

Splunk® Enterprise 6.5.0

知识管理器手册

生成时间：2016 年 9 月 26 日，下午 10:24

Table of Contents

欢迎使用知识管理	4
Splunk 知识是什么？	4
为什么管理 Splunk 知识？	4
知识管理的前提条件	5
组织和管理知识对象	6
通过“设置”页面管理知识对象	6
监视和组织知识对象	6
制定知识对象的命名约定	7
了解和使用通用信息模型加载项	8
管理知识对象权限	8
解决孤立的搜索、报表和告警	10
禁用或删除知识对象	12
搜索时间操作的顺序	13
数据解释方式：字段和字段提取	18
关于字段	18
当 Splunk Enterprise 提取字段时	21
使用字段提取器构建字段提取	22
字段提取器：“选择示例”步骤	25
字段提取器：“选择方法”步骤	28
字段提取器：“选择字段”步骤	29
字段提取器：“重命名字段”步骤	33
字段提取器：“验证”步骤	35
字段提取器：“保存”步骤	35
使用“字段提取”页面	36
使用“字段转换”页面	39
通过配置文件创建和维护搜索时间字段提取	42
通过 fields.conf 配置多值字段	52
关于已计算字段	53
通过 Splunk Web 创建已计算字段	54
通过 props.conf 配置已计算字段	54
使用默认字段	55
关于 Splunk 正则表达式	58
数据分类：事件类型和交易	61
关于事件类型	61
在 Splunk Web 中配置事件类型	62
关于事件类型优先级	64
自动查找和构建事件类型	65
在 eventtypes.conf 中配置事件类型	67
配置事件类型模板	68
关于交易	68
搜索交易	69
配置交易类型	70
数据浓缩：查找和工作流动作	72

关于查找和工作流动作	72
使用字段查找将信息添加到事件中	73
查找配置简介	77
配置 CSV 查找	78
配置外部查找	81
配置 KV 存储查找	82
配置地理空间查找	85
在查找配置中添加字段匹配规则	87
配置基于时间的查找	88
设为自动查找	88
在 Splunk Web 中创建和维护工作流动作	91
数据标准化：标记和别名	98
关于标记和别名	98
在“搜索”中为字段值对设置标记	99
在“设置”中定义和管理标记	101
为主机字段设置标记	103
为事件类型设置标记	103
在 Splunk Web 中创建字段别名	104
通过 props.conf 创建字段别名	104
搜索快捷方式：搜索宏	105
在搜索中使用搜索宏	105
在“设置”中定义搜索宏	105
搜索宏示例	107
使用数据集	108
关于数据集	108
查看和管理已有数据集	109
数据集扩展	112
构建数据模型	114
关于数据模型	114
管理数据模型	118
设计数据模型	123
定义数据集字段	127
添加自动提取字段	129
添加 eval 表达式字段	131
添加查找字段	132
添加正则表达式字段	134
添加地理 IP 字段	136
使用数据摘要加速搜索	137
基于摘要的搜索和数据透视表加速概述	137
管理报表加速	139
加速数据模型	149
使用摘要索引提高报表效率	156
管理摘要索引间隙	160
配置摘要索引	162
配置批处理模式搜索	165

欢迎使用知识管理

Splunk 知识是什么？

Splunk 软件是一款功能强大的搜索和分析引擎，它可以帮助您详细了解 IT 数据信息以及查看更大规模的数据模式。使用 Splunk 软件，您不仅可以查看日志文件中的各个条目；还可以利用这些文件集中保留的信息来进一步了解 IT 环境。

Splunk 软件会从您的 IT 数据中提取各种不同类型的知识（事件、字段、时间戳等），以帮助您用更佳、更智能、更集中的方式处理这些信息。其中的部分信息在索引时提取，即 Splunk 软件为您的 IT 数据创建索引时。但大部分信息都由 Splunk 软件及其用户在“搜索时”创建。与用于决定预先提取或分析哪些信息的基于数据库或架构的分析工具不同，Splunk 软件可以根据需要从原始数据中动态提取知识。

当您的组织使用 Splunk 软件时，会陆续创建其他类别的 Splunk 软件知识对象，包括事件类型、标记、查找、字段提取、工作流动作和保存的搜索。

您可以将 Splunk 软件知识视为一个多功能工具，可用于发现和分析 IT 数据的各个方面。例如，您可以通过事件类型快速轻松地对相似事件进行分类和分组；然后可以使用这些事件来对精确定义的子事件组执行分析搜索。

《知识管理器》手册介绍了如何通过 Splunk Web 和配置文件维护组织的各组知识对象，还演示了如何利用 Splunk 知识来解决组织的实际问题。

Splunk 软件知识分为五类：

- **数据解释方式：字段和字段提取** - 字段和字段提取是 Splunk 软件知识的首要组成部分。Splunk 软件自动从 IT 数据中提取的字段有助于理解原始数据的含义，并对初看上去难以理解的内容进行了说明。您手动提取的字段将基于这一层含义进行扩展和完善。
- **数据分类：事件类型和交易** - 您可以使用事件类型和交易将所需的相似事件分为一组。事件类型将通过搜索发现的若干事件分为一组，而交易是指一定时间跨度内概念相关事件的集合。
- **数据浓缩：查找和工作流动作** - 查找和工作流动作是知识对象的类别，它们以各种方式扩展了数据的可用性。字段查找使您可以将来自外部数据源（如静态表（CSV 文件）或基于 Python 的命令）的字段添加到数据中。工作流动作使您可以在您的数据字段和其他应用程序或网络资源之间交互操作，例如，对包含 IP 地址的字段进行 WHOIS 查找。
- **数据标准化：标记和别名** - 标记和别名用于管理和规范化各组字段信息。您可以使用标记和别名将若干组相关字段值组合在一起，并为所提取的字段赋予标记，以反映其特性的不同方面。例如，您可以将来自某个特定位置（如某一座建筑物或某个城市）的主机的事件分为一组，只需为每个主机赋予相同的标记即可。或者，您可能有两个不同的数据来源，它们使用不同的字段名称来引用相同数据，此时您可以使用别名将数据规范化（例如，将 `clientip` 的别名设置为 `ipaddress`）。
- **数据模型** - 数据模型表示一个或多个数据集，可驱动数据透视表工具，使数据透视表用户能够快速生成有用的表格、复杂的可视化以及功能强大的报表，而无需与 Splunk 软件搜索语言进行交互。数据模型是由知识管理员设计的，知识管理员完全了解其索引数据的格式和语义。典型的数据模型会使用本手册中介绍的其他知识对象类型，包括查找、交易、搜索时间字段提取和已计算字段。

《知识管理器》手册包含了以下主题的相关信息：

- **搜索和数据透视表任务** - 搜索和数据透视表任务是搜索或数据透视表每次运行的项目。如果这些任务未保存或未与其他人共享，则系统将自动在 10 分钟内删除它们。知识管理员可以通过“任务”页面检查和管理最近运行并保存的任务。
- **基于摘要的报表和数据模型加速** - 如果搜索和数据透视表的完成速度很慢，可使用 Splunk 软件来加速任务。本章讨论报表加速（用于搜索）、数据模型加速（用于数据透视表）和摘要索引（用于特殊情况搜索）。

此时，您可能会问“为什么还要‘管理’Splunk 知识？”答案请参阅本章的下一个主题“[为什么管理 Splunk 知识？](#)”。

知识管理员应该对数据导入设置、事件处理以及索引概念有基本的了解。有关更多信息，请参阅本章的第三个主题“[知识管理的前提条件](#)”。

为什么管理 Splunk 知识？

如果您需要在您的 Splunk 部署中维护大量知识对象，您就会知道管理这些知识很重要。这对于拥有大量 Splunk 用户的组织而言尤其如此，而且如果您有多个用户团队都在使用 Splunk 软件，情况更是如此。这只是因为用户的增长将导致额外 Splunk 知识的增长。

如果对这种情况置之不顾的话，您的用户可能需要对大量具有误导性名称或冲突名称的对象进行排序，艰难地查找和使用对应用分配和权限应用不均衡的对象，并且浪费宝贵的时间来创建已经存在于系统中其他位置的对象（如报表和字段提取）。

Splunk 知识管理器可让您对 Splunk 软件知识进行集中监视。知识管理器可以提供的益处包括：

- **监视各团队、部门和部署之间知识对象的创建和使用情况**。如果您的大型 Splunk 部署跨若干个用户团队，您总会找到进行“重复性操作”的团队，他们所设计的对象是其他团队已经开发出来的对象。通过监视对象

创建过程并确保在各部署之间全局共享有用的“通用”对象，知识管理器可以缓解这些情况。

有关更多信息，请参阅本手册中的[“监视和组织知识对象”](#)。

- **规范化事件数据。**简而言之：知识对象会激增。尽管 Splunk 软件所基于的是数据索引，而不是数据库，但是规范化的基本原则仍然适用。对于任何功能强大、运行良好的 Splunk 实现而言，很容易产生数十个标记（所有标记均添加到同一字段），但是这些多余的知识对象会堆叠在一起，进而导致部分用户产生混淆和工作效率低下。我们将为您提供一些提示，通过应用统一的命名标准并使用 Splunk 的“通用信息模型”来规范化知识对象库。

有关更多信息，请参阅本手册中的[“制定知识对象的命名约定”](#)。

- **通过配置文件来管理知识对象。**在管理 Splunk 知识方面，真正的知识管理专家知道如何以及何时利用强大配置文件功能。某些方面的知识对象设置最好通过配置文件来处理。本手册将向您演示如何以此方式来处理知识对象。

有关如何通过配置文件来管理 Splunk 知识的示例，请参阅本手册中的[“通过索引文件创建和维护搜索时间字段提取”](#)。

- **为数据透视表用户创建数据模型。**Splunk 软件为想要快速创建表格、图表和仪表板的用户提供了数据透视表工具，使用户无需编写搜索字符串（有时较长且复杂）。数据透视表工具由**数据模型**来驱动 - 没有数据模型，数据透视表无法生成报表。数据模型由 Splunk 知识管理员设计：他们了解其索引数据的格式和语义，且熟悉 Splunk 搜索语言。

有关数据模型架构和使用情况的概念性概述，请参阅本手册中的[“关于数据模型”](#)。

- **管理基于摘要的搜索和数据透视表加速工具的设置和使用。**无论是启动搜索、运行报表或尝试使用数据透视表，大量数据都将导致 Splunk 软件的性能减缓。要加速任务，知识管理员可使用**报表加速**、**数据模型加速**和**摘要索引**以有助于确保您部署中的团队可快速、高效地获得结果。本手册介绍如何对这些加速策略进行集中监视，以便您可确保合理、高效地使用它们。

有关更多信息，请参阅本手册中的[“基于摘要的搜索和数据透视表加速概述”](#)。

知识管理的前提条件

大部分知识管理任务是以“搜索时间”事件操作为中心的。也就是说，典型的知识管理器通常关注的不是在为事件创建索引之前所发生的工作，例如，设置数据导入、调整事件处理活动、纠正默认字段提取问题、创建和维护索引、设置转发和接收等等。

但是，我们建议所有知识管理员都应该很好地理解这些“Splunk 管理员”概念。在这些方面打下坚实基础之后，知识管理员就能够更好地为其部署来规划管理知识对象的方法，进而更好地解决随着时间不可避免地产生的问题。

以下是知识管理员应当熟悉的一些“管理员”主题，以及供您开始了解的相关链接：

- **使用 Splunk 应用：**如果您的部署使用多个 Splunk 应用，您应当了解一些有关这些应用的组织方式以及应用对象管理如何在多应用部署中工作的背景信息。请参阅《管理员手册》中的“什么是应用”、“应用架构和对象所有权”以及“管理应用对象”。
- **配置文件管理：**配置文件位于何处？其组织方式如何？配置文件相互间的优先顺序如何？请参阅《管理员手册》中的“关于配置文件”和“配置文件优先顺序”。
- **为传入数据创建索引：**什么是索引？其工作原理是什么？“索引时间”与“搜索时间”之间有何区别？为什么会存在如此显著的区别？从《管理索引器和群集手册》中的“关于索引和索引器”开始阅读章节的其余部分。特别注意“索引时间对比搜索时间”。
- **将事件数据导入到您的 Splunk 部署：**首先应对 Splunk 数据导入有一个基本的了解，这点非常重要。请根据阅读《数据导入手册》中的“Splunk 可为哪些内容创建索引”以及其他主题。
- **了解转发和接收设置：**如果您的 Splunk 部署使用转发器和接收器，建议掌控转发器和接收器的实现方式，因为这会影响到您的知识管理策略。有关此主题的概述，请参《转发数据》手册中的“关于转发和接收”。
- **了解事件处理：**建议清楚地了解 Splunk 软件为数据创建索引前，为“分析”数据所进行的步骤。这方面的知识可帮助您解决事件数据的问题并发现“索引时间”事件处理问题。从《数据导入手册》中的“事件处理概述”开始阅读整个章节。
- **默认字段提取：**大多数字段提取发生在搜索时间，但某些默认字段的提取发生在索引时间。身为知识管理员，您大部分时间都在关注搜索时间字段提取，但是建议您在必要时需要了解默认字段的管理方式。这有助于您解决 Splunk 软件应用于每个事件的 `host`、`source` 和 `sourcetype` 字段的相关问题。请从《数据导入手册》中的“关于默认字段”开始阅读。
- **管理用户和角色：**知识管理员通常不直接设置用户和角色。但是，建议了解如何在您的部署中设置用户和角色，因为这直接影响到您在各用户组之间共享和提升知识对象的工作。有关更多信息，请从《管理员手册》中的“关于用户和角色”开始并根据需要阅读章节的其余部分。

组织和管理知识对象

通过“设置”页面管理知识对象

当组织使用 Splunk 软件时，用户会添加知识到其中已建立索引的一组基本事件数据中。您和您的同事可以：

- 保存和计划搜索。
- 给字段添加标记。
- 定义组合各组事件的事件类型和交易。
- 创建查找和工作流动作。

创建知识对象的过程最初很缓慢，但当 Splunk 软件使用一段更长的时间后会变得复杂。用户很容易发生创建已经存在的搜索、添加不必要的标记、设计多余的事件类型等等。如果您的用户群很小，这些问题可能并不明显。但当这些问题随时间累积增多时，它们可能会导致不必要的混乱和重复性工作。

本章介绍知识管理器如何使用设置中的知识页面控制其 Splunk 部署中的知识对象。“设置”可以让细心的知识管理器了解正在创建的知识对象、这些对象的创建者以及（在一定程度上了解）这些对象当前的使用方式。

使用“设置”，您可以轻松地完成以下任务：

- 当您需要时，“从头”或通过对象复制来创建知识对象。
- 检查其他人创建的知识对象，以便减少多余的重复对象，同时确保遵循命名标准。
- 删除不必要的或不明确的知识对象，以免这些对象在下游形成依赖性。
- 确保值得在某一特定工作组、角色或应用间共享的知识对象可供其他组、角色和其他应用的用户使用。

注意：本章假定您作为知识管理员具有 管理员角色或拥有等效权限集的角色。

本章中的主题将介绍如何：

- [保持知识对象集合规范化且整齐有序。](#)
- [制定知识对象的命名约定](#)，这更便于理解和使用知识对象。
- [使用通用信息模型加载项规范化事件数据。](#)
- [管理知识对象权限。](#)使知识对象可供特定应用的用户、具有特定角色的用户或所有应用（具有“全局”权限）的用户使用。
- [禁用或删除知识对象。](#)了解删除知识对象的限制，以及了解删除具有下游依赖性的知识对象的风险。

使用配置文件（而不是“设置”）管理知识

在以前版本中，Splunk Enterprise 用户直接编辑配置文件来添加、更新或删除知识对象。现在，用户可以使用“设置”中的知识页面，它提供了一个图形界面用于更新那些配置文件。

注意：Splunk Cloud 用户必须使用“设置”中的 Splunk Web 知识页面来维护知识对象。

Splunk 建议 Splunk Enterprise 管理员了解如何修改配置文件。由于下列原因，了解配置文件是有好处的：

- 如果您了解任务在配置文件级别的工作方式，有些 Splunk Web 功能会更为有效。这尤为适用于 Splunk Web 中的[字段提取](#)和[字段转换](#)页面。
- 管理需要更改配置文件的某些知识对象类型。
- 只能使用配置文件批量删除已过时、多余或定义不正确的知识对象。
- 您可能会发现您更喜欢直接使用配置文件。例如，如果您是一位已经使用 Splunk Enterprise 很长时间、且已经熟悉配置文件系统的管理员，您可能已经熟悉如何使用配置文件管理 Splunk 知识。其他用户依赖于配置文件可以提供的粒度和控制级别。

《知识管理器》手册中提供了有关通过配置文件处理各种知识对象类型的说明。有关更多信息，请参阅相关类型的文档。

有关 Splunk Enterprise 配置文件的一般信息，请参阅《管理员手册》中的以下主题：

- 关于配置文件
- 配置文件优先顺序

《管理员手册》中也包含一份配置文件参考，包括了用于 Splunk Enterprise 所有配置文件的 .spec 和 .example 文件。

监视和组织知识对象

身为知识管理员，您应该定期检查 Splunk 部署中的知识对象集合。您应该监视是否存在以下类型的知识对象：

- 不符合命名标准
- 重复/多余
- 值得与更多的受众共享
- 因已过时或设计不当而应被禁用或删除

定期检查系统中的知识对象将有助于您发现以后可能会演变成问题的异常情况。

注意：本主题假定您作为知识管理员具有 *管理员* 角色或拥有等效权限集的角色。

示例 - 保持标记简明易懂

大多数运行状况良好的 Splunk 部署最后都将产生很多**标记**，这些标记用于对字段/值对的集合执行搜索。但是随着时间的推移，最后很容易产生名称相似但结果却惊人的不同的标记。这可能会让人十分困惑和沮丧。

您可以遵循如下过程来管理标记。您可以轻松地针对通过 Splunk Web 处理的其他知识对象类型调整此过程。

- 1. 转到 **设置 > 标记 > 按标记名称排列的列表**。
- 2. 查找属于相同应用（或已全局推广可供所有用户使用）的名称相似或重复的标记。例如，您可能在同一应用中找到诸如 `authentication` 和 `authentications` 之类的一组标记，这两个标记分别链接到一组完全不同的字段/值对。
- 或者，您可能遇到除了字母大小写形式不同之外名称完全相同的标记，如 `crash` 和 `Crash`。标记区分大小写，因此 Splunk 软件会将这些标记视为两个独立的知识对象。
- 请记住，如果将**应用上下文**设置为*所有*，而属于不同应用的标记具有相同的名称，那么您可能会发现一些合法的重复标记。这通常是允许的，毕竟，比如 Windows 应用的 `authentication` 标记与 UNIX 应用的 `authentication` 标记所关联的字段/值对必须不同。
- 3. 尝试禁用或删除找到的重复或已过时标记（如果您有相应权限）。**但请注意，这可能会影响依赖此类标记的对象。**如果在报表、仪表板搜索、其他事件类型或交易中使用此类标记，则在删除或禁用该标记之后，这些对象将终止执行。如果对象属于一个应用上下文，而您尝试将其移动到另一个应用上下文，也可能发生这种情况。
- 有关更多信息，请参阅[禁用或删除知识对象](#)。
- 4. 如果要使用更具唯一性的新名称创建一个替代标记，请确保此替代标记连接到与被替换标记相同的字段/值对。

使用命名约定来预防对象命名规则问题

如果您在 Splunk 部署的初期设定好知识对象命名约定，就可以避免发生某些较为棘手的对象命名问题。有关更多信息，请参阅[制定知识对象的命名约定](#)。

制定知识对象的命名约定

最佳方式是，在必要时为知识对象制定有意义的命名约定。如果组织内的所有 Splunk 用户都始终遵循您制定的命名约定，您将会发现这些命名约定更易于使用，而且可以相当方便地一眼就辨别出知识对象的用途。

您可以为 Splunk 部署中几乎所有类型的知识对象制定命名约定。命名约定不但有助于组织对象，还有助于用户区分具有类似用途的各个报表、事件类型以及标记组。另外，您可以借助命名约定来识别对象的各种相关内容，包括可能甚至未体现在对象定义中的内容，例如，哪些团队或位置使用此对象、涉及哪些技术以及此对象是用来做什么的。

在您的 Splunk 部署初期制定命名约定将有助于避免以后发生无序和混乱的情况。

示例 - 为报表设置命名约定

您在公司的系统工程组工作，身为 Splunk 部署的知识管理员，由您负责为团队生成的报表定义一个命名约定。

您制定了一个包含以下项目的命名约定：

- **组**：对应于保存该搜索的用户所在的工作组。
- **搜索类型**：指明搜索的类型（告警、报表、摘要索引填充）。
- **平台**：对应于执行搜索的平台。
- **类别**：对应于主流平台的关注领域。
- **时间间隔**：在其间运行搜索的间隔（或者，如果是计划的搜索，此为搜索的运行间隔）。
- **描述**：有关搜索上下文和目的的有意义的描述，尽可能限制在一到两个词。确保搜索名称是唯一的。

组	搜索类型	平台	类别	时间间隔	描述
SEG NEG OPS NOC	告警 报表 摘要	Windows iSeries 网络	磁盘 Exchange SQL 事件日志 CPU 任务 子系统 服务 安全	<任意>	<任意>

使用此命名约定的可能报表可以为：

- SEG_Alert_Windows_Eventlog_15m_Failures
- SEG_Report_iSeries_Jobs_12hr_Failed_Batch
- NOC_Summary_Network_Security_24hr_Top_src_ip

了解和使用通用信息模型加载项

“通用信息模型加载项”所基于的理念是您可以将大多数日志文件细分为两个部分：

- 字段
- 事件类别标记

有了这两大部分，知识管理器能够规范化搜索时的日志文件，使其遵循类似的方案。“通用信息模型”包含有关 Splunk 软件用来处理大多数 IT 数据的标准字段和事件类别标记的详细信息。

过去，“通用信息模型”在这里显示为一组表格，通过确保为来自于不同来源或供应商的等效事件使用相同的字段名称和事件标记，可使用该表格对数据进行规范化。

如今，“通用信息模型”作为加载项提供，用于将 CIM 表实现为**数据模型**。您可以通过两种方式使用这些数据模型：

- 最初，您可使用它们来测试您的字段和标记是否已进行正确的规范化。
- 在验证您的数据规范化后，您可使用该模型来通过数据透视表生成报表和仪表板面板。

您从此处的 Splunkbase 中下载“通用信息模型加载项”。要深入了解 CIM 加载项的更多概述，请参阅“通用信息模型加载项”的产品文档。

管理知识对象权限

注意：本主题假定您作为知识管理员具有 *管理员* 角色或拥有等效权限集的角色。

身为知识管理员，您可以设置知识对象权限，以限制或扩展对您的 Splunk Enterprise 实现中各种知识对象的访问权限。

在某些情况下，您需要决定某些专门的知识对象只应该供特定应用内具有特定角色的人员使用。而在其他情况下，您将需要进行相反的调整，以使普遍有用的知识对象可供所有应用的所有用户全局使用。与知识管理的所有方面一样，您需要认真考虑这些访问权限限制和扩展的影响。

当某个 Splunk Enterprise 用户首次新建报表、事件类型、交易或类似的知识对象时，对象将只对该用户可用。为使该对象可供更多的人员使用，Splunk Web 提供了以下选项，您必须具有相应的权限才能使用这些选项。您可以：

- 使知识对象可供所有应用的用户全局使用（也称为“提升”对象）。
- 使知识对象可供某一应用的所有用户使用。
- 按用户或角色限制（或扩展）对全局对象或特定于应用的对象访问权限。
- 为角色设置应用级别的读取/写入权限，以使用户能够共享或删除不属于自己的对象。

默认情况下，只有具有 *超级用户* 或 *管理员* 角色的用户可以共享和推广知识对象。这使您以及您同组的知识管理员成为对新知识对象的共享具有批准权限的审核人。

有关扩展设置其他角色权限的功能的更多信息，请参阅下文的子主题[“使管理员和超级用户以外的角色能够设置权限和共享对象”](#)。

权限如何影响知识对象使用情况？

为了说明这些选择如何影响知识对象的使用情况，我们假设 Bob（虚构的网络安全应用用户，具有 *管理员* 级别的“防火墙管理员”角色）将新建一个名为 `firewallbreach` 的事件类型，用于查找指示防火墙违规的事件。以下是可能会产生的一系列权限相关问题，以及操作和结果：

问题	操作	结果
当 Bob 第一次创建 <code>firewallbreach</code> 时，此对象仅对他可用。其他用户无法看到或使用此对象。Bob 决定要将此对象共享给其他网络安全应用用户。	Bob 对 <code>firewallbreach</code> 事件类型的权限进行了更新，使其对网络安全应用的所有用户可用，无论什么角色。他还对这一新的事件类型进行了设置，使所有网络安全应用用户都可以编辑其定义。	所有在网络安全应用上下文中使用 Splunk Enterprise 的用户都可以看到、使用及编辑 <code>firewallbreach</code> 事件类型。同一 Splunk Enterprise 实现内的其他 Splunk 应用的用户不会知道此对象的存在。
不久后，Mary（知识管理员）意识到只有具有防火墙管理员角色的用户才应编辑或更新 <code>firewallbreach</code>	Mary 将编辑此事件类型的权限限制为防火墙管理员角色。	网络安全应用的用户可以在交易、搜索、仪表板等内容中使用 <code>firewallbreach</code> 事件类型，但现在只有具有防火墙管理员角色和具有 <i>管理员</i> 级别权限的人员（如知识管理员）可以编辑此知识对象。在其他

事件类型。		应用上下文中使用 Splunk Enterprise 的人员仍对该事件类型一无所知。
在某一时刻，几个已经习惯了在网络安全应用中使用非常方便的 <code>firewallbreach</code> 事件类型的人决定想在 Windows 应用上下文中也使用该事件类型。	他们将各自的情况报告给了知识管理员，知识管理员立即将 <code>firewallbreach</code> 事件类型提升到全局可用。	现在，所有使用此 Splunk Enterprise 实现的人员都可以使用 <code>firewallbreach</code> 事件类型，而无论是在哪个应用上下文中。但仍然只有 管理员级别用户 以及具有防火墙管理员角色的用户能够更新该事件类型的定义。

权限 - 入门

要更改知识对象的权限，应遵循以下步骤：

1. 在 Splunk Web 中，导航到要更新权限的知识对象类型所对应的页面（例如，搜索和报表或事件类型）。
2. 找到已创建的知识对象（如有必要，使用页面顶部的筛选字段），然后单击其**权限**链接。
3. 在所需知识对象的“权限”页面上，根据更改对象权限的方式，执行以下小节中的操作。



使对象对所有应用的用户可用

要使某对象对 Splunk Enterprise 实现内的所有应用的用户全局可用，请执行以下操作：

1. 导航到该知识对象的“权限”页面（按照上述说明进行）。
2. 在**对象应显示于**下方，选择**所有应用**。
3. 在“权限”部分中，对于**每个人**选择**读取**或**写入**权限：
 - **读取**可以让用户看到并使用该对象，但不能更新其定义。换言之，如果用户对某一特定报表仅具有**读取**权限，他们可以在顶级导航中看到和运行该报表。但是不能更新搜索字符串、更改其时间范围以及保存所做的更改。
 - **写入**可以让用户根据需要查看、使用并更新对象的详细定义。
 - 如果既未选择**读取**，也未选择**写入**，那么用户将无法看到或使用该知识对象。
4. 保存权限更改。

使对象对其应用的所有用户可用

所有知识对象都与某个应用相关。当您创建新知识对象时，它此时与您所在的应用上下文相关。换言之，如果您创建对象时正在使用搜索和报表应用，则此对象将列入“设置”中，**应用**列的值为**搜索和报表**。这意味着，如果您将它的共享权限限制在应用级别，则它仅将对搜索和报告应用的用户可用。

当您创建新对象时，您可以选择保持对象专用、将对象共享给您当前正在使用的应用的用户或将对象全局共享给所有用户。可选择使用应用对“仅此应用”可用，以将其使用限制为此应用的用户（当用户在此应用上下文中时）。

如果您对已经存在的对象具有写入权限，则通过以下步骤，您可以更改该对象的权限，以便它仅对其应用的用户可用。

1. 导航到该知识对象的“权限”页面（按照上述“权限 - 入门”中的说明进行）。
2. 在**对象应显示于**下方，选择**仅此应用**。
3. 在“权限”部分中，对于**每个人**选择**读取**或**写入**权限：
 - **读取**可以让用户看到并使用该对象，但不能更新其定义。换言之，如果用户对某一特定报表仅具有**读取**权限，他们可以在顶级导航中看到和运行该报表。但是不能更新搜索字符串、更改其时间范围以及保存所做的更改。
 - **写入**可以让用户根据需要查看、使用并更新对象的详细定义。
 - 如果既未选择**读取**，也未选择**写入**，那么用户将无法看到或使用该知识对象。
4. 保存权限更改。

移动或复制知识对象

您可能会遇到这样的情况：您希望应用的用户能够访问属于不同应用的特定知识对象，但您又不希望将此对象全局共享给所有应用。您可以通过两种方式来实​​现：复制对象或移动对象。

- **复制** - 创建知识对象的副本。此副本具有和原始对象一样的设置，您可以保留或修改这些设置。您可以将它保留在与正在复制对象的应用相同的应用中，或者您可以将它放入新应用中。如果您将复制的对象添加到原始对象的同一应用中，则将其命名为其他名称。如果您将对象添加到不具同名同类型知识对象的应用中，则您可保留原始名称。您可复制任何对象，即使您的角色对其不具有写入权限。
- **移动** - 将现有的知识对象移动到另一个应用中。将对象从其当前应用中删除，并将其置入想要放在其中的应用中。置入后，您可设置其权限，以使它专用、全局可用或仅对该应用的用户可用。移动应用的功能与决定您是否可删除应用的同一权限有关：仅当您已创建知识对象并对其所属的应用具有写入权限时，您才可移动该知识对象。

注意：通过移动知识对象的应用上下文来切换它，可能会影响与该对象关联的对象的下游。有关更多信息，请参阅本手册中的[“禁用或删除知识对象”](#)。

您可在各种知识对象类型的“设置”页面中找到**复制**和**移动**控制。要复制或移动对象，在其列表中找到该对象，并单击**复制**或**移动**。

按应用和角色限制知识对象访问权限

您可以使用此方法来按特定角色锁定各种知识对象，以防被改动。您可以安排让特定角色的用户可以使用该知识对象，但不能进行更新，也可以进行相关设置，使这些用户根本看不到此对象。在后一种情况中，该对象将不会显示在 Splunk Web 中，即使这些用户对其进行搜索也不会找到任何结果。

如果您希望按角色限制查看或更新知识对象的功能，只需导航到该对象的“权限”页面。如果您希望某角色的成员：

- **能够使用该对象并更新其定义**，应为该角色赋予**读取**和**写入**访问权限。
- **能够使用该对象但不能更新它**，应该只为该角色赋予**读取**访问权限（并确保取消选中**每个人**角色的**写入**权限）。
- **根本无法看到或使用该知识对象**，应取消选中该角色的**读取**和**写入**权限（并同时取消选中**每个人**角色的这两种权限）。

有关 Splunk Enterprise 中基于角色的权限的更多信息，请参阅《安全性手册》中的“关于基于角色的用户访问权限”。

使管理员和超级用户以外的角色能够设置权限和共享对象

如果以默认配置使用 Splunk Enterprise，则只有**超级用户**和**管理员**角色可以设置知识对象的权限。如果您要为其角色赋予设置知识对象权限的功能，应选择**应用**，并单击特定应用的**权限**以转到其“权限”页面。在此页面上，您可以决定哪些角色对该应用所包含的知识对象具有读取或写入访问权限。如果您希望某角色完全能够设置知识对象权限，包括能够在通过 Splunk Web 创建搜索、告警和仪表板时共享这些内容，应将该角色赋予写入访问权限。

设置知识对象权限和共享知识对象的功能是在应用级别控制的，但不能在此级别控制诸如计划搜索或更改默认输入设置之类的其他操作**功能**。

有关使角色在应用内为知识对象设置权限的详细信息，请参阅 Splunk 开发人员门户中的“在 Splunk 应用中为对象设置权限”。

关于删除具有非共享对象的用户和角色

如果某个用户已离开您的团队，您需要将该用户或角色从 Splunk Enterprise 系统中删除，请注意，您将丢失属于这些用户或角色的所有共享状态为**专用**的知识对象。如果要保留这些知识对象，应在删除该用户或角色之前，在应用级别或全局级别共享这些对象。

解决孤立的搜索、报表和告警

当用户离开其工作部门或公司时，他们的帐户会被停用，但他们创建的搜索、报表和告警仍会留存于系统中，从而造成“孤立的搜索”。孤立的搜索是指配置为按计划运行（如计划的报表或告警）但却失去有效拥有者的搜索。

搜索计划程序无法运行孤立的计划搜索。计划程序不知道如何代表一个不存在的拥有者来正确运行此类搜索。此程序不知道孤立搜索拥有者的角色，从而无法得知要对孤立搜索应用什么配置、也不知道该拥有者具有什么搜索配额限制。

孤立的搜索也会造成安全方面的隐患：如果有用户离开了您的公司，而您随后删除了他们的帐号，所有以他们的名义运行的运行中搜索即变为孤立的搜索，必须找出这些搜索并停止运行。

如何解决孤立的搜索

默认情况下，当 Splunk 软件检测到计划的搜索、报表或告警变为孤立时，会发出一个通知，其中列出孤立搜索的名

称。

解决孤立搜索应采取的措施取决于您希望此搜索接下来如何操作。

- 如果您希望此搜索继续按计划运行，即仍作为计划的报表或告警，您可以重新启用其拥有者或者为其指定一个新拥有者。此方式要求您具有访问文件系统的权限。
- 如果您希望此搜索再次运行，但不再作为计划搜索，您可以删除其运行计划（仅适用于已共享给其他用户的孤立搜索）。
- 如果您希望此搜索永远不再运行，您可以完全将其禁用。

重新启用孤立的搜索

如果您希望孤立的搜索继续按计划运行，可以重新启用该搜索。此时，您有两个选择：一是让失效的搜索拥有者再次变为有效；二是将此搜索的拥有者重新指派给另一个有效用户。

让失效的搜索拥有者再次变为有效

在 Splunk 部署中，将失效的搜索拥有者添加为新用户。请参阅《管理员手册》中的“关于用户和角色”。

将搜索的拥有者重新指派给另一个有效用户

将孤立搜索重新指派给另一个有效用户的最稳妥方式是进行一次 REST API 调用。如果您的 Splunk 部署使用搜索头群集，则使用此方式。

以下为 REST API 调用的一个示例。您可使用此示例将孤立搜索重新指派给另一个有效用户。此 REST API 调用考虑了应用上下文，且要求目标搜索用户具有一样的共享等级。

```
curl https://<host>:<mgmt_port>/servicesNS/<user_context>/<app_context>/saved/searches/<entity_name>/acl -d owner=<desired_owner> -d sharing=<sharing_level>
```

或者，您可以手动修改 `local.meta` 和 `save searches.conf` 文件来将孤立搜索进行重新指派。要修改的具体文件取决于此孤立搜索是否已共享给其他用户。

这些方式具有以下限制和注意事项：

- 如果您的 Splunk 部署使用搜索头群集，则不建议使用。
- 这些方式要求您具有 Splunk 部署中文件系统的访问权限（Splunk Cloud 用户没有此权限）。
- 这些方式要求重启 Splunk 部署。

通过编辑 `.meta` 文件重新指派共享的孤立搜索

当用户在应用级或全局将搜索共享给其他用户时，此搜索在应用上下文中进行了共享。

修改保存孤立搜索拥有者信息的 `.meta` 文件。大部分情况下，此文件是 `local.meta`，但 `default.meta` 文件中也可能包含孤立搜索的拥有者信息。

1. 在 Splunk 部署的文件系统中，打开 `etc/apps/<name_of_app>/metadata/local.meta`。
2. 找到孤立搜索的 `save searches` 段落，并用有效的搜索拥有者名称取代 `owner` 值。
3. 重启 Splunk 部署使修改生效。

例如，假设有一个计划的搜索名为 Important Report，而且此搜索的原始拥有者 John Vincent 已将其共享给搜索应用的其他用户。现在，Vincent 离开了公司，所以 Important Report 如今变成了孤立的搜索。您前往 `etc/apps/Search/metadata/local.meta` 并发现：

```
[savedsearches/important%20report]
access = read : [ * ], write : [ admin ]
export = none
owner = jvincent
version = 6.4.0
modtime = 1461111154.871686000
```

有一个有效用户可以作为此搜索的拥有者。她的名字是 Mary Bee。将 Important Report 段落修改为：

```
[savedsearches/important%20report]
access = read : [ * ], write : [ admin ]
export = none
owner = mbee
version = 6.4.0
modtime = 1461111154.871686000
```

保存修改并重启系统。现在，Mary Bee 即为 Important Report 的拥有者。

通过编辑 `savedsearches.conf` 文件重新指派未共享的孤立搜索

如果孤立的报表并未共享给其他用户，则其定义为用户等级且完全位于 `savedsearches.conf` 文件中。

从失效用户的 `savedsearches.conf` 文件中剪下孤立搜索的段落，然后将其粘贴到另一有效用户的 `savedsearches.conf` 文件中。

1. 在 Splunk 部署的文件系统中，打开位于 `etc/users/<name_of_invalid_user>/search/local/savedsearches.conf` 的、失效用户的 `savedsearches.conf` 文件。
2. 找到孤立的计划搜索的段落，并剪下。
3. 保存对此文件所做的修改并将其关闭。
4. 打开位于 `etc/users/<name_of_valid_user>/search/local/savedsearches.conf` 的、有效用户的 `savedsearches.conf` 文件。
5. 将前面剪下的搜索段落粘贴到此有效用户的 `savedsearches.conf` 文件中。
6. 保存对此文件所做的修改并将其关闭。
7. 重启 Splunk 部署使修改生效。

删除孤立搜索的计划

您可以在 **设置 > 搜索、报表和告警** 中从孤立搜索的定义中删除所有计划信息，以此方式删除孤立搜索的计划。这一操作之后，此搜索仍然存在。如果此搜索已共享给应用中的其他用户，则那些用户仍然可以运行此搜索。这点很重要，比如如果此搜索用于一个仪表板。但是，您可能需要确保其他用户不会为此搜索添加计划。要确保这一点，您可以限制具有此搜索编辑权限的角色数。

禁用孤立的搜索

如果您希望永远不再运行孤立的搜索，则可在 **设置 > 搜索、报表和告警** 的列表页面中将其禁用。成功禁用搜索后，此搜索将无法用于任何其他管理操作（如重新启用）。

删除孤立的搜索

如果您有相关权限而且当前和以后都不会再运行某个孤立的计划搜索，则可以通过报表列表页面将其删除。

关闭孤立搜索的通知

默认情况下，Splunk 软件会在发现孤立的搜索时发出通知。如果您不想收到此类通知，打开 `limits.conf`，查找 `[system_checks]` 段落，并将 `orphan_searches` 设置为 `disabled`。

禁用或删除知识对象

删除 Splunk Web 中知识对象的功能取决于诸多因素：

- 您不能删除 Splunk 软件（或应用）所附带的默认知识对象。

如果知识对象定义位于默认目录中，则不能通过 Splunk Web 删除此知识对象。只能禁用该对象（方法是在“设置”中单击该对象的**禁用**）。只能删除存在于应用的本地目录中的对象。

- 您始终可以删除您已经创建但尚未（由您或具有管理级别权限的用户）共享的知识对象。

所创建的知识对象与其他用户共享之后，如果您对该对象所属的应用不具有写入权限，将取消您删除此对象的功能（请参阅下一点）。

- 要删除任意其他知识对象，您的角色必须要具有对于下列项的写入权限：
 - 该知识对象所属的应用。
 - 知识对象本身。

这适用于全局共享的知识对象以及仅共享于某一应用内的知识对象。所有知识对象都属于特定的应用，无论其共享方式如何。

通常，仅向具有管理员等效角色的用户授予应用级别的写入权限。

注意：如果一个角色对于应用不具有写入权限但是对于属于该应用的知识对象具有写入权限，则能够禁用那些知识对象。对于知识对象，单击**禁用**与删除知识对象具有相同的功能，不同之处在于 Splunk 软件不会从系统中删除禁用的知识对象。对于禁用的知识对象，具有写入权限的角色可在任何时候重新启用它。

数据模型有类似的规则。要允许某个角色创建数据模型并与其他用户共享，必须为该角色赋予对应用的写入访问权限。这意味着，能够创建并共享数据模型的用户也可能能够删除知识对象。有关此方面的详细信息，请参阅[管理数据模型](#)（并查找“允许角色创建数据模型”子主题）。

为角色授予应用的写入权限

如果您的角色具有管理级别权限，您可以在 Splunk Web 中为角色授予应用的写入权限。

1. 单击页面顶部的**应用**下拉列表，选择**管理应用**转到“应用”页面。
2. 在“应用”页面上，找到要为其授予写入权限的应用，然后单击**权限**。
3. 在应用的“权限”页面上，为应能够删除应用知识对象的角色选择**写入**。
4. 单击**保存**以保存更改。

您也可通过更新应用的 `local.meta` 文件为应用管理基于角色的权限。有关更多信息，请参阅《确保 Splunk Enterprise 安全》手册中的“设置管理器控制台和应用的访问权限”。

赋予具有应用写入权限的角色删除属于该应用的知识对象的能力

一个角色具有应用的写入权限之后，只要具有该角色的用户也有属于该应用的任何知识对象的写入权限，他们就能够删除那些对象。不管该知识对象是否只在该应用中共享，或是在全局所有应用中共享，用户都可以执行该操作。即使知识对象是全局共享，它们也属于特定的应用。

这个过程假设：

- 您的角色具有管理级别权限。
- 设置了对对象级别权限的角色能够向对象属于的应用中写入数据。

1. 导航到在**设置**中的该知识对象的列表页面。
2. 为了确保正在查看的对象是属于该角色具有写入权限的应用的，请选择应用名称或**应用上下文的全部**。
3. 在列表页面上，找到该角色需要能够删除的知识对象并单击它的**权限**链接。
4. 在知识对象的权限页面上，为该角色选择**写入**。

如果您遵循了这个过程和在此之前的过程，您的角色应该能够删除知识对象了。

删除具有下游依赖性的知识对象

删除具有下游依赖性的知识对象时必须谨慎操作，因为这可能会产生不利影响。

例如，您的某个标记可能看起来好像与另一个更为常见的标记重复。从表面上看，删除重复标记似乎无关紧要。但您可能不知道，此重复标记恰好也是一种非常常用的事件类型所基于的搜索的一部分。而且，此常用事件类型用于两个重要的报表：一个报表是常用仪表板面板的基本报表，另一个用于填充**摘要索引**，以便搜索可以运行若干其他仪表板面板。因此，如果您删除该标记，该事件类型将中断，并且该事件类型的所有下游内容也将中断。

这就是为什么一定要在命名或定义不当的知识对象给您的部署工作造成不利硬性影响之前就把这些对象扼杀在萌芽期的原因。识别特定知识对象的下游依赖性的唯一方式是，对该知识对象进行搜索，弄清它的使用位置，然后搜索相应内容以了解其使用位置（这可能需要进行一些查证）。目前还没有便捷的方法来显示知识对象下游依赖性的列表。

如果您确定必须删除某一知识对象，但不确定您是否已追踪并修复了该对象的所有下游依赖性，您可以首先尝试禁用该对象以查看影响效果。如果一天左右之后似乎没有任何内容出现严重差错，可删除该对象。

在配置文件中删除知识对象

请注意，在 Splunk Web 中，一次只能禁用或删除一个知识对象。如果您需要删除大量对象，则最有效的方式是通过配置文件直接删除知识对象段落。请牢记，您的系统中可能存在某一特定配置文件的若干个版本。在大多数情况中，要进行基于网站范围的本地更改，只应编辑 `$SPLUNK_HOME/etc/system/local/` 中的配置文件；要进行仅应用于特定应用的更改，应编辑 `$SPLUNK_HOME/etc/apps/<App_name>/local/` 中的配置文件。

在尝试编辑配置文件之前，请先阅读并理解《管理员手册》中的下列主题：

- 关于配置文件
- 配置文件优先顺序

搜索时间操作的顺序

当您运行一个搜索时，Splunk 软件会运行多个操作以派生各种知识对象并将他们应用到此搜索返回的事件中。这些知识对象包括提取的字段、已计算字段、查找字段、字段别名、标记和事件类型。

Splunk 软件以特定的顺序执行这些操作。如果您为处理顺序中较为靠前的操作配置了一个定义，但此定义引用了处理顺序中较为靠后的操作的配置结果，则可能会导致问题。

搜索时间操作顺序示例

考虑已计算字段。已计算字段操作位于搜索时间操作顺序的中段。Splunk 软件会在此操作之前先执行一些操作，也

会在此操作之后执行另一些操作。已计算字段通过 `eval` 公式运行已存在于事件内的字段的值，通过此方式来派生新字段。这意味着，对于由搜索时间操作顺序中位于已计算字段操作之后的操作所加入到事件中的字段，已计算字段的公式无法将其包含在内。

例如，当为已计算字段设计 `eval` 表达式时，可以将提取的字段包含在此表达式中，因为字段提取操作位于搜索时间操作顺序中的一开始。当 Splunk 软件开始处理已计算字段时，字段提取已完成，所以已计算字段操作可以成功完成。

但是，已计算字段的 `eval` 表达式不可包含通过查找操作添加的字段。Splunk 软件始终会先执行已计算字段操作，后执行查找操作。所以，当 Splunk 软件开始处理已计算字段时，通过搜索时间查找操作添加的字段还并不存在。如果您的已计算字段 `eval` 表达式包含由查找操作添加的字段，则会收到一个错误消息。

搜索时间操作顺序

下表以列表的形式列出了搜索时间操作的顺序。列表下面提供了更多关于此操作中各项操作的相关信息。

每项操作可以包含一些配置，这些配置引用由此顺序中位于前面的操作所派生的字段。但是，同样的配置不能包含由此顺序中位于后面的操作所派生的字段。

除了其中一项操作之外的其他所有操作都可通过 Splunk Web 进行配置，尽管有些配置选项只能通过手动 `.conf` 编辑实现。所有基于文件的手动操作配置都应该在搜索头层级进行。

此列表不包含索引时间操作，如默认和索引字段提取。索引时间操作会优先于所有搜索时间操作。

请参阅《管理索引器和索引器群集》手册中的“索引时间对比搜索时间”。

搜索时间操作顺序	操作名称	可通过 Splunk Web 配置？	文件配置位置
第一	内联字段提取（无字段转换）	是	<code>props.conf</code> 段落中的 <code>EXTRACT-<class></code>
第二	采用字段转换的字段提取	是	<code>props.conf</code> 段落中的 <code>REPORT-<class></code>
第三	自动键值字段提取	否	<code>props.conf</code> 段落，其中 <code>KV_MODE</code> 设置为 <code>none</code> 之外的有效值。如果没有为段落指定 <code>KV_MODE</code> 值，则默认会设置为 <code>auto</code> 。
第四	字段别名	是	<code>props.conf</code> 段落中的 <code>FIELDALIAS-<class></code>
第五	已计算字段	是	<code>props.conf</code> 段落中的 <code>EVAL-<fieldname></code>
第六	查找	是	<code>props.conf</code> 段落中的 <code>LOOKUP-<class></code> ，引用 <code>transforms.conf</code> 查找段落。
第七	事件类型	是	<code>eventtypes.conf</code> 段落
第八	标记	是	<code>tags.conf</code> 段落

内联字段提取

内联字段提取指不包含字段转换引用的显式字段提取。显式字段提取指配置为提取特定字段或字段集的字段提取。

每个内联字段提取配置都针对属于特定主机、来源或来源类型的事件。

此操作不包含自动键值字段提取。自动键值字段提取有自己单独的操作类别。

Splunk Web 管理

在“设置”中创建和管理内联字段提取。导航到 **设置 > 字段 > 字段提取**。您也可以使用字段提取器实用工具来设计内联字段提取。

配置信息

在 `props.conf` 段落中创建 `EXTRACT-<class>` 配置。

限制

Splunk 软件根据内联字段提取的 `<class>` 值，以字母数字顺序处理所有属于特定主机、来源或来源类型的内联字段提取。这意味着在 `EXTRACT-aaa` 的字段提取定义中，您无法引用由 `EXTRACT-ddd` 提取的字段。请参阅 [以字母数字顺序处理字段提取配置](#)。

由于他们位于搜索时间操作顺序的顶端，内联字段提取配置无法引用由其他搜索时间操作所派生或添加的字段。

相关信息

在本手册中：

- [使用字段提取器构建字段提取](#) - 您可使用字段提取器在 Splunk Web 中创建内联字段提取。这种方式并不要求您熟悉如何编写正则表达式。
- [使用“字段提取”页面](#) - 通过“设置”中的“字段提取”页面在 Splunk Web 中创建内联字段提取。
- [通过配置文件创建和维护搜索时间字段提取](#) - 在 `props.conf` 中配置内联字段提取。

采用字段转换的字段提取

Splunk 软件始终会在处理完内联字段提取后即开始处理引用了字段转换的字段提取配置。和内联字段提取一样，每个引用了转换的字段提取都显式配置为提取特定字段或字段集。

每个引用了转换的字段提取配置都针对属于特定主机、来源或来源类型的事件。

此操作不包含自动键值字段提取。自动键值字段提取有自己单独的操作类别。

Splunk Web 管理

您可以在“设置”中创建和管理采用字段转换的字段提取。导航到 **设置 > 字段**，并通过“字段提取”和“字段转换”页面设置字段提取。

配置信息

在 `props.conf` 段落中创建 `REPORT-<class>` 配置。`REPORT-<class>` 配置引用了 `transforms.conf` 中的一个额外配置。

限制

Splunk 软件根据引用了转换的字段提取的 `<class>` 值，以字母数字顺序处理所有属于特定主机、来源或来源类型的此类字段提取。这意味着在 `REPORT-aaa` 的字段提取定义中，您无法引用由 `REPORT-ddd` 提取的字段。请参阅 [以字母数字顺序处理字段提取配置](#)。

引用了转换的字段提取配置可引用通过内联字段提取操作提取的字段，但不能引用由自动键值字段提取所派生和添加的字段，或其他在搜索时间操作顺序中排在较后面的操作所派生和添加的字段。

相关信息

在本手册中：

- [使用“字段转换”页面](#) - 创建引用了转换的搜索时间字段提取的 `transforms.conf` 部分。
- [使用“字段提取”页面](#) - 创建引用了转换的搜索时间字段提取的 `props.conf` 部分。
- [通过配置文件创建和维护搜索时间字段提取](#) - 在 `transforms.conf` 和 `props.conf` 中配置引用了转换的字段提取。

自动键值字段提取

一种字段提醒配置，使用了 `KV_MODE` 属性以自动为特定主机、来源或来源类型的事件提取字段。

自动键值字段提取是非显式的，您无法配置此类提取查找特定字段或字段集。此类提取查找的是其可找到的所有事件中的键=值模式，然后将他们提取为字段/值对。可以将其配置为从结构化数据格式（如 JSON、CSV 和表格形式的事件）中提取字段。

自动键值字段提取始终在显式字段提取方式（内联字段提取和引用了转换的字段提取）完成之后才会运行。

Splunk Web 管理

目前在 Splunk Web 中尚无法配置自动键值字段提取。

配置信息

在 `props.conf` 中找到（或创建）相应段落，并将 `KV_MODE` 属性设置为 `auto`、`auto_escaped`、`multi`、`json` 或 `xml`，通过这种方式为特定主机、来源或来源类型设置自动键值字段提取。

当 `props.conf` 段落没有设置 `KV_MODE` 的值时，此段落默认采用 `KV_MODE=auto`。必须设置 `KV_MODE=none` 才能针对特定主机、来源或来源类型禁用自动键值字段提取。自动键值字段提取禁用之后，显式字段提取仍会运行。

当 `KV_MODE` 设置为 `auto` 或 `auto_escaped` 时，自动 JSON 字段提取可能会与其他自动键/值字段提取一起运行。如果您想禁用 JSON 字段提取但又不想更改当前 `KV_MODE` 的设置值 `auto`，则可以添加 `AUTO_KV_JSON=false` 段落。未设置时，`AUTO_KV_JSON` 采用默认值 `true`。

限制

Splunk 软件以在事件找到自动键值字段提取的先后顺序对他们进行处理。

相关信息

在本手册中：

- [通过配置文件创建和维护搜索时间字段提取](#) - 在 `.conf` 文件中配置自动键值字段提取。

字段别名

字段别名是字段别名配置的应用，使您可以在一个搜索中以多种名称或别名来引用同一个字段。

每个字段别名配置都针对属于特定主机、来源或来源类型的事件。

Splunk Web 管理

在“设置”中创建和管理字段别名。导航到 **设置 > 字段 > 字段别名**。

配置信息

在 `props.conf` 段落中创建 `FIELDALIAS-<class>` 配置。

限制

Splunk 软件以字母数字顺序处理属于特定主机、来源或来源类型的字段别名。

可为在索引时间或搜索时间提取的字段创建别名。对于搜索时间操作顺序中位于字段别名操作之后的操作（如查找或已计算字段），由这些操作添加到事件中的字段则无法创建别名。

相关信息

- [在 Splunk Web 中创建字段别名](#)
- [通过 props.conf 创建字段别名](#)

已计算字段

通过 `eval` 表达式的计算创建一个或多个字段并将这些字段添加到事件中的配置。由于索引时间或搜索时间字段提取操作，`eval` 表达式可使用事件中已存在的字段值。

每个已计算字段配置都针对属于特定主机、来源或来源类型的事件。

Splunk Web 管理

您可以在“设置”中创建和管理已计算字段。导航到 **设置 > 字段 > 已计算字段**。

配置信息

在 `props.conf` 段落中添加 `EVAL-<fieldname>` 配置，从而实现已计算字段的创建。

限制

单个 `props.conf` 段落内的所有 `EVAL-<fieldname>` 配置都并行排列，而不是以特定顺序排列。这意味着您不能将各已计算字段表达式“链接”在一起，因为一个已计算字段的求值结果将用于另一个已计算字段的表达式。

已计算字段可引用所有类型的字段提取和字段别名，但不能引用查找、事件类型或标记。

相关信息

在本手册中：

- [关于已计算字段](#)
- [通过 Splunk Web 创建已计算字段](#)
- [通过 props.conf 配置已计算字段](#)

查找

当查找表字段与事件中的一个或多个字段匹配时将查找表字段添加到事件中的配置。查找配置有四种不同的类型：CSV 查找、外部查找、KV 存储查找和地理空间查找。

每个查找配置都针对属于特定主机、来源或来源类型的事件。

Splunk Web 管理

在“设置”中创建和管理您的查找。导航到 **设置 > 查找**。

配置信息

在 `props.conf` 中创建 `LOOKUP-<class>` 配置，以此方式定义会自动将字段添加到搜索结果事件中的查找。每个 `LOOKUP-<class>` 都包含一个引用至 `transforms.conf` 中的 `[<lookup_name>]` 段落。

限制

Splunk 软件以字母数字顺序处理属于特定主机、来源或来源类型的查找。

查找配置可引用由字段提取、字段别名及已计算字段添加到事件中的字段，但不能引用事件类型和标记。

相关信息

在本手册中：

- [使用字段查找将信息添加到事件中](#)
- [查找配置简介](#)

事件类型

将事件类型字段/值对添加到与定义该事件类型的搜索字符串匹配的事件中的配置。

Splunk Web 管理

运行搜索后，将该搜索保存为事件类型。也可以在“设置”中定义和维护事件类型。导航到**设置 > 事件类型**。

配置信息

在 `eventtypes.conf` 段落中配置事件类型。

限制

Splunk 软件先按优先级分数后按字母数字顺序处理事件类型。因此，软件会先处理**优先级**为 **1** 的所有事件类型，然后按字母数字顺序将他们应用到事件中。然后会处理**优先级**为 **2** 的事件类型，依此类推。

定义事件类型的搜索字符串无法引用标记。事件类型始终会先于标记被处理和添加到事件中。

相关信息

在本手册中：

- [在 Splunk Web 中定义和维护事件类型](#)
- [自动查找和构建事件类型](#)
- [直接在 eventtypes.conf 中配置事件类型](#)

标记

添加标记到事件中特定字段/值对的配置。

Splunk Web 管理

您可以直接添加标记至搜索结果的字段/值对中。也可以在“设置”中定义和维护标记。导航到**设置 > 标记**。

配置信息

在 `tags.conf` 段落中配置标记。

限制

Splunk 软件以字母数字顺序将标记应用到字段/值对中，先按字段值，再按字段名称。

您可以将标记应用到事件中的所有字段/值对，不管是在索引时间提取的、在搜索时间提取的、还是通过其他方式添加的，如事件类型、查找或已计算字段。

相关信息

在本手册中：

- [在“搜索”中为字段值对设置标记](#)
- [在“设置”中定义和管理标记](#)

以字母数字顺序处理知识对象配置

Splunk 软件基于知识对象所属的主机、来源或来源类型，以字母数字顺序处理以下知识对象：

- 内联字段提取
- 采用字段转换的字段提取
- 字段别名
- 事件类型（在按优先级进行保存之后）
- 查找

Splunk 软件以字母数字顺序处理标记，但他们并未与特定主机、来源或来源类型相关联。

示例

例如，Splunk 软件以字母数字顺序提取内联字段提取至特定主机、来源或来源类型的字段别名。这意味着，当软件处理属于 `access_combined_wcookies` 来源类型的内联字段提取时，会在 `REPORT-bbb` 之前先处理名为 `REPORT-aaa` 的提取，在 `REPORT-ddd` 之前先处理 `REPORT-bbb`，依此类推。

这意味着在 `REPORT-aaa` 的字段提取定义中，您无法引用由 `REPORT-ddd` 提取的字段。

例如，此配置没有作用，因为根据字段提取处理顺序 (`zzz < aaa`)，会先提取 `first_two` 字段，后提取 `first_ten` 字段。

```
[splunkd]
EXTRACT-zzz = ^(?<first_ten>.{10})
EXTRACT-aaa = (?<first_two>.{2}) in first_ten
```

此配置会起作用，因为根据字段提取处理顺序 (`aaa > mmm`)，会先提取 `first_ten` 字段，后提取 `first_two` 字段

```
[mongod]
EXTRACT-aaa = ^(?<first_ten>.{10})
EXTRACT-mmm = (?<first_two>.{2}) in first_ten
```

您可使用以下搜索确认这些配置问题。

```
index=_internal (sourcetype=splunkd OR sourcetype=mongod) | stats values(first_ten) values(first_two) by sourcetype
```

数据解释方式：字段和字段提取

关于字段

显示在事件数据中的字段为可搜索名称/值对，例如 `user_name=fred` 或 `ip_address=192.168.1.1`。他们是 Splunk Enterprise 中的搜索、报表和数据模型的构建块。当您在事件数据中运行搜索时，Splunk Enterprise 查找数据中的字段。

注意：字段名称经常称为键。首字母缩写 `kv` 是 `key/value`（键/值）的简称。

让我们看一下下面的搜索示例。

```
status=404
```

这个搜索要查找的是带 `status` 字段且字段值为 `404` 的事件。当运行这个搜索时，Splunk Enterprise 不会查找 `status` 字段值为其他值的事件，也不会查找其他字段的值为 `404` 的事件。因此，与在搜索字符串中使用 `404` 的搜索结果相比，此搜索会返回一组更具体的搜索结果。

字段在事件中通常为 `key=value` 对的形式显示，例如 `user_name=Fred`。但在许多事件中，字段值会显示在没有标识键的固定的、带分隔符的位置。例如，可能在有些事件中，`user_name` 值始终单独显示在时间戳和 `user_id` 值的后面。

```
Nov 15 09:32:22 00224 johnz
Nov 15 09:39:12 01671 dmehta
Nov 15 09:45:23 00043 sting
Nov 15 10:02:54 00676 lscott
```

Splunk Enterprise 能够使用自定义字段提取来标识这些字段。

关于字段提取

在 Splunk Enterprise 处理事件数据时，它会从中提取字段。此过程称为**字段提取**。

Splunk Enterprise 自动提取一些字段

Splunk Enterprise 可以在没有帮助的情况下从事件中提取一些字段。当为传入事件创建索引时，自动提取 `host`、`source` 和 `sourcetype` 值、时间戳和若干个其他的**默认字段**。

也会提取在事件数据中以 `key=value` 对形式显示的字段。识别并提取 `k/v` 对的过程称为**字段发现**。您可以禁用字段发现来改进搜索性能。

当显示在事件中的字段没有键时，Splunk Enterprise 使用称为正则表达式的模式匹配规则来提取那些字段以作为

完整的 k/v 对。通过正确配置的正则表达式，Splunk Enterprise 能够从之前的示例事件中提取 `user_id=johnz`。Splunk Enterprise 附带若干个字段提取配置，能够使用正则表达式来从事件数据中标识并提取字段。

有关字段发现和自动字段提取示例的更多信息，请参阅本手册中的[“当 Splunk Enterprise 提取字段”](#)。

有关 Splunk Enterprise 如何使用正则表达式提取字段的更多信息，请参阅本手册中的[“关于 Splunk Enterprise 正则表达式”](#)。

为获得数据中的所有字段，创建自定义字段提取

为了使用 Splunk Enterprise 搜索的强大功能，创建其他的字段提取。自定义字段提取允许您捕获和跟踪对您很重要、但 Splunk Enterprise 未自动发现并提取的信息。您提供的任何字段提取配置必须包括一个能告诉 Splunk Enterprise 如何找到您想提取的字段的正则表达式。

所有字段提取，包括自定义字段提取，均与特定的 `source`、`sourcetype` 或 `host` 值相关联。例如，如果要创建一个 `ip` 字段提取，可能需要将 `ip` 的提取配置与 `sourcetype=access_combined` 关联。

自定义字段提取应该在搜索时间发生，但在极少数的情况下，您可安排一些自定义字段提取在索引时间发生。请参阅本手册中的[“当 Splunk Enterprise 提取字段”](#)。

在创建自定义字段提取前，需要了解您的数据

在开始创建字段提取之前，确保您熟悉所处理的与 `source`、`sourcetype` 或 `host` 关联的事件数据的格式和模式。一种方式是调查带有模式选项卡的数据的主要事件模式。请参阅《搜索手册》中的[“使用模式选项卡识别事件模式”](#)。

以下是来自同一来源类型，即 `apache` 服务器 `web` 访问日志的两个事件。

```
131.253.24.135 - - [03/Jun/2014:20:49:53 -0700] "GET /wp-content/themes/aurora/style.css HTTP/1.1" 200 7464
"http://www.splunk.com/download" "Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.0; Trident/5.0; Trident/5.0)"

10.1.10.14 - - [03/Jun/2014:20:49:33 -0700] "GET / HTTP/1.1" 200 75017 "-" "Mozilla/5.0 (compatible; Nmap Scripting
Engine; http://nmap.org/book/nse.html)"
```

虽然这些事件包含不同的字符串和字符，它们仍是以一致的方式格式化。它们都以一个可靠的顺序显示字段值，例如 `clientIP`、`status`、`bytes`、`method` 等。

可靠意味着 `method` 值后面始终跟着 `URI` 值，`URI` 值后面始终跟着 `status` 值，`status` 值后面始终跟着 `bytes` 值，依此类推。当事件具有一致可靠的格式时，可以创建一个能准确从其中捕获多个字段值的字段提取。

为方便对照，请查看这组 Cisco ASA 防火墙日志事件：

1	Jul 15 20:10:27 10.11.36.31 %ASA-6-113003: AAA group policy for user AmorAubrey is being set to Acme_techoutbound
2	Jul 15 20:12:42 10.11.36.11 %ASA-7-710006: IGMP request discarded from 10.11.36.36 to outside:87.194.216.51
3	Jul 15 20:13:52 10.11.36.28 %ASA-6-302014: Teardown TCP connection 517934 for Outside:128.241.220.82/1561 to Inside:10.123.124.28/8443 duration 0:05:02 bytes 297 Tunnel has been torn down (AMOSORTILEGIO)
4	Apr 19 11:24:32 PROD-MFS-002 %ASA-4-106103: access-list fmVPN-1300 denied udp for user 'sdewilde7' outside/12.130.60.4(137) -> inside1/10.157.200.154(137) hit-cnt 1 first hit [0x286364c7, 0x0] "

虽然这些事件包含始终以空格分隔的字段值，它们并不像前面的两个事件一样共享一个可靠的格式。这些事件依次表示：

1. 组策略更改
2. IGMP 请求
3. TCP 连接
4. 防火墙拒绝来自特定 IP 的访问请求

因为这些事件差异巨大，所以如果要创建单个字段提取，用于每个这些事件模式和提取相关字段数值，则非常困难。

在这种情况下，当特定主机、来源类型或来源包含多个事件模式时，您可能想要定义匹配每个模式的字段提取，而不是设计单个的能适用于所有模式的提取。检查事件来确定对于每个模式都通用和可靠的文本。

在字段提取中使用必填文本

在上面的四个事件中，跟在 `%ASA-#-` 后的字符串数字有特定的意义。您可在 Cisco 文档中找到它们的定义。当在您的数据中有如此的唯一事件标识符时，在字段提取中指定它们为必填文本。必填文本字符串限制了能匹配字段提取中

的正则表达式的事件。

指定必填文本是可选的，但它提供了多种好处。因为必填文本减少了扫描的事件集，提高了字段提取效率和降低了假阳性字段提取的数目。

“字段提取器”实用工具允许您在示例事件中突出显示文本并指定其为必填文本。

Splunk Enterprise 中的自定义字段提取的方法

作为知识管理器，您需要监视 Splunk Enterprise 实现的用户所创建的自定义字段提取集，您也可能会自行定义一组专门的自定义字段提取。实现此目的的方式包括：

- 能生成用于字段提取的正则表达式的“字段提取器”实用工具。
- 通过“设置”中的页面添加字段提取。必须提供正则表达式。
- 在 `.conf` 文件级别手动添加字段提取配置。提供最灵活的字段提取。

下面的部分介绍了对于 Splunk Enterprise 用户可用的字段提取方法。所有这些方法允许您创建搜索时间字段提取。要创建索引时间字段提取，选择第三个选项：直接在配置文件中配置字段提取。

使用字段提取器为您构建提取

字段提取器实用工具引导您逐步完成字段提取设计过程。它提供了字段提取的两种方法：正则表达式和基于分隔符的字段提取。对于从非结构化事件数据中提取字段，其中事件可能遵循多种不同的事件模式，正则表达式的方法会很有用。如果您不熟悉正则表达式语法和用法，这也会很有用，因为它会生成正则表达式并允许您验证它们。

基于分隔符的字段提取方法适用于结构化事件数据。结构化事件数据来自如 SQL 数据库和 CSV 文件一类的源，并且所产生的事件中，所有字段都由一个共同的分隔符（例如逗号、空格或管道符号）隔开。对于来自共同源的结构化数据事件，通常不需要使用正则表达式。

使用字段提取器的正则表达式方法，您可以：

- 通过选择示例事件和突出显示从事件中提取的字段来设置字段提取。
- 创建能捕获多个字段的单个提取。
- 通过检测和删除假阳性匹配来提高提取的准确度。
- 使用搜索过滤器验证提取结果来确保提取出特定值。
- 指定仅能从具有必填文本特定字符串的事件中提取的字段。
- 查看通过提取发现的字段值的 stats 表。
- 自行手动配置正则表达式用于字段表达式。

使用字段提取器的分隔符方法，您可以：

- 确定分隔符以提取事件中的所有字段。
- 适当地重命名特定字段。
- 验证提取结果。

字段提取器只能构建与数据中的特定来源或来源类型（非主机）关联的**搜索时间**字段提取。

有关使用字段提取器的更多信息，请参见本手册中的[“使用字段提取器构建字段提取”](#)。

通过“字段提取”和“字段转换”页面定义字段提取

您可以使用“设置”中的“字段提取”和“字段转换”页面来定义和维护 Splunk Web 中复杂的提取字段。

字段提取创建的方法允许您创建比用字段提取器实用工具生成的更广泛的字段提取。您需要具备以下知识。

- 了解如何设计正则表达式。
- 基本了解如何在 `props.conf` 和 `transforms.conf` 中配置字段提取。

如果您想创建能从 `_raw` 中提取字段并且不需要字段转换的自定义字段提取，可以使用字段提取器实用工具。字段提取器能够生成正则表达式，并且对于您所定义的字段提取的准确度给予反馈信息。

使用“字段提取”页面创建基本字段提取，或与“字段转换”页面结合使用，定义执行下面操作的字段提取配置。

- 在多个来源、来源类型或主机之间重复使用相同的正则表达式。
- 对同一个来源、来源类型或主机应用多个正则表达式。
- 使用正则表达式从其他字段值中提取字段。

“字段提取”和“字段转换”页面只定义**搜索时间**字段提取。

请参阅本手册中的以下主题：

- [使用 Splunk Web 的字段提取页面](#)
- [使用 Splunk Web 的字段转换页面](#)

直接在配置文件中配置字段提取

为了获得对于字段提取的完全控制，可直接向 `props.conf` 和 `transforms.conf` 中添加配置。使用这种方法创建的字段提取的功能超出了使用 Splunk Web 方法例如字段提取器实用工具或“设置”页面创建的字段提取的功能。例如，可以使用配置文件来设置以下内容：

- 基于分隔符的字段提取。
- 多值字段提取。
- 名称以数字或下划线开头的字段提取。如果未禁用键清理功能，通常不允许执行此类操作。
- 所提取字段的格式设置。

请参阅本手册中的[“通过配置文件创建和维护搜索时间提取”](#)。

只能通过在 `props.conf` 和 `transforms.conf` 中配置索引时间字段提取的方式来创建此类提取。向默认索引字段集中添加字段提取会导致搜索性能和索引问题。但是如果必须创建其他的索引时间字段提取，请参阅《数据导入》手册中的“在索引时间创建自定义字段”。

创建自定义已计算字段和多值字段

通过 `.conf` 文件可以持续配置两种自定义字段：已计算字段和多值字段。

多值字段在单个事件中可多次出现，每次均具有不同值。要配置自定义多值字段，更改 `fields.conf` 以及 `props.conf`。请参阅本手册的[“配置多值字段”](#)。

已计算字段提供了通过 `eval` 表达式从事件中显示的其他字段值计算出的值。在 `props.conf` 中配置它们。请参阅本手册的[“关于已计算字段”](#)。

在搜索字符串中构建字段提取

可通过以下搜索命令以不同方式提取搜索时间字段：

- `rex`
- `extract`
- `multikv`
- `spath`
- `xmlkv`
- `xpath`
- `kvform`

请参阅《搜索手册》中的“使用搜索命令提取字段”。或者您可以在[搜索参考](#)中查找每条搜索命令。

由搜索命令进行的字段提取仅适用于使用那些命令进行的搜索的返回结果。不能使用这些搜索命令创建在搜索结束后仍保留的可重复使用的提取。为此，使用字段提取器实用工具，通过“设置”页面配置提取，或在 `.conf` 文件中直接设置配置。

当 Splunk Enterprise 提取字段时

Splunk Enterprise 首先在**索引时间**内提取字段，其次在**搜索时间**内提取字段。在运行搜索后，此搜索提取的字段会列在字段边栏中。

索引时间的字段提取

在索引时间中，Splunk 软件会为每个事件提取一小组**默认字段**，包括 `host`、`source` 和 `sourcetype`。默认字段通用于所有事件。请参阅[使用默认字段](#)。

Splunk 软件也能在索引时间提取自定义**索引字段**。这些字段已针对索引时间提取进行明确配置。

警告：不要将自定义字段添加到 Splunk 软件在**索引时间**内提取并建立索引的一组默认字段中。添加此字段列表会降低索引性能和减慢搜索时间，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，因为无论何时更改索引字段集，都必须为整个数据集重新创建索引。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”。

搜索时间的字段提取

在搜索时间中，Splunk 软件可以提取其他字段，具体取决于其**搜索模式**设置，以及对于所运行搜索的类型该设置是否启用**字段发现**。

如果启用字段发现，Splunk 软件会执行以下操作：

- 确定并提取其在事件数据中找到的与指定 `key=value` 对匹配的前 50 个字段。这 50 个字段的限制是默认值。如果您有 Splunk Enterprise，可以通过编辑 `limits.conf` 中的 `[kv]` 段落来修改此限值。
- 提取搜索中明确指出的所有字段，它可能会通过自动提取找到这些字段，但不是已在确定的前 50 个字段范围

- 内。
- 通过字段提取器、“设置”中的“提取的字段”页面、配置文件编辑或 `rex` 等搜索命令执行您所定义的自定义字段提取。

如果禁用字段发现，Splunk 软件会提取：

- 搜索中明确指出的所有字段。
- 上文所述的默认和索引字段。
- 其 `CAN_OPTIMIZE` 参数**设为 true**（位于 `transforms.conf` 中）的所有自定义字段提取。

只有在以下情况下，Splunk 软件会发现除默认字段和在搜索字符串中明确指出的字段以外的其他字段：

- 在 **智能**搜索模式下运行**非转换**搜索。
- 在 **详细**搜索模式下运行任何搜索。

请参阅 *搜索手册* 中的“设置搜索模式以调整搜索体验”。

有关搜索时间和索引时间的说明，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”。

自动字段提取示例

下面是一个关于在没有用户帮助的情况下 Splunk 软件如何自动提取字段的示例，与自定义字段提取相反，因为它遵循您所定义的事件提取规则。

假设您要搜索 `sourcetype`，即 Splunk 软件会在索引时间为每个事件提取的一个默认字段。如果您的搜索为

```
sourcetype=veeblefetzter
```

（时间范围为过去 24 小时），Splunk 软件将返回该时间范围内的 `sourcetype` 为 `veeblefetzter` 的所有事件。Splunk 软件会从该组事件中提取其可自行确定的前 50 个字段。然后根据配置文件执行自定义字段提取。当搜索完成时，所有这些字段会显示在字段边栏中。

