对象符号 (JSON) 数据来源）和地图字段。要了解 Splunk Enterprise 如何运作，请参阅本手册中的从具有标头的文件中提取数据。

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。

为了正确创建事件，Splunk 软件会在为事件建立索引时使用某些字段的值，尤其是 `sourcetype`。创建数据索引之后，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创建自定义字段。

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。更多有关主机的信息，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。更多有关来源类型的信息，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。更多有关文件输入的信息，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。

在索引时间创建自定义字段

通常，您应试图在索引时间内提取字段。但是，您有时可能会需要添加自定义索引字段。例如，可能有这样一种情形：某些搜索时间字段提取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` 字段几乎*始终*具有 `bar` 值，也会发生这种情况。

相反地，如果搜索时间提取字段的值时常位于字段之外，则您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，您可能想要把 `foo` 添加到 Splunk 在索引时间提取的字段列表。

更多相关信息请参阅《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。

警告：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新创建索引。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"。

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。

注意：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

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。前导下划线预留用于 Splunk 的内部变量。
- 不允许使用国际字符。

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` 是一个正则表达式，用于处理您的数据以提取字段。

    - 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。用于指定您目前正在提取的字段/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 行为会因为提取发生在搜索时间或索引时间而不同。
    - 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，您还可以设置创建连接字段的索引时间字段提取：

      ```
      FORMAT = ipaddress::$1.$2.$3.$4
      ```

    使用 FORMAT 创建连接字段时，$ 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 把提取到的字段名称和值写入 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 及其在索引字段创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 或未设置。该字段指定 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。
    - 更多有关 _meta 及其在索引字段创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。
    - 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 亦即 _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。
    - 一旦上一个匹配停止，REPEAT_MATCH 就会开始并继续，直至找不到任何匹配。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内的字段名称必须遵循字段名称语法限制。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，用于辨识您正在提取的字段（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 中您的段落的名称。

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

针对新的索引字段，把一个条目添加到 fields.conf：

```
[<your_custom_field_name>]
INDEXED=true
```

请注意以下事项：

- <your_custom_field_name> 是您在唯一字段（您已将其添加至 transforms.conf）中设置的自定义字段的名称。
- 设置 INDEXED=true 即说明已为该字段创建索引。

注意：若在搜索时间提取了某一同名字段，您**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，您*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 中的所有匹配转换进行修改，这些匹配转换包含 DEST_KEY = _meta 或 WRITE_META = true。
- 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。

注意：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 创建字段名称时

请牢记：当 Splunk 创建字段名称时，它会应用字段名称语法限制。

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

此基础示例将创建一个名为 `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` 将指示软件把 `FORMAT` 的内容写入索引。

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

本示例创建了两个索引字段，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，以创建单一字段。

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，您必须创建新的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

- 如需了解如何在导入结构化数据文件时设置来源类型，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。

要使用 Splunk Web 从结构化数据文件中提取字段：

1. 请选择您想要添加数据的方式。
2. 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. 审阅显示于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. 如果您不想将设置保存为新的来源类型，请转到步骤 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. 请转到步骤 4a 以进入“修改输入设置”页面。

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
	指定文件类型以及 Splunk Enterprise 应对文件使用的提取和/或	

INDEXED_EXTRACTIONS = {CSV|W3C|TSV|PSV|JSON}	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_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

编辑配置文件以创建和引用来源类型

要创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所述的启用基于标头字段的自动提取。如果 `props.conf` 文件不存在，您必须创建一个。
2. 定义新 Sourcetype，方法是创建一个告诉 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，请在相同的目录中创建一个。
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。

转发从结构化数据文件中提取的数据

您也可以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，请遵循本主题稍早所述的使用 Splunk Web 从结构化数据文件中提取字段一文中的说明。
    - 要使用配置文件，请遵循本主题稍早所述的编辑配置文件以创建和引用 sourcetype 一文中的说明。

4. （可选）如您需要在为数据创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 软件不会解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、过滤、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。

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。

注意：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,vqmburstcount,vqmbu
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,,,,,6
60,975,488,179,192,999.3,0,0,4.07,,4.12,,4.2,,4.03,,0.02,63,76,76,,,,43,0,6.8,0,520,10054,87,87,89,93,9,79,12,12,12,6
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,BW16350227029061159456629 9@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,,,,,6
62,993,496.5,126,139,3404.7,0,0,4.04,,4.07,,4.2,,3.94,,0.36,58,64,69,,,,49,0,286,0,529,17839,86,86,87,93,9,137,8,8,8,
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。

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。由于这是一个**默认字段**，意味着 Splunk 软件会给它为其建立索引的所有事件分配一个主机，因此，您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主机的主机名、IP 地址或完全限定的域名。

如何确定主机值

Splunk 软件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` 中指定的事件特定的任何主机分配。

2. （如有）创建了事件的输入的默认主机值。

3. 最初获取数据的 Splunk 索引器或转发器的默认主机值。

之后概述了这些分配方法及其用例。后续主题将更加详细地介绍这些方法。

默认主机值

如果没有为数据来源指定任何其他主机规则，Splunk 软件将为主机字段分配一个默认值，该值将应用到从任何输入进入到实例中的所有数据。默认主机值是最初获取数据的 Splunk 索引器或转发器的主机名或 IP 地址。Splunk 实例在发生事件的服务器上运行属于正常情况，无需进行手动干预。

更多信息请参阅本手册中的设置 Splunk 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 Enterprise，或使用了从您环境中其他主机转发来的文件，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 软件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。

更多信息请参阅本手册中的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，若您用中央日志主机向您的 Splunk 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。

更多信息请参阅本手册中的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，请不要担心。您可以通过多种方法解决或处理此问题。

详情请参阅本手册中的创建索引后更改主机值。

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《知识管理器》手册中的“关于标记和别名”。

设置 Splunk 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。

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（使用本章中后续主题中的信息），则事件的默认主机值为服务器的主

机名或 IP 地址，该服务器运行获取事件数据的 Splunk 实例（转发器或索引器）。如果事件源自 Splunk 实例在上面运行的服务器，则该主机分配正确，无需做任何更改。但是，如果您的所有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`。

使用 Splunk Web 设置默认主机值

1.在 Splunk Web 中单击**设置**。

3.在“设置”页面上单击**常规设置**。

4.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**。

5.保存更改。

该操作将为传入 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，如本章后续部分所述。

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 中设置。您可以通过编辑位于 `$SPLUNK_HOME/etc/system/local/` 内或 `$SPLUNK_HOME/etc/apps/` 中您自己的自定义应用程序内的文件来修改主机值。

主机分配指定于 `[default]` 段落中。

这是 `inputs.conf` 中默认主机分配的格式：

```
[default]
host = <string>
```

把 `<string>` 设置为您选择的默认主机值。`<string>` 默认为数据源自其内的主机的 IP 地址或域名。

警告：请*勿*在 `<string>` 值上加引号。`host=foo`，非 `host="foo"`。

编辑 `inputs.conf` 之后，Splunk 实例必须重启，这样您所做的更改才能生效。

注意：根据默认设置，`host` 属性被设置为变量 `$decideOnStartup`，意味着该属性被设置为运行 `splunkd` 的计算机的主机名称。Splunk 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。

更多信息请参阅本手册中的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 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。

更多信息请参阅本手册中的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

您可以设置来自特定文件或目录输入的所有数据的主机值。您既可以静态设置主机，也可以动态设置主机。

若您静态设置主机值，相同的主机将被分配给从指定文件或目录输入中接收到的所有事件。

若您动态设置主机值，Splunk 软件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。

此时，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"定义和管理标记"。

使用 Splunk Web

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。

要在创建新输入时设置默认主机，请参阅设置新输入的默认主机。

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"常量值"选项。
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

创建新输入时，设置默认主机的进程会有所不同。

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，并指定任意白名单或黑名单。
5. 单击**下一步**。
6. （可选）设置新输入的来源类型。

   注意：如您指定了目录，"设置 Sourcetype"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**按钮。
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**来创建输入。

编辑 inputs.conf

如需为受监视的文件或目录输入指定主机值，请通过编辑 inputs.conf 来为受监视的文件或目录输入指定主机值。编辑 `inputs.conf` 时，请在定义输入的段落中设置 `host` 属性。如果您使用的是 Splunk Cloud，请在运行 Splunk 通用转发器的计算机上配置该设置。

```
[monitor://<path>]
host = <your_host>
```

编辑 `inputs.conf`，其位于 `$SPLUNK_HOME/etc/system/local/` 内或 `$SPLUNK_HOME/etc/apps/` 中您自己的自定义应用程序目录内。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"。

更多有关输入和输入类型的信息，请参阅本手册中的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
```

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。

有关正则表达式语法和用法的入门，请参阅 Regular-Expressions.info。您可以用这两种方法测试正则表达式：通过 rex 搜索命令在搜索中使用正则表达式；以及使用第三方工具来编写并测试正则表达式。

使用 Splunk Web

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：

    - **路径的正则表达式** - 如果要通过正则表达式提取主机名，则可以选择此选项。然后在**正则表达式**字段中输入要提取的主机的正则表达式。

    - **路径中的段** - 如果要从数据来源路径中的段提取主机名，则可以选择此选项。然后在**段编号**字段中输入段编号。例如，若前往来源的路径为 `/var/log/<host server name>`，而您想要使第三段（主机服务器名称）成为主机值，请输入 "3"。
5. 单击**保存**。

动态设置新输入的默认主机

创建新输入时，动态设置默认主机的进程会有所不同。

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，并指定任意白名单或黑名单。
5. 单击**下一步**。
6. （可选）设置新输入的来源类型。**注意：**如您指定了目录，“设置 Sourcetype”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**来创建输入。

编辑 inputs.conf

配置 `inputs.conf` 可让您设置动态主机提取规则。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 host_regex 属性设置事件主机

1. 编辑 `inputs.conf`，其位于 `$SPLUNK_HOME/etc/system/local/` 内或 `$SPLUNK_HOME/etc/apps/` 中您自己的自定义应用程序目录内。
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 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 `inputs.conf`，其位于 `$SPLUNK_HOME/etc/system/local/` 内或 `$SPLUNK_HOME/etc/apps/` 中您自己的自定义应用程序目录内。
2. 把 `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>` 且您想要使第三个段（主机服务器名称）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 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
```

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），Splunk 软件将根据您在 host_segment 中指定的段编号提取主机值。
  - 若 source 不含 /，或您指定的 host_segment 值大于 source 中可用的段数量，则 Splunk 软件无法提取主机值，只能使用提取数据的主机的名称。请参阅以下示例：

示例 1：主机名称为 server01，来源路径为 /mnt/logs/server01，inputs.conf 包含：

```
[monitor:///mnt/logs/]
host_segment = 3
```