现在，如果搜索某一名称/值组合（如 `userlogin=fail`）时第一次显示 25,000 个事件，并且 `userlogin` 不在预先配置的自定义字段集内，则它可能不是 Splunk 软件自行找到的前 50 个字段之一。

但是，如果您将搜索更改为

```
sourcetype=veeblefetzter userlogin=*
```

那么 Splunk 软件可以找到并返回包含 `userlogin` 字段且 `sourcetype` 值为 `veeblefetzter` 的所有事件。该字段将与此搜索提取的其他字段一起显示在字段边栏中。

使用字段提取器构建字段提取

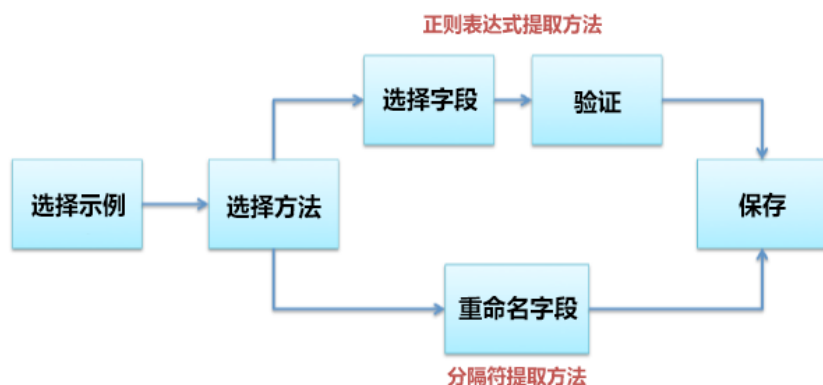
使用**字段提取器**实用工具创建新字段。字段提取器提供两种字段提取的方法：正则表达式和分隔符。

正则表达式方法最适用于非结构化事件数据。您选择一个示例事件，并突出显示要从该事件中提取的一个或多个字段，并且字段提取器会生成一个匹配您数据集中类似事件的正则表达式，并从这些事件中提取字段。正则表达式方法为测试和优化正则表达式的准确度提供了若干个工具。它还允许您手动编辑正则表达式。

分隔符方法设计用于结构化事件数据：数据来自具有标头的文件，在这些文件中事件的所有字段都是由一个共同的分隔符隔开的，例如逗号或空格。您选择一个示例事件，确定分隔符，然后重命名字段提取器查找的字段、驻留在具有标头的文件中的数据并由特定字符分隔的字段

字段提取器概述

为帮助您创建新字段，字段提取器会带您了解这一系列步骤。字段提取器工作流分散在“选择方法”步骤中，该步中您会选择想要使用的字段提取方法。



此表提供了所需步骤的概述。有关步骤的详细信息，请单击**步骤名称**一栏中的链接。

步骤名称	描述	字段提取方法
选择示例	选择与事件（具有您想要提取的一个或多个字段）相关联的 来源类型 或 来源 。然后选择一个具有该字段或这些字段的示例事件。	两者皆可
选择方法	选择一种字段提取方法。您可以使用字段提取器生成字段提取正则表达式，或者采用基于分隔符的字段提取。您所做的选择取决于您是否正尝试从非结构化或结构化的事件数据中提取字段。	两者皆可
选择字段	突出显示事件中的一个或多个字段值，以确定它们作为字段。字段提取器会生成一个与事件匹配的正则表达式，并提取字段。（可选）您可以： <ul style="list-style-type: none"> 提供其他的示例事件以提高提取准确度。 确定必填文本使字段提取专注于包含该文本的事件。 检查字段提取结果。 手动更新底层正则表达式。 	正则表达式
重命名字段	确定将事件中所有字段分隔开的分隔符，然后重命名这些字段中的一个或多个字段。	分隔符
验证字段	<ul style="list-style-type: none"> 检查字段提取结果。 确定不正确的提取字段作为反例以提高字段提取的准确度。 	正则表达式
保存	给新的字段提取命名，设置权限，并保存。	两者皆可

访问字段提取器

有多种方法可以访问字段提取器实用工具。您使用的访问方法决定了您从字段提取器工作流的哪一步开始。

在运行返回事件的搜索后，所有用户都能访问字段提取器。对于字段提取器，有三个搜索后入口点：

- 字段边栏的底部
- “全部字段”对话框
- 搜索结果中的任何事件

您也可以访问字段提取器：

- 通过“设置”的“字段提取”页面。
- 当添加具有固定来源类型的数据时。
- 通过 Splunk Web 主页（如果您拥有“管理员”角色权限）。

从字段边栏的底部访问字段提取器

当您使用此方法访问字段提取器时，它仅在所运行搜索返回的事件集上进行。要获得您的 Splunk 部署中的全部来源类型，[可使用“设置”中的“字段提取”页面](#)。

1. 运行一个返回事件的搜索。
2. 向下滚动至字段边栏的底部，然后单击**提取新字段**。

字段提取器在“[选择示例](#)”这步启动。

# timestartpos 7	>	14/06/17
a uri 100+		18:20:55.000
a uri_path 14		
a uri_query 100+		
a user 1		
a useragent 26	>	14/06/17
# version 1		18:20:55.000
	>	14/06/17
		18:20:55.000

还有 4 个字段
提取新字段

从“全部字段”对话框访问字段提取器

当您使用该方法访问字段提取器时，您只能从搜索返回的数据中提取字段。要获得您的 Splunk 部署中的全部来源类型，可使用“[设置](#)”中的“[字段提取](#)”页面。

1. 运行一个返回事件的搜索。
2. 在字段边栏的顶部，单击**全部字段**。
3. 在“全部字段”对话框中，单击**提取新字段**。

字段提取器在“[选择示例](#)”步骤启动。

值的数目	事件覆盖范围	类型
3	100%	字符串
3	100%	字符串
1	100%	字符串
>100	100%	字符串

从特定事件访问字段提取器

使用此方法选择搜索结果中的事件，并创建字段提取：

- 提取事件中找到一个或多个字段。
- 与事件的来源类型相关联。

当您使用此方法访问字段提取器时，字段提取器在所运行搜索返回的事件集上进行。

1. 运行一个返回事件的搜索。
2. 找到您想提取字段的事件，单击时间戳左侧的箭头符号打开它。
3. 单击**事件操作**，然后选择**提取字段**。

字段提取器在新的浏览器选项卡中，在“[选择方法](#)”步骤中启动。您已经定义了来源类型和示例事件。

i	时间	事件
✓	14/06/17 18:22:16.000	91.205.189.15 - - [17/Jun/2014:18:22:16] "GET /oldlin L7FF7ADFF53113 HTTP 1.1" 200 1665 "http://www.butterc 14" "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/ e/19.0.1084.46 Safari/536.5" 159
		事件操作
		构建事件类型
		提取字段
		显示来源
		个值
		www2
		tutorialdata.zip:/www2/access.log
		access_combined_wcookie
		JSESSIONID
		SD6SL7FF7ADFF53113
		bytes
		1665

通过“设置”的“字段提取”页面访问字段提取器

该访问方法对所有用户可用。

1. 选择 **设置 > 字段 > 字段提取**。

2. 单击 **打开字段提取器** 按钮。

字段提取器在“[选择示例](#)”步骤启动。

通过主页访问字段提取器

该访问方法仅对其角色具备 `edit_monitor` 操作的用户可用，例如“管理员”。

在主页上，单击 **添加数据** 图标下方的 **提取字段** 链接。

字段提取器在“[选择示例](#)”步骤启动。

浏览 Splunk Enterprise



在添加数据后访问字段提取器

该访问方法仅对其角色具备 `edit_monitor` 操作的用户可用，例如“管理员”。

在向 Splunk Enterprise 中添加数据后，只要有固定的来源类型，就可以使用字段提取器从数据中提取字段。

例如：向 Splunk 部署中添加名为 `vendors.csv` 的文件并赋予其自定义来源类型 `vendors`。在保存此输入后，您可以进入字段提取器并从与 `vendors` 来源类型关联的事件中提取字段。

另一个示例：为 `/var/log` 目录创建监视器输入，并将来源类型设置为 **自动**，则 Splunk 软件会逐个事件地自动确定来自此输入的数据的来源类型值。当保存此导入时不会提示您从新数据导入中提取字段，因为在目录上建立索引的事件可以有各种来源类型值。

1. 进入“添加数据”页面。

请参阅《[数据导入](#)》手册中的“您想如何添加数据？”。

2. 定义具有固定来源类型的数据导入。

可以是现有的来源类型，或是您定义的自定义来源类型。请参阅《[数据导入](#)》手册中的“为事件数据查看和设置来源类型”。

3. 保存新的数据导入。

注意：在进行下一步之前等待 30 秒。这给了 Splunk 软件一些时间来索引数据，并为字段提取做好准备。

4. 在“文件已成功上载”对话框中，单击 **提取字段**。

字段提取器在“[选择示例](#)”这步启动。

字段提取器：“选择示例”步骤

在字段提取器“选择示例”这步中，执行两个操作：

- 首先您为字段提取确定数据类型。您的数据类型选择会产生一个具备所选来源或来源类型值的事件列表。
- 然后您从列表中选择带有您想提取的一个或多个字段的事件。

当您从搜索结果中的特定事件进入字段提取器时，字段提取器会跳过“选择示例”这步。执行此操作时，字段提取器在“[选择方法](#)”这步启动。

选择数据类型和示例事件

注意：如果您在进入字段提取器之前选择了您的来源类型，则字段提取器会跳过该过程的第一步（选择数据类型）。

Splunk 字段提取器在示例事件中限制为二十行。

当您进入字段提取器时，会发生底下这种情况：

- 在您运行搜索（在搜索字符串中确定了特定来源类型）之后，并且随后单击了[字段边栏](#)或[“全部字段”对话框](#)中的[提取新字段](#)链接。
- 在您运行搜索（返回所有具有相同来源类型的事件集）之后，并且随后单击了字段边栏或“全部字段”对话框中的[提取新字段](#)链接。
- 在您[添加具有固定来源类型的数据导入](#)之后。

1. 为字段提取选择数据类型。

每个字段提取与特定的**来源类型**或**来源值**关联。当运行搜索后进入字段提取器时，您能选择的来源和来源类型集只限于那些搜索返回结果中发现的来源和来源类型。为了查看您的 Splunk Enterprise 实例中的所有来源和来源类型集，[可使用“设置”的“字段提取”页面](#)。

如果您选择 **sourcetype**，则会显示**来源类型**列表。在其中选择来源类型。如果您未找到想要使用的来源类型，则尝试在该搜索中指定您想要使用的来源类型，并重新运行它。

如果您选择 **source**，则会显示**来源名称**字段。在其中输入 **source** 值。

此屏幕截图是您从“设置”的“字段提取”页面进入字段提取器时查看的来源类型列表的示例。

选择示例事件

选择数据源或来源类型，选择示例事件，然后单击“下一步”进入下一步骤。该字段提取器将使用

数据类型 sourcetype ▾

来源类型 -- 选择来源类型 -- ▾

过滤器

Web	access_combined 国家超级计算应用中心 (NCSA) 合并的格式 HTTP web 服务器日志 (可由 apache 或其他 web 服务器生成)
应用	
操作系统	
数据库	apache_error 由 Apache web 服务器生成的错误日志格式 (通常为 *nix 系统上的 error_log)
未分类	
杂项	iis 由 Microsoft Internet 信息服务 (IIS) web 服务器生成的 W3C 扩展日志格式
电子邮件	
结构化	
网络和安全	
自定义	

如果您运行搜索，然后通过[单击字段边栏底部的提取新字段](#)进入字段提取器，则您的**来源类型**列表选项可能会减少。这是因为列表只显示搜索返回的数据中出现的来源类型。

在您提供来源类型或数据来源后，会显示“事件”选项卡。如果具有您所提供的来源或来源类型的事件存在，则它们会在该选项卡中列出。

2. 在事件列表中，选择一个具有您想作为字段提取的一个或多个值的示例事件。示例事件限制为二十行。

所选事件就显示在“事件”选项卡的上方。

提取字段

选择示例

选择方法

选择字段

保存

下一步 >

现有字段 >

选择示例事件

选择数据源或来源类型，选择示例事件，然后单击“下一步”进入下一步骤。该字段提取器将使用该事件来提取字段。[了解更多信息](#)
[我更喜欢自己编写正则表达式 >](#)

数据类型

sourcetype

来源类型

access_combined

24.185.15.226 - - [11/Jun/2015:13:03:06] "GET /product.screen?productId=SF-BVS-01&JSESSIONID=SD9SL8FF5ADFF4958 HTTP/1.1" 503 3308
"http://www.buttercupgames.com/cart.do?action=changequantity&itemId=EST-15" "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/536.5
(KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 724

事件

✓ 1,000 个事件 (15/11/03 9:34:55.000 之前)

每页 20 个 >

过滤器

应用

示例: 1,000 个事件 >

所有事件 >

_raw

24.185.15.226 - - [11/Jun/2015:13:03:06] "GET /product.screen?productId=SF-BVS-01&JSESSIONID=SD9SL8FF5ADFF4958 HTTP/1.1" 503 3308
"http://www.buttercupgames.com/cart.do?action=changequantity&itemId=EST-15" "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/536.5
(KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 724

当对于来源类型或所选数据来源存在字段提取时，四周会呈现事件列表中所选事件的颜色轮廓。鼠标悬停在圈出的值上，以查看字段名称。

注意：当两个或多个字段提取在所选的事件中重叠时，只会突出显示其中一个。当字段提取器检测到重叠字段时，一个红色的三角形警告图标会出现在**现有字段**按钮旁边。请参阅[“使用字段边栏来控制现有字段提取的突出显示”](#)。

3. 单击下一步转到[“选择方法”](#)这步。

使用字段边栏来控制现有字段提取的突出显示

这是您可以在每一个字段提取器步骤（除了[保存](#)这步）执行的可选操作。

您选择的来源或来源类型可能已经与搜索时间字段提取相关联。在这种情况下，字段提取器会使用彩色轮廓突出显示在示例事件中提取的字段值。

字段提取器的突出显示功能不会突出显示重叠的字段值。当两个或多个提取的字段共享事件文本时，它一次只能突出显示其中一个字段。

例如，如果字段提取器从一个事件的相同文本本位提取 `phone_number` 值（对象是 (555) 789-1234）和 `area_code` 值（对象是 555），它可以突出显示 `phone_number` 值或 `area_code` 值，但无法一次同时突出显示两个值。

当两个或多个现有字段提取重叠时，字段提取器会自动禁用所有字段的突出显示。如果您选择了一个带有重叠的字段提取的示例事件时，字段提取器会在**现有字段**按钮的旁边显示一个红色的三角形警告标记。

注意：当您使用字段边栏手动关闭不与其他字段重叠的提取字段的突出显示时，不会出现该警告。

现有字段按钮可打开字段边栏。使用字段边栏：

- 确定在示例事件中突出显示哪个现有字段提取。
- 如果您想要定义一个与现有字段提取重叠的新字段提取，请关闭现有字段提取的突出显示。
- 确定现有字段提取是否准确地提取了字段值。

1. 单击屏幕右上方的**现有字段**。

字段边栏打开。您选择的数据来源或来源类型的现有字段提取会显示在表格中。

27

字段

来源类型: access_combined

下面字段提取先前已为此来源类型定义过。有关字段对象的完整列表，请参阅 [字段页面](#)。

字段名称	模式名称	操作	突出显示
HTTP_Status	EXTRACT-HTTP_Status	打开	<input checked="" type="checkbox"/>
product_id	EXTRACT-product_id	打开	<input type="checkbox"/>
referrer	EXTRACT-referrer	打开	<input checked="" type="checkbox"/>

一个字段可能会伴随不同的**模式名称**值多次出现。

如果没有现有的字段提取，则不会出现表格。

2. (可选) 单击**打开**进行提取，以查看关于它的详细信息。

在新选项卡中会打开一个页面。该页面会显示提取字段的正则表达式。它还提供字段提取匹配的事件示例和正则表达式提取的值。

如果字段提取匹配的**事件模式**与您想从中提取字段的事件模式不同，只要有唯一的**模式名称**，您就可以创建一个带有相同名称的新提取。在“[保存](#)”步骤，您可以定义字段提取的模式名称。

3. (可选) 使用**突出显示**复选框来管理示例事件中提取字段的突出显示。

取消选中**突出显示**复选框来关闭字段的突出显示，反之亦然。

当两个或多个字段提取彼此重叠时，在任何给定的时间只有一个字段提取能启用突出显示。要使得不可用的字段提取再次可用，取消选择与它重叠的字段提取。如果随后您选择了其他提取，则您刚取消选择的提取会变得不可用。

如果您想要创建一个与现有字段提取重叠的新字段提取，则必须首先取消选择现有提取。有关更多信息，请参阅“[选择字段](#)”这步的文档。

4. 通过单击角落的 **X** 或通过单击边栏的外面来关闭边栏。

字段提取器：“选择方法”步骤

在字段提取器的“选择方法”步骤，您可以选择适合于您所用数据的字段提取方法。

该步显示了您的**数据来源**或**来源类型**以及您的示例事件。在该步的底部，您会看到两种字段提取方法：**正则表达式**和**分隔符**。

提取字段

选择示例

选择方法

选择字段

验证

保存

< 下一步 >

现有字段 >

选择方法

指示您想用来提取您的字段的方法。 [了解更多信息](#)

来源类型 access_combined

```
196.28.38.71 - - [11/Jun/2015:17:36:40] "GET /cart.do?action=addtocart&itemId=EST-14&productId=DB-SG-G01&JSESSIONID=SD4SL3FF4ADFF4953 HTTP/1.1" 200 1572 "http://www.buttercupgames.com" Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.2.28) Gecko/20120306 YFF3 Firefox/3.6.28 (.NET CLR 3.5.30729; .NET4.0C)" 448
```

(.*?)

正则表达式

Splunk Enterprise 将使用正则表达式提取字段。

x|y|z

分隔符

Splunk Enterprise 将使用分隔符（例如逗号、空格或字符）提取字段。使用该方法将逗号分隔值（CSV 文件）一样分隔数据。

1. 单击适合于您数据的字段提取方法。

如果您选择的事件是来自非结构化数据，例如系统日志，则单击**正则表达式**。字段提取器会尝试生成一个与类似事件匹配的正则表达式，并提取您的字段。

单击**分隔符**，如果您所选事件中的字段是：

- 通过一个共同的分隔符（例如空格、逗号或管道符号）清晰地分隔开。
- 在多个事件间一致（从一个事件到另一个事件，每个值都是在相同的位置）。

这通常是使用结构化、基于表格的数据的情况，例如 .csv 文件或从数据库索引的事件。

以下是一个使用逗号分隔符分隔其字段的事件示例。它的来源是来自 USGS 地震网站的 .csv 文件，该网站提供过去 30 天内全球发生的地震数据。

```
2015-06-01T20:11:31.560Z,44.4864,-129.851,10,5.9,mwb,,158,4.314,1.77,us,us2000213n,2015-06-01T21:38:31.455Z,Off the coast of Oregon
```

您会看到，在两个逗号彼此相邻的地方有一个字段缺失。

在您的字段由分隔符分隔开，但在多个事件之间不一致的情况下，您应该结合使用**正则表达式**方法和必填文本。以下是两个事件使用逗号分隔符清晰地分隔开但其字段并不一致的示例：

- indexer.splunk.com,jesse,pwcheck,fail
- Indexer.splunk.com,usercheck,greg

第二个字段提取将包括 jesse 和 usercheck，即使这些是两个不同字段的值。因此这组事件不是使用基于分隔符的字段提取的好的候选者。

2. 单击下一步转到下一步。

如果您选择**正则表达式**方法，则转到[“选择字段”](#)这步。

如果您选择**分隔符**方法，则转到[“重命名字段”](#)这步。

字段提取器：“选择字段”步骤

字段提取器的“选择字段”这步只用于基于正则表达式的字段提取。

在字段提取器的“选择字段”这步，突出显示您想用字段提取器提取的示例事件的值作为字段。

为提高字段提取的准确度，您可以选择：

- 正则表达式返回的[结果预览](#)。
- [确定其他示例事件](#)来扩大正则表达式的范围。
- [确定必填文本字符串](#)使字段提取专注于包含该文本的事件。
- [手动编辑正则表达式](#)。

确定一个或多个字段值

为您选择的来源或来源类型定义至少一个字段提取。

1. 在示例事件中，突出显示您想作为字段提取的值。

在突出显示的值下面会出现一个带有字段的对话框。

注意：字段提取器会使用彩色轮廓来标识示例事件中现有的字段提取。如果您想要选择与现有字段提取重叠的文本，则必须关闭它的突出显示，然后才能选择重叠的文本。您可以使用“现有字段”边栏来关闭以前提取字段的突出显示。请参阅“选择示例”这步中的[“使用字段边栏来控制现有字段提取的突出显示”](#)。

2. 在字段名称区域输入名称。

字段名称必须以字母开头并且其中只能包含字母、数字和下划线。

3. 单击添加提取来保存提取。

当您添加您的第一个字段提取时，字段提取器会生成一个与您已经选择的事件类似的事件相匹配的正则表达式，并试图从这些事件中提取您已经定义的字段。

字段提取器还会在示例事件下显示“预览”部分。该部分将显示与您选择的来源或来源类型匹配的事件列表，并指示这些事件中的哪一个与字段提取器已经生成的正则表达式相匹配。字段提取器使用彩色的突出显示来标识提取的字段。对于所选数据来源或来源类型，之前提取的字段通过彩色轮廓表示。

4. （可选）预览字段提取的结果，以查看是否正确提取了字段。

这可以帮助您确定是否需要采取措施，通过添加示例事件或确定必填文本来改进您的字段提取。

请参阅[“预览字段提取的结果”](#)。

5. (可选) 重复步骤 1 至 4 直到您确定了想提取的所有值。

字段提取器给每个提取的值一个不同的突出显示颜色。

当您为了提取而在事件中选择多个字段时，更可能的是字段提取器将不能生成一个能可靠提取所有字段的正则表达式。通过[添加示例事件](#)和[确定必填文本](#)能够提高多字段提取的可靠性。也可以通过[手动编辑](#)改进正则表达式。

6. (可选) 通过单击示例事件中的字段提取和选择**删除**或**重命名**操作将其删除或进行重命名。

7. 单击下一步转到[“验证字段”](#)这步。

预览字段提取的结果

此操作对于“选择字段”这步和“验证字段”这步都是可选的。

在您添加第一个字段提取后，会出现“预览”部分。它显示与您选择的来源或来源类型匹配的事件列表。它还会显示您正尝试从示例事件中提取的每个字段的选项卡。

事件列表具有用来检查字段提取准确度的功能。列表默认为来源类型显示示例中的所有事件。

- 用最左边一列来确定哪些事件与正则表达式匹配，哪些不匹配。
- 如果正则表达式匹配了一小部分示例事件，可切换视图到**匹配**来移除列表中不匹配的事件。您也可以选择不**匹配**仅查看未与正则表达式匹配的事件。
- 单击字段选项卡查看字段的值分布统计。每个字段选项卡从高到低依次显示在事件示例字段中查找的每个值计数的条形图。



- 单击图表上的一个值来过滤出含该值的字段列表。例如，在 `status` 图表上，单击 `503` 值会导致过滤器设为 `status=503`，字段提取器返回主“预览”字段列表视图。它仅列出包含 `status` 值的事件。

您可能会找到未正确匹配事件的生成的字段提取。或者您会发现它正在提取错误的字段值。当这种情况发生时，您可以采取措施来改进字段提取。

您可以：

- [添加示例事件来扩大正则表达式的范围](#)。这可以帮助它匹配更多的事件。
- [确定必填文本来创建与特定事件模式匹配的提取](#)。这缩小了正则表达式所匹配的事件集。
- 在“[验证字段](#)”这步中，提交不正确的提取字段值作为反例。
- 当提取失败时，从包含多个字段的提取中移除字段。您可以为这些移除的字段创建其他的字段提取。

添加示例事件来扩大正则表达式的范围

此操作对于“选择字段”这步是可选的。

当您在示例事件中选择字段集时，您会发现带有这些字段的事件并不匹配。当字段提取器生成的正则表达式匹配了模式与示例事件类似的事件，但是错过了模式略有不同的其他事件的时候，这种情况就会发生。

通过将错过的事件之一添加为补充示例事件来尝试扩大正则表达式的范围。在您突出显示错过的事件后，字段提取器会尝试生成新的涵盖两种事件模式的字段提取。

1. 在字段列表中，单击一个与正则表达式不匹配但是具有从首个示例事件中提取的所有字段值的事件。

当其他示例事件的格式或模式与原始示例事件最匹配的时候，最可能提高字段提取的准确度。所选择的示例事件会显示在原始示例事件的下面。

2. 在其他示例事件中，突出显示从首个示例事件中提取的字段值。

3. 选择正确的字段名称。

只查看在首个示例事件中确定的字段名称。

4. 单击添加提取。

字段提取器尝试扩大正则表达式的范围，这样就能查找两种事件模式中的字段值。它与新正则表达式而非示例事件匹配，然后在事件表中显示结果。

5. (可选) 如果您在提取多个字段，为每个字段重复执行步骤 2 到 4。

不需要突出显示在首个示例事件中突出显示的所有字段。例如，当其他示例事件只突出显示在原始示例事件中突出显示的两个字段之一时，您可以得到一个更可靠的字段提取结果。

6. (可选) 添加其他示例事件。

7. (可选) 通过单击事件旁边的灰色 "X" 删除示例事件。

选择字段

突出显示示例事件中的一个或多个值以创建字段。您可指示一个值为必填，这意味着该值必须存在于事件中，以匹配正则表达式。单击示例事件中突出显示值以修改这些值。 [了解更多信息](#)

显示正则表达式 >

预览

如果您看到下面的不正确结果

Events

sent_to

✓ 1,000 个事件 (15/01/08 15:17:05.000 之前)

每页 20 个 < 预览 1 2 3 4 5 6 7 8 9 ... 下一步 >

过滤器 应用

示例: 前 1,000 个事件 < 所有事件 > 所有事件 匹配 不匹配

	_raw	sent_to
✓	Sun Sep 28 2014 14:46:03 Sent to Accounting System 100303	Accounting
✗	Sun Sep 28 2014 14:46:03 TransactionID=107387 AcctCode=4400-4383	
✓	Sun Sep 28 2014 14:46:01 ecomm engine response TransactionID=107387 CustomerID=5i31kpk5 accepted	response

有时字段提取器不能构建一个匹配示例事件以及原始示例事件的正则表达式。您可以使用以下方法之一解决问题：

- 如果您在提取多个字段，删除您尝试提取的一些字段。此操作会产生一个在全部所选择事件间运行的字段提取。您该删除的首个字段值是那些嵌入到更长文本字符串中的那些值。您可以为删除的字段设置单独的字段提取。
- 为每个包含您想提取的字段值的事件模式定义单独字段提取，且使用必填文本分别设置提取。关于必填文本的信息，请参阅下一个主题。

确定必填文本来创建与特定事件模式匹配的提取

此操作对于“选择字段”这步是可选的。

有时一个来源类型包含不同种类的事件，这些事件包含了同样的字段或您想提取的字段。设计一个单一的能匹配多个事件模式的字段提取是很困难的。处理这种情况的一种方式是为每个事件模式定义不同的字段提取。

您可以让提取专注于带有必填文本的特定事件模式。必填文本的行为与搜索过滤条件相同。它是一个文本字符串，必须包含在 Splunk 软件事件中，以与提取相匹配。

例如，对于 `access_combined` 来源类型，可能有通过字符串 `action=addtocart`、`action=changequantity`、`action=purchase` 和 `action=remove` 区分的不同事件模式。您可以创建四个提取，每个字符串一个，每个都提取同样的字段，但是其必填文本是不同的字符串。

您也可以使用必填文本来确保仅从特定事件中提取值。

必填文本定义有两个限制：

- 对于单个字段提取，只能定义一个必填文本字符串。
- 不能将必填文本字符串应用于突出显示为提取字段值的文本字符串，反之也不行。

1. 在示例事件中，突出显示您需要的文本。

2. 选择需要。



3. 单击**添加必填文本**来向字段提取中添加必填文本。

4. (可选) 通过单击示例事件中的必填文本并选择**删除必填文本**来将其删除。

选择字段
突出显示示例事件中的一个或多个值以创建字段。您可指示一个值为必填，这意味着该值必须存在于事件中，以匹配正则表达式。单击示例事件中突出显示值以修改这些值。 [了解更多信息](#)

```
66.69.195.226 - - [22/Sep/2014:12:58:24] "POST /cart.do?action=purchase&itemId=EST-11&JSESSIONID=SD2SL7FF7ADFF4961 HTTP 1.1" 200 1219 "http://www.buttercupgames.com/cart.do?action=addtocart&itemId=EST-11&categoryId=ACCESSORIES&productId=WC-SH-A02" "Mozilla/5.0 (compatible; NetcraftSurveyAgent/1.0/cc-prepass-https; +info@netcraft.com)" 933
```

[显示正则表达式](#)

预览

如果您看到下面的不正确结果，则单击其他事件，以将其添加到示例事件集中。突出显示该事件值以改善提取。您可在下一步骤删除不正确的值。

事件	http_method	status
✓ 1,000 个事件 (15/01/08 15:23:36.000 之前)		
过滤器 应用 示例: 前 1,000 个事件 所有事件 匹配 不匹配		
194.215.205.19 - - [22/Sep/2014:13:35:27] "POST /product.screen?productId=DC-SG-G02&JSESSIONID=SD7SL7FF1ADFF4961 HTTP 1.1" 200 1303 "http://www.buttercupgames.com/oldlink?itemId=EST-19" "Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.2.28) Gecko/20120306 YFF3 Firefox/3.6.28 (.NET CLR 3.5.30729; .NET4.0C)" 291		200
194.215.205.19 - - [22/Sep/2014:13:35:15] "GET /category.screen?categoryId=ARCADE&JSESSIONID=SD7SL7FF1ADFF4961 HTTP 1.1" 200 3863 "http://www.buttercupgames.com" "Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.2.28) Gecko/20120306 YFF3 Firefox/3.6.28 (.NET CLR 3.5.30729; .NET4.0C)" 468		200
128.241.220.82 - - [22/Sep/2014:13:34:11] "POST /cart.do?action=purchase&itemId=EST-18&JSESSIONID=SD7SL7FF7ADFF4961 HTTP 1.1" 503 2053 "http://www.buttercupgames.com/cart.do?action=addtocart&itemId=EST-18&categoryId=STRATEGY&productId=PZ-SG-G05" "Opera/9.20 (Windows NT 6.0; U; en)" 279	POST	503

本示例显示一个字段提取，它提取名称为 `http_method` (绿色) 和 `status` (黄色) 的字段，并且将 `action=purchase` 定义为必填文本。在字段列表中，前两个事件并不与提取匹配，因为它们没有必填文本。第三个事件与正则表达式匹配且有必填文本。提取的字段已经突出显示。

过滤器功能是设置和测试必填文本的有用工具。

手动编辑正则表达式

此操作对于“选择字段”这步和“验证”这步都是可选的。

您可以手动编辑正则表达式。然而，执行此操作会使您退出字段提取器的工作流。当您将其更改保存到字段提取器时，字段提取器将带您转到最后的“保存”步骤。

1. 单击**显示正则表达式**。

2. 单击**编辑正则表达式**。

如果您想要放弃手动编辑正则表达式并返回到字段提取器工作流，请单击页面左上方的**返回**按钮。如果您还没有尝试预览正则表达式的更改，则只可以返回。

3. 编辑正则表达式。

4. 单击**预览**来针对示例事件匹配所编辑的提取。

返回按钮消失。**预览**按钮显示为灰色，直到您对于字段提取做出更多的编辑。

使用**过滤**、**示例**和**匹配**以及**不匹配**控件来帮助您评估正则表达式的质量。

重复步骤 3 到 4 直到正则表达式与事件匹配并能提取相应字段。

5. 单击**保存**来保存新的字段提取。

字段提取器带您转到“保存”步骤。
当您进入“保存”这步，单击返回继续编辑正则表达式。在您为提取输入一个名字或做了权限选择之后，返回按钮就会消失。

提取字段

< 上一步

现有字段 >

1

如果您手动编辑，然后预览下面的正则表达式，则您无法返回到自动提取 workflow。

使用下面列出的事件验证由您的正则表达式生成的字段提取。

正则表达式

[正则表达式参考](#) [在搜索中查看](#)

预览

保存

事件

http_method

status

url

✓ 1,000 个事件 (15/01/08 15:23:36.000 之前)

每页 20 个

过滤器

应用

示例: 前 1,000 个事件

所有事件

值	计数	%
200	890	89.000
408	23	2.300
503	21	2.100
500	16	1.600
...

请参阅本手册中的[“关于 Splunk 正则表达式”](#)。

字段提取器：“重命名字段”步骤

字段提取器的“重命名字段”这步只用于基于分隔符的字段提取。如果您正在使用正则表达式提取字段，请参阅[“选择字段”](#)和[“验证”](#)这步的相关主题。

在“重命名字段”这步中，您可以：

- 确定示例事件中分隔字段的分隔符，例如空格、逗号、制表符、管道或其他字符或字符组合。字段提取器根据您选择的分隔符将事件分解为字段。
- 重命名您想要从这些事件中提取的一个或多个字段。
- 可选择预览基于分隔符的字段提取的结果。这可以帮助您验证提取并确定哪些字段要重命名。

确定分隔符和重命名一个或多个字段

确定分隔符。重命名至少一个字段。

1. 在“重命名字段”下方，选择一个可用的分隔符选项，或提供一个您自己的分隔符。

字段提取器将示例事件替换为显示它在事件中的找到的字段（使用您选择的分隔符）。它给每个字段一种颜色和一个临时名称（field1、field2、field3，依此类推）。

如果您选择空格、逗号、制表符或管道，则字段提取器会将事件分解为基于分隔符的字段。例如，如果您选择管道作为分隔符，一个类似 2015-06-01T14:07:50:170Z|Jones|Alex|555-922-1212|324 Bowie Street|Alexandria, Va 的字符串将会分解成六个单独的字段。

如果分隔符不是这四个选项之一，选择其他，在提供的字段中输入一个或多个分隔符字符。然后单击“返回”键，让字段提取器基于该分隔符将事件分解为字段。

字段提取器也可以在字段显示的下方创建“预览”区域，来预览基于分隔符的字段提取如何用于数据集中的其他事件（由您的来源或来源类型选择所代表的）。请参阅“预览字段提取的结果”。

33



2. (可选) 查看“预览”部分的内容来确定基于分隔符提取的准确度和识别应该重命名的字段。

这可以帮助您决定哪些字段要重命名。

3. 单击您想要重命名的字段。

字段名称字段会显示。输入正确的字段名称。

您必须选择并重命名至少一个字段以转到“保存”步骤。

4. 单击**重命名字段**来重命名字段。

字段提取器将字段的临时名称替换为您通过该页面提供的名称。

重命名字段

选择一个分隔符。在显示的表中通过单击字段名称或值来重新命名字段。 [了解](#)



5. (可选) 对于您从事件中选择要重命名的所有其他字段，重复步骤 3 和 4。

注意：您不必重命名字段提取器发现的每个字段。

6. 单击**下一步**转到[保存](#)这步。

预览字段提取的结果

这些操作对于“重命名字段”这步是可选的。

在字段提取器将基于分隔符的字段提取应用于您的示例事件之后，此页面的下半部分会变为“预览”。您可以在“预览”部分，针对由您选择的来源或来源类型所代表的数据集，来预览该提取的结果。

“预览”部分具备一些功能，您可以用来检查字段提取的准确度和确定您可能想要重命名的字段。它由一张表组成，该表显示根据您选择的分隔符分解为字段的事件。它也会为字段提取器发现的每个字段提供信息选项卡。

1. (可选) 更改预览数据集的样本大小以查看更大范围的事件统计数据。

默认情况下，预览部分显示数据集中**前 1,000 个事件**的结果。您可以更改预览设置为前 10,000 个事件或过去五分钟、24 小时或 30 天的事件。

- 2. (可选) 查阅第一列，以查看是否有任何事件匹配所选事件的模式失败。**

“预览”事件列表的第一列对于匹配模式的事件显示绿色对勾标记，对于不匹配的事件显示红色的“X”。

如果您有不匹配的事件，这意味着那些事件与您的示例事件相比可能有更多或更少的字段，而且您可能要尝试使用不同的分隔符，或调查为什么您选择的分隔符只能用于您的事件集中的某些事件。

通过使用**匹配**和**不匹配**过滤器，您可以快速查找很少匹配或不匹配事件。

- 3. (可选)** 单击“字段”选项卡以查看其信息。

每个字段信息选项卡提供该字段的值分布，按最常见到最罕见来组织。它基于所选的事件示例。如果默认的样本值 1,000 没有提供您期望查看的值，则尝试将它更改为更大的样本值。

字段提取器：“验证”步骤

字段提取器的“验证”这步只用于基于正则表达式的字段提取。

在字段提取器的“验证”这步验证您的字段提取。字段提取器提供下列验证方法：

- 查阅事件列表来查看哪个事件与字段提取匹配或者哪个不匹配。请参阅[“预览字段提取的结果”](#)。
- 通过提供反例向字段提取器举报不正确的提取。相应地，字段提取器会尝试提高正则表达式的准确度。
- 手动编辑正则表达式。请参阅[“手动编辑正则表达式”](#)。

当验证字段提取完成的时候，单击**保存**来[保存提取](#)。

提供反例反馈信息

对于“验证”这步，这是可选操作。

如果您发现事件中包含不正确的提取字段，可将它们作为反例的反馈信息提交。

- ### 1. 查找含有不正确提取的字段值的事件。

突出显示文本不是荧光笔所代表字段的正确值。

- 2.单击不正确字段值旁边灰色的 "X"。**

字段提取器在表上方显示反例事件，并用红色删除线标记不正确的值。它也会更新正则表达式及其预览结果。

验证字段

验证您的字段提取，并删除事件选项卡错误突出显示的值。在字段选项卡上，检查每个字段的提取值，并根据需要单击某个值，以将其作为搜索过滤器应用到事件选项卡事件列表上。

```
127.0.0.1 - splunk-system-user [26/Sep/2014:14:30:01.314 -0700] "GET /services/NS-/sso/admin/summarization?_nop_sid=scheduler_nobody&_sso_RWD59d4672721e98f163_at1411767000_61&use_normalized=yes&noProxy=true&noDetails=1&search=summary.hash%3DNS4844c4fe836790802AND%20ae1:acl.app%3Dsos&sort_key=summary.size&sort_dir=desc&sort_mode=num HTTP/1.0" 200 1689 - - 6ms
```

```
127.0.0.1 - admin [26/Sep/2014:14:29:07.240 -0700] "GET /services/search/jobs/1411766946.143/results?max_lines=100&count=44&output_mode=xml&time_format=%25Y-%25m-%25dT%25H%25M%25S%25Z&show_empty_fields=True&offset=0&field_list= HTTP/1.0" 200 13735 - - 4ms
```

[显示正则表达式 >](#)

The screenshot shows the Splunk search interface. At the top, the search bar contains the query 'summary'. Below the search bar, the results are displayed in a table. The first two rows of the table are visible, showing log entries from the 'splunk-system-user' on 26/Sep/2014. The first row shows a GET request to '/servicesNS/-/sos/admin/summarization?' with a search hash of 30dbd6f0f9a5afb262%. The second row shows a similar GET request with a search hash of 3087aec77304e257c3%. The interface includes a sidebar with '事件' (Events) and 'sort_key' tabs, a search bar with a filter icon, and a table of results with columns for time, source, and raw data.

- 3. 如果一个反例没有帮助作用，可通过单击反例事件左侧蓝色的“X”删除它。**

字段提取器：“保存”步骤

在字段提取器的“保存”这步，您可以定义新字段提取定义的名称、设置其权限和保存提取。

- 1.如果字段没有名称，请定义一个，或者验证字段提取器提供的名称是否正确。**

如果您使用正则表达式模式创建了您的字段提取定义，则**名称**将由定义提取的字段的逗号分隔列表组成。您可以更改该名称。

如果您使用分隔符模式创建了您的字段提取定义，则**名称**将为空白。您必须提供一个名称来保存该字段提取定义。

注意：提取名称不能包含空格。

2. (可选) 对于任一**应用**或者**全部应用**更改字段提取的**权限**并且更新基于角色的读取/写入权限。

您只能更改字段提取权限，如果您的角色包括允许您执行此操作的**功能**的话。

字段提取权限设置为**所有者**，这意味着它只提取由创建提取的人所运行的搜索中的字段。

权限设置为**应用**会使该提取仅能由字段提取所属应用的用户使用。

权限设置为**全部应用**会使所有应用的全部用户都能在运行搜索时受益于该字段提取。

当您更改应用权限为**应用**或者**全部应用**时，可以为每个角色设置读取和写入权限。请参阅本手册中的[“管理知识对象权限”](#)。

3. 单击**完成**来保存提取。

您可以管理您创建的字段提取。它们列在“设置”的“字段提取”页面上。请参阅本手册中的[使用“字段提取”页面](#)。

使用“字段提取”页面

使用“设置”中的“字段提取”页面来管理搜索时间**字段提取**。您可以通过三种方法来添加搜索时间字段提取。您可以：

- [使用字段提取器](#)创建提取。此方法相对容易，不需要您了解正则表达式如何工作。
- [如果您有 Splunk Enterprise，可以直接编辑 props.conf。](#)
- 使用“字段提取”页面添加新的字段提取（请参阅下文）。

通过“字段提取”页面，您可以执行以下操作：

- 针对您的 Splunk 部署中的所有应用检查您所创建或您有权查看的整个搜索时间提取集。
- 创建新的搜索时间字段提取。
- 更新字段提取权限。通过字段提取器和“字段提取”页面创建的字段提取在与其他人共享前，最初仅创建者可用。
- 删除字段提取（前提是，您的应用级别权限允许您这样做并且这些提取不是随产品一起提供的默认提取）。不能删除默认知识对象。有关删除知识对象的更多信息，请参阅本手册中的[禁用或删除知识对象](#)。

如果对于特定的搜索时间字段提取，您拥有“写入”**权限**，您可以通过“字段提取”页面执行以下操作：

- 更新其正则表达式（如果是内联交易）。
- 添加或删除已经在 transforms.conf 或 [Splunk Web 的字段交易页面](#)中定义的命名提取（如果它使用交易）。

注意：不能在 Splunk Web 中管理索引时间字段提取。我们建议您不要更改索引时间字段提取集，但是如果您发现必须这样做，则需要手动修改 props.conf 和 transforms.conf 配置文件。有关索引时间字段提取配置的更多信息，请参阅《数据导入手册》中的“配置索引时间字段提取”。

选择**设置 > 字段 > 字段提取**可导航到“字段提取”页面。

在 Splunk Web 中检查搜索时间字段提取

要更好地了解“字段提取”页面如何显示您的字段提取，应了解如何在 props.conf 和 transforms.conf 文件中设置字段提取。

props.conf 中可完成字段提取的全部设置，此时这些提取在“字段提取”页面中将被标识为**内联**字段提取。但是一些字段提取包含 transforms.conf 组件（称为**字段转换**）。要通过 Splunk Web 创建/编辑该字段提取组件，应使用 Splunk Web 的“字段转换”页面。

有关转换和“字段转换”页面的更多信息，请参阅本手册中的[“管理字段转换”](#)。

有关直接在 props.conf 和 transforms.conf 文件中设置字段提取的更多信息，请参阅本手册中的[“通过配置文件创建和维护搜索时间字段提取”](#)。

名称列

“字段提取”页面中的**名称**列显示字段提取的完整名称（或“类”）。字段提取格式为：

```
<spec> : [EXTRACT-<class> | REPORT-<class>]
```

- **<spec>** 可以是：
 - **<sourcetype>**，事件的来源类型。
 - **host::<host>**，其中 **<host>** 为事件所在的主机。

- `source::<source>`，其中 `<source>` 为事件的来源。

EXTRACT-`<class>` 字段提取是指在 `props.conf` 中完整定义的提取（即，未引用 `transforms.conf` 中的转换的提取）。它们是由通过 IFX 和某些搜索命令进行的字段提取自动创建的。如果您有 Splunk Enterprise，可以[通过直接更新 props.conf 文件](#)的方式来添加这些提取。此类提取始终与字段提取正则表达式关联。在“字段提取”页面上，此正则表达式显示在**提取/转换**列中。

REPORT-`<class>` 字段提取引用 `transforms.conf` 中的字段转换段落。这是其字段提取正则表达式所在的位置。在“字段提取”页面上，所引用的字段转换段落显示在**提取/转换**列中。

在 Splunk Web 中，您可以通过“字段转换”页面使用转换。有关更多信息，请参阅本手册中的[“使用 Splunk Web 的字段转换页面”](#)。

类型列

字段提取有两种类型：**内联**和 `transforms.conf`。

- **内联**提取始终具有 `EXTRACT-<class>` 配置。这样标识的原因是此类提取完全在 `props.conf` 中定义；而不引用外部字段转换。
- **使用转换**提取始终具有 `REPORT-<class>` 名称配置。因此，它们引用 `transforms.conf` 中的字段转换。您可以直接在 `transforms.conf` 中定义字段转换，也可以通过 Splunk Web 的“字段转换”页面定义。

提取转换列

在**提取/转换**列中，Splunk Web 会根据字段提取**类型**显示不同的内容。

- 对于**内联**提取类型，Splunk Web 显示 Splunk 软件用于提取字段的正则表达式。正则表达式包含的一个或多个命名组会显示其所提取的字段。

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。

- 对于**使用转换**提取类型，Splunk Web 显示一个或多个 `transforms.conf` 字段转换段落的名称，字段提取通过 `props.conf` 链接到这些段落。如果您要对相同来源、来源类型或主机应用多个字段提取正则表达式，则字段提取可以引用多个字段转换。如果您要提取的一个或多个字段以两个或更多个截然不同的事件模式显示，这是非常必要的。

例如，某一**使用转换**提取在“表达式”列中可能显示两个值：`access-extractions` 和 `ip-extractions`。在 `props.conf` 中，这两个值可能显示为：

```
[access_combined]
REPORT-access = access-extractions, ip-extractions
```

在本例中，`access-extractions` 和 `ip-extractions` 都是 `transforms.conf` 中的字段转换段落的名称。要通过 Splunk Web 使用这些字段转换，请[转到“字段转换”页面](#)。

添加新字段提取

单击“字段提取”页面顶部的**新建**按钮，添加新字段提取。此时将显示“新增”页面。

如果您知道如何在 `props.conf` 中设置字段提取，会发现此操作相当简单。

下文所述的所有字段均为必填字段。

1. 为字段提取定义**目标应用**上下文。默认为您当前所在的应用上下文。
2. 指定字段提取的**名称**（使用下划线来分隔单词）。在 `props.conf` 中，对于 `EXTRACT` 或 `REPORT` 字段提取类型，为 `<class>` 值。**注意：**`<class>` 值不需要遵循字段名称语法限制（请参阅下文的“**重要提示**”注意事项）。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。另外，`<class>` 值也不受“键清理”功能的制约。
3. 定义提取所应用于的来源类型、来源或主机。选择**来源类型**、**来源**或**主机**并输入值。这将映射到 `<spec>` 值（位于 `props.conf` 中）。
4. 定义提取类型。

如果选择**使用转换**，则输入**提取/转换**字段中涉及的转换并用逗号进行分隔。然后，可以[通过“字段转换”页面](#)创建或更新转换。

如果选择**内联**，则在**提取/转换**字段中输入用于提取字段的正则表达式。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 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。

重要提示：正则表达式中捕获组内的字段名称必须仅包含字母数字字符或下划线。

- 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 Enterprise 的内部变量。
- 不允许使用国际字符。

无论是默认情况还是自定义配置，Splunk Enterprise 都会对所有提取的字段应用以下“键清理”规则：

- 所有**非** a-z、A-Z 和 0-9 字符均替换为下划线 (**_**)。
- 从提取的字段名称中删除所有前导下划线和 0-9 字符。

要为某一特定字段提取禁用此行为，必须同时手动修改 `props.conf` 和 `transforms.conf`。有关更多信息，请参阅本手册中的[“通过配置文件创建和维护搜索时间字段提取”](#)。

注意：对于内联字段提取（不需要字段转换组件的字段提取），如果未编辑 `props.conf` 中的提取段落，则不能关闭键清理功能。

示例 - 添加新的错误代码字段

本例介绍如何为新增的 `err_code` 字段定义提取。您可以通过 `device_id=` 后跟带括号的单词和一个以冒号结尾的文本字符串来找到此字段。应与 `testlog` 来源类型相关的事件中提取此字段。

在 `props.conf` 中，该提取类似如下所示：

```
[testlog]
EXTRACT-errors = device_id=\[w+\] (?<err_code>[^\:]+)
```

以下介绍如何通过“添加新字段提取”页面来设置此提取：

新增

字段 » 字段提取 » 新增

目标应用 *

search

名称 *

errors

应用到 *

sourcetype

已命名 *

testlog

类型 *

嵌入

提取/转换 *

device_id=\[w+\] (?<err_code>[^\:]+)

如果嵌入了字段提取，请提供正则表达式。如果字段提取使用转换，请指定转换名称。

取消

保存

注意：您可以在本手册中的[“创建和维护搜索时间字段提取”](#)主题中找到此示例，此示例向您介绍了如何使用 `props.conf` 文件设置字段提取。

更新现有字段提取

要编辑现有字段提取，请在**名称**列中单击其名称。

PerformanceMonitor : REPORT-MESSAGE

字段 » 字段提取 » PerformanceMonitor : REPORT-MESSAGE

提取/转换 *

perimon-kv

如果输入了字段提取，请提供正则表达式。如果字段提取使用转换，请指定转换名称。

取消

保存

这会将您带到该字段提取的详细信息页面。在**提取/转换**字段中，可以执行的操作取决于所处理的提取类型。

- 如果字段提取为内联提取，可以编辑用于提取字段的正则表达式。
- 如果字段提取使用一个或多个转换，可以更新所涉及的一个或多个转换（如果存在多个转换，将这些转换放入一个逗号分隔列表中）。然后，可以[通过“字段转换”页面](#)创建或更新转换。

上文描述的字段提取使用了三个转换，名为 `wel-message`、`wel-eq-kv` 和 `wel-col-kv`。要了解如何设置这些转换的更多信息，请导航到**设置 > 字段 > 字段转换**或直接转到 `transforms.conf`。

注意：使用转换字段提取必须包含至少一个有效的 `transforms.conf` 字段提取段落名称。

更新字段提取权限

如果字段提取是通过内联方法（如 IFX 或搜索命令）创建的，则最初只有创建者可以使用该提取。要使该字段提取对其他用户可用，需要更新其**权限**。为此，需要在“字段提取”页面上找到该字段提取，并选择其**权限**链接。随即打开 Splunk Web 中用于**知识对象**的标准权限管理页面。

在此页面中，您可以为字段提取设置**基于角色**的权限，并确定该提取是对一个特定应用的用户可用还是对所有应用的用户全局可用。有关使用 Splunk Web 管理权限的更多信息，请参阅本手册中的[“管理知识对象权限”](#)。

删除字段提取

在 Splunk Web 的“字段提取”页面上，如果您的权限允许，可以删除字段提取。不能删除默认的字段提取（即随产品一起提供的提取以及存储在应用的“默认”目录中的提取）。

单击要删除的字段提取所对应的**删除**。

注意：删除具有下游依赖性的对象时须谨慎。例如，如果使用您的字段提取的搜索又是某个事件类型的基础，并且该事件类型由其他五个保存的搜索（其中两个搜索是仪表板面板的基础）使用，则将该提取从系统中删除将会对所有其他知识对象产生负面影响。有关删除知识对象的更多信息，请参阅本手册中的[“禁用或删除知识对象”](#)。

使用“字段转换”页面

通过“设置”中的“字段转换”页面，您可以管理 `transforms.conf` 中包含的搜索时间字段提取的**字段转换**组件。可以通过直接编辑 `transforms.conf` 创建字段转换，也可以通过在“字段转换”页面添加来创建字段转换。

注意：每个字段转换至少包含一个字段提取组件。但“内联”字段提取不需要包含字段转换组件。

通过“字段转换”页面，您可以执行以下操作：

- 针对您的 Splunk 部署中的所有应用检查您所创建或您有权查看的整个字段转换集。
- 创建新的搜索时间字段转换。有关要求使用字段转换的情况的更多信息，请参阅下文的“何时使用字段转换页面”。
- 更新字段转换权限。通过“字段转换”页面创建的字段转换，最初仅仅创建者可用，而现在已经可以与其他人共享。仅当您拥有该转换或您的角色权限允许的情况下，才能更新字段转换权限。
- 删除字段转换（前提是，您的应用级别权限允许您这样做并且这些转换不是随产品一起提供的默认字段转换）。不能删除默认知识对象。有关删除知识对象的更多信息，请参阅本手册中的[“禁用或删除知识对象”](#)。

如果对于特定的字段转换您拥有“写入”权限，通过“字段转换”页面您可以执行以下操作：

- 更新其正则表达式并更改该正则表达式所应用于的键。
- 定义或更新字段转换格式。

选择**设置 > 字段 > 字段转换**可导航到“字段转换”页面。

为何为字段提取设置字段转换？

虽然大多数搜索时间字段提取可以完全在 `props.conf`（或 Splunk Web 的“字段提取”页面）中进行定义，但是有些高级搜索时间字段提取需要名为**字段转换**的组件 `transforms.conf`。该组件可通过“字段转换”页面进行定义和管理。

当您需要执行以下操作时，应设置具有字段转换组件的搜索时间字段提取：

- **在多个来源、来源类型或主机之间重复使用相同的字段提取正则表达式**（即，配置一个字段转换供多个字段提取引用）。如果您发现自己需要使用同一正则表达式来提取不同来源、来源类型和主机的字段，可能希望将该表达式设置为一个转换。然后，如果您发现自己需要更新该正则表达式，只需执行一次此操作，即使它是由多个字段提取所使用也是如此。
- **对同一个数据来源、来源类型或主机应用多个字段提取正则表达式**（即，将多个字段转换应用于同一字段提取）。有时，如果您要从特定来源/来源类型/主机中提取的字段以两个或更多个截然不同的事件模式显示，这是非常必要的。
- **使用正则表达式从其他字段值中提取字段**（也称为“来源键”）。例如，您可能从 `url` 字段值中提取一个字符串并将该字符串设置为一个新字段的值。

如果直接在 `transforms.conf` 中进行配置，您可以使用搜索时间字段转换完成更多的操作（例如，设置基于分隔符的字段提取和配置多值字段提取）。有关更多信息，请参阅本手册[“通过配置文件创建和维护搜索时间字段提取”](#)中的字段转换设置部分。

注意：所有索引时间字段提取都与一个或多个字段转换关联。但您不能通过 Splunk Web 来管理索引时间字段提取，而是必须使用 `props.conf` 和 `transforms.conf` 配置文件。在正常情况下，我们建议您不要更改索引时间字段提取，但是如果有必要这样做的话，请参阅《数据导入手册》中的“在索引时创建自定义字段”。

在 Splunk Web 中查看和更新搜索时间字段转换

要更好地了解 Splunk Web 的“字段转换”页面如何显示您的字段转换，应了解如何在 `props.conf` 和 `transforms.conf` 文件中设置搜索时间字段提取。

在 `transforms.conf` 中，典型的字段转换类似如下所示：

```
[banner]
REGEX = /js/(?<license_type>[^/]*)(?<version>[^/]*)/login/(?<login>[^/]*
SOURCE_KEY = uri
```

此转换会将其正则表达式与 `uri` 字段值进行匹配，并提取三个字段作为命名的组：`license_type`、`version` 和 `login`。

在 `props.conf` 中，该转换将与来源 `.../banner_access_log*` 进行匹配，如下所示：

```
[source:.../banner_access_log*]
REPORT-banner = banner
```

这意味着，正则表达式将仅与事件中的 `uri` 字段（来自 `.../banner_access_log` 来源的事件）进行匹配。

但是如有必要，您可以将其与其他来源、来源类型和主机相匹配。使用内联字段提取（即完全在 `props.conf` 中设置的字段提取）则无法实现。

注意：默认情况下，转换与 `SOURCE_KEY` 值（`_raw` 的值）相匹配，此时其正则表达式将应用于整个事件，而不是仅仅应用于该事件内的字段。

名称列

“字段转换”页面的**名称列**显示您有权查看的搜索时间字段转换的名称。这些名称是 `transforms.conf` 所包含字段转换的实际段落名称。上文所示的转换示例将在转换列表中显示为 `banner`。

单击转换名称可查看该特定转换的详细信息。

查看和编辑转换详细信息

通过字段转换的详细信息页面，可以查看并更新其正则表达式、键及事件格式。下面是我们在本子主题开头部分介绍的 `banner` 转换的详细信息页面：

Banner

字段 » 字段转换 » Banner

正则表达式 *

/js/(?<license_type>[^\s]*)/(?<version>[^\s]*)/login/(?<login>[^\s]*) SOURCE

来源键 *

_raw

指定应用正则表达式的键，默认为 _raw。

格式

根据字段名称和值指定事件格式。
使用 \$n (例如 \$1, \$2 等)指定正则表达式输出匹配。默认为 <transform_stanza_name>::\$1

☐ 创建多值字段

如算式中，在已经提取字段时，提取器将创建多值字段。

☒ 自动清除字段名称

如算式中，将清除字段名称，使其只包含 a-zA-Z0-9_

取消

保存

如果您具有相应的权限，则可以编辑正则表达式、键和事件格式。请记住，如果转换已经应用于多个数据来源、来源类型或主机，那么这些编辑可能会影响在 `props.conf` 和“字段提取”页面中定义的多个字段提取。

创建新字段转换

创建新字段转换：

1. 首先，导航到“字段转换”页面并单击**新建**按钮。

2. 确定字段转换的**目标应用**（如果不是当前使用的应用）。

3. 为字段转换指定一个**名称**。

此名称与 `transforms.conf` 中指定的转换段落名称一致。保存此转换后，此名称将显示在“字段转换”页面上的**名称**列中。（此字段为必填字段。）

4. 输入转换的**正则表达式**。

请参阅[“正则表达式语法和用法”](#)。

5. （可选）定义转换的**键**。

此字段对应于 `SOURCE_KEY` 选项（位于 `transforms.conf` 中）。默认情况下，此字段设置为 `_raw`（表示正则表达式将应用于整个事件）。

要将正则表达式应用于特定字段的值，将 `_raw` 替换为该字段的名称。只能使用执行字段转换时存在的字段。

6. （可选）指定**事件格式**。

此字段对应于 `FORMAT` 选项（位于 `transforms.conf` 中）。使用 `$n` 表示正则表达式所捕获的组。例如，如果您所设计的正则表达式捕获两个组，可以将**格式**设置为如下形式：`$1::$2`，其中第一个组是字段名称，第二个组是字段值。或者，可将**格式**设置为 `username::$1 userid::$2`，表示正则表达式提取 `username` 和 `userid` 字段的值。“格式”字段的默认值为 `<transform_stanza_name>::$1`。

7. （可选）选择**创建多值字段**，如果相同字段可以从您的事件中多次提取。

这使得 Splunk 软件将字段作为单个多值字段提取。

8. （可选）选择**自动清理字段名称**来确保提取的字段具有有效名称。

它会从字段名称中移除前导下划线字符和 0-9 数字字符，并且会将除了 a-z、A-Z 和 0-9 范围内的字符之外的其他字符替换为下划线。请参阅[“正则表达式语法和用法”](#)。

正则表达式语法和用法

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。

重要提示：正则表达式中捕获组内的字段名称必须包含字母数字字符或下划线。

- 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 软件的内部变量。
- 不允许使用国际字符。

当为字段转换选择**自动清理字段名称**时，Splunk 软件会将以下“键清理”规则应用于该转换提取的字段名称：

- 所有非 a-z、A-Z 和 0-9 字符均替换为下划线 (**_**)。
- 从名称中删除所有前导下划线和 0-9 字符。

注意：对于内联字段提取（不需要字段转换组件的字段提取），不能关闭键清理功能。

示例 - 从事件中提取字段名称及其相应的字段值

您可以将**事件格式**属性与设计正确的正则表达式结合使用来设置字段转换，以便从每个匹配事件中同时提取字段名称及其相应的字段值。

以下示例使用了 Splunk 软件所附带的一个转换。

bracket-space 字段转换有一个正则表达式，用于在事件数据中查找括在括号内的字段名称/值对。本转换将重复应用该正则表达式，直到事件中的所有匹配字段/值对都被提取出来为止。

bracket-space
字段 » 字段转换 » bracket-space

正则表达式 *

来源键 *

指定应用正则表达式的键。默认为 `_raw`。

格式

根据字段名称和值指定事件格式。
使用 `$n` (例如 `$1`、`$2` 等)指定正则表达式输出匹配。默认为 `<transform_stanza_name>::$1`

☐ 创建多值字段

如果选中，在已经提取字段时，提取器将创建多值字段。

☒ 自动清除字段名称

如果选中，将清除字段名称，使其只包含 `a-zA-Z0-9_`

正如我们在本主题前面曾经提到的，字段转换始终与字段提取相关联。在 Splunk Web 的“字段提取”页面上，您会看到 **bracket-space** 字段转换与 **osx-as1:REPORT-as1** 提取关联。

更新字段转换权限

首次创建字段转换时，默认情况下该转换仅对其创建者可用。要使该字段转换对其他用户可用，需要更新其**权限**。为此，需要在“字段转换”页面上找到该字段转换，并选择其**权限**链接。随即打开 Splunk Web 中用于**知识对象**的标准权限管理页面。

在此页面中，您可以为字段转换设置**基于角色**的权限，并确定该转换是对一个特定应用的用户可用还是对所有应用的用户全局可用。有关使用 Splunk Web 管理权限的更多信息，请参阅本手册中的[“管理知识对象权限”](#)。

删除字段转换

在 Splunk Web 的“字段转换”页面上，如果您的权限允许，可以删除字段转换。

单击要删除的字段提取所对应的**删除**。

注意：删除具有下游依赖性的知识对象时须谨慎。例如，如果使用您的字段转换所提取的字段的搜索是某个事件类型的基础，并且该事件类型由其他五个报表（其中两个报表是仪表板面板的基础）使用，则将该转换从系统中删除将会对所有其他知识对象产生负面影响。有关删除知识对象的更多信息，请参阅本手册中的[“禁用或删除知识对象”](#)。

通过配置文件创建和维护搜索时间字段提取

虽然可以通过 Splunk Web 来设置和管理搜索时间**字段提取**，但也一定要了解如何在 `props.conf` 和 `transforms.conf` 级别处理此类提取，因为 Splunk Web 的“字段提取”和“字段转换”页面将在这两个配置文件中进行读写操作。

很多知识管理员，尤其是已经使用 Splunk 软件一段时间的管理员，会发现通过配置文件更容易管理自定义字段，因为使用配置文件不但可以添加和维护自定义字段，还可以检查为其团队添加的自定义字段库。配置文件还可为字段提取启用比您使用“设置”页面获得的更广泛的字段提取选项。

本主题向您介绍如何：

- 通过编辑 `props.conf` 设置基本“内联”搜索时间字段提取。
- 通过同时编辑 `props.conf` 和 `transforms.conf` 设计更为复杂的搜索时间字段提取。

正则表达式和字段名称语法

Splunk 软件使用正则表达式从事件数据中提取字段。当您使用交互式字段提取器 (IFX) 时，Splunk 软件会尝试为您生成字段提取正则表达式，但它所创建的正则表达式一次只能从匹配的事件中提取一个字段。

另一方面，如果您通过配置文件手动设置字段提取，您必须自行提供正则表达式，但是可以根据需要对其进行相应设计以使其从匹配的事件中提取两个或更多个字段。

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。

重要提示：正则表达式中捕获组内的字段名称必须包含字母数字字符或下划线。请参阅下文中的“使用正确的字段名称语法”。

使用正确的字段名称语法

字段名称只能包含字母数字字符和下划线：

- 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 Enterprise 的内部变量。
- 不允许使用国际字符。

无论是默认情况还是自定义配置，Splunk 软件都会对在搜索时间提取的所有字段应用以下“键清理”规则：

1. 所有非 **a-z**、**A-Z** 和 **0-9** 字符均替换为下划线 (`_`)。
2. 启用键清理时（默认情况下启用此功能），Splunk Enterprise 会从提取的字段中删除所有前导下划线和 **0-9** 字符。

您可以禁用特定搜索时间字段提取的键清理功能，方法是将其配置为一种高级 REPORT 提取类型，然后在所引用的字段转换段落中加入 `CLEAN_KEYS=false` 设置。有关 REPORT 提取配置的更多信息，请参阅下文。

注意：您不能关闭基本 EXTRACT（仅限 `props.conf`）字段提取配置的键清理功能。

通过编辑 `props.conf` 创建基本搜索时间字段提取

如果您有 Splunk Enterprise，可以通过编辑 `props.conf` 配置文件来创建基本搜索时间字段提取（即完全在 `props.conf` 中定义的字段提取，而不是引用 `transforms.conf` 中**字段转换**的提取）。`props.conf` 位于 `$SPLUNK_HOME/etc/system/local/` 或您自定义的应用目录 `$SPLUNK_HOME/etc/apps/` 中。（如果您希望能够更方便地将数据自定义项传输到其他搜索服务器，建议您使用后一个目录。）

注意：请勿编辑 `$SPLUNK_HOME/etc/system/default/` 中的文件。

有关配置文件的一般详细信息，请参阅《管理员手册》中的“关于配置文件”。

通过 `props.conf` 定义基本搜索时间字段提取的步骤

基本搜索时间字段提取使用 `props.conf` 中的 EXTRACT 提取配置。每个 EXTRACT 提取段落都包含在搜索时间提取一个或多个字段所使用的正则表达式，以及用于控制这些字段提取方式的其他属性。

按以下步骤来创建基本搜索时间字段提取：

1. `props.conf` 中的所有提取配置受特定来源、来源类型或主机限制。首先确定应从中提取字段的事件的来源类型、来源或主机。

注意：有关主机、来源和来源类型的信息，请参阅《数据导入》手册中的“关于默认字段（主机、来源、来源类型等）”。

2. 创建正则表达式以确定事件中的字段。使用命名的捕获组为提取的值提供字段名称。使用先前部分中所述的字段名称语法。

3. 在 `props.conf` 中按 EXTRACT 字段提取类型的格式（在下一节中定义）创建包含所确定的主机/来源/来源类型和

正则表达式的字段提取段落。编辑 `props.conf` 文件（位于 `$SPLUNK_HOME/etc/system/local/` 或您自定义的应用目录 `$SPLUNK_HOME/etc/apps/` 中）。

注意：请勿编辑 `$SPLUNK_HOME/etc/system/default/` 中的文件。

4. 如果您的字段值是一个单词的一部分，则还必须向 `fields.conf` 中添加一个条目。请参阅下文的“从子令牌创建字段”示例。

5. 重新启动 Splunk Enterprise 使更改生效。

向 `props.conf` 添加 `EXTRACT` 字段提取段落

按以下格式将 `EXTRACT` 字段提取添加到 `props.conf` 中：

```
[<spec>]
EXTRACT-<class> = [<regular_expression>|<regular_expression> in <source_field>]
```

- `<spec>` 可以是：
 - `<source type>`，事件的来源类型。
 - `host::<host>`，其中 `<host>` 为事件所在的主机。
 - `source::<source>`，其中 `<source>` 为事件的来源。
 - `rule::<rulename>`，其中 `<rulename>` 为来源类型分类规则的唯一名称。
 - `delayedrule::<rulename>`，其中 `<rulename>` 为延迟来源类型分类规则的唯一名称。

注意：只有在根据 Splunk 软件发现的来源生成新来源类型之前，才会考虑 `rule` 和 `delayedrule`。

- `<class>` 是唯一的文字字符串，用于标识所提取字段（键）的命名空间。
 - **注意：**`<class>` 值不需要遵循[字段名称语法限制](#)（请参阅上文）。您可以使用 a-z、A-Z 和 0-9 以外的其他字符，也可以使用空格。`<class>` 值不受键清理功能的制约。
- `<regular_expression>` 必须包含命名的捕获组；每个组表示一个不同的提取字段。当 `<regular_expression>` 与某一事件匹配时，命名的捕获组及其值会添加到该事件中。
- 使用 `<regular_expression> in <source_field>` 使正则表达式与指定字段的值匹配。否则表达式将与 `_raw`（所有原始事件数据）相匹配。
 - **注意：**`<src_field>` 是字段名称，表示它必须遵循字段名称语法。只能包含字母数字字符（a-z、A-Z 和 0-9）。
- 如果您的正则表达式要以 `in <string>` 结尾，且 `<string>` **不是** 字段名称，应将正则表达式更改为以 `[i]n <string>` 结尾，以确保 Splunk 软件不会尝试将 `<string>` 与字段名称匹配。

适用于 `EXTRACT` 字段提取类型的优先级规则：

- 对于每个字段提取，Splunk 软件会从最高优先级的配置段落中获取配置。
- 如果存在多个类别的匹配 `[<spec>]` 段落，`[host::<host>]` 设置将覆盖 `[<sourcetype>]` 设置。
- `[source::<source>]` 设置将同时覆盖 `[host::<host>]` 和 `[<sourcetype>]` 设置。
- 同样，如果在 `../local/` 中为 `<spec>` 指定了特定字段提取，该提取将覆盖 `../default/` 中的相应类。

`[<spec>]` 段落还有很多优先级规则；请参阅 `props.conf.spec` 了解所有详细信息。

注意：与配置在索引时间提取的默认字段集的过程不同，`transforms.conf` 不需要 `DEST_KEY`，因为在搜索时间字段提取期间并不会向索引写入任何内容。在搜索时间提取的字段不会以键的形式保留在索引中。

Splunk 软件将按照优先级规则来运行搜索时间字段提取。它会首先运行内联字段提取 (`EXTRACT-<class>`)，然后运行引用字段转换 (`REPORT-<class>`) 的字段提取。

为搜索时间数据设置 `KV_MODE`

您可以使用 `KV_MODE` 属性为您的数据指定字段/值提取模式。可以将 `KV_MODE` 添加到 `EXTRACT` 或 `REPORT` 段落中。其格式为：

```
KV_MODE = [none|auto|auto_escaped|multi|json|xml]
```

KV_MODE 值	描述
none	禁用段落名称所标识的来源、来源类型或主机的字段提取。您可以使用此设置来确保已创建的其他正则表达式不会被特定来源、来源类型或主机的自动字段/值提取所覆盖。另外，还可以使用此设置来禁用常见但不重要的字段的提取，以提高搜索性能。本主题的末尾提供了一些字段提取示例，演示了如何在各种不同的情况下禁用字段提取。
auto	提取字段/值对并用等号分隔这些值。如果您的字段提取段落中不包含此属性，这将是默认的字段提取行为。
auto_escaped	提取字段/值对并用等号分隔这些值。另外，此设置可确保 Splunk 软件优先采用 <code>\</code> 和 <code>\\</code> 作为加引号的值中的转义序列。例如： <code>field="value with \"nested\" quotes"</code> 。

multi	这将调用 <code>multikv</code> 搜索命令，该命令将从表格形式的事件中提取字段值。
xml	如果您要使用字段提取段落来从 XML 数据中提取字段，应使用此设置。此模式不会提取非 XML 数据。
json	<p>如果您要在搜索时间使用字段提取段落来从 JSON 数据中提取字段，应使用此设置。此模式不会提取非 JSON 数据。</p> <p>注意：如果要设置 <code>KV_MODE = json</code>，确保没有为同一来源类型同时设置 <code>INDEXED_FIELDS = JSON</code>。如果两个同时设置会导致 JSON 字段被提取两次，一次在索引时间而另一次在搜索时间。</p>

内联（仅使用 `props.conf`）搜索时间字段提取示例

下面是一组仅使用 `props.conf` 设置的搜索时间自定义字段提取的示例。

添加新的错误代码字段

本例介绍如何通过在 `props.conf` 中配置字段提取来创建新的“错误代码”字段。您可以通过 `device_id=` 后跟带括号的单词和一个以冒号结尾的文本字符串来找到此字段。应从与 `testlog` 来源类型相关的事件中提取此字段。

在 `props.conf` 中，添加以下内容：

```
[testlog]
EXTRACT-errors = device_id=\[w+\](?<err_code>[^\:]+)
```

使用一个正则表达式提取多个字段

本字段提取示例将提取五个单独的字段。然后，您可以将这些字段与某些事件类型结合使用，以帮助您查找端口不断开关的事件并进行报告。

以下为从中提取字段的事件数据示例：

```
#%LINEPROTO-5-UPDOWN: Line protocol on Interface GigabitEthernet9/16, changed state to down
```

`props.conf` 中此提取所对应的段落类似如下所示：

```
[syslog]
EXTRACT-port_flapping = Interface\s(?<interface>(?(media>[^\d]+) (?(slot>\d+)\/(?(port>\d+)))\,\schanged
\sstate\sto\s(?(port_status>up|down)
```

注意，将提取五个单独的字段，提取出的组分别命名为：`interface`、`media`、`slot`、`port` 和 `port_status`。

字段提取不需要这两个步骤，这两个步骤只是告诉您对提取的字段进行哪些操作，以便找到端口不断开关的事件并进行报告。

使用标记在 `eventtypes.conf` 中定义几种事件类型：

```
[cisco_ios_port_down]
search = "changed state to down"

[cisco_ios_port_up]
search = "changed state to up"
```

最后，在 `savedsearches.conf` 中创建报表以将上述大部分事件类型结合在一起，进而找到不断开关的端口并报告结果：

```
[port flapping]
search = eventtype=cisco_ios_port_down OR eventtype=cisco_ios_port_up starthoursago=3 | stats count by
interface,host,port_status | sort -count
```

从子令牌创建字段

如果您所提取的字段值是一个子令牌（即一个更长令牌的一部分），可能会遇到问题。令牌是在建立索引之前通过事件处理运行的事件数据块。在事件处理期间，事件将被分为若干段，即创建令牌，所创建的每个段都是一个令牌。