结果主机值：server01

示例 2：主机名称为 server01，来源路径为 /mnt/logs/server01，inputs.conf 包含：

```
[monitor:///mnt/logs/server01]
source = /mnt/logs/server01
host_segment = 3
```

结果主机值：server01

示例 3：主机名称为 server02，来源路径为 /mnt/logs/server02，inputs.conf 包含：

```
[monitor:///mnt/logs/server02]
source = serverlogs
host_segment = 3
```

结果主机值：server02

基于事件数据设置主机值

通过对 Splunk 软件进行配置，您可以根据事件中的数据把主机名分配到您的事件。本主题向您展示如何借助事件数据，利用 props.conf、transforms.conf 和正则表达式覆盖默认的主机分配。

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 社区 wiki 还有一个列表，其中列出了可用于编写和测试正则表达式的有用工具。

配置

要逐一配置事件覆盖，您需要创建两个段落，一个在 transforms.conf 中，另一个在 props.conf 内。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。如果您使用的是 Splunk Cloud，请在运行 Splunk 通用转发器的计算机上编辑这些设置。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”。

transforms.conf

在本语法后的 transforms.conf 中创建一个段落：

```
[<unique_stanza_name>]
REGEX = <your_regex>
FORMAT = host::$1
DEST_KEY = MetaData:Host
```

请注意以下事项：

- `<unique_stanza_name>` 应反映其涉及一个主机值。您稍后将在 `props.conf` 段落中使用此名称。
- `<your_regex>` 是一个正则表达式，用来识别您在事件中想要提取主机值的位置。
- `FORMAT = host::$1` 将 `REGEX` 值写入 `host::` 字段。

props.conf

然后在 `props.conf`（其引用了 `transforms.conf` 段落）中创建一个段落：

```
[<spec>]
TRANSFORMS-<class> = <unique_stanza_name>
```

请注意以下事项：

- `<spec>` 可为：
  - `<sourcetype>`，事件的来源类型。
  - `host::<host>`，其中 `<host>` 为事件的主机值。
  - `source::<source>`，其中 `<source>` 为事件的来源值。
- `<class>` 为任何您想要赋予转换的唯一标识符。
- `<unique_stanza_name>` 为您在 `transforms.conf` 中创建的段落的名称。

示例

假设从来自 `houseness.log` 文件的下列事件组开始。主机位于第三个位置（"fflanda" 等）。

```
41602046:53 accepted fflanda
41602050:29 accepted rhallen
41602052:17 accepted fflanda
```

首先，利用提取主机值的正则表达式在 `transforms.conf` 中创建一个新的段落：

```
[houseness]
DEST_KEY = MetaData:Host
REGEX = \s(\w*)$
FORMAT = host::$1
```

之后，把您的 `transforms.conf` 段落引用到 `props.conf` 段落。例如：

```
[source::.../houseness.log]
TRANSFORMS-rhallen=houseness
SHOULD_LINEMERGE = false
```

以上段落包含传统的属性/值对 `SHOULD_LINEMERGE = false`，可在每个换行符处划分事件。

此时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，您可能会发现一些事件的主机值不正确。例如，您可能直接在 Splunk Enterprise 服务器上将某些 Web 代理日志收集到目录中，并且将该目录作为输入添加而删除覆盖主机字段的值的记忆，则将导致主机值与 Splunk Enterprise 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：

- 请删除并重新为数据建立索引。
- 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。
- 标记不正确的主机值，然后使用标记进行搜索。

- 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。
- 将主机字段化名为新的字段（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）。

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清晰的替代方案。

配置来源类型

来源类型为何重要

来源类型是 Splunk 软件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 Splunk 软件您所获数据的类型，方便 Splunk 软件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 软件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 软件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，则您可能需要创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，因此您可能将在搜索中频繁使用来源类型。

典型来源类型

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 软件会自动识别的常见来源类型包括：

- **access_combined**，对于 NCSA 组合格式 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，通常经由远程 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。
- 重新创建新来源类型，或者通过修改现有来源类型来创建新来源类型。

分配来源类型

大多数情况下，Splunk 软件会为您的数据决定最佳来源类型，并自动把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：

- 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
- 创建来源类型
- 重命名来源类型

本主题后续将有一个部分介绍 Splunk Enterprise 如何分配来源类型。

创建新来源类型

如果现有来源类型都不符合您的数据需求，则您可以创建一个新类型。

Splunk Web 允许用户调整来源类型，使其更好地符合数据。事实上，Splunk Web 就是一个可视化的来源类型编辑器。请参阅"设置 Sourcetype"页面。

如果您使用的是 Splunk Enterprise，您也可以通过直接编辑 props.conf 和添加来源类型段落来创建新的来源类

型。请参阅创建来源类型。 如果您使用的是 Splunk Cloud，请使用 Splunk Web 定义来源类型。

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"设置 Sourcetype"页面。

搜索来源类型

`sourcetype` 为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（Splunk 术语也称主机）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 软件如何分配来源类型

Splunk 软件采用各种方法在索引时间为事件数据分配来源类型。处理事件数据过程中，Splunk 软件将按照定义好的优先顺序逐步执行这些方法。先从 `inputs.conf` 和 `props.conf` 中的硬编码来源类型配置开始，然后移动到基于规则的来源类型关联，最后再通过诸如来源类型自动识别和来源类型自动学习等方法工作。这些方法可以在 Splunk 软件把来源类型值自动分配给其他事件的同时，让您配置 Splunk 软件把来源类型值应用到特定事件类型的方法。