令牌决不会小于一个完整的单词或数字。例如，您的事件可能包含 `foo123` 一词。如果已经过事件处理和索引，则该词即为一个令牌，而且可以是一个字段值。但是，如果您的提取是提取 `foo` 作为字段值，此时提取的就是子令牌。问题在于，虽然索引中存在 `foo123`，但并不存在 `foo`，也就是说，如果您搜索该子令牌，可能获得的结果很少，即使看

似在搜索结果中提取正确也是如此。

令牌不能小于字符串内的单个“单词”，因此在执行子令牌（单词的一部分）字段提取时，由于子令牌本身并不存在于索引中，而只是一个更大单词的一部分，可能会引发问题。

如果字段值是一个标记的较小部分，则必须按照上述说明配置 `props.conf`。然后，向 `fields.conf` 添加一个条目：

```
[<fieldname>]
INDEXED = False
INDEXED_VALUE = False
```

- 在 `<fieldname>` 中填入字段名称。
 - 例如，如果已经配置了一个名为 `"url"` 的字段，则填入 `[url]`。
- 将 `INDEXED` 和 `INDEXED_VALUE` 设置为 `false`。
 - 此设置表示所搜索的值不是索引中的令牌。

注意：自版本 4.3 开始，您不再需要将此条目添加到 `fields.conf` 中，前提是从某一未索引（因此未标签化）的默认字段（如 `host`、`source`、`sourcetype` 或 `timestamp`）中提取字段值。

有关事件数据标签化的更多信息，请参阅《数据导入》手册中的“关于分段”。

通过字段转换创建高级搜索时间字段提取

如果您有 Splunk Enterprise，可以在 `props.conf` 中即可完成大部分搜索时间字段提取的定义，但是有些高级搜索时间字段提取需引用其他组件（称为**字段转换**）。本部分向您介绍如何在 `transforms.conf` 中配置字段转换。

字段转换包含字段提取正则表达式以及用于控制转换提取字段的方式的其他属性。字段转换始终与 `props.conf` 中的字段提取段落一同创建，而不能单独创建。

在以下情况下，搜索时间字段提取需要字段转换组件：

- **在多个数据来源、来源类型或主机之间重复使用同一字段提取正则表达式**（即，为多个字段提取配置一个字段转换）。如果您发现自己需要使用同一正则表达式来提取不同来源、来源类型和主机的字段，可能希望将该表达式设置为一个转换。然后，如果您需要更新该正则表达式，只需执行一次此操作，即使此表达式由多个字段提取使用。
- **对同一个数据来源、来源类型或主机应用多个字段提取正则表达式**（即，将多个字段转换应用于同一字段提取）。有时，如果您要从特定来源/来源类型/主机中提取的字段以两个或更多个截然不同的事件模式显示，这是非常必要的。
- **设置基于分隔符的字段提取**。如果您的事件数据中有以逗号、冒号、竖线、换行符和制表符等分隔符进行分隔的字段/值对（或只是字段值），此时基于分隔符的提取将会很有用。
- **为多值字段配置提取**。执行此操作时，Splunk 软件会将事件数据中找到的其他字段值附加到字段中。
- **提取名称以数字或下划线开头的字段**。通常，键清理功能将删除字段名称中的前导数字字符和下划线，但是您可以根据需要将转换配置为禁用此功能。

还可以将转换配置为：

- 从其他字段（非 `_raw`）的值中通过 `SOURCE_KEY` 属性提取字段。
- 使用 `FORMAT` 属性管理所提取字段的格式设置（如果是提取多个字段或同时提取字段名称和字段值）。

现在，这两种配置也可以直接在正则表达式中进行设置。有关如何执行上述操作的更多信息，请参阅下文的“定义字段转换”部分。

注意：如果您需要将一组正则表达式提取连接成单个字段值，您可以使用 `FORMAT` 属性来实现，但这仅适用于该属性被设置为索引时间提取的情况。例如，如果您的事件数据中包含诸如 `192(x)0(y)2(z)1` 之类的字符串，可以在索引时间提取该字符串作为 `ip address` 字段值，格式为 `192.0.2.1`。有关更多信息，请参阅《数据导入》手册中的“配置索引时间字段提取”。但是，我们建议不要对索引字段集进行大量更改，若有必要进行更改，应少量更改。

定义引用字段转换的自定义搜索时间字段提取的步骤

高级搜索时间字段提取使用 `props.conf` 中的 `REPORT` 提取配置。每个 `REPORT` 提取段落都引用一个字段转换，该转换在 `transforms.conf` 中单独定义。字段转换包含 Splunk Enterprise 在搜索时提取字段所用的正则表达式，以及用于控制转换提取这些字段的方式的其他属性。

按以下步骤来创建高级搜索时间字段提取：

1. `props.conf` 中的所有提取配置受特定来源、来源类型或主机限制。首先确定应从中提取字段的事件的来源类型、来源或主机。（先不要更新 `props.conf`。）

注意：有关主机、来源和来源类型的信息，请参阅《数据导入手册》中的“关于默认字段（主机、来源、来源类型等）”。

2. 创建正则表达式以确定事件中的字段。使用命名的捕获组为提取的值提供字段名称。使用先前部分中所述的字段名称语法。

注意：如果您的事件列出字段/值对或仅列出字段值，可以创建基于分隔符的字段提取（不需要正则表达式）；有关更多信息，请参阅下文 `DELIMS` 属性的相关信息。）

3. 在 `transforms.conf` 中创建使用此正则表达式（或分隔符配置）的字段转换。此转换可以定义源键和/或事件值格式。

编辑 `transforms.conf` 文件（位于 `$SPLUNK_HOME/etc/system/local/` 或您自定义的应用目录 `$SPLUNK_HOME/etc/apps/` 中）。

注意：请勿编辑 `$SPLUNK_HOME/etc/system/default/` 中的文件。

4. 遵循 REPORT 字段提取类型（在下面的两个部分中定义）的格式，以便在使用步骤 1 确定的主机、数据来源或来源类型的 `props.conf` 中创建字段提取段落。如有必要，可以为引用同一字段转换的其他主机、来源和来源类型创建其他字段提取段落。

编辑 `props.conf` 文件（位于 `$SPLUNK_HOME/etc/system/local/` 或您自定义的应用目录 `$SPLUNK_HOME/etc/apps/` 中）。

注意：请勿编辑 `$SPLUNK_HOME/etc/system/default/` 中的文件。

5. 重启 Splunk 部署使修改生效。

首先，定义字段转换

按以下格式在 `transforms.conf` 中定义搜索时间字段转换：

```
[<unique_transform_stanza_name>]
REGEX = <regular expression>
FORMAT = <string>
SOURCE_KEY = <string>
DELIMS = <quoted string list>
FIELDS = <quoted string list>
MV_ADD = [true|false]
CLEAN_KEYS = [true|false]
KEEP_EMPTY_VALS = [true|false]
CAN_OPTIMIZE = [true|false]
```

- `<unique_transform_stanza_name>` 对于所有搜索时间转换都是必需的。**注意：**`<unique_transform_stanza_name>` 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。这些值不受键清理功能的制约。
- `REGEX` 是处理数据以提取字段的正则表达式。它对于所有搜索时间字段转换都是必需的，除非您设置的是基于分隔符的交易，此时应改用 `DELIMS` 属性（请参阅下文的 `DELIMS` 属性描述）。
 - 默认为空字符串。
- `REGEX` 和 `FORMAT` 属性：
 - `REGEX` 中的名称捕获组将直接提取到字段，这意味着简单的字段提取无需指定 `FORMAT`。
 - 如果 `REGEX` 提取字段名称和其相应值，可以使用以下特殊的捕获组以跳过在 `FORMAT` 中指定映射的操作：

```
<_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]+)
```

- 在这两种情况中，正则表达式都会反复应用于事件的源文本，以提取它可以找到的所有字段/值组合。
- `FORMAT` 为可选。用于指定所提取的字段/值对的格式。无需指定 `FORMAT`，如果是包含名称捕获组的简单 `REGEX`。
 - 对于搜索时间提取，`FORMAT` 字段的模式如下：

```
FORMAT = <field-name>::<field-value>(<field-name>::<field-value>)*
```

其中：

```
field-name = [<string>|<$<extracting-group-number>>]
field-value = [<string>|<$<extracting-group-number>>]
```

搜索时间 `FORMAT` 用法示例：

```
1.FORMAT = firstfield::$1 secondfield::$2 thirdfield::other-value
```

2.FORMAT = \$1::\$2

- 如果使用可变字段名称配置 `FORMAT`（如上面的示例 2，`$1` 表示字段名称），正则表达式会反复应用于源事件文本，以匹配并提取它可以找到的所有字段/值对。
 - **注意：**在搜索时，不能使用 `FORMAT` 创建连接的字段。此功能仅适用于索引时间字段转换。
 - `FORMAT` 默认为空字符串。
- `SOURCE_KEY` 为可选。用于从其他字段的值中提取一个或多个值。您可以使用执行此字段提取时可用的任意字段。
 - 要配置 `SOURCE_KEY`，请确定转换的 `REGEX` 要应用到哪个字段。
 - 默认情况下，`SOURCE_KEY` 设置为 `_raw`，这意味着它将应用于所有事件的未经处理的原始文本。
- `DELIMS` 为可选。可代替 `REGEX` 来处理基于分隔符的字段提取（即字段值或字段/值对以逗号、冒号、空格、制表符、换行符等分隔符进行分隔）。
 - 分隔符两侧必须有引号 `"`。如果需要，可以使用反斜线来转义值两侧的双引号 `(\)`。
 - 重要提示：如果值可能包含嵌入的非转义双引号字符，如 `"foo"bar"`，我们建议您使用 `REGEX` 而不是 `DELIMS`。
 - 分隔符字符串中的每个字符将用作拆分事件的分隔符。
 - 如果事件包含完整的以分隔符分隔的字段/值对，应为 `DELIMS` 输入两组带引号的分隔符。第一组带引号的分隔符用于分隔字段/值对。第二组带引号的分隔符用于分隔字段名称与其相应值。
 - 如果事件只包含以分隔符分隔的值（无字段名称），应使用一组带引号的分隔符来分隔这些值。然后使用 `FIELDS` 属性将字段名称应用于所提取的值（请参阅下文的 `FIELDS`）。或者，Splunk 软件将偶数令牌读取为字段名称，而将奇数令牌读取为字段值。
 - 如果您未指定字段名称的列表，Splunk 软件将使用连续分隔符字符。
 - 默认为空字符串。
 - 本示例中的 `DELIMS` 用法适用于字段/值对以 `|` 符号分隔，而字段名称以 `=` 符号与其相应值分隔的事件：

```
[pipe_eq]
DELIMS = "|", "="
```

- `FIELDS` 与 `DELIMS` 结合使用可执行基于分隔符的字段提取，但仅提取字段值。使用 `FIELDS` 可按照值的提取顺序为所提取的字段值以列表格式提供字段名称。
 - **注意：**如果字段名称包含空格或逗号，则必须在两侧添加引号 `"`（要进行转义，应使用 `\`）。
 - 默认为空字符串。
 - 以下是基于分隔符的提取示例，其中的事件有三个字段值。这些值用逗号后跟空格进行分隔。

```
[commalist]
DELIMS = ", "
FIELDS = field1, field2, field3
```

- `MV_ADD` 为可选。如果在您的事件中同一字段出现多次但每次的值有所不同，而且您希望保留字段的每个值，应使用此属性。
 - 对于 `MV_ADD = true`，Splunk 软件会将事件中多次以不同值出现的字段转换为多值字段（即字段名称出现一次，`=` 符号后跟此字段的多个值）。
 - 对于 `MV_ADD=false`，Splunk 软件将保留为事件中某一字段找到的第一个值，而丢弃为此事件中为此字段找到的所有后续值。
 - 默认为 `false`。
- `CLEAN_KEYS` 为可选。用于控制系统是否从提取的键（字段名称）中去除前导下划线和 0-9 字符（有关更多信息，请参阅上文的子主题“[使用正确的字段名称语法](#)”）。“键清理”功能会将字段名称中的所有非字母数字字符（非 a-z、A-Z 和 0-9 范围内的字符）替换为下划线，以及从字段名称中去除前导下划线和 0-9 字符。
 - 如果您需要保持字段名称不变（不删除前导下划线和/或 0-9 字符），应将 `CLEAN_KEYS = false` 添加到转换中。
 - 默认情况下，始终为转换将 `CLEAN_KEYS` 设置为 `true`。
- `KEEP_EMPTY_VALS` 为可选。用于控制当值为空字符串时 Splunk 软件是否保留字段/值对。
 - 此选项不适用于由 Splunk 软件的 `autokv` 提取（自动字段提取）过程所生成的字段/值对。Autokv 将忽略含有空值的字段/值对。
 - 默认为 `false`。
- `CAN_OPTIMIZE` 为可选。用于控制 Splunk 软件是否可以优化提取（或者换言之，是否禁用提取）。
 - 如果您在运行搜索时搜索模式的设置禁用了字段发现（此功能可确保 Splunk 软件始终发现特定字段），则可能需要使用此属性。
 - 只有可以确定成功计算搜索不需要提取过程找到的任何字段时，Splunk 软件才会禁用提取。
 - **注意：**此属性应尽量不要设置为 `false`。
 - 默认为 `true`。

其次，配置 `props.conf` 中的 `REPORT` 字段提取段落并将其与字段转换进行关联

在 `props.conf` 中设置与字段转换关联的搜索时间字段提取时，使用 `REPORT` 字段提取类。遵循以下格式。

您可以将多个字段转换段落关联到一个字段提取，方法是在初始的 `<unique_transform_stanza_name>` 之后列出这些段落

并用逗号进行分隔。（有关更多信息，[请参阅本主题后面的示例。](#)）

```
[<spec>]
REPORT-<class> = <unique_transform_stanza_name1>, <unique_transform_stanza_name2>,...
```

- `<spec>` 可以是：
 - `<sourcetype>`，事件的来源类型。
 - `host::<host>`，其中 `<host>` 为事件所在的主机。
 - `source::<source>`，其中 `<source>` 为事件的来源。
- `<class>` 是唯一的文字字符串，用于标识所提取字段（键）的命名空间。**注意：**`<class>` 值不需要遵循[字段名称语法限制](#)（请参阅上文）。您可以使用 a-z、A-Z 和 0-9 以外的其他字符，也可以使用空格。`<class>` 值不受键清理功能的制约。
- `<unique_transform_stanza_name>` 是来自 `transforms.conf` 的字段转换段落的名称。
- **REPORT 字段提取类的优先级规则：**
 - 对于每个类，Splunk 软件会从最高优先级的配置块中获取配置。
 - 如果为 `source` 和 `sourcetype` 指定了特定类，则 `source` 的类将优先。
 - 同样，如果在 `../local/` 中为 `<spec>` 指定了特定类，该类将覆盖 `../default/` 中的相应类。

如果您的一组转换必须按特定顺序运行并且属于同一主机、来源或来源类型，可以逗号分隔列表的形式将这组转换放入同一 `props.conf` 段落中。转换会以指定的顺序进行应用。例如，以下序列可确保首先应用 `[yellow]` 字段转换，然后依次应用 `[blue]` 和 `[red]`：

```
[source::color_logs]
REPORT-colorchange = yellow, blue, red
```

如果需要更改顺序，可重新排列列表。

自定义搜索时间字段提取（使用字段转换）示例

以下示例显示自定义字段提取使用案例，这些案例需要在 `transforms.conf` 中配置一个或多个字段转换段落，然后在 `props.conf` 字段提取段落中引用这些转换。

配置使用多个字段转换的字段提取

本搜索时间字段转换设置示例演示如何：

- 创建从事件中提取不同字段名称/值对的转换。
- 创建引用两个或更多个字段转换的字段提取。

假设您的日志包含多个字段名称/字段值对。虽然字段在各事件之间不同，但是字段名称/值对始终按照以下两种格式之一显示。

日志通常为以下格式：

```
[fieldName1=fieldValue1] [fieldName2=fieldValue2]
```

不过，有时日志会较为复杂一些，即以列表形式记录多个名称/值对，此时格式类似如下所示：

```
[headerName=fieldName1] [headerValue=fieldValue1], [headerName=fieldName2] [headerValue=fieldValue2]
```

请注意，列表项目用逗号分隔，并且每个 `fieldName` 均与相应的 `fieldValue` 匹配。在第二种情况中，您仍然需要提取字段名称和值，使搜索结果

```
fieldName1=fieldValue1
fieldName2=fieldValue2
```

等等。

为了更清楚地说明，下面的 HTTP 请求事件示例结合了上述两种格式。

```
[method=GET] [IP=10.1.1.1] [headerName=Host] [headerValue=www.example.com], [headerName=User-Agent]
[headerValue=Mozilla], [headerName=Connection] [headerValue=close] [byteCount=255]
```

您要开发单个字段提取以从该事件中获取以下字段/值对：

```
method=GET
IP=10.1.1.1
Host=www.example.com
User-Agent=Mozilla
Connection=close
```

```
byteCount=255
```

解决方案

为了有效、可靠地提取两种格式的字段/值对，您将需要设计两个不同的正则表达式并针对每种格式进行优化。一个正则表达式将确定具有第一种格式的事件并提取所有匹配字段/值对。另一个正则表达式将确定具有另一种格式的事件并提取这些字段/值对。

然后，在 `transforms.conf` 中创建两个唯一的转换（每个正则表达式对应一个转换），然后将它们合并到 `props.conf` 中的相应字段提取段落中。

添加到 `transforms.conf` 中的第一个转换用于获取相当常规的 `[fieldName=fieldValue1] [fieldName2=fieldValue2]` 实例。

```
[myplaintransform]
REGEX=[ (?!(?:headerName|headerValue)) ([^\s\=]+)\s*([^\s\=]+)\s*]
FORMAT=$1::$2
```

第二个转换（也添加到 `transforms.conf` 中）用于获取略微复杂的 `[headerName=fieldName1] [headerValue=fieldValue1], [headerName=fieldName2] [headerValue=fieldValue2]` 实例：

```
[mytransform]
REGEX= \[headerName=(\w+)\],\s\[headerValue=([^\s\=]+)\s*\]
FORMAT= $1::$2
```

两个转换都使用 `<fieldName>::<fieldValue> FORMAT` 将事件中的每个字段名与其相应值进行匹配。FORMAT 中的设置使 Splunk 软件可以一直根据匹配事件匹配正则表达式，直到提取所有匹配的字段/值组合为止。

最后，以下字段提取段落（在 `props.conf` 中创建）将引用两个字段转换：

```
[mysourcetype]
KV_MODE=none
REPORT-a = mytransform, myplaintransform
```

请注意，除了使用多个字段转换之外，此字段提取段落还设置了 `KV_MODE=none`。这将禁用所确定来源类型的自动字段/值提取（同时允许您手动定义的提取继续进行）。从而可确保这些新正则表达式不会被自动字段提取覆盖，而且还有助于提高您的搜索性能。（有关禁用键/值提取的更多信息，请参阅以下小节。）

配置基于分隔符的字段提取

您可以在字段转换中使用 `DELIMS` 属性来为字段值或字段/值对用逗号、冒号、制表符等分隔符进行分隔的事件配置字段提取。

例如，假设您有一个重复性多行事件，其中的不同字段/值对位于单独行上，并且各个字段/值对用冒号后跟制表符的形式进行分隔。下面是一个示例事件：

```
ComponentId:      Application Server
ProcessId:        5316
ThreadId:         00000000
ThreadName:       P=901265;O=0:CT
SourceId:         com.ibm.ws.runtime.WsServerImpl
ClassName:
MethodName:
Manufacturer:     IBM
Product:          WebSphere
Version:          Platform 7.0.0.7 [BASE 7.0.0.7 cf070942.55]
ServerName:       sfeserv36Node01Cell\sfeserv36Node01\server1
TimeStamp:        2010-04-27 09:15:57.671000000
UnitOfWork:
Severity:         3
Category:         AUDIT
PrimaryMessage:   WSVR0001I: Server server1 open for e-business
ExtendedMessage:
```

现在，您可以在 `props.conf` 中设置一个庞大、冗长的搜索时间字段提取段落来处理所有这些字段：

```
[activityLog]
LINE_BREAKER = [-]{8,}([\\r\\n]+)
SHOULD_LINEMERGE = false
```

```
EXTRACT-ComponentId = ComponentId:\t(?:.*)
EXTRACT-ProcessId = ProcessId:\t(?:.*)
EXTRACT-ThreadId = ThreadId:\t(?:.*)
EXTRACT-ThreadName = ThreadName:\t(?:.*)
EXTRACT-SourceId = SourceId:\t(?:.*)
EXTRACT-ClassName = ClassName:\t(?:.*)
EXTRACT-MethodName = MethodName:\t(?:.*)
EXTRACT-Manufacturer = Manufacturer:\t(?:.*)
EXTRACT-Product = Product:\t(?:.*)
EXTRACT-Version = Version:\t(?:.*)
EXTRACT-ServerName = ServerName:\t(?:.*)
EXTRACT-TimeStamp = TimeStamp:\t(?:.*)
EXTRACT-UnitOfWork = UnitOfWork:\t(?:.*)
EXTRACT-Severity = Severity:\t(?:.*)
EXTRACT-Category = Category:\t(?:.*)
EXTRACT-PrimaryMessage = PrimaryMessage:\t(?:.*)
EXTRACT-ExtendedMessage = ExtendedMessage:\t(?:.*)
```

但是，该解决方案过于繁琐了。是否有更好的处理方法而不需使用这些 `EXTRACT` 行？是的，有！

在 `transforms.conf` 中配置以下段落：

```
[activity_report]
DELIMS = "\n", ":\t"
```

该段落声明事件中的字段/值对位于单独行上（“\n”），然后指定每一行上的字段名称和字段值用冒号和制表符（“:\t”）进行分隔。

要完成此配置，应将上文所述的冗长 `props.conf` 段落重写为：

```
[activitylog]
LINE_BREAKER = [-]{8,}([\r\n]+)
SHOULD_LINEMERGE = false
REPORT-activity = activity_report
```

这两个简洁的配置将提取与之前相同的一组字段，但是降低了出错的几率，而且更为灵活。

处理具有多值字段的事件

如果同一字段在事件中使用多次，但每次都使用不同的值，此时可以使用 `MV_ADD` 属性来提取字段。通常，Splunk 软件会提取某个字段在事件中出现的第一个实例，丢弃所有后续实例。但是，如果在 `transforms.conf` 中将 `MV_ADD` 设置为 `true`，Splunk 软件会将该字段视为多值字段，并提取事件中的所有唯一字段/值对。

假设您有类似如下所示的一组事件：

```
event1.epochtime=1282182111 type=type1 value=value1 type=type3 value=value3
event2.epochtime=1282182111 type=type2 value=value4 type=type3 value=value5 type=type4 value=value6
```

了解 `type` 和 `value` 字段是如何在每个事件中多次重复出现的？您想要做的是搜索 `type=type3` 并返回这两个事件。或者对这两个事件运行 `count(type)` 报表以返回 5。

因此，您要执行的操作是为这些事件创建 `type` 字段的自定义多值提取。下面介绍如何设置 `transforms.conf` 和 `props.conf` 文件来启用：

首先，`transforms.conf`：

```
[mv-type]
REGEX = type=(?<type>\s+)
MV_ADD = true
```

然后，在来源类型或数据源的 `props.conf` 中设置：

```
REPORT-type = mv-type
```

禁用特定来源、来源类型或主机的自动搜索时间提取

您可以通过编辑 `props.conf` 禁用特定来源、来源类型或主机的自动搜索时间字段提取。添加 `KV_MODE = none`，添加对象为 `[<spec>]`（位于 `props.conf`）。

注意：如果 `KV_MODE = none`，对于受影响的来源、来源类型或主机，手动通过配置文件或 Splunk Web 设置的自定

义字段提取仍将受到处理。

```
[<spec>]
KV_MODE = none
```

<spec> 可以是：

- <sourcetype> - 事件来源类型。
- host::<host>，其中 <host> 为事件所在的主机。
- source::<source>，其中 <source> 为事件的来源。

通过 fields.conf 配置多值字段

多值字段是指在某一事件中多次出现并且每次出现时均具有不同值的字段。多值字段的一个较为常见的示例是电子邮件地址字段，该字段通常会在单个 sendmail 事件中出現兩到三次--一次用于发件人，另一次用于收件人列表，如果存在抄送地址，第三次可能会出现在抄送地址列表中。如果所有这些字段都具有相同的标签（例如，"AddressList"），它们将失去当被分别标识为 "From"、"To" 和 "Cc" 时本应拥有的含义。

多值字段会在搜索时间进行分析，以便您能在搜索管道中处理这些值。用于处理多值字段的搜索命令包括 makemv、mvcombine、mvexpand 和 nomv。有关这些命令和其他命令的详细信息，请参阅《搜索手册》中的操作多值字段相关主题。《搜索参考》手册中提供了完整的命令参考。

在 fields.conf 中使用 TOKENIZER 键配置多值字段。TOKENIZER 使用正则表达式来指示 Splunk 软件如何识别并提取事件中某一重复性字段的多个字段值。如果您有 Splunk Enterprise，可以编辑 \$SPLUNK_HOME/etc/system/local/ 中或您自己的自定义应用目录 \$SPLUNK_HOME/etc/apps/ 中的 fields.conf。

有关配置文件的一般详细信息，请参阅《管理员手册》中的“关于配置文件”。

有关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。

通过 fields.conf 配置多值字段

如果您有 Splunk Enterprise，可以在 fields.conf 中添加一个段落来定义多值字段。然后添加一个包含 TOKENIZER 键和相应正则表达式的行，以显示该字段如何取得多个值。

注意：如果要为多值字段设置其他属性，应在此相同段落中的 TOKENIZER 行下方设置这些属性。有关更多信息，请参阅《管理员手册》中的 fields.conf 主题。

```
[<field name 1>]
TOKENIZER = <regular expression>
```

```
[<field name 2>]
TOKENIZER = <regular expression>
```

- <regular expression> 应指示相关字段如何取得多个值。
- TOKENIZER 默认为空。如果 TOKENIZER 为空，该字段只能取得单个值。
- 否则，从每个匹配项中获取第一个组，从而形成字段值集。
- TOKENIZER 键适用于 where、timeline 和 stats 命令。它还可提供异步搜索 API 的摘要和 XML 输出。

注意：不支持对索引字段（在索引时提取的字段）进行标签化。如果您已经为某字段设置 INDEXED=true，则不能同时对该字段使用 TOKENIZER 键。可以使用在 props.conf 和 transforms.conf 中定义的搜索时间提取将索引字段拆分为多个值。

示例

假设您有一个格式糟糕的电子邮件日志文件，其中包含的所有地址被都分组到 AddressList 下：

```
From: sender@splunkexample.com
To: recipient1@splunkexample.com, recipient2@splunkexample.com, recipient3@splunkexample.com
CC: cc1@splunkexample.com, cc2@splunkexample.com, cc3@splunkexample.com
Subject: Multivalue fields are out there!
X-Mailer: Febooti Automation Workshop (Unregistered)
Content-Type: text/plain; charset=UTF-8
Date: Wed, 3 Nov 2014 17:13:54 +0200
X-Priority: 3 (normal)
```

该示例（来自 \$SPLUNK_HOME/etc/system/README/fields.conf.example）将 email 字段 To、From 和 CC 拆分为多个值。

```
[To]
TOKENIZER = (\w[\w\.\-]*@[\w\.\-]*\w)

[From]
TOKENIZER = (\w[\w\.\-]*@[\w\.\-]*\w)

[Cc]
TOKENIZER = (\w[\w\.\-]*@[\w\.\-]*\w)
```

关于已计算字段

已计算字段指在搜索时间添加到事件中的字段。这些字段以事件中已有的两个或多个字段的值来运行计算。将已计算字段作为一种快捷方式用于以 `eval` 命令执行重复的、较长的或复杂的转换。

利用 `eval` 命令可以编写表达式，使此表达式使用**提取的字段**并创建一个以此表达式的计算结果作为值的新字段。有关更多信息，请参阅 [eval](#)。

`Eval` 表达式也可以很复杂。如果您经常要使用一个长而复杂的 `eval` 表达式，每次都要准确地键入该表达式是一件乏味的工作。

已计算字段使您可以使用 `eval` 表达式来定义字段。在编写搜索时，您可以将整个 `eval` 表达式去掉，并按照其他提取的字段的引用方式来引用该字段。这些字段会在搜索时间进行提取并添加到将这些字段包含在 `eval` 表达式的事件中。

您可以在 Splunk Web 中和 `props.conf` 中创建已计算字段。有关在 Splunk Web 中创建已计算字段的信息，请参阅[通过 Splunk Web 创建已计算字段](#)。有关通过 `props.conf` 创建已计算字段的信息，请参阅[通过 props.conf 创建已计算字段](#)。

已计算字段限制

当您运行一个搜索时，Splunk 软件会运行多个操作以派生各种知识对象并将他们应用到此搜索返回的事件中。因为这些操作会以一定顺序进行应用，如果一个定义引用了处理顺序中较为靠后的操作的配置结果，则无法将此定义配置给处理顺序中较为靠前的操作。已计算字段操作位于搜索时间操作顺序的中段，所以不可引用查找、事件类型或标记。

有关更多信息，请参阅[搜索时间操作顺序](#)。

防止覆盖现有字段

如果已计算字段的名称与某一通过普通方式提取的字段名称相同，则此已计算字段将覆盖提取的字段，即使 `eval` 语句的计算结果为空值也是如此。您可以通过将 `eval` 的 `coalesce` 函数与 `eval` 表达式配合使用来取消此类覆盖。`Coalesce` 可获取任意数量的参数并返回第一个非空值。

如果您不希望当 `eval` 语句返回值时已计算字段覆盖现有字段，则使用：

```
EVAL-field = coalesce(field, <eval expression>)
```

如果您不希望当 `eval` 语句返回空值时已计算字段覆盖现有字段，则使用：

```
EVAL-field = coalesce(<eval expression>, field)
```

有关 `coalesce` 和其他 `eval` 函数的更多信息，请参阅《搜索参考》中的[评估函数](#)。

已计算字段的独立性

当 Splunk 软件对已计算字段进行求值时，它会独立于所有其他表达式来计算每个表达式。您不能将各个已计算字段表达式“链接”在一起，因为一个已计算字段的求值结果将用于另一个已计算字段的表达式。

在以下示例中，对于任何单个事件，`x` 的值与已计算字段 `y` 的值相等，因为两个计算是彼此独立进行的。两个表达式都是使用 `x` 来计算 `x*2`。

```
[<foo>]
EVAL-x = x * 2
EVAL-y = x * 2
```

对于特定事件，如果 `x=4`，这些已计算字段会将 `x` 的值替换为 `8`，并将 `y=8` 添加到事件中。

涉及提取的字段 `response_time` 的另一个示例。第一次提取时，`response_time` 值的单位是毫秒。以下是两个已计算字段，它们以不同的方式使用 `response_time`。

```
[<access_common>]
EVAL-response_time = response_time/1000
EVAL-bitrate = bytes*1000/response_time
```

在本例中，`access_common` 来源类型发生了两件事情。

- 第一个 EVAL 更改 `response_time` 值（所有 `sourcetype=access_common` 事件中的该值），以将该值单位从毫秒更改为秒。新值（秒）将覆盖旧值（毫秒）。
- 第二个 EVAL 计算一个名为 `bitrate` 的新字段（针对所有 `sourcetype=access_common` 事件）。该字段单位为 bytes 每秒。Bytes 是另一个提取的字段。

在两个计算中，`response_time` 的最初单位都是毫秒，因为两个 EVAL 是彼此独立计算的。

通过 Splunk Web 创建已计算字段

通过 Splunk Web 创建已计算字段。

前提条件

- 要了解有关已计算字段的信息，请参阅[关于已计算字段](#)。

在“设置”中创建一个新的已计算字段

1. 选择设置 > 字段。
2. 选择已计算字段 > 新建。
3. 选择要使用已计算字段的应用。
4. 选择要应用到已计算字段的主机、来源或来源类型，并指定名称。
5. 为已计算字段命名。
6. 定义 eval 表达式。

通过 props.conf 配置已计算字段

要创建已计算字段，将已计算字段键添加到新的或现有 `props.conf` 段落中。`props.conf` 位于 `$SPLUNK_HOME/etc/system/local/` 或您自定义的应用目录 `$SPLUNK_HOME/etc/apps/` 中。将您的数据自定义传送给其他搜索服务器的最佳方式是使用您自己的自定义应用目录。

请勿编辑 `$SPLUNK_HOME/etc/system/default/` 中的文件。

有关配置文件的更多信息，请参阅“关于配置文件”。

在 `props.conf` 中，已计算字段键的格式为：

```
[<stanza>]
EVAL-<field_name> = <eval statement>
```

- `<stanza>` 可以是：
 - `<source type>`，事件的来源类型。
 - `host::<host>`，其中 `<host>` 为事件所在的主机。
 - `source::<source>`，其中 `<source>` 为事件的来源。
- 已计算字段键必须以 "EVAL-" 开头（包括连字符），但 "EVAL" 不区分大小写（例如，可以是 "eVaL"）。
- `<field_name>` 是区分大小写的。这与 Splunk 软件中的所有其他字段名称一致。
- `<eval_statement>` 的灵活性堪比 `eval` 搜索命令。其计算结果可以为任何值类型，包括多值、布尔或空值。

通过 props.conf 配置已计算字段的示例

以下搜索示例（取自《搜索参考》中有关 `eval` 命令的介绍）检查地震数据，并通过创建 `Description` 字段按地震深度对地震进行分类：

```
source=eqs7day-M1.csv | eval Description=case(Depth<=70, "Shallow", Depth>70 AND Depth<=300, "Mid", Depth>300 AND Depth<=700, "Deep") | table Datetime, Region, Depth, Description
```

使用已计算字段，您可以为 `props.conf` 中的 `Description` 字段定义 `eval` 表达式。在 `props.conf` 中创建以下段落：

```
<Stanza>
Eval-Description = case(Depth<=70, "Shallow", Depth>70 AND Depth<=300, "Mid", Depth>300 AND Depth<=700, "Deep")
```

并将搜索编写为如下形式：

```
source=eqs7day-M1.csv | table Datetime, Region, Depth, Description
```

现在您可将 `Description` 视为提取的其他字段并对它执行搜索。Splunk 软件将查找已计算字段键，并针对包含 `Depth` 字段的每个事件对该键进行求值。还可以运行如下所示的搜索：

```
source=eqs7day-M1.csv Description=Deep
```

在定义已计算字段键之后，Splunk 软件会在搜索时针对将提取的字段包含在 `eval` 语句中的事件计算该字段。已计算字段的计算发生在**搜索时间字段提取**和**设置字段别名**之后，但在**查找字段**派生之前。

使用默认字段

字段是事件数据中的可搜索名称/值对。执行**搜索**时，是将搜索术语与**事件数据**的片段相匹配；使用字段可进行更为精确的搜索。字段是在索引时或搜索时从事件数据**提取**而来的。在索引时自动提取的字段称为**默认字段**。

默认字段有许多用途。例如，默认字段 `index` 标识事件所位于的索引。默认字段 `linecount` 描述事件所包含的行数，而 `timestamp` 指定事件发生的时间。为了正确创建事件，Splunk 软件在为事件创建索引时会使用某些字段的值，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。

有关在搜索命令中使用默认字段的更多信息，请参阅《[搜索手册](#)》中的“关于搜索语言”。有关配置默认字段的信息，请参阅《[数据导入手册](#)》中的“关于默认字段”。

字段类型	字段列表	描述
内部字段	<code>_raw</code> , <code>_time</code> , <code>_indextime</code> , <code>_cd</code>	包含事件的一般信息。
默认字段	<code>host</code> , <code>index</code> , <code>linecount</code> , <code>punct</code> , <code>source</code> , <code>sourcetype</code> , <code>splunk_server</code> , <code>timestamp</code>	这些字段包含有关事件来源、事件所属的索引、事件类型、事件所包含的行数以及事件发生时间的信息。默认情况下，会为这些字段建立索引并将其添加到“字段”菜单中。
默认日期时间字段	<code>date_hour</code> , <code>date_mday</code> , <code>date_minute</code> , <code>date_month</code> , <code>date_second</code> , <code>date_wday</code> , <code>date_year</code> , <code>date_zone</code>	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 <code>date_*</code> 字段。如果某事件包含 <code>date_*</code> 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。

一个字段可以具有多个值。请参阅“操作和评估多值字段”。

可以通过 Splunk Web 或使用提取搜索命令来提取其他（非默认）字段。请参阅[关于字段](#)。

您还可能需要更改字段的名称，或者将其与其他相似字段组合在一起。这可通过字段和字段值的标记或别名来轻松实现。请参阅在[“搜索”中为字段值对设置标记](#)。

本主题介绍在您创建数据索引时 Splunk 软件会自动添加的内部字段及其他默认字段。

内部字段

以下划线开头的字段是内部字段。

注意：建议您不要覆盖内部字段，除非您对自己进行的操作有十足的把握。

`_raw`

`_raw` 字段包含事件的原始数据。`search` 命令使用 `_raw` 中的数据来执行搜索和数据提取。

您无法总是直接对 `_raw` 值执行搜索，但可以使用诸如 `regex` 或 `sort` 等命令执行筛选。

示例：返回包含以 "10" 开头的 IP 地址的 `sendmail` 事件。

```
eventtype=sendmail | regex _raw="*10.\d\d\d\d.\d\d\d\d.\d\d\d\d\*
```

`_time`

`_time` 字段包含以 Unix 时间表示的事件时间戳。此字段用于在 Splunk Web 中创建事件时间线。

注意：`_time` 字段在内部存储为 UTC 格式。当 Splunk 软件呈现搜索结果时，会将此字段转换成人工可读的 Unix 时间（这是**搜索时间**事件处理的最后一步）。

示例：在类型为 "mail" 的所有来源中搜索发送给用户 "strawsky@bigcompany.com" 的邮件，然后按时间戳对搜索结果进行排序。

```
sourcetype=mail to=strawsky@bigcompany.com | sort _time
```

`_indextime`

`_indextime` 字段包含为事件创建索引的时间（以 Unix 时间表示）。您可以使用此字段专注于或筛选出在特定时间范围内创建索引的事件。

`_cd`

`_cd` 字段实际上是为索引内的事件提供“地址”。它由两个数字组成，即一个短数字和一个长数字。短数字表示事件所位于的特定索引数据桶。长数字表示索引数据桶偏移量。它提供了事件在其数据桶中的精确位置。`_cd` 仅供内部参考之用，因此我们建议不要设置包含此字段的搜索。

其他默认字段

`host`

`host` 字段包含生成相应事件的网络设备的原始主机名或 IP 地址。可使用 `host` 字段指定事件必须匹配的 `host` 值，以缩小搜索范围。您可以使用通配符在一个表达式内指定多个主机（示例：`host=corp*`）。

可以使用 `host` 在数据生成命令中筛选结果，或将其用作数据处理命令中的参数。

示例 1：在所有 "corp" 服务器上搜索用户 "strawsky" 访问的事件。随后，报告了 20 个最近发生的事件。

```
host=corp* eventtype=access user=strawsky | head 20
```

示例 2：搜索包含 "404" 且来自以 "192" 开头的主机的事件。

```
404 | regex host=*192.\d\d\d\.\d\d\d\.\d\d\d\.*
```

`index`

`_index` 字段包含用于为给定事件创建索引的索引的名称。使用以下形式指定要在搜索中使用的索引：`index="name_of_index"`。默认情况下，所有事件的索引都将在 `main` 索引 (`index="main"`) 中创建。

示例：在 `myweb` 索引中搜索具有 ".php" 扩展名的事件。

```
index="myweb" *.php
```

`linecount`

`linecount` 字段包含事件所含的行数。这是在为事件建立索引之前事件所包含的行数。可使用 `linecount` 搜索与指定行数匹配的事件，或将其用作数据处理命令中的参数。要指定匹配范围，请使用大于或小于表达式（示例：`linecount>10 linecount<20`）。

示例：在 `corp1` 中搜索包含 "40" 且行数为 40 的事件，并忽略包含 400 的事件。

```
40 linecount=40 host=corp1 NOT 400
```

`punct`

`punct` 字段包含从事件中提取的标点符号模式。每种事件类型的标点符号模式是唯一的。可使用 `punct` 在搜索期间筛选事件，或将其用作数据处理命令中的字段参数。

可在 `punct` 字段中使用通配符来搜索多个具有您要搜索的一些公用字符的标点符号模式。必须使用引号来定义 `punct` 字段的标点符号模式。

示例 1：搜索所有以 : 开头和结束的标点符号模式

```
punct=":":
```

示例 2：在 `php_error.log` 中搜索具有以下标点符号模式的 php 错误事件：`"[--::]_:___:/-.////.____"`。

```
source="/var/www/log/php_error.log" punct="[--::]_:___'/-.-'///.____"
```

source

`source` 字段包含事件来源的文件、流或其他输入的名称。可使用 `source` 在搜索期间筛选事件，或将其用作数据处理命令中的参数。您可以使用通配符在一个表达式内指定多个来源（示例：`source=*php.log*`）。

可以使用 `source` 在数据生成命令中筛选结果，或将其用作数据处理命令中的参数。

示例：搜索来自来源 `/var/www/log/php_error.log` 的事件。

```
source="/var/www/log/php_error.log"
```

sourcetype

`sourcetype` 字段指定事件来源的数据导入格式，例如 `access_combined` 或 `cisco_syslog`。可使用 `sourcetype` 在搜索期间筛选事件，或将其用作数据处理命令中的参数。您可以使用通配符在一个表达式内指定多个来源（示例：`sourcetype=access*`）。

示例：搜索来源类型为 `"access log"` 的所有事件。

```
sourcetype=access_log
```

splunk_server

`splunk_server` 字段包含事件所属的 Splunk 服务器的名称。在分布式 Splunk 环境尤其有用。

示例：将搜索限制在名为 `"remote"` 服务器上的 `main` 索引。

```
splunk_server=remote index=main 404
```

timestamp

`timestamp` 字段包含事件的时间戳值。您可以配置时间戳的提取方式。可将 `timestamp` 用作一个 `search` 命令参数来筛选搜索。

例如，您可以在搜索中添加 `timestamp=none` 来筛选搜索结果，以仅包含没有可识别时间戳值的事件。

示例：返回数据中没有可识别时间戳的事件的数量。

```
timestamp=none | stats count(_raw) as count
```

默认日期时间字段

可使用日期时间字段在搜索期间筛选事件，或将其用作数据处理命令中的字段参数。

如果您所在的时区与 Splunk 服务器不同，基于时间的搜索将使用创建事件索引时的服务器所指定的事件时间戳。日期时间值是在创建事件索引时从事件分析而来的文本值，与事件的时区无关。因此，诸如 `"05:22:21"` 之类的字符串将被分析为以下索引字段：`"date_hour::5 date_minute::22 date_second::21"`。

date_hour

`date_hour` 字段包含事件发生时的小时值（范围：0-23）。该值是从事件时间戳（`_time` 中的值）中提取而来的。

示例：搜索包含术语 `"apache"` 且发生在当前日期上午 12 点到晚上 10 点之间的事件。

```
apache (date_hour >= 12 AND date_hour <= 22)
```

date_mday

`date_mday` 字段包含事件发生时的日期值（范围：1-31）。该值是从事件时间戳（`_time` 中的值）中提取而来的。

示例：搜索包含 `"apache"` 术语且发生在当月 1 号到 15 号之间的事件。

```
apache (date_mday >= 1 AND date_mday <= 15)
```

date_minute

`date_minute` 字段包含事件发生时的分钟值（范围：0-59）。该值是从事件时间戳（`_time` 中的值）中提取而来的。

示例：搜索包含 `"apache"` 术语且发生在当前小时的 15 到 20 分钟之间的事件。

```
apache (date_minute >= 15 AND date_minute <= 20)
```


date_month

`date_month` 字段包含事件发生时的月份值。该值是从事件时间戳（`_time` 中的值）中提取而来的。

示例：搜索包含 "apache" 术语且发生在一月份的事件。

```
apache date_month=1
```

date_second

`date_second` 字段包含事件时间戳的秒数值部分（范围：0-59）。该值是从事件时间戳（`_time` 中的值）中提取而来的。

示例：搜索包含 "apache" 术语且发生在当前分钟的 1 秒到 15 秒之间的事件。

```
apache (date_second >= 1 AND date_second <= 15)
```

date_wday

`date_wday` 字段包含事件发生时的星期值（星期日、星期一等）。日期会从事件时间戳中提取（`_time` 中的值），并确定将该日期转换成星期几。然后将该星期值放入 `date_wday` 字段中。

示例：搜索包含 "apache" 术语且发生在星期一的事件。

```
apache date_wday="sunday"
```

date_year

`date_year` 字段包含事件发生时的年份值。该值是从事件时间戳（`_time` 中的值）中提取而来的。

示例：搜索包含 "apache" 术语且发生在 2008 年的事件。

```
apache date_year=2008
```

date_zone

`date_zone` 字段包含事件本地时区的时间值（以 Unix 时间格式的小时数表示）。该值是从事件时间戳（`_time` 中的值）中提取而来的。可使用 `date_zone` 以分钟为单位指定偏移量，以偏移事件的时区（范围：-720 至 720）。

示例：搜索包含 "apache" 术语且发生在当前时区（本地）的事件。

```
apache date_zone=local
```

关于 Splunk 正则表达式

本入门指导旨在帮助您创建有效的正则表达式。对于正则表达式语法和用法的讨论，请参阅在线资源，例如 www.regular-expressions.info 或者关于本主题的手册。

正则表达式匹配文本字符模式，并用于提取默认字段、识别二进制文件类型以及自动分配来源类型。当您定义自定义字段提取、过滤器事件、发送数据和关联搜索时，也可以使用正则表达式。使用正则表达式的搜索命令包括 `rex` 和 `regex`，以及评估函数，例如 `match` 和 `replace`。

Splunk 正则表达式均为 PCRE（Perl 兼容正则表达式），且使用 PCRE C 库。

正则表达式术语和语法

术语	描述
文本	使用正则表达式匹配的精确字符文本。
正则表达式	定义 Splunk 软件用于根据文本匹配的模式元字符。
组	正则表达式可以通过用于括起正则表达式字符的括号类型来表示分组。组可以定义字符类、重复匹配、命名的捕获组、模块化正则表达式等等。您可以对括起的组应用量词，并可在该组内使用替换。
字符类	方括号括起的字符。用于匹配字符串。要设置字符类，可使用连字符定义范围，例如 <code>[A-Z]</code> ，来匹配任何大写字母。使用脱字符（ <code>^</code> ）作为字符类的首字符将定义反向匹配，例如 <code>[^A-Z]</code> 用于匹配任何小写字母。
字符类型	与通配符相类似，字符类型是特定字面匹配的表示。例如，句点 <code>.</code> 匹配任意字符， <code>\w</code> 匹配单词或字母数字字符（包括下划线等）。
锚点	匹配文本格式位置的字符类型，例如，回车（ <code>\r</code> ）和换行符（ <code>\n</code> ）。

替换	是指在正则表达式中提供替代匹配模式。使用竖线或管道符 () 分隔各替代模式，替代模式可以包含完整的正则表达式。例如， <code>grey gray</code> 匹配 <code>grey</code> 或 <code>gray</code> 。
量词或重复	(<code>*</code> , <code>+</code> , <code>?</code>) 用于定义如何将组与字面模式进行匹配。例如， <code>*</code> 匹配 0 次或多次， <code>+</code> 匹配 1 次或多次， <code>?</code> 匹配 0 次或 1 次。
反向引用	您可以重新调用以供今后使用的文本组。要表示值的反向引用，指定一个美元符号 (\$) 和一个非零数字。
环视结构	定义组的一种方法，可确定组在字符串中的位置。此定义将匹配组中的正则表达式，但将放弃匹配项以保留结果。例如，您可以使用环视结构来匹配 <code>x</code> (后跟 <code>y</code>)，而不匹配 <code>yo</code> 。

字符类型

字符类型是字面匹配的简单表示。

术语	描述	示例	说明
<code>\w</code>	匹配单词字符（字母、数字、或下划线字符）。	<code>\w\w\w</code>	匹配任意三个单词字符。
<code>\W</code>	匹配非单词字符。	<code>\W\W\W</code>	匹配任意三个非单词字符。
<code>\d</code>	匹配数字字符。	<code>\d\d\d-\d\d\d-\d\d\d\d\d</code>	匹配社会保险号或是类似于 3-2-4 的数字字符串。
<code>\D</code>	匹配非数字字符。	<code>\D\D\D</code>	匹配任意三个非数字字符。
<code>\s</code>	匹配空白字符。	<code>\d\s\d</code>	匹配由一个数字、一个空白字符然后是另一个数字组成的一个序列。
<code>\S</code>	匹配非空白字符。	<code>\d\S\d</code>	匹配由一个数字、一个非空白字符和另一个数字组成的一个序列。
<code>.</code>	匹配任意字符。少量使用。	<code>\d\d.\d\d.\d\d</code>	匹配诸如 12/31/14 或 01.01.15 的日期字符串，但也可以匹配 99A99B99。

组、量词和替换

正则表达式可以通过用于括起正则表达式字符的括号类型来表示分组。您可以对括起的组应用量词 (`*`, `+`, `?`)，并可在该组内使用替换。

术语	描述	示例	说明
<code>*</code>	匹配零次或多次。	<code>\w*</code>	匹配零个或多个单词字符。
<code>+</code>	匹配一次或多次。	<code>\d+</code>	匹配至少一个数字。
<code>?</code>	匹配零次或一次。	<code>\d\d\d-?\d\d-?\d\d\d\d\d</code>	匹配带或不带短划线的社会保险号。
()	括号用于定义匹配或捕获组、原子组和环视结构。	<code>(H..).(o..)</code>	当给出字符串 <code>Hello World</code> 时，其匹配 <code>Hel</code> 和 <code>o Wo</code> 。
[]	方括号用于定义字符类。	<code>[a-z0-9#]</code>	匹配从 <code>a</code> 到 <code>z</code> 、 <code>0</code> 到 <code>9</code> 或 <code>#</code> 中的任意字符。
{ }	大括号用于定义重复。	<code>\d{3,5}</code>	匹配 3 到 5 个数字长度的字符串。
< >	尖括号用于定义命名的捕获组。使用 <code>(?P<var> ...)</code> 语法来设置已命名的字段提取。	<code>(?P<ssn>\d\d\d-\d\d-\d\d\d\d\d)</code>	提取一个社会保险号并将它分配给 <code>ssn</code> 字段。
[[]]	两个方括号用于定义 Splunk 特定的模块化正则表达式。	<code>[[octet]]</code>	0-255 范围内的整数有效。

组、量词和替换的简单示例

本例显示了匹配 `to` 或 `too` 的两种方法。

第一个正则表达式使用 `?` 量词以在匹配第一个 "o" 之后最多再匹配 1 次 "o"。

第二个正则表达式使用替换来指定模式。

```
to(o)?
(to|too)
```

正则表达式中的捕获组

命名的捕获组是一个当正则表达式匹配事件时提取字段值的正则表达式分组。捕获组包含字段的名称。它们以尖括号表示如下：

```
matching text (?<field_name>capture pattern) more matching text
```

例如，您有如下事件文本：

```
131.253.24.135 fail admin_user
```

下面是两个在捕获组中使用不同语法的正则表达式要从事件中提取同样的字段集。

- 表达式 A： `(?<ip>\d+\.\d+\.\d+\.\d+) (?<result>\w+) (?<user>.*)`
- 表达式 B： `(?<ip>\S+) (?<result>\S+) (?<user>\S+)`

在表达式 A 中，用于首个捕获组 (ip) 的模式匹配字符是特定的。`\d` 表示“数字”，而 `+` 表示“一个或多个”。因此 `\d+` 表示“一个或多个数字”。`\.` 表示句点。

ip 的捕获组想要匹配一个或多个数字，后跟句点，后跟一个或多个数字，后跟句点，后跟一个或多个数字，后跟句点，后跟一个或多个数字。这描述了 ip 地址的语法。

在表达式 A 中的第二个捕获组用于 `result` 字段，且该字段具有 `\w+` 模式，表示“一个或多个字母数字字符”。在表达式 A 中的第三个捕获组用于 `user` 字段，且该字段具有 `.*` 模式，表示“匹配余下的所有字符”。

表达式 B 使用一种称为反向匹配的常用技术。使用反向匹配，正则表达式不会尝试定义要匹配的文本。相反它定义不要匹配的文本。在此表达式 B 中，应该从示例事件中提取的值是“非空格”字符 (`\s`)。可使用 `+` 来指定“一个或多个”“非空格”字符。

因此表达式 B 假设：

1. 先为 ip 字段值提取出非空格字符的首个字符串。
2. 忽略下面的空格。
3. 然后为 result 字段值提取出非空格字符的第二个字符串。
4. 忽略第二个空格。
5. 为 user 字段值提取出非空格字符的第三个字符串。

非捕获组匹配

使用语法 `(?: ...)` 创建非捕获的匹配组。注意，此时无需使用尖括号括起字段名。`?` 号之后的冒号将其识别为非捕获组。

例如，`(?:Foo|Bar)` 匹配 `Foo` 或 `Bar`，但两个字符串都是非捕获的。

模块化正则表达式

模块化正则表达式是指被定义为用于较长正则表达式定义的较小正则表达式部分。模块化正则表达式在 `transforms.conf` 中进行定义。

例如，可以先定义整数，然后使用该正则表达式定义来定义浮点数。

```
[int]
# matches an integer or a hex number
REGEX = 0x[a-fA-F0-9]+\d+

[float]
# matches a float (or an int)
REGEX = \d*\.\d+|[int]
```

在 `[float]` 的正则表达式中，应使用两个方括号 `[int]` 来调用于整数或十六进制数匹配的模块化正则表达式。

也可以在字段提取中使用模块化正则表达式。

```
[octet]
# this would match only numbers from 0-255 (one octet in an ip)
REGEX = (?:2(?:5[0-5]|[0-4][0-9])|[0-1][0-9][0-9]|[0-9][0-9]?)
```

```
[ipv4]
# matches a valid IPv4 optionally followed by :port_num the
# octets in the ip would also be validated 0-255 range
# Extracts: ip, port
REGEX = (?<ip>[[octet]](?:\.[[octet]]){3})(?::[[int:port]])?
```

[octet] 正则表达式使用两个嵌套的非捕获组来完成这一工作。请参阅本主题的子章节中关于非捕获组匹配的相关内容。

数据分类：事件类型和交易

关于事件类型

事件类型是一种分类系统，旨在帮助您理解数据。事件类型可让您通过大量数据进行筛选，查找类似的模式，并创建告警和报表。

注意：不建议使用事件类型作为搜索的快捷方式。如果您想缩短搜索的某个环节，用搜索宏会有效许多。搜索宏在可表达的内容方面更具灵活性，可包含其他搜索命令而非基本查询词，可进行参数化，并且在检索到事件时不会产生费用。有时候这有助于简化管理，比如，因为单个搜索宏可以取代多个事件类型。

有关使用搜索宏的更多信息，请参阅[在搜索中使用搜索宏](#)。

事件类型如何工作

该搜索可以返回的每个事件都与事件类型相关联。例如，假设搜索如下：

```
sourcetype=access_combined status=200 action=purchase
```

如果将搜索保存为名为 `successful_purchase` 的事件类型，则 `eventtype=successful_purchase` 会在搜索时间添加到该搜索可以返回的任何事件中。即使您正搜索的内容完全不同，结果也是一样。

注意：使用事件类型可能会获取大量数据，因为所有搜索都会尝试将事件与任一已知事件类型相关联。随着定义的事件类型越来越多，搜索性能的成本也随之增加。您可以用 `command.search.typer` 参数查看搜索命令的执行成本。请参阅搜索任务查看器。

要构建一个搜索来结合匹配该事件类型的所有事件，添加 `eventtype=access_combined` 作为搜索词。

单个事件可以匹配多种事件类型。如果一个事件匹配两个或多个事件类型，`eventtype` 则作为多值字段。

重要事件类型定义限制

事件类型也不能基于以下搜索：

- 在简单搜索后包括管道符。
- 包括子搜索。
- 是由使用 `savedsearch` 命令来引用报表名的简单搜索所定义的。例如，如果报表名为 `failed_login_search`，则不应使用以下搜索来定义事件类型：`| savedsearch failed_login_search`。此时，应使用将 `failed_login_search` 定义为事件类型定义的搜索字符串。

最后一点更像是最佳实践，而非严格的限制。您要避免 `failed_login_search` 下面的搜索字符串以后被另一个用户修改（可能是以一种会破坏事件类型的方式）的情形。如果在事件定义中使用实际搜索字符串，则可以更好的控制事件类型的持续有效性。

注意：如果您想使用事件类型作为搜索的快捷方式，则使用搜索宏。有关事件类型对比搜索宏的更多信息，请参阅[关于事件类型](#)。

创建事件类型

创建新事件类型的最简单方法是通过 Splunk Web。在运行返回有效事件类型的搜索之后，单击**另存为**并选择**事件类型**。此时将打开**另存为事件类型**对话框，您可在其中提供相应事件类型名称并选择对其应用标记。有关将搜索保存为事件类型的更多信息，请参阅[在 Splunk Web 中定义和维护事件类型](#)。

也可以通过修改 `eventtypes.conf` 创建新的事件类型。有关以此方式手动配置事件类型的更多信息，请参阅[直接在 eventtypes.conf 中配置事件类型](#)。

事件类型标记

事件类型可有一个或多个标记与之关联。将搜索另存为事件类型时可从事件类型管理器（位于**设置 > 事件类型**）中添加这些标记。在此窗口的事件类型列表中，选择您要编辑的事件类型。

为事件类型设置标记类型可将您的数据组织为各种类别。每个事件可以有多个标记。您可以在 Splunk Web 中为事件类型设置标记或在 `tags.conf` 中进行配置。有关事件类型标记的更多信息，请参阅[为事件类型设置标记](#)。

事件类型标记示例 #1

为抽象的字段值赋予更为直观易读的名字，然后使用事件类型标记帮助跟踪这些字段值，如 HTTP 访问日志、IP 地址或 ID 编号。前往**设置 > 事件类型**，为事件类型添加标记。从此菜单的事件类型列表中，选择所需事件类型。

将标记添加到事件类型后，可按照搜索任意标记的相同方式来搜索它们。

假设我们已将查找未找到页面的搜索另存为事件类型 `status=404`，然后将查找验证失败的搜索另存为事件类型 `status=403`。如果将这两种事件类型均标记为**客户端错误**，则可通过使用以下搜索来检索任一事件类型的所有事件：

```
tag::eventtype=HTTP client error
```

有关使用标记的更多信息，请参阅[在“搜索”中为字段值对设置标记](#)。

事件类型标记示例 #2

事件类型标记一般用于 Splunk 平台的通用信息模型 (CIM) 加载项，旨在将来自陌生来源类型的新索引数据规范化。我们可以使用标记来识别单一数据源中的不同事件类型。

您可以在数据中使用符合 CIM 的标记。

1. 从 Splunk Web 中，选择**设置 > 数据模型**。找到您的数据要映射到的数据模型数据集，然后识别其关联的标记。例如，Performance 数据模型中的 `cpu_load_percent` 对象包含以下关联标记：

```
tag = performance
tag = cpu
```

2. 在 Splunk Web 的事件类型管理器中，前往**设置 > 事件类型**，然后创建所需的事件类型。您也可以直接编辑 `eventtypes.conf` 文件。
3. 在 Splunk Web 中创建所需的标记。选择**设置 > 事件类型**，找到要添加标记的事件类型并单击其名称。您也可以直接编辑 `tags.conf` 文件。

有关通用信息模型 (CIM) 和标记事件的更多信息，请参阅“配置符合 CIM 的事件标记”。

在 Splunk Web 中配置事件类型

事件类型代表会返回特定类型事件或有用事件集合的搜索。该搜索可以返回的每个事件都与事件类型相关联。例如，假设搜索如下：

```
sourcetype=access_combined status=200 action=purchase
```

如果将搜索保存为名为 `successful_purchase` 的事件类型，则 `eventtype=successful_purchase` 会在搜索时间添加到该搜索可以返回的任何事件中。即使您正搜索的内容完全不同，结果也是一样。

之后，如果要构建一个搜索来结合匹配该事件类型的所有事件，在搜索字符串中加入 `eventtype=access_combined`。

单个事件可以匹配多种事件类型。如果一个事件匹配两个或多个事件类型，`eventtype` 则作为多值字段。

将所运行的搜索保存为事件类型

运行搜索时，可以将该搜索保存为事件类型。事件类型通常代表会返回特定类型事件或各种有用事件的搜索。

在创建事件类型时，事件类型定义会添加到 `$SPLUNK_HOME/etc/users/<your-username>/<app>/local/` 中的 `eventtypes.conf`，其中 `<app>` 是当前的应用上下文。如果您更改了事件类型的权限，使其对所有用户可用（仅在该应用中可用，或者对所有应用全局可用），Splunk 平台会将该事件类型移至 `$SPLUNK_HOME/etc/apps/<App>/local/`。

1. 在“搜索”视图中，运行一个搜索。
2. 单击**另存为**并选择**事件类型**。
3. 为事件类型取唯一的**名称**。
4. （可选）添加一个或多个以逗号分隔的**标记**。

可以将同样的标记应用于产生类似结果的所有事件类型。带此标记的搜索会返回一组共同属于那些事件类型的事件。

5. （可选）选择一个**颜色**。

如果进行了此操作，匹配该事件类型的所有事件列表开头会显示一条彩色带。例如，此事件匹配**颜色为紫色**的事件类型。

```

> 16/06/14 12.130.60.5 - - [14/Jun/2016:17:57:58] "POST /cart/error.do?msg=CreditDoesNotMatch&JSESSIONID=SD5SL6FF7ADFF53001
17:57:58.000 HTTP 1.1" 200 1167 "http://www.buttercupgames.com/cart.do?action=purchase&itemId=EST-19" "Mozilla/5.0
(compatible; MSIE 9.0; Windows NT 6.1; WOW64; Trident/5.0; BOIE9;ENUS)" 232
host = www1 | source = tutorialdata.zip./www1/access.log | sourcetype = access_combined_wcookie

> 16/06/14 12.130.60.5 - - [14/Jun/2016:17:57:58] "POST /cart/error.do?msg=CreditDoesNotMatch&JSESSIONID=SD5SL6FF7ADFF53001
17:57:58.000 HTTP 1.1" 200 1167 "http://www.buttercupgames.com/cart.do?action=purchase&itemId=EST-19" "Mozilla/5.0
(compatible; MSIE 9.0; Windows NT 6.1; WOW64; Trident/5.0; BOIE9;ENUS)" 232
host = www1 | source = tutorialdata.zip./www1/access.log | sourcetype = access_combined_wcookie

```

您可以在“设置”中修改事件类型的颜色（或不设置任何颜色）。

6. （可选）为事件类型指定**优先级**。

优先级会影响匹配两个或多个事件类型的事件的显示方式。**1** 表示**优先级**最高，而 **10** 表示最低。请参阅[关于事件类型优先级](#)。

7. 单击**保存**以保存新的事件类型。

您可以访问您自己和其他用户在**设置 > 事件类型**中创建的事件类型列表。

用此方式创建的所有事件类型也会显示于“设置”中的“事件类型”列表页面。如果需要更新事件类型，则可以前往该页面操作。

“设置”中的“事件类型”页面

“设置”中的“事件类型”页面会列出您有权限查看或编辑的所有事件类型。您可以通过“事件类型”页面来创建事件类型或维护现有事件类型。

在“设置”中添加事件类型

还可以通过“事件类型”页面创建新的事件类型。

前提条件

- 查看[将所运行的搜索保存为事件类型](#)。该主题涵盖了大部分字段的介绍。
- 有关事件类型标记的信息，请参阅[关于标记和别名](#)。
- 有关**颜色**和**优先级**字段的更多信息，请参阅[关于事件类型优先级](#)。

在“设置”中添加新事件类型

1. 选择**设置 > 事件类型**。
2. 单击**新建**。

新增
事件类型 > 新增

目标应用 *

search

名称 *

external-referer

搜索字符串 *

sourcetype=access_combined referer!="-" referer_domain!=" . splunk.com referer_domain!=" . splunkbase.com source!="var/log/httpd/banner_access_log" NOT eventtype=clientip-noroutable NOT eventtype=clientip-internal NOT eventtype=file-image NOT eventtype=file-resource*

标记

输入以逗号分隔的标记列表。

颜色

洋红色

优先级

1 (最高)

最高优先级在结果中优先显示。

取消 保存

3. （可选）如果**目标应用**不是当前应用上下文，请将其值修改为适用于事件类型的正确应用。
4. 提供该事件类型的唯一的**名称**。
5. 输入该事件类型的**搜索字符串**。

此搜索持续返回特定类型的事件。

6. （可选）添加一个或多个以逗号分隔的**标记**。

可以将同样的标记应用于产生类似结果的所有事件类型。带此标记的搜索会返回一组共同属于那些事件类型的事件。

7. (可选) 选择一个颜色。

如果进行了此操作，匹配该事件类型的所有事件列表开头会显示一条彩色带。

8. (可选) 为事件类型指定优先级。

优先级会影响匹配两个或多个事件类型的事件的显示方式。**1** 表示**优先级**最高，而 **10** 表示最低。

优先级决定了展开事件中事件类型列表中的顺序。如果匹配某事件的两个或多个事件类型还定义了**颜色**值，它同时还会决定事件类型要显示什么颜色。

有关更多信息，请参阅[关于事件类型优先级](#)。

9. 单击**保存**以保存事件类型。

注意：所有事件类型最初都是为特定 Splunk 应用而创建的。要使特定事件类型对所有用户全局可用，您必须给所有角色赋予读取或写入 Splunk 应用的权限，并使它对所有 Splunk 应用可用。有关设置事件类型（及其他知识对象类型）权限的更多信息，请参阅本手册中的[管理知识对象权限](#)。

在“设置”中更新事件类型

您可以更新您创建或您有权限编辑的任何事件类型的定义。

1. 导航到**设置 > 事件类型**。
2. 在“事件类型”列表页面中找到要更新的事件类型并单击其名称。
3. 根据需要修改事件类型的**搜索字符串**、**标记**、**颜色**和**优先级**。
4. 单击**保存**以保存更改。

关于事件类型优先级

您为某一事件类型选择的**优先级**值会影响匹配该事件类型的事件的显示方式，当这些事件同时也匹配其他事件类型时。

优先级影响展开事件中的事件类型排列顺序。

事件类型匹配在搜索时间内进行。当您运行一个搜索且此搜索返回的事件匹配某事件类型时，Splunk 软件会将相应地 `eventtype` 字段/值/值对添加给该事件，这个值就是事件类型的名称。

在查看搜索结果时，可以看到已添加到一个事件中的所有事件类型。展开该事件，并查看是否有列出 `eventtype` 字段。如果有列出，则表示该事件至少匹配一种事件类型。

如果该事件匹配两种或多种事件类型，`eventtype` 则变为多值字段，其值按字母顺序排序，但具有**优先级**设置的事件类型除外。具有**优先级**设置的事件类型会列在没有此设置的事件类型上方，并按他们的**优先级**值排序。

如果您有许多重叠事件类型，或者您的事件类型是更大事件类型的子集，您可能希望为所关注的事件类型赋予更好的优先级。例如，您可以轻松获得一组属于一个大范围 `all_system_errors` 事件类型的事件。在该大型事件集中，您的事件可能还属于更为关注的事件类型（如 `critical_disc_error` 和 `bad_external_resource_error`）。

以下是匹配 `all_system_errors` 和 `critical_disc_error` 事件类型事件的示例。

事件操作 ▾		
类型	✓ 字段	↑ 值
选定的	✓ host ▾	docs-unix-4
	✓ source ▾	/tmp/tutorialdata.zip:/www1/access.log
	✓ sourcetype ▾	access_combined_wcookie
事件	<input type="checkbox"/> JSESSIONID ▾	SD10SL4FF1ADFF53066
	<input type="checkbox"/> bytes ▾	268
	<input type="checkbox"/> categoryId ▾	ACCESSORIES
	<input type="checkbox"/> clientip ▾	91.205.189.15
	<input type="checkbox"/> eventtype ▾	critical_disc_error (crit_disc_error syserror) all_system_errors (syserror)
	<input type="checkbox"/> file ▾	msdunint screen

在本例中，`critical_disc_error` 事件类型的优先级是 3，而 `all_system_errors` 事件类型的优先级是 7。3 相对于 7 来说是更好的优先级值，因此 `critical_disc_error` 会优先显示在列表顺序中。

优先级决定事件要显示的事件类型颜色

每个事件只能显示一种事件类型颜色。当一个事件匹配多个事件类型且其中有两个或多个事件类型都设置了颜色值，则会显示带最佳优先级的事件类型的颜色。

接之前讨论的示例，以下为带事件类型颜色设置的两个事件的示例。

		host = docs-unix-4 source
>	16/06/14 17:57:58.000	91.205.189.15 - - [14/06/2016:17:57:58.000] "Chrome/19.0.1084.46" www.buttercupgames.com
		host = docs-unix-4 source
>	16/06/14 17:57:58.000	91.205.189.15 - - [14/06/2016:17:57:58.000] "Chrome/19.0.1084.46" www.buttercupgames.com
		host = docs-unix-4 source
>	16/06/14	[14/Jun/2016:17:57:58]

两个事件都匹配 `all_system_errors` 事件类型，且该事件类型的颜色值为橙色。以 `all_system_errors` 为主导事件类型的所有事件均以橙色事件类型的颜色设置显示。其中一个事件也匹配 `critical_disk_error` 事件类型，且该事件类型的优先级值优于 `all_system_errors`。 `critical_disk_error` 事件类型的颜色设置为紫色，则匹配此类型的事件具有紫色事件类型的颜色设置，而非橙色。

自动查找和构建事件类型

以下实用工具自动找到并创建事件类型，协助您确认数据中是否存在任何可能有用的事件类型：

- **查找事件类型：** `findtypes` 搜索命令分析事件集，并识别事件中可转换为有用事件类型的模式。
- **构建事件类型：** **构建事件类型实用工具** 基于单个事件创建事件类型。该实用工具还可以为事件类型分配特定颜色。例如，如果您将 "sendmail error" 事件类型设置为红色，则当您下一次运行搜索以返回适合该事件类型的事件时，这些事件将很容易发现，因为它们在事件列表中显示为红色。

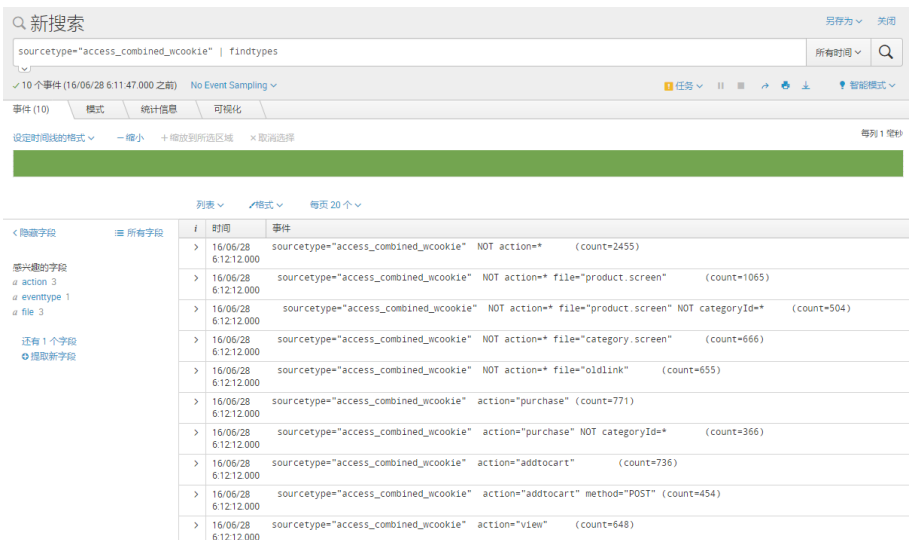
通过 `findtypes` 命令查找搜索数据中的事件类型

要查看搜索返回数据中的事件类型，将 `findtypes` 命令添加到搜索的结尾处：

```
... | findtypes
```

使用 `findtypes` 命令的搜索返回在搜索结果中发现的最常见事件组的明细。它们是：

- 按照“范围”（频率）的顺序排序。这有助于您方便地确定大型事件分组的各种事件子集类型。
- 与可用作事件类型基础的搜索关联，从而有助于您查找类似事件。



< 隐藏字段	所有字段	i	时间	事件	
		>	16/06/28 6:12:12.000	sourcetype="access_combined_wcookie" NOT action="*" (count=2455)	
		>	16/06/28 6:12:12.000	sourcetype="access_combined_wcookie" NOT action="*" file="product.screen" (count=1065)	
		>	16/06/28 6:12:12.000	sourcetype="access_combined_wcookie" NOT action="*" file="product.screen" NOT categoryId="*" (count=504)	
		>	16/06/28 6:12:12.000	sourcetype="access_combined_wcookie" NOT action="*" file="category.screen" (count=666)	
		>	16/06/28 6:12:12.000	sourcetype="access_combined_wcookie" NOT action="*" file="oldlink" (count=655)	
		>	16/06/28 6:12:12.000	sourcetype="access_combined_wcookie" action="purchase" (count=771)	
		>	16/06/28 6:12:12.000	sourcetype="access_combined_wcookie" action="purchase" NOT categoryId="*" (count=366)	
		>	16/06/28 6:12:12.000	sourcetype="access_combined_wcookie" action="addtocart" (count=736)	
		>	16/06/28 6:12:12.000	sourcetype="access_combined_wcookie" action="addtocart" method="POST" (count=454)	
		>	16/06/28 6:12:12.000	sourcetype="access_combined_wcookie" action="view" (count=648)	

默认情况下，`findtypes` 根据与所发现的每种事件类型匹配的事件数量，返回在样本中找到的前 10 个潜在事件类型。要增加此数量，可添加 `max` 参数。例如，`findtypes max=30` 返回样本中前 30 个潜在事件类型。