以下列表显示 Splunk 软件如何为数据导入决定来源类型。Splunk 软件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。该列表还概述了如何为每个级别配置来源类型分配。

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 软件会停在此处。

在 inputs.conf 或 Splunk Web 中完成此配置。以下为 `inputs.conf` 语法，用于将来源类型分配到文件输入：

```
[monitor://<path>]
sourcetype=<sourcetype>
```

在 Splunk Web 中定义输入时也可以分配来源类型。针对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 软件会停在此处。

使用以下语法在 props.conf 中完成此配置：

```
[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 软件接下来会为来源类型查找您已经创建的规则。

您可以在 `props.conf` 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 软件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 软件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 软件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。

若需查看 Splunk 软件可直接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 软件会查寻延迟的规则。

运行方式和基于规则的关联相似。创建一个 `delayedrule::` 段落，创建位置为 `props.conf`。这种“获取全部”来源类型的方法很有用，以免 Splunk 软件在使用智能匹配时错过任何来源类型（见上文）。

延迟规则关联的一个范例就是上述第 3 步中使用 `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::` 把通用的 "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
```

更多有关设置或移除来源类型识别延迟规则的信息，请参阅配置基于规则的来源类型识别。

自动来源类型学习

如果无法使用前述方法为事件分配来源类型，Splunk 软件会为事件签名创建一个新的来源类型（见上文第 4 步）。Splunk 软件将学习到的模式信息存储在 sourcetypes.conf 中。

覆盖自动来源类型分配

Splunk 软件会尝试为您的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 软件进行配置，这样它就可以根据数据导入或数据来源分配来源类型。

Splunk 软件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 软件如何分配来源类型。

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，可转而为来源类型创建一个标记。

本主题介绍如何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的数据分配来源类型，例如 `/var/log/`。如果使用的是 Splunk Enterprise，您可以在 Splunk Web 中执行此操作，或者也可以通过编辑 `inputs.conf` 配置文件来执行此操作。如果您使用的是 Splunk Cloud，请使用 Splunk Web 定义来源类型。

注意：虽然通过输入分配来源类型看起来是一种简单的处理方法，但结果不够细致；使用这种方法时，Splunk 软件会为**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，您可以按照本主题后续部分中的介绍，根据数据的来源为其分配来源类型。

使用 Splunk Web

定义数据导入时，您可以设置一个来源类型值，将其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：

1. 单击 Splunk Web 右上角的**设置**。

2. 在“设置”弹出窗口的“数据”部分中，单击**数据导入**。

3. 单击**文件和目录**。

4. 单击**新建**按钮以添加输入。

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 Sourcetype”页面，请单击 "Sourcetype" 下拉列表并从预置来源类型列表中选择。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“设置 Sourcetype”页面。

8. 如果您想要把来源类型另存为其它名称，请单击**另存为...**即可打开保存新来源类型的对话框。否则，请转到步骤 10。

9. 如果您选择保存来源类型，Splunk Web 将显示“保存 Sourcetype”对话框。请输入名称、描述、类别和来源类型应适用的应用。请参阅将修改保存为新来源类型。

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。

Splunk 软件现在会把您选择的来源类型分配给 Splunk 软件针对该输入为其建立索引的所有事件。

使用 inputs.conf 配置文件

在 inputs.conf 中配置输入时，可以为该输入指定来源类型。编辑 `inputs.conf`，其位于 `$SPLUNK_HOME/etc/system/local/` 内或 `$SPLUNK_HOME/etc/apps/` 中您自己的自定义应用程序目录内。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"。

警告：请勿在属性值两侧加引号：`sourcetype=log4j`，非 `sourcetype="log4j"`。

指定来源的来源类型

使用 props.conf 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 `props.conf`，其位于 `$SPLUNK_HOME/etc/system/local/` 内或 `$SPLUNK_HOME/etc/apps/` 中您自己的自定义应用程序目录内。有关配置文件的一般信息，请参阅“关于配置文件”。

注意：若要转发数据且想为某来源分配一个来源类型，您必须在**转发器**上的 `props.conf` 中分配该来源类型。若您在**接收器**上的 `props.conf` 中执行此操作，则覆盖无效。

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]`）应尽可能具体一点。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
```

这样非常危险。它将指示 Splunk 软件把 `/home/fflanda` 中的所有 gzip 文件处理为 "mytype" 文件，而非 gzip 文件。

相反，请将正则表达式编写为：

```
[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。您可以在 `props.conf` 中创建一个可以把特定来源类型和一组限定条件关联起来的 `rule::` 段落。获取数据时，Splunk 软件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。

您可以在 `props.conf` 中创建两种规则：规则和延迟规则。二者之间唯一的区别是来源键入过程期间 Splunk 软件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 软件会使用多种方法确定来源类型：

- 根据数据导入或来源检查显示来源类型定义后，Splunk 软件会查看所有 `rule::` 段落（定义于 `props.conf`

中），并尝试根据这些段落内指定的分类规则把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 软件将尝试使用来源类型自动匹配，试着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 软件接下来将检查所有 `delayedrule::` 段落（位于 `props.conf` 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 软件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 软件如何分配来源类型。

有关正则表达式语法和用法的入门，请参阅 Regular-Expressions.info。您可以用这两种方法测试正则表达式：通过 rex 搜索命令在搜索中使用正则表达式；以及使用第三方工具来编写并测试正则表达式。-->

您可以配置自己的系统，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 软件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。

配置

要设置来源键入规则，请编辑 `props.conf`，该文件位于 `$SPLUNK_HOME/etc/system/local/` 内或 `$SPLUNK_HOME/etc/apps/` 中您自己的自定义应用程序目录内。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”。

通过把 `rule::` 或 `delayedrule::` 段落添加到 `props.conf` 即可创建一个规则。在段落标题中提供该规则的名称，在段落正文中声明来源类型。在来源类型声明之后列出来源类型分配规则。这些规则使用一个或多个 `MORE_THAN` 和 `LESS_THAN` 语句在数据（按特定百分比符合给定的正则表达式）中查找模式。

要创建规则，可使用以下语法：

```
[rule::<rule_name>] OR [delayedrule::<rule_name>]
sourcetype=<source_type>
MORE_THAN_[0-99] = <regex>
LESS_THAN_[1-100] = <regex>
```

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` 表示不到 20% 的行包含关联的表达式。

注意：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”。

一规则可以包含任意数量的 `MORE_THAN` 和/或 `LESS_THAN` 条件。仅当数据满足规则中的所有语句时，该规则的来源类型才会分配给数据文件。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。

示例

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 软件随附一组名为“预置来源类型”的内置来源类型。

Splunk 软件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。Splunk 软件还包括一些它无法自动识别的预置来源类型，但您可以利用覆盖自动来源类型分配中介绍的方法，通过 Splunk Web 或 `inputs.conf` 手动分配。

如果预置的来源类型与您的数据相匹配，建议使用预置的来源类型，因为 Splunk 软件已经知道如何正确地为预置来源类型建立索引。不过，如果您的数据不符合任何预置来源类型，您可以根据创建来源类型中的介绍，创建自己的来源类型。即使没有自定义属性，Splunk 软件实际上也可以为任何格式的数据建立索引。

有关来源类型的简介，请参阅来源类型为何重要。

自动识别的来源类型

来源类型名称	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（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	Linux 安全日志	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 Unix/Linux 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 Unix/Linux 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 服务器日志	####<Sep 26, 2005 7:27:24 PM MDT> <Warning> <WebLogicServer> <bea03> <asiAdminServer> <ListenThread.Default> <<WLS 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 Corefile 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，但包含较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 Unix 或 Linux 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

来源类型名称	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、.dat 等。旨在获取明显的非文本文件

预置来源类型

以下为所有的预置来源类型，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
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
Webserver	access_combined、access_combined_wcookie、access_common、apache_error、iis†
Splunk	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†

† 这些来源类型使用`INDEXED_EXTRACTIONS` 属性，该属性把 `props.conf` 中的其他属性设置为特定默认值，且要求特殊处理，以便转发到另一个 Splunk 实例。请参阅转发从结构化数据文件中提取的数据。

了解如何配置预置来源类型的工作

若想了解 Splunk 软件在为给定的来源类型建立索引时使用了哪些配置信息，您可以调用 `btool` 实用工具列出属性。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

本主题介绍如何以每个事件为基础来覆盖来源类型。在 Splunk 软件按照来源类型为何重要中的说明完成初始分配后，您可以在分析时间执行此操作。

要逐一配置事件覆盖，可以结合使用 `transforms.conf` 和 `props.conf`。

由于此类覆盖在分析时发生，因此，仅在索引器或重型转发器上有效，在通用转发器上则无效。若想了解在输入/分析/创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”。

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（不是每个事件）来源类型覆盖，请参阅本手册中的覆盖来源类型自动分配。

配置

要逐一配置事件覆盖，您需要创建两个段落，一个在 `transforms.conf` 中，另一个在 `props.conf` 内。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”。

transforms.conf

在本语法后的 `transforms.conf` 中创建一个段落：

```
[<unique_stanza_name>]
REGEX = <your_regex>
FORMAT = sourcetype::<your_custom_sourcetype_value>
DEST_KEY = MetaData:Sourcetype
```

请注意以下事项：

- `<unique_stanza_name>` 应反映其涉及的来源类型。您稍后将在 `props.conf` 段落中使用此名称。

- `<your_regex>` 是一个正则表达式，可以识别您想对其应用自定义来源的事件（例如，具有特定主机名称或其他字段值的事件）。
- `<your_custom_sourcetype_value>` 就是您想要应用到正则表达式所选事件的来源类型。

注意：有关正则表达式语法和用法的入门，请参阅 Regular-Expressions.info。您可以用这两种方法测试正则表达式：通过 rex 搜索命令在搜索中使用正则表达式；以及使用第三方工具。

props.conf

然后在 `props.conf`（其引用了 `transforms.conf` 段落）中创建一个段落：

```
[<spec>]
TRANSFORMS-<class> = <unique_stanza_name>
```

请注意以下事项：

- `<spec>` 可为：
  - `<sourcetype>`，事件的来源类型。
  - `host::<host>`，其中 `<host>` 为事件的主机值。
  - `source::<source>`，其中 `<source>` 为事件的来源值。
- `<class>` 为任何您想要赋予转换的唯一标识符。
- `<unique_stanza_name>` 为您在 `transforms.conf` 中创建的段落的名称。

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"。您的 Splunk 部署可以为通过此输入传入众多主机的大量数据建立索引。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 到达 Splunk 部署的数据。

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 `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
```

您可以轻松地修改此正则表达式，这样就可以把匹配范围缩小到来自目标主机名（本例中为 host1、host2 和 host3）的事件：

```
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
```

之后，您可以在 `props.conf` 段落中指定该转换，该段落可以识别事件的特定输入：