`findtypes` 命令还会说明所发现的事件分组是否匹配其他事件类型。

注意：为了返回这些结果，`findtypes` 命令分析了多达 5,000 个事件。若希望搜索的效率更高（但准确率可能更低），则可以使用 `head` 命令来减少这一数字：

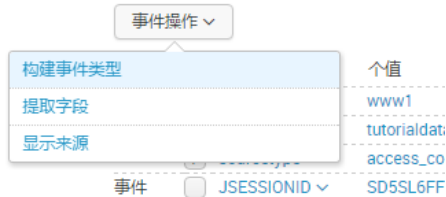
... | head 1000 | findtypes

使用构建事件类型实用工具创建事件类型

构建事件类型实用工具（通常称为“事件类型构建器”）会引导您完成基于搜索结果中的事件创建事件类型的整个过程。

1. 运行一个搜索，使它返回您希望事件类型所基于的事件。
2. 从返回的搜索结果中识别可作为事件类型的事件，并展开该事件。
3. 单击“事件操作”，并选择“构建事件类型”。

```
16/06/14      12.130.60.5 - - [14/Jun/2016:17:57:58] "POST /cart/error.do?msg=CreditDoesNotMatch&JSESSIONID=SD5SL6FF7ADFF53001
17:57:58.000 HTTP 1.1" 200 1167 "http://www.buttercupgames.com/cart.do?action=purchase&itemId=EST-19" "Mozilla/5.0
(compatible; MSIE 9.0; Windows NT 6.1; WOW64; Trident/5.0; BOIE9;ENUS)" 232
host = www1 | source = tutorialdata.zip./www1/access.log | sourcetype = access_combined_wcookie
```



通过构建事件类型实用工具，您可以设计一个可返回特定结果集的搜索。此搜索字符串显示在实用工具界面顶部的**生成的事件类型**的下方。

此实用工具还会显示一个示例事件列表。在您调整事件类型搜索字符串时，此列表会进行动态更新。

4. 在**事件类型功能**边栏中，选择可缩小事件类型搜索范围的字段/值对。

在您进行选择时，**生成的事件类型**搜索会相应地更新以包含这些选择。示例事件列表也会更新，以反映匹配所设计的事件类型的事件。

5. （可选）您可以随时单击**编辑**来直接编辑事件类型搜索。
6. （可选）当您认为某搜索可能是有用的事件类型时，可单击**测试**进行测试。

该搜索会在单独的窗口中运行。

7. 当您的搜索可返回正确的事件集时，单击**保存**以打开**保存事件类型**对话框。

保存事件类型

注意：要确保事件类型正常工作，请在保存之前先对其进行测试。

名称:

样式:

红色

优先级:

1 (最高)

当事件具有多个事件类型时，决定占优势的样式。

取消

保存

8. 输入事件类型的**名称**。
9. （可选）为事件类型指定**样式**。

样式就是其他事件类型定义工作流中的**颜色**。如果进行了此操作，匹配该事件类型的所有事件列表开头会显示一条彩色带。例如，此事件匹配**样式为紫色**的事件类型。

```
> 16/06/14      12.130.60.5 - - [14/Jun/2016:17:57:58] "POST /cart/error.do?msg=CreditDoesNotMatch&JSESSIONID=SD5SL6FF7ADFF53001
17:57:58.000 HTTP 1.1" 200 1167 "http://www.buttercupgames.com/cart.do?action=purchase&itemId=EST-19" "Mozilla/5.0
(compatible; MSIE 9.0; Windows NT 6.1; WOW64; Trident/5.0; BOIE9;ENUS)" 232
host = www1 | source = tutorialdata.zip./www1/access.log | sourcetype = access_combined_wcookie

> 16/06/14      12.130.60.5 - - [14/Jun/2016:17:57:58] "POST /cart/error.do?msg=CreditDoesNotMatch&JSESSIONID=SD5SL6FF7ADFF53001
17:57:58.000 HTTP 1.1" 200 1167 "http://www.buttercupgames.com/cart.do?action=purchase&itemId=EST-19" "Mozilla/5.0
(compatible; MSIE 9.0; Windows NT 6.1; WOW64; Trident/5.0; BOIE9;ENUS)" 232
host = www1 | source = tutorialdata.zip./www1/access.log | sourcetype = access_combined_wcookie
```

您可以在“设置”中修改事件类型的颜色（或不设置任何颜色）。

10. （可选）为事件类型指定**优先级**。

优先级会影响匹配两个或多个事件类型的事件的显示方式。**1** 表示**优先级**最高，而 **10** 表示最低。

优先级决定了展开事件中事件类型列表中的顺序。如果匹配某事件的两个或多个事件类型还定义了**颜色值**，它同时还会决定事件类型要显示什么颜色。

请参阅[关于事件类型优先级](#)。

11. 单击**保存**以保存事件类型。

在 eventtypes.conf 中配置事件类型

您可以通过配置 eventtypes.conf 添加新事件类型及更新现有事件类型。\$SPLUNK_HOME/etc/system/default/eventtypes.conf 中定义了几个默认事件类型。任何[通过 Splunk Web 创建的事件类型](#)均会自动添加到 \$SPLUNK_HOME/etc/system/local/eventtypes.conf 中。

重要事件类型定义限制

事件类型也不能基于以下搜索：

- 在简单搜索后包括**管道符**。
- 包括**子搜索**。
- 是由使用 savedsearch 命令来引用报表名的简单搜索所定义的。例如，如果报表名为 failed_login_search，则不应使用以下搜索来定义事件类型：| savedsearch failed_login_search。此时，应使用将 failed_login_search 定义为事件类型定义的搜索字符串。

最后一点更像是最佳实践，而非严格的限制。您要避免 failed_login_search 下面的搜索字符串以后被另一个用户修改（可能是以一种会破坏事件类型的方式）的情形。如果在事件定义中使用实际搜索字符串，则可以更好的控制事件类型的持续有效性。

配置

在 eventtypes.conf 中更改事件类型。将 \$SPLUNK_HOME/etc/system/README/eventtypes.conf.example 用作示例，或创建自己的 eventtypes.conf。

编辑 eventtypes.conf（位于 \$SPLUNK_HOME/etc/system/local/ 或您自定义的应用目录 \$SPLUNK_HOME/etc/apps/）。请参阅《[管理员手册](#)》中的“[关于配置文件](#)”。

```
[$EVENTTYPE]
```

- 事件类型的**标头**
- \$EVENTTYPE 是事件类型的名称。
 - 您可以有任意多的事件类型，每个事件类型由一个段落和以下任意多的属性/值对表示。

注意：如果事件类型的名称包括用百分比字符包围的字段名称（例如，%\$FIELD%），则在搜索时 \$FIELD 的值将替换为该事件的事件类型名称。例如，如果事件类型的标头为 [cisco-%code%] 且 code=432，则此事件类型将标记为 [cisco-432]。

```
disabled = <1 or 0>
```

- 启用或禁用事件类型。
- 设置为 1 时禁用。

```
search = <string>
```

- 此事件类型的搜索术语。
- 例如：error OR warn。

```
description = <string>
```

- （可选）易于人们理解的事件类型描述。

```
priority = <integer>
```

- 确定与事件匹配的事件类型的显示顺序。1 表示最高，10 表示最低。
 - 请参阅[关于事件类型优先级](#)。

```
color = <string>
```

- 此事件类型的颜色。
- 支持的颜色包括：无、et_blue、et_green、et_magenta、et_orange、et_purple、et_red、et_sky、et_tea 和 et_yellow。

注意：您可以按照为任何其他字段/值组合设置标记的方式来为 eventtype 字段值设置标记。有关更多信息，请参阅 tags.conf 规范文件。

示例

有两个事件类型；一个称为 `web`，另一个称为 `fatal`。

```
[web]
search = html OR http OR https OR css OR htm OR html OR shtml OR xls OR cgi
```

```
[fatal]
search = FATAL
```

禁用事件类型

要禁用某事件类型，将 `disabled = 1` 添加到相应事件类型段落 `eventtypes.conf` 中：

```
[$EVENTTYPE]
disabled = 1
```

`$EVENTTYPE` 是要禁用的事件类型的名称。

因此，如果要禁用 `web` 事件类型，应将以下条目添加到其段落中：

```
[web]
disabled = 1
```

配置事件类型模板

在搜索时，事件类型模板会创建事件类型。如果您有 Splunk Enterprise，则可以在 `eventtypes.conf` 中定义事件类型模板。编辑 `eventtypes.conf`（位于 `$SPLUNK_HOME/etc/system/local/` 或您自定义的应用目录 `$SPLUNK_HOME/etc/apps/`）。

有关配置文件的一般详细信息，请参阅《管理员手册》中的“关于配置文件”。

事件类型模板配置

在搜索时，事件类型模板会使用以百分比字符包围的字段名称来创建事件类型，其中 `%%$FIELD%` 值将替换为事件类型的名称。

```
[$NAME-%%$FIELD%]
$SEARCH_QUERY
```

因此，如果模板中的搜索查询返回 `%%$FIELD%=bar` 的事件，则会为该事件创建名为 `$NAME-bar` 的事件类型。

示例

```
[cisco-%%code%]
search = cisco
```

如果针对 "cisco" 的搜索返回 `code=432` 的事件，Splunk Enterprise 将创建名为 "cisco-432" 的事件类型。

关于交易

交易是指一定时间跨度内的概念相关事件组。**交易类型**是指已在 `transactiontypes.conf` 中配置并以**字段**形式保存的交易。

交易可以包括：

- 来自相同数据来源和相同主机的不同事件。
- 来自相同主机的不同数据来源的不同事件。
- 来自不同主机和不同数据来源的类似事件。

例如，在网上商店购物的客户可能将来自若干个来源的事件结合在一起，生成一笔交易：

- **一组 Web 访问事件**与....
-**应用程序服务器日志中的相应事件**共享会话 ID，应用程序服务器日志还包含相关帐户 ID、产品 ID 和交易 ID。该应用程序服务器事件中的交易 ID 还会显示在...
- 包含消息 ID 的...**消息队列事件**。该消息 ID 又通过...被共享
- 履行应用程序所记录的...**购买履行事件**共享，履行应用程序还包含客户所购买项目的发货状态。

将此处所有突出显示的内容组合在一起，即表示一笔用户交易。如果您要定义为交易类型，可以称为“项目购买”交易。其他类型的交易包括 Web 访问、应用程序服务器下载、电子邮件、安全性违规和系统故障。

交易搜索

通过交易搜索，您可以确定每个交易事件在多个已记录事件内的执行情况。使用 `transaction` 命令及其选项来定义返回交易（事件组）的搜索。有关以下操作的各种示例，请参阅《搜索参考》中的关于此命令的文档：

- 查找第一个事件与最后一个事件之间的时间跨度不超过特定数量（使用 `maxspan` 选项设置）的事件组
- 查找两个所含事件之间的时间跨度不超过特定值（使用 `maxpause` 选项设置）的事件组。
- 查找事件总数不超过特定数量（使用 `maxevents` 选项设置）的相关事件组
- 设计一笔交易，以查找最终事件包含特定文本字符串（使用 `endswith` 选项设置）的事件组。

学习 `transaction` 命令主题，以获取可用于该命令的选项的完整列表。

还可以使用 `transaction` 命令来覆盖已在 `transactiontypes.conf` 中配置的交易选项。

要了解使用 `transaction` 执行搜索的更多信息，请参阅《搜索手册》中的“确定事件并将其分组为交易”。

配置交易类型

在创建交易搜索之后，如果您发现值得重复使用该搜索，可以将该搜索作为一种交易类型添加到 `transactiontypes.conf` 中，以将其设置成可持续使用。

要了解有关配置交易类型的更多信息，请参阅本手册中的[“配置交易类型”](#)。

何时使用 stats 替代交易

交易并不是计算交易数据的聚合统计信息的最有效方法。如果您要在由单个字段的数据所定义的交易内计算聚合统计信息，可使用 `stats` 命令。

例如，如果您想要计算字段 `session_id` 定义交易的持续时间统计信息：

```
* | stats min(_time) AS earliest max(_time) AS latest by session_id | eval duration=latest-earliest | stats min(duration) max(duration) avg(duration) median(duration) perc95(duration)
```

同样，如果您想要按访问日志中的 `clientip` 计算点击次数：

```
sourcetype=access_combined | stats count by clientip | sort -count
```

另外，如果会话已通过 `cookie` 参数化，且您想要按访问日志中的 `clientip` 计算不同会话的数量：

```
sourcetype=access_combined | stats dc(cookie) as sessions by clientip | sort -sessions
```

有关使用搜索命令的更多信息，请阅读 `stats` 命令参考。

搜索交易

在 Splunk Web 或 CLI 使用交易搜索命令搜索交易。`transaction` 命令会生成可在报表中使用的事件分组。要使用 `transaction`，请调用交易类型（通过 [transactiontypes.conf 配置](#) 的类型），或者通过设置 `transaction` 命令的搜索选项定义搜索中的交易约束。

搜索选项

在搜索时返回的交易包含每个事件的原始文本、共享的事件类型以及字段值。交易还包含其他数据，它们存储在以下字段中：`duration` 和 `transactiontype`

- `duration` 包含交易的持续时间（交易的第一个事件与最后一个事件的时间戳之间的差异）。
- `transactiontype` 是交易的名称（根据交易的段落名称在 `transactiontypes.conf` 中定义）。

可以将 `transaction` 添加到任何搜索中。要获得最佳搜索性能，请精心创建搜索，然后将其发送到交易命令。有关更多信息，请参阅《搜索参考》手册中 `transaction` 命令的相关主题。

在 `transaction` 命令后加上以下选项。**注意：**有些 `transaction` 选项不能与其他选项结合使用。

[field-list]

- 这是一个用逗号分隔的字段列表，例如 `...|transaction host,cookie`
- 如果设置此选项，则所有事件都必须具有相同的字段，才会被视为属于相同交易。
- 具有通用字段名称但值不同的事件将不分在同一组。
 - 例如，如果您添加了 `...|transaction host`，则含有 `host=mylaptop` 的搜索结果决不能与含有 `host=myserver` 的搜索结果处于相同交易之中。
 - 不具有 `host` 值的搜索结果可以与具有 `host=mylaptop` 的结果处于相同交易之中。

match=closest

- 指定要用于交易定义的匹配类型。
- 当前唯一受支持的值为 closest。

maxspan=[<integer> s|m|h|d]

- 设置交易中各事件之间的最大跨度。
- 可以为秒、分钟、小时或天数。
 - 例如：5s、6m、12h 或 30d。
- 默认为 maxspan=-1，适用于“所有时间”的时间范围。

maxpause=[<integer> s|m|h|d]

- 指定各交易之间的最长暂停时间。
- 交易内各事件之间的暂停时间不得超过 maxpause。
- 如果为负值，将禁用 maxpause 约束。
- 默认为 maxpause=-1。

startswith=<string>

- 搜索或 eval 过滤表达式，如果某事件满足条件，则标志新交易的开始。
- 例如：
 - startswith="login"
 - startswith=(username=foobar)
 - startswith=eval(speed_field < max_speed_field)
 - startswith=eval(speed_field < max_speed_field/12)
- 默认为 ""。

endswith=<transam-filter-string>

- 搜索或 eval 过滤表达式，如果某事件满足条件，则标志交易的结束。
- 例如：
 - endswith="logout"
 - endswith=(username=foobar)
 - endswith=eval(speed_field < max_speed_field)
 - endswith=eval(speed_field < max_speed_field/12)
- 默认为 ""。

对于 startswith 和 endswith，<transam-filter-string> 使用以下语法进行定义："<search-expression>" | (<quoted-search-expression>) | eval(<eval-expression>)

- <search-expression> 是一个不包含引号的有效搜索表达式。
- <quoted-search-expression> 是一个包含引号的有效搜索表达式。
- <eval-expression> 是一个求值结果为布尔值的有效 eval 表达式。

示例：

- 搜索表达式：(name="foo bar")
- 搜索表达式："user=mildred"
- 搜索表达式：("search literal")
- eval 布尔表达式：eval(distance/time < max_speed)

交易和宏搜索

交易和宏搜索是一个功能非常强大的组合，可以替代交易搜索。创建交易搜索，然后使用 `$field$` 对其进行保存以允许替代。

若需搜索宏和交易组合的示例，请参阅[搜索宏示例](#)。

交易搜索示例

运行搜索，以将单个用户（或客户端 IP 地址）在时间范围内查看的所有网页分为一组。

该搜索从访问日志中获取事件，然后使用共享相同 `clientip` 值和彼此 5 分钟内发生的事件（在 3 小时时间跨度内）创建一个交易。

```
sourcetype=access_combined | transaction clientip maxpause=5m maxspan=3h
```

配置交易类型

任意事件系列都可以转化为交易类型。有关更多信息，请阅读本手册[“关于交易”](#)中的使用案例。

您可以通过 transactiontypes.conf 创建交易类型。请参阅下文了解配置详细信息。

有关配置文件的一般详细信息，请参阅《管理员手册》中的“关于配置文件”。

在 transactiontypes.conf 中配置交易类型

1. 创建 transactiontypes.conf 文件（位于 \$SPLUNK_HOME/etc/system/local/ 或您自己的自定义应用目录 \$SPLUNK_HOME/etc/apps/）。

2. 创建段落并在相应段落中列出每个交易的规范，以定义交易。使用以下属性：

```
[<transactiontype>]
maxspan = [<integer> s|m|h|d|-1]
maxpause = [<integer> s|m|h|d|-1]
fields = <comma-separated list of fields>
startswith = <transam-filter-string>
endswith=<transam-filter-string>
```

```
[<TRANSACTIONTYPE>]
```

- 创建任意多的交易类型，每个交易类型由一个段落名称和以下任意多的属性/值对表示。
- 使用段落名称 [<TRANSACTIONTYPE>] 在 Splunk Web 中搜索交易。
- 如果您没有为以下各属性指定一个条目，Splunk Enterprise 会使用默认值。

```
maxspan = [<integer> s|m|h|d|-1]
```

- 设置交易的最大时间跨度。
- 可以为秒、分钟、小时或天数。设置为 -1 时表示无限制。
 - 例如：5s、6m、12h 或 30d。
- 默认为 -1。

```
maxpause = [<integer> s|m|h|d|-1]
```

- 设置交易中各事件之间的最长暂停时间。
- 可以为秒、分钟、小时或天数。设置为 -1 时表示无限制。
 - 例如：5s、6m、12h 或 30d。
- 默认为 -1。

```
maxevents = <integer>
```

- 一个交易中的最大事件数。如果此值为负整数，则禁用此约束。
- 默认为 1000。

```
fields = <comma-separated list of fields>
```

- 如果设置此选项，则所有事件都必须具有相同的字段，才会被视为属于相同交易。
 - 例如：fields = host,cookie
- 默认为 " "。

```
connected= [true|false]
```

- 仅当 fields 不为空时相关。控制与某个交易字段不抵触且不一致的事件是打开新交易 (connected=true) 还是被添加到此交易中。
- 如果某个事件包含交易所需的字段，但这些字段都没有在此交易中实例化（通过前一个事件添加），则此事件可能不抵触且不一致。
- 默认为：connected = true

```
startswith = <transam-filter-string>
```

- 搜索或 eval 过滤表达式，如果某事件满足条件，则标志新交易的开始
- 例如：
 - startswith="login"
 - startswith=(username=foobar)
 - startswith=eval(speed_field < max_speed_field)
 - startswith=eval(speed_field < max_speed_field/12)
- 默认为：" "。

```
endswith=<transam-filter-string>
```

- 搜索或 eval 过滤表达式，如果某事件满足条件，则标志交易的结束
- 例如：
 - endswith="logout"
 - endswith=(username=foobar)
 - endswith=eval(speed_field > max_speed_field)

- `endswith=eval(speed_field > max_speed_field/12)`
- 默认为：" "。

对于 `startswith` 和 `endswith`，`<transam-filter-string>` 的语法如下：

```
"<search-expression>" | (<quoted-search-expression> | eval(<eval-expression>))
```

其中：

- `<search-expression>` 是一个不包含引号的有效搜索表达式。
- `<quoted-search-expression>` 是一个包含引号的有效搜索表达式。
- `<eval-expression>` 是一个求值结果为布尔值的有效 `eval` 表达式。例如，`startswith=eval(foo<bar*2)` 将匹配 `foo` 小于 `2 x bar` 的事件。

示例：

- `"<search-expression>" : startswith="foo bar"`
- `<quoted-search-expression> : startswith=(name="foo bar")`
- `<quoted-search-expression> : startswith=("search literal")`
- `eval(<eval-expression>) : eval(distance/time < max_speed)`

3. 在 Splunk Web 中使用交易命令来调用您所定义的交易（通过其交易类型名称）。在搜索期间，您可以覆盖配置规范。

有关搜索交易的更多信息，请参阅本手册中的[“搜索交易”](#)。

其他交易配置属性

`transactions.conf` 包含更多的几组属性，设计的目的是处理诸如多值字段和内存约束问题之类的情况。

内存约束问题的交易选项

```
maxopentxn=<int>
```

- 使用 LRU（最近最少使用的内存高速缓存算法）策略指定在开始退出交易之前保留在开放池中的尚未关闭交易的最大数量。
- 该属性的默认值从 `limits.conf` 的交易段落中读取。

```
maxopenevents=<int>
```

- 使用 LRU（最近最少使用的内存高速缓存算法）策略指定在交易开始发生之前属于开放交易一部分的最大事件数。
- 该属性的默认值从 `limits.conf` 的交易段落中读取。

```
keepevicted=[true|false]
```

- 是否输出已退出的交易。可以通过检查 `evicted` 字段的值（设置为 1 时代表已退出的交易）来区分已退出的交易与未退出的交易。
- 默认为 `keepevicted=false`。

呈现多值字段的交易选项

```
mvlist=[true|false] | <field-list>
```

- `mvlist` 属性控制交易的多值字段是 (1) 按照到达顺序排序的原始事件的列表还是 (2) 按照词典顺序排序的唯一字段值的集合。如果提供了以逗号/空格分隔的字段列表，则只有这些字段以列表形式呈现。
- 默认为：`mvlist=false`。

```
delim=<string>
```

- 用于分隔交易事件字段中的原始事件值的字符串。
- 默认为：`delim=" "`

```
nullstr=<string>
```

- 当将缺少的字段值呈现为交易中多值字段的一部分时使用的字符串值。
- 此选项仅适用于以列表形式呈现的字段。
- 默认为：`nullstr=NULL`

数据浓缩：查找和工作流动作

关于查找和工作流动作

查找和工作流动作使您可以通过与外部资源的交互丰富和扩展事件数据的实用性。

查找表

查找表使用您事件中的信息来确定如何从外部数据源（如静态表（CSV 文件）、基于 Python 和二进制的脚本和“应用键值存储”（KV 存储）集合）添加其他字段。CSV 内联查找表文件和使用 CSV 文件的内联查找定义均为一种数据集类型。基于 Python 和二进制的脚本以及应用键值存储（KV 存储）集合并非数据集类型。每个查找类型都可以选择性地基于时间信息添加字段。有关数据集的更多信息，请参阅“关于数据集”。

此功能的一个例子是 CSV 查找，它将获取事件中的 `http_status` 值，将该值与 CSV 文件中的相应定义匹配，然后将该定义作为新 `status_description` 字段的值添加到事件中。因此，对于 `http_status = 503` 的事件，该查找会将 `status_description = Service Unavailable, Server Error` 添加到该事件中。

当然，您还可以通过一些更加高级的方法来使用查找。例如，您可以：

- 安排使用报表的结果来填充静态查找表。
- 定义基于外部 Python 脚本（而不是查找表）的字段查找。例如，您可以创建使用 Python 脚本的查找，使其在给定主机名称时返回 IP 地址，并在给定 IP 地址时返回主机名称。
- 定义一个能将事件中的字段与 KV 存储查找中的字段匹配的查找，然后将字段返回给事件。您也可以设计能将搜索结果写入 KV 存储集合的搜索。
- 创建基于时间的查找（前提是您使用的查找表包含表示时间的字段值）。例如，如果您需要使用 DHCP 日志来根据用户的 IP 地址和事件时间戳识别网络中的用户，此方法会对您有所帮助。

有关更多信息，请参阅[配置 CSV 和外部查找](#)和[配置 KV 存储查找](#)。

工作流动作

工作流动作使您可以设置您数据中的特定字段与其他应用程序或 Web 资源之间的交互。一个非常简单的工作流动作是与 `IP_address` 字段关联的工作流动作，该动作启动后会根据 `IP_address` 值在单独的浏览器窗口中打开外部 WHOIS 搜索。

您还可以设置如下工作流动作：

- 仅应用于特定字段（而不是事件中的所有字段）。
- 仅应用于属于特定事件类型或事件类型组的事件。
- 通过事件下拉菜单和/或字段下拉菜单访问。
- 执行 HTTP GET 请求，以便将信息传递到外部 Web 资源（如搜索引擎或 IP 查找服务）。
- 执行 HTTP POST 请求以便将字段值发送到外部来源。例如，您可以设计如下工作流动作：将状态值发送到一个外部问题跟踪应用程序。
- 从所选事件中获取某些字段值并将其插入辅助搜索，以便用这些字段值填充辅助搜索并在辅助浏览器窗口中启动该搜索。

有关在 Splunk Web 中设置工作流动作的信息，请参阅[在 Splunk Web 中创建和维护工作流动作](#)。

使用字段查找将信息添加到事件中

通过此查找功能，您可以引用外部 CSV 文件中与您的事件数据字段匹配的字段。使用这种匹配，您可以向您的事件数据中添加更多的可搜索字段，从而丰富数据。字段查找可以基于任何字段（包括时间字段）或 Python 脚本的输出。CSV 内联查找表文件和使用 CSV 文件的内联查找定义均为一种数据集类型。

本主题介绍如何使用 Splunk Web 中的“查找”页面（位于**设置 > 查找**）执行以下操作：

- 列出现有查找表或上传新文件。
- 编辑现有查找定义或定义新的基于文件的查找或外部查找。
- 编辑现有自动查找或将新查找配置为自动运行。

有关查找的更多详细信息，请参阅本手册中的[配置 CSV 和外部查找](#)和[配置 KV 存储查找](#)。

列出现有查找表或上传新文件

在**设置 > 查找 > 查找表文件**中查看现有查找表文件，或单击“新增”上传更多 CSV 文件，以在基于文件的查找定义中使用。

注意：不支持行尾为 OSX OS9 及更早版本“Classic Macintosh”样式（也就是只有一个回车“\r”）的 CSV 文件。

上传新文件：

1. 从列表中选择**目标应用**。查找表文件保存在应用所在的目录中。例如：`$SPLUNK_HOME/etc/users/<username>/<app_name>/lookups/`
2. 为查找表文件指定一个**名称**。这将是您在查找定义中用来参考该文件的名称。
3. **浏览**要上传的 CSV 文件。
4. 单击**保存**。

编辑现有查找定义或定义新的基于文件的查找或外部查找

使用 **设置 > 查找 > 查找定义** 页面来定义查找表或编辑现有查找定义。您可以指定查找类型（基于文件的查找或外部查找）以及该查找是否基于时间。定义完查找表之后，您可以在搜索中调用查找（使用查找命令），也可以将查找配置为自动进行。

这等效于在 `transforms.conf` 中定义查找。

配置基于时间的查找

如果字段匹配取决于时间信息（即查找表中表示时间戳的字段），则基于文件的查找和外部查找也可以基于时间。

要配置基于时间的查找，选择 **配置基于时间的查找**，然后指定 *时间字段的名称*。还可以为此时间信息指定 `strftime` 格式，并为时间匹配指定偏移。

包括高级选项

在高级选项下，您可以指定：

- 每个输入查找值的最少匹配数。
- 每个输入查找值的最多匹配数。
- 指定输入少于最少匹配数时输出的默认值。

编辑现有自动查找或将新查找配置为自动运行

如果您希望向事件应用字段查找，可以将查找设置为自动运行，而无需调用查找命令。使用 **设置 > 查找 > 查找定义** 页面编辑或配置自动查找：

要编辑现有自动查找，选择相应查找并修改为该查找显示的字段中的值。

添加新查找以自动运行：

1. 在 **自动查找** 页面中选择 **新建**：
2. 选择 **目标应用**。
3. 选择要在字段查找中使用的 **查找表**。

这是您在“查找定义”页面上定义的查找定义的名称。

4. 在 **应用到** 菜单中，选择要应用查找的主机、来源或来源类型值并为其指定名称。
5. 在 **查找输入字段** 下方提供一对或多对输入字段。

第一个字段是您想在查找表中匹配的字段。第二个字段是应该匹配查找表字段的事件中的字段。例如，在事件中含有的 `ip_address` 字段与查找表中的 `ip` 字段匹配。因此您可以在自动查找定义中输入 `ip = ip_address`。

6. 在 **查找输出字段** 下方提供一对或多对输出字段。

第一个字段是您想输出给事件的相应字段。第二个字段是事件应该含有的输出字段的名称。例如，查找表可能有名为 `country` 的字段，您可能想将其作为 `ip_city` 输出给事件。

7. 您也可以选择 **覆盖字段值** 在每次运行查找时覆盖字段值。**注意：**这等效于在 `props.conf` 中配置字段查找。

HTTP 状态查找示例

本示例将引导您定义一个向 Web 访问事件添加两个信息字段（即 `status_description` 和 `status_type`）的静态查找。这样，您就可以在可能不知道具体错误代码的情况下搜索所需的事件。例如，您可以使用 `status="Server Error"`，而不是搜索所有服务器错误代码。

将查找表上载到 Splunk Enterprise

1. 下载 `http_status.csv` 文件：http_status.csv 以下为此文件的示例：

```
status,status_description,status_type
100,Continue,Informational
101,Switching Protocols,Informational
200,OK,Successful
201,Created,Successful
202,Accepted,Successful
203,Non-Authoritative Information,Successful
...
```

2. 返回到“搜索”应用，然后选择 **设置 > 查找**。



3. 在查找页面中，为查找表文件选择新增。

查找

创建和配置查找。

	操作
查找表文件 列出现有查找表或上传新文件。	新增
查找定义 编辑现有查找定义或定义新的基于文件的查找或外部查找。	新增
自动查找 编辑现有自动查找或将新查找配置为自动运行。	新增

4. 在新增页面中，
 - 为目标应用选择 `search`。
 - 浏览您之前下载的 CSV 文件。
 - 将查找表命名为 `http_status`。
 - 单击 **保存**。

新增

[查找](#) > [查找表文件](#) > 新增

目标应用 *

search

上传查找文件

[Choose File](#) [product_lookup](#)

选择纯文本 CSV 文件或使用 gzip 压缩的 CSV 文件。
可以通过浏览器上传的最大文件大小为 500MB。

目标文件名 *

http_status.csv

输入查找表文件在 Splunk 服务器上的名称。如果要上传 gzip 压缩的 CSV 文件，请输入以 ".gz" 结尾的文件名。如果您要上传纯文本 CSV 文件，建议使用以 ".csv" 结尾的文件名。

[取消](#)

[保存](#)

保存此文件之后，Splunk Enterprise 会将您带到以下视图中：

查找表文件

[查找](#) > [查找表文件](#)

应用上下文 [Search & Reporting \(search\)](#) 所有者 [任何](#)

☐ 仅显示在此应用上下文中创建的对象 [了解更多信息](#)

[新建](#)

显示 4 个项目中的 1-4

每页数量 [25](#)

路径	所有者	App	共享	状态	操作
/opt/splunk/etc/apps/simple_xml_examples/lookups/geomaps_data.csv	无所有者	simple_xml_examples	全局 权限	已启用	移动 删除
/opt/splunk/etc/users/admin/search/lookups/http_status.csv	admin	search	专用 权限	已启用	移动 删除
/opt/splunk/etc/users/admin/search/lookups/http_status_CN.csv	admin	search	专用 权限	已启用	移动 删除
/opt/splunk/etc/apps/search/lookups/prices.csv	admin	search	全局 权限	已启用	移动 删除

现在，让我们返回到 **设置 > 查找** 视图。要执行此操作，单击页面痕迹中的 **查找** 链接。您始终可以通过此方式导航回之前的视图。

查找表文件

查找 » 查找表文件

定义查找

1. 从 **设置 > 查找** 中，为 **查找定义** 选择 **新增**。在 **新增** 页面中：

新增

[查找](#) » [查找定义](#) » 新增

目标应用 *

search

名称 *

http_status

类型 *

基于文件

查找文件 *

http_status.csv

[创建和管理查找表文件。](#)

☐ 配置基于时间的查找

☐ 高级选项

取消

保存

2. 为 **目标应用** 选择 **search**。
3. 将查找定义的 **名称** 设置为 **http_status**。
4. 在 **类型** 下方，选择 **基于文件**。
5. 单击 **保存**。保存查找定义之后，Splunk Enterprise 会将您带到以下页面中：

查找定义

[查找](#) » [查找定义](#)

应用上下文 Search & Reporting (search) 所有者 任何

☐ 仅显示在此应用上下文中创建的对象 [了解更多信息](#)

新建

显示 10 个项目中的 1-10

每页结果数 25

名称	类型	所有者	App	共享	状态	操作
dnslookup	external	无所有者	system	全局 权限	已禁用 启用	复制
guid_lookup	file	无所有者	system	全局 权限	已禁用 启用	复制
http_status	file	admin	search	专用 权限	已启用 禁用	复制 移动 删除
http_status_CN	file	admin	search	全局 权限	已启用 禁用	复制 移动 删除
price_lookup	file	admin	search	专用 权限	已禁用 启用	复制 移动 删除
prices_lookup	file	admin	search	专用 权限	已禁用 启用	复制 移动 删除
prices_lookup_J	file	admin	search	专用 权限	已禁用 启用	复制 移动 删除
prices_lookup_K	file	admin	search	专用 权限	已启用 禁用	复制 移动 删除
prices_lookup_KO	file	admin	search	专用 权限	已启用 禁用	复制 移动 删除
sid_lookup	file	无所有者	system	全局 权限	已禁用 启用	复制

注意，您可以对查找定义执行一些操作。**权限**使您可以更改查找表的可访问性。可以**禁用**查找定义，以及将其**复制**和**移动**到其他应用。也可以**删除**查找定义。定义查找之后，您可以使用 `lookup` 命令在搜索中调用该查找，或者将其配置为自动运行。

将查找设置为自动运行

1. 返回到 **设置 > 查找** 视图，并为 **自动查找** 选择 **新增**。在 **新增** 页面中：

新增

[查找](#) » [自动查找](#) » 新增

目标应用 *

search

名称 *

http_status

查找表 *

http_status

应用到 *

sourcetype

已命名 *

access_combine

查找输入字段

status

=

status

删除

[添加另一个字段](#)

查找输出字段

status_description

=

status_description

删除

status_type

=

status_type

删除

[添加另一个字段](#)

☐ 重写字段值

取消

保存

2. 为目标应用选择 **search**。
3. 将查找命名为 **http_status**。
4. 从**查找表**下拉列表中选择 **http_status**。
5. 将该查找应用于 **sourcetype**，名为 **access_combined**。

应用到 *

sourcetype

已命名 *

access_combine

6. **查找输入字段**是指事件中要与查找表匹配的字段。此处，两者均被命名为 **status**（左侧是 CSV 列名称，右侧

查找输入字段

status

=

status

删除

是要匹配的字段）：

7. **查找输出字段**是指查找表中要添加到事件中的字段：**status_description** 和 **status_type**。左侧是 CSV 列

查找输出字段

status_description

=

status_description

删除

status_type

=

status_type

删除

名称，右侧是要匹配的字段。

8. 单击**保存**。

自动查找

[查找](#) » [自动查找](#)

已成功保存了 search 中的 "http_status"。

应用上下文

Search & Reporting (search)

所有者

任何

Q

☐ 仅显示在此应用上下文中创建的对象 [了解更多信息](#)

新建

显示 4 个项目中的 1-4

每页结果 25

名称	查找	所有者	App	共享	状态	操作
access_combined : LOOKUP-http_status	http_status status AS status OUTPUTNEW status_description AS status_description status_type AS status_type	admin	search	专用 权限	已启用	复制 移动 删除
access_combined : LOOKUP-http_status	http_status status AS status OUTPUTNEW status_description AS status_description status_type AS status_type	admin	search	专用 权限	已启用	复制 移动 删除
access_combined : LOOKUP-http_status_CN	http_status_CN status AS status OUTPUTNEW status_description AS status_description status_type AS status_type	admin	search	专用 权限	已启用	复制 移动 删除
access_combined_wcookie : LOOKUP-price_lookup_K	prices_lookup_K productId AS productId OUTPUTNEW price AS price productName AS productName	admin	search	专用 权限	已启用	复制 移动 删除

查找配置简介

基于事件中现有的字段值，查找从外部来源往事件中添加字段。一个简单的查找示例就是使用 CSV 文件的查找，且

该文件将可能的 HTTP 状态值（303、404、201 等）与各自定义相结合。如果有一个事件包含有 HTTP 状态值，则查找可以将此 HTTP 状态描述添加到该事件中。

您也可以使用查找来执行反向操作，以使查找将事件中的字段添加到查找表的行中。

您可以配置不同类型的查找。查找按两种方式加以区分：按数据来源和按信息类型。

有关数据集类型的更多信息，请参阅[关于数据集](#)。

查找类型	数据来源	描述
CSV 查找	CSV 文件	<p>用从 CSV 文件中提取的字段填充事件。因为 CSV 文件代表数据的静态表，也称其为“静态查找”。将 CSV 表中的每一列解释为字段的可能值。</p> <p>CSV 内联查找表文件和使用 CSV 文件的内联查找定义均为一种数据集类型。</p>
外部查找	外部源，如 DNS 服务器。	<p>使用 Python 脚本或二进制可执行文件往事件中填充来自外部来源的字段值。也称为“脚本式查找”。</p> <p>不是一种数据集类型。</p>
KV 存储查找	KV 存储集合	<p>将事件中的字段匹配到 KV 存储集合中的字段，并将该集合中的相应字段输出到事件中。</p> <p>不是一种数据集类型。</p>
地理空间查找	KMZ（即压缩的 Keyhole 标记语言）文件用于定义映射区域的边界，如国家、美国州与美国县。	<p>您使用地理空间查找来创建一个查询。Splunk 软件用此查询来配置 Choropleth 地图。地理空间查找将事件中的位置坐标匹配到 KMZ（Keyhole 标记语言）文件中的地理特征集合，并将提供在 KMZ 中编码的相应地理特征信息的字段输出到事件中，例如国家、州或县名。</p> <p>不是一种数据集类型</p>

配置 CSV 查找

CSV 查找将事件中的字段值与 CSV 文件所表示的静态表中的字段值匹配。然后将相应的字段值从静态表输出到事件中。也称为“静态查找”。CSV 内联查找表文件和使用 CSV 文件的内联查找定义均为一种数据集类型。有关数据集类型的更多信息，请参阅[关于数据集](#)。

将 CSV 表中的每一列解释为字段的可能值。

关于 CSV 文件

对于可用于 CSV 查找的 CSV 文件类型有一些限制：

- 由 CSV 文件表示的表必须至少有两列。其中一列应该表示那些属于事件中字段值集合的字段。该列不需要用与事件字段相同的名称。任何一列当它表示多值字段时都可以有同样值的多个实例。
- CSV 文件中不能包含非 utf-8 字符。支持纯 ascii 文本，因为任何字符集同时也是有效的 utf-8。
- 不支持以下内容：
 - 行尾为 OS X（OS 9 或更早版本）Macintosh 样式（只有一个回车 "\r"）的 CSV 文件
 - 标头行超过 4096 个字符的 CSV 文件。

创建 CSV 查找

- 为在 Splunk 部署中执行查找，添加 CSV 文件。CSV 文件必须位于以下位置之一：

```
$SPLUNK_HOME/etc/system/lookups
$SPLUNK_HOME/etc/apps/<app_name>/lookups
```

如果查找目录不存在，需创建该目录。

- 向 `transforms.conf` 中添加 CSV 查找段落。

如果想查找在全局范围可用，请向 `transforms.conf` 版本（位于 `$SPLUNK_HOME/etc/system/local/`）添加查找段落。如果想查找专用于特定的应用，请向 `transforms.conf` 版本（位于 `$SPLUNK_HOME/etc/apps/<app_name>/local/`）添加段落。

警告： 请不要在 `$SPLUNK_HOME/etc/system/default` 中编辑配置文件。

CSV 查找段落命名查找表和提供查找所使用的 CSV 文件的名称。它使用这些必需字段。

- `[<lookup_name>]`：查找的名称。
 - `filename = <string>`：查找参考的 CSV 文件的名称。
3. （可选）使用 `filter` 字段预过滤大的 CSV 查找表。

您可以需要预过滤非常大的 CSV 文件。若有此需求，可使用 `filter` 字段来限制搜索。请参阅[预过滤大的 CSV 查找表](#)。

4. （可选）为 CSV 查找设置字段/值匹配规则。

请参阅[在查找配置中添加字段匹配规则](#)。

5. （可选）如果 CSV 文件包含时间字段，则定义为时间限制的 CSV 查找。

请参阅[配置时间限制查找](#)。

6. （可选）在 `props.conf` 添加配置将 CSV 查找设为自动查找。

如果想自动查找在全局范围内可用，请向 `props.conf` 版本（位于 `$SPLUNK_HOME/etc/system/local/`）添加查找段落。如果想查找专用于特定的应用，请向 `props.conf` 版本（位于 `$SPLUNK_HOME/etc/apps/<app_name>/local/`）添加段落。

警告： 请不要在 `$SPLUNK_HOME/etc/system/default` 中编辑配置文件。

请参阅本手册中的[“设为自动查找”](#)。

7. 重新启动 Splunk Enterprise，使更改生效。

如果已设置自动查找，在重新启动后应该能在列于字段边栏的查找表中看到 `output` 字段。您可以从中选择要在每个匹配搜索结果中显示的字段。

预过滤大的 CSV 查找表

当您的 CSV 查找文件非常大时，如果您的查找必须搜索整个表来检索匹配的字段值，则性能会受到影响。如果您知道您只需要查找表中一个记录子集的结果，则可通过使用 `filter` 字段过滤掉所有不需要查看的记录，来提高搜索性能。`filter` 字段需要一个字符串，该字符串包含一个带有布尔表达式和/或比较运算符（`=`、`!=`、`>`、`<`、`<=`、`>=`、`OR`、`AND` 和 `NOT`）的搜索查询。当您运行一个调用该查找的搜索时，该查询将运行。

例如，如果您的查找配置具有 `filter = (CustID>500) AND (CustName="P*")`，则它只会尝试从表中的特定记录（`CustID` 值大于 500 且 `CustName` 值的开头字母为 P）里检索值。

注意： 当您使用这些命令来搜索 CSV 文件时，如果您结合使用 `WHERE` 子句和 `inputlookup` 以及 `inputcsv` 命令，则也可以从 CSV 表中过滤记录。

CSV 查找示例

此示例介绍如何为 `access_combined` 日志中的 HTTP 状态代码设置查找。在此示例中，您设计一个查找。此查找将事件中的 `status` 字段匹配查找表中的 `status` 列，且该查找表名为 `http_status.csv`。然后查找会向事件中输出相应的 `status_description` 和 `status_type` 字段。

以下是 `http_status.csv` 文件的文本。

```
status,status_description,status_type
100,Continue,Informational
101,Switching Protocols,Informational
200,OK,Successful
201,Created,Successful
202,Accepted,Successful
203,Non-Authoritative Information,Successful
204,No Content,Successful
205,Reset Content,Successful
206,Partial Content,Successful
300,Multiple Choices,Redirection
301,Moved Permanently,Redirection
302,Found,Redirection
303,See Other,Redirection
```

```

304,Not Modified,Redirection
305,Use Proxy,Redirection
307,Temporary Redirect,Redirection
400,Bad Request,Client Error
401,Unauthorized,Client Error
402,Payment Required,Client Error
403,Forbidden,Client Error
404,Not Found,Client Error
405,Method Not Allowed,Client Error
406,Not Acceptable,Client Error
407,Proxy Authentication Required,Client Error
408,Request Timeout,Client Error
409,Conflict,Client Error
410,Gone,Client Error
411,Length Required,Client Error
412,Precondition Failed,Client Error
413,Request Entity Too Large,Client Error
414,Request-URI Too Long,Client Error
415,Unsupported Media Type,Client Error
416,Requested Range Not Satisfiable,Client Error
417,Expectation Failed,Client Error
500,Internal Server Error,Server Error
501,Not Implemented,Server Error
502,Bad Gateway,Server Error
503,Service Unavailable,Server Error
504,Gateway Timeout,Server Error
505,HTTP Version Not Supported,Server Error

```

1. 将 `http_status.csv` **文件放入** `$SPLUNK_HOME/etc/apps/search/lookups/` **中。这表示查找是“搜索应用”特定的。**

2. 在 `transforms.conf` **文件（位于** `$SPLUNK_HOME/etc/apps/search/local` **）中放入：**

```

[http_status]
filename = http_status.csv

```

3. 重新启动 Splunk Enterprise，使更改生效。

现在您可以通过以下命令在搜索字符串中调用此查找：

- `lookup`：用于将字段添加到搜索结果的事件中。
- `inputlookup`：用于搜索查找表的内容。
- `outputlookup`：用于将搜索结果中的字段写入指定的 CSV 文件中。

有关如何操作的详细信息，请参阅《搜索参考》中关于这些命令的主题。

例如，您可以运行此搜索将 `status_description` 和 `status_type` 字段添加到包含 `status` 值的事件中，且这些值匹配 CSV 表中的 `status` 值。

```
... | lookup http_status status OUTPUT status_description, status_type
```

使用搜索结果填充 CSV 查找表

要使用报表结果来填充查找表，您可以编辑 `savedsearches.conf` 的本地副本或特定于应用的副本。

在报表段落（其中，搜索返回结果表）中：

1. 添加以下行以启用查找填充动作。

```
action.populate_lookup = 1
```

这将指示 Splunk 软件将结果表保存到 CSV 文件中。

2. 添加以下行以指定将查找表复制到何处。

```
action.populate_lookup.dest = <string>
```

`action.populate_lookup.dest` 值是 `transforms.conf` 中的查找名称，或搜索结果要复制到的 CSV 文件的路径。如果是 CSV 文件的路径，则此路径应相对于 `$SPLUNK_HOME`。

例如，如果您要将结果保存到全局查找表，则可能包括：

```
action.populate_lookup.dest = etc/system/lookups/myTable.csv
```

目标目录 `$SPLUNK_HOME/etc/system/lookups` 或 `$SPLUNK_HOME/etc/<app_name>/lookups` 应该已经存在。

3. 如果您希望在 Splunk Enterprise 启动时运行此搜索，添加以下行。

```
run_on_startup = true
```

如果搜索未在启动时运行，它将在下一个计划时间运行。对于填充查找表的计划的搜索，我们建议您设置为

```
run_on_startup = true
```

由于报表结果会复制到 CSV 文件，因此您可以按照设置 CSV 查找的方式来设置此查找。

配置外部查找

外部查找调用脚本来匹配事件中的字段和外部来源中的字段，并从外部来源中输出相应字段且将它们添加到您的事件中。

外部查找也常称为“脚本式查找”，因为他们都使用脚本。关于此类脚本如何工作的信息，请参阅[关于外部查找脚本](#)。

创建外部查找

以下介绍如何为 Splunk Enterprise 部署创建外部查找。如果您有 Splunk Cloud 并想定义外部查找，向 Splunk 支持提交问题。

1. 为在 Splunk 部署中执行查找，添加脚本。

其脚本必须位于以下两个位置之一：

- `$SPLUNK_HOME/etc/searchscripts`
- `$SPLUNK_HOME/etc/apps/<app_name>/bin`

2. 向 `transforms.conf` 中添加外部查找段落。

如果想查找在全局范围可用，请向 `transforms.conf` 版本（位于 `$SPLUNK_HOME/etc/system/local/`）添加查找段落。如果想查找专用于特定的应用，请向 `transforms.conf` 版本（位于 `$SPLUNK_HOME/etc/apps/<app_name>/local/`）添加段落。

警告：请不要在 `$SPLUNK_HOME/etc/system/default` 中编辑配置文件。

外部查找段落命名查找表、提供脚本和参数来执行查找、确定脚本类型和提供脚本支持的字段列表。它使用这些必需属性。

- `[<lookup_name>]`：查找的名称。
- `external_cmd = <string>`：为执行查找所发出的命令和参数。此处的命令必须为脚本名称，例如 `external_lookup.py`。此处的参数为要传递给脚本的字段名称，以空格分隔，如下所示：`clienthost clientip`。
- `external_type = [python|executable|kvstore|geo]`：用于查找的脚本类型。对于 Python 脚本，值为 `python`；对于二进制可执行文件，值为 `executable`。`kvstore` 和 `geo` 的值分别保留给 KV 存储查找和地理空间查找。
- `fields_list = <string>`：是外部查找支持的所有字段列表。字段必须用一个逗号后跟一个空格来分隔开。

4. （可选）为外部查找设置字段/值匹配规则。

请参阅本手册中的[“在查找配置中添加字段匹配规则”](#)。

5. （可选）如果外部查找的数据来源包含时间字段，则定义为时间限制的外部查找。

请参阅本手册中的[“配置时间限制的查找”](#)。

6. （可选）在 `props.conf` 添加配置将外部查找设为自动查找

如果想自动查找在全局范围内可用，请向 `props.conf` 版本（位于 `$SPLUNK_HOME/etc/system/local/`）添加查找段落。如果想查找专用于特定的应用，请向 `props.conf` 版本（位于 `$SPLUNK_HOME/etc/apps/<app_name>/local/`）添加段落。

警告：请不要在 `$SPLUNK_HOME/etc/system/default` 中编辑配置文件。

请参阅本手册中的[“设为自动查找”](#)。

7. 重新启动 Splunk Enterprise，使更改生效。

如果已设置自动查找，在重新启动后应该能在列于字段边栏的查找表中看到 `output` 字段。您可以从中选择要在每个匹配搜索结果中显示的字段。

外部查找示例

下面是 Splunk 软件所附带的一个外部查找示例。它匹配来自 DNS 服务器的信息。因其不含 `props.conf` 组件，所以不是自动查找。您可通过运行带 `[[Documentation: Splunk: SearchReference: Lookup] lookup` 命令的搜索访问它。

Splunk Enterprise 在 `$SPLUNK_HOME/etc/system/bin/` 中随附了一个名为 `external_lookup.py` 的脚本。这个 DNS 查找脚本将：

- 给定主机时，返回 IP 地址。
- 给定 IP 地址时，返回主机名称。

此脚本的配置位于 `$SPLUNK_HOME/etc/system/default/transforms.conf`。

```
[dnslookup]
external_cmd = external_lookup.py clienthost clientip
fields_list = clienthost,clientip
```

您可运行带 `lookup` 命令的搜索，该命令使用默认 `transforms.conf` 文件中的 `[dnslookup]` 段落。

```
sourcetype=access_combined | lookup dnslookup clienthost AS host | stats count by clientip
```

该搜索：

- 将外部查找表中的 `clienthost` 字段与事件 `</code>` 中的 `host` 字段匹配
- 返回一个表，该表提供每个 `clientip` 值的计数，这些值对应于 `clienthost` 匹配。

此搜索不会向您事件中添加字段。

您也可以设计一个执行反向查找的搜索，它此时会返回收到的每个 IP 地址的主机值。

```
sourcetype=access_combined | lookup dnslookup clientip | stats count by clienthost
```

注意此反向查找搜索不包含 AS 子句。这是因为 Splunk 自动提取 IP 地址作为 `clientip`。

关于外部查找脚本

您的外部查找脚本务必记得输入部分为空的 CSV 文件并输出一个已填充的 CSV 文件。传递给脚本的参数是这些输入和输出文件的标头。

在上面的 DNS 查找示例中，CSV 文件包含两个字段：`clienthost` 和 `clientip`。传递给此脚本的字段是在 `transforms.conf` 中指定使用 `external_cmd` 属性的字段。如果您不传递这些参数，此脚本返回一个错误。

```
external_cmd = external_lookup.py clienthost clientip
```

运行该搜索字符串后：

```
... | lookup dnsLookup clienthost
```

即是指示 Splunk 软件：

1. 使用您在 `transforms.conf` 中定义的查找表作为 `[dnsLookup]`
2. 向外部命令脚本中传递 `clienthost` 字段值来作为 CSV 文件。CSV 文件如下所示：

```
clienthost,clientip
work.com
home.net
```

此 CSV 文件含有作为列标题的 `clienthost` 和 `clientip`，但不含 `clientip` 值。此脚本包含两个标头，因为它们是您 `fields_list` 属性中指定的字段。该属性属于 `[dnslookup]` 段落，位于默认 `transforms.conf` 中。

然后，脚本会输出以下 CSV 文件，用于填充结果中的 `clientip` 字段：

```
host,ip
work.com,127.0.0.1
home.net,127.0.0.2
```

注意：编写脚本时，如果引用任何外部来源（如文件），则该引用必须相对于脚本所在的目录。

另请参阅

除了使用外部查找将外部源的字段添加到事件中之外，您可能也可以使用脚本式输入发送来自非标准源的数据去建立索引或将此数据准备用于解析。有关更多信息，请参阅《开发用于 Splunk Web 的视图和应用》中的“脚本式输入概述”。

配置 KV 存储查找

KV 存储查找会从“应用键值存储”（KV 存储）集合中提取字段填充您的事件。KV 存储查找可通过 REST 端点或通过使用下面的搜索命令调用：`lookup`、`inputlookup` 和 `outputlookup`。

也可将 KV 存储查找设置为“自动”查找。自动查找于搜索时间在后台运行，并自动将输出字段添加到具有正确匹配字段的事件中。您无需使用 `lookup` 命令来调用自动查找。请参阅本手册中的[“设为自动查找”](#)。

本主题介绍如何通过配置 `props.conf` 中的查找段落来设置和管理 KV 存储查找。与您使用 Splunk Web 设置查找文件的时候相比，配置文件为您在查找设计和行为上提供了更大程度的控制。但是，如果您没有访问 `.conf` 文件的权限或您倾向于尽可能通过 Splunk Web 维护查找，则可以使用设置 > 查找的页面来配置 KV 存储查找。请参阅本手册中的[“使用查找将信息添加到事件中”](#)。

Splunk Cloud 用户： 必须使用 Splunk Web 来定义查找。如果您的 Splunk Web 部署是受管部署，则必须在上载查找文件后请求 Splunk Support 进行重后，以使新上载的文件出现在用于定义查找的可用文件列表中。

您也可以定义查找：

- 将 CSV 文件中提取的字段填充到您的事件中。
- 使用 Python 脚本或二进制可执行文件将来自外部来源的字段值填充到您的事件中。

请参阅本手册的[“配置 CSV 和外部查找”](#)。

有关开发人员关注的 KV 存储查找的配置说明，请参阅 Splunk 开发人员门户的[“使用查找与 KV 存储数据”](#)。

关于 KV 存储集合

在创建 KV 存储查找前，Splunk 部署必须至少有一个已在 `collections.conf` 中定义的 KV 存储集合。请参阅 Splunk 开发人员门户的[“使用配置文件创建 KV 存储集合存储”](#)。

KV 存储集合是类似于数据库的数据容器。它们将您的数据存为键/值对。当创建 KV 存储查找时，集合应该至少有两个字段。其中一个字段应该拥有一组能与您的事件数据中的字段值相匹配的值，这样查找匹配才会发生。

当您通过 `lookup` 命令在一个搜索中调用查找时，可在您的搜索数据中指定一个字段来与您的 KV 存储集合中的字段匹配。当一个事件中的该字段值与您的 KV 存储集合中的指定字段值相匹配的时候，您的 KV 存储集合中其他字段的相应值会被添加到此事件中。

KV 存储字段不需要用与事件字段相同的名称。每个 KV 存储字段可以是多值字段。

注意： 当将 CSV 文件复制到索引器时，KV 存储集合继续在搜索头上工作。如果查找数据频繁更改，您会发现 KV 存储查找的性能优于同等 CSV 查找。

在 transforms.conf 中定义 KV 存储查找段落

`transforms.conf` KV 存储查找段落提供用作查找表的 KV 存储集合的位置。可以选择添加字段匹配规则和时间限制的查找规则。

若要 KV 存储查找在全局范围可用，将其查找段落添加到 `transforms.conf` 版本（位于 `$SPLUNK_HOME/etc/system/local/`）。如果想查找专用于特定的应用，请向 `transforms.conf` 版本（位于 `$SPLUNK_HOME/etc/apps/<app_name>/local/`）添加段落。

警告： 请不要在 `$SPLUNK_HOME/etc/system/default` 中编辑配置文件。

KV 存储查找段落格式

当您添加 KV 存储查找段落到 `transforms.conf` 时，应遵循以下格式。

```
[<lookup_name>]
external_type = kvstore
collection = <string>
fields_list = <string>
filter = <string>
```

- `[<lookup_name>]` 是查找的名称。
- `external_type` 应设置为 `kvstore`（如果您正在定义 KV 存储查找）。
- `collection` 是与查找关联的 KV 存储集合的名称。
- `fields_list` 是 KV 存储查找支持的所有字段列表。字段必须用一个逗号后跟一个空格来分隔开。字段可以是您在 KV 存储集合中拥有的键和值的任意组合。

默认情况下，每条 KV 存储记录都有唯一的键 ID，并存储在内部的 `_key` 字段中。如果您希望能够通过 KV 存储查找修改特定的记录，则将 `_key` 添加到 `fields_list` 中的字段列表。然后您可以在查找操作中指定键 ID 值。

当您使用 `outputlookup` 命令写入 KV 存储而未指定键 ID 时，会为您自动生成一个键 ID。

- `filter`：针对特别大的 KV 存储集合，可以选择性的使用此属性来提高搜索性能。请参阅[“预过滤大的 KV 存储集合”](#)。

配置 KV 存储查找

如果您有 Splunk Cloud 并想定义 KV 存储查找，向 Splunk 支持提交问题。如果您有 Splunk Enterprise，可以执行以下操作。

1. 在 `collections.conf` 中定义 KV 存储集合。请参阅 Splunk 开发人员门户的“使用配置文件创建 KV 存储集合存储”。

2. 遵循[上文介绍](#)的段落格式，在 `transforms.conf` 中创建 KV 存储查找段落。

如果想查找在全局范围可用，请向 `transforms.conf` 版本（位于 `$SPLUNK_HOME/etc/system/local/`）添加查找段落。如果想查找专用于特定的应用，请向 `transforms.conf` 版本（位于 `$SPLUNK_HOME/etc/apps/<app_name>/local/`）添加段落。

警告：请不要在 `$SPLUNK_HOME/etc/system/default` 中编辑配置文件。

3. （可选）使用 `filter` 属性预过滤非常大的 KV 存储查找表。

对于特别大的 KV 存储集合，您可以使用 `filter` 属性来限制搜索，从而加速查找搜索过程。请参阅[“预过滤大的 KV 存储集合”](#)。

4. （可选）为 KV 存储查找设置字段/值匹配规则。

请参阅本手册中的[“在查找配置中添加字段匹配规则”](#)。

5. （可选）如果 KV 存储集合包含时间字段，则定义为时间限制的 KV 存储查找。

请参阅本手册中的[“配置时间限制的查找”](#)。

6. （可选）在 `props.conf` 添加配置将 KV 存储查找设为自动查找。

如果想自动查找在全局范围内可用，请向 `props.conf` 版本（位于 `$SPLUNK_HOME/etc/system/local/`）添加查找段落。如果想查找专用于特定的应用，请向 `props.conf` 版本（位于 `$SPLUNK_HOME/etc/apps/<app_name>/local/`）添加段落。

警告：请不要在 `$SPLUNK_HOME/etc/system/default` 中编辑配置文件。

请参阅本手册中的[“设为自动查找”](#)。

7. 保存对 `.conf` 文件的更改。

8. 重新启动 Splunk Enterprise，使更改生效。

如果已设置自动查找，在重新启动后应该能在列于字段边栏的查找表中看到 `output` 字段。您可以从中选择要在每个匹配搜索结果中显示的字段。

预过滤大的 KV 存储集合

当您的 KV 存储集合非常大时，如果您的查找必须搜索整个集合来检索匹配的字段值，则性能会受到影响。如果您知道您只需要查找表中一个记录子集的结果，则可通过使用 `filter` 属性过滤掉所有不需要查看的记录，来提高搜索性能。

`filter` 属性需要一个字符串，该字符串包含一个带有布尔表达式和/或比较运算符（`==`、`!=`、`>`、`<`、`<=`、`>=`、`OR`、`AND` 和 `NOT`）的搜索查询。当您运行一个调用该查找的搜索时，该查询将运行。

例如，如果您的查找配置具有 `filter = (CustID>500) AND (CustName="P*")`，则它只会尝试从 KV 存储集合中的特定记录（`CustID` 值大于 500 且 `CustName` 值的开头字母为 P）里检索值。

注意：如果您不想在查找定义中安装过滤器，则可以通过结合使用 `where` 子句与 `inputlookup` 命令来获得类似的效果。

KV 存储查找示例

有一个称为 `employee_info` 的 KV 存储查找。它位于 `$SPLUNK_HOME/etc/system/bin/` 中。

```
[employee_info]
external_type = kvstore
collection = kvstorecoll
fields_list = _key, CustID, CustName, CustStreet, CustCity, CustZip
filter = (CustID>500) AND (CustName="P*")
```

`employee_info` 查找获取事件中的员工 ID 然后向事件中输出相应的员工信息，例如员工姓名、住址、城市和邮政编码。该查找针对称为 `kvstorecoll` 的 KV 存储集合。`filter` 将查找查询限制为客户 ID 大于 500 和客户名称开头字母为 "P" 的记录。

要了解如何通过添加配置到 `props.conf` 的方式来将此 KV 存储查找设为“自动”查找，请参阅本手册中的[“设为自动查找”](#)。

搜索命令和 KV 存储查找

在您保存 KV 存储查找段落并重新启动 Splunk Enterprise 后，您可以通过搜索命令与新的 KV 存储查找进行交互。

使用 `lookup` 来匹配 KV 存储集合中的值与搜索结果中的字段值，然后将相应的字段值输出至这些结果中。本搜索使用在前面的使用示例中定义的 `employee_info` 查找。

```
... | lookup employee_info CustID AS ID OUTPUT CustName AS Name | ...
```

它匹配 `kvstorecoll` 中的员工 id 值与您事件中的员工 id 值，然后将相应的员工姓名值输出到您的事件中。

您可使用 `inputlookup` 搜索命令在 KV 存储集合的内容中进行搜索。有关相关示例，请参阅《搜索参考》中关于 `inputlookup` 的主题。

您可使用 `outputlookup` 搜索命令将搜索管道的搜索结果写入 KV 存储集合。有关相关示例，请参阅《搜索参考》中关于 `outputlookup` 的主题。

您也可以在 Splunk 开发人员门户的“使用查找与 KV 存储数据”中查找 KV 存储查找搜索的若干个示例。

配置地理空间查找

您使用地理空间查找来创建查询。Splunk 软件可利用查询返回的结果来生成 Choropleth 地图可视化。如果没有相应地理空间查找生成的数据，则无法呈现 Choropleth 地图。

地理空间查找将事件中的位置坐标匹配到地理特征集合（KMZ 文件，也称为 Keyhole 标记语言文件）中的位置坐标范围，并将提供在特征集合中编码的相应地理特征信息的字段输出到事件中。此地理信息通常代表一个地理区域，且该区域可与同类型的其他地理区域共享边界，如国家、州、省或县。

Splunk 针对美国和其他国家提供两种地理空间查找，使您可以呈现以下对象的 Choropleth 地图：

- 美国，划分为多个州行政区。
- 而世界划分为多个国家。

本主题介绍如何创建其他地理空间查找，以将 Choropleth 地图细分为其他类型的区域（县、省、时区等）。

有关 Choropleth 地图和地理数据可视化的详细信息，请参阅《仪表板和可视化》手册中的“映射数据”。

FeatureId 和 featureCollection 字段

地理空间查找与其他查找类型不同，因为他们被设计为始终输出以下两个字段：`featureId` 和 `featureCollection`。`featureId` 指特征的“名称”，如 "California"、"CA" 或其他在特征集合中编码的内容。`featureCollection` 字段提供包含该特征的查找的名称。

如果您将地理空间查找的输出通过管道直接传递给 `geom` 命令，则无需为此命令提供此查找的名称。`geom` 命令检测事件中的 `featureId` 和 `featureCollection` 字段，并使用该查找来生成 Splunk 软件生成 Choropleth 地图所需的地理数据结构。但是，注意地理数据结构可能很大；因为太大，所以强烈建议不要将事件通过管道传递给 `geom` 命令，因为每个事件都会附加地理数据结构。相反，您应先针对地理空间查找结果执行 `stats` 命令，然后只针对“归结后”的聚合统计数据执行 `geom` 命令，如按 `featureId` 计数。

在 transforms.conf 中定义地理空间查找段落

地理空间查找段落提供用作查找表的地理特征集合的位置。可以选择性的添加：

- `feature_id_element` 属性。
- 字段匹配规则。
- 时间限制的查找规则。

有关详细信息，请参阅“地理空间查找段落格式”。

若要地理空间查找在全局范围可用，将其查找段落添加到 `transforms.conf` 版本（位于 `$SPLUNK_HOME/etc/system/local/`）。如果想查找专用于特定的应用，请向 `transforms.conf` 版本（位于 `$SPLUNK_HOME/etc/apps/<app_name>/local/`）添加段落。

警告：请不要在 `$SPLUNK_HOME/etc/system/default` 中编辑配置文件。

地理空间查找段落格式

当您添加地理空间查找段落到 `transforms.conf` 时，应遵循以下格式。

```
[<lookup_name>]
external_type = geo
filename = <name_of_KMZ_file>
feature_id_element = <XPath_expression>
```

- `[<lookup_name>]` 是查找的名称。
- `external_type` 应设置为 `geo`（如果您正在定义地理空间查找）。
- `filename` 是所使用的 KMZ 文件的名称。KMZ 文件也称为“地理特征集合”。
 - 提供两个特征集合：`geo_us_states` 针对美国；`geo_countries` 针对世界其他各国。
 - 您可以选择性地地上载其他地区的地理特征集合和特征类型，如美国各县或欧洲各省。
- `feature_id_element` 是可选属性。这是一个 XPath 表达式，定义了从 KML 文件中的多边形元素到其他包含此特征名的 XML 元素和属性的路径。它适用于未使用名为 Placemark 元素的典型风格的场景。
 - 有些情况需要使用 `feature_id_element`，如查找生成的 `featureID` 字段为空字符串，或特征集合默认返回错误的特征。在后一种情况中，您所需的特征可能是默认特征的对等体或位于默认特征的相对位置。
 - 要确定所需的路径，请学习地理特征集合的相关内容。集合中的每个特征都带有 `<Placemark>` 标签，且每个 `<Placemark>` 包含一个 `<name>`。查找会将此字段写出为 `featureId` 字段。
 - `feature_id_element` 的默认设置是 `/<placemark>/<name>`。

配置地理空间查找

- 1.（可选）如果您需要使用一个集合而不用 `geo_us_states` 或 `geo_countries`，请将所需地理特征集合上载至 Splunk 部署。

地理特征集合编码为 KMZ（Keyhole 标记语言）文件。

在“设置”中上载特征集合。导航到 **设置 > 查找 > 查找表文件**。

如果您有 KML 文件，可将它压缩并将其后缀 `.zip` 修改为 `.kmz`，从而将它转换为 KMZ 文件。

2. 在 `transforms.conf` 中创建一个地理空间查找段落，并遵循上文[“地理空间查找段落格式”](#)中所介绍的段落格式。

如果想查找在全局范围可用，请向 `transforms.conf` 版本（位于 `$SPLUNK_HOME/etc/system/local/`）添加查找段落。如果想查找专用于特定的应用，请向 `transforms.conf` 版本（位于 `$SPLUNK_HOME/etc/apps/<app_name>/local/`）添加段落。

警告：请不要在 `$SPLUNK_HOME/etc/system/default` 中编辑配置文件。

- 3.（可选）为地理空间查找设置字段/值匹配规则。

请参阅本手册中的[“在查找配置中添加字段匹配规则”](#)。

- 4.（可选）在 `props.conf` 添加配置将地理空间查找设为自动查找。

如果想自动查找在全局范围内可用，请向 `props.conf` 版本（位于 `$SPLUNK_HOME/etc/system/local/`）添加查找段落。如果想查找专用于特定的应用，请向 `props.conf` 版本（位于 `$SPLUNK_HOME/etc/apps/<app_name>/local/`）添加段落。

警告：请不要在 `$SPLUNK_HOME/etc/system/default` 中编辑配置文件。

请参阅本手册中的[“设为自动查找”](#)。

5. 保存对 `.conf` 文件的更改。

6. 重新启动 Splunk Enterprise，使更改生效。

如果已设置自动查找，在重新启动后应该能在列于字段边栏的查找表中看到 `output` 字段。您可以从中选择要在每个匹配搜索结果中显示的字段。

地理空间查找示例

有一个称为 `geo_us_states` 的地理空间查找。它位于 `$SPLUNK_HOME/etc/system/bin/` 中。

```
[geo_us_states]
external_type=geo
filename=geo_us_states.kmz
```

此查找处理包含美国州的地理特征集合。

要使用此查找构建 Choropleth 地图，您需要创建一个搜索对它进行查询，并让此搜索返回可用于生成该地图的结果。

果。此搜索需要完成以下操作：

- 指示一个事件数据来源。
- 通过将事件中的字段与 KMZ 文件中的字段相匹配的方式查询该查找。
- 包含一个使用查找的地理输出字段 `featureId` 聚合数据的转换命令。
- 使用 `geom` 搜索命令来生成可用于创建 Choropleth 地图的数据。

这是一个局部的 Choropleth 地图查询。它符合以上列出的 Choropleth 地图查找搜索的四大要求中的前两个要求。它会针对特征集合的特征返回经度和纬度。

```
sourcetype=crime_data cc=USA | lookup geo_us_states latitude, longitude
```

该查找的输出应该类似如下示例：

_featureIdField	featureId	featureCollection	纬度	经度
featureId	AK	geo_us_states	y	x
featureId	AL	geo_us_states	y	x

您可以更新此搜索以显示 `geom` 命令的结果。请注意，`geom` 命令之前要加一个转换命令操作，如本例中涉及 `count` 的这个。

这是一个全局的 Choropleth 地图查询。它按美国各州检索犯罪事件的计数，并将每个州的几何体添加为 `geom` 列。

```
sourcetype=crime_data cc=USA | lookup geo_us_states latitude, longitude | stats count by featureId | geom
```

该搜索的输出应该类似如下示例：

_featureIdField	featureId	geom	count
featureId	AK	{...}	10
featureId	AL	{...}	15

`_featureIdField` 是一个与 `geomfilter` 后期处理搜索命令结合使用的隐藏字段。当您运行包含该字段的搜索时，它使 `geomfilter` 可以获知哪个字段包含 `featureId` 值，即使 `featureId` 已经重新命名为其他名称。

例如，假设您将 `featureId` 重新命名为 `state`。如果您运行 `geomfilter`，它会咨询存储在搜索 `dispatch` 文件夹中的搜索结果，查看 `_featureIdField` 列，并会在此找到 `state` 值。这会导致它去查找其计算所需的 `featureId` 值，查找目标为 `state` 列。

有关地理空间查找搜索查询的详细信息，请参阅《仪表板和可视化》手册中的“映射数据”。

在查找配置中添加字段匹配规则

这些属性提供查找中的字段匹配规则。它们适用于所有三种查找类型。将它们添加到查找的 `transforms.conf` 段落中。

属性	类型	描述	默认
<code>max_matches</code>	整数	从您的事件输入到查找表中的每个值可能匹配的最大数目。范围是 1-1000。如果没有指定 <code>time_field</code> 属性，Splunk 软件使用按文件次序排在第一个的 <code><integer></code> 条目。如果没有指定 <code>time_field</code> 属性（因为这是一个时间限制查找），Splunk 软件使用按时间降序排列排在第一个的 <code><integer></code> 条目。换句话说，最多可以匹配 <code><max_matches></code> 。当超过该数目时，Splunk 软件使用与查找值最接近的匹配。	1000（如果未指定 <code>time_field</code> 属性）。1（如果指定 <code>time_field</code> 属性）。
<code>min_matches</code>	整数	从您的事件输入到查找表中的每个值可能匹配的最小数目。可使用 <code>default_match</code> 来帮助处理对于任意给定输入其匹配数小于 <code>min_matches</code> 的情况。	0（用于非时间限制查找和时间限制查找，表示如果没有找到匹配的话，什么都不会输出到您的事件中）。
<code>default_match</code>	字符串	当 <code>min_matches</code> 大于 0 而对于任意给定输入 Splunk 平台找到的匹配数小于 <code>min_matches</code> 时，它会一次或多次提供 <code>default_match</code> 值直到达到 <code>min_matches</code> 阈值。	空字符串

case_sensitive_match	布尔	若要在匹配查找表字段时考虑大小写，则设置为 true；否则，设置为 false。 注意： 您无需为 KV 存储查找设置该属性。KV 存储查找始终区分大小写。	True
match_type	字符串	允许一个或多个字段（以一个逗号后跟一个空格分隔的列表形式排列）的非精确匹配。格式是 match_type = <match_type>(<field_name1>, <field_name2>,...<field_nameN>)。将 match_type 设为 WILDCARD 以应用通配符匹配，或将其设为 CIDR 以应用 CIDR 匹配（专门用于 IP 地址值）。	EXACT（无需指定）

配置基于时间的查找

如果查找表中有表示时间的字段，可使用其来创建时间限制查找（也称为时间查找）。您可将所有四种类型的查找配置为时间限制查找。

要创建时间限制查找，将以下行添加到 transforms.conf 的查找段落中：

```
time_field = <field_name>
time_format = <string>
```

如果存在 time_field 属性，默认 max_matches = 1 并且 Splunk 平台应用降序排列中的首个匹配条目。有关 max_matches 的详细信息，请参阅本手册中的“在查找配置中添加字段匹配规则”。

time_format 属性指定 time_field 属性的 strftime() 格式。time_format 属性的默认值是 %s.%Q，您可在此处输入以秒为单位的 Unix epoch 时间值 (%s)，也可选择包括以毫秒为单位的时间值 (%Q)。

注意：您可以使用部分非标准的日期时间 strftime() 格式。例如，当您定义 ISO 8601 时间戳（以秒为单位的 Unix epoch 时间值）时，您可以使用 time_format = '%s.%Q'，其中，%s 代表秒，%Q 代表毫秒。请参阅《数据导入手册》中的“配置时间戳识别”的子标题“增强的 strftime() 支持”。

时间限制查找发生匹配时，您还可以指定事件可能晚于查找条目的最短和最长时间的偏移。要执行此操作，将以下行添加到段落中：

```
max_offset_secs = <integer>
min_offset_secs = <integer>
```

默认没有最大偏移。默认最小偏移是 0。

基于时间的查找示例

下面是一个 CSV 查找示例。在本示例中，该查找使用 DHCP 日志来根据用户的 IP 地址和时间戳识别网络中的用户。DHCP 日志位于 dhcp.csv 文件中。此文件包含时间戳、IP 地址以及用户的名称和 MAC 地址。

1.在 transforms.conf 文件中，放入：

```
[dhcpLookup]
filename = dhcp.csv
time_field = timestamp
time_format = %d/%m/%y %H:%M:%S
```

2.在 props.conf 文件中，将该查找设为自动查找：

```
[dhcp]
LOOKUP-table = dhcpLookup ip mac OUTPUT user
```

有关更多信息，请参阅本手册中的[“设为自动查找”](#)。

3.重新启动 Splunk Enterprise。

设为自动查找

当您在 transforms.conf 中创建查找配置时，您通过运行引用它的搜索来调用它。但是，您也可以选择性地创建一个额外 props.conf 配置来将该查找设为“自动”查找。这意味着它会于搜索时间在后台运行，并自动将输出字段添加到具有正确匹配字段的事件中。

您可以将所有类型的查找设为自动查找。但是，对于 KV 存储查找，您必须在通过 props.conf 将它们配置为自动查找之前完成一个额外的设置步骤。请参阅[启用 KV 存储集合的复制](#)。

您创建的每个自动查找配置仅限于属于特定主机、数据来源或来源类型的事件。自动查找可以访问属于您或您已共享的查找表中的所有数据。

当您的查找是自动查找时，您无需调用其 `transforms.conf` 配置（通过 `lookup` 命令）。

props.conf 中的自动查找格式

`props.conf` 中的自动查找配置：

- 参考您在 `transforms.conf` 中配置的查找表。
- 指定查找应该在查找表中匹配的您事件中的字段。
- 指定查找应该从查找表输出到您事件中的相应字段。

在搜索时间，`LOOKUP-<class>` 配置识别查找，并描述应如何将该查找应用于您的事件。要创建一个自动查找，请遵循以下语法：

```
[<spec>]
LOOKUP-<class> = $TRANSFORM <match_field_in_lookup_table> OUTPUT|OUTPUTNEW <output_field_in_lookup_table>
```

- 段落标题是 `[<spec>]`。`<spec>` 可以是：
 - `<sourcetype>`，事件的来源类型。
 - `host::<host>`，其中 `<host>` 是事件的主机或主机匹配模式。
 - `source::<source>`，where `<source>` 是事件的数据来源或来源匹配模式。
- `<spec>` 不能使用正则表达式语法。
- `$TRANSFORM`：参考定义查找表的 `transforms.conf` 段落。
- `match_field_in_lookup_table`：此变量是查找表中的字段，该字段匹配具有和 `props.conf` 段落同样的主机、数据来源或来源类型的事件字段。如果事件中的匹配字段名称与查找表中的匹配字段名称不一样，使用下面第 3 步中指定的 AS 子句。
- `output_field_from_lookup_table`：查找表中您想要添加到您事件中的相应字段。如果事件中的输出字段名称与查找表中的输出字段名称不一样，使用下面第 3 步中指定的 AS 子句。

您可以在查找的任一侧设置多个字段。例如，您可以设置：

```
$TRANSFORM <match_field_in_lookup_table1>, <match_field_in_lookup_table>OUTPUT|OUTPUTNEW
<output_field_from_lookup_table1>, <output_field_from_lookup_table2>
```

还可以让一个匹配字段返回两个输出字段、三个匹配字段返回一个输出字段，等等。

如果不包括 `OUTPUT|OUTPUTNEW` 子句，Splunk 软件会将查找表中的所有字段名称和值添加到您的事件中。如果使用 `OUTPUTNEW`，Splunk 软件仅能添加对事件来说是“全新”的输出字段。如果使用 `OUTPUT`，则会覆盖事件中已经存在的输出字段。

如果查找表中的“匹配”字段名与事件中的不一致，或您想“重命名”输出字段或已添加到您事件中的字段，使用 `AS` 子句：

```
[<stanza name>]
LOOKUP-<class> = $TRANSFORM <match_field_in_lookup_table> AS <match_field_in_event>OUTPUT|OUTPUTNEW
<output_field_from_lookup_table> AS <output_field_in_event>
```

例如，如果查找表中有一个字段名为 `dept` 且您想让自动查找将它添加到您的事件中并重命名为 `department_name`，则将 `department_name` 设置为 `<output_field_in_event>` 的值。

注意：您可以有多个 `LOOKUP-<class>` 配置，即使是在单个 `props.conf` 段落中。每个查找都应该具有自己的唯一查找名称。例如，如果您有多个查找，可以将它们命名为 `LOOKUP-table1`、`LOOKUP-table2` 等，依此类推。

您也可以有多个不同的 `props.conf` 自动查找段落，每一个都引用 `transforms.conf` 中的同一个查找段落。

在 props.conf 中创建自动查找段落

1. 创建一个引用查找所关联的主机、数据来源或来源类型的段落标题。

2. 向您已确定或创建的段落中添加 `LOOKUP-<class>` 配置。

正如前面的内容中所述，此配置指定以下事项：

- 它应将事件中的哪些字段匹配到查找表中的字段。
- 它应从查找表中将哪些相应的输出字段添加到您的事件中。

务必确保 `<class>` 值唯一。如果两个或多个自动查找配置的 `<class>` 名称一样，您可能就会遇到麻烦。请参阅[“在自动查找配置中不可使用一样的名称”](#)。

3. (可选) 如果查找表中的“匹配”字段名与事件中的不一致, 在配置中添加 AS 子句, 或如果想“重命名”输出字段或已添加到您事件中的字段, 使用 AS 子句。

4. 重新启动 Splunk Enterprise, 使更改生效。

如果已设置自动查找, 在重新启动后应该能在列于字段边栏的查找表中看到 output 字段。您可以从中选择要在每个匹配搜索结果中显示的字段。

启用 KV 存储集合的复制

在 Splunk Enterprise 中, KV 存储集合默认不会将软件包复制给索引器, 且查找会运行于本地搜索头而非远程节点。在您启用了 KV 存储集合的复制后, 可以在索引器上运行查找。这反过来又可以让您针对 KV 存储集合使用自动查找。

要启用 KV 存储集合的复制并允许将针对此集合的查找设为自动:

1. 打开 collections.conf。

2. 在该集合的段落中, 将 replicate 设为 true。

此参数的默认设置为 false。

3. 重新启动 Splunk Enterprise, 使更改生效。

警告: 如果您的索引器所运行的 Splunk Enterprise 版本低于 6.3, 则运行自动查找的尝试会失败且会返回“查找不存在”错误。您必须将索引器的版本升级到 6.3 或以上, 才能使用此功能。

有关更多信息, 请参阅 Splunk 开发人员门户中的“使用配置文件创建 KV 存储集合”。

自动 KV 存储查找的配置示例

此配置引用了本手册[“配置 KV 存储查找”](#)中的 KV 存储查找配置示例。KV 存储查找在 transforms.conf 中进行了定义, 具体位置为 employee_info 段落。

```
[access_combined]
LOOKUP-http = employee_info cust_ID AS CustID OUTPUT CustName AS cust_name, CustCity AS cust_city
```

此配置使用 transforms.conf 中的 employee_info 查找将字段添加到您的事件中。具体来说, 它将 cust_name 和 cust_city 字段添加到所有 access_combined 事件, 只要该事件的 cust_ID 值匹配 custID 值 (位于 kvstorecoll KV 存储集合)。它也使用 AS 子句来:

- 查找 KV 存储集合中的匹配字段。
- 在将输出字段添加到事件时重命名这些字段。

在自动查找配置中不可使用一样的名称

在 props.conf 中, 您通过 LOOKUP-<class> 属性确定表格定义。如果所有 props.conf 查找定义都具有不同的 <class> 名称, 这是最佳效果。该做法降低了出错率。

如果您为两个或更多个查找指定了相同的 <class> 名称, 可能会陷入麻烦 (除非您清楚自己在做什么):

- 如果两个或更多个同名查找共用相同的 props.conf 段落 (相同主机、数据来源或来源类型), 则 fields.conf 中包含该段落的第一个查找将覆盖其他查找。所有具有相同主机、来源或来源类型的查找都应该具有不同的名称。
- 如果具有不同主机、来源或来源类型的查找共享相同名称, 结果可能是在任意指定时间点上看上去似乎只有一个查找正常工作。您可能需要这样设置, 但在大多数情况下并不方便。

例如, 假设您有以下两个名为 LOOKUP-table 的查找:

```
[host::machine_name]
LOOKUP-table = logs_per_day host OUTPUTNEW average_logs AS logs_per_day

[sendmail]
LOOKUP-table = location host OUTPUTNEW building AS location
```

这两个查找之间的所有重叠事件仅受其中一个查找影响。换言之:

- 与主机匹配的事件获得主机查找。
- 与来源类型匹配的事件获得来源类型查找。
- 同时与上述两者匹配的事件获得主机查找。

如果将查找命名为 LOOKUP-table, 则表示该查找将实现 "table" 所描述的某种目的或操作。在本例中, 这些查找实现

不同的目标。一个查找确定每天的日志信息，另一个查找与位置有关。对它们进行重命名。

```
[host::machine_name]
LOOKUP-table = logs_per_day host OUTPUTNEW average_logs AS logs_per_day

[sendmail]
LOOKUP-location = location host OUTPUTNEW building AS location
```

现在，您有了两个不同配置，但并不冲突。

在 Splunk Web 中创建和维护工作流动作

使用工作流动作可以在索引或提取字段和其他 Web 资源之间进行广泛的交互。工作流动作具有十分广泛的应用。例如，可以定义工作流动作来执行以下操作：

- 基于在某个事件中找到的 IP 地址执行外部 WHOIS 查找。
- 使用 HTTP 错误事件中的字段值在外部问题管理系统中创建新条目。
- 启动将使用选定事件的一个或多个字段值的辅助搜索。
- 针对在某个事件中找到的特定字段值执行外部搜索（使用 Google 或类似的 Web 搜索应用）。

此外，还可以定义以下工作流动作：

- 以包含一个特定字段或一组字段的事件，或属于特定事件类型的事件为目标。
- 显示在搜索结果的字段菜单或事件菜单中。也可以将其设置为仅显示在特定字段的菜单中，或显示在符合条件的事件中的所有字段菜单中。
- 选中后，在当前窗口或新窗口中打开。

使用 Splunk Web 定义工作流动作

您可以使用 Splunk Web 设置本章顶部项目符号列表中的所有工作流动作，以及许多其他工作流动作。要开始设置，请导航到**设置 > 字段 > 工作流动作**。在“工作流动作”页面中您可以通过单击其名称查看和更新现有工作流动作。也可以通过单击**新增**来创建新的工作流动作。这两种方法都可以将您带到工作流动作详细信息页面，您可以在其中定义各个工作流动作。

如果正在创建新的工作流动作，则需要为其指定一个**名称**并确定其**目标应用**。

您可以设置以下三种类型的工作流动作：

- **GET 工作流动作**，用于创建典型的 HTML 链接以执行一些操作，如针对特定值执行 Google 搜索或针对外部 WHOIS 数据库运行域名查询。
- **POST 工作流动作**，用于生成对特定 URI 的 HTTP POST 请求。此动作类型可用来执行一些操作，如使用一组相关字段值在外部问题管理系统中创建条目。
- **搜索工作流动作**，用于启动将使用某个事件的特定字段值的辅助搜索，如在索引中查找特定时间范围内 `ipaddress` 和 `http_status` 字段值特定组合的搜索。

使工作流动作以小范围内的一组事件为目标

在 Splunk Web 中创建工作流动作时，可以选择使工作流动作以小范围内的一组事件为目标。可以通过字段、事件类型或两者的组合来限制工作流动作范围。

通过字段缩小工作流动作范围

可以设置仅应用于具有一个指定字段或一组字段的事件的工作流动作。例如，如果您有一个 `http_status` 字段，并且您希望某个工作流动作仅应用于包含该字段的事件，则可在**只应用到以下字段**设置中声明 `http_status`。

如果您希望工作流动作仅应用于具有一组字段的事件，则可以在**只应用到以下字段**中声明以逗号分隔的字段列表。列出了多个字段时，只有事件中存在完整的字段列表时才会显示工作流动作。

例如，假设您希望工作流动作仅应用于具有 `ip_client` 和 `ip_server` 字段的事件。要达到此目的，应在**只应用到以下字段**中输入 `ip_client, ip_server`。

也可以使用星号通配符来限定工作流动作的字段范围。例如，如果您声明了一个简单的字段列表 `ip *`，Splunk 软件会对具有 `ip_client` 或 `ip_server` 以及二者组合的事件（以及任何具有与 `ip_*` 相匹配的字段的其他事件）应用所得到工作流动作。

默认情况下字段列表设置为 `*`，表示它与所有字段相匹配。

如果您需要更复杂的选择逻辑，建议您通过事件类型来限定范围，而不要通过字段限定范围，或将二者结合起来使用。

通过事件类型缩小工作流动作范围

通过事件类型限定范围的工作方式与通过字段限定范围的完全一样。您可以在**只应用到以下事件类型**设置中输入一个事件类型或多个以逗号分隔的事件类型，以创建一个工作流动作，仅将该工作流动作应用于属于相应的一个或一组事件类型的事件。也可以使用通配符匹配来找出属于某个事件类型范围的事件。

还可以通过字段和事件类型组合来缩小工作流动作的范围。例如，您可能有一个字段 `http_status`，但仅希望当 `http_status` 大于或等于 500 时，得到的工作流动作才显示在包含该字段的事件中。要达到此目的，您首先应设置一个事件类型 `errors_in_500_range`，该事件类型将应用于与以下搜索相匹配的事件

```
http_status >= 500
```

接下来您应定义一个工作流动作，其**只应用到以下字段**设置为 `http_status`，**只应用到以下事件类型**设置为 `errors_in_500_range`。

有关事件类型的详细信息，请参见本手册中的[“关于事件类型”](#)。

控制工作流动作在字段和事件菜单中的显示

正确设置了工作流动作之后，它们会显示在与搜索结果中的字段和事件相关联的菜单中。您可以将工作流动作安排成事件级（意味着它们应用于整个事件）、字段级（意味着它们应用于事件中的特定字段），或这两者。

要选择事件级工作流动作：

- 运行搜索。
- 转到**事件**选项卡。
- 在搜索结果中展开事件并单击**事件动作**。

下面是“显示来源”（事件级工作流动作，单击时会显示原始搜索数据中事件的来源）的示例。

The screenshot shows the Splunk search results interface. A search has been performed, and the results are displayed in a table. The '事件' (Event) column is expanded, showing a list of events. The '事件操作' (Event Actions) menu is open, and the '显示来源' (Show Source) action is highlighted with a red circle. The '显示来源' action is described as '显示原始搜索数据中事件的来源' (Show the source of the event in the original search data).

也可以让工作流动作显示在事件中的字段的**操作**菜单中。下面是工作流动作针对所选字段和值在单独窗口中打开 Google 搜索的示例。

The screenshot shows the Splunk search results interface. A search has been performed, and the results are displayed in a table. The '事件' (Event) column is expanded, showing a list of events. The '事件操作' (Event Actions) menu is open, and the 'Google Search' action is highlighted with a red circle. The 'Google Search' action is described as '谷歌这个字段和值' (Google this field and value).

这两种示例都属于使用 "GET" 链接方法的工作流动作。这是您可以在 Splunk Enterprise 中实现的三种类型的工作流动作之一（其他两个是 "POST" 链接工作流动作和搜索工作流动作）。请继续阅读，了解有关设置全部三种类型的说明。

您还可以定义显示在事件级和字段级两者中的工作流动作。例如，您可以为工作流动作进行定义，这些工作流动作可以使用事件（如 `User_ID`）中的特定字段值进行某些操作。

设置 GET 工作流动作

GET 链接工作流动作会将一个或多个值输入到 HTML 链接中。单击该链接即会在浏览器中执行 HTTP GET 请求，

您可用来自外部 Web 资源传递信息，如搜索引擎或 IP 查找服务。

注意：在传输期间，在 GET 动作中通过 URI 传递的变量会编码成 URL 格式。这表示您可以包含在单词或标点符号之间有空格的值。但是，如果您使用的是以 HTTP 地址作为值的字段，且您希望将整个字段值作为一个 URI 来传递，则应该使用 `$!` 前缀以防止 Splunk 软件对该字段值进行转义。有关更多信息，请参阅下面的[“使用 `\$!` 前缀防止转义 URL 或 HTTP 形式的字段值”](#)。

要定义 GET 工作流动作：

1. 导航到设置 > 字段 > 工作流动作。

2. 单击新建以打开一个新的工作流动作表单。

3. 为动作定义一个标签。

标签字段可用于定义将显示在字段或事件工作流菜单中的文本。标签可以是静态的，或者包含相关字段的值。

4. 确定是否将工作流动作应用于您数据中特定的字段或事件类型。

使用**仅应用于下列字段**以确定一个或多个字段。当您确定字段时，工作流动作只会出现在拥有这些字段的事件中，在事件的事件菜单或字段菜单中出现。如果您将其留空或输入一个星号，则动作会显示在所有字段的菜单中。

使用**仅应用于下列事件类型**以确定一个或多个事件类型。如果您确定了一个事件类型，则工作流动作只会出现在属于该事件类型的事件的事件菜单中。

5. 对于显示动作于，确定您是否想要动作出现在**事件菜单、字段菜单或两者中**。

6. 将动作类型设置为 **link**。

7. 在 URI 中，为字段值要发送到的外部资源的位置提供 URI。

与**标签**设置类似，在声明字段值时，应使用以美元符号括起来的字段名称。

在 GET 动作中通过 URI 传递的变量在传输期间会自动编码成 URL 格式。这表示您可以包含在单词或标点符号之间有空格的值。

8. 在打开链接于的下方，确定是否在当前窗口中显示工作流动作，或者在新窗口中打开链接。

9. 设置链接方法为 **get**。

10. 单击保存以保存您的工作流动作定义。

示例 - 对于字段值进行 Google 搜索

下面是一个 GET 链接工作流动作的设置示例，该示例将启动一个 Google 搜索，在搜索结果中查找 `topic` 字段的值：

Google this topic
字段 » 工作流动作 » Google this topic

标签 *

Google \$topic\$

输入为此动作显示的标签。或者，通过用美元符号括起字段名称来加入字段值。例如，"搜索票证编号:\$ticketnum\$".

只应用到以下字段

*

指定以逗号分隔的字段列表。事件中必须出现这些字段才能应用工作流动作。指定字段时，工作流动作只显示在这些字段的字段菜单中；否则会显示在所有字段菜单中。

只应用到以下事件类型

指定以逗号分隔的事件类型列表。事件必须与其中的事件类型相关联才能应用工作流动作。

显示动作于

两者

动作类型 *

链接

链接配置

URI *

http://www.google.com/search?q=\$topic\$

输入要链接的位置。或者，通过用美元符号括起字段名称来指定字段。例如，http://www.google.com/search?q=\$host\$.

打开链接于

新窗口

链接方法

get

取消

保存

在该示例中，我们将**标签**值设置为 `Google $topic$`，因为在我们的事件中有一个字段称为 `topic`，而且我们希望在这个工作流动作的标签中包含 `topic` 的值。例如，如果事件中 `topic` 的值为 `CreatefieldactionsinSplunkWeb`，则字段动作在 `topic` 字段菜单中将显示为 `Google CreatefieldactionsinSplunkWeb`。

`Google $topic$` 动作适用于所有事件。

`Google $topic$` 动作 **URI** 使用 `GET` 方法将 `topic` 值提交到 `Google` 进行搜索。

示例 - 提供外部 IP 查找

您已将 `Splunk` 应用配置为从 `Web` 服务日志中提取域名并将它们指定为名为 `domain` 的字段。您希望能够搜索外部 `WHOIS` 数据库，以获得有关所出现的域的更多信息。

下面介绍了如何设置 `GET` 工作流动作来帮助您实现这个目标。

在工作流动作详细信息页面中，将**动作类型**设置为 `link`，将**链接方法**设置为 `get`。

接下来将使用**标签**和 **URI** 字段来标识所涉及的字段。将**标签**值设为 `WHOIS : $domain$`。将 **URI** 值设为 `http://whois.net/whois/$domain$`。

之后，您可以确定以下项：

- 链接是显示在字段菜单中，还是事件菜单中，或者同时显示在两个菜单中。
- 链接是在相同窗口中还是在窗口中打开 `WHOIS` 搜索。
- 为事件显示工作流动作链接的限制。可以使工作流动作以具有特定字段、属于特定事件类型或使用两者某种组合的事件为目标。

设置 POST 工作流动作

您可以使用类似于 `GET` 链接动作的设置方式来设置 `POST` 工作流动作。但是，`POST` 请求一般由 `HTML` 中的一个格式元素以及一些将转换成 `POST` 参数的输入来定义。这表示您需要确定要发送至所标识的 `URI` 的 `POST` 参数。

注意：在传输期间，在 `GET` 动作中通过 `URI` 传递的变量会编码成 `URL` 格式。这意味着您可以包含在单词或标点符号之间有空格的值。但是，如果您使用的是以 `HTTP` 地址作为值的字段，且您希望将整个字段值作为一个 `URI` 来传递，则应该使用 `!` 前缀以防止 `Splunk` 软件对该字段值进行转义。有关更多信息，请参阅下面的[“使用 `!` 前缀防止转义 URL 或 HTTP 形式的字段值”](#)。

1. 导航到 **设置 > 字段 > 工作流动作**。

2. 单击 **新建** 以打开一个新的工作流动作表单。

3. 为动作定义一个 **标签**。

标签 字段可用于定义将显示在字段或事件工作流菜单中的文本。标签可以是静态的，或者包含相关字段的值。

4. 确定是否将工作流动作应用于您数据中特定的字段或事件类型。

使用 **仅应用于下列字段** 以确定一个或多个字段。当您确定字段时，工作流动作只会出现在拥有这些字段的事件中，在事件的事件菜单或字段菜单中出现。如果您将其留空或输入一个星号，则动作会显示在所有字段的菜单中。

使用 **仅应用于下列事件类型** 以确定一个或多个事件类型。如果您确定了一个事件类型，则工作流动作只会出现在属于该事件类型的事件的事件菜单中。

5. 对于 **显示动作于**，确定您是否想要动作出现在 **事件菜单、字段菜单或两者中**。

6. 将 **动作类型** 设置为 **Link**。

7. 在 **URI** 下方，为响应 POST 请求的 Web 资源提供 URI。

8. 在 **打开链接于** 的下方，确定是否在当前窗口中显示工作流动作，或者在新窗口中打开链接。

9. 设置 **链接方法** 为 **Post**。

10. 在 **Post 参数** 下方，定义应该发送到在确定的 URI 上的 Web 资源的参数。

这些参数是关键字和值的组合。在参数的关键字和值端，您可以使用以美元符号括起来的字段名称来标识应发送到资源的来自事件的字段值。您可以在一个 POST 工作流动作中定义多个关键字/值参数。

在第一个字段中输入关键字，并在第二个字段中输入值。单击 **添加另一个字段**，以创建额外的 POST 参数。

11. 单击 **保存** 以保存您的工作流动作定义。

Splunk 软件将 POST 链接动作中通过 URI 传递的变量自动编码成 HTTP 格式。这表示您可以包含在单词或标点符号之间有空格的值。

示例 - 允许 http 错误在问题跟踪应用中创建一个条目

您已将 Splunk 应用配置为从 Web 服务日志作为 `http_status` 字段提取 HTTP 状态代码。除 `http_status` 字段以外，事件一般还包含标准单行描述请求，或来源于生成了错误的 Python 进程的多行 Python Stacktrace。

您希望设计一个工作流动作，仅当错误事件的 `http_status` 在 500 范围内时才显示出来。您希望此工作流动作将相关的 Python Stacktrace 和 HTTP 状态代码发送至外部问题管理系统，以生成新的故障报告。但是，问题管理系统仅接受对特定端点的 POST 请求。

下图描述了如何设置满足您需求的 POST 工作流动作。

新增

字段 » 工作流动作 » 新增

目标应用 *

search

名称 *

submit_error_report

输入不含空格或特殊字符的唯一名称。此名称用于以后在 Splunk 设置中标识您的工作流动作。

标签 *

submit error report

输入为此动作显示的标签。或者，通过用美元符号括起字段名称来加入字段值，例如，搜索票据编号:Sticketnum\$。

只应用到以下字段

指定以逗号分隔的字段列表。事件中出现这些字段才能应用工作流动作。指定字段时，工作流动作只显示在这些字段的字段菜单中；否则会显示在所有字段菜单中。

只应用到以下事件类型

errors_in_500_range

指定以逗号分隔的事件类型列表。事件必须与其中的事件类型相关才能应用工作流动作。

显示动作于

两者

动作类型 *

链接

链接配置

URI *

http://jira.pwnyinc.com/bugs/new

输入要链接的位置。或者，通过用美元符号括起字段名称来指定字段，例如，http://www.google.com/search?q=\$host\$。

打开链接于

新窗口

链接方法

post

Post 参数

title

=

server error \$http_status\$

删除

description

=

\$_raw\$

删除

添加另一个字段

取消

保存

请注意，第一个 POST 参数将 `server error $http_status$` 发送至外部问题跟踪系统中的 `title` 字段。如果对 `http_status` 为 500 的事件您选择了此工作流动作，则它会打开问题跟踪系统中标题为 `server error 500` 的问题。

第二个 POST 参数使用 `_raw` 字段来包含新问题 `description` 字段中的多行 Python Stacktrace。

最后，请注意工作流动作已设置为仅应用于属于 `errors_in_500_range` 事件类型的事件。此事件类型仅应用于 `http_error` 值在典型 HTTP 错误范围 500 或更高范围内的事件。HTTP 错误代码低于 500 的事件不会在其事件或字段菜单中显示 **提交错误报告** 工作流动作。

设置使用事件的字段值进行动态填充的辅助搜索

要设置工作流动作以后启动动态填充的辅助搜索，请先在工作流动作详细信息页面上将 **动作类型** 设置为 **搜索**。此时会展现一组 **搜索配置** 字段，您可以使用这些字段来定义辅助搜索的细节。

在 **搜索字符串** 中输入一个搜索字符串，其中包含以美元符号括起的一个或多个字段值占位符。例如，如果要设置一个工作流动作，以对事件中出现的客户端 IP 进行搜索，您可以仅在该字段中输入 `clientip=$clientip$`。

标识在其中运行搜索的应用。如果希望搜索在当前视图之外的视图中运行，可选中该视图。与所有工作流动作一样，您可以决定是在当前窗口中还是在新的窗口中打开它。

务必为搜索设置一个时间范围（或者表明它是否应与创建了字段列表的搜索使用相同的时间范围），方法是在 **最早时间** 和 **最晚时间** 字段中输入相对时间修饰符。如果将这些字段留空，则默认情况下将对所有时间运行搜索。

最后，与其他工作流动作类型一样，您可以限制搜索工作流动作，使其只应用于包含特定字段集和/或属于特定事件类型的事件。

示例 - 启动辅助搜索以查找源自特定 Ruby On Rails 控制器的错误

假设您的公司使用的是建构于 Ruby on Rails 之上的 Web 基础设施。您已设置了一个事件类型，用于辨别与 Ruby 控制器相关的错误（标题为 `controller_error`），但有时您只想查看与特定控制器相关的所有错误。下面介绍了如何设置一个工作流动作来达到此目的：

1. 在工作流动作详细信息页面上，设置一个具有下列**标签**的动作：See other errors for controller `$controller$` over past 24h

2. 将**动作类型**设置为 *Search*。

3. 输入下列**搜索字符串**：`sourcetype=rails controller=$controller$ error=*`

4. 将**最早时间**设置为 *-24h*。将**最晚时间**留空。

5. 使用**仅应用于下列...**设置，安排工作流动作仅显示在属于 `controller_error` 事件类型以及包含 `error` 和 `controller` 字段的事件中。

新增

字段 » 工作流动作 » 新增

目标应用 *

search

名称 *

shox

输入不含空格或特殊字符的唯一名称。此名称用于以后在 Splunk 设置中标识您的工作流动作。

标签 *

see other errors for controller `$controller$` over past 24h

输入为此动作显示的标签。或者，通过用美元符号括起字段名称来加入字段值。例如，搜索票据编号 `$ticketnum$`。

只应用到以下字段

error, controller

指定以逗号分隔的字段列表。事件必须出现这些字段才能应用工作流动作。指定字段时，工作流动作只显示在这些字段的字段菜单中；否则会显示在所有字段菜单中。

只应用到以下事件类型

errors_in_500_range

指定以逗号分隔的事件类型列表。事件必须与其中的事件类型相关联才能应用工作流动作。

显示动作于

事件菜单

动作类型 *

搜索

搜索配置

搜索字符串 *

sourcetype=rails controller=`$controller$` error=*

输入对此动作的搜索。或者，将字段指定为 `$fieldname$`。例如，`sourcetype=rails controller=$controller$ error=*`。

在应用中运行

选择一个在其中运行搜索的应用。默认为当前应用。

在视图中打开

输入在其中打开搜索的视图的名称。默认为当前视图。

运行搜索于

新窗口

时间范围

最早时间

-24h

最晚时间

☐ 使用与创建了字段列表的搜索相同的时间范围

这些是基础知识。您也可以决定工作流动作将在哪个应用或视图中运行（例如，可以为此信息指定一个名为 `ruby_errors` 的视图），并明确动作是在当前视图中执行还是打开一个新视图。

在工作流动作中使用特殊参数

工作流动作可使用以 "@" 符号开头的特殊参数。其中有两个特殊参数仅用于字段菜单。可使用它们来设置工作流动作，以应用于它们所应用于的事件中的所有字段。

- **@field_name** - 表示所单击的字段的名称。
- **@field_value** - 表示所单击的字段的价值。

其他特殊参数有：

- **@sid** - 表示返回了事件的任务的 SID
- **@offset** - 表示任务中事件的偏移量
- **@namespace** - 表示从中派遣任务的命名空间
- **@latest_time** - 表示事件发生的最晚时间，可用于区分相似的事件，并不始终可用于所有字段。

示例 - 创建应用于事件中所有字段的工作流动作

可以更新上面讨论的 Google 搜索示例（在 GET 链接工作流动作部分），使其能够搜索它所应用于的事件中每个字段的字段名称和字段值。您只需将标题更改为 `Google this field and value`，并将该动作的 URI 替换为

```
http://www.google.com/search?q=${@field_name}${@field_value}
```

这会导致工作流动作对您要在字段菜单中查看的每一个字段/值组合执行搜索。如果您要在字段菜单中查找 `sourcetype=access_combined`，并且选择了**对此字段和值执行 Google 搜索**字段动作，则得到的 Google 搜索为 `sourcetype accesscombined`。

请牢记：使用 **@field_name** 和/或 **@field_value** 参数的工作流动作与事件级菜单不兼容。

示例 - 显示事件的来源

此工作流动作使用了其他特殊参数来显示原始搜索数据中事件的来源。

动作类型为 link，其**链接方法**为 `get`。其**标题**为 `Show source`。**URI** 为 `/app/${@namespace}/show_source?sid=${@sid}&offset=${@offset}&latest_time=${@latest_time}`。它仅应用于包含 `_cd` 字段的事件。

尝试在您的应用中设置此工作流动作（如果尚未安装），并查看其工作方式。

使用 \$! 前缀防止转义 URL 或 HTTP 形式的字段值

当为工作流动作定义字段时，可以将这些字段进行转义，以便能安全地将它们通过 HTTP 传递给外部端点。但是，有些情况不需要这种转义。在这些情况中，使用 `$!` 前缀来防止对字段值进行转义。此前缀防止对 GET 工作流动作进行的 URL 转义以及对 POST 工作流动作进行 HTTP 形式的转义。

示例 — 将 HTTP 地址传递给单独的浏览器窗口

您有一个 GET 工作流动作，且该动作包含一个名为 `http` 的字段。`http` 字段的值是完整格式的 HTTP 地址。此工作流动作打开一个新的浏览器窗口，指向 `http` 字段的 HTTP 地址值。如果以转义后的 HTTP 地址打开此新窗口，则此工作流动作无法正常运行。

为防止 HTTP 地址被转义，使用 `$!` 前缀。在“设置”中，您可能通常会一样将此工作流动作的 **URI** 设为 `$http$`，但现在应将其设为 `$!http$`。

数据标准化：标记和别名

关于标记和别名

在您的数据中，可能具有包含相关字段值的事件组。为了帮助您更有效地搜索这些特定的事件数据组，可以在数据中分配标记和别名。

如果要标记成千上万个项目，最好采用字段查找。使用大量标记不会影响索引建立，但如果用查找使事件具有更好的分类，则搜索会更具意义。关于字段查找的更多信息，请参阅[使用字段查找将信息添加到事件中](#)。

标记

您可使用**标记**为特定字段和值组合（包括事件类型、主机、来源或来源类型）分配名称。

您可使用标记来协助跟踪抽象字段值，如 IP 地址或 ID 编号。例如，您可能有一个与您的主办公室相关的 IP 地址，其值为 `192.168.1.2`。将该 `IPaddress` 值标记为 `mainoffice`，然后搜索该标记即可找到具有该 IP 地址的事件。

也可以使用一个标记将一组字段值组合在一起，这样就可以用一个简单的命令对它们进行搜索。例如，您可能发现您有两个主机名与同一台计算机相关。此时可以给这两个值分配相同的标记。当搜索此标记时，会返回两个主机名值都包含的事件。

您也可以给提取的特定字段分配多个标记以反映其标识的不同方面，这样您便可以基于标记执行搜索，以帮助快速缩小所需结果的范围。

标记示例

您有一个提取的字段，名为 `IPAddress`，该字段表示您公司内网中数据来源的 IP 地址。您可以基于每个 IP 地址的功能或位置为其设置标记，使 `IPAddress` 变得有用。您可以将您的所有路由器的 IP 地址标记为 `router`。也可以基于其位置为每个 IP 地址设置标记，例如：`SF` 或 `Building1`。Building 1 中位于 San Francisco 的一个路由器 IP 地址可能具有 `router`、`SF` 和 `Building1` 标记。

要搜索 San Francisco 中不在 `Building1` 内的所有路由器，可以执行以下搜索：

```
tag=router tag=Sf NOT (tag=Building1)
```

字段别名

您可使用**字段别名**将多个来源的数据规范化。可为一个字段名添加多个别名，或使用这些字段别名来规范化不同的字段名称。这不会重命名或删除原始字段名称。在为字段设置了别名之后，可使用其任一名称别名来进行搜索。您可以在 Splunk Web 或 `props.conf` 中定义字段别名。请参阅[在 Splunk Web 中创建字段别名](#)。

可使用别名将提取的不同字段名分配给同一个字段名。

注意：所有搜索都会用到所有来源类型的字段别名。如果提取的字段名不太多，那没有问题，但随着时间的推移可能会产生许多开销。

字段别名示例

一个数据模型可能会有一个名为 `http_referrer` 的字段。在源数据中，此字段可能错误的拼写为 `http_referer`。使用字段别名来捕获原始来源数据中命名不同的字段，并将其映射为所需的字段名。

在“搜索”中为字段值对设置标记

您可能会有包含相关字段/值对的事件组。为帮助您更有效地搜索这些事件，可为相关字段/值对分配同一个标记。

请参阅[关于标记和别名](#)。

为字段值对设置标记

可以直接从搜索结果中为任意字段/值对设置标记。

1. 查找具有您想要标记的字段/值对的事件。
2. 通过单击 *i* 列中的箭头来打开事件，以查看从事件中提取的字段的完整列表。
3. 选择您想要为其创建标记的字段值对的**动作**箭头，并选择**编辑动作**。

将打开“创建动作”对话框。



4. 在“创建动作”对话框中，定义字段值对的一个或多个**标记**。

标记字段的值决不能放在双引号之内。



5. 单击**保存**以保存标记。

从标记定义中删除 URL 编码的值

当您标记字段/值时，值中的“值”部分不能为 URL 编码。如果您的标记有任何 `%##` 形式的 URL 编码，则将其解码，然后使用解码的 URL 保存标记。

例如，假定您想要给该字段值使用标记 "Useful"：

```
url=http%3A%2F%2Fdocs.splunk.com%2FDocumentation
```

要进行管理，则必须在“设置”中定义该标记。

1. 选择**设置 > 标记 > 按标记名称排列的列表**，并单击 **Useful** 标记名以打开该标记的详细信息页面。
2. 在**字段值对**下方，将 `url=http%3A%2F%2Fdocs.splunk.com%2FDocumentation` 替换为解码的版本：`url=http://docs.splunk.com/Documentation`
3. 单击**保存**以保存更改。

请参阅[在“设置”中定义和管理标记](#)。

搜索加标记字段值

可使用两种方法来搜索标记。如果要搜索与任何字段中的值相关联的标记，可使用以下语法：

```
tag=<tagname>
```

或者，如果要搜索与特定字段中的值相关联的标记，可使用以下语法：

```
tag::<field>=<tagname>
```

使用通配符搜索标记

搜索关键字和字段值（包括事件类型和标记）时，可使用星号 (*) 通配符。

例如，如果具有不同类型 IP 地址的多个事件类型标记（如 `IP-src` 和 `IP-dst`），可使用以下语法来搜索所有这些事件：

```
tag::eventtype=IP-*
```

如果想要找到标记包含 "local" 的所有主机，可搜索标记：

```
tag::host=*local*
```

如果想要搜索事件类型无标记的事件，可搜索布尔表达式：

```
NOT tag::eventtype=*
```

禁用和删除标记

您可以通过“搜索”应用删除特定字段值的标记关联。您可以在“设置”中禁用或删除标记，即使它们与多个字段值相关。

请参阅[在“设置”中定义和管理标记](#)。

删除搜索结果中特定字段值的标记关联

如果您不再需要与搜索结果中某个特定字段值关联的标记，请执行以下操作。

1. 单击该事件旁的箭头。
2. 在**操作**下方，单击该字段值旁的箭头。
3. 选择“编辑标记”以打开**创建标记**弹出窗口。
4. 在**创建标记**弹出窗口中，清除要**从标记**字段禁用的一个标记或多个标记。
5. 单击**保存**以保存更改。

这会从系统中删除此特定的标记与字段值的关联。如果只有这一个字段值与此标记相关联，则也会从系统中删除此标记。

在“设置”中标记事件类型

在创建或编辑事件类型中可以添加标记。请参阅[为事件类型设置标记](#)。

重命名来源类型

在 `props.conf` 中配置来源类型时，可以重命名来源类型。多个来源类型可以共享相同的名称；如果为了便于搜索您要将一组来源类型组合在一起，则这样十分有用。例如，您可以规范化包含 `-too_small` 的来源类型名称，以删除分类符。请参阅“重命名来源类型”。

在“设置”中定义和管理标记

Splunk 软件为创建和管理标记提供了多种方法。大多数用户会使用最简单的方法：直接在搜索结果中为字段/值对设置标记。在[“搜索”中为字段值设置标记和别名](#)中对此方法进行了详细描述。

但是，作为一个知识管理员，您可能会使用“设置”中的“标记”页面来管理由 Splunk 部署用户所创建的标记。本主题介绍如何使用“设置”中的“标记”页面执行以下操作：

- 管理 Splunk 部署中的标记。
- 创建新标记。
- 禁用或删除标记。

要导航到“标记”页面，选择 **设置 > 标记**。

使用设置中的标记页面

“设置”中的“标记”页面提供三种标记视图。每个视图的标记分类均不同：

- 按字段值对排列
- 按标记名称排列
- 所有唯一的标记对象，按其唯一 ID 对标记进行排列。

使用这些页面，可以通过不同方式管理您的标记集合，并快速访问随着时间的推移标记和字段/值对之间所建立的关联。还可以创建和删除标记之间的关联。选择满足您的标记管理需求的页面。

管理与特定字段值对关联的标记集

如果您要查看系统中与标记相关联的所有字段/值对列表，该如何操作？此外，如果要查看甚至更新与特定字段/值对关联的标记集，该如何操作？或者如果要为特定字段/值对定义一组标记，该怎么做？

按字段值对排列的列表 页面使您能够查看和编辑与特定字段/值对关联的标记集。

通过此页面，还可以使用标记管理特定字段/值组合的权限。

要查看特定字段/值对的标记列表，找到该对，并单击相应字段/值对。这将带您前往该对的关联详细信息页面。

以下是为 `eventtype=auditd_create` 字段/值对定义的一组标记示例：

按字段值对列出

标记 » 按字段值对列出

应用上下文

Search & Reporting (search)

所有者

任何

☐ 仅显示在此应用上下文中创建的对象 [了解更多信息](#)

新建

显示 1 个项目中的 1-1

每页每页 25

字段值对	标记名	App	共享	状态	操作
eventtype=auditd_create	auditd, create, file, resource, success	search	专用 权限	已启用 禁用 所有或对标记	复制 移动 删除

您可以从此详细视图中添加和删除标记（如果您有相应权限）。

单击**按字段值对排列的列表**页面上的**新建**，可为新字段/值对定义一组标记。

在为字段/值对创建或更新标记列表时，要记住您可能正在创建新的标记，或正在将现有标记与不同于最初设计用于的字段/值对类型相关联。作为一个知识管理器，您应该始终小心谨慎地设计和维护标记集。此做法有助于实现数据标准化，并可减少用户方的混淆。

请参阅[通过“设置”页面管理知识对象](#)。

注意：您可能需要确认添加到“按字段/值对排列的标记”页面中的字段/值对确实存在。系统不会阻止您为一个不存在的字段/值对定义标记列表。

查看和更新与特定标记关联的字段值对集

如果要查看系统中有一个或多个标记与其关联的所有标记的列表，该如何操作？而且，如果要查看甚至更新与某个特

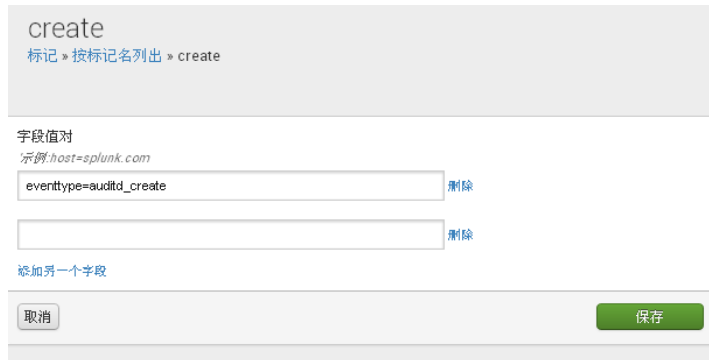
定标记关联的字段/值对集，该如何操作？或者如果要为新标记定义一组字段/值对，该如何操作？

Splunk Web 中的[按标记名称排列的列表](#)页面回答了这些问题。可查看和编辑与特定标记关联的字段/值对集。

但是，在此页面中，不能管理与标记关联的字段/值对集的权限。

可以查看特定标记的字段/值对列表。在[按标记名称排列的列表](#)页面中找到该标记，然后单击**标记名列**中的标记名。此任务会将您带到该标记的详细信息页面。

下面的示例显示了 `modify` 标记与之关联的各个字段/值对。



create
标记 » 按标记名列出 » create

字段值对
示例: host=splunk.com

eventtype=auditd_create [删除](#)

[删除](#)

[添加另一个字段](#)

[取消](#) [保存](#)

可以添加和删除字段/值关联（如果您有相应权限）。

要为新标记定义字段/值集对，单击“按标记名称排列的列表”页面上的**新建**。

在为标记创建或更新字段/值对集时，请注意您可能正在创建新的字段/值对。您可能需要确认与标记关联的字段/值对确实存在。系统不会阻止您添加不存在的字段/值关联。

需小心创建新标记。具有您正尝试构建的用途的标记可能已经存在。作为一个知识管理器，您应该始终小心谨慎地设计和维护标记集。此做法有助于实现数据标准化，并可减少用户方的混淆。请参阅[通过“设置”页面管理知识对象](#)。

查看所有唯一字段/值对和标记组合

所有唯一的标记对象页面分隔系统中的所有唯一标记名称和字段/值对。不像前两页面，此页面仅允许您编辑标记和字段/值对之间的一对一关系。

您可以搜索特定标记以快速查看与其关联的所有字段/值对，反之亦然。如果您要禁用或复制特定标记与字段/值的关联时，或者要以该粒度级别维护权限，则使用本页面。

禁用和删除标记

如果有不再需要使用，或不再需要与特定字段/值对相关联的标记，则可以选择禁用或删除它。如果您有权限执行此操作，可以：

- 删除搜索结果中特定字段/值对的标记关联。
- 通过“按标记名称排列的列表”页面批量禁用或删除标记，即使它与多个字段值关联。
- 通过“按字段值对排列的列表”页面批量禁用或删除某个字段/值对与一组标记之间的关联。

有关删除搜索结果中与特定字段/值对的标记关联的信息，请参见在[“搜索”中为字段值对设置标记](#)。

删除与多个字段/值对关联的标记

可以使用 Splunk Web 从系统中完全删除一个标记，即使它与许多字段/值对相关联。此方法使您通过一个步骤即可摆脱所有这些关联。

导航到**设置 > 标记 > 按标记名称排列的列表**。删除标记。如果未看到标记的删除链接，则表示您没有权限执行删除操作。删除标记时，要注意删除操作会对其产生影响的下游相关性。请参阅[通过“设置”页面管理知识对象](#)。

注意：您也可以进入特定标记的编辑视图，直接删除字段/值对关联。

禁用或删除某个字段/值对与一组标记之间的关联

可使用此方法来批量删除与某个字段/值对关联的标记集。此方法使您通过一个步骤即可摆脱这些关联。但它不会从您的数据中删除字段/值对。

导航到**设置 > 标记 > 按字段值对排列的列表**。删除字段/值对。如果未看到字段/值对的删除链接，则表示您没有权限执行删除操作。删除这些关联时，要注意可能会因删除操作而受到不良影响的下游相关性。请参阅[通过“设置”页面管理知识对象](#)。

注意：您也可以直接在特定字段/值对的编辑视图中删除标记关联。

禁用标记

取决于您的相关权限，您也可以使用“设置”中的三个“标记”页面禁用标记与字段/值的关联。禁用了标记与字段/值对之间的关联后，该关联仍存在于系统中，只是处于非活动状态，直到再次启用。

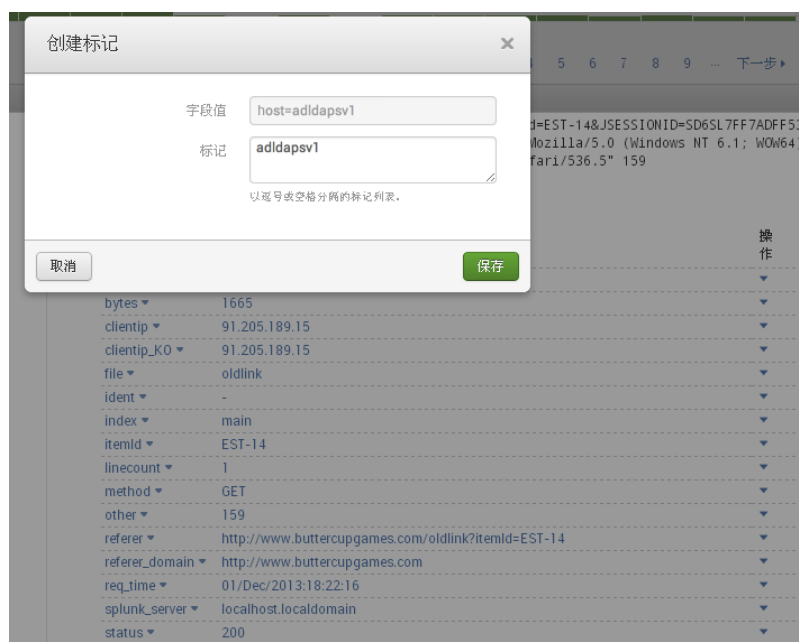
为主机字段设置标记

为主机字段设置标记对于知识捕获和共享，以及精心创建更精确的搜索十分有用。您可以使用一个或多个单词为主机字段设置标记。这允许您按功能或类型分组主机，这样用户便能更轻松地搜索相似服务器组上的所有活动。如果您已更改了给定输入的主机字段值，您还可以使用新主机名称为索引中已存在的事件设置标记，以便于跨数据集执行搜索。

在搜索结果中为主机字段添加标记

在搜索结果中为主机字段/值组合添加标记：

1. 针对主机中您要为其设置标记的数据执行搜索。
2. 在搜索结果中，单击包含要为其设置标记的字段的事件所关联的箭头。在展开的列表中，单击与该字段关联的操作下方的箭头，然后选择**编辑标记**。



3. 在“创建标记”对话框中，输入要为其设置标记的主机字段值，例如，在“字段值”中输入 **Tag host=<current host value>**。输入以逗号或空格分隔的标记，然后单击**保存**。

“主机名称”对比“为主机字段设置标记”

主机字段的值是在索引事件时设置的。默认情况下可以基于 Splunk 服务器主机名设置值，为给定输入设置值，或从每个事件数据中提取值。使用替代主机名为主机字段设置标记不会更改主机字段的实际值，但可以搜索您指定的标记，而不必使用主机字段值。每个事件只能有一个主机名，但可以有多个主机标记。

例如，如果 Splunk 服务器要从某个特定主机接收合规性数据，使用**合规**为主机设置标记将有利于合规性搜索。使用主机标记，您可以创建一个宽松的数据组，而不必标记或更改基本主机名。

如果您检索到一些来自特定输入来源的数据，然后决定更改该输入的主机字段的值--所有来自该输入的新数据都将具有新的主机字段值，而索引中已存在的数据仍具有原有值，此时您可能还需要用其他主机名为主机字段设置标记。通过为现有数据的主机字段设置标记，您可以搜索新的主机值，而不必排除所有现有数据。

为事件类型设置标记

可通过为事件类型设置标记的方式向数据中添加信息。一个事件类型可以有多个标记。例如，可将所有防火墙事件类型标记为**防火墙**，将防火墙事件类型的一个子集标记为**拒绝**，将另一个子集标记为**允许**。为事件类型设置了标记后，与标记模式相匹配的任何事件类型也会带有标记。

注意：在 [Splunk Web 中创建事件类型](#)或在 eventtypes.conf 中配置事件类型时，可为其设置标记。

使用 Splunk Web 为事件类型添加标记

可使用 Splunk Web 来查看和编辑事件类型列表。

- 导航到**设置 > 事件类型**。
- 找到要为其设置标记的事件类型，单击其名称转到其详细信息页面。
 - **注意：**记住，事件类型常与特定的 Splunk 应用相关联。它们还具有基于角色的权限，可防止您对其进行查看和/或编辑。
- 在事件类型的详细信息页面上，在**标记**字段中添加或编辑标记。
- 单击**保存**确认更改。

为事件类型设置了标记后，便可以使用语法 `tag::<field>=<tagname>` 或 `tag=<tagname>` 在搜索栏中对其进行搜索：

```
tag=foo
```

```
tag::host=*local*
```

在 Splunk Web 中创建字段别名

在您的数据中，可能具有包含相关字段值的事件组。为帮助您更有效地搜索这些字段组，可为其字段值分配字段别名。可为任何提取的字段（包括事件类型、主机、来源或来源类型）分配一个或多个标记。

分配给字段的另一个名称，您可用此名称来搜索包含该字段的事件。一个字段可以有多个别名，而同一个别名也可应用于多个字段。别名不会取代或删除原始字段名称。

为字段设置别名在提取关键字/值之后但在查找字段之前执行。因此，您可以指定一个基于字段别名的查找表。如果查找表中有一个或多个字段与您数据中的字段相同，但命名方式不同，则这样很有帮助。有关更多信息，请参阅[配置 CSV 和外部查找](#)和[配置 KV 存储查找](#)。

有关别名的更多信息，请参阅[关于标记和别名](#)。

使用字段别名来规范化数据

您可以使用 Splunk Web 直接从搜索结果中为任何字段值对设置标记。

1. 找到搜索中您想要添加别名的字段。
2. 选择**设置 > 字段 > 字段别名**。
3. 选择要使用别名的应用。
4. 设置别名。
5. 选择要应用到已计算字段的主机、来源或来源类型。
6. 选择要创建别名的字段和要创建的别名。
7. 单击保存。

通过 props.conf 创建字段别名

可为一个字段创建多个别名。系统不会删除原始字段。通过此过程可以使用其任何一个别名来搜索原始字段。

为字段设置别名在提取关键字/值之后但在查找字段之前执行。因此，您可以指定一个基于字段别名的查找表。如果查找表中有一个或多个字段与您数据中的字段相同，但命名方式不同，则这样很有帮助。有关更多信息，请参阅[配置 CSV 和外部查找](#)和[配置 KV 存储查找](#)。

可为在索引时间提取的字段以及在搜索时间提取的字段定义别名。

可将字段别名添加到 `props.conf` 中，可在 `$SPLUNK_HOME/etc/system/local/` 或您自己的自定义应用目录 (`$SPLUNK_HOME/etc/apps/`) 中编辑此文件。（如果您希望能够更方便地将数据自定义项传输到其他索引服务器，建议您使用后者一个目录。）

注意：Splunk Enterprise 的为字段设置别名功能目前不支持多值字段。

通过 props.conf 配置字段别名

1. 将下面这行添加到 `props.conf` 的段落中：

```
FIELDALIAS-<class> = <orig_field_name> AS <new_field_name>
```

- `<orig_field_name>` 是字段的原始名称。
 - `<new_field_name>` 是要分配给字段的别名。
 - 您可以在一个段落中包含多个字段别名重命名。
- 重新启动 Splunk Enterprise 使更改生效。

为查找添加字段别名示例

假设您要为外部静态表 CSV 文件创建一个查找，在此查找中将在搜索时间提取的字段 "ip" 称为 "ipaddress"。在已定义了提取的 `props.conf` 文件中，您将添加一行，以将 "ipaddress" 定义为 "ip" 的一个别名，如下所示：

```
[accesslog]
EXTRACT-extract_ip = (?<ip>\d{1,3}\.\d{1,3}\.\d{1,3}\.\d{1,3})
FIELDALIAS-extract_ip = ip AS ipaddress
```

在 `props.conf` 中设置查找时，可以只使用 `ipaddress`，虽然本来应该使用的是 `ip`：

```
[dns]
lookup_ip = dnsLookup ipaddress OUTPUT host
```

有关搜索时间字段提取配置的更多信息，请参阅[通过配置文件创建和维护搜索时间字段提取](#)。

有关查找配置的更多信息，请参阅[配置 CSV 和外部查找](#)和[配置 KV 存储查找](#)。

搜索快捷方式：搜索宏

在搜索中使用搜索宏

搜索宏指可插入其他搜索并可重复使用的搜索处理语言 (SPL) 数据块。搜索宏可以是搜索的任意部分（如 `eval` 语句或搜索术语），不必是完整的命令。您还可以指定宏字段是否使用任何参数。

将搜索宏插入到搜索字符串中

使用反引号字符 (```) 可在搜索字符串插入搜索宏。在大多数英文键盘上，该字符与波形符 (`~`) 位于相同按键上。也可以使用此相同语法在其他搜索宏内引用某搜索宏。如果您有一个名为 `mymacro` 的搜索宏，在搜索中引用时的形式如下：

```
sourcetype=access_* | `mymacro`
```

带引号的值内的宏不会扩展。在以下示例中，搜索宏 `bar` 没有扩展。

```
"foo`bar`baz"
```

包含生成命令的搜索宏

诸如 `search`、`metadata`、`inputlookup`、`pivot` 和 `tstats` 之类的**生成命令**始终会出现在带前导管道字符的搜索字符串的起始处。如果搜索宏的定义以生成命令开头，则搜索宏应插入到搜索字符串的起始处，前面是前导管道字符。不要将前导管道字符放到以生成命令开头的搜索宏的定义中。以下是一个示例：

```
| `mygeneratingmacro`
```

请参阅[在“设置”中定义搜索宏](#)。

搜索宏何时要用参数

如果您的搜索宏要用到参数，则要在将搜索宏插入到搜索字符串时定义这些参数。例如，如果搜索宏 `argmacro(2)` 包含两个整数参数，您可以用以下方式将此宏插入到搜索字符串中：``argmacro(120,300)``。

如果搜索宏参数包含引号，则在搜索中调用宏时需要对引号进行转义。例如，如果您要将具有引号的字符串作为宏的参数进行传递，应使用以下形式：``mymacro("He said \"hello!\"")``。

您的搜索宏定义可包含验证表达式，用于确定您输入的参数是否有效，并会在发现无效参数时发出验证错误消息。

额外资源

有关更多信息，请参阅以下资源。

- [在“设置”中定义搜索宏](#)
- [搜索宏示例](#)
- [《搜索参考》](#)中的生成命令。

在“设置”中定义搜索宏

搜索宏指可插入其他搜索并可重复使用的搜索处理语言 (SPL) 数据块。搜索宏可以是搜索的任意部分（如 `eval` 语句或搜索术语），不必是完整的命令。您还可以指定宏字段是否使用任何参数。

前提条件

- 了解如何[将搜索宏插入到搜索字符串中](#)。
- 了解如何[设计搜索宏定义](#)。
- 如果搜索宏要求搜索编写人员提供参数变量，您可以设计验证表达式，以便了解何时提交了无效参数。请参阅[验证搜索宏参数](#)。

步骤

1. 导航到**设置 > 高级搜索 > 搜索宏**。
2. 单击**新建**即可创建新的搜索宏。
3. **目标应用程序**的默认值如果不对，则将其更改为您要将搜索宏限制到的目标应用。
4. 提供该搜索宏的**唯一名称**。
如果搜索宏包含参数，则需要在名称后附加参数的数目来进行表示。例如，如果搜索宏 `mymacro` 包含两个参数，则将其名称设置为 `mymacro(2)`。
5. 如果您在另一个搜索中引用了该搜索宏，则在**定义**中提供此宏扩展的搜索字符串。
6. (可选) 选择**使用基于 eval 的定义？**以表示定义值是一个 `eval` 表达式。
7. (可选) 为搜索宏选择适当的**参数**。这是一个字符串，包含以逗号分隔的参数名称，不包含重复的元素。参数名称只能包含字母数字字符 (a-Z、A-Z、0-9)，下划线和短划线。
8. (可选) 提供**验证表达式**，用于验证调用搜索宏的参数值是否可接受。验证表达式是一个 `eval` 表达式，用于为布尔值或字符串生成求值结果。
9. (可选) 如果已定义验证表达式，则会提供**验证错误消息**。当调用搜索宏的参数值没有通过验证表达式的验证时，会返回此消息。
10. 单击“保存”以保存搜索宏。

设计搜索宏定义

搜索宏的基础部分是其定义，即当您在另一个搜索中引用搜索宏时此宏扩展出的 SPL 数据块。在设计搜索宏定义之前应了解一些事项。

如果搜索宏定义中包含必须由宏用户输入的变量，则必须将这些变量作为标记放入定义中而且要在两边加上美元符号。例如，`$arg1$` 可以是搜索宏定义中的第一个参数。

宏定义中的管道字符和生成命令

当在搜索中使用诸如 `search`、`inputlookup` 或 `tstats` 之类的**生成命令**时，务必始终将他们放在搜索的起始处，并且带前导管道字符。

但是，如果您想让搜索宏使用生成命令，则必须从宏定义中删除前导管道字符，并将其放在搜索宏的目标搜索字符串的起始处，位于搜索宏引用之前。

例如，假设您有一个名为 `mygeneratingmacro` 的搜索宏，其定义如下：

```
tstats latest(_time) as latest where index!=filemon by index host source sourcetype
```

`mygeneratingmacro` 的定义以生成命令 `tstats` 开头。此时，不可在宏定义中的 `tstats` 前面放置管道字符，而应将管道字符放在搜索字符串中，位于搜索宏引用之前，如下所示：

```
| `mygeneratingmacro`
```

宏定义中的 Eval 表达式

要以 `eval` 命令表达式创建宏定义，请选择**使用基于 eval 的表达式？**。此设置指定了搜索宏定义是一个返回字符串的 `eval` 表达式。此字符串即为宏最终会扩展成的字符串。

验证搜索宏参数

如果搜索宏包含必须由用户输入的参数，则在定义搜索宏时可以定义一个**验证表达式**，用于确定用户提供的参数是否有效。您还可以定义一个**验证错误消息**，在搜索宏参数没有通过验证时显示这一消息。

验证表达式必须是一个 `eval` 表达式，用于为布尔值或字符串生成求值结果。如果验证表达式为布尔表达式，则当验证表达式返回 `true` 时表示验证成功。如果返回 `false` 或为空，则表示验证失败。

如果验证表达式不是布尔表达式，则当验证表达式返回空值时表示验证成功。如果返回一个字符串，则表示验证失败。

额外资源

有关更多信息，请参阅以下资源。

- [搜索宏示例](#)
- 《[管理员手册](#)》中的 `macros.conf`。`macros.conf` 配置文件即为 Splunk 软件存放搜索宏定义的地方。本主题提供更多搜索宏示例。
- 《[搜索参考](#)》中的生成命令。

搜索宏示例

以下为一些搜索宏使用案例及其相应解决方案。

前提条件

- 了解如何[在搜索字符串中使用搜索宏](#)。
- 了解如何在[“设置”中创建或更新搜索宏](#)。

带参数的简单搜索宏

假设您有一组几乎一样的部分搜索：

```
sourcetype="iis" cs_username!="-" /TM/ .pdf
```

```
sourcetype="iis" cs_username!="-" /TD/ .pdf
```

```
sourcetype="iis" cs_username!="-" /TDB/ .pdf
```

您要创建一个搜索宏，让其使用此片段的共同部分并为斜线之间的可变内容传递参数。

步骤

1. 用以下定义创建一个名为 `iis_search(1)` 的搜索宏：

```
sourcetype="iis" cs_username!="-" /$fragment$/ .pdf
```

2. 在**参数字段**，输入**片段**作为参数。
3. 保存新宏。

现在，例如，您可以在搜索字符串中插入 ``iis_search(fragment=TM)`` 以便为 TM 片段调用此搜索宏。

合并搜索宏和交易

交易和宏搜索是一个功能强大的组合，可用来简化交易搜索和报表。本示例演示如何使用搜索宏基于定义的交易构建报表。

名为 `makesessions` 的搜索宏定义了特定事件的交易会话，这些事件拥有相同的 `clientip` 值且是在彼此相隔 30 分钟内所发生的。以下为 `makesessions` 的定义：

```
transaction clientip maxpause=30m
```

此搜索使用 `makesessions` 搜索宏获取 Web 流量事件，并将其拆分为会话：

```
sourcetype=access_* | `makesessions`
```

此搜索使用 `makesessions` 搜索宏每天返回一个报表，其中列出每个会话的 `pageview` 数量：

```
sourcetype=access_* | `makesessions` | timechart span=1d sum(eventcount) as pageviews count as sessions
```

如果您要构建相同的报表，但使用不同的跨度长度，只需将其保存为一个含有跨度长度参数的搜索宏。以下为新搜索宏 `pageviews_per_session(1)` 的定义。注意，此宏引用了原始 `makesessions` 宏。

```
sourcetype=access_* | `makesessions` | timechart $span$ sum(eventcount) as pageviews count as sessions
```

现在，您可以在将其插入搜索字符串时指定跨度长度：

```
`pageviews_per_session(span=1h)`
```

验证参数以确定参数是否为数字

此示例介绍搜索宏的参数验证。

步骤

1. 导航到**设置 > 高级搜索 > 搜索宏**，并单击**新建**即可创建新的搜索宏。
2. 将搜索宏命名为：**newrate(2)**。这个名称表示此宏包含两个参数。
3. 为搜索宏 `newrate(2)` 提供以下定义：

```
eval new_rate=$val*$rate$
```

此定义包含参数变量 `"val"` 和 `"rate"`。

4. 在**参数字段**中，列出 **val** 和 **rate**。
5. `"rate"` 参数只能采用数字值，所以您要设计一个**验证表达式**来确认 `"rate"` 的值是数字。以下为您输入的表

达式：

`isnum($rate$)`

6. 提供以下**验证错误消息**：所提供的 "rate" 值非数字。请输入一个数字值。
7. 保存搜索宏定义。

当在搜索中使用 `newrate(2)` 宏时，可能要按以下方式补充参数：``newrate(revenue, 0.79)``。

注意，如果您没有加零 ``newrate(revenue, .79)``，则该宏无效，因为值 ".79" 前面没有前导零所以被解读为字符串。要确保此参数被解读为浮点值，使用 `tonumber` 函数：``newrate(revenue, tonumber(.79))``

使用数据集

关于数据集

数据集是您为特定商业目的定义和维护的数据集合。如果查看数据集的内容，会看到数据集内的数据以表格形式呈现，其中字段为列、字段值为单元格。“数据集”列表页面用于查看和管理现有数据集。

本主题是数据集的简要概述，主要介绍了三种数据集类型并总结说明了数据集的用途。本主题也介绍了 Splunk 数据集加载项的内容。即使您对 Splunk 搜索处理语言 (SPL) 不熟，也可以用此加载项来自行设计复杂的表数据集并进行持续管理。

数据集的用途

数据集为您提供有用的数据视图。如果已经有了数据集，那么可以用数据集来进行什么操作呢？下表列出了通过“数据集”列表页面可用数据集进行的操作。

数据集活动	为何有用的原因
查看数据集内容	检查数据集以确定它是否包含您想要使用的字段和值。例如，您可以直接查看查找表文件，而不用在搜索视图中搜索这些文件的内容。
在数据透视表中打开数据集	通过数据透视表，您可以基于自己的数据集设计富含可视化元素的分析报表或仪表板面板。数据透视表也可以用于发掘数据趋势和数据集内的字段相关性。
探索搜索中的数据集	将数据集扩展为搜索，按需要修改其搜索字符串，然后将此搜索保存为报表、告警或仪表板面板。请参阅 扩展数据集 。

有关“数据集”列表页面的详细信息，请参阅[查看和管理数据集](#)。

如果您下载并安装了 Splunk 数据集加载项，则可以用数据集进行更多操作。请参阅[在获取 Splunk 数据集加载项之后](#)。

数据集类型

在 Splunk 中可使用的数据集有三种。其中两种数据集类型，即查找和数据模型，是 Splunk 平台中早就存在的知识对象。表数据集，或称为表格，是一种新的数据集类型。在下载并安装了 Splunk 数据集加载项之后即可创建和维护这一数据集类型。

“数据集”列表页面用于查看和管理已有数据集。请参阅[查看和管理已有数据集](#)。

查找

“数据集”列表页面显示两类查找数据集：查找表文件和查找定义。此时，它仅会列出基于文件的 .csv 查找的文件和定义。其他查找类型，如外部查找，KV 存储查找和地理空间查找，目前并未作为数据集列出。

您可前往**设置 > 查找**，通过此页面上载查找表文件和创建基于文件的查找定义。请参阅[使用字段查找将信息添加到事件中](#)。

数据模型数据集

数据模型由一个或多个数据模型数据集组成。当一个数据模型包含多个数据集时，这些数据集分层排列，其中根数据集位于顶部，子数据集位于其下。在数据模型数据集层次结构中，子数据集继承父数据集的字段，但也可以有自己的额外字段。

您可使用“数据模型编辑器”创建和编辑数据模型数据集。请参阅[关于数据模型](#)。

注意：在旧版本的 Splunk 平台中，数据模型数据集称为数据模型对象。

表数据集

表数据集，或称为表格，是具有特定商业目的事件数据的集中而有组织的集合。通过一个简单搜索、索引和来源类型的组合或一个任何类型的已有数据集即可派生出他们的初始数据。例如，您可以创建一个新的表数据集，其初始数据来自特定数据模型数据集。创建好之后，您可以对其进行修改，如更新字段名称、添加字段等等。

您可用表编辑器对数据集进行定义和维护。表编辑器将复杂的搜索命令转换为简单的 UI 编辑器交互。即使您不大了解 SPL 知识，使用起来也非常简单。

Splunk 数据集加载项可用于创建和编辑表数据集。请参阅[在获取 Splunk 数据集加载项之后](#)。

在获取 Splunk 数据集加载项之后

如果您使用的是 Splunk Enterprise，并已从 Splunkbase 下载和安装 Splunk 数据集加载项，则可以创建和管理表数据集（表格）。下表列出了在安装此加载项之后您可以通过数据集实现的其他操作。

如果您使用的是 Splunk Cloud，默认即可拥有以下 Splunk 数据集加载项功能。

数据集活动	为何有用的原因
使用表编辑器创建表格	即使您对 SPL 不熟，也可以针对特定商业需求，设计复杂且高度集中的事件数据集。
持续分享和优化表格	在创建表格之后，您可以赋予其他用户读或写的权限，以便这些用户可以对表格进行管理和优化。例如，您可以快速创建一个简单的数据集，然后将其提供给另一个对源数据更加了解的用户，以便他们因应特定用途对表格进行调整。您也可以扩展您的数据集——创建一个将您的数据集作为起点的新数据集——并让其他人优化此扩展。
查看字段分析	表编辑器提供一个摘要字段视图，其中列出数据集所包含字段的各种分析信息。您可使用此知识来确定要对数据集进行哪些修改以便更符合您的需求。
将任意数据集扩展为表格	您可通过数据集扩展来创建表格，并使用任意数据集类型的定义作为表格的基础。这使您能够创建基于查找和数据模型数据集的表格，并根据自己特定的使用案例对这些表格进行修改。请参阅 扩展数据集 。
复制表格	您可以完全复制表数据集，并用新名字另存。只有表数据集可以进行复制。
加速表格	您可以以类似加速报表和数据模型的方式来加速表格。如果您使用一个很大的数据集作为数据透视表报表或仪表板面板的基础，则加速会起到很大帮助。一旦加速，表格会以较之前更快的速度返回结果。

有关此功能的更多信息，请参阅《*Splunk 数据集加载项的安装和使用*》中的“关于 Splunk 数据集加载项”。

查看和管理已有数据集

“数据集”列表页面用于查看和管理已有数据集。通过“数据集”列表页面，您可以：

- 查看数据集
- 在数据透视表中打开数据集
- 探索搜索中的数据集
- 在数据集原本的编辑环境中对其进行编辑
- 管理查找和表数据集的权限
- 删除数据集

如果您使用的是 Splunk Enterprise，并已安装 Splunk 数据集加载项，则可以通过“数据集”列表页面实现更多操作。Splunk Cloud 用户已配备此加载项。

查看数据集

查看数据集以便了解其结构并判断其是否包含您要使用的信息。数据集以表格形式呈现，其中字段为列、字段值为单元格。数据模型数据集和表数据集以行显示事件。查找以行显示记录。

1. 在搜索和报表应用中，打开“数据集”列表页面。
2. 找到要查看的数据集。

3. (可选) 单击 > 符号展开数据集的行以显示数据集的具体信息。您无须前往查看页面即可查看数据集中包含的字段的列表。
4. 单击数据集名称即可在数据集查看页面中查看其结构和内容。

通过查看页面，您可以执行在列表页面可执行的所有数据集操作。您可以管理权限、编辑数据集描述、在数据透视表中打开数据集，等等。

在数据透视表中打开数据集

数据透视表工具让您精心创建数据集备份报表和仪表板面板，而无须用到 Splunk 搜索处理语言 (SPL)。您可以使用其拖放 UI 来设计复杂表格、图表和可视化，以显示数据集的趋势和模式。

数据透视表中可以打开所有数据集类型。

- 在“数据集”列表页面中，单击**数据透视表**在数据透视表中打开一个数据集，然后开始设计一个基于此数据集的表格或可视化。
- 在查看数据集时，单击**数据透视表**即可在数据透视表中打开此数据集。
- 如果您使用的是 Splunk Cloud 或 Splunk Enterprise，并已安装 Splunk 数据集加载项，则也可以通过表编辑器在数据透视表中打开表数据集。

有关使用数据透视表创建数据集备份报表和仪表板面板的更多信息，请参阅《[数据透视表手册](#)》中的“数据透视表简介”。

浏览搜索中的数据集

在搜索视图中可以浏览数据集中的内容。采用此方式时，使用以 `from` 命令引用原始数据集的搜索字符串打开搜索视图。此搜索返回的结果会列出此数据集中的内容。

此搜索可保存为报表、告警或仪表板面板。新保存的搜索是原始搜索的**扩展**。扩展数据集与其来源父数据集不同但又相关联。如果您对父数据集进行了更改，则此更改会传播给此父数据集扩展出的所有数据集。

前提条件

- 有关数据集扩展的机制和后果，请参阅[扩展数据集](#)。

步骤

1. 在搜索和报表应用中，打开“数据集”列表页面。
2. 找到要在“搜索”中浏览的数据集。
3. (可选) 单击数据集的名称以便在查看页面中查看此数据集。
4. 单击在“**搜索**”中浏览。
此搜索默认以事件列表格式返回结果。将显示格式切换为**表格**，以表格视图查看此数据集。
5. (可选) 用其他 SPL 语言更新搜索字符串。不要删除 `from` 引用。
6. (可选) 单击**另存为**并选择**报表**、**仪表板面板**或**告警**保存您的搜索。
7. (可选) 单击**新表格**基于此搜索字符串创建一个新的表数据集。
只有当您是 Splunk Cloud 用户或已安装 Splunk 数据集加载项时，此选项才可用。