```
[source::udp:514]
TRANSFORMS-changesourcetype = set_sourcetype_my_log_for_some_hosts
```

创建来源类型

您可以通过以下几种方式创建新来源类型：

- 请使用 Splunk Web 中的“设置 Sourcetype”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”管理页面中创建一个来源类型。
- 直接编辑 `props.conf` 配置文件。

尽管通过编辑驻留在转发器计算机上的 .conf 文件也可以配置单个转发器，实现创建来源类型的目的；但是，创建来源类型的最佳做法是使用 Splunk Web，该方法可以确保创建于 Splunk 部署上的来源类型前后一致。

在 Splunk Web 中设置 Sourcetype

Splunk Web 中的“设置 Sourcetype”页面可以轻松查看为数据应用来源类型的效果，并可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。至于其他修改，该页面允许您直接编辑基础的 `props.conf` 文件。更改设置时，您可以立即查看事件数据所发生的更改。

该页面仅当您指定或上传单个文件时显示。当您指定任何其他来源类型时，不会显示该页面。

要了解更多有关此页面的信息，请参阅本手册中的"设置 Sourcetype"页面。

创建来源类型

您可使用"来源类型"管理页面创建新的来源类型。请参阅本手册中的添加来源类型。

编辑 props.conf

如果您使用的是 Splunk Enterprise，您也可以通过编辑 `props.conf` 和添加新的段落来创建新的来源类型。有关 `props.conf` 的详细信息，请参阅《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《管理员》手册中的"关于配置文件"。

以下是 props.conf 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指定多个文件或目录。

```
[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 = iis
```

要编辑 props.conf：

1. 在您想要创建来源类型的主机上创建 `$SPLUNK_HOME/etc/system/local/props.conf`
   注意：您可能需要创建 `local` 目录。如您使用应用，请转到 `$SPLUNK_HOME/etc/apps` 中的应用目录。
2. 请用文本编辑器打开 `props.conf` 文件（位于 `$SPLUNK_HOME/etc/system/local` 中）。
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应使用的属性。

   ```
   [my_sourcetype]
   attribute1 = value
   attribute2 = value
   ```

   注意：请参阅属性列表的 props.conf 规范及它们的使用方式。
4. （可选）若您知道来源类型将被应用于其中的一个（或多个）文件的名称，请在 `[source::<source>]` 段落中指定这些文件：

   ```
   [my_sourcetype]
   attribute1 = value
   attribute2 = 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

当您创建某来源类型时，以下是您应指定的一些关键属性：

- **事件换行。**要了解如何使用 `props.conf` 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` 指定时间戳，请参阅配置时间戳识别及本手册"配置时间戳"一章中的其他主题。

您也可配置许多其他设置。请参阅 props.conf 规范了解更多信息。

管理来源类型

本主题介绍 Splunk Web 来源类型管理页面，您可以在该页面中创建、编辑和删除来源类型。当从“设置”菜单中选择“来源类型”时会将其加载。

“来源类型”页面显示已在实例上配置的所有来源类型。页面上会显示 Splunk 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”管理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

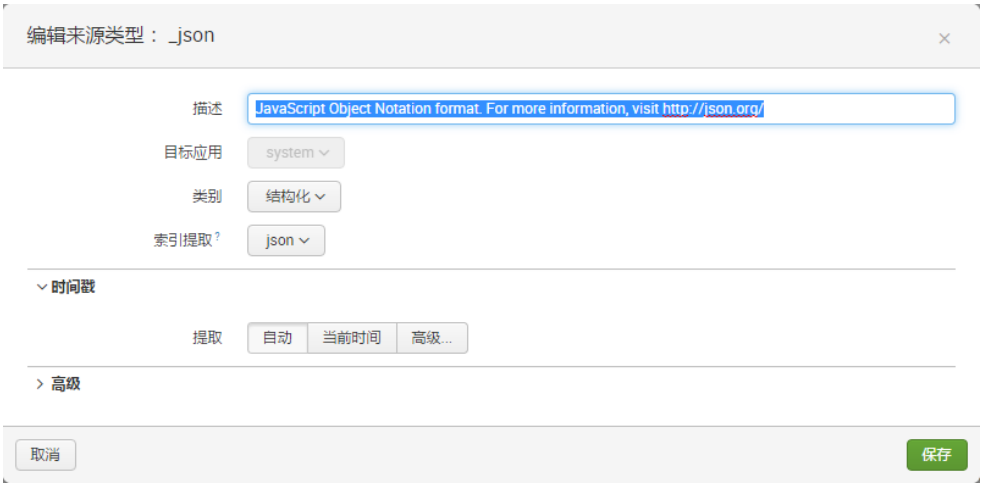
您可对您在“来源类型”管理页面中查看的来源类型数目进行筛选。

- 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”。
- 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”。
- 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“过滤器”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“过滤器”框中键入的内容相匹配的来源类型会显示。要再次设置所有来源类型，请单击“过滤器”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”列单击其“编辑”链接。“编辑来源类型”页面显示。



“编辑来源类型”对话框允许您更改来源类型的配置。您可以进行以下更改：

描述：在“描述”字段中键入来源类型的描述。

目标应用：来源类型应用到的应用程序上下文。

注意：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。

已建立索引的提取：在索引时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。

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。

- 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。

高级时间戳提取配置

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，如果来源文件包含的日志具有此格式的时间戳：

```
6 Jun 2015 18:35:05
```