编辑数据集

在“数据集”列表页面和数据集查看页面中，您可以单击进入查找和数据模型数据集的编辑工作流程。

只有安装了 Splunk 数据集加载项，您才能创建和编辑表数据集。

查找

前提条件

- 要了解有关在“设置”中管理查找表文件和编辑查找定义的信息，请参阅[使用字段查找将信息添加到事件中](#)。

步骤

1. 在搜索和报表应用中，打开“数据集”列表页面。
2. 找到要编辑的查找表文件或查找定义。
3. (可选) 单击查找表文件或查找定义的名称以便在数据集查看页面中进行查看。
4. 单击**管理**。根据所选查找数据集的类型，选择**编辑查找表文件**或**编辑查找定义**。

选择	描述	此选择的其他可选步骤
编辑查找表文件	带您前往“设置”页面，其中列出了已上载至 Splunk 平台部署中的查找表文件。	<ul style="list-style-type: none"> ◦ 更新文件权限 ◦ 将文件移至其他应用上下文 ◦ 删除文件

		<ul style="list-style-type: none"> ◦ 上载新的 .csv 文件
编辑查找定义	带您前往“设置”页面，其中列出了查找定义。	<ul style="list-style-type: none"> ◦ 配置查找为基于时间的查找 ◦ 设置高级字段匹配规则

数据模型数据集

前提条件

- 要了解有关使用“数据模型编辑器”更新数据模型及其数据集的更多信息，请参阅[设计数据模型数据集](#)。

步骤

1. 在搜索和报表应用中，打开“数据集”列表页面。
2. 找到要编辑的数据模型数据集。
3. （可选）单击数据集的名称以便在数据集查看页面中进行查看。
4. 选择**管理 > 编辑数据模型**。
5. （可选）使用“数据模型编辑器”更新数据模型数据集的约束和属性（字段）。

管理数据集权限

拓宽或缩小其他用户对数据集的访问权限。可以按角色设置读和写权限，也可以决定数据集是否全局都可访问、只有特定应用上下文可访问、还是只有单个用户可以访问。

有关 Splunk 平台权限功能的工作原理简介，请参阅[管理知识对象权限](#)。

查找和表数据集

您可以通过“数据集”列表页面为查找和表数据集设置权限。

1. 在搜索和报表应用中，打开“数据集”列表页面。
2. 找到要查看或更新权限的查找或表数据集。
3. 选择**管理 > 编辑权限**。
4. （可选）更改此数据集的**为...显示**的目标用户。此数据集可以仅显示给特定**应用**或**所有应用**的用户（全局可访问性）。表数据集也可以设为**专用**，表示只有创建者可以访问。
5. （可选）如果数据集设置为显示给特定**应用**或**所有应用**，您可以修改**读**和**写**的设置，决定哪些角色可以查看或编辑此数据集。
6. 单击**保存**以保存您所做的修改，或**取消**以撤消您所做的修改。

数据模型数据集

数据模型数据集的权限是在数据模型级设置的。数据模型内的所有数据集具有相同的权限设置。设置数据模型权限的方式有两种：

- 通过“数据模型编辑器”
- 通过“设置”中的“数据模型”列表页面

前提条件

- 要了解设置数据模型权限的相关信息，请参阅[管理数据模型](#)。

步骤

通过	操作
数据模型编辑器	<ol style="list-style-type: none"> 1. 在搜索和报表应用中，打开“数据集”列表页面。 2. 识别要更新权限的数据模型数据集。 3. 选择管理 > 编辑数据模型。 4. 选择编辑 > 编辑权限。此操作将设置所选数据模型数据集所属的数据模型的权限。 5. （可选）更改此数据模型的为...显示的目标用户。此数据模型可以仅显示给特定应用或所有应用的用户。 6. （可选）如果数据模型设置为显示给特定应用或所有应用，您可以修改读和写的设置，决定哪些角色可以查看或编辑此数据模型。 7. 单击保存以保存您所做的修改，或取消以撤消您所做的修改。
	<ol style="list-style-type: none"> 1. 选择设置 > 数据模型 2. 识别要更改权限的数据模型。

设置 > 数据模型

3. 单击**编辑** > **编辑权限**。此操作将设置所选数据模型数据集所属的数据模型的权限。
4. (可选) 更改此数据模型的**为...显示**的目标用户。此数据模型可以仅显示给特定**应用**或**所有应用**的用户。
5. (可选) 如果数据模型设置为显示给特定**应用**或**所有应用**，您可以修改**读**和**写**的设置，决定哪些角色可以查看或编辑此数据模型。
6. 单击**保存**以保存您所做的修改，或**取消**以撤消您所做的修改。

删除数据集

您可以通过“数据集”列表页面删除查找和表数据集。要删除数据模型数据集，请转到“数据模型编辑器”。

查找和表数据集

1. 在搜索和报表应用中，打开“数据集”列表页面。
2. 找到要删除的查找或表数据集。
3. 选择**管理** > **删除**。
4. 在**删除数据集**对话框中，再次单击**删除**确认您要删除此数据集。

您也可以通过数据集查看页面删除相关查找和表。

数据模型数据集

1. 在搜索和报表应用中，打开“数据集”列表页面。
2. 找到要删除的数据模型数据集。
3. 选择**管理** > **编辑数据集**。
4. 在“数据模型编辑器”中，针对目标数据模型数据集单击**删除**。

如果您已安装 Splunk 数据集加载项

如果您已安装 Splunk 数据集加载项，则可以通过“数据集”列表页面实现更多操作。您可以：

- 访问表编辑器，并创建新的表数据集。
- 在编辑器中打开已有表数据集进行编辑。
- 将任意数据集扩展为新的表数据集。
- 复制已有表数据集。
- 编辑表数据集描述。

有关此功能的更多信息，请参阅《Splunk 数据集加载项的安装和使用》中的“关于 Splunk 数据集加载项”。

数据集扩展

数据集扩展是一种可用于创建新搜索、报表、数据集或其他对象的快捷方式，这些对象都基于到已有数据集的引用构建。此引用表示，对于基础数据，这些对象始终会“回顾”原始数据集。如果原始数据集的定义发生了改变，这些改变会传递到其所有的扩展数据集中。

数据集扩展和数据集复制有所不同。复制数据集时，您建立的是一个与原始数据集一样但却无关联的不同独立数据集。而扩展时，您创建的数据集、报表、仪表板面板或告警通过引用是与原始数据集绑定在一起的。

数据集扩展为报表示例

例如，假设您有一个名为 Alpha 的数据集。如果您针对 Alpha 数据集在“数据集”列表页面中单击在“**搜索**”中**浏览**，则会显示搜索视图。在此视图中，您可以运行一个搜索以显示 Alpha 的内容。此搜索字符串使用 `from` 命令引用 Alpha。您可以添加其他 Splunk 搜索处理语言 (SPL) 至此搜索字符串中，以便在需要时通过查找和 `eval` 计算添加新字段。

如果您将此搜索字符串另存为报表并命名为 Beta，此报表仍然包含至 Alpha 的引用。这意味着，如果有人决定要修改 Alpha 时，这些修改会向下传递给 Beta 报表。而这可能会造成问题，但也可能不会，具体取决于实际修改点。

例如，您可能修改带查找和 `eval` 表达式的 Beta 搜索字符串，而这些搜索和表达式在其定义中使用了来自 Alpha 数据集的字段。如果有人从 Alpha 数据集中删除了这些字段，则在 Beta 报表中的相关查找和 `eval` 表达式会损坏。

数据集扩展链

如果您已安装 Splunk 数据集加载项，则可以将任意数据集扩展为表数据集。这意味着有可能会扩展数据集链。例如，您可以将 Alpha 数据集扩展为 Beta 数据集，然后将 Beta 数据集扩展为 Gamma 数据集，并依此类推。那么，Alpha 数据集中的任何更改都可能会通过链中的其他数据集向下传递。

如果您要使用数据集扩展，则必须小心使用。Splunk 数据集加载项可让您从链尾了解数据集扩展链，但无法从链头

了解。继续以之前段落中的示例为例，如果您位于数据集 Gamma，您可以发现它由 Beta 扩展而来，而 Beta 由 Alpha 扩展而来。但如果您查看的是 Alpha，则无法得知通过它扩展出了哪些数据集。

要了解一个数据集扩展出了哪些数据集

在“数据集”列表页面中找到该数据集并展开行。如果此数据集扩展了一个或多个数据集，则此页面会包含一个扩展行，其中从上至下的列出所扩展的数据集。例如， 以下为 Gamma 的详细信息，显示它扩展了 Alpha 和 Beta。

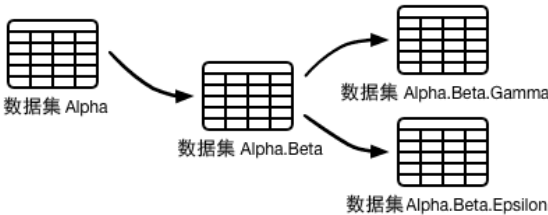
▼	Gamma
数据集类型	表格
应用	search
权限	专用。由 admin 拥有。 编辑
字段	_time, sourcetype, _raw
延伸	Alpha > Beta

在数据集的查看页面也能找到此信息。单击[更多信息](#)了解您所查看的数据集扩展了哪些数据集。

对扩展的数据集采用命名约定

当操作一个数据集时，很难了解哪些数据集是其扩展出来的。例如，一个用户在操作 Alpha 数据集，此用户无法得知 Beta 和 Gamma 数据集是其扩展数据集。

此时，可以采用一种命名约定，当数据集是从另一个数据集扩展而来时，通过名称进行说明。例如，如果您从数据集 Alpha 扩展出一个新数据集，则可命名为 Alpha.Beta。之后，如果您从 Alpha.Beta 扩展出两个数据集，则可分别命名为 Alpha.Beta.Gamma 和 Alpha.Beta.Epsilon。此命名方式与数据模型数据集的相似，其中数据集名称可表示它在数据模型数据集的大层次结构中所处的位置。



当扩展数据集时，也可以更新其描述，说明此数据集已被扩展。您只需指出直接从此数据集扩展出的知识对象，而非完整的扩展链，如果有的话。在数据集描述中添加类似如下示例的句子：“此数据集已扩展为一个名为<数据集_名称>的表数据集和一个名为<报表_名称>的报表。”

from 命令

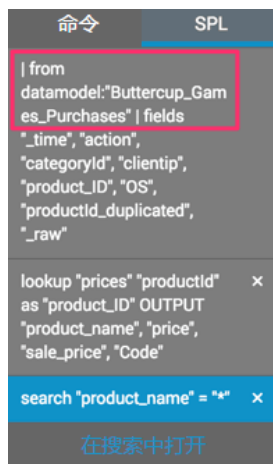
不管是在搜索视图中打开并扩展数据集还是通过表编辑器来扩展，`from` 命令都使数据集扩展变得更为简单。

当在搜索视图中打开数据集时，会看到以 `from` 命令从此数据集检索数据的搜索字符串。例如，假设您有一个名为 Buttercup Games Purchases 的数据集。在“数据集”列表页面中，如果您针对此数据集单击在“搜索”中浏览，则 Splunk 会显示搜索视图。在此视图中，您可以看到类似下面的搜索字符串：

```
| from datamodel:"Buttercup Games Purchases"
```

如果您使用的是 Splunk Cloud，或虽然使用的是 Splunk Enterprise 但已安装 Splunk 数据集加载项，则可以将任何数据集扩展为表数据集。这样操作时，表编辑器在后台使用 `from` 命令。单击命令历史边栏中的 **SPL** 标记，查看表编辑器如何使用 `from` 命令。

在这个对于表编辑器中命令历史边栏的特写中，您可以了解此表数据集的初始数据由 Buttercup Games Purchases 数据集的 `from` 命令扩展提供。



有关更多信息，请参阅《[搜索参考](#)》中的 `from`。

构建数据模型

关于数据模型

数据模型用来驱动数据透视表工具。借助数据模型，Pivot 用户可创建引人注目的报表和仪表板，而无需对生成它们的搜索进行设计。数据模型可有一些其他用途，尤其适合 Splunk 应用开发人员使用。

Splunk 知识管理员设计和维护数据模型。知识管理员了解其索引数据的格式和语义，并熟悉 Splunk Enterprise 搜索语言。在构建典型数据模型中，知识管理员使用知识对象类型，如[查找](#)、[交易](#)、[搜索时间字段提取](#)和[已计算字段](#)。

什么是数据模型？

数据模型是有关一个或多个数据集语义知识的分层结构搜索时间映射。它将编码构建这些数据集的各种专门搜索所需的域知识。这些专门搜索被 Splunk 软件用于为数据透视表用户生成报表。

数据透视表用户设计数据透视表报表前，会选择用来表示将要使用的事件数据类别的数据模型，例如“Web 智能”或“电子邮件日志”。然后在该数据模型中选择用来表示要报告的特定数据集的[数据集](#)。数据模型由数据集（可按父/子数据集的分层结构排列）组成。每个子数据集代表其父数据集所涵盖的数据集的一部分。

如果熟悉关系数据库设计，则将数据模型视为数据库方案的类似结构。将数据模型连入“数据透视表编辑器”时，您可以根据所选的列和行配置生成统计表格、图表和可视化。

要创建有效的数据模型，您必须了解您的数据源和数据语义。本信息会影响您的数据模型架构——组成数据模型的数据集的组织方式。

例如，如果数据集以基于表格的数据格式的内容为依据（如 .csv 文件），则生成的数据模型会相对简单，只用一个顶级根数据集来涵盖表格各列表示的字段。根数据集下面可能具有子数据集。但這些子数据集不包含从根数据集继承的字段集之外的其他字段。

同时，源自异类系统日志的数据模型可能具有多种根数据集（事件、搜索和交易）。每个根数据集都可以是带父/子关系的数据集层次结构中的第一个数据集。数据集层次结构中的每个子数据集除具有继承自其上级数据集的字段外，还可以拥有一些新字段。

数据模型数据集可从您定义的自定义[字段提取](#)中获取字段。数据模型数据集可通过基于正则表达式的字段提取、[查找](#)和 `eval` 表达式使其在搜索时获得其他字段。

数据模型所使用的字段分成上文所述的类别（自动提取、`eval` 表达式和正则表达式）及其他（[查找](#)和地理 IP）。请参阅[数据集字段](#)。

数据模型属于[知识对象](#)类别，因此具有充分的权限。数据模型的权限涵盖其所有数据模型数据集。

请参阅[管理数据模型](#)。

数据模型生成搜索

考虑什么是数据模型以及它们如何工作时，将数据模型视为用来生成不同搜索类的结构化信息的集合非常有益。数据模型中的每个数据集都可用来生成将返回特定数据集的搜索。

以下小节我们将更加详细地介绍数据模型、数据模型数据集和搜索之间的这种关系。

- **数据集约束** 通过以下方法确定搜索的第一部分：
 - 简单搜索过滤器（根事件数据集和所有子数据集）。
 - 复杂搜索字符串（根搜索数据集）。
 - `transaction` 定义（根交易数据集）。
- 为数据透视表选择数据集时，针对该对象定义的未隐藏字段包含了您将在决定要报告的数据集时从数据透视表中选择的字段的列表。所选的字段会添加到数据集生成的搜索。这些字段包括**已计算字段**、用户定义的字段提取以及通过查找添加到数据中的字段。

数据集生成的搜索的最后部分由“数据透视表编辑器”选项决定。他们会添加转换命令至将结果聚合为统计表格的搜索中。随后数据透视表会将本表作为图表和其他可视化类型的基础。

有关如何使用“数据透视表编辑器”创建基于数据模型数据集的数据透视表表格、图表和可视化的更多信息，请参阅《[数据透视表手册](#)》中的“数据透视表简介”。

数据集

数据模型由一个或多个数据集组成。以下列出了数据模型数据集的一些基本事实：

- **每个数据模型数据集都对应索引中的一组数据。** 您可以将数据模型应用于不同的索引以获得不同的数据集。
- **数据集分成四种类型。** 这些类型包括：[事件数据集](#)、[搜索数据集](#)、[交易数据集](#)和[子数据集](#)。
- **数据集具有层次结构。** 可以使用父/子关系分层排列数据模型中的数据集。数据模型中的顶级事件、搜索和交易数据集统称为“根数据集”。
- **数据集具有继承性。** 数据模型数据集由一般分为**约束**和**字段**的特性进行定义。子数据集会继承其父数据集的约束和字段，同时也包含自己的附加约束和字段。

我们将在以下小节中更加详细地介绍这些信息以及数据模型数据集的其他方面。

- **子数据集提供筛选来自父数据集事件的方式** - 由于子数据集始终在继承自其父数据集的约束之外提供附加约束，因此它所表示的数据集始终是其父数据集所表示数据集的子集。

根数据集和数据模型数据集类型

数据模型中的顶级数据集称为“根数据集”。数据模型可包含多个不同类型的根数据集，其中的每个根数据集可以是多个子数据集的父数据集。基本数据集与子数据集之间的关联称为“数据集树”。数据集树代表的整组数据首先由根数据集选择，然后由其子数据集优化和扩展。

根数据集可通过搜索约束、搜索或交易来定义：

- **根事件数据集**是最常用的根数据模型数据集类型。每个根事件数据集都广泛地表示某一类型的事件。例如，[HTTP 访问](#)根事件数据集对应于访问日志事件，而[错误](#)事件则对应于包含错误消息的事件。根事件数据集通常由简单的约束定义。经验丰富的 Splunk 用户会将此约束视为搜索在应用管道符、命令和参数之前的第一部分。例如，`status > 600` 和 `sourcetype=access_* OR sourcetype=iis*` 为可能的事件数据集定义。请参阅[数据集约束](#)。
- **根搜索数据集**使用任意 Splunk 搜索来定义其所表示的数据集。如果要定义一个包括在整个数据集内聚合的一个或多个字段的基本数据集，则可能需要使用在其搜索中包含转换命令的根搜索数据集。例如，各种系统入侵事件在不同时段按类别分开的系统安全数据集。
- **根交易数据集**可用于创建表示**交易**的数据模型：某一时段内的相关事件组。交易数据集定义通过事件或搜索数据集利用已经添加到模型的字段，这意味着您无法创建仅由交易数据集及其子数据集组成的数据模型。创建交易数据集之前，模型中必须已经具备一些事件或搜索数据集树。

事件、交易和搜索这三种根数据集类型的子数据集都通过简单的约束来定义，简单约束可以缩小子数据集从其上级数据集继承的数据集范围。

数据集类型和数据模型加速

您可以选择使用[数据模型加速](#)来加速数据透视表表格和图表的生成。在您认为用户可受益于数据模型加速时，请注意此功能对数据模型构建方式存在一定限制。

要加速数据模型，必须包含至少一个根事件数据集，或一个仅使用流命令的根搜索数据集。加速只会影响这些数据集类型和这些根数据集的子数据集。使用非流命令（包括转换命令）的根搜索数据集、根交易数据集和这些数据集的子数据集都无法加速。数据模型可同时包括加速的数据集和未加速的数据集。

请参阅[管理数据模型](#)。

要了解有关流命令和其他命令类型的详细信息，请参阅《[搜索参考](#)》中的“命令类型”。

数据模型数据集层次结构示例

以下示例介绍了“呼叫详细记录”数据模型中的前几个数据集。显示了四个顶级根数据集：**全部呼叫**、**全部交换机记录**、**通话**和**呼出**。

选择对象	
上一步	
	23个对象 Call Detail Records
	All Calls
	Voice
	SMS
	Data
	Roaming
	All Switch Records
	ATT Carrier
	Metro Carrier
	SWB Carrier
	VER Carrier
	Virgin Carrier
	Conversations (1 day maxspan, 5 hours maxpause)
	Outgoing Calls (1 day maxspan)
	All Calls and Switches

全部呼叫和**全部交换机记录**为根事件数据集，分别表示所有呼叫记录和所有运营商交换机记录。这两个根事件数据集都具有子数据集，是其父项所拥有数据的子集。**全部呼叫**根事件数据集的子数据集分成不同的呼叫分类：语音、SMS、数据和漫游。如果数据透视表用户只想报告手机的数据使用方面，则应选择“数据”数据集。不过，如果要创建这四个呼叫类型的对比报表，则应选择**全部呼叫**根事件数据集。

通话和**呼出**为根交易数据集。二者都表示交易，即某一时段内的相关事件分组。“通话”数据集只包含两名或更多名人员之间通话的呼叫记录，其中，通话呼叫记录事件之间的最大暂停时间不超过两小时，通话总时长不超过一天。

有关定义不同数据模型数据集定义的详细信息，请参阅“设计数据模型数据集”。

数据集约束

所有数据模型数据集都由**约束集**定义。数据集约束可以筛选出与数据无关的事件。

- **对于根事件数据集或任意类型的子数据集**，约束都类似于没有附加管道和搜索命令的简单搜索。例如，“Web 智能”数据模型的根事件数据集之一 HTTP Request 的约束为 `sourcetype=access_*`。
- **对于根搜索数据集**，约束为数据集搜索字符串。
- **对于根交易数据集**，约束为交易定义。交易数据集定义必须指定**组合数据集**（一个或多个事件数据集、一个搜索数据集或交易数据集）和一个或多个**分组依据**字段。其中也可以选择包括**最大暂停时间**和**最大跨度**值。

子数据集会继承约束。约束继承特性可确保每个子数据集都代表其父数据集所表示数据的子集。这样，数据透视表用户即可使用这些子数据集，通过已经预先筛选出多余数据的数据集设计报表。

Buttercup Games

Tutorial

< 所有数据模型

数据集

添加数据集

事件

HTTP Requests

Client Errors

Server Errors

Purchases

Successful Purchases

Failed Purchases

Successful Purchases

Successful_Purchases

重命名 删除

约束

sourcetype=access_*

继承

action=purchase

继承

status=200

约束

编辑

批量编辑

添加字段

假定您有一个名为 Buttercup Games 的数据模型。它的“成功购买”数据集是根事件数据集 HTTP Requests 的子数据集，并且设计用于只包含那些代表了成功的客户购买操作的事件。“成功购买”继承了 HTTP Requests 和另一个名为“购买”的父数据集的约束。

1. HTTP Requests 通过设置只查找 webserver 访问事件的搜索来启动。

```
sourcetype=access_*
```

2. “购买”数据集将关注范围进一步缩小为包含购买操作的 webserver 访问事件。

```
action=purchase
```

3. 最后，“成功购买”会添加一个约束，将数据集事件集减少为代表了成功购买事件的 Web 访问事件。

```
status=200
```

将所有约束添加到一起后，“成功购买”数据集的基本搜索如下：

```
sourcetype=access_* action=purchase status=200
```

当数据透视表用户确定只需要报告成功的购买操作时，即可使用此数据集进行报告。

有关数据集和数据集约束的详细信息，请参阅主题[设计数据模型数据集](#)。

数据集字段

数据集字段有五个类别：

- **自动提取：**在搜索时间派生的字段。自动提取的字段只能添加到根数据集。子数据集可以继承自动提取的字段，但无法添加自己的新自动提取的字段。自动提取的字段可以是：
 - 自动提取的字段，比如 `uri` 或 `version`。这包括通过结构化数据输入建立索引的字段，如从索引 CSV 文件标头提取的字段。
 - 您已在“设置”中定义或在 `props.conf` 中配置的**字段提取、查找或已计算字段**。
 - 您已手动添加的字段，因为它们虽然当前未在数据集中，但将来应该在其中。可以包含通过生成命令（如 `inputcsv` 或 `dbinspect`）添加到数据集的字段。
- **Eval 表达式：**一个源自字段定义中所输入 `eval` 表达式的字段。Eval 表达式通常涉及一个或多个提取的字段。
- **查找：**一个利用在字段定义中配置的**查找**添加到数据集中事件的字段。“查找”从 CSV 文件和脚本等外部数据源添加字段。定义查找字段时，可使用已经在**设置中定义**的任意查找，并将其与其他任意已经关联到同一数据集的字段相关联。
- **正则表达式：**使用您在字段定义中所提供的正则表达式从数据集事件数据中提取的字段。正则表达式字段定义可使用提取多个字段的正则表达式；每个字段都将在数据集字段列表中显示为单独的正则表达式字段。
- **地址 IP：**一种特定的**查找**类型，可将纬度、经度、国家/地区和城市等地理字段添加到数据集中具有有效 IP 地址字段的事件。这对于地图相关的可视化非常有用。

请参阅[设计数据模型数据集](#)。

字段类别

“数据模型编辑器”将字段分为三种类别：

- **已继承：**所有数据集具有至少几个已继承字段。子字段继承来自其父数据集的字段，且这些继承的字段始终显示在“已继承”类别中。根事件、搜索和交易数据集也具有分类为已继承的默认字段。
- **已提取：**任何添加到数据集的自动提取字段都被列入“已提取”字段类别。
- **已计算：**Splunk 软件通过计算、查找定义或字段匹配正则表达式来派生已计算字段。当您添加 Eval 表达式、正则表达式、查找和地理 IP 字段类型添加到数据集时，它们都会显示于此字段类别。

“数据模型编辑器”可用于排列已计算字段的顺序。当您有一组必须要按特定顺序处理的字段时，这个工具就非常有用了。例如，您可以定义一个 Eval 表达式，将一组字段添加到数据集的事件中。然后，您可以创建一个查找，并使其定义使用此 eval 表达式所计算的其中一个字段。此查找使用此定义将另一组字段添加到同一批事件中。

字段会被继承

所有数据集都有继承字段。

子数据集将自动拥有属于其父项的所有字段。所有这些已继承的字段将显示在子数据集的“已继承”类别中，即使这些字段已在父数据集中进行了分类。

您可以将其他字段添加到子数据集。根据字段类型，“数据模型编辑器”会将这些数据集分类为已提取字段或已计算字段。

您可以设计一个相对简单的数据模型，在其根数据集中定义数据集树的所有必要字段，这就意味着数据集树中所有的子数据集都与其根数据集拥有完全相同的字段集。在此类数据模型中，子数据集之间以及子数据集与根数据集之间只能通过约束区分。

根事件、搜索和交易数据集也具有已继承字段。这些已继承字段是从每个事件（如 `_time`、`host`、`source` 和 `sourcetype`）中提取的默认字段。

您无法删除已继承字段，也无法编辑其定义。编辑或删除属于子数据集的已继承字段的唯一方式是将源自父数据集的字段作为已提取或已计算字段进行编辑或删除。对于源自于根数据集的已继承字段，您无法进行删除或编辑。

您可对数据透视表用户隐藏此字段，作为字段删除的替代方法。

您还可以决定已继承字段对数据集是可选的还是必需的。

字段用于几个目的

它们最明显的功能是提供了一组字段，以便数据透视表用户用于定义和生成数据透视表报表。数据透视表用户拥有访问权限的字段集由用户进入“数据透视表编辑器”时选择的数据集决定。您可以添加字段到子数据集，以提供字段给该数据集的特定数据透视表用户。

此外，您也可以设计仅用于设置其他字段或约束定义的已计算字段。以下为**字段列出顺序很重要**的原因所在：字段会按他们在“数据模型编辑器”中列出的顺序进行处理。这就是“数据模型编辑器”允许您重新排列已计算字段排列顺序的原因。

例如，您可以设计三个 Eval 表达式字段的**链式集**。前两个 Eval 表达式字段实际上用于创建**已计算字段**。第三个 Eval 表达式字段则在其 eval 表达式中使用这两个已计算字段。

可设置字段对数据透视表用户可见还是隐藏

定义字段时，可决定该字段对数据透视表用户是**可见**还是**隐藏**。如果数据模型中的每个数据集都具有很多字段，但每个数据集只有部分字段对数据透视表用户有用，则此方法会对您有所帮助。

注意：字段可在某些数据集中可见，而在其他数据集中隐藏。隐藏父数据集中的字段不会导致该字段在其子数据集中也隐藏。

默认情况下，字段均可见。数据集的已隐藏字段在数据集字段列表中标记为隐藏。

在数据透视表中，决定在模型中要包含哪些字段以及对特定对象显示哪些字段可简化数据集的使用。通常情况下，如果每个数据集只显示与该数据集相关的数据，则有利于数据透视表用户简化有用报表的构建。例如，这意味着您可以向根数据集添加一些特定的字段，这些字段除了对层次结构中某处的特定数据集可见外（在层次结构中，它们的可见性在此数据集及其特定数据集上下文中可起到特殊作用），在整个模型中的其他位置都被隐藏。

以上一小节中提到的示例为例。此示例包含一组三个“链式”Eval 表达式字段。您要隐藏前两个 Eval 表达式字段，因为它们只是第三个字段的“输入”。您已使第三个字段可见，因为它是最终“输出”，即对数据透视表非常有用的字段。

对于数据集，字段可为必需或可选

字段设计过程中，您还可以决定字段为**必需**还是**可选**。这一操作可以作为一种过滤器来过滤数据集所表示的事件集。如果指定字段**必需**，则表示此数据集所表示的每个事件都必须包含该字段。如果将字段定义为**可选**，则此数据集的部分事件可能不包含该字段。

注意：通过字段可见性（参阅上文），字段对某些数据集来说是必需的，而对其他数据集来说是可选的。在父数据集中将字段标记为**必需**不会自动使该字段对于其子数据集来说是**必需**。

默认情况下，字段均为可选。数据集中状态已更改为**必需**的字段在此数据集的字段列表中也标记为必需。

管理数据模型

“数据模型”管理页面可供用户创建数据模型并维护权限和加速等某些“高优先级”功能。您可以在此页面上：

- **创建新的数据模型** - 该过程如单击按钮一样简单。
- **设置权限** - 数据模型为知识对象，因此具有权限。权限用于确定哪些人员可以查看和更新数据模型。
- **启用数据模型加速** - 此功能可提升涵盖大型数据集的数据模型的数据透视表性能。
- **复制数据模型** - 可用于快速创建基于现有数据模型的新数据模型，或将数据模型复制到其他应用。
- **上传和下载数据模型** - 下载数据模型（将其导出到 Splunk 外部）。将已导出的数据模型上传到不同的 Splunk 安装系统。
- **删除数据模型** - 删除不再有用的数据模型。

本主题我们将讨论数据模型管理的上述方面。需要定义组成数据模型的数据集层次结构时，请转到“数据模型编辑器”。有关更多信息，请参阅[设计数据模型数据集](#)。

导航到“数据模型”管理页面

“数据模型”管理页面实际上是一个列表页面，与“告警”、“报表”和“仪表板”列表页面相似。该页面可用于管理权限和加速，也可用于复制和删除数据模型。该页面与您首次进入数据透视表时看到的“选择数据模型”页面（仅当具有多个数据模型时才可见）有所不同，后者仅供数据透视表用户用来选择创建数据透视表时要使用的数据模型。

“数据模型”管理页面通过分页表格列出系统中的所有数据模型。该表格可按应用、所有者和名称过滤。也可以显示所有对选定应用的用户可见的数据模型，或者仅显示应用中实际创建的数据模型。

如果您使用的是 Splunk Cloud 或者是 Splunk Enterprise 但已安装 Splunk 数据集加载项，则也可以通过“数据模型”管理页面查看表数据集。

请参阅[关于数据集](#)了解更多关于表数据集的信息。

有两种方式可以打开“数据模型”管理页面。使用“设置”列表，或通过“数据集”列表页面和“数据模型编辑器”打开。

通过“设置”列表

导航到 **设置 > 数据模型**。

通过“数据集”列表页面

- 1. 在搜索和报表应用中，打开“数据集”列表页面。
- 2. 找到一个数据模型数据集。
- 3. （可选）单击数据模型数据集的名称以便在数据集查看页面中进行查看。
- 4. 针对此数据集选择**管理 > 编辑数据模型**。
- 5. 在“数据模型编辑器”中，单击左上角的 **< 所有数据模型** 链接前往“数据模型”管理页面。

创建新数据模型

前提条件

只有在权限允许的情况下才能创建数据模型。您的角色必须有权限向至少一个应用中写入数据。如果您的角色具备的权限不足，则系统不会显示**新数据模型**按钮。

请参阅[允许角色创建数据模型](#)。

步骤

- 1. 导航到“数据模型”管理页面。
- 2. 单击**新建数据模型**创建一个新的数据模型。
- 3. 输入此数据模型的**标题**。
标题字段可接受除星号外的任何字符。它还可接受字符间的空格。
在您输入标题时，数据模型 **ID** 字段会自动填入。不要进行修改。数据模型 **ID** 必须是数据模型的唯一标识符。其中只能包含字母、数字和下划线。同时不允许在两个字符之间使用空格。在单击**创建**之后，即无法更改 **ID** 值。
- 4. （可选）输入此数据模型的**描述**。
- 5. （可选）如果要使数据模型属于其他应用上下文，可更改**应用**的值。**应用**显示的是您当前所处的应用上下文。
- 6. 单击**创建**在“数据模型编辑器”中打开新数据模型，您可以在该编辑器中添加和定义组成数据模型的数据集。

创建新数据模型

标题

Web Intelligence

ID ?

Web_Intelligence

只能包含字母、数字和下划线。

应用

Search & Reporting ▾

描述

Enables data analytics and reporting for webserver activity

取消

创建

首次进入新数据模型的“数据模型编辑器”时，数据模型不包含任何数据集。要定义数据模型的首个数据集，单击**添加数据集**并选择数据集类型。有关数据集定义的更多信息，请参阅有关添加字段、搜索、交易和子数据集的以下各章节内容。

有关“数据模型编辑器”和数据模型数据集创建操作的所有详细信息，请参阅[设计数据模型数据集](#)。

允许角色创建数据模型

默认情况下，仅拥有 **管理员**或**高级用户**角色的用户可以创建数据模型。对于其他用户，创建数据模型的能力与他们的角色是否拥有对应用程序的“写入”访问权限相关。要为其授予应用写入访问权限，请遵循下列步骤。

步骤

- 1. 单击页面顶部的**应用**下拉列表，选择**管理应用**转到“应用”页面。
- 2. 在“应用”页面上，找到要为其授予数据模型创建权限的应用，然后单击**权限**。
- 3. 在应用的“权限”页面上，为能够为应用创建数据模型的角色选择**写入**。

4. 单击**保存**以保存更改。

为角色赋予创建数据模型的功能会产生一些其他影响。

请参阅[禁用或删除知识对象](#)。

关于数据模型权限

数据模型为知识对象，因此其查看和编辑功能由基于角色的权限控制。在您首次创建数据模型时，该数据模型为您的专用模型，即，其他任何用户都无法在“选择数据模型”页面或“数据模型”管理页面查看此模型，也无法通过任何方式更新此模型。

如果要加速一个数据模型，则首先要将其共享。专用数据模型无法进行加速。请参阅[启用数据模型加速](#)。

将数据模型权限与相关知识对象的权限对齐

当您共享数据模型时，与该数据模型关联的知识对象（如查找或字段提取）必须具有相同的权限。否则，其他人在使用此数据模型时可能会发生错误。

例如，如果您的数据模型共享给“搜索”应用的所有用户，但该数据模型使用的查找表和查找定义仅共享给具有“管理员”角色的用户，则对于“管理员”角色用户，一切将正常工作，但所有其他用户尝试在数据透视表中使用时，系统会发出错误消息。解决方案是将数据模型限定于“管理员”用户或将查找表和查找定义共享给“搜索”应用的所有用户。

编辑数据模型的权限

前提条件

- [管理知识对象权限](#)

步骤

1. 前往“数据模型”管理页面。
2. 找到要为其编辑权限的数据模型。使用以下各选项中的其中一项。

选项	此选项的其他步骤
选择 编辑 > 编辑权限 。	无
展开此数据集的行。	单击 编辑 以编辑权限。

- 3.
4. 编辑数据集权限并单击**保存**以保存所做的更改。

将打开**编辑权限**对话框，您可使用它来与其他人共享专用数据模型和确定不同角色对于数据模型的访问级别。

启用数据模型加速

在启用数据模型加速后，使用该数据模型的数据透视表、报表和仪表板面板返回报表的速度会比之前快。

数据模型加速由**高性能分析存储**提供支持。借助高性能分析存储的强大功能，数据模型可在索引级别为数据模型构建数据摘要。该摘要实际可由多个更小型的摘要组成，分布在整个索引器当中。

当摘要完成构建之后，使用加速的数据模型数据集的数据透视表将尽可能地针对此摘要而不是全部 `_raw` 数据运行。这样可以大大缩短数据透视表返回结果的时间。

虽然数据模型加速可用于加速非常大型的数据集，但同时也应谨记几个要点。

- **默认情况下，只有具备管理员权限的用户才能加速数据模型。**数据模型加速会耗用大量资源，因此应适当地供有限数量的 Splunk 用户使用此功能。数据模型加速功能与 `accelerate_datamodel` 操作相关。
- **无法为专用数据模型启用加速。**必须与应用的各用户共享数据模型才能实现加速。执行此操作时，您还需要以完全相同的方式共享相关知识对象（如您的查找字段所依赖的查找表和查找定义）。有关更多信息，请参阅“关于数据模型权限”。
- **一旦您加速了数据模型，则无法对其进行编辑。**如果您要对加速的数据模型进行更改，则需要禁用其加速。重新加速数据模型会耗用大量资源，因此，最好尽可能避免禁用加速。
- **数据模型加速仅可应用于根事件数据集、仅使用流命令的根搜索数据集、及其子数据集。**基于根交易数据集或使用非流命令的根搜索数据集的数据集层次结构无法加速。使用未加速数据集的数据透视表将回退到 `_raw` 数据。
- **如果所加速的根事件数据集或根搜索数据集的初始约束搜索中包含待搜索的索引，则数据模型加速是最有效的方式。**否则，会搜索此数据模型的所有可用索引，而这会浪费时间加速不必要的数据。

请参阅本手册中的[加速数据模型](#)。

要了解有关流命令、生成命令和转换命令的更多信息，请参阅《搜索参考》中的“命令类型”。

启用数据模型加速

如果您的权限足以加速数据模型，请遵循下列步骤：

- 1. 导航到“数据模型”管理页面。
- 2. 找到要加速的数据模型，并打开其加速控制。使用以下各选项中的其中一项。

选项	此选项的其他步骤
选择编辑 > 编辑加速。	无
展开此数据模型的行。	单击加速所对应的添加。

- 3. 选择加速为数据模型启用加速。

编辑加速

数据模型: SF Food Trucks

加速: ☒

加速可提高存储和处理成本。

摘要范围: 3个月

取消

保存

- 4. 根据您计划针对哪个时间范围运行使用数据模型中加速数据集的数据透视表，将摘要范围设置为 1 天、7 天、1 个月、3 个月、1 年或所有时间。

例如，如果您仅计划对最近七天内的时间段运行数据透视表，请选择 7 天。

如果您需要摘要范围与摘要范围字段所提供的有所不同，则您可在 datamodels.conf 中为您的数据模型进行配置。

- 5. 单击保存以保存加速设置。

一旦加速数据模型，“数据模型”管理页面上的模型“闪电”符号就会以黄色点亮。



检查数据模型加速指标

加速数据模型后，即可在“数据模型”管理页面上查看有关模型加速的详细信息。只需展开加速的数据模型所在行，查看加速下面显示的信息。

加速	
重建	更新
编辑	
状态	构建
访问计数	0。上次访问: 1970-01-01
	T01:00:00+01:00
磁盘上的大小	0.00MB
摘要范围	2592000
数据桶	0

- **状态** 指示数据模型的加速摘要是否完成。如果处于构建状态，则将指示摘要已完成的百分比。请记住，很多数据模型摘要都会不断更新为新数据；摘要此刻已“完成”并不意味着后续不会处于“构建”状态。
- **访问计数** 指示数据模型摘要自创建后访问过多少次以及上一次访问的时间。这对于确定哪些数据模型不常用非常有用。由于数据模型加速会使用系统资源，因此，对于不定期访问的数据模型，最好不要进行加速。
- **磁盘上的大小** 指示数据模型的加速摘要占用了多少存储空间。您可以使用此指标与访问计数一起确定哪些摘要不是系统中的必要负载，而应将其删除。如果数据模型的加速摘要占用了磁盘上的大量空间，也可考虑缩小其摘要范围。
- **摘要范围** 以秒为单位显示数据模型的范围，始终与当前时刻相关。为数据模型定义加速时可设置此范围。
- **数据桶** 显示数据模型加速摘要所涵盖的索引数据桶数量。

单击**重建**以从头开始重建摘要。当您怀疑有数据因系统崩溃或类似事故而丢失时，可以重建摘要。Splunk Enterprise 将在您先禁用再重新启用摘要加速（例如，为了编辑数据模型）后自动重建摘要。

单击**更新**可刷新加速摘要详细信息。

单击**编辑**可打开**编辑加速**对话框并更改**摘要范围**或完全禁用数据模型加速。

复制数据模型

数据模型复制这种方法可以快速创建基于现有数据模型的数据模型。之后可通过编辑此数据模型，使其侧重于完全不同的数据集，或者采用不同于原始数据集分隔方式的数据集结构。

步骤

- 1. 使用以下各选项中的其中一项。

选项	此选项的其他步骤
前往“数据模型”管理页面。	找到要复制的数据模型，然后选择 编辑 > 复制 。
打开待复制的数据模型的“数据模型编辑器”。	选择 编辑 > 复制 。

- 2. 在**新标题**中为复制的数据模型输入一个唯一的名称。
- 3. （可选）为新数据模型提供**描述**。
- 4. （可选）如果您的权限允许，选择**复制**可为复制的数据模型赋予与原数据模型同样的权限。
- 5. 单击**复制**创建数据模型副本。

您可以通过“数据模型”管理页面（如本主题所述）和“数据模型编辑器”（如[设计数据模型数据集](#)中所述）编辑复制的数据模型。

上传和下载数据模型

您可以使用上传/下载功能将数据模型从 Splunk 部署中导出，然后将其上传到另一个 Splunk 部署中。您可使用此功能备份重要的数据模型，或通过将它们以电子邮件形式发送到其他 Splunk 用户，与他们共同就数据模型相关内容进行合作。您还可以使用它在 Splunk 的分段和生产实例间移动数据模型。

您可在 Splunk 部署间手动移动数据模型 JSON 文件，但这是一个未受支持的过程，很有可能出现错误。

请参阅[手动数据模型管理](#)。

下载数据模型

从“数据模型编辑器”下载数据模型。您一次只能下载一个数据模型。

步骤

- 1. 在“数据模型编辑器”中打开数据模型。
- 2. 单击右上方的**下载**按钮。

Splunk 会将数据模型的 JSON 文件下载到您指定的下载目录。如果您还没有指定此目录，将显示对话框，要求您指定要将文件保存至其中的目录。



下载的 JSON 文件的名称和数据模型 **ID** 相同。您只需在首次创建数据模型时提供一次 **ID**。与数据模型**标题**不同的是，一旦 **ID** 在模型创建时保存，您不可更改它。

当您在“数据模型编辑器”中查看模型时，您可看到现有数据模型的 **ID**。**ID** 显示在“编辑器”左上角附近，在模型标题下。

当您上传数据模型时，您将有机会为它提供一个新的、与原始数据模型 **ID** 不同的 **ID**。

上传数据模型

从“数据模型”管理页面上载数据模型。您一次只能上传一个数据模型。

Splunk 软件会验证您尝试上传的所有文件。它不会上传包含任何有效 JSON 数据模型代码以外的代码的文件。

步骤

- 1. 导航到“数据模型”管理页面。

2. 单击**上传数据模型**。
 3. 找到要上传的 JSON 文件。
此 ID 字段将填充数据模型的原始 ID。
 4. （可选）将数据模型 ID 修改为一个全新而唯一的值。
请牢记，一旦您将数据模型文件保存到系统中，您将无法更改此 ID。在将数据模型标题保存到系统中后，您仍然可以对其进行编辑。
 5. 提供此数据模型所属的**应用**的名称。
 6. （可选）如果您的操作权限允许，将所上传的数据模型的权限从**专用**改为**在应用中共享**。
 - **在应用中共享**表示此数据模型会共享给此**应用**中的所有用户。
 - 如果您选择**在应用中共享**，则还可通过选择**加速**和**摘要范围**为数据模型启用加速。
 7. 单击**上传**将数据模型上传。
- 所上传的数据模型如果通过验证，会出现在“数据模型”管理页面列表中。

请参阅[关于数据模型权限](#)。

请参阅[启用数据模型加速](#)。

删除数据模型

您可以从“数据模型”管理页面或“数据模型编辑器”删除数据模型。只需单击**编辑**并选择**删除**。

注意：如果您的角色授予您创建数据模型的功能，也应同时授予您删除这些模型的功能。有关更多信息，请参阅[允许角色创建数据模型](#)。

手动管理数据模型

Splunk 不建议您通过手动移动文件或手动对数据模型文件编码来手动管理数据模型。您应尽量在 Splunk Web 中创建和编辑数据模型。在 Splunk Web 中编辑模型时，“数据模型编辑器”将对您的更改进行验证。“数据模型编辑器”无法对手动创建或编辑的模型进行验证。

数据模型以 JSON 文件形式存储在磁盘上。其关联配置存储在 `datamodels.conf` 中，元数据存储在 `local.meta` 中（适用于用户创建的模型）和 `default.meta` 中（适用于随产品一起提供的模型）。

用户创建的模型存储在 `<yourapp>/local/data/models` 中，而随产品一起提供的模型则位于 `<yourapp>/default/data/models` 中。

您可在 Splunk 安装系统间手动移动模型文件，但使用 Splunk Web 中的“数据模型下载/上传”功能（如上所述）要容易得多。如果一定要手动移动模型文件，则移动过程中请注意移动其 `datamodels.conf` 段落和 `local.meta` 元数据。

删除数据模型时也应如此。通常最好通过 Splunk Web 删除模型，这样便可以执行适当的清理工作。

设计数据模型

在 Splunk Web 中，可使用“数据模型编辑器”来设计新**数据模型**和编辑现有模型。本主题介绍如何使用“数据模型编辑器”来进行以下操作：

- 通过向数据模型添加根数据集和子数据集，构建**数据模型数据集**层次结构。
- 定义数据集（通过提供**约束**、搜索字符串或交易定义）。
- 重命名数据集。
- 删除数据集。

您还可使用“数据模型编辑器”创建和编辑数据集**字段**。有关更多信息，请参阅[定义数据集字段](#)。

注意：本主题不会过多地介绍基本数据模型概念。如果您之前没有在 Splunk Enterprise 中使用过数据模型，请参阅主题[关于数据模型](#)。到底什么是数据模型，什么是数据模型数据集，他们的工作方式又是如何，该主题围绕这些问题提供了大量详细背景信息。

有关创建和管理新数据模型的信息，请参阅本手册中的[管理数据模型](#)。除了通过“数据模型”管理页面创建新数据模型外，本主题还将介绍如何管理数据模型权限和加速。

数据模型编辑器

数据模型是按照分层结构排列的数据模型数据集的集合。要设计新数据模型或重新设计现有数据模型，请转到“数据模型编辑器”。在“数据模型编辑器”中，您可以为数据模型创建数据集，定义其约束和**字段**，以及以逻辑数据集层次结构排列模型并维护模型。

Splunk's Internal Server Logs - SAMPLE
internal_server

编辑 下载 数据透视图 文档

数据集
事件
Splunk Server
Scheduler
Alerts
Scheduled Reports
Summary Indexing Searches
Acceleration
Data Model Acceleration
Report Acceleration
Licenser
Daily Usage Summary
Daily Slave Warning Summary
Quota Usage
Pool Warnings
Performance and System Data
Pipeline
Queue

Quota Usage
quota

重命名 删除

约束
index=internal source=*scheduler.log* OR source=*metrics.log* OR source=*splunkd.log* OR source=*license_usage.log* OR source=*splunkd_access.log* 继承
source=*license_usage.log* 继承
type=Usage 约束 编辑
批量编辑 添加字段
继承
_time 时间

alert actions 字符串 已隐藏 覆盖

app 字符串 已隐藏 覆盖

clientip 字符串 已隐藏 覆盖

cpu seconds 数字 已隐藏 覆盖

current size (KB) 数字 已隐藏 覆盖

executes 数字 已隐藏 覆盖

historical searches 数字 已隐藏 覆盖

Host 字符串 已隐藏 覆盖

host 字符串 已隐藏 覆盖

Index 字符串 已隐藏 覆盖

只有在权限允许的情况下才能编辑特定数据模型。

导航到“数据模型编辑器”

要打开已有数据模型的“数据模型编辑器”，选择以下选项之一。

选项	此选项的其他步骤
导航到设置 > 数据模型。	找到要编辑的数据模型，并选择编辑 > 编辑数据集。
打开“数据集”列表页面	找到要编辑的数据模型数据集，并选择管理 > 编辑数据模型。
导航到数据模型数据集的“数据透视表编辑器”。	单击编辑数据集。

为数据模型添加根事件数据集

数据模型主要由构建在根事件数据集上的数据集层次结构组成。每个根事件数据集代表由约束定义的一组数据：筛选与数据集不相关事件的简单搜索。约束类似于搜索在添加管道符和其他搜索命令之前的第一部分。

根事件数据集的约束通常用于返回相对宽泛的数据。在大型数据集中，将子事件数据集与根事件数据集关联后将为您带来一些优势，由于每个子事件数据集都会向继承自其上级数据集的约束添加一个附加约束，因此缩小了数据集所表示的范围。

有关约束如何缩小数据集层次结构中数据集范围的更多信息，请参阅数据集约束。

要为数据模型添加根事件数据集，在“数据模型编辑器”中单击添加数据集，然后选择根事件。此时将进入“添加事件数据集”页面。

添加事件数据集
数据集模型: Buttercup Games
文档

数据集名称
HTTP request

约束
sourcetype=access.* OR sourcetype=iiis*
示例
uri="*.php*" OR uri="*.py*" NOT (referrer=null OR referrer="")

数据集 ID
HTTP_request
只能包含字母、数字和下划线。

取消 预览 保存

✓ 1,000 个事件 (17/01/03 3:49:31.000 之前)

每页 20 个 < 上一个 1 2 3 4 5 6 7 8 9 ... 下一步 >

示例 1,000 个事件

事件

91.205.189.15 - - [23/Nov/2016:18:22:16] "GET /oldlink?itemId=EST-14&JSESSIONID=SD6SL7FF7ADFF53113 HTTP 1.1" 200 1665 "http://www.buttercupgames.com/oldlink?itemId=EST-14" "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 159
91.205.189.15 - - [23/Nov/2016:18:22:15] "GET /category.screen?categoryID=SHOOTER&JSESSIONID=SD6SL7FF7ADFF53113 HTTP 1.1" 200 1369 "http://www.google.com" "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 779
182.236.164.11 - - [23/Nov/2016:18:20:56] "GET /cart.do?action=addtocart&itemId=EST-15&productId=BS-AG-G09&JSESSIONID=SD6SL8FF10ADFF53101 HTTP 1.1" 200 2252 "http://www.buttercupgames.com/oldlink?itemId=EST-15" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_4) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 506
182.236.164.11 - - [23/Nov/2016:18:20:55] "POST /oldlink?itemId=EST-18&JSESSIONID=SD6SL8FF10ADFF53101 HTTP 1.1" 408 893 "http://www.buttercupgames.com/product.screen?productId=SF-BVS-601" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_4) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 134

124

为根事件数据集指定**数据集名称**、**数据集 ID** 以及一个或多个**约束**。

数据集名称字段可接受除星号外的任何字符。它还可接受字符间的空格。您将在此看到“选择数据集”页面和其他列出数据模型数据集的位置。

数据集 ID 必须是数据集的唯一标识符。其中不得包含空格或任何不是字母数字、下划线或连字符（a-z、A-Z、0-9、_ 或 -）的字符。同时不允许在两个字符之间使用空格。一旦保存了**数据集 ID** 值，便无法再进行更改。

为根事件数据集提供**约束**后，即可单击**预览测试**所提供的约束是否返回您希望的事件类型。

为数据模型添加根搜索数据集

根搜索数据集可用于创建基本数据集为任意搜索结果的数据集层次结构。您可以在搜索字符串中使用定义了根搜索数据集的任意 SPL。

使用转换命令的根搜索数据集无法加速。转换搜索使用转换命令来定义包括一个或多个字段在整个数据集内聚合的基本数据集。

要为数据模型添加根搜索数据集，在“数据模型编辑器”中单击**添加数据集**，然后选择**根搜索**。此时将进入“添加搜索数据集”页面。

添加基本搜索
数据模型: Buttercup Games

数据集名称: user 数据集 ID: user
只能包含字母、数字和下划线。

搜索字符串
time=* host=* source=* uri=* status <600 clientip=* useragent=* (sourcetype = access* OR source = 8.log) | eval userid=clientip | stats first(_time) as earliest, last(_time) as latest, list(uri_path) as uri_list by userid

取消 保存

✓ 237,192 个事件 (17/01/03 3:57:44.000 之前) 每页 20 个 < 上一个 1 2 3 4 5 6 7 8 9 10 下一步 >

示例 1,000 个事件 >

userid	earliest	latest	uri_list
107.3.146.207	1479932715	1465348937	/cart/success.do /cart.do /cart.do /product.screen /oldlink /oldlink /cart.do /oldlink /cart.do /oldlink

为根搜索数据集指定**数据集名称**、**数据集 ID** 以及搜索字符串。要在页面底部部分预览搜索结果，请单击放大图标运行搜索，或者只需在光标位于搜索栏中时敲击键盘上的回车键。

数据集名称字段可接受除星号外的任何字符。它还可接受字符间的空格。您将在此看到“选择数据集”页面和其他列出数据模型数据集的位置。

数据集 ID 必须是数据集的唯一标识符。其中不得包含空格或任何不是字母数字、下划线或连字符（a-z、A-Z、0-9、_ 或 -）的字符。同时不允许在两个字符之间使用空格。一旦保存了**数据集 ID** 值，便无法再进行更改。

有关设计搜索字符串的更多信息，请参阅《搜索手册》。

如果搜索是简单的 `transaction` 搜索，则不要为此搜索创建根搜索数据集。将其设置为根交易数据集。

在根搜索数据集中使用转换搜索

对于不直接映射到 Splunk 事件的搜索，您可以为其创建根搜索数据集，只要您明白他们无法加速。即，涉及到非事件格式的输入或输出的搜索。其中包括如下搜索：

- 利用 `stats`、`chart` 和 `timechart` 等**转换命令**。转换命令将其返回的数据组织置于表格而非事件列表中。
- 使用其他不返回事件的命令。
- 使用 `lookup` 以外的命令从外部非 Splunk 来源获取数据。此数据无法确保具有 `host`、`source`、`sourcetype` 或 `_time` 等默认字段，因此可能无法映射到事件。使用 `inputcsv` 命令从外部 `.csv` 文件获取信息便属于这种情况。

为数据模型添加根交易数据集

根交易数据集可用于创建数据集层次结构，且该结构基于由**交易**事件组成的数据集。交易事件实际上是一段时间跨度内概念相关事件的集合，例如与单个客户酒店预订会话相关的所有网站访问事件，或者与防火墙入侵事件相关的所有事件。定义根交易数据集层次结构时，即意味着定义提取一组交易事件的交易。

如果不熟悉交易的工作方式，请阅读交易和 `transaction` 命令的相关信息。从《搜索手册》中的**关于交易**开始学习。有关 `transaction` 命令的详细信息，请参阅《搜索参考》手册中该命令的相关条目。

根交易数据集及其子数据集不会受益于数据模型加速。

要为数据模型添加根交易数据集，在“数据模型编辑器”中单击**添加数据集**，然后选择**根交易**。此时将进入“添加交易数据集”页面。

添加交易数据集

数据模型: Buttercup Games

您必须指定至少一个可选字段。

数据集名称

Web Session

数据集 ID

Web_Session

只能包含字母、数字和下划线。

根数据集

HTTP Requests

持续时间

最大暂停时间:

秒

最大跨度:

1

分钟

分组依据

client IP

取消

预览

保存

✓ 1,000 个事件 (17/01/03 3:50:34 000 之前)

每页 20 个

← 上一个

1

2

3

4

5

6

7

8

9

...

下一步 →

示例 1,000 个事件

事件

182.236.164.11 - - [23/Nov/2016:18:20:54] "POST /cart.do?action=purchase&itemId=EST-6&JSESSIONID=SD6SL8FF10ADFF53101 HTTP 1.1" 200 1803 "http://www.buttercupgames.com/cart.do?action=addtocart&itemId=EST-6&categoryId=ARCADE&productId=MB-AG-G07" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_4) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 524

182.236.164.11 - - [23/Nov/2016:18:20:54] "POST /cart/success.do?JSESSIONID=SD6SL8FF10ADFF53101 HTTP 1.1" 200 356 "http://www.buttercupgames.com/cart.do?action=purchase&itemId=EST-6" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_4) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 220

198.35.1.75 - - [23/Nov/2016:18:18:57] "POST /cart.do?action=purchase&itemId=EST-27&JSESSIONID=SD10SL2FF4ADFF53099 HTTP 1.1" 200 3577 "http://www.buttercupgames.com/cart.do?action=addtocart&itemId=EST-27&categoryId=TEE&productId=MB-AG-T01" "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 827

198.35.1.75 - - [23/Nov/2016:18:18:57] "POST /cart/success.do?JSESSIONID=SD10SL2FF4ADFF53099 HTTP 1.1" 200 613 "http://www.buttercupgames.com/cart.do?action=purchase&itemId=EST-27" "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 328

198.35.1.75 - - [23/Nov/2016:18:18:58] "POST /cart.do?action=purchase&itemId=EST-16&JSESSIONID=SD10SL2FF4ADFF53099 HTTP 1.1" 200 821 "http://www.buttercupgames.com/cart.do?action=addtocart&itemId=EST-16&categoryId=SIMULATION&productId=SC-MG-G10" "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 178

198.35.1.75 - - [23/Nov/2016:18:18:59] "POST /cart/success.do?JSESSIONID=SD10SL2FF4ADFF53099 HTTP 1.1" 200 2568 "http://www.buttercupgames.com/cart.do?action=purchase&itemId=EST-16" "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 386

根交易数据集的定义要求指定**数据集名称**和**数据集 ID**以及至少一个**组合数据集**。**分组依据**、**最大暂停时间**和**最大跨度**字段为可选字段，但在定义三个字段中至少一个字段前，交易定义都是不完整的。

数据集名称字段可接受除星号外的任何字符。它还可接受字符间的空格。您将在此看到“选择数据集”页面和其他列出数据模型数据集的位置。

数据集 ID必须是数据集的唯一标识符。其中不得包含空格或任何不是字母数字、下划线或连字符（a-z、A-Z、0-9、_ 或 -）的字符。同时不允许在两个字符之间使用空格。一旦保存了**数据集 ID**值，便无法再进行更改。

所有根交易数据集的定义都要求指定一个或多个**组合数据集名称**，以定义交易数据集要用来派生其交易的数据池。不过，在**组合数据集**下添加哪些内容具有一些限制。**组合数据集**可包含以下三个选项之一：

- 一个或多个事件数据集（根事件数据集或子事件数据集）
- 一个交易数据集（根数据集或子数据集）
- 一个搜索数据集（根数据集或子数据集）

此外，您只能处理属于当前选定数据模型中的数据集。

如果您熟悉 `transaction` 命令的工作方式，则可将**组合数据集**视为我们在交易命令之前所提供的搜索字符串部分。以上文屏幕截图为例，我们已经将 *Apache 访问搜索* 数据集添加到根交易数据集“Web 会话”的定义中。*Apache 访问搜索* 表示一组成功的 webserver 访问事件，它的两个约束是 `status < 600` 和 `sourcetype = access_* OR source = *.log`。因此，此根交易数据集所表示的交易搜索的开头部分为：

```
status < 600 sourcetype=access_* OR source=*.log | transaction...
```

现在我们只需要定义其余的 `transaction` 参数。

为数据模型添加子数据集

您可以为根数据集和其他子数据集添加子数据集。子数据集继承属于其父数据集的所有约束和字段。一个数据集可以与多个子数据集关联。

当定义新子数据集时，您将指定一个或多个额外约束，以进一步专注于子数据集所代表的数据集。例如，如果“Web 智能”数据模型有一个名称为“HTTP 请求”的根事件数据集，并且该请求用于捕获所有 webserver 访问事件，则可为该对象指定三个子事件数据集：“HTTP 成功”、“HTTP 错误”和“HTTP 重定向”。每个子事件数据集都专注于“HTTP 请求”数据集的某个特定子集：

- 子事件数据集 *HTTP 成功* 使用附加约束 `status = 2*` 专注于成功的 webserver 访问事件。
- *HTTP 错误* 使用附加约束 `status = 4*` 专注于失败的 webserver 访问事件。
- *HTTP 重定向* 使用附加约束 `status = 3*` 专注于重定向的 webserver 访问事件。

可以选择添加除继承自父数据集字段以外的字段。有关字段定义的更多信息，请参阅[使用“数据模型编辑器”管理数据集字段](#)。

要为数据模型添加子数据集，在左侧数据集层次结构中选择父数据集，然后在“数据模型编辑器”中单击**添加数据**

集并选择子项。此时将进入“添加子数据集”页面。

添加子数据集

数据模型: Buttercup Games

数据集名称

Client Errors

数据集 ID

Client_Errors

其他约束

status = 4*

继承自

HTTP Requests

取消

预览

保存

✓ 888 个事件 (17/01/03 3:54:57.000 之前)

每页 20 个 < 上一个 1 2 3 4 5 6 7 8 9 ... 下一步 >

示例 1,000 个事件 >

事件
125.89.78.6 - - [23/Nov/2016:17:42:06] "GET /rush/signals.zip?JSESSIONID=SD10SL8FF3ADFF52952 HTTP/1.1" 404 1252 "http://www.buttercupgames.com/cart.do?action=purchase&itemId=EST-7" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_4) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 594
201.28.109.162 - - [23/Nov/2016:16:16:08] "GET /oldlink?itemId=EST-15&JSESSIONID=SD6SL4FF10ADFF52464 HTTP/1.1" 406 2327 "http://www.buttercupgames.com/cart.do?action=purchase&itemId=EST-15" "Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.1; WOW64; Trident/5.0; BOIE9;ENUS)" 464
87.194.216.51 - - [23/Nov/2016:15:51:14] "GET /cart.do?action=purchase&itemId=EST-12&JSESSIONID=SD10SL9FF8ADFF52336 HTTP/1.1" 408 3014 "http://www.buttercupgames.com/category.screen?categoryId=NULL" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" 800
203.172.197.2 - - [23/Nov/2016:15:18:56] "GET /cart.do?action=purchase&itemId=EST-19&JSESSIONID=SD5SL6FF6ADFF52150 HTTP/1.1" 408 2312 "http://www.buttercupgames.com/cart.do?action=purchase&itemId=EST-19" "Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 5.1; Trident/4.0; .NET CLR 2.0.50727; MS-RTC LM 8; InfoPath.2)" 397

为子数据集指定一个数据集名称和数据集 ID。

数据集名称字段可接受除星号外的任何字符。它还可接受字符间的空格。您将在此看到“选择数据集”页面和其他列出数据模型数据集的位置。

数据集 ID 必须是数据集的唯一标识符。其中不得包含空格或任何不是字母数字、下划线或连字符（a-z、A-Z、0-9、_ 或 -）的字符。同时不允许在两个字符之间使用空格。在保存了数据集 ID 值之后，便无法再进行更改。

为子数据集定义约束后，即可单击预览测试所提供的约束是否返回您希望的事件类型。

数据模型设计的一些最佳实践

确定适用的数据模型设计之前可以先进行试验并查看是否发生错误。以下提示有助于实现良好的开端。

- 尽可能使用根事件数据集和仅使用流命令的根搜索数据集来利用数据模型加速（并受益于其易于优化的特性）。
- 为要加速的根事件数据集定义约束或为要加速的根搜索数据集定义搜索时，请指定要从中进行选择的索引。当数据模型针对特定索引或索引集进行搜索时，数据模型加速的有效性和准确性都会提升。如果没有指定索引，数据模型将针对所有可用索引进行搜索。
- 尽可能降低数据集层次结构的深度。基于约束的筛选功能在对象树中靠下的位置效果不佳。
- 使用字段标志选择性地显示每个数据集的各小组字段。您可以显示和隐藏不同数据集的不同字段。子字段所显示的字段集可与其父字段所显示的字段集完全不同。数据透视表用户将从这种选择中受益，因为开始创建数据透视表图表或表格时无需处理大量眼花缭乱的字段，而是只会看到那些在所选数据集上下文中有用的字段。
- 将现有仪表板和搜索反向设计成数据模型。这样可以快速掌握数据模型。使用源自数据透视表的面板构建的仪表板更容易维护。
- 设计新数据模型时，应事先了解数据透视表用户要通过该模型实现哪些功能。由此展开反向设计。模型结构应由用户需求和预期决定。

定义数据集字段

在本主题中，我们会讨论添加和编辑数据模型数据集的字段。数据集字段提供了一组数据透视表用户在定义和生成数据透视表报表时使用的字段。

字段可存在于数据集中，或可通过使用查找和 eval 表达式派生和添加到数据集。

您可使用“数据模型编辑器”创建和管理数据集字段。它使您能够：

- 创建新字段。
- 更新或删除现有的非继承字段。
- 覆盖继承字段的某些设置。

注意：本主题不会详细介绍数据集字段所涉及的相关概念。如果您尚没有使用过数据模型字段，则应查阅主题[关于数据模型](#)。

您也可使用“数据模型编辑器”构建数据模型数据集层次结构、定义数据集（通过提供约束、搜索字符串或交易定义）、重命名数据集和删除数据集。有关使用“数据模型编辑器”执行这些任务的更多信息，请参阅[设计数据模型数据集](#)。

有关创建和管理新数据模型的信息，请参阅[管理数据模型](#)。除了通过“数据模型”管理页面创建新数据模型外，本主题还会介绍如何管理数据模型权限和加速。

数据集字段概览

数据集字段是供数据透视表用户用于定义和生成数据透视表报表的一组字段。

您可在数据模型中为数据集定义五种不同类型的字段：

- **自动提取：**这些字段表示在索引时间和搜索时间期间提取的字段。自动提取的字段只能添加到根数据集。子数据集只能继承自动提取的字段，而无法添加自己的新自动提取字段。自动提取的字段可包括：
 - 自动提取的字段，比如 `uri` 或 `version`。这包括为 CSV、IIS 和 JSON 文件通过结构化数据输入建立索引的字段。
 - 您已在“设置”中定义或在 `props.conf` 中配置的**字段提取、查找或已计算字段**。
 - 您已手动添加到字段的字段，因为它们虽然当前未在数据集中，但将来应该在其中。可以包含通过生成命令（如 `inputcsv` 或 `dbinspect`）添加到数据集中的字段。
- **Eval 表达式：**一个源自字段定义中所输入 `eval` 表达式的字段。Eval 表达式通常涉及一个或多个提取的字段。
- **查找：**一个利用在字段定义中配置的**查找**添加到数据集中事件的字段。定义查找字段时，可使用已经在**设置中定义**的任意查找，并将其与其他任意已经关联到同一数据集的字段相关联。
- **正则表达式：**使用您在字段定义中所提供的正则表达式从数据集事件数据中提取的字段。
- **GeoIP：**一种特定的查找类型，可将纬度、经度、国家/地区和城市等地理字段添加到数据集中具有有效 IP 地址字段的事件。这对于地图相关的可视化非常有用。

有关数据集字段的更多概况，例如什么是数据集字段、数据集字段如何工作以及为何使用数据集字段，请参阅[关于数据模型](#)中的相关小节。

字段类别

“数据模型编辑器”将字段分为三类别：

- **已继承：**所有数据集具有至少几个已继承字段。子数据集继承属于其父数据集的所有字段。根事件、搜索和交易数据集包含一组默认的已继承字段。
- **已提取：**所有已添加到数据集的自动提取字段显示在此类别中。
- **已计算：**所有通过计算或查找派生的字段显示在此类别中。当您为 Eval 表达式、正则表达式、查找和地理 IP 字段类型添加到数据集时，它们会显示于此字段类别。

字段顺序和字段链接

“数据模型编辑器”可用于重新排列已计算字段的顺序。当您有一组必须要按特定顺序处理的字段时，这非常有用，因为字段都会按列表从上到下的降序顺序进行处理。

例如，您可设计使用两个自动提取字段值的 Eval 表达式字段。已提取字段会位于已计算字段之上，以便在此情况下字段可按正确顺序进行处理而无需您进行任何操作。但您可能还会将 `eval` 表达式字段用作查找字段的输入。因为“数据模型编辑器”将 Eval 表达式字段和查找字段分类为已计算字段，所以您需要确保将已计算字段列表排序，使 Eval 表达式字段显示在查找字段之上。

因此，这些字段的顺序是：

- 自动提取字段 1
- 自动提取字段 2
- Eval 表达式字段（通过两个自动提取字段值计算一个字段）
- 查找字段（使用 Eval 表达式字段作为输入字段）

将字段标记为隐藏或必需

默认情况下，所有数据集字段都为**显示**和**可选**。

- 当**显示**字段在其所属的数据集的上下文中时，此字段可见且对数据透视表用户可用。例如，假设 HTTP Requests 数据集的 `uri` 字段标记为显示。当用户进入数据透视表并选择 HTTP Requests 数据集时，他们在定义数据透视表报表时可使用 `uri` 字段。
- **可选**字段不需要在其数据集所表示的数据集中的每个事件中显示。这意味着，数据集中可能有很多未包含此字段的事件。

您可将这些设置分别更改为**隐藏**和**必需**。执行此操作时，字段将在数据集字段列表中标记为**隐藏**和/或**必需**。

- 当数据透视表用户在数据透视表上下文中选择数据集时，**隐藏**字段不会显示。他们将无法将它用于定义数据透视表报表。
 - 此设置可让您针对数据模型中的每个数据集显示不同的字段子集，即使所有数据集都从同一个父数据集继承了相同的字段集。这样有助于确保数据透视表用户仅接触数据集所表示的数据集上下文中有意义的字段。
 - 您可隐藏仅能被添加到数据集的字段，因为它们用于定义其他字段（请参阅上文中的“字段顺序和字段链接”）。数据透视表用户可能不需要接触字段链中的第一个字段。
- **必需**字段**必须**在数据集所表示的每个事件中显示。这会过滤出所有不包含此字段的事件。从效果上看，这是另一种在数据集所关联的所有正式约束之上的**约束**。

数据模型中每个数据集的字段设置都是特定的。这意味着您可以在父数据集中将 `ip_address` 字段设置为必需，但在该父数据集的子数据集中仍设置为可选。即使数据模型中所有数据集都具有相同的字段（这意味着字段在顶层根数据集中设置，然后仅通过继承传递到层次结构中的其他所有数据集），该数据模型中各数据集之间标记为隐藏或必需的字段仍然可以有所不同。

注意：为数据模型的不同数据集中的相同字段提供不同的“显示/隐藏”和“可选/必需”设置的功能有一个例外。**您无法更新已继承字段（在其第一次出现的父数据集中被分类为“已计算”字段）的设置。**对于此类字段，您只能通过更新其父数据集的字段来更改设置。您的更改将会复制到该父数据集所产生的子数据集中。

当您第一次定义已提取和已计算字段时，可为它们设置这些值。您还可在定义字段名称或类型后对它们进行编辑。

1. 针对已继承类别中的字段单击**覆盖**或针对已提取和已计算类别中的字段单击**编辑**。
2. 将**标记**字段的值更改为相应的值。
3. 单击**保存**以保存更改。

使用**批量编辑**列表，您可一次更改多个字段的“显示/隐藏”和“可选/必需”值。

1. 选择您要编辑的字段。
2. 单击**批量编辑**，并选择 *可选*、*必需*、*隐藏*或*显示*。

如果您选择*必需*或*隐藏*，则相应字段会进行更新以显示所选字段的所选状态。您无法更新已继承字段（在其第一次出现的父数据集中被分类为“已计算”字段）的值。有关更多信息，请参阅上文中的**备注**。

输入或更新字段名称和类型

“数据模型编辑器”可用于为**已提取**和**已计算**类别中的字段指定一个显示**名称**。即使字段已自动分配了**类型**值，您仍可通过编辑器为此类字段设置新的**类型**。

Splunk 软件自动为自动提取的字段分配类型。如果为自动提取字段分配的**类型**值不正确，则您将其修改为正确的值。例如，Splunk 软件可能根据自动提取字段的可用值将该字段归为**数字**类型的字段，而您所了解的实际情况却是**字符串**类型。如果确实是这样，则可以将**类型**值更改为**字符串**。

更改自动提取字段的显示**名称**并不会更改其相关联字段在索引中的命名方式，而是仅在此数据模型的上下文中重命名该字段。

1. 为待更新其**名称**或**类型**的字段单击**编辑**。
2. **更新名称**或**更改类型**。**名称**值不可以包含星号字符。
3. 单击**保存**以保存更改。

使用**批量编辑**列表为多个字段指定相同的**类型**值。

1. 选择您要编辑的字段。
2. 单击**批量编辑**并选择 *布尔*、*IPv4*、*数字*或*字符串*。
您不可为已继承字段更改**类型**值。如果您选择任何已继承字段，则**批量编辑**列表中的**类型**值将不可用。

所有已选择字段应将其**类型**值更新为您选择的值。

添加自动提取字段

您可向数据模型中的任何根数据集添加自动提取字段。

添加自动提取的字段

示例: 前 1,000 个事件 | ✓ 1,000 个事件 (14/07/28 10:23:06.000 之前) | 字段缺失? 按名称添加

字段	重命名	类型
<input type="checkbox"/> ISBN	ISBN	数字 可选
<input checked="" type="checkbox"/> JSESSIONID Example values: SD10SL6FF4ADFF3 SD1SL1FF9ADFF10 SD1SL4FF4ADFF2 SD1SL6FF6ADFF4 SD1SL8FF9ADFF1 SD3SL10FF2ADFF7 SD3SL7FF8ADFF10 SD3SL8FF6ADFF2 SD4SL8FF9ADFF10 SD4SL8FF9ADFF9	JSession ID	字符串 可选
<input type="checkbox"/> action		
<input type="checkbox"/> app		
<input checked="" type="checkbox"/> categoryId	Category ID	字符串 可选
<input type="checkbox"/> cookie		
<input type="checkbox"/> date_hour		
<input type="checkbox"/> date_minute		

取消 保存

1. 在“数据模型编辑器”中，打开要向其添加自动提取字段的根数据集。
2. 单击**添加字段**并选择**自动提取**以定义自动提取字段。

将显示“添加自动提取的字段”的对话框。它包括一个字段列表，其中列出了可添加到数据模型数据集的字段。

3. 通过勾选相应复选框的方式选择您要向数据模型添加的字段。

您可选择标头中的复选框以选择列表中的所有字段。

如果您查看列表且未发现您需要的字段，尝试更改事件示例的大小（在默认情况下，它设置为**前 1,000 个事件**）。更大的事件示例可能包含未在前一千个事件中显示的罕见字段。例如，您可选择**一个示例大小，如前 10,000 个事件或过去 7 天**。

4. （可选）**重命名**自动提取字段。

如果您使用**重命名**，不要在新字段名称中包含星号字符。

5. （可选）纠正自动提取字段**类型**。
6. （可选）根据需要更新自动提取字段的**状态**（**可选**、**必需**、**隐藏**或**隐藏和必需**）。
7. 单击**保存**向根数据集添加所选的字段。

注意：您无法向子数据集添加自动提取字段。子数据集从其数据集层次结构顶部的根数据集继承自动提取字段。

“添加自动提取字段”对话框显示的字段列表包括：

- 自动提取的字段，比如 `uri` 或 `version`。这包括通过结构化数据输入建立索引的字段，如从索引 CSV 文件标头提取的字段。
- 您已在“设置”中定义或在 `props.conf` 中配置的**字段提取**、**查找**或**已计算字段**。

展开字段所在字段行，查看字段的前十个示例值。

手动向自动提取字段集添加字段

构建数据模型时，您可能发现您缺少某些自动提取字段。由于多种原因，它们可能会缺少。例如：

- 您可能先构建数据模型，之后才对组成数据集的数据建立索引。
- 您正在为数据建立索引，但某些您要查看的罕见字段最终未被索引到。
- 您正在使用生成搜索命令，如 `inputcsv`。该命令会添加此列表中未显示的字段。

您可以手动将自动提取字段添加到根数据集中。

注意：在手动添加字段前，尝试增加事件示例的大小（正如以上过程所述），以添加前一千个事件中未找到的罕见字段。

1. 单击“添加自动提取字段”对话框右上角的**按名称添加**。

这会向字段表添加行。注意，在本主题顶部的示例中，已为手动添加的 ISBN 字段添加行。

2. 在该行中，手动为自动提取字段确定**字段名称**、**类型**和**状态**。
3. 再次单击**按名称添加**以添加其他字段行。
4. 单击已添加行右上角的 **X** 以删除它。
5. 单击**保存**以保存更改。

添加到表格中的所有字段（与任何选定的自动提取字段一起）都作为“已提取”类别中的“已提取”添加到根数据集中。

添加 eval 表达式字段

您可向数据模型中的任何数据集添加 eval 表达式字段。使用 eval 表达式创建字段并将其添加到事件中，方式与处理**已计算字段**的方式类似。

添加具有 Eval 表达式的字段

数据模型: Buttercup Games 数据集: Purchase Requests > Successful Purchases

Eval 表达式

字段

case((id LIKE "DM.%") OR savedsearch_name LIKE "ACCELERATE_DM%"), "dm_acceleration", search_id LIKE "scheduled%", "scheduled", search_id LIKE "rt%", "realtime", search_id LIKE "subsearch%", "subsearch", (search_id LIKE "SummaryDirector%" OR search_id LIKE "summarize_SummaryDirector%"), "summary_director", 1=1, "adhoc")

示例
case(error == 404, "Not found", error == 500, "Internal Server Error")
if(cidrmatch("192.0.0.0/16", clientip), "local", "other")

了解更多信息

字段名称

显示名称

类型

标志

search_type

search type

字符串

可选

取消

预览

保存

事件

值

✓ 1,000 个事件 (17/01/03 3:53:21.000 之前)

每页 10 个 < 上一个 1 2 3 4 5 6 7 8 9 ... 下一步 >

示例 1,000 个事件

_time	search_type	host	source	sourcetype	JSESSIONID	action	apiEndTime	apiStartTime	autojoin	bu
2016/11/23 18:20:54	adhoc	www1	tutorialdata.zip./www1/access log	access_combined_wcookie	SD6SL8FF10ADFF53101	purchase				
2016/11/23 18:20:54	adhoc	www1	tutorialdata.zip./www1/access log	access_combined_wcookie	SD6SL8FF10ADFF53101	purchase				
2016/11/23 18:18:59	adhoc	www1	tutorialdata.zip./www1/access log	access_combined_wcookie	SD10SL2FF4ADFF53099	purchase				
2016/11/23 18:18:58	adhoc	www1	tutorialdata.zip./www1/access log	access_combined_wcookie	SD10SL2FF4ADFF53099	purchase				
2016/11/23 18:18:57	adhoc	www1	tutorialdata.zip./www1/access log	access_combined_wcookie	SD10SL2FF4ADFF53099	purchase				

1. 在“数据模型编辑器”中，打开要向其添加字段的数据集。
2. 单击**添加字段**。选择 *Eval 表达式* 定义 eval 表达式字段。

随即显示“使用 Eval 表达式添加字段”对话框。

3. 输入用来定义字段值的 **Eval 表达式**。

Eval 表达式文本区域应仅包含 `<eval-expression>` 部分（属于 eval 语法）。无需输入“搜索”中使用的所有语法 (`eval <eval-field>=<eval-expression>`)。

4. 在**字段**下，输入**字段名称**和**显示名称**。

字段名称是数据集中的字段名称。**显示名称**是数据透视表用户创建数据透视表时所见的字段名称。**注意**：**字段名称**不可包含空格、单引号、双引号、大括号或星号。**字段显示名称**不可包含星号。

5. 定义**字段类型**并设置其**标记**。

有关**标记**值的更多信息，请参阅**定义数据集字段**中介绍将字段标记为隐藏或必需的小节内容

6. （可选）单击**预览**以确认 eval 表达式按预期运行。

您应看到以表格形式列出的事件，其中以列的形式包含新的 eval 字段。例如，如果您正在使用基于事件的数据集，并添加了名为 **gb** 的 eval 字段，则预览事件表格应在第一列 (**_time**) 右边显示标有 **gb** 的列。

预览窗格有两个选项卡。**事件**是默认选项卡。它以表格形式显示事件。新 eval 字段应显示在第一列 (**_time** 列) 的右边。

如果您在此列中未看到值，或您看到相同的值在列表顶部的事件中重复，则这可能意味着，更多的值在示例的更后面显示。单击**值**选项卡，以查看 eval 字段值在所选事件示例中的分布情况。您还可更改**示例值**以增加预览示例中事件的数量 - 这有时会特别显示由 eval 表达式创建的罕见字段值。

在以下示例中，当**示例**从**前 1,000 个事件**扩展到**前 10,000 个事件**时，三个实时搜索仅在值分布情况中显示。

事件

值

✓ 10,000 个事件 (14/07/28 21:36:19.000 之前)

示例 前 10,000 个事件

每页 20 个

值	计数	%
scheduled	6390	63.900
adhoc	2291	22.910
subsearch	664	6.640
summary_director	652	6.520
realtime	3	0.030

7. 单击**保存**以保存您的更改并返回到“数据模型编辑器”。

有关 `eval` 命令及 `eval` 表达式格式的更多信息，请参阅 `eval` 页面以及《搜索参考》手册中的“评估函数”主题。

`Eval` 表达式可使用已定义或已计算的字段，这意味着您可将字段链接在一起。字段会以其在列表中出现的顺序从上到下的进行处理。这意味着，在 `eval` 表达式中，您必须将前提字段放置在使用这些字段的 `eval` 表达式字段之上。换言之，如果计算 B 取决于另一个计算 A，则应在字段顺序中确保计算 A 位于计算 B 之前。有关更多信息，请参阅[定义数据集字段](#)中有关字段顺序和链接的小节内容。

您可以在 `eval` 表达式字段定义中使用任何类型的字段。例如，在 `eval` 表达式中，您可以创建一个使用自动提取字段和另一个 `eval` 表达式字段的 `eval` 表达式字段。只要这些字段列出在当前创建的字段之前便有效。

在创建 `eval` 表达式字段且此字段在其定义中使用了其他字段的值时，可以选择性地将这些字段的**标志**设置为**隐藏**来“隐藏”这些字段）。这确保数据透视表用户只能使用最后的 `eval` 表达式值。

添加查找字段

您可向数据模型中的所有数据集添加查找字段。

要创建查找字段，在**设置 > 查找 > 查找定义**中必须至少定义了一个**查找定义**。查找定义说明了查找表的位置和访问方式，是读取上载的 CSV 文件还是通过 Python 脚本连接到外部查找表。一旦查找定义就位，Splunk 软件可以将所选字段的值与查找表中字段的值进行匹配，然后返回相应的字段/值组合并作为查找字段应用到您的数据集中。

注意：查找字段中使用的任何查找表文件和查找定义所具有的权限必须与数据模型相同。如果数据模型的权限为全局（即，共享给“所有应用”），而查找表文件或定义为专用，则可能破坏字段。（通常，数据模型及其关联的查找表文件和定义都应全局共享给所有应用。）

关于创建查找定义（以及上载 CSV 文件）的更多信息，请参阅[使用字段查找将信息添加到事件中](#)。

1. 在“数据模型编辑器”中，打开要向其添加查找字段的数据集。
2. 单击**添加字段**，并选择**查找**。

此时将进入“使用查找添加字段”页面。

3. 在**查找表**下选择要与输入字段匹配的查找表。

查找表列表中的所有值都是之前在“设置”中定义的**查找定义**。

当您选择有效的查找表时，会显示和填充页面的**输入**和**输出**部分。**输出**部分应显示已选择**查找表**中所有列的列表。

4. 在**输入**下，定义查找输入字段。从正在编辑的数据集中，选择**查找中的字段**（您已选择的**查找表**中的字段）和相应的**字段**。

输入查找表字段/值组合是一个密钥，用于选择查找表中的行。对于此输入密钥选择的每行，您可从该行导入输出字段值并将其添加到匹配事件。

例如，您的数据集在查找表中有一个名为 `productId` 的字段，且该字段与数据集事件数据中的自动提取字段 `Product ID` 相匹配。此查找表字段和数据集字段应具有相同（或相似）的值集。换句话说，如果您在查找表中有一个行，此行中的 `productId` 值为 `PD3Z002`，则数据集中应该包含 `Product ID = PD3Z002` 的事件。将使用 `productId` 值为 `PD3Z002` 的行的输出字段/值组合来更新这些匹配事件。有关此过程的详细分步说明，请参阅下文中的“查找字段设置示例”。

在一个特定输入密钥匹配多个查找表行的情况下，仅会返回第一行匹配的字段值。要缩小匹配的行集的范围，您可选择定义输入字段的多个对。要选择一行，所有这些输入密钥都必须匹配。当您具有多个输入时，您不可重复使用**查找中的字段**值。

5. 在**输出**下，确定哪个查找字段将作为新查找字段添加到数据集中符合条件的事件。

您应该在此找到一个从已选择的查找表的列中提取的字段列表。首先选择要添加到事件的字段。所有您已指定为输入的查找字段将不可用。必须至少定义一个输出字段才能使查找字段定义生效。

如果您在此未找到任何字段，则指定的**查找表**可能有问题。

6. 在**字段名称**下，提供查找字段在数据中应具有的字段的名称。

字段名称值不可包含空格、单引号、双引号、大括号或星号。

7. 在**显示名称**下，提供查找字段在“数据模型编辑器”中和数据透视表中的显示名称。

显示名称值不可以包含星号字符。

8. 为您定义的每个查找字段设置相应的**类型**和**标记**值。

有关**类型**字段的更多信息，请参阅[定义数据集字段](#)主题中的“标记字段为隐藏或必需”小节。

9. （可选）单击**预览**确认正在向符合条件的事件添加输出字段。

（符合条件的事件是指其输入字段值与查找表中输入字段值相匹配的事件）。有关更多信息，请参阅下文的“预览查找字段”。

10. 如果您认为查找运行良好且符合预期结果，则单击**保存**以保存您的字段，并返回到“数据模型构建器”。

新查找字段将添加到数据集字段列表的底部。

预览查找字段

在设置查找字段后，您可单击**预览**以查看查找字段是否已添加到符合条件的事件（指定的输入字段值与查找表中相应的输入字段值相匹配的事件）。Splunk Web 在两个或更多选项卡式页面中显示结果。

第一个选项卡显示由基本搜索返回的事件的示例。新查找字段应显示在第一列（_time 列）的右边。如果您在前几个页面的查找字段列中未看到任何值，则可能表明这些值非常罕见。您可通过查看剩余的预览选项卡进行检查。

取消预览保存

事件	Product Name	Price	Sale Price	Code									
✓ 1,000 个事件 (14/07/28 21:39:11.000 之前)													
示例 前 1,000 个事件 每页 20 个													
_time	product_name	price	sale_price	Code	host	source	sourcetype	bytes	Category ID	client IP	Product ID	status	
2014-04-29 17:05:25	Puppies vs. Zombies	4.99	1.99	F	www2	/opt/log/www2/access.log	access_combined	3868					21
2014-04-29 17:05:17	World of Cheese	24.99	19.99	D	www2	/opt/log/www2/access.log	access_combined	224					21
2014-04-29 17:05:02					www2	/opt/log/www2/access.log	access_combined	1522					21

Splunk Web 为每个您在**输出**部分选择的查找字段显示一个选项卡。每个字段选项卡提供一个快速摘要，说明所选事件示例的值的分布情况。它是由**计数**和百分比形式组织的上限值列表。

取消预览保存

事件	Product Name	Price	Sale Price	Code									
✓ 1,000 个事件 (14/07/28 21:40:50.000 之前)													
示例 前 1,000 个事件 每页 20 个													
Values	Count	%											
World of Cheese	71	13.004											
Mediocre Kingdoms	54	9.890											
Final Sequel	52	9.524											
Fire Resistance Suit of Provolone	51	9.341											
Orvil the Wolverine	49	8.974											
SIM Cubicle	49	8.974											
Dream Crusher	46	8.425											
Manganiello Bros.	42	7.692											
Holy Blade of Gouda	40	7.326											
Benign Space Debris	33	6.044											
Curling 2014	31	5.678											
Puppies vs. Zombies	28	5.128											

查找字段设置示例

假设以下事件是真的：

- 您有一个数据模型数据集，其中包含名为产品 ID 的自动提取字段和另一个名为产品名称的自动提取字段。您想要使用查找表向提供产品价格的数据集添加新字段。
- 您有一个 .csv 文件，名为 product_lookup。该表格包括与产品相关的几个字段，包括 productId 和 product_name（它们与您数据集中类似命名的字段有非常类似的值集），以及 price（查找表中的字段，您要将其作为查找字段添加到数据集）。
- 您了解有些产品具有相同的产品名称，但却有不同的产品 ID 值和价格。这意味着您不可以将仅取决于产品名称的查找定义设置为输入字段，因为它将尝试将查找表的相同 price 值应用于两个或更多产品中。您必须设计将产品名称和产品 ID 用作输入字段的查找字段定义，将匹配事件中的每个值组合与具有相同名称/ID 组合的查找表中的行进行匹配。

如果出现这种情况，要按以下步骤将 price 作为字段正确地添加到您的数据集中。

1. 在“设置”中，[创建查找定义](#)，使其指向 product_lookup.csv 查找文件。将此查找定义命名为 product_lookup。
2. 转到数据透视表，并打开您要添加查找字段的数据集的“数据模型编辑器”。
3. 单击添加字段，并选择查找。

随后会显示“使用查找编辑字段”页面。

4. 在查找表中选择 product_lookup。

所有在查找表中跟踪的字段将会在输出下显示。

5. 在输入下，定义两个查找中的字段/数据集中的字段对。第一个对应包括查找中的字段值为 ProductId，数据集中的字段值为产品 ID。第二个对应包括查找中的字段值为 product_name，数据集中的字段值为 Product Name。

第一个对将查找表的 productId 字段与数据集的产品 ID 字段相匹配。第二个对将查找表的 product_name 字段与数据集的 Product Name 字段相匹配。请注意，执行此操作时，在输出下，productId 和 product_name 字段的行变为不可用。

6. 在输出下，选择 price 字段的复选框。

此设置表示您要将其添加到在具有匹配输入字段的数据集中的事件。

7. 为 price 字段指定“显示名称”为价格。

price 字段的类型值应该已设置为数字。

8. 单击预览以测试 price 是否添加到您的事件中。

预览事件以表格形式显示，且 price 字段是时间戳之后的第二个列。

9. 如果 price 字段按预期在预览结果中显示，则单击保存以保存查找字段。

现在，数据透视表用户在构建数据透视表报表和仪表板时，可将价格用作字段选项。

添加正则表达式字段

您可向数据模型中的任何数据集添加正则表达式字段。正则表达式字段将正则表达式字符串命名的组变成单独的数据模型字段。您可安排正则表达式从 _raw 事件文本以及特定字段值中提取字段。

添加具有正则表达式的属性

数据模型: Buttercup Games 对象: HTTP Requests

文档

提取自

source

正则表达式

(?<Path>+)(?<File>+)(?<Extension>+)

示例

http://<uri_domain>/<uri_query_string>?<uri_query>

From: (?<from>+) To: (?<to>+)

了解更多信息

属性

字段名称

显示名称

类型

标志

Path

Path

字符串

可选

File

File

字符串

可选

Extension

Extension

字符串

可选

取消

预览

保存

所有

匹配

不匹配

Path

File

Extension

✓ 1,000 个事件 (14/07/28 21:47:31.000 之前)

示例 前 1,000 个事件

每页 20 个

	uri_path	Path	File	Extension
✓	/cart/success.do	/cart/	success	do
✗	/cart.do			
✗	/cart.do			
✗	/product.screen			
✗	/cart.do			

1. 在“数据模型编辑器”中，打开要向其添加正则表达式字段的数据集。

有关“数据模型编辑器”的概述，请参阅[设计数据模型数据集](#)。

2. 单击**添加字段**，并选择**正则表达式**。

此时将进入“使用正则表达式添加字段”页面。

3. 在**提取自**下，选择要从中进行提取的字段。

提取自列表应包括当前在数据集中查找到的所有字段，外加 `_raw`。如果您的正则表达式设计为从特定字段值中提取一个或多个字段，则从**提取自**列表中选择该字段。另一方面，如果您的正则表达式设计为分析整个事件字符串，则从**提取自**列表中选择 `_raw`。

4. 提供**正则表达式**。

正则表达式必须至少具有一个命名组。正则表达式中每个命名的表达式都作为单独字段提取。字段名称不可包含空格、单引号、双引号、大括号或星号。

提供正则表达式后，命名组在**字段**下显示。

注意：正则表达式字段目前不支持 `sed` 模式或 `sed` 表达式。

5. （可选）为字段提供不同的**显示名称**值。

字段**显示名称**值不可以包含星号字符。

6. （可选）纠正字段**类型**值。

默认情况下，它们将被指定为**字符串**。

7. （可选）将字段**标记**值更改为更适合您需要的其他任何值。

8. （可选）单击**预览**以查看字段在数据集中是否正常显示。

有关预览字段的更多信息，请参阅下文中的“预览正则表达式字段表示”。

9. 单击**保存**以保存更改。

您将返回到“数据模型编辑器”。正则表达式字段将被添加到已计算数据集字段的列表中。

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 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。

预览正则表达式字段表示

当您在定义一个或多个字段提取字段后单击**预览**时，Splunk 软件针对数据集中具有您所选**提取自**字段的数据集（若您正在从 `_raw` 中进行提取，则针对原始数据）运行正则表达式，并显示结果。在一组四个或更多选项卡式页面中，预览结果在设置字段下方显示。其中每个选项卡向您显示来自数据集中事件示例的信息。通过选择**示例**列表中的选项，如**前 1,000 个事件**或**前 24 小时**，您可决定该示例确定的方式。您还可决定每页显示多少个事件（默认为 20）。

如果预览未返回任何事件，则这可能表明您需要调整正则表达式或您选择了错误的**提取自**字段。

所有选项卡

所有选项卡使您快速了解与正则表达式相匹配的事件在事件数据中的普遍程度。您可在本主题顶部附近的屏幕截图的操作中看到**所有**选项卡的示例。

它向您显示数据中具有**提取自**字段的事件的未筛选示例。例如，如果您选定的**提取自**字段为 `uri_path`，则此选项卡仅显示具有 `uri_path` 值的事件。

第一列指示事件是否与正则表达式相匹配。匹配的事件具有绿色选中标记。非匹配的事件具有红色 "x" 标记。

第二个列显示事件中**提取自**字段的值。如果**提取自**字段为 `_raw`，则显示整个事件字符串。剩余的列显示由正则表达式提取的字段值（如果有）。

匹配和非匹配选项卡

“匹配”和“非匹配”选项卡与“所有”选项卡类似，只不过它们经筛选，仅显示与正则表达式**匹配**的事件或与正则表达式**不匹配**的事件。这些选项卡有助于您更好地了解示例中字段分布情况，尤其是在示例中的大多数事件都属于匹配事件集或非匹配事件集的情况下更是如此。

所有	匹配	不匹配	Path	File	Extension
✓ 1,000 个事件 (14/07/28 22:50:27.000 之前)					
<div> <div>示例 前 1,000 个事件</div> <div>每页 20 个</div> <div> < 预览 <div>1 2 3 4 5 6 7 8 9 ... 下一步</div> </div> </div>					
uri_path	Path	File	Extension		
/cart/success.do	/cart/	success	do		
/cart/success.do	/cart/	success	do		
/cart/success.do	/cart/	success	do		
/cart/success.do	/cart/	success	do		
/cart/success.do	/cart/	success	do		
/cart/success.do	/cart/	success	do		
/numa/numa.html	/numa/	numa	html		
/cart/success.do	/cart/	success	do		
/cart/success.do	/cart/	success	do		
/cart/success.do	/cart/	success	do		
/cart/error.do	/cart/	error	do		

字段选项卡

每个在正则表达式中已命名的字段都有一个自己的选项卡。字段选项卡中会简要给出所选事件示例的值的分布情况。它是由**计数**和百分比形式组织的上限值列表。如果您未看到您期望的值，或您正在查看的值分布情况看起来不对，这可能表示您需要微调您的正则表达式。

您还可增加示例大小以查找罕见字段值或在很久以前显示的值。在以下示例中，将**示例**设置为 *前 10,000 个事件* 则会针对 `path` 字段显示很多值，而这些值在仅显示前 1,000 个事件示例时并不会显示。

所有	匹配	不匹配	Path	File	Extension
✓ 10,000 个事件 (14/07/28 22:56:45.000 之前)					
<div> <div>示例 前 10,000 个事件</div> <div>每页 20 个</div> </div>					
值	计数	%			
/flower_store/	6441	95.112	<div></div>		
/cart/	223	3.293	<div></div>		
/flower_store/images/	85	1.255	<div></div>		
/numa/	6	0.089			
/rush/	6	0.089			
/stuff/	6	0.089			
/hidden/	5	0.074			

字段选项卡中的上限值表启用了钻取功能。您可单击某行以查看该行所表示的所有事件。例如，如果您正在查看 `path` 字段，且您看到有 6 个事件具有路径 `/numa/`，您可单击 `/numa/` 行以转到显示 `path="/numa/"` 的 6 个事件的列表。

添加地理 IP 字段

您可以将地理 IP 字段添加到其字段列表中已包含**类型**为 `ipv4` 字段的数据模型中的任意数据集。`ipv4` 字段必须显示在地理 IP 字段位置的上方，且要求尚未用于其他地理 IP 字段计算。

地理 IP 字段是一种查找。该字段可读取数据集事件中的 IP 地址值并向这些事件添加相关的经度、纬度、城市、区域和国家/地区值。

- 在“数据模型编辑器”中，打开要向其添加字段的数据集。
- 单击**添加字段**，并选择**地理 IP**，以定义地理 IP 字段。

这一操作会打开“通过 IP 查找添加地理字段”页面。

- 如果所选对象存在多个匹配的字段，则选择您要匹配的那个 **IP** 字段。
- 选择要添加到数据集的字段。
- （可选）通过更改选定字段的**显示名称**来重命名所选字段。

显示名称不可以包含星号字符。

- （可选）单击**预览**以验证地理 IP 字段正在通过所选的地理 IP 字段更新事件。

您应看到以表格形式（新地理 IP 字段显示为列）呈现的事件。例如，如果您正在使用基于事件的数据集且您已选择**城市、区域和国家/地区**地理 IP 地理 IP，则预览事件表应显示**城市、区域和国家/地区**列，位于第一列（**_time**）右侧。

预览窗格有两个选项卡。**事件**是默认选项卡。它以表格形式显示事件。选择**值**选项卡，以查看地理 IP 字段值在事件中的分布情况。

如果您未看到您期待的值范围，则尝试增加预览事件示例。默认情况下，示例设置为前一千个事件。您可通过将**示例**值设置为 *前 10,000 个事件* 或 *过去 7 天* 来增加它。

事件	Longitude	Latitude	City	Region	Country
✓ 10,000 个事件 (14/07/28 22:35:41.000 之前)					
示例: 前 10,000 个事件 每页 20 个					
值	计数	%			
Mexico	2424	24.240			
United States	2355	23.550			
France	1330	13.300			
Brazil	1099	10.990			
China	767	7.670			
United Kingdom	406	4.060			
Russia	361	3.610			
South Korea	209	2.090			
Germany	120	1.200			
Finland	97	0.970			
Canada	96	0.960			
Ukraine	74	0.740			
Thailand	73	0.730			
Argentina	55	0.550			
India	53	0.530			
Turkey	42	0.420			
Hong Kong	37	0.370			
Australia	35	0.350			
Spain	33	0.330			
Israel	32	0.320			

7. 单击**保存**以保存更改。

您将返回到“数据模型编辑器”。您已定义的地理 IP 字段值将被添加到数据集的“已计算”字段集中。

注意：地理 IP 字段作为**必填**字段添加到数据集，而且其**类型**值已经预先确定。您无法对这些值加以更改。

使用数据摘要加速搜索

基于摘要的搜索和数据透视表加速概述

Splunk Enterprise 能够基于大量数据生成报表。但是，它计算此类报表所用的时间量与其所汇总的事件数量成正比。简而言之，它可能会花费大量时间基于非常大的数据集生成报表。如果只是偶尔为之，则时间长度也许不是问题。但定期运行此类报表（或将其用作常用仪表板中面板的基础）是不切实际的，随着您组织中越来越多的用户运行相似的报表，这种不适宜性会迅速日益凸显。

要有效地基于大量数据生成报表，您需要创建使用报表所基于的搜索的后台运行结果填充的数据摘要。下次根据以此方式汇总的数据运行报表时，其完成速度应显著加快，因为摘要比生成摘要时所用的原始事件要小得多。

创建数据摘要的方式有三种：

- **报表加速** - 使用自动创建的摘要来加速某些类型报表的完成时间。
- **数据模型加速** - 使用自动创建的摘要来加速数据透视表的完成时间。
- **摘要索引** - 通过手动创建独立于主索引的单独摘要索引实现搜索和报表加速。

报表加速

报表加速用于加速单个报表。可以很容易地针对在大型数据集上运行的任何**转换搜索**或报表设置加速。

在 Splunk 软件的早期版本中，摘要索引用于加速报表。由于下列原因，报表加速比摘要索引更合适：

- **启动报表加速只需选中一个复选框并选择一个时间范围这样简单。**此后的所有操作都在后台进行。只要加速报表在选定的时间范围内运行（至少是部分），后续运行加速报表时其完成速度应该会加快。对于摘要索引，您需要设计搜索来填充索引，包括特殊的搜索命令；您可能还需要创建摘要索引。
- **Splunk 软件会自动与相似的搜索共享报表加速摘要。**假设员工 Mary 为一个报表设置了报表加速，这会导致 Splunk 软件为其构建一个摘要。几天之后，Joe 设计了一个报表，与 Mary 的报表十分相似，只有几处不同。Joe 为报表启用报表加速并将其保存起来时，Splunk 软件会自动将此报表分配给为 Mary 的报表所构建的摘要，这表示 Joe 不必等待构建摘要。
- **报表加速具有自动回填的特征。**如果由于某种原因您的数据发生中断，Splunk 软件可以侦测到中断，并在必要时自动更新或重新构建摘要。
- **报表加速摘要与您索引中的数据桶存储在一起。**而摘要索引则驻留在搜索头。以数据桶级在索引中存储摘要使 Splunk Enterprise 能够轻松处理迟来事件所造成的窘境，这种状况可能会导致完全重建摘要索引。由于 Splunk 软件摘要可同时覆盖热和温数据桶，它们可以汇总迟来数据，因为此类数据只能添加到热数据桶中。

必须注意，**不是所有搜索都符合报表加速的条件**。只有使用**转换命令**的搜索（此类搜索可将其结果转换为统计表格和图表）才符合条件。另外，在转换命令之前搜索中使用的任何命令都必须是**流命令**。此限制与摘要是在索引级而不是在搜索头构建的这一事实相关。

在 Splunk Web 中，可在将符合条件的搜索保存为报表时为该搜索启动报表加速。可通过以下方式符合条件的现有搜索启用报表加速：

- 在“报表”页面上，展开某报表所对应的行并单击**编辑**，以打开**编辑加速**对话框。如果您的报表符合加速的条件并且您的权限允许报表加速，则“编辑加速”对话框将会显示一个标有“加速报表”的复选框。选中此复选框。此时应该会显示“摘要范围”字段。选择计划针对哪个时间范围运行此报表，然后单击**保存**。
- 在**设置 > 搜索和报表**中，打开某报表的详细信息页面，单击**加速此搜索**，然后设置摘要范围。

请参阅“加速报表”。

可使用系统中的“报表加速摘要”页面来查看和管理通过报表加速创建的摘要。

请参阅[管理报表加速](#)。本主题还介绍了摘要的工作方式，并包含了符合条件的和不符合条件的搜索的示例。

应何时使用报表加速？

报表加速适用于有 100k 或更多**热数据桶**事件、完成速度缓慢且满足上面列出的合格条件的报表。

数据模型加速

可使用数据模型加速来加速在数据模型中定义的所有字段。数据模型加速后，由该数据模型生成的任何数据透视表或报表的完成速度应该比没有加速时快得多，即使数据模型表示很大的数据集。

数据模型加速有两种类型：临时和持续。临时加速应用于单个数据集、始终在运行且在用户数据透视表会话的持续时间内存在；而持续加速由管理员启用、在后台发生且可限制到更短的时间范围，如一周或一个月。搜索在针对启用加速数据模型中的数据集运行时，可随时使用持续加速。

数据模型加速利用 Splunk 的高性能分析存储 (HPAS) 技术。此技术采用类似于报表加速的方式来一起构建摘要和索引中的**数据桶**。和报表加速一样，持续数据模型加速也可以方便地启用；只需单击要加速的数据模型所对应的复选框并选择摘要范围即可。执行此操作之后，Splunk 软件便会开始构建一个涵盖所指示范围的摘要。当摘要完成之后，任何使用加速数据模型数据集的数据透视表、报表或仪表板面板都将尽可能地针对此摘要而不是 `_raw` 的完整数组运行，并且结果返回时间应该会大幅改善。

持续数据模型加速有一些限制。

- **持续数据模型加速仅适用于事件数据集层次结构和基于使用流命令的根搜索数据集的搜索数据集层次结构。**基于使用非流命令的根搜索数据集和根交易数据集的数据集层次结构无法进行加速。
 - 所有数据模型数据集都可以从“临时”数据模型加速中受益。请参阅下一小节。
- **数据模型持续加速之后，便不能再对其进行编辑。**在为数据模型启用加速之后，编辑此模型的唯一方法是禁用其加速。
- **默认情况下，只有具有管理员权限的用户可以持续加速数据模型。**
- **专用数据模型无法实现持续加速。**您必须与至少一个应用的用户共享某一数据模型，该数据模型才符合加速的条件。

在 Splunk Web 中，您可以在“数据模型”管理页面上为符合条件的数据模型启用数据模型加速，可通过多种方式访问此页面（包括导航到**设置 > 数据模型**）。

有关启用持续数据模型加速的更多信息，请参阅[管理数据模型](#)。

有关数据模型加速以及高性能分析存储后台工作原理的技术后台信息，请参阅[加速数据模型](#)。

临时数据模型加速

临时数据模型加速是指在后台运行的进程，针对所有未预先“持续”加速的数据模型数据集进行加速。与持续数据模型加速不同，临时数据模型加速应用于所有数据集类型，包括根搜索数据集、根交易数据集及其子数据集。

每当您基于尚未加速的数据集构建数据透视表时，Splunk 软件都将使用临时数据模型加速来在 `dispatch` 目录中构建一个临时加速摘要，此目录仅在“数据透视表编辑器”中定义了数据透视表时存在。结果是，当您在“数据透视表编辑器”中微调特定数据透视表时，您将会发现数据透视表性能提高，结果的返回速度比您首次进入此编辑器时要快。

这比不上持续数据模型加速（此类加速会持续维护数据模型数据集的摘要，并从您进入“数据透视表编辑器”时起就确保快速的性能），但是仍然很有用。

请参阅[加速数据模型](#)。

应何时使用持续数据模型加速？

如果您正在因“数据透视表编辑器”中数据透视表完成速度缓慢的问题而苦不堪言，并且这些数据透视表的来源数据集属于最高根事件数据集层次结构，则应该考虑启用加速该数据模型。这将确保基于这些数据集的数据透视表的结果返回速度比之前的方式要快。

此外，任何引用持续加速的数据模型数据集的报表或仪表板面板也将受益于这种加速（临时数据模型加速将不会发生这种情况）。

报表加速对比数据模型加速

通常，数据模型加速快于报表加速。然而，有些特定类型的搜索允许报表加速在数据模型加速完成前先完成。

您的转换搜索聚合越多，它就越快。当运行的搜索聚合为每个索引数据桶一项的时候，报表加速尤其快。例如，如果每天有上亿的事件发生而您只想要每个月的总数和平均值，报表加速返回的结果将优于数据模型加速。在本例中，您将最大限度地利用报表加速的聚合功能。

报表加速与数据模型加速采用类似的方式进行加速搜索。它们都自动预处理索引器中的事件和创建数据桶级别的加速摘要。但是数据模型加速的常规优势在于它的摘要不同于由报表加速创建的那些。

报表加速设计为创建包括预计统计的摘要。另一方面，数据模型加速用一种能更有效阅读的格式构建摘要，它使得 Splunk 软件能在不放弃性能的要求下进行统计计算。因此如果搜索相对复杂，您最好使用数据模型加速。

摘要索引

摘要索引是一种可用来加速长期运行的、不符合报表加速条件的报表的方法，如在转换命令之前使用**非流式**搜索命令的报表。它类似于报表加速，因为它也需要使用搜索结果填充数据摘要，但在这种情况下数据摘要实际上是一个在搜索头创建并存储的特殊**摘要索引**。此摘要索引由计划的报表填充，该计划的报表基于您要加速的报表，并且已在**设置 > 搜索和报表**中针对摘要索引选中了**启用**。

例如，如果要加速的报表使用的是**转换命令**，则可以使用将转换命令替换成以下相似的 "si-" 前缀摘要索引转换命令的报表来填充其摘要索引：sichart、sitimechart、sistats、sitop 和 sirareo

本文包含两个有关摘要索引设置的主题。

- [使用摘要索引提高报表效率](#)介绍了一种通过使用 si- 命令的计划搜索设置摘要索引的简便方法。
- [配置摘要索引](#)介绍了一种棘手而且困难的、使用 addinfo、collect 和 overlap 命令设置摘要索引的方法。如果您可以很轻松地设置涵盖了汇总统计信息的搜索，则应仅使用后一种方法。

即使发生许可证违规，摘要索引量也不用于计算许可证，因为摘要索引将如任何其他非内部搜索行为一样停止下来。

应何时使用摘要索引？

如果您所使用的报表符合报表加速的条件，则几乎可以始终优先使用加速大数据量搜索性能的方法。

在以下情况下，您可能希望使用摘要索引而不是报表加速：

- 要加速的主报表在转换命令（与报表加速一样，摘要索引填充报表必须包含转换命令）前包含非流式命令。
- 您希望仅通过在搜索字符串中包含 index=<summary_index_name> 来根据特定摘要索引运行任何报表。（在报表加速下，Splunk 软件会自动决定是否针对特定数据摘要运行报表。）
- 您的原始数据比报表窗口滚动更频繁（例如，您的保存策略是 6 个月，但您想用过去一年的数据操纵仪表板面板）。摘要索引通常比其聚合的事件占用的空间小、可被单独保留且持续时间较长。

批处理模式搜索

批处理模式搜索功能可提高转换搜索的性能和可靠性。对于不要求事件按时间顺序的转换搜索，以批处理模式运行意味着搜索按数据桶（分批）执行，而不是基于时间来执行。在某些报表情况中，这意味着转换搜索可以更快速地完成。另外，批处理模式搜索可以提高长时间运行的分布式搜索的可靠性，当索引器在搜索运行期间关闭时，可能无法进行这种分布式搜索。在这种情况下，Splunk 软件会尝试重新连接到丢失的对等节点并重试搜索。

满足批处理模式搜索条件的转换搜索包括：

- 搜索中不包括 localize 或 transaction 命令的生成和转换搜索（stats、chart 等）。
- 非实时和非摘要搜索。
- 非状态流的非分布式搜索。（streamstats 搜索是状态流搜索的一个示例。）

可以从配置文件调用批处理模式搜索，具体位置为 [search] 段落（位于 limits.conf 中）。使用搜索查看器来确定是否以批处理模式运行转换搜索。

请参阅[配置批处理模式搜索](#)。

管理报表加速

对于因要覆盖大量数据而需要花费很长时间来完成的**转换搜索**和报表而言，报表加速是最简单的加速方式。可以在将转换搜索保存为报表时为该搜索启用加速。也可以加速使用转换搜索的基于报表的仪表板面板。

本主题更为详细地介绍了报表加速的各个方面。其中包括：

- 为转换报表启用自动加速的快速指南。
- 符合条件的和不符合条件的报表的示例（仅特定报表类型符合报表加速的条件）。
- 有关在哪里创建和维护报表加速摘要的详细信息

- “设置”中“报表加速摘要”页面的概述，您可以使用该页面来查看和维护用于执行自动报表加速的数据摘要。

报表加速的限制

如果出现以下情况，您将无法加速报表：

- 报表是通过数据透视表创建的。通过数据模型加速来加速数据透视表报表。请参阅本手册中的[“管理数据模型”](#)。
- 您的权限不允许您加速搜索。如果您的角色不具有 `schedule_search` 和 `accelerate_search` 操作，则无法加速报表。
- 您的角色没有此报表的写入权限。
- 报表所基于的搜索不符合加速条件。有关更多信息，请参阅本主题中的[报表如何才能符合加速的条件](#)。

此外，如果报表基本搜索包含**标记、事件类型、搜索宏**和其他知识对象，则加速这些报表时应当格外小心，因为这些知识对象的定义可能会在加速报表后单独进行更改。如果此情况发生，则加速报表可能会返回无效的结果。

如果怀疑您所加速的报表正返回无效结果，则可验证其摘要，确认摘要中包含的数据是否一致。请参阅本主题中的[验证摘要](#)。

启用报表加速

您可以在创建报表时或创建报表之后启用报表加速。

有关此过程更全面的描述，请参阅《报表手册》中的“创建和编辑报表”。

在创建报表时启用报表加速

要为符合条件的运行缓慢的搜索启用报表加速，您只需执行以下操作：

1. 在“搜索”中，运行该搜索。
2. 单击**另存为**，然后选择**报表**。将打开“另存为报表”对话框。
3. 为您的报表指定一个**名称**，也可以指定**描述**。单击**保存**以将该搜索保存为报表。这会将您带到“已创建报表”对话框。
4. 单击**加速**以加速您的报表。这会将您带到“编辑加速”对话框。

在以下情况下，“编辑加速”对话框将显示“无法加速此报表”：

- 您的权限不允许您加速报表。您的角色必须具有 `schedule_search` 和 `accelerate_search` 操作。
 - 您的报表不符合报表加速的条件。有关更多信息，请参见下面的子主题[“报表如何才能符合加速的条件”](#)。
5. 如果您的报表符合加速的条件并且您的权限允许报表加速，则“编辑加速”对话框将会显示一个标有**加速报表**的复选框。选中此复选框。
 6. 此时应该会显示**摘要范围**字段。根据您计划针对哪个时间范围运行报表，选择 **1 天**、**7 天**、**1 个月**、**3 个月**、**1 年**或**所有时间**。例如，如果您仅计划对最近七天内的时间段运行报表，请选择 **7 天**。
 7. 单击**保存**以保存加速设置。

为现有报表启用报表加速

要为现有报表启用报表加速，请在“报表”页面上找到该报表，并展开其所在的行以显示报表详细信息。如果该报表尚未启用，**加速**值将处于**禁用**状态。单击**编辑**打开“编辑加速”对话框。按照上面列表中的步骤 5-7 定义并保存您的报表加速设置。

注意：如果您的权限不允许加速报表或者报表不符合报表加速的条件，则“编辑加速”对话框将显示“无法加速此报表”。

也可以在**设置 > 搜索和报表**中为现有报表启用报表加速。

在为报表启用加速之后

当为报表启用加速时，如果 Splunk 软件确定该报表可从摘要获益，它将开始为报表构建一个报表加速摘要。如果为报表启用了加速，发现系统没有为其构建摘要（如果转到**设置 > 报表加速摘要**，您会注意到**摘要状态**长时间保持在已完成 0% 状态），请参见下面的子主题[“在什么条件下 Splunk 软件不会构建或更新摘要”](#)。

摘要构建完成之后，后续运行加速报表时其完成速度要比以前快。请参见下面的子主题以了解有关摘要及其工作方式的更多信息。

注意：在“搜索”中运行报表时，要记住报表加速仅对其**搜索模式**设置为**智能**或**快速**的报表起作用。如果为受益于报表加速的报表选择了**详细**搜索模式，则它运行起来就像没有摘要那样慢。（**搜索模式**不会影响搜索操纵仪表板面板。）有关**搜索模式**设置的更多信息，请参见《搜索手册》中的“设置搜索模式以调整搜索体验”。

报表如何才能符合加速的条件

要使报表符合加速的条件，其搜索必须满足以下三个要求：

- 搜索字符串必须使用**转换命令**（如 `chart`、`timechart`、`stats` 和 `top`）。
- 如果搜索字符串在第一个转换命令之前还包含其他任何命令，这些命令必须为**非流式**。
- 此搜索不可使用事件示例。

注意：可以在第一个转换命令后使用非流式命令，此时报表仍符合自动加速的条件。仅在第一个转换命令前使用非流式命令的报表不符合加速的条件。

有关事件示例的更多信息，请参阅《搜索手册》中的“事件示例”。

符合条件的搜索字符串的示例

下面是符合报表加速条件的搜索字符串示例：

```
index=_internal | stats count by sourcetype

index=_audit search=* | rex field=search "'(?.*)'" | chart count by user search

test foo bar | bin _time span=1d | stats count by _time x y

index=_audit | lookup usertogroup user OUTPUT group | top searches by group
```

不符合条件的搜索字符串的示例

下面是不符合报表加速条件的搜索字符串示例：

思考以下搜索字符串失败的原因：这是一个简单的事件搜索，没有转换命令。

```
index=_internal metrics group=per_source_thruput
```

思考以下搜索字符串失败的原因：`eventstats` 不是转换命令。

```
index=_internal sourcetype=splunkd *thruput | eventstats avg(kb) as avgkb by group
```

思考以下搜索字符串失败的原因：`transaction` 不是流命令。其他非流命令包括 `dedup`、`head`、`tail`，以及任何其他未出现在流命令列表中的搜索命令。

```
index=_internal | transaction user maxspan=30m | timechart avg(duration) by user
```

符合报表加速的条件但不会有很大帮助的搜索字符串

另外，报表也可能从技术上符合报表加速的条件，但不会从中得到很大帮助。对于数据基数很高的报表这是常见情形，当搜索字符串中有两个或更多转换命令并且第一个转换命令生成许多（50,000 以上）输出时您会发现这种情况。例如：

```
index=* | stats count by id | stats avg(count) as avg, count as distinct_ids
```

设置报表加速摘要的时间范围

报表加速摘要涵盖近似的时间范围。从**摘要范围**列表选择一个值即确定了此时间范围。有时，报表加速摘要会有一个稍微超出其摘要范围的数据存储，但是摘要总是尽力符合该范围，除非是在第一次构建它的那段时期。

例如，如果您将一个加速报表的摘要范围设为 **7 天**，则会创建一个能够大概涵盖过去 7 天的数据摘要。每 10 分钟会运行一个搜索，以确保摘要始终涵盖了所选的范围。这些维护搜索会添加新的摘要数据，并删除超出范围的更旧的摘要数据。

当您之后运行过去 7 天范围内的加速报表时，报表会搜索其摘要而不是来源索引（报表最初搜索的索引）。在大多数情况下，摘要包含的数据比来源索引小得多，另外对于部分搜索管道而言，报表摘要包含预先计算的结果，这表示报表的完成速度应该要比首次运行时快。

当您针对摘要仅涵盖其中一部分的时间段运行加速的报表时，则报表不会完成得那么快。这是因为对于报表时间范围未落在摘要范围内的部分，Splunk 软件必须转到来源索引。

如果报表的**摘要范围**设置为 **7 天**，而您要针对过去 9 天运行报表，则只有在 Splunk 软件中涵盖过去 7 天的那部分搜索会受益于加速。针对第 8 天和第 9 天运行的那部分报表将以正常速度进行。

在设置**摘要范围**值时要记住这一点。如果您时常计划针对超出过去 7 天但没有超过 30 天的时间范围运行报表，则在为该报表设置报表加速时，应为**摘要范围**选择 **1 个月**。

Splunk 平台如何构建报表加速摘要

在您为合格的报表启用了加速之后，Splunk 软件会决定是否为该报表构建摘要。在所选**摘要范围**所涵盖的**热数据桶**中的事件数量大于或等于 100,000 时，才会为合格的报表生成摘要。有关更多信息，请参见下面的子主题“[在什么条件下 Splunk 平台不会构建或更新摘要](#)”。

当 Splunk 软件决定为该报表构建摘要时，它便开始运行报表以使用数据填充摘要。当摘要完成时，此报表仍会每 10 分钟运行一次，以使摘要保持在最新状态。每次更新都能确保摘要涵盖了所配置的整个时间范围，且数据中没有明显的间隙。这种摘要构建方法也能确保为迟来的数据构建摘要，而不会引起其他复杂情况。

报表加速摘要需要时间来构建和维护

构建报表摘要可能会花费一些时间。创建时间取决于摘要中所涉及的事件数量、总体摘要范围以及摘要时间跨度（数据块）的长度。

可以在“设置”的“报表加速摘要”页面上跟踪摘要完成进度。可以在主页上查看**摘要状态**以了解摘要完成百分比。

注意：就像普通计划的报表一样，定期自动填充报表加速摘要的报表由报表计划程序管理。默认情况下，最多只允许报表计划程序将其总搜索带宽的 25% 用在报表加速摘要的创建上。

报表计划程序也会以最低优先级运行填充报表加速摘要的报表。如果这些“自动摘要”报表的计划与用户定义的告警、摘要索引报表以及定期运行的报表相冲突，则始终最先运行用户定义的报表。这意味着您可能会遇到，因为正在运行更高优先级的报表而未创建或更新摘要的情况。

有关搜索计划程序的更多信息，请参阅《报表手册》中的主题“配置计划的报表优先级”。

使用并行摘要来加快报表摘要的创建和维护

如果您觉得一些加速报表的摘要构建或更新太慢，您可以打开这些报表的并行摘要来加快该过程。为此，您可以在存在问题的一个或多个报表的 `savedsearches.conf` 中添加一个参数。

如果使用了并行摘要，则会并发运行多个搜索任务来构建报表加速摘要。它也会以 10 分钟的计划间隔来运行相同数量的并发搜索，以维护这些摘要文件。并行摘要降低了构建和维护报表加速摘要所需要的时间量。

这类摘要搜索性能改善也是一项成本支出。并发搜索的数量对比 Splunk 部署可以运行的并发搜索任务的总数量，这意味着它们会导致索引器的资源使用率增加。

1. 打开 `savedsearches.conf` 文件，该文件包含要为其更新摘要设置的报表。

2. 查找该报表的段落。

3. 添加 `auto_summarize.max_concurrent = 2`，如果该参数未出现在段落中。

4. 保存更改。

如果您打开一些报表的并行摘要，并发现您的整体搜索性能受到了影响，那么或者是因为您一次运行了太多的搜索，或者因为达到了您的并发搜索限制，此时您可以轻松地再将加速报表的 `auto_summarize.max_concurrent` 值恢复为 1。

通常我们不建议将 `auto_summarize.max_concurrent` 增加为大于 2 的值。然而，如果您的 Splunk 部署具备进行大量搜索并发的能力，则可以尝试将所选的加速报表的 `auto_summarize.max_concurrent` 设置为 3 或更大值。

有关更多信息，请参阅：

- 有关并发搜索对于搜索性能的影响的信息，请参阅《容量规划手册》中的“容纳许多同时进行的搜索”。
- 有关如何根据您的实现确认并发搜索限制的更多信息，请参阅“配置计划报表的优先级”。

使用常规时间跨度将摘要数据分成多个数据块

Splunk 软件在构建和维护摘要时，它会根据基于总体摘要范围自动确定的“时间跨度”将数据分成多个数据块，以确保统计信息正确无误。例如，当报表的摘要范围是 1 个月时，则可能会选择一个 1d（一天）的时间跨度。

摘要时间跨度代表了摘要将包含其中精确统计数据的最小时间范围。如果针对时间跨度为一小时的摘要运行报表，当您希望报表使用摘要数据时，您为报表选择的时间范围将按该时间跨度进行均匀划分。如果使用 1h 的时间跨度，则针对过去 24 小时的报表可以很好地运行，但针对过去 90 分钟的报表可能无法使用摘要数据。

摘要可以有多个时间跨度

在必要时会为报表加速摘要分配多个时间跨度，以尽可能使其可搜索。例如，摘要范围为 3 个月的摘要可以具有 1 个月和 1 天这两种时间跨度。另外，当摘要涵盖多个索引数据桶而这些数据桶所涵盖的时间量又有很大差别时，可能会分配更多的时间跨度。例如，如果摘要涵盖两个数据桶，第一个数据桶涵盖两个月，第二个数据桶涵盖 40 分钟，则摘要会有时间跨度为 1 天和 1 个月的数据块。

可以手动设置摘要的时间跨度（但不建议这样做）

您可以在 `savedsearches.conf` 中手动设置报表级别的摘要时间跨度，具体做法为更改 `auto_summarize.timespan` 参数的

值。如果要手动设置摘要的时间跨度，请记住，非常小的时间跨度会导致极其缓慢的摘要创建速度，尤其在摘要范围很长时。另一方面，时间范围较大时，虽然摘要构建速度很快，但时间范围较短的报表不能使用此摘要。在几乎所有情况下，为了达到最佳性能和最高可用性，最好要让 Splunk 软件决定摘要的时间跨度。

Splunk 软件为加速报表收集数据的方式会导致在很短的时间内产生大量的文件

因为报表加速摘要收集多个时间跨度的信息，对于同样的摘要，在短时间内会创建许多文件。如果您对来说，文件和文件夹管理是一个问题，则有一些事项需要注意。

对于您的系统中每个加速报表和搜索头的组合，您将获得：

- 对于每个 1 天跨度，获得 2 个文件（数据 + 信息）
- 对于每个 1 小时跨度，获得 2 个文件（数据 + 信息）
- 对于每个 10 分钟跨度，获得 2 个文件（数据 + 信息）
- 有时：对于每个 1 分钟跨度，获得 2 个文件（数据 + 信息）

因此，如果您有一个 30 天范围和 10 分钟粒度的加速报表，结果是：

```
(30x1 + 30x24 + 30x144)x2 = 10,140 files
```

如果您切换到 1 分钟粒度，结果是：

```
(30x1 + 30x24 + 30x144 + 30x1440)x2 = 96,540 files
```

如果您使用了部署监视器，默认情况下它随附 12 个加速报表，对于每个启用了它的搜索头，一个立即回填会在 `$SPLUNK_HOME/var/lib/splunk/_internaldb/summary` 中的每个索引器上产生 12.2 万到 120 万个文件。

在哪里创建和存储报表加速摘要

索引器上会创建加速的报表摘要，与涵盖摘要所涉及的时间范围的数据桶平行。例如，对于 "index1" 索引，它们驻留在 `$SPLUNK_HOME/var/lib/splunk/index1/summary` 下。

注意：数据模型加速摘要以相同的方式存储，但是存储在标记为 `datamodel_summary` 而非 `summary` 的目录中。

请记住，报表加速摘要不是索引。索引器群集不复制报表加速摘要或适应它们的存在。如果主要性从数据桶的原始副本重新分配到另一个副本（例如，由于拥有主要副本的对等节点发生故障），则摘要不会移动到拥有新主要副本的对等节点。因此，它会变得不可用。直到下一次 Splunk 软件尝试更新摘要，发现它已丢失，并重新生成它，摘要才再次可用。

注意：在 Splunk 软件重新生成摘要之前，加速报表的结果虽然正确，但运行速度较慢。

有关索引器群集如何处理报表加速摘要的更多信息，请参阅《管理索引器和索引器群集》手册中的“在索引器群集中搜索如何工作”。

为报表加速摘要配置基于大小的保存

您为您的索引设置基于大小的保存限制以便它们不会占用太多的磁盘存储空间了吗？默认情况下，报表加速摘要理论上可占用无限量磁盘空间。如果您还锁定了您的索引或索引量的最大数据大小，则这可能是个问题。好消息是，您可以选择为您的报表加速摘要配置类似的保存限制。

注意：虽然报表加速摘要默认情况下不受大小限制，但它们与您的热/温数据桶中的原始数据关联，且与它一起老化。当事件从热/温数据桶传递到冷数据桶时，它们也从相关摘要中删除。

重要提示：在尝试为报表加速摘要配置基于大小的保存前，首先您应了解如何使用卷跨多个索引配置索引大小限制，因为很多原则是相同的。有关更多信息，请参阅《管理索引器和群集》中的“配置索引大小”。

默认情况下，报表加速摘要与热/温数据桶一起存在于位于 `homePath/./summary/` 的索引中。换句话说，如果在 `indexes.conf` 中，索引中热/温数据桶的 `homePath` 是：

```
homePath = /opt/splunk/var/lib/splunk/index1/db
```

则映射到该索引中数据桶的摘要将创建于：

```
homePath/opt/splunk/var/lib/splunk/index1/summary
```

您可以执行以下步骤来为该索引中的摘要设置基于大小的保存。所有描述的配置都在 `indexes.conf` 中进行。

1. 检查您的卷定义并确定一个将成为您的报表加速摘要数据主页的卷。

如果正确的卷不存在，需创建一个。

如果您想利用控制索引原始数据的现有卷，则您可使该卷引用托管热/温数据桶目录的文件系统，因您的报表加速摘要将与它一起存在。

不过，如果您愿意，您也可将报表加速摘要放到它们自己的文件系统中。仅有的规则如下：每卷只可引用一个文件系统，但每个文件系统可以引用多个卷。

2. 对于将成为报表加速数据的主页的卷，添加 `maxVolumeDataSizeMB` 参数以设置卷的最大大小。

这样您就可以在索引中管理报表加速摘要数据的基于大小的保存。

3. 更新索引定义。

为每个处理摘要数据的索引设置 `summaryHomePath`。确保路径正在引用您在步骤 1 中确定的摘要数据卷。

`summaryHomePath` 覆盖摘要的默认路径。它的值应补充索引中热/温数据桶的 `homePath`。例如，以下 `summaryHomePath` 就是补充以上确定的 `homePath` 值：

```
summaryHomePath = /opt/splunk/var/lib/splunk/index1/summary
```

该示例配置显示基于全局、按卷和按索引设置的数据大小限制。

```
#####
# Global settings
#####

# Inheritable by all indexes: No hot/warm bucket can exceed 1 TB.
# Individual indexes can override this setting. The global
# summaryHomePath setting indicates that all indexes that do not explicitly
# define a summaryHomePath value will write report acceleration summaries
# to the small_indexes # volume.
[global]
homePath.maxDataSizeMB = 1000000
summaryHomePath = volume:small_indexes/$_index_name/summary

#####
# Volume definitions
#####

# This volume is designed to contain up to 100GB of summary data and other
# low-volume information.
[volume:small_indexes]
path = /mnt/small_indexes
maxVolumeDataSizeMB = 100000

# This volume handles everything else. It can contain up to 50
# terabytes of data.
[volume:large_indexes]
path = /mnt/large_indexes
maxVolumeDataSizeMB = 50000000

#####
# Index definitions
#####

# The report_acceleration and rare_data indexes together are limited to 100GB, per the
# small_indexes volume.
[report_acceleration]
homePath = volume:small_indexes/report_acceleration/db
coldPath = volume:small_indexes/report_acceleration/coldddb
thawedPath = $SPLUNK_DB/summary/thaweddb
summaryHomePath = volume:small_indexes/report_acceleration/summary
maxHotBuckets = 2

[rare_data]
homePath = volume:small_indexes/rare_data/db
coldPath = volume:small_indexes/rare_data/coldddb
thawedPath = $SPLUNK_DB/rare_data/thaweddb
summaryHomePath = volume:small_indexes/rare_data/summary
maxHotBuckets = 2

# Splunk constrains the main index and any other large volume indexes that
# share the large_indexes volume to 50TB, separately from the 100GB of the
# small_indexes volume. Note that these indexes both use summaryHomePath to
```

```
# direct summary data to the small_indexes volume.
[main]
homePath = volume:large_indexes/main/db
coldPath = volume:large_indexes/main/coldddb
thawedPath = $SPLUNK_DB/main/thaweddb
summaryHomePath = volume:small_indexes/main/summary
maxDataSize = auto_high_volume
maxHotBuckets = 10

# Some indexes reference the large_indexes volume with summaryHomePath,
# which means their summaries are created in that volume. Others do not
# explicitly reference a summaryHomePath, which means that the Splunk platform
# directs their summaries to the small_indexes volume, per the [global] stanza.
[idx1_large_vol]
homePath=volume:large_indexes/idx1_large_vol/db
coldPath=volume:large_indexes/idx1_large_vol/coldddb
homePath=$SPLUNK_DB/idx1_large/thaweddb
summaryHomePath = volume:large_indexes/idx1_large_vol/summary
maxDataSize = auto_high_volume
maxHotBuckets = 10
frozenTimePeriodInSecs = 2592000

[other_data]
homePath=volume:large_indexes/other_data/db
coldPath=volume:large_indexes/other_data/coldddb
homePath=$SPLUNK_DB/other_data/thaweddb
maxDataSize = auto_high_volume
maxHotBuckets = 10
```

当报表加速摘要卷达到它的大小限制时，Splunk 卷管理器会删除卷中最旧的摘要以腾出空间。当卷管理器删除某摘要时，它将 marker 文件放置到其相应的数据桶中。此 marker 文件告诉摘要生成器不要重新构建摘要。

数据模型加速摘要具有名为 `_splunk_summaries` 的默认卷，所有索引都引用该默认卷，用于数据模型加速摘要基于大小的保存。默认情况下，此卷不具有 `maxVolumeDataSizeMB` 设置，这意味着它具有无限的保存空间。

您可使用此现有卷在一个位置上管理数据模型加速摘要和报表加速摘要。您需要：

- 使您的报表加速摘要的 `summaryHomePath` 参考引用 `_splunk_summaries` 卷。
- 将 `maxVolumeDataSizeMB` 设为 `_splunk_summaries` 的值。

有关数据模型加速摘要基于大小的保存的更多信息，请参阅本手册中的[“加速数据模型”](#)。

一个摘要多个报表

当搜索满足以下两个条件时，单个的报表摘要可以与多个搜索相关联。

- 搜索直到并包括首条报表命令都完全相同。
- 搜索的所有者相同，并与同样的应用相关联。

满足第一个条件但是其所有者不同或属于不同的应用，这样的搜索不能共享相同的摘要。

例如，下面是两个使用同一报表加速摘要的报表。

```
sourcetype=access_* status=2* | stats count by price

sourcetype=access_* status=2* | stats count by price | eval discount = price/2
```

下面这两个报表使用不同的报表加速摘要。

```
sourcetype=access_* status=2* | stats count by price

sourcetype=access_* status=2* | timechart by price
```

如果两个报表除语法差异外完全相同，而且语法差异不会导致它们输出不同的结果，则它们也可以使用相同的摘要。

下面是两个使用同一报表加速摘要的搜索。

```
sourcetype = access_* status=2* | fields - clientip, bytes | stats count by price

sourcetype = access_* status=2* | fields - bytes, clientip | stats count by price
```

也可以根据摘要运行非保存的搜索，只要基本搜索与填充的已保存搜索直到首条报表命令都相匹配，并且搜索时间范围在摘要跨度内。

可以通过导航到**管理器 > 报表加速摘要**来查看哪些搜索与您的摘要相关联。请参阅本主题中的[“使用‘报表加速摘要’”](#)。

要'页面'。

在什么条件下 Splunk 平台不会构建或更新摘要

Splunk 软件在要为报表创建摘要的数据符合下列条件时，不会为报表构建摘要：

- 选定**摘要范围**所涵盖的**热数据桶**中的事件数量小于 100,000。当该条件存在时，您会看到一个**摘要状态警告**，表示**没有足够的数据用来创建摘要**。
- Splunk 软件估计完成的摘要不会超过部署中总数据桶大小的 10%。当 Splunk 软件做出这个评估时，它会将摘要暂停 24 小时（此时您会看到**摘要状态为暂停**）。

可在**设置 > 报表加速摘要**中查看摘要的**摘要状态**。

如果您定义了一个摘要，但 Splunk 软件因满足了上述条件而不创建它，Splunk Enterprise 会继续定期核查，以确定条件是否改进。当摘要在其范围内有 100,000 个或更多的热数据桶事件，且在完成时不会变得过大，Splunk 软件便开始创建或更新摘要。

如何判断报表是否正在使用其摘要？

报表正在使用其摘要的明显提示是，如果在运行报表时发现其报表性能有了提高（完成速度比以前快），就表示报表正在使用摘要。

但如果这还不够，或您不能确定性能是否有改善，则可以在《[搜索手册](#)》的“查看搜索任务属性”中查找指示报表是否正在使用特定报表加速摘要的调试消息。下面提供了一个示例：

```
DEBUG: [thething] Using summaries for search, summary_id=246B0E5B-A8A2-484E-840C-78CB43595A84_search_admin_b7a7b033b6a72b45, maxtimespan=
```

在本示例中，最后一个数字字符串 b7a7b033b6a72b45 与“报表加速摘要”页面上显示的**摘要 ID** 相对应。

使用“报表加速摘要”页面

可以使用“设置”中的“报表加速摘要”页面来查看您的报表加速摘要，甚至管理它们的各个方面。请转到**设置 > 报表加速摘要**。

报表加速摘要					
显示 3 个项目中的 1-3					
摘要 ID *	规范化摘要 ID *	使用摘要的报告	摘要负载 *	访问计数 *	摘要状态 *
7e8cb0de72523dac	NS8149b5333fa5bed3	ACC	0.0007	0 上次访问: Never	20% 完成 更新时间: < 1 min ago
44044717708bd1d0	NS5dba11481031d5af	Buttercupgames	0.0005	0 上次访问: Never	没有足够的数据进行摘要* 更新时间: Never
a79cc7a337e35ad3	NS2045392fa55435f4	Stats	0.0000	0 上次访问: Never	正在构建摘要 - 0% 更新时间: Never
显示 3 个项目中的 1-3					

可以使用主“报表加速摘要”页面来查看有关您有查看权限的摘要的基本信息。

摘要 ID 和**规范化摘要 ID** 列显示分配给这些摘要的唯一的哈希值。ID 源自报表的远程搜索字符串。它们作为摘要文件创建的目录名称的一部分。单击一个摘要 ID 或规范化摘要 ID 可以查看摘要详细信息，并执行摘要管理操作。有关此详细视图的更多信息，请参见下面的子主题“[查看摘要详细信息](#)”。

使用摘要的报表列列出了与您的每个摘要相关联的已保存报表。它表明与特定摘要关联的每个报表都将因该摘要而受益于报表加速。单击一个报表标题可深入到该报表的详细信息页面。

查看**摘要负载**可以了解 Splunk 软件需要在更新摘要上投入的工作量。它是通过用运行填充报表所用的秒数除以填充报表的间隔而计算出来的。所以，如果报表每 10 分钟（600 秒）运行一次，每次需花 30 秒来运行，摘要负载就是 0.05。如果摘要负载很高，且摘要的**访问计数**显示摘要很少被使用或长时间内从未使用过，您可能会考虑删除摘要以减少系统的压力。

摘要状态列表明了摘要的总体状态，以及它最后一次使用新数据进行更新的时间。可能的状态值有**完成**、**待定**、**暂停**、**没有足够数据用来创建摘要**或者当前摘要的完成百分比。如果您希望将摘要更新到当前时刻，请单击其摘要 ID 转到其详细信息页面，然后单击**更新**启动一个填充摘要的新报表。

如果**摘要状态**是**待定**，这意味着摘要可能稍微过时了，搜索头准备为其计划一个新的更新任务。

如果**摘要状态**为**暂停**，则表示因要创建的摘要过大，不值得为报表构建摘要。Splunk 软件将规划摘要的大小，此摘要可由报表创建。如果它确定摘要将大于它所包含的索引数据桶的 10%，则会将该摘要暂停 24 小时。例如，如果摘要包含完整索引中 90% 的数据，则创建摘要没有意义。

您不能覆盖摘要暂停，但可以通过更改 `savedsearches.conf`，中 `auto_summarize.suspend_period` 属性的值来调整摘要暂停的时间长度

如果**摘要状态**为**没有足够的数据进行摘要**，这表示 Splunk 软件当前没有生成或更新摘要，因为与其相关联的报表

从由摘要范围涵盖的**热数据桶**返回的事件数少于 100,000。有关更多信息，请参见上面的子主题“[在什么条件下 Splunk 平台不会构建或更新摘要](#)”。

查看摘要详细信息

可以使用摘要详细信息页面来查看有关特定摘要的详细信息，并对该摘要执行操作。通过单击“设置”中“报表加速摘要”页面上的**摘要 ID** 即可转到此页面。

摘要：7e8cb0de72523dac

摘要状态

待决

更新时间：3h 23m ago

已验证 6d 19h 51m ago

操作

验证

更新

重建

删除

使用此摘要的报告

搜索名称	所有者	应用
dfdsf	admin	search
top checkins by venue, past 24h	admin	search

详细信息

了解更多信息。

摘要加载	0.0001
访问计数	0 上次访问：Never
磁盘上的大小	11.62MB
摘要范围	7 days
时间跨度	10min, 1d, 1h
数据桶	3
组块	744

摘要状态

在**摘要状态**下，您会看到该摘要的基本状态信息。它是“报表加速摘要”页面上列出的**摘要状态**的镜像（请参见上面的章节），同时提供了有关该摘要的确认状态的信息。

如果要摘要更新到当前时刻，请单击**操作**下的**更新**按钮以启动一个填充摘要的新报表。

如果您从未对摘要执行确认，则不会显示确认状态。在执行确认后，此状态将显示确认完成百分比。否则此状态将显示对摘要确认的最后一次尝试的结果；可能值有**已确认**和**未确认**，同时还表明此尝试发生在多久以前。

有关摘要确认的更多信息，请参见下面的“[确认摘要](#)”。

使用摘要的报表

使用此摘要的**报表**部分列出了与摘要关联的报表，及其所有者和主应用。单击一个报表标题可深入到该报表的详细信息页面。相似的报表（例如，搜索字符串全都使用不同转换命令转换同一根搜索的报表）可使用同一摘要。

摘要详细信息

详细信息部分提供了一组有关摘要的指标。

摘要负载和访问计数是主**报表加速摘要**页面上相应信息的镜像。有关更多信息，请参见上面的子主题“[使用‘报表加速摘要’页面](#)”。

磁盘上的大小显示了摘要占用了多少存储空间。可以使用此指标以及**摘要负载**和**访问计数**来确定哪些摘要应删除。

注意：如果**大小**值保持在 *0.00MB*，这表示 Splunk 软件当前没有生成此摘要，因为与其关联的报表没有足够多的事件。要求至少 100,000 个热数据桶事件。也可能预计的摘要大小超过报表与之关联的数据桶的 10%。Splunk 软件会定期检查此报表，当报表满足创建摘要的条件时会自动为其创建摘要。

摘要范围是摘要所涵盖的时间范围，始终与当前时刻相关。您将在定义填充摘要的报表时设置此项。有关更多信息，请参见上面的子主题“[设置报表加速摘要的时间范围](#)”。

时间跨度显示了组成摘要的数据块的大小。摘要时间跨度代表了摘要将包含其中精确统计数据的最小时间范围。所以，如果根据时间跨度为一小时的摘要运行报表，为了得到满意的结果，您为报表选择的时间范围应按该时间跨度进

行均匀划分。因此，如果使用 1h 的时间跨度，则针对过去 24 小时的报表可以很好地运行，但针对过去 90 分钟的报表可能会发生问题。有关更多信息，请参见上面的小节“[Splunk 平台如何构建摘要](#)”。

数据桶显示了摘要所涵盖的索引**数据桶**个数，**块数**则表明了构成摘要的数据块个数。这两个指标大多只用做参考信息，但是当您遇到摘要方面的问题时，可以帮助您排除故障。

确认摘要

有时您可能会发现加速的报表返回的结果似乎与在首次创建时报表返回的结果不一致。当报表的某些方面发生更改（如报表使用的标记、事件类型或字段提取规则的定义有更改）而您并不知情，此时就会发生这种情况。

如果您怀疑您的一个加速报表出现了这种情况，请转到报表与之关联的摘要的详细信息页面。可以运行确认过程来检查摘要的一个子集，确认所检查的全部数据是否一致。如果它发现数据不一致，将通知您确认已失败。

例如，假设您有一个报表使用了与特定网络安全事件类型关联的**数据类型** `netsecurity`。在为此报表启用加速后，Splunk 软件将为其构建一个摘要。后来，事件类型 `netsecurity` 的定义有更改，以至于返回了一组完全不同的事件，这表示用于填充摘要的数据集已不同于以前。您注意到加速的报表返回的结果似乎有所不同，所以在“设置”的“报表加速摘要”页面中对报表运行了确认过程。摘要没有通过确认，所以您开始调查根报表以查明发生了什么。

理论上为了节省时间，确认过程仅需查看摘要数据的一个子集，整个摘要的完整确认会花费与构建摘要本身一样长的时间才能完成。但在某些情况下需要执行更全面的确认。



单击**确认**打开**确认摘要**对话框。“确认摘要”提供了两个确认选项：

- **快速确认**，用于设置为以全面性为代价，快速确认摘要数据的一个小型子集。
- **全面确认**，用于设置为以速度为代价，全面地检查摘要数据。

对于这两种情况，均提供了预计确认时间。

单击**启动**开始确认过程后，可以在摘要的详细信息页面中的**摘要状态**下检查其进度。确认过程完后时，您可以在这里看到它是成功还是失败。不管怎样，您都可以单击确认状态来查看有关所发生情况的详细信息。

确认失败时，**确认已失败**对话框会告知您什么发生了错误：



在确认过程中，会跳过热数据桶以及正在构建中的数据桶。

当摘要未通过确认时，您可以检查根搜索字符串确定是否可以修复，以便提供正确结果。报表可以工作后，单击**重建**即可重建摘要，以使其完全一致。或者，如果您对当前报表很满意，则仅重建报表。如果您宁愿重新开始，请删除摘要，使用一个全新的报表重新开始。

更新、重建和删除摘要

如果**摘要状态**显示摘要已有一段时间未进行更新，您希望将其更新到最新状态，请单击**更新**。**更新**会启动一个标准的摘要更新报表来获取事件，这样便不会丢失数据，例如最后几个小时内的数据。

注意：当摘要的**摘要状态**为**暂停**时，您无法使用**更新**来更新摘要。

单击**重建**以从头开始重建索引。当您怀疑有数据因系统崩溃或类似事故而丢失时，或摘要未通过确认，且您已修复了基本报表或已确定摘要当前所包含的数据完好时，在这样的情况下您可能希望重建索引。

单击**删除**可从系统中删除摘要（并且以后不会重新生成摘要）。如果摘要很少被使用，并且它所占用的空间如果用于其他用途会更好，则此时您可能希望这样做。可以使用“设置”中的“搜索和报表”页面为与摘要关联的报表重新启用报表加速。

加速数据模型

数据模型加速工具可用于加速表示巨大数据集的数据模型。基于加速的数据模型数据集的数据透视表在加速之后，其完成速度要比之前快，基于这些数据透视表的报表和仪表板面板也是一样。

数据模型加速可借助“高性能分析存储”功能执行此操作，“高性能分析存储”功能采用类似于**报表加速**的方式在后台构建数据摘要。与报表加速摘要一样，数据模型加速摘要可方便地启用和禁用，它们与包含为其建立摘要的事件的索引数据桶一起存储在索引器上。

本主题涉及：

- 数据模型加速、报表加速和摘要索引之间的差异。
- 您如何启用数据模型的持续加速。
- Splunk 软件如何构建数据模型加速摘要。
- 如何使用 `tstats` 命令查询加速的数据模型加速摘要。
- 持续加速的数据模型的高级配置。