...在“时间戳格式”字段中指定以下内容：

```
%d %b %Y %H:%M:%S
```

另一个示例：

```
Tue Jun 4 2:55:18 PM 2015
```

映射到

```
%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 man 页面网站上的 strptime(3) (http://linux.die.net/man/3/strptime)。

- 时间戳前缀：代表在时间戳之前显示的字符的正则表达式。在事件中看到该字符集时，Splunk 软件期望前缀之后会出现一个时间戳。

- 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

对话框的“高级”部分向您显示该来源类型的所有配置（按键/值格式）。这代表定义来源类型的 `props.conf` 文件中的内容。您可直接编辑每个设置或添加与删除设置。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“新设置”链接。这将显示字段的键/值对。请在“名称”字段中输入键名称并在“值”字段中输入键值。

警告：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。

添加来源类型

要创建一个新来源类型，请单击屏幕右上方的“新来源类型”按钮。“创建来源类型”对话框打开。

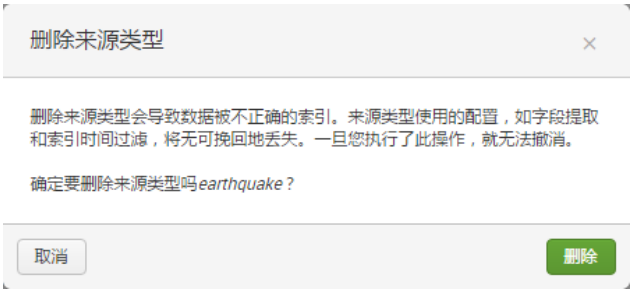
该对话框与“编辑来源类型”对话框完全一样。关于控制的信息，请参阅对话框中的 "Managesourcetypes"。

当您已完成配置来源类型，请单击“保存”。

删除来源类型

要删除来源类型，请单击您想要删除的来源类型“操作”列中的“删除”链接。您无法删除内置的来源类型，只能删除您创建的或应用随附的来源类型。

当您删除来源类型时，会显示以下对话框：



警告：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。

- 当您删除来源类型时，可能会对数据错误地建立索引。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。本案例中可用的选项仅包括从备份中恢复定义了来源类型的 `props.conf` 文件，或重新手动创建该来源类型。

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。

搜索时重命名来源类型

某些情况下，您可能希望重命名来源类型。例如，假设您将输入意外分配给错误的来源类型。或者意识到搜索时两个名称不同的来源类型应按照完全相同的名称处理。

如果使用的是 Splunk Enterprise，您可以利用 `rename` 属性（位于 `props.conf` 中）在搜索时间为事件分配新的来源类型。以防您需要搜索新的来源类型，原始的来源类型已被移动至单独的字段 `_sourcetype`。

注意：索引的事件仍包含原始来源类型名称。重命名操作只在搜索时发生。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，把 `rename` 属性添加到您的来源类型段落中：

```
rename = <string>
```

注意：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 `_`（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"。您可以使用 `props.conf` 段落，把 "whoops" 重命名为 "cheese_shop"：

```
[whoops]
```

```
rename=cheese_shop
```

现在，搜索 "cheese_shop" 时将找出所有 "whoops" 事件以及一开始具有 "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
```

重要提示：来自重命名的来源类型的数据仅使用目标来源的搜索时间配置（本例中为 "cheese_shop"）。原始来源类型的所有字段提取（本例中为 "whoops"）都将被忽略。

管理事件分段

关于事件分段

分段功能可以在**索引**时间和**搜索**时间把事件划分成可搜索的**段**。段可以分类为**主要**或**次要**。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`" 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创建搜索的能力。

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 Enterprise，您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。

如果使用的是 Splunk Cloud，您可以在重型转发器计算机上配置索引时间分段，但若要配置搜索时间分段，则必须提交 Splunk 支持问题。

事件分段的类型

分段有三种主要类型或级别，可在索引或搜索时间内配置：

- **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**与内部分段相反。在外部分段的情况下，Splunk 软件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**是内部和外部分段的组合。完全分段的情况下，系统在为 IP 地址创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。

`segmenters.conf` 文件（位于 `$SPLUNK_HOME/etc/system/default`）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。

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。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。

有关可用的默认分段类型的信息，请查看 `segmenters.conf` 文件（位于 `$SPLUNK_HOME/etc/system/default` 内）。

重要提示：不要修改默认文件。若想更改现有的分段段落或创建全新的分段段落，您可以把默认文件复制到 `$SPLUNK_HOME/etc/system/local/` 或 `$SPLUNK_HOME/etc/apps/` 中的自定义应用目录。有关配置文件和目录位置的信息，请参阅“关于配置文件”。

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` 事件创建索引，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，您可以创建其他类型。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>`
  - 定义于 `segmenters.conf` 中的分段类型或 "rule"
  - 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` 创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>` 可以为下列各项中的一项：`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>`
  - 定义于 `segmenters.conf` 中的分段类型或 "rule"
  - 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` 创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。必须为数据重新创建索引，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。虽然这与索引时间分段无关，但 Splunk Web 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：

1.执行搜索。查看结果。

2.单击返回的事件集上方的**选项...**。

3.在**事件分段**下拉框中，从可用选项中选择：完全、内部、外部或原始。默认为“完全”。

可以按照“设置事件数据的分段”所述，配置这些下拉选项的含义。

改善数据导入过程

使用测试索引测试输入

将新输入添加到生产索引之前，最好对其进行测试。将输入添加到测试索引中。验证接收的数据导入正确且生成的事件格式可用后，可以将输入指向默认 "main" 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，或者索引事件并未采用需要的结构，则可以一直处理测试索引，直至对结果满意。情况好转时，您可以编辑输入以改为指向 main 索引。