本主题还介绍了临时数据模型加速。当您使用未加速的数据集构建数据透视表时，Splunk 软件会应用临时数据模型加速。这甚至也适用于无法以持续的方式加速的、使用转换命令的交易数据集和搜索数据集。但是，当您退出“数据透视表编辑器”或在与“数据透视表编辑器”的会话期间切换数据集时，您所获得的所有加速优势将会丢失。“持续”加速的数据集没有这些弊端，这些数据集在通过“数据透视表”进行访问时始终都将加速加载。另外，与“持续”数据模型加速不同的是，临时加速不应用于使用数据透视表构建的报表或仪表板面板。

数据模型加速与报表加速和摘要索引的不同之处

这是数据模型加速与报表加速和摘要索引的不同之处：

- 报表加速和摘要索引按报表**加速**各个**搜索**。它们通过构建预先计算的搜索结果聚合集合来执行此操作。
- 数据模型加速针对在数据模型中定义的，以及您和您的数据透视表用户想要为其制作报表的，一整组**字段**加速报表。实际上，它加速由该字段集合表示的数据集，而非针对该数据集的特定搜索。

什么是高性能分析存储？

数据模型加速摘要由多个**时间序列索引文件**（文件扩展名为 `.tsidx`）组成。每个 `.tsidx` 文件包含所选数据集中索引字段::值组合的记录以及这些字段::值组合的所有索引位置。这些 `.tsidx` 文件构成了高性能分析存储。`.tsidx` 文件集体进行优化，以加速涉及在加速的数据模型中所定义的字段集的一系列分析搜索。

加速的数据模型的高性能分析存储涵盖“摘要范围”。这是您在为数据模型启用加速时选择的时间范围。在加速的数据集上运行数据透视表时，该数据透视表的时间范围必须至少一部分在此摘要范围内，才能获得加速优势。例如，如果您的数据模型是加速上个月的数据，但您使用此数据模型中在过去一年内运行的一个数据集创建了一个数据透视表，则最初在该数据透视表中在过去一个月内运行的那部分搜索会受益于加速。

构成单个数据模型的高性能分析存储的 `.tsidx` 文件始终跨一个或多个索引器进行分布。这是因为 Splunk 软件会在索引器上创建 `.tsidx` 文件，与包含该文件中引用的事件的数据桶以及涵盖摘要所涉及的时间范围的数据桶平行。

通过持续数据模型加速创建的高性能分析存储与通过临时数据模型加速创建的摘要不同。临时摘要始终创建于搜索头的 `dispatch` 目录中。

请参阅[关于临时数据模型加速](#)。

为数据模型启用持续加速

前提条件

- [数据模型加速警告](#)

步骤

1. 打开“编辑加速”对话框。使用以下各选项中的其中一项。

选项	此选项的其他步骤
导航到“数据模型”管理页面。	找到要加速的模型，并选择 编辑 > 编辑加速 。
导航到“数据模型”管理页面。	展开要加速的数据模型所对应的行，然后单击 加速 所对应的 添加 。
打开数据模型的“数据模型编辑器”。	选择 编辑 > 编辑加速 。

2. 选择**加速**为数据模型启用加速。
3. 选择**摘要范围**。

摘要范围的跨度可以是 *1 天、7 天、1 个月、3 个月、1 年或所有时间*。它代表您计划要针对数据模型中的加速数据集运行数据透视表的时间范围。例如，如果您仅希望对最近七天内的时间段运行数据透视表，请选择 *7 天*。

时间范围较小时，意味着 `.tsidx` 文件较小，构建所需的时间较少，占用的磁盘空间也较小，因此您可能希望在可行的情况下使用较短的范围。

如果您需要摘要范围与**摘要范围**字段所提供的有所不同，则您可在 `datamodels.conf` 中为您的数据模型进行配置。

请参阅[关于摘要范围](#)。

数据模型加速警告

可以加速的数据模型数据集类型有许多限制。

- 只有包含至少一个根事件层次结构或一个仅使用流命令的根搜索层次结构的数据集可以进行加速。基于使用非流命令的根搜索数据集和根交易数据集的数据集层次结构不会进行加速。
 - 使用非加速数据集的数据透视表将回退到 `_raw` 数据，这意味着最初它们运行比较慢。但是，它们可受益于临时数据模型加速中的某些加速。请参阅[关于临时数据模型加速](#)。
- 如果所加速根事件数据集和根搜索数据集的初始约束搜索中包含 Splunk 软件应对其执行搜索的索引，则最有效的加速方式是数据模型加速。单个高性能分析存储可跨越多个索引器中的几个索引。如果您知道要为其制作数据透视表的所有数据都位于一个特定索引或一组索引中，则可以通过指示 Splunk 软件要查看的位置来加速任务。否则，Splunk 软件会浪费不必要的时间来加速对您无用的数据。

有关数据模型使用情况的限制和警告的完整列表，请参阅[管理数据模型](#)。

在为数据模型启用加速之后

在为数据模型启用持续加速之后，Splunk 软件会开始为涵盖指定摘要范围的数据模型，构建数据模型加速摘要。Splunk 软件为索引中的摘要创建 `.tsidx` 文件，这些索引包含了拥有数据模型中指定字段的事件。这些 `.tsidx` 文件将与其相应的索引数据桶并行存储，存储方式与报表加速摘要相同。

在 Splunk 软件构建数据模型加速摘要之后，它会每隔 5 分钟运行一次计划的搜索，以使其保持更新。Splunk 软件每隔 30 分钟会删除一次过时的旧 `.tsidx` 摘要文件。您可以分别在 `datamodels.conf` 和 `limits.conf` 中调整这些时间间隔。

关于数据模型加速摘要有几个事实：

- Splunk 部署的每个索引中的每个数据桶都可具有一个或多个数据模型加速摘要 `.tsidx` 文件，每个拥有与数据模型摘要相关数据的加速数据模型也有一个摘要。这些摘要在数据收集过程中创建
- 摘要会限制到特定的搜索头（或搜索头池 ID），以体现可能会为相同搜索字符串生成不同结果的不同提取。
- 您只能加速已经共享给应用的所有用户的数据模型，或者全局共享给您的 Splunk 部署的所有用户的数据模型。您不能加速私有的数据模型。这可防止个别用户占用含有专用数据模型加速摘要的磁盘空间。

注意：如有必要，您可以通过 `indexes.conf` 配置数据模型加速摘要的位置。

关于摘要范围

数据模型加速摘要范围涵盖近似的时间范围。有时，数据模型加速摘要会有一个稍微超出其摘要范围的数据存储，但是摘要总是尽力符合该范围，除非是在第一次构建它的那段时期。

当 Splunk 软件完成构建数据模型加速摘要时，它的数据模型摘要过程会确保摘要始终涵盖了它的摘要范围。该过程会定期删除超出摘要范围的更旧的摘要数据。

如果您有一个与加速的数据模型数据集相关联的数据透视表，那么当您在处于数据模型的摘要范围内的时间段里运行它时，该数据透视表会最快完成。数据透视表的运行会根据数据模型加速摘要而非来源索引 `_raw` 数据。摘要中的数据远少于来源索引，这意味着数据透视表的完成速度应该比其初始运行时更快。

如果您针对摘要范围仅涵盖其中一部分的时间段运行相同的数据透视表，则数据透视表会完成得较慢。Splunk 软件

必须在索引中的来源索引 `_raw` 数据上运行至少部分的数据透视表搜索，这意味着它必须分析更大的事件集。因此，最好设置足够宽的数据模型摘要范围，以便它能捕获您计划对其运行的所有搜索。

注意：如果您具有更大的涉及多 TB 数据集的 Splunk 部署，则有一些摘要范围的高级设置可用。这会导致构建初始数据模型加速摘要所需的搜索运行时间过长，以及/或者耗费的资源更多。有关更多信息，请参阅子主题[持续加速数据模型的高级配置](#)。

摘要范围示例

您可以创建一个数据模型，并用 **7 天的摘要范围** 对其进行加速。Splunk 软件为您的数据模型构建一个大约跨越过去 7 天的摘要，然后不断维护它，定期使用新的数据进行更新并删除超过 7 天的数据。

您在处于过去一周内的时间段里运行数据透视表，那么它应该完成地相当快。但是，如果您在过去 3 到 10 天内运行相同的数据透视表，它则不会完成地如此快，即使该搜索也涵盖了 7 天的数据。只有在过去 3 到 7 天内运行的部分搜索会受益于数据模型加速摘要的运行。在过去 8 到 10 天内运行的一部分搜索运行于原始数据上，并没有进行加速。在此类情况下，Splunk 软件首先从摘要返回加速结果，然后以更慢的速度填充间隙。

在设置摘要范围值时要记住这一点。如果您时常计划针对超出过去 7 天但没有超过 30 天的时间范围运行报表，则为该报表设置数据模型加速时，应为摘要范围选择 **1 个月**。

Splunk 平台如何构建数据模型加速摘要

在为数据模型启用加速时，Splunk 软件会为该数据模型构建一组初始的 `.tsidx` 文件摘要，然后每隔 5 分钟在后台运行一次计划的搜索，以使这些摘要保持最新。每次更新都能确保摘要涵盖了所配置的整个时间范围，且数据中没有明显的间隙。这种摘要构建方法也能确保为迟来的数据构建摘要，而不会引起其他复杂情况。

并行摘要

默认情况下，数据模型加速摘要使用并行摘要。这意味着 Splunk 软件会运行两个并发搜索任务而非一个搜索任务来构建 `.tsidx` 摘要文件。它也会以 5 分钟的计划间隔来运行两个并发搜索，以维护这些摘要文件。并行摘要降低了构建和维护数据模型加速摘要所需要的时间量。

这类摘要搜索性能改善也是一项成本支出。并发搜索的数量对比 Splunk 部署可以运行的并发搜索任务的总数量，这意味着它们会导致索引器的资源使用率增加。

如果您发现默认的并行摘要设置（使用两个并发摘要构建和维护每个摘要的搜索）是一个负担，则可以通过更改存在问题的一个或多个数据模型的 `datamodels.conf` 中的设置来将其减少为单个搜索。

1. 打开 Splunk 部署中的 `datamodels.conf` 文件，该实例拥有您想要为其更新摘要设置的数据模型。
2. 查找该数据模型的段落。
3. 添加 `acceleration.max_concurrent = 1`，如果该参数未出现在段落中。

如果它出现在段落中，则将其值更改为 `1`。

4. 保存更改。

通常我们不建议将 `acceleration.max_concurrent` 增加为大于 2 的值。然而，如果您的 Splunk 部署具备进行大量搜索并发的能力，则可以尝试将所选的加速数据模型的 `acceleration.max_concurrent` 设置为 3 或更大值。

- 有关并发搜索对于搜索性能的影响的信息，请参阅《容量规划手册》中的“容纳许多同时进行的搜索”。
- 有关如何根据您的 Splunk 部署确定并发搜索限制的更多信息，请参阅“配置计划报表的优先级”。

查看摘要创建指标

创建摘要的速度取决于所涉及的事件数量以及摘要范围的大小。您可以在“数据模型”管理页面上跟踪摘要的完成进度。找到要检查的加速的数据模型，展开其所对应的行，然后查看**加速**下方显示的信息。

加速	
重建	更新
编辑	
状态	构建
访问计数	0。上次访问: 1970-01-01
	T01:00:00+01:00
磁盘上的大小	0.00MB
摘要范围	2592000
数据桶	0

状态 指示数据模型的加速摘要是否完成。如果处于 **构建** 状态，则将指示摘要已完成的百分比。数据模型加速摘要会不断地使用新数据进行更新。此时“已完成”的摘要可能会在更新新数据时返回至“构建”状态。

当 Splunk 软件计算数据模型的加速状态时，其计算会基于您为些数据模型设置的**计划窗口**。但是，如果您为此数据模型设置了回填相对时间范围，则会用此时间范围来计算加速状态。

当填充数据模型加速摘要的搜索需要特别长的运行时间时，您可以为此数据模型设置回填时间范围。请参阅[持续加速的数据模型的高级配置](#)。

验证 Splunk 平台是否正在计划摘要更新搜索

您可以验证 Splunk 软件是否正在计划搜索以更新您的数据模型加速摘要。在 `log.cfg` 中，设置 `category.SavedSplunker=DEBUG`，然后观察事件的 `scheduler.log`，例如：

```
04-24-2013 11:12:02.357 -0700 DEBUG SavedSplunker - Added 1 scheduled searches for accelerated datamodels to the end of ready-to-run list
```

当数据模型定义有更改而您的摘要尚未更新以匹配时

当您更改了一个加速数据模型的定义时，Splunk 软件需要一段时间才会更新其摘要以便反映这种更改。同时，当您运行使用该数据模型的数据透视表搜索（或 `tstats` 搜索）时，默认情况下，它不会使用旧于新定义的摘要。这确保您从该数据模型的数据透视表所获得的输出始终反映您的当前配置。

如果您知道旧数据已经“足够好”，则可以利用一个高级性能功能。该功能使用一个名为 `allow_old_summaries` 的设置（默认值为 `false`）让数据模型返回尚未进行更新以匹配该数据模型当前定义的摘要数据。

- **按搜索：**当运行从加速数据模型进行选择的 `tstats` 搜索时，设置参数 `allow_old_summaries=t`。
- **针对整个 Splunk 部署：**转到 `limits.conf`，并将 `allow_old_summaries` 参数更改为 `"true"`。

磁盘上的数据模型加速摘要大小

您可以使用“数据模型”管理页面上的数据模型指标来跟踪磁盘上数据模型摘要的总大小。摘要确实会占用一些空间，有时需要大量的空间，因此有必要避免过度使用数据模型加速。例如，您可能希望为其数据透视表在仪表板面板中大量使用的数据模型保留数据模型加速。

数据模型占用的空间量与针对所选的摘要范围收集的事件数量相关。如果数据模型包含基数很高（具有大型唯一值集）的字段，例如 `Name` 字段，这也可能会给数据模型占用的空间量带来负面影响。

如果存在特定大小限制，您可能希望测试数据模型加速摘要将占用的空间量，方法是先对小摘要范围启用加速，然后移至较大的范围（如果您认为自己可以承担得起的话）。

Splunk 平台创建和存储数据模型加速摘要的位置

默认情况下，Splunk 软件在索引器上创建每个数据模型加速摘要，与涵盖摘要所涉及的时间范围的数据桶平行，而不管落在该范围内的数据桶是热、温还是冷数据桶。如果此摘要范围内的数据桶变为冻结状态，Splunk 软件会在删除或归档该数据桶内的数据时，将对应于该数据桶的摘要信息删除。

默认情况下，数据模型加速摘要驻留在标题为 `_splunk_summaries` 的预定义的卷中，该卷位于以下路径：

```
$SPLUNK_DB/<index_name>/datamodel_summary/<bucket_id>/<search_head_or_pool_id>/DM_<datamodel_app>_<datamodel_name>
```

此卷最初没有最大大小规范，这意味着它具有无限的保存空间。

默认情况下，会将 `tstatsHomePath` 参数在 `indexes.conf` 中指定为全局设置，且仅指定一次。其路径由所有索引继承。在 `etc/system/default/indexes.conf` 中：

```
[global]
[....]
tstatsHomePath = volume:_splunk_summaries/$_index_name/datamodel_summary
[....]
```

您可以选择：

- 提供一个替代卷和文件路径作为 `tstatsHomePath` 参数的值，从而覆盖此默认文件路径。
- 为特定索引设置不同的 `tstatsHomePath` 值。
- 向任何卷（包括 `_splunk_summaries`）添加大小限制，具体做法为在卷配置中设置 `maxVolumeDataSizeMB` 参数。

请参阅 [为数据模型加速摘要配置基于大小的保存](#)中的基于大小的保存示例。

有关索引数据桶及其老化过程的更多信息，请参阅《[管理索引器和索引器群集](#)》手册中的“索引器如何存储索引”。

群集如何处理数据模型加速摘要

默认情况下，索引器群集不会复制数据模型加速摘要。这意味着只有主要数据桶副本具有关联摘要。在此默认设置下，如果主要性从数据桶的原始副本重新分配到另一个副本（例如，由于拥有主要副本的对等节点发生故障），则数据模型摘要不会移动到拥有新主要副本的对等节点。因此，它会变得不可用。直到下一次 Splunk 软件尝试更新数

据模型摘要时发现摘要丢失并重新生成之后，摘要才再次可用。

您可以配置主节点，使群集复制数据模型加速摘要。在您进行此操作后，所有可搜索数据桶副本即会具有关联摘要。**建议采用这种行为。**

请参阅《[管理索引器和索引器群集](#)》手册中的“群集如何处理报表和数据模型加速摘要”。

为数据模型加速摘要配置基于大小的保存

您为您的索引设置基于大小的保存限制以便它们不会占用太多的磁盘存储空间了吗？默认情况下，数据模型加速摘要可占用无限量磁盘空间。如果您还锁定了您的索引或索引量的最大数据大小，则这可能是个问题。但是，您可选择为您的数据模型加速摘要配置类似的保存限制。

虽然数据模型加速摘要在默认情况下不受大小限制，但它们与您索引数据桶中的原始数据关联，且与它一起老化。当汇总事件从冷数据桶传递到冻结数据桶时，这些事件会从相关摘要中删除。

重要提示：在尝试为数据模型加速摘要配置基于大小的保存前，您应了解如何使用卷跨多个索引配置索引大小限制。有关更多信息，请参阅《[管理索引器和索引器群集](#)》手册中的“配置索引大小”。

您可以执行以下步骤来为数据模型加速摘要设置基于大小的保存。所有描述的配置都在 `indexes.conf` 中进行。

1. （可选）如果您想让数据模型加速摘要结果进入卷，而不是 `_splunk_summaries`，则创建它们。

如果您想使用控制索引原始数据的现有卷，则可使该卷引用托管数据桶目录的文件系统，因为数据模型加速摘要将与它放在一起。

如果需要，您也可以将数据模型加速摘要放到它们自己的文件系统中。每卷只可引用一个文件系统，但每个文件系统可以引用多个卷。

2. 向将成为数据模型加速摘要数据的（如 `maxVolumeDataSizeMB`）主卷添加 `_splunk_summaries` 参数。

这样您就可以在索引中管理数据模型加速摘要数据的基于大小的保存。当数据模型加速摘要卷达到其最大大小时，Splunk 软件卷管理器将删除卷中最旧的摘要以腾出空间。之后，它会留下一个“已完成”文件。此“已完成”文件的出现会防止 Splunk 软件重新构建该摘要。

3. 更新索引定义。

为每个处理数据模型加速摘要数据的索引设置 `tstatsHomePath`。如果您在步骤 1 中选择一个替代卷而不是 `_splunk_summaries`，确保该路径引用那个卷。

如果您为数据模型加速摘要定义了多个卷，确保索引的 `tstatsHomePath` 设置指向相应的卷。

您可为报表加速摘要设置基于大小的保存，其设置方式与您为数据模型加速摘要设置的方式大致相同。主要差异在于报表加速摘要没有默认卷。有关管理报表加速摘要基于大小的保存的详细信息，请参阅本手册中的[“管理报表加速”](#)。

数据模型加速基于大小的保存的配置示例

此配置示例以默认的、按卷和按索引的方式，为 `_splunk_summaries` 卷上的数据模型加速摘要设置了数据大小限制。

```
#####
# Default settings
#####

# When you do not provide the tstatsHomePath value for an index,
# the index inherits the default volume, which gives the index a data
# size limit of 1TB.
[default]
maxDataSize = 1000000
tstatsHomePath = volume:_splunk_summaries/$_index_name/datamodel_summary

#####
# Volume definitions
#####

# Indexes with tstatsHomePath values pointing at this partition have
# a data size limit of 100GB.
[volume:_splunk_summaries]
path = $SPLUNK_DB
maxVolumeDataSizeMB = 100000

#####
# Index definitions
```

```
#####
```

```
[main]
homePath    = $SPLUNK_DB/defaultdb/db
coldPath    = $SPLUNK_DB/defaultdb/colddb
thawedPath  = $SPLUNK_DB/defaultdb/thaweddb
maxMemMB    = 20
maxConcurrentOptimizes = 6
maxHotIdleSecs = 86400
maxHotBuckets = 10
maxDataSize = auto_high_volume

[history]
homePath    = $SPLUNK_DB/historydb/db
coldPath    = $SPLUNK_DB/historydb/colddb
thawedPath  = $SPLUNK_DB/historydb/thaweddb
tstatsHomePath = volume:_splunk_summaries/historydb/datamodel_summary
maxDataSize = 10
frozenTimePeriodInSecs = 604800

[dm_acceleration]
homePath    = $SPLUNK_DB/dm_accelerationdb/db
coldPath    = $SPLUNK_DB/dm_accelerationdb/colddb
thawedPath  = $SPLUNK_DB/dm_accelerationdb/thaweddb

[_internal]
homePath    = $SPLUNK_DB/_internaldb/db
coldPath    = $SPLUNK_DB/_internaldb/colddb
thawedPath  = $SPLUNK_DB/_internaldb/thaweddb
tstatsHomePath = volume:_splunk_summaries/_internaldb/datamodel_summary
```

查询数据模型加速摘要

您可以在“搜索”中使用 `tstats` 命令来查询特定加速数据模型的高性能分析存储。

`tstats` 可以排序属于加速的数据模型的完整 `.tsidx` 文件摘要集，即使这些摘要分布在多个索引之间。

借此可以快速针对特定数据模型运行基于 `stats` 的搜索，查看为选定摘要范围捕获的数据是否符合您的预期。

为此，使用 `FROM datamodel=<datamodel-name>` 来标识数据模型：

```
| tstats avg(foo) FROM datamodel=buttercup_games WHERE bar=value2 baz>5
```

上述查询返回 "Buttercup Games" 数据模型加速摘要中字段 `foo` 的平均值，具体来说，其中 `bar` 为 `value2`，且 `baz` 的值大于 `5`。

注意：您不必指定数据模型的应用，因为 Splunk 软件会从搜索上下文（您所在的应用）中获取此信息。但是，您不能从应用 A 中查询应用 B 中的加速的数据模型，除非应用 B 中的数据模型处于全局共享状态。

使用 *summariesonly* 参数

`summariesonly` 参数（当用于 `tstats` 命令时）使您可以获得有关数据模型加速摘要的特定信息。

此示例使用 `summariesonly` 参数为名为 `mydm` 的加速数据模型获得摘要的时间范围。

```
| tstats summariesonly=t min(_time) as min, max(_time) as max from datamodel=mydm | eval prettymin=strftime(min, "%c") | eval prettymax=strftime(max, "%c")
```

此示例使用 `summariesonly` 和 `timechart` 以显示在一段选定的时间范围内为标题为 `mydm` 的加速数据模型汇总了什么数据。

```
| tstats summariesonly=t prestats=t count from datamodel=mydm by _time span=1h | timechart span=1h count
```

有关 `tstats` 命令的更多信息（包括用于查询普通索引数据的 `tstats` 的使用情况），请参阅《搜索参考》中的 `tstats` 条目。

启用多-Eval 来提高数据模型加速

通过许多 `Eval` 命令在数据模型遍历中针对根事件数据集进行搜索，要在数据模型加速期间完成它会是一个很耗资源的操作。您可以通过为 `limits.conf` 中的搜索启用多-Eval 计算，来提高数据模型搜索效率。

```
enable_datamodel_meval = <bool>
```



```
* Enable concatenation of successively occurring evals into a single
  comma separated eval during generation of datamodel searches.
* default true
```

如果您对于任意加速数据模型都禁用了自动重建，则在启用多-Eval 计算之后，将需要手动重建该数据模型。有关重建数据模型的更多信息，请参阅[管理数据模型](#)。

持续加速的数据模型的高级配置

在少数情况下，可能需要您为 `datamodels.conf` 中的持续加速数据模型设置高级配置。

当填充摘要的搜索的运行时间过长

如果 Splunk 部署定期处理极其大量的数据，您可能发现持续数据模型加速摘要的初始创建耗费的资源很多。构建这些摘要的搜索可能运行很长时间，从而导致它们无法汇总传入事件。要处理此情况，Splunk 软件为您提供了两个配置参数，都在 `datamodels.conf` 中。这些参数是 `acceleration.max_time` 和 `acceleration.backfill_time`。

重要提示：大多数 Splunk 用户不需要调整这些参数。默认的 1 小时 `max_time` 设置应确保长期运行的摘要创建搜索不会阻碍向摘要添加新事件。建议您不要更改这些高级的摘要范围配置，除非您知道它是解决您的摘要创建问题的唯一解决方案。

更改填充摘要的搜索可运行的最大时间段

在超过指定时间量后，`max_time` 会导致填充摘要的搜索退出。在填充摘要的搜索停止后，Splunk 软件运行搜索以获取自初始填充摘要的搜索开始后搜索到的所有事件，然后它继续向最后的填充摘要搜索停止的位置添加摘要。默认情况下，`max_time` 参数设置为 3,600 秒（60 分钟），该设置应为大多数 Splunk 部署确保正确的摘要创建。

例如：您已为数据模型启用加速，且您想要它的摘要保留过去三个月的事件。因为组织索引大量数据，最初创建此摘要的搜索应花费四小时才能完成。不幸的是，您不能让搜索运行中断这么长时间，因为它可能无法索引在这四个小时的搜索过程中搜索到的一些新事件。

`max_time` 参数在一个小时后停止搜索，之后由另一个搜索取而代之的为在这段时间内搜索到的新事件建立索引。然后，它继续运行以向摘要添加过去三个月中发生的事件。第二个搜索也在一个小时后停止，且该过程重复直到摘要完成。

注意：`max_time` 参数是大概的时间限制。60 分钟过后，Splunk 软件必须在启动摘要搜索前完成当前数据桶的汇总。这将省去不必要的工作。

设置比摘要时间范围短的回填时间范围

如果您正在使用 Splunk 部署索引大量的数据，且您不想为运行缓慢的填充摘要的搜索调整 `max_time` 范围，则您有一个替代选项：`acceleration.backfill_time` 参数。

`acceleration.backfill_time` 参数创建第二个您在摘要范围内设置的“回填时间范围”。Splunk 软件会构建一个最初仅涵盖此更短时间范围的部分摘要。之后，摘要通过每个新汇总的事件扩展，直到其达到更大摘要时间范围的限制。此时，完整摘要已完成，且不再保留脱离摘要范围的事件。

例如，假设您想设置摘要范围为 1 个月，但您知道系统会被构建该时间范围的摘要的搜索所占用。要处理此情况，您可设置 `acceleration.backfill_time = -7d` 以运行创建部分摘要（最初仅涵盖过去一周）的搜索。达到限制后，Splunk 软件将仅向摘要添加新事件，这会导致摘要涵盖的时间范围扩展。但是，完整摘要将仍然只保留事件一个月，因此，一旦部分摘要扩展到过去一个月的完整摘要范围，它开始丢弃最旧的事件，就像普通数据模型加速摘要所做的一样。

当您不希望持续加速数据模型自动重新构建时

默认情况下，每当 Splunk 软件发现持续加速数据模型过时，它会自动重新构建这些模型。当储存在 `savesearches.conf` 中的数据模型配置中的搜索不再匹配实际数据模型的搜索时，数据模型将过时。如果在磁盘上编辑加速模型的 JSON 文件时未首先禁用模型的加速，可能发生这种情况。

在非常特定的情况下，您可能想要禁用特定加速数据模型的此功能，以便这些数据模型在变为过时不会自动重新构建。相反，它将由管理员负责手动启动重新构建。通过扩展受影响的数据模型的行并单击**重新构建**，管理员可从“数据模型管理器”页面手动重新构建数据模型。

请参阅[管理数据模型](#)。

要禁用特定持续加速数据模型的自动重新构建，打开 `datamodels.conf`，找到数据模型的段落，并设置

```
acceleration.manual_rebuilds = true
```

关于临时数据模型加速

即使您所构建的数据透视表基于非持续加速的数据模型数据集，该数据透视表仍可以从“临时”数据模型加速中受益。在这些情况中，当您使用某个数据集在“数据透视表编辑器”中构建数据透视表时，Splunk 软件会在一个搜索头

dispatch 目录中构建摘要。

在您选择数据集并进入数据透视表编辑器之后，搜索头便开始构建临时数据模型加速摘要。您可以通过进度栏跟踪临时摘要的构建进度：



在进度栏显示**完成**之后，临时摘要即已构建，搜索头继续用其来更快速地返回数据透视表结果。但是仅当您在“数据透视表编辑器”中使用该数据集时，此摘要才显示。如果您退出编辑器后又返回，或者切换到其他数据集然后又返回到第一个数据集，搜索头将需要重新构建临时摘要。

临时数据模型加速摘要在更短时间范围内收集数据时完成速度更快。通过重置“数据透视表编辑器”中的时间**筛选**，您可更改根数据集及其子数据集的范围。有关更多信息，请参见下文中的“关于临时数据模型加速摘要时间范围”。

临时数据模型加速适用于所有数据集类型，包括使用转换命令的根搜索数据集和根交易数据集。与持续数据模型加速相比，临时数据模型加速的主要缺点是，使用持续数据模型加速时，摘要始终存在，数据透视表性能始终很快，直到禁用数据模型加速为止。使用临时数据模型加速时，您每次进入“数据透视表编辑器”都需要等待重新构建摘要。

关于临时数据模型加速摘要时间范围

搜索头始终尝试使临时数据模型加速摘要与由“数据透视表编辑器”中的时间**筛选**设置的范围相符合。当您首次进入数据集的“数据透视表编辑器”时，数据透视表时间范围设置为**所有时间**。如果您的数据集代表某个大型数据集，这可能意味着初始数据透视表的完成速度缓慢，因为它在后台构建临时摘要。

当您为数据透视表指定的时间范围为非**所有时间**时，搜索头会尽可能高效地构建符合该范围的临时摘要。对于任何指定的数据模型数据集，搜索头完成具有短时间范围的数据透视表的临时摘要比完成具有更长时间范围的相同的数据透视表的速度更快。

如果您使用具有不同“最晚”时间的新时间范围来替换当前时间范围，则搜索头从开始到结束仅重新构建临时摘要。这是因为搜索头反向（从最晚时间到最早时间）构建每个临时摘要。如果您保持最晚时间相同，但更改最早时间，搜索头大多数时候将收集所需的任何额外数据。

根搜索数据集及其子数据集为特殊情况，因为它们的数据透视表中不具有时间范围筛选（它们未提取 `extract time` 作为字段）。基于这些数据集的数据透视表始终为搜索返回的所有事件构建摘要。不过，您可设计根搜索数据集的搜索字符串，使其包含“最早”和“最晚”日期，这会限制由根搜索数据集及其子数据集所代表的数据集。

临时数据模型加速与持续数据模型加速有何差异

以下是临时数据模型加速与持续数据模型加速差异汇总：

- **临时数据模型加速发生在搜索头，而不是索引器。**这允许它加速所有三种数据集类型（事件、搜索和交易）。
- **Splunk 软件在搜索头的 dispatch 目录中创建临时数据模型加速摘要。**它在您的索引和索引数据桶中创建并存储持续数据模型加速摘要。
- **当您离开“数据透视表编辑器”或更改您在“数据透视表编辑器”中正在处理的数据集时，Splunk 软件将删除临时数据模型加速摘要。**当您返回到相同数据集的“数据透视表编辑器”时，搜索头必须重新构建其临时摘要。您不能保留临时数据模型加速摘要以供今后使用。
 - 数据透视表任务 ID 保留在数据透视表 URL 中，因此，如果您在离开数据透视表（或通过永久链接返回数据透视表任务）后，快速使用返回按钮，则您无需等待重新构建就可使用该任务的临时摘要。在您离开数据透视表或切换到数据透视表中的不同模型几分钟后，搜索头从 dispatch 目录中删除临时数据模型加速摘要。
- **临时加速不应用于基于数据透视表的报表或仪表板面板。**如果您想让基于数据透视表的报表和仪表板面板从数据模型加速中受益，则将它们基于来自于持续加速事件数据集层次结构的数据集。
- **临时数据模型加速在搜索头上比持续数据模型加速在索引器上潜在创建更多负载。**这是因为搜索头会为访问未持续加速的数据透视表中特定数据模型数据集的每个用户创建单独的临时数据模型加速摘要。另一方面，持续加速数据模型数据集的摘要会共享给相关联的数据模型的每个用户。此数据模型加速摘要的重复使用导致索引器的工作减少。

使用摘要索引提高报表效率

使用**摘要索引**可以高效率地针对大量数据创建报表。使用摘要索引，可以设置一个频繁运行的搜索，以提取所需的精确信息。每次搜索运行时，其结果都将保存到您指定的摘要索引中。您随后便可针对这一大小显著减少（因此速度也更快）的摘要索引运行搜索和报表。更重要的是，由于填充索引搜索的频率（例如，如果要手动运行涵盖过去七天的搜索，您可能会针对每小时更新一次的摘要索引运行这些搜索），这些报表在统计方面十分准确。

通过摘要索引，可以将计算量很大的报表成本分摊到一段时期内。在我们正在讨论的示例中，每小时运行一次的搜索只需片刻即可完成用前一小时的数据来填充摘要索引。如果不利用摘要索引，生成完整的报表大约会花费 168 小时（7 天 * 24 小时/天）。

摘要索引的一个更重要的优势或许在于：它能够将成本分摊到不同报表上以及同一报表中不同但重叠的时间范围上。在周二生成的相同摘要数据可用于周三、周四或下周一生成的前七天报表。它也可以用于需要每天平均响应大小信息的月度报表。

注意：摘要索引量不用于计算许可证，即使您有多个摘要索引。然而，发生许可证违规时，摘要索引将如任何其他非内部搜索行为一样停止下来。

摘要索引使用案例

示例 1 - 更高效地在很长的时间范围内针对大型数据集运行报表：您正在公司里使用 Splunk Enterprise，每天需要将数百万个（或者更多）事件编入索引。除了其他事情外，您希望为雇员设置一个仪表板，以显示一个按站点生成的报表，其中包含每个网站在过去 30 天内的页面视图数量以及访问者数信息。

您可以针对主要数据卷运行此报表，但其运行时间会相当长，因为 Splunk 软件需要排序大量与网站流量完全不相关的事件，以提取所需数据。但还不止如此，将报表包含在常用仪表板中还意味着此报表会经常运行，因而可能会大大延长其平均运行时间，这会让很多用户感到失望。

但是如果使用摘要索引，可以设置一个保存的搜索，按周、天甚至小时将网站页面视图和访问者信息收集到指定的摘要索引中。随后您可以针对此较小的摘要索引运行月末报表，此报表的完成速度应比没有此摘要索引时快得多，因为它是针对更小的、更清晰的数据集进行搜索。

示例 2 - 构建滚动的报表：假设您要运行一个报表，以显示在一段很长的时间段内某个汇总统计信息的实时计数，例如您所管理网站的文件下载实时计数。

首先，安排一个保存的搜索返回某个指定时间段内的下载总数。然后，使用摘要索引，将该搜索的结果保存到摘要索引中。随后可以根据需要针对摘要索引中的数据随时运行报表以获取下载总次数的最新计数。

要了解其他信息，可以观看此有关摘要索引理论和实践的 Splunk 开发人员视频。

使用摘要索引报表命令

如果您并不熟悉摘要索引，请在定义要摘要索引填充搜索时使用摘要索引报表命令（`sichart`、`sitimechart`、`sistats`、`sitop` 和 `sirare`）。如果使用这些命令，可以使用在最终对摘要索引运行的搜索中使用的同一搜索字符串，除非您在对摘要索引运行的搜索中使用了常规报表命令。

注意：如果您很熟悉以“传统”方式创建摘要索引填充搜索，则不必使用 `si` 摘要索引搜索命令。如果您使用这些方法创建了摘要索引，并且这些索引可以正常使用，则没有必要更新它们。实际上，它们可能更高效：使用 `si` 命令不会对性能带来影响，因为它们创建的索引只比“手动”方法创建的索引稍大一点儿。

在多数情况下，影响不明显，但是如果您创建的摘要索引本身相当大，可能会看到差别。如果您设置了多个搜索，以便针对由 `si` 命令搜索填充的索引创建报表，则可能会发生性能问题。

如果您对不使用 `si` 搜索命令设计摘要索引很感兴趣，请参见下面的部分。

不使用特殊命令定义填充索引的搜索

在以前的 Splunk Enterprise 版本中，对于如何设计用于摘要索引填充搜索您必须十分小心，尤其是在要对已完成摘要索引运行的搜索调用了汇总统计信息时，因为这意味着您必须十分小心地设置“索引填充”搜索，以便提供正确的结果。例如，如果您想对已完成的摘要索引运行一个搜索，以提供服务器发出的平均响应时间，您需要将摘要索引填充搜索设置成：

- 安排为以更高的频率运行（与计划针对摘要索引运行的搜索相比）。
- 对更多的数据进行取样（与计划针对摘要索引运行的搜索相比）。
- 包含其他搜索命令，以确保填充索引的搜索生成加权平均值（仅在首先查找平均值时需要）。

摘要索引报表命令会为您完成后两点，它们会自动确定需要做出的调整，以确保使用不会生成不准确统计结果的数据来填充摘要索引。但是，您仍应将摘要索引填充搜索的运行频率安排成高于以后针对摘要索引进行搜索的运行频率。

您是否对不使用 `si` 命令设置摘要索引感兴趣？请了解 `addinfo`、`collect` 和 `overlap` 命令，学习如何设计提供加权平均值的搜索，并查看本手册主题“[配置摘要索引](#)”中的通过 `savedsearches.conf` 配置摘要索引的示例。

摘要索引报表命令使用示例

假设您正在运行下列搜索，时间范围为过去一年：

```
eventtype=firewall | top src_ip
```

此搜索可提供过去一年排名前几位的来源 IP，但它却无休止地运行，因为它每次都对整个索引进行扫描。

您需要创建一个由来自“防火墙”事件类型的前几个来源 IP 组成的摘要索引。您可以使用下列搜索来构建该摘要索引。您计划每天运行该搜索，每次仅收集前 24 小时内的前几个 `src_ip` 值。由每个每日运行一次的搜索产生的结果将添加到一个名为“摘要”的索引中：

```
eventtype=firewall | sitop src_ip
```

注意：如果您将运行摘要索引填充搜索并对信息进行取样的频率安排成比计划针对已完成摘要索引进行的搜索的频率高，则这些搜索在统计方面的准确度会更高。在本示例中，因为计划运行的搜索涵盖了过去一年的时间跨度，所以设置了一个每天都对信息进行取样的摘要索引填充搜索。

重要提示：在定义摘要索引填充搜索时，不要在主摘要索引报表命令后使用其他搜索运算符。换句话说，不要包含其他 `| eval` 命令及类似内容。将其他搜索运算符留给 *针对*摘要索引运行的搜索，而不是要用来填充摘要索引的搜索。

重要提示：经过摘要索引优化过的搜索所生成的结果将以特殊格式存储，在执行最终转换之前无法修改。这意味着，如果您使用 `... | sistsats <args>` 填充摘要索引，唯一有效的数据检索是：`index=<summary> source=<saved search name> | stats <args>`。针对摘要索引的搜索不能创建或修改 `| stats <args>` 命令之前的字段。

现在，假设您使用名称 "Summary - firewall top src_ip" 保存此搜索（所有保存的摘要索引填充搜索都应使用这样的名称来标识）。在摘要索引中填充了结果后，将使用指定了摘要索引以及用于填充此摘要索引的搜索名称的搜索，针对摘要索引进行搜索并创建报表。例如，您可使用下面的搜索来获得去年排名前几位的 `source_ip`：

```
index=summary search_name="summary - firewall top src_ip" |top src_ip
```

因此此搜索指定了搜索名称，所以它会筛选掉由其他摘要索引搜索填入到摘要索引中的其他数据。此搜索会运行得相当快，即使时间范围为一年或更长。

注意：如果针对摘要索引运行一个搜索，以查询具有特定 `sourcetype` 值的事件，注意此时需要改为使用 `orig_sourcetype`。所以，不针对摘要索引运行诸如 `...|stats timechart avg(ip) by sourcetype` 的搜索，而应使用 `...|stats timechart avg(ip) by orig_sourcetype`

为什么需要这样做呢？当事件被收集到摘要索引中时，其 `sourcetype` 值会更改为 "stash"，而原始来源类型值也会移到 `orig_sourcetype`。

在 Splunk Web 中设置摘要索引搜索

可以通过 Splunk Web 设置摘要索引搜索。对于计划的报表，摘要索引是一个告警选项。决定要用于填充摘要索引的报表后，请遵循下列步骤：

1. 导航到 **设置 > 搜索、报表和告警**。
2. 选择报表的名称（或单击**新建**创建一个报表）。
3. 在“计划和告警”下，如果还没有为报表排定计划，则选择**计划此搜索的时间**。

您必须选择**计划此搜索**以查看报表计划选项。

4. 将报表安排为按适当的时间间隔运行。

摘要索引填充搜索应以非常高的频率运行，这样才能创建在统计上十分准确的最终报表。如果针对摘要索引运行的报表要收集过去一周的信息，您应让摘要报表每小时运行一次，以收集每小时内的信息。如果要针对过去一年的数据运行报表，您可能会让摘要索引每天收集一次过去一天的数据。有关更多信息，请参阅《报表手册》中的“计划报表”。

注意：请务必妥善计划报表的时间，以确保数据中不存在间隙和重叠。有关这方面的更多信息，[请参见下面有关此问题的子主题](#)。

5. 在告警下，将**条件**设置为**始终**。
6. 将**告警模式**设置为**每次搜索一次**。

这可确保每次运行报表时都触发告警。

摘要索引

☒ 启用

Enabling summary indexing will set the alert condition to 'always'.

选择摘要索引

summary

只列出您可以写入的索引。

添加字段

report	=	summary_top_source_ip	删除
--------	---	-----------------------	----

[添加另一个字段](#)

7. 在“摘要索引”下，选择**启用**。

当您选择**启用**时，告警**条件**会设置为**始终**，而**告警模式**设置为**每次搜索一次**。在未禁用摘要索引时，您不可选

择这些字段的其它值。

8. 从选择摘要索引列表选择报表填充的摘要索引的名称。

默认的摘要索引名称为 *摘要*。此列表仅显示您具有写入权限的索引。

9. (可选) 如果您计划运行多种摘要索引报表, 则可能需要创建其他摘要索引。

有关创建新索引的信息, 请参见《管理索引器和群集》手册中的“创建自定义索引”。建议创建专门用于收集摘要数据的索引。

10. (可选) 在添加字段下, 可以将字段/值对添加到摘要索引定义中。

这些键/值对作为批注附加到每个已建立摘要索引的事件中。这样将方便以后通过搜索进行查找。例如, 您可以添加填充摘要索引的报表的名称 (*report=summary_firewall_top_src_ip*) 或报表填充的索引的名称 (*index=summary*), 以后便可以针对这些术语进行搜索。

注意: 您也可以将字段/值对添加到 `savedsearches.conf` 中的摘要索引配置中。有关更多信息, 请参阅《知识管理手册》中的[“配置摘要索引”](#)。

有关将搜索保存为报表和告警的更多信息, 请参阅: “创建和编辑报表” (位于《报表手册》) 和《告警手册》。

计划填充报表的时间以避免数据出现间隙和重叠

为了将数据间隙和重叠减至最少, 您应确保在用来填充摘要索引的报表的计划中设置适当的时间间隔和延迟。

摘要索引中的 *间隙* 是指摘要索引无法将事件编入索引的时间段。在以下情况下可能会出现间隙:

- **填充摘要索引的报表的运行时间过长, 以至于超过了下一个计划运行时间。**例如, 如果您将填充摘要的报表安排为每 5 分钟运行一次, 当该报表一般会花费大约 7 分钟来运行时, 就会出现间隙, 因为系统不会在前一个报表正在运行时, 再次运行搜索。
- **您已强制填充摘要索引的报表使用实时计划。**因为您错误地更改了 `savedsearches.conf` 中的报表定义, 将 `realtime_schedule` 属性设置为 1, 从而启用了实时计划。如果您正在同时运行数个报表, 此设置可能会导致数据集中出现间隙。通过为摘要索引选择 **启用** 并保存报表来在 Splunk Web 中定义填充摘要索引的计划报表时, `realtime_schedule` 会设置为 0, 以确保报表从不跳过计划的运行。有关更多信息, 请参阅《报表手册》中的“配置计划的报表优先级”。
- **splunkd 发生故障。**如果 Splunk Enterprise 无法将事件编入索引, 摘要索引中就会出现间隙。

当摘要索引中的事件 (来自同一报表) 共享相同时间戳时就会发生 **重叠**。重叠的事件会使通过摘要索引创建的报表和统计信息变得不准确。如果您将报表的时间范围设置得比报表计划的频率长时, 就可能发生重叠。换句话说, 不要安排每小时运行一次的报表来收集过去 90 分钟内的数据。

注意: 有关检测和修复重叠数据和数据间隙的信息, 请参阅本手册中的[“管理摘要索引间隙和重叠”](#)。

摘要索引如何工作

在 Splunk Web 中, 摘要索引是已保存计划搜索的一个告警选项。在启用了摘要索引的情况下运行保存的搜索时, 其搜索结果会临时存储在一个文件 (`$SPLUNK_HOME/var/spool/splunk/<savedsearch_name>_<random-number>.stash`) 中。在此文件中, Splunk 软件会使用 `addinfo` 命令将有关当前搜索的常规信息以及在配置期间指定的字段添加到每个结果中。Splunk Enterprise 随后会将您已为其指定的摘要索引 (默认情况下是 `index=summary`) 中的结果事件数据编入索引。

注意: 可使用 `addinfo` 命令将包含当前搜索常规信息的字段添加到搜索结果, 进而编入到摘要索引中。所添加的有关搜索的常规信息将有助于您针对摘要索引中的结果运行报表。

为没有时间戳的数据建立摘要索引

要为摘要索引事件设置时间, Splunk 软件将按以下优先顺序使用下列信息:

1. 要为其建立摘要的事件的 `_time` 值。
2. 填充摘要索引的计划的搜索的最早 (或最小) 时间。例如, 如果填充摘要索引的搜索涵盖在其搜索每次启动前的两分钟, 其最早时间为 `-2m`。
3. 当前系统时间 (如果是“所有时间”搜索, 则没有指定的“最早”值)

在大多数情况下, 您的事件都具有时间戳, 所以首要方法是识别摘要索引的时间戳。但是如果您要为其建立摘要的数据 (如来自查找的数据) 不包含 `_time` 字段, 结果事件的时间戳将为填充摘要索引的搜索的最早时间。

例如, 如果您每天子夜为查找 `asset_table` 建立摘要, 并且此资产表不包含 `_time` 列, 则今天晚上的摘要所包含的 `_time` 值将等于搜索的最早时间。如果已将搜索的时间范围设置在 `-24h` 和 `+0s` 之间, 则对于每个建立摘要的事件, 其 `_time` 值均为 `now()-86400` (即搜索的开始时间减去 86,400 秒, 也就是 24 小时)。这表示此摘要索引填充搜索找到的每个没有 `_time` 字段值的事件都会被分配确切的相同 `_time` 值: 搜索的最早时间。

为没有时间戳的数据建立摘要的最佳做法是在搜索中手动创建一个 `_time` 值。继续使用上面的示例：

```
|inputlookup asset_table | eval _time=now()
```

通过 `si` 摘要索引命令添加到摘要索引数据的字段

警告：通过 `si` 摘要索引命令以外的任何搜索命令，使用这些字段和其编码的数据都是不支持的。这些字段的格式和内容可以随时更改，而不会发出警告。

使用 `si` 命令运行搜索以填充摘要索引时，Splunk 软件会向摘要索引数据添加一组特殊字段，这些字段全部以 `psrsvd` 开头，如 `psrsvd_ct_bytes` 和 `psrsvd_v` 等。当您使用 `chart`、`timechart` 和 `stats` 等报表命令针对摘要索引运行搜索时，会使用 `psrsvd*` 字段来为表格和图表计算正确的统计结果。`psrsvd` 代表 "prestats reserved"（预留的 `prestat`）。

大多 `psrsvd` 类型都会提供数据集中原始（在建立摘要索引前）文件中某个特定字段的相关信息，但某些 `psrsvd` 类型并不仅限于一个字段。常规模式为 `psrsvd_[type]_[fieldname]`。例如，`psrsvd_ct_bytes` 可提供 `bytes` 字段的计数信息。

下面是可用的 `psrsvd` 类型列表：

- `ct` = 计数
- `gc` = 组计数（stats“分组”的计数，不仅限于一个字段。
- `nc` = 数字计数（数字值的个数）
- `nn` = 最小数字值
- `nx` = 最大数字值
- `rd` = `rdigest` 值（值出现的次数）
- `sm` = 总和
- `sn` = 最小字典值
- `ss` = 平方和
- `sx` = 最大字典值
- `v` = 版本（不仅限于一个字段）
- `vm` = 值映射（字段的所有非重复值和它们出现的次数）
- `vt` = 值类型（包含关联字段的精度）

管理摘要索引间隙

如果调用的摘要索引在其所收集的数据中有间隙，则摘要索引搜索的精确度可能会受到负面影响。

摘要索引数据中的间隙可能由多种原因导致：

- **摘要索引最初仅包含您开始收集数据那个时间点的事件：**不要忽视摘要索引不包含摘要索引收集开始日期之前的数据这一事实，除非您使用回填脚本来自将相应数据放在索引中。
- **Splunk 部署中断：**如果 Splunk 部署中断了很长时间，则摘要索引数据中很有可能出现间隙，这取决于填充索引的搜索的计划运行时间。
- **搜索的运行时间比其计划间隔长：**如果用于填充计划搜索的搜索运行时间长于为其计划的运行间隔时间，则很有可能出现间隙，因为 Splunk 软件不会在前一个搜索仍在运行时再次运行计划搜索。例如，如果您安排填充索引的搜索每五分钟运行一次，当此搜索的运行时间长于五分钟时，您的索引数据集中将会出现间隙。

注意：有关创建和维护摘要索引的常规信息，请参见《知识管理器》手册中的[“使用摘要索引提高报表效率”](#)。

使用回填脚本添加其他数据或填充摘要索引间隙

如果您有 Splunk Enterprise，则可以使用 `fill_summary_index.py` 脚本。该脚本可通过运行填充摘要索引的已保存搜索，就像它们按其固定计划时间针对给定时间范围执行那样，回填摘要索引集合中的间隙。换句话说，即使新的摘要索引仅在本周初才开始收集数据，在必要时也可以使用 `fill_summary_index.py` 上个月的数据填充摘要索引。

另外，在运行 `fill_summary_index.py` 时，可以指定一个应用，并为与该应用关联的一系列摘要索引搜索计划回填操作，或仅选择回填与该应用关联的所有已保存搜索。

通过 CLI 输入 `fill_summary_index.py` 命令时，必须为回填操作指定“最早时间”和“最晚时间”，以提供回填时间范围。可以通过使用相对时间标识符（如代表“3 天前午夜”的 `-3d@d`）或通过使用 UTC epoch 数字指定精确的时间。此脚本会自动计算摘要索引搜索将在其中运行的时间范围的时间。

注意：为了确保 `fill_summary_index.py` 脚本仅偶尔执行与丢失的数据相对应的摘要索引搜索，您必须在调用它时使用 `-dedup trueo`

`fill_summary_index.py` 脚本要求您提供必要的验证（用户名和密码）。如果您在调用该脚本时知道有效的 Splunk Enterprise 密钥，可以通过 `-sk` 选项跳过它。

该脚本用于提示您输入您未在命令行中提供的必需信息，包括摘要索引搜索的名称、验证信息以及时间范围。

fill_summary_index.py 调用示例

如果您的情况如下：

您需要为 splunkdotcom 应用回填过去一个月的所有摘要索引搜索，但您还需要跳过在摘要索引中已有数据的任何搜索：

可以将下列内容输入到 CLI：

```
./splunk cmd python fill_summary_index.py -app splunkdotcom -name "*" -et -mon@mon -lt @mon -dedup true -auth admin:changeme
```

如果您的情况如下：

您需要回填过去一年的 my_daily_search 摘要索引搜索，在任何给定时间运行的并发搜索不能超过 8 个（以便在系统收集回填数据时减少对性能的影响）。您不希望脚本跳过在摘要索引中已有数据的搜索。my_daily_search 摘要索引搜索由“管理员”角色所有。

可以将下列内容输入到 CLI：

```
./splunk cmd python fill_summary_index.py -app search -name my_daily_search -et -y -lt now -j 8 -owner admin -auth admin:changeme
```

注意：您需要为由特定用户或角色所有的搜索指定 -owner 选项。

如果 fill_summary_index.py 在运行时中断怎么办

如果 fill_summary_index.py 中断，在您从其调用此过程的应用（例如 Search）中查找 log 目录。在此目录，您应该能看到一个名为 fsidx*lock 空临时文件。

删除此临时文件，然后您应该就可以重新启动 fill_summary_index.py。

fill_summary_index.py 用法和命令

在 CLI 中，开始先输入以下内容：

```
python fill_summary_index.py
```

...并从上表中添加必需的以及可选的字段。

注意：<boolean> 选项接受以下值：1、t、true，或表示 "true" 的 yes 和 0、f、false，或表示 "false" 的 no。

字段	值
-et <string>	最早时间（必需）。输入 UTC 时间或相对时间字符串。
-lt <string>	最晚时间（必需）。输入 UTC 时间或相对时间字符串。
-app <string>	要使用的应用上下文（默认为 None）。
-name <string>	指定一个已保存搜索的名称。可以指定多次以提供多个名称。使用通配符 ("*") 可指定所有具有摘要索引操作的已启用、已保存的计划搜索。
-names <string>	指定一个以逗号分隔的已保存搜索名称列表。
-namefile <filename>	指定一个包含一列已保存搜索名称（一个名称占一行）的文件。以 # 开头的行被视为注释行，将被忽略。
-owner <string>	要使用的用户上下文（默认为 "None"）。
-index <string>	标识已保存搜索要填充的摘要索引。如果未提供索引，回填脚本会自动尝试确定它。如果尝试自动检测索引失败，则索引默认为“摘要”。
-auth <string>	验证字符串可以为 <username> 或 <username>:<password>。如果仅提供了用户名，脚本会以交互方式要求输入密码。
-sleep <float>	在搜索之间的休眠秒数。默认为 5 秒。
-j <int>	可以运行的最多并发搜索数（默认为 1）。
-dedup <boolean>	当此选项设置为 true 时，如果数据已存在于该时间跨度的摘要索引中，则脚本不会在计划时间跨度中运行保存的搜索。默认情况下，此选项设置为 false。 注意： 该选项与搜索语言中的 dedup 命令没有关系。脚本没有执行事件级别数据分析的能力。它不能确定某些事件是否与其他事件重复。
-nolocal <boolean>	如果您正在分布式环境中工作，指定摘要索引不在搜索头上，而是在索引中。要与 -dedup 结合使用。
-showprogress <boolean>	当此选项设置为真时，脚本会定期显示它所衍生的每个当前正在运行的搜索的完成进度。如果未使用此选项，则其默认值为假。

高级选项：几乎在任何情况下，均不应使用这些选项。	
-trigger <boolean>	当此选项设置为假时，脚本会运行每个搜索，但不触发摘要索引操作。如果未使用此选项，则其默认值为真。
-dedupsearch <string>	指示用于确定在特定计划时间对应于特定已保存搜索的数据是否存在的搜索。
-distdedupsearch <string>	与 -dedupsearch 相同，除了这是一个分布式搜索字符串。它不会将其范围限制在搜索头。它也在索引器上查找摘要数据。
-namefield <string>	指示摘要索引数据中包含生成该数据的已保存搜索名称的字段。
-timefield <string>	指示摘要索引数据中包含生成该数据的已保存搜索的计划时间的字段。

配置摘要索引

有关摘要索引的一般概述以及通过 Splunk Web 设置摘要索引的说明，请参见《知识管理器手册》中的主题[“使用摘要索引提高报表效率”](#)。

只有将保存的报表设置为按固定时间间隔运行、每次运行时都触发并且选中了**启用摘要索引**告警选项的计划报表，您才能在 `savedsearches.conf` 中为保存的报表手动配置摘要索引。

另外，您还需要输入报表将填充的摘要索引的名称。在选择**启用摘要索引**之后，可通过报表的详细信息页面（**设置 > 搜索和报表**）完成此操作。*摘要索引*是默认的摘要索引（如果您没有指定其他摘要索引，Splunk Enterprise 将使用此索引）。

如果您计划运行多种摘要索引报表，则可能需要创建其他摘要索引。有关创建新索引的信息，请参见《管理索引器和群集》手册中的“创建自定义索引”。建议创建专门用于收集摘要数据的索引。

摘要索引量不用于计算许可证，即使您有若干个摘要索引。发生许可证违规时，摘要索引将如任何其他非内部搜索行为一样停止下来。

注意：如果输入了不存在的索引名称，Splunk Enterprise 将按您定义的计划运行此报表，但它不会将其数据保存到摘要索引中。

有关创建和管理报表的更多信息，请参阅本手册中的“创建和编辑报表”。

有关定义可填充摘要索引的报表的更多信息，请参阅本手册[“使用摘要索引提高报表效率”](#)中有关通过 Splunk Web 设置摘要索引报表的子主题。

注意：在定义要用来构建索引的报表时，大多数情况下您应在报表的搜索字符串中使用[摘要索引转换命令](#)。这些命令以 "si" 为前缀：`sichart`、`sitimechart`、`sistats`、`sitop` 和 `sirareo`。使用这些命令创建的报表应是您最终用来查询已完成摘要索引的报表版本。

摘要索引转换命令会自动考虑下面“摘要索引报表定义注意事项”中所涵盖的问题，例如为填充报表计划更短的时间范围以及将填充报表设置为以更大规模进行取样。仅当用来构建索引的报表不包含摘要索引转换命令时，您才需要考虑这些问题。

如果不使用摘要索引转换命令，您可以使用 `addinfo` 和 `collect` 搜索命令来创建 Splunk Enterprise 保存并为其设置计划、用于填充预先创建的摘要索引的报表。有关该方法的更多信息，请参见本主题中的“手动填充摘要索引”。

为计划的报表自定义摘要索引

使用 Splunk Web 为已启用摘要索引的计划的报表启用摘要索引时，Splunk Enterprise 会自动在 `$SPLUNK_HOME/etc/system/local/savedsearches.conf` 中生成一个段落。您可以通过编辑此段落为报表自定义摘要索引。

如果您已使用 Splunk Web 来保存报表并为其设置计划，但没有使用 Splunk Web 为报表启用摘要索引，只要有新的索引供报表填充，就可以轻松通过 `savedsearches.conf` 为其启用摘要索引。有关手动索引配置的更多信息，请参见《管理索引器和群集手册》中的主题“关于管理索引”。

```
[ <name> ]
action.summary_index = 0 | 1
action.summary_index._name = <index>
action.summary_index.<field> = <value>
```

- `[<name>]`：Splunk Enterprise 会基于您为摘要索引启用的计划的报表的名称为段落命名。
- `action.summary_index = 0 | 1`：设置为 1 时将启用摘要索引。设置为 0 时将禁用摘要索引。
- `action.summary_index._name = <index>` - 用于显示此报表填充的摘要索引的名称。如果已为此报表创建了一个特定的摘要索引，请在 `<index>` 中输入其名称。默认为 `summary`，即 Splunk Enterprise 所附带的摘要索引。
- `action.summary_index.<field> = <value>`：指定将添加到每个由此报表建立摘要索引的事件的字段/值对。可为一

个摘要索引报表定义多个字段/值对。

此字段/值对可以作为分类的“标记”，当您针对更大规模的事件数据集运行报表时，通过它可以更方便地识别将编入摘要索引的事件。此关键字是可选的，但建议您绝不要设置一个字段/值对也没有的摘要索引。

例如，添加用于填充摘要索引的报表的名称 (`action.summary_index.report = summary_firewall_top_src_ip`) 或者报表填充的索引的名称 (`action.summary_index.index = search`)。

可用于摘要索引的搜索命令

摘要索引使用了一组专用的转换命令，如果您要在没有 Splunk Web 界面或[摘要索引转换命令](#)的帮助下手动创建摘要索引，则需要使用这些命令。

- `addinfo`：摘要索引将使用 `addinfo` 将包含当前报表常规信息的字段添加到报表结果中，进而编入到摘要索引中。将 `| addinfo` 添加到任意报表即可看到结果在编入到摘要索引后的大致外观。
- `collect`：摘要索引将使用 `collect` 将报表结果编入到摘要索引中。使用 `| collect` 可将任意报表结果编入到其他索引中（使用 `collect` 命令选项）。
- `overlap`：可使用 `overlap` 来识别摘要索引中的间隙和重叠。`overlap` 可在摘要索引中查找具有重叠时间戳值的相同 `query_id` 的事件，或识别丢失了事件的时间段。

手动配置用于填充摘要索引的报表

如果您希望不使用 Splunk Web 中的报表选项对话框和[摘要索引转换命令](#)配置摘要索引，您必须先配置一个摘要索引，就像通过 `indexes.conf` 配置任何其他索引那样。有关手动索引配置的更多信息，请参见《管理索引器和群集手册》中的主题“关于管理索引”。

重要提示：您必须重新启动 Splunk Enterprise，`indexes.conf` 中的更改才能生效。

1. 设计要从 Splunk Web 为结果建立摘要的搜索字符串。

- 务必要限制报表的时间范围。报表生成的结果数量必须在您已为报表设置的最大报表结果限制以内。
- 确保选择一个适用于您的数据的时间间隔，如 10 分钟、2 小时或 1 天。（有关使用 Splunk Web 计划报表时间间隔的更多信息，请参阅《报表手册》中的“为报表设定计划”主题。）

2. 使用 `addinfo` 搜索命令。将 `| addinfo` 附加到报表的搜索字符串结尾。

- 此命令将有关报表的信息添加到 `collect` 命令为了将其编入到摘要索引中而要求的事件中。
- 可以始终向任意搜索字符串添加 `| addinfo`，以预览其结果在摘要索引中的大致外观。

3. 将 `collect` 搜索命令添加到报表的搜索字符串。将 `|collect index=<index_name> addtime=timestamp="report_name=\"<summary_report_name>\""` 附加到搜索字符串结尾。

- 用摘要索引的名称替换 `index_name`。
- 用关键字替换 `summary_report_name`，以在索引中查找此报表的结果。
- 如果您希望对生成的事件使用 `overlap` 搜索命令，则*必须*设置 `summary_report_name`。

注意：对于一般情况，建议您使用提供的 `summary_index` 告警操作。通过 `addinfo` 和 `collect` 进行配置需要一些额外的步骤，从计划的报表生成摘要索引事件时则不需要这些步骤。当您回填已确定的时间范围内的摘要索引时，必须进行手动配置。

摘要索引报表定义注意事项

如果出于某种原因，您要设置不使用[摘要索引转换命令](#)的填充摘要索引的报表，则应花些时间计划您的方法。使用摘要索引，可以先有蛋再有鸡。可使用您实际上要运行并查看结果的报表来帮助定义要用来填充摘要索引的报表。

当 `main` 索引每天增加数百万事件时，许多摘要搜索报表都会涉及汇总统计信息，例如用来搜索过去一天内与防火墙攻击相关的前 10 个 IP 地址的报表。

如果您使用针对摘要索引运行的同一报表的结果填充摘要索引，则可能会得到不准确的统计结果。在定义用于填充摘要索引的报表时应遵循这些规则，以提高由摘要索引报表生成的汇总统计信息的准确度。

为填充报表计划更短的时间范围

计划用于填充摘要索引的报表，以便与最终针对索引运行的报表相比，运行的间隔时间更短（因此更频繁）。您应使用可能的最短时间范围。例如，如果您每天都需要生成一个“前几位”报表，则填充摘要索引的报表应每小时取样一次。

将填充报表设置为以更大规模进行取样

与您要针对摘要索引运行的报表相比，用于填充摘要索引的报表应找出一个更大规模的样本。例如，如果您计划在摘要索引中搜索每日前 10 个攻击 IP 地址，则可设置一个报表，以每小时前 100 个攻击 IP 地址填充摘要索引。

此方法有两个好处：它可确保前 10 报表具有更高的统计准确度（因为取样总体上规模更大也更频繁了），并且如果您决定要针对前 20 或 30 个攻击 IP 建立报表，这会给您留有余地。

与运行用来查询已完成摘要索引的报表相比，[摘要索引转换命令](#)可自动以更大的规模进行取样，因此创建摘要索引所使用的事件数据不会不正确地影响索引准确性。如果您没有使用这些命令，则可以使用 head 命令为填充摘要索引的报表选择一个比针对摘要索引运行的报表所用样本更大的样本。换句话说，对于每小时运行一次的填充摘要索引的报表，可以让 | head=100；对于每天运行一次的已完成摘要索引的报表，可以让 | head=10。

设置报表以获得加权平均值

如果填充摘要索引的报表涉及平均值，并且您没有使用[摘要索引转换命令](#)，则需要将该报表设置成获取加权平均值。

例如，假设您要构建基于小时、天或周的平均响应时间报表。要这样做，您需通过将“每小时平均值”加在一起取平均值来生成“每日平均值”。不幸的是，如果每个“每小时平均值”中的事件数不等，则每日平均值会变得不准确。您可以通过使用加权平均值函数来获得正确的“每日平均值”。

下列表达式通过使用 stats 和 eval 命令并结合 sum 统计聚合器，使用加权平均值计算出每日平均响应时间。在本示例中，eval 命令创建了一个 daily_average 字段，该字段值是用平均响应时间总和除以平均响应时间计数所得的结果。

```
| stats sum(hourly_resp_time_sum) as resp_time_sum, sum(hourly_resp_time_count) as resp_time_count | eval
daily_average= resp_time_sum/resp_time_count | .....
```

为填充摘要索引的报表计划时间以避免数据出现间隙和重叠

除上面的两个规则外，要将数据间隙和重叠减少到最少，您还应确保在用来填充摘要索引的报表计划中设置适当的时间间隔和延迟。

摘要索引中的间隙是指摘要索引无法将事件编入索引的时间段。在以下情况下可能会出现间隙：

- splunkd 发生故障。
- 已保存的计划的报表（正在建立摘要索引的报表）花费了太长的时间来运行，以至于覆盖了下一个计划运行时间。例如，如果您将填充摘要的报表计划为每 5 分钟运行一次，当该报表一般会花费大约 7 分钟来运行时，就会出现间隙，因为系统不会在前一个报表正在运行时，再次运行报表。

当摘要索引中的事件（来自同一报表）共享相同时间戳时就会发生重叠。重叠的事件会使通过摘要索引创建的报表和统计信息变得不准确。如果您将保存的报表的时间范围设置得比报表的计划间隔时间长，或者您手动使用 collect 命令运行摘要索引，就可能发生重叠。

摘要索引配置示例

本示例显示了 Apache 服务器统计信息摘要索引的配置，此配置可能会显示在 savedsearches.conf 中。下面列出的关键字为报表 "Apache Method Summary" 启用了摘要索引。

注意：如果设置了 action_summary.index=1，则不需要在报表的搜索字符串中使用 addinfo 或 collect 命令。

```
#name of the report = Apache Method Summary
[Apache Method Summary]
# sets the report to run at each interval
counttype = always
# enable the report schedule
enableSched = 1
# report interval in cron notation (this means "every 5 minutes")
schedule = */5****
# id of user for report
userid = jsmith
# search string for summary index
search = index=apache_raw startminutesago=30 endminutesago=25 | extract auto=false | stats count by method
# enable summary indexing
action.summary_index = 1
#name of summary index to which report results are added
action.summary_index._name = summary
# add these keys to each event
action.summary_index.report = "count by method"
```

受摘要索引影响的其他配置文件

除在 savedsearches.conf 中配置的设置外，indexes.conf 和 alert_actions.conf 中也有用于摘要索引的设置。

Indexes.conf 可指定摘要索引的索引配置。Alert_actions.conf 可控制与报表关联的告警操作（包括摘要索引）。

警告：没有 Splunk 技术支持人员的明确指示，不要编辑 `alert_actions.conf` 中的设置。

配置批处理模式搜索

以批处理模式运行的搜索，分批地一次搜索一个数据桶，而不是基于时间通过事件搜索。符合批处理模式处理要求的**转换搜索**会比不符合的完成地更快。

批处理模式搜索还可以提高长时间运行的**分布式搜索**的可靠性，当**索引器**在搜索运行期间关闭时，可能无法进行这种分布式搜索。在这种情况下，Splunk 软件会尝试重新连接到丢失的**对等节点**或在其余对等节点之间重新分布搜索，以完成搜索。

默认启用批处理模式搜索功能。有关配置或禁用批处理模式搜索的信息，请参阅本主题中的[“在 limits.conf 中配置批处理模式搜索”](#)。

您甚至可以通过启用批处理模式搜索并行化，来使您的批处理模式搜索变得更快。使用了批处理模式搜索并行化，则对于符合条件的搜索会启动两个或多个搜索管道，并且它们会并发地处理搜索结果。请参阅本主题中的[“配置批处理模式搜索并行化”](#)。

批处理模式搜索要求

满足下列条件的转换搜索可在批处理模式下运行。

- 搜索需要使用**生成命令**，例如 `search`、`loadjob`、`datamodel`、`pivot` 或 `dbinspect`。
- 搜索可以包括转换命令，例如 `stats`、`chart` 等等。然而搜索不能包括如 `localize` 和 `transaction` 一类的命令。
- 如果搜索是非分布式的，则不能使用要求事件按时间排序的命令，例如 `streamstats`、`head` 和 `tail`。

通过使用**搜索任务查看器**来确认搜索是否以批处理模式运行。批处理模式搜索可通过布尔参数 `isBatchModeSearch` 来判断。请参阅《搜索手册》中的“查看搜索任务属性”。

在 limits.conf 中配置批处理模式搜索

如果您有 Splunk Enterprise 部署（而非 Splunk Cloud），您可以更改 `limits.conf` 配置文件中 `[search]` 段落下的设置，从而在整个实现中配置批处理模式搜索。

当您有几个批处理模式搜索线程并发运行时，它们会变成内存使用的负担。您可以通过禁用整个实现的批处理模式搜索，或者通过限制批处理模式搜索线程一次可以从索引数据桶读取的事件数量，来解决该问题。

```
[search]
allow_batch_mode = <bool>
batch_search_max_index_values = <int>
```

- `allow_batch_mode` 默认值为 `true`，表示对于符合条件的转换搜索启用批处理模式搜索。通过设置 `allow_batch_mode = false` 来禁用批处理模式搜索。
- 如果 `allow_batch_mode = true`，则使用 `batch_search_max_index_values` 来限制从索引文件（数据桶）读取的事件数量。这些条目很小，大约 72 字节；但是，如果可以同时读取更多的条目，则批处理模式会更为高效。默认为 10000000（或 10M）。

例如，如果您的批处理模式搜索导致您在系统内存中运行很慢，可减少 `batch_search_max_index_values` 为 1000000（1M），以减少内存使用量。将此参数设置为较小的值可能会导致搜索速度减缓。您希望找到有效的批处理模式搜索和系统内存保护之间的平衡点。

设置搜索对等节点重试周期

其他 `limits.conf` 设置用于控制在故障期间（如连接错误）重试搜索对等节点的周期。故障与第一次重试以及后续重试（如果进一步故障）之间有一定的时间间隔。

```
[search]
batch_retry_min_interval = <int>
batch_retry_max_interval = <int>
batch_retry_scaling = <double>
batch_wait_after_end = <int>
```

- 可使用 `batch_retry_min_interval` 和 `batch_retry_max_interval` 参数来指定在批处理模式尝试对故障对等节点重试搜索之前要等待的最小或最大时间间隔（以秒为单位）。最小时间间隔默认为 5 秒。最大时间间隔默认为 300 秒。
- 在重试尝试失败之后，可按比例因子 `batch_retry_scaling`（获取大于 1.0 的值）延长在再次重试之前需等待的时间。默认为 1.5。
- 如果所有对等节点都表示没有发生故障，并已发送完整的响应，则批处理模式会认为搜索已完成。如果搜索已完成但一个或多个对等节点出现问题，批处理模式会在 `batch_wait_after_end` 指定的秒数内尝试重新与出现问题的对等节点建立连接。如果批处理模式无法在此时间段内重新建立连接，则会声明搜索结果不完整。默认为

900 秒。

为了进行批处理模式搜索，重新启动搜索对等节点

根据对等节点是否为群集节点，批处理模式会以不同方式重新启动搜索对等节点。

- 如果搜索对等节点为群集节点，则批处理模式会等待群集主节点衍生新的生成。
- 如果搜索对等节点不是群集节点，并且与其之间的连接已丢失，则批处理模式会按照如上所述的重试周期参数尝试进行重新连接。当批处理模式重新建立了到搜索对等节点的连接时，它会恢复批处理模式搜索，直到搜索完成。

配置批处理模式搜索并行化

您甚至可以选择利用批处理模式搜索并行化，来使您的批处理模式搜索更加高效。当您启用批处理模式搜索并行化时，会并发运行批处理搜索的两个或多个搜索管道来读取索引数据桶和处理事件。这种方式提高了批处理模式搜索的速度和效率，但增加了系统内存消耗的开销。

您可以使用额外的 `limits.conf` 参数集，来启用和配置批处理模式搜索并行化。这是索引器侧的设置。它需要配置在您所有的索引器上，而不是搜索头上。

```
[search]
batch_search_max_pipeline = <int>
batch_search_max_results_aggregator_queue_size = <int>
batch_search_max_serialized_results_queue_size = <int>
```

- 使用 `batch_search_max_pipeline` 设置当运行符合批处理模式要求的搜索时会启动的批处理模式搜索管道数量。此参数默认值为 1。将其设置为 2 或更大值，在整个 Splunk 部署中并行化批处理模式搜索。设置为更大值提高了搜索性能，代价是增加了线程的使用和内存消耗。
- `batch_search_max_results_aggregator_queue_size` 参数控制结果队列的大小。结果队列是搜索管道放置处理的搜索结果的地方。它的默认大小为 100MB。永远不要将它设为零。
- `batch_search_max_serialized_results_queue_size` 参数控制序列化结果队列的大小，批处理搜索过程从中传输序列化的搜索结果。它的默认大小为 100MB。永远不要将它设为零。