您可以预览 Splunk 软件为您的数据建立索引并生成测试索引的方式。在预览期间，您可交互调整一些事件处理设置。有关详细信息，请参阅“‘设置 Sourcetype’页面”。

使用测试索引

要了解如何创建和使用自定义索引，请参阅《管理索引器和群集》手册中的“创建自定义索引”。此主题中详细介绍的有几个基本步骤：

1.使用 Splunk Web，或者如果您有 Splunk Enterprise，也可以使用 CLI 来创建测试索引；直接编辑 `indexes.conf` 亦可创建测试索引。详细信息请参阅“创建自定义索引”。

2.配置数据导入时，将事件发送到测试索引。通常可以在 Splunk Web 中执行此操作。对于每个输入：

a. 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。

b. 在**索引**下拉框中，选择您的测试索引。该数据导入的所有事件现在都将转到此索引。

c. 对要发送到测试索引的每个数据导入都重复此过程。

如此处的说明所示，您还可以在 `inputs.conf` 中配置输入时，指定一个索引。

3.搜索时，在搜索命令中指定测试索引。（默认情况下，Splunk 软件会搜索 "main" 索引。）使用 `index=` 命令：

```
index=test_index
```

注意 ：若要搜索事件的测试索引，而该些事件又来自您新创建的输入，请在字段边栏中使用*实时 > 全时（实时）*时间范围。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创建索引，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，这项功能尤其有用。

删除索引数据并重新开始

根据此处的说明，若想清除测试索引并重新开始，请使用 CLI `clean` 命令。

将输入指向默认索引

对结果满意并准备真正开始创建索引时，最好编辑数据导入以使其指向默认的 "main" 索引，而非测试索引。此过程比较简单，只是颠倒您起初使用测试索引时所采用的步骤。对于已经设置的每个数据导入：

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：

a. 选择**系统 > 系统配置 > 数据导入**。

b. 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。

c. 选择要编辑的特定数据导入。这会将您带到可对其进行编辑的屏幕。

d. 选择**显示高级设置**选项。转到名为**索引**的字段。

e. 在**索引**下拉框中，选择 **main** 索引。该数据导入的所有事件现在都将转到此索引。

根据此处的说明，若您改用 `inputs.conf` 配置输入，可以直接在该文件中更改索引。

2.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。这有助于防止备份**转发器**或**索引器**时数据丢失。

默认情况下，转发器和索引器有 500KB 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，则在队列用尽时会出现不希望的后果。如果是 UDP，数据会在队列中减少且会丢失。对于其他输入类型，会备份生成数据的应用程序。

通过实施保留队列，可以帮助防止此问题发生。使用保留队列时，一旦内存中队列已满，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，它处理来自队列（内存中和磁盘）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。

注意：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 Splunk 软件崩溃，您仍可能会丢失数据。例如，Splunk 软件会将一些输入数据保留在内存队列中以及保留在永久性队列文件中。如果发生崩溃，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，它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（例如文件监视）。

保留队列可用于以下输入类型：

- TCP
- UDP
- FIFO
- 脚本式输入
- Windows 事件日志输入

保留队列不可用于以下输入类型：

- 监视器
- 批处理
- 文件系统更改监视器
- splunktcp（来自 Splunk 转发器的输入）

配置保留队列

使用 inputs.conf 文件配置保留队列。

输入不共享队列。可以在特定输入的段落中配置保留队列。

语法

要创建保留队列，请在特定输入的段落中指定以下两个属性：

```
persistentQueueSize = <integer>(KB|MB|GB|TB)
* Max size of the persistent queue file on disk.
* Defaults to 0 (no persistent queue).
```

示例

下面是为 tcp 输入指定保留队列的示例：

```
[tcp://9994]
persistentQueueSize=100MB
```

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，保留队列位于以下位置：

```
$SPLUNK_HOME/var/run/splunk/[tcpin|udpin]/pq__<port>
```

注意：文件名中有*两个*下划线：`pq__<port>`，*非* `pq_<port>`。

例如：

- TCP 端口 2012 的保留队列：`$SPLUNK_HOME/var/run/splunk/tcpin/pq__2012`
- UDP 端口 2012 的保留队列：`$SPLUNK_HOME/var/run/splunk/udpin/pq__2012`

若为 FIFO 输入，保留队列驻留在 `$SPLUNK_HOME/var/run/splunk/fifoin/<encoded path>` 下。

若为脚本式输入，保留队列驻留在 `$SPLUNK_HOME/var/run/splunk/exec/<encoded path>` 下。`inputs.conf` 中的 FIFO/脚本式输入段落派生出 `<encoded path>`。

输入过程故障排除

本主题介绍在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

把某输入添加到您的 Splunk 部署时，该输入相对地添加到了您所处的应用中。一些应用会把输入数据写入到指定的索引中。如果您确定数据就在 Splunk 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。

- 更多有关角色的信息，请参阅*《确保 Splunk Enterprise 安全》*手册中有关角色的主题。
- 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！”

注意：如果您使用的是 Splunk Enterprise 而且通过编辑 `inputs.conf` 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` 后重启 Splunk 服务。

尾文件故障排除

您可以使用 `FileStatus` 表述性状态转移 (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，用于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》。

监视器输入故障排除

有关处理监视器输入问题的各种信息，请参阅社区 Wiki 上的“监视器输入故障排除”。

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。详细内容请参阅*监视 Splunk Enterprise*。