数据分析与决策技术丛书

HZ BOOKS 华章IT

[PACKT] PUBLISHING

Splunk Operational Intelligence Cookbook

Splunk智能运维实战

乔史·戴昆（Josh Diakun）
[美] 保罗 R. 约翰逊（Paul R Johnson） 著
德莱克·默克（Derek Mock）

宫鑫 康宁 刘法宗 译

Splunk智能运维权威参考，包含70多条实用技巧。通过简单易学、循序渐进的操作技巧，引导读者掌握Splunk Enterprise的关键特性，获取智能运维能力。

机械工业出版社
China Machine Press

数据分析与决策技术丛书

Splunk Operational Intelligence Cookbook

Splunk智能运维实战

乔史·戴昆（Josh Diakun）
[美] 保罗 R. 约翰逊（Paul R Johnson） 著
德莱克·默克（Derek Mock）

宫鑫 康宁 刘法宗 译

机械工业出版社
China Machine Press

图书在版编目（CIP）数据

Splunk 智能运维实战 /（美）戴昆（Diakun，J.），（美）约翰逊（Johnson，P. R.），（美）默克（Mock，D.）著；宫鑫，康宁，刘法宗译 . —北京：机械工业出版社，2015.9
（数据分析与决策技术丛书）
书名原文：Splunk Operational Intelligence Cookbook

ISBN 978-7-111-51549-4

I. S… II. ①戴… ②约… ③默… ④宫… ⑤康… ⑥刘… III. 数据处理软件 IV. TP274

中国版本图书馆 CIP 数据核字（2015）第 220331 号

本书版权登记号：图字：01-2014-8121

Splunk 智能运维实战

出版发行：机械工业出版社（北京市西城区百万庄大街 22 号　邮政编码：100037）
责任编辑：陈佳媛　　责任校对：殷　虹
印　　刷：三河市宏图印务有限公司　　版　　次：2015 年 10 月第 1 版第 1 次印刷
开　　本：186mm × 240mm　1/16　　印　　张：18.75
书　　号：ISBN 978-7-111-51549-4　　定　　价：69.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换
客服热线：（010）88379426　88361066　　投稿热线：（010）88379604
购书热线：（010）68326294　88379649　68995259　　读者信箱：hzit@hzbook.com

The Translator's Words 译者序

物联网、大数据、云计算、工业 4.0，这些新潮词汇越来越频繁地出现在我们的生活中，读者也一定有所耳闻。科学的发展日新月异，我们已经步入信息时代，大数据已成为传统行业顺应潮流进行变革的一种资源。企业该如何处理应用程序、服务器和各式各样传感器生成的无尽数据呢？ Splunk 提供了简单易行的解决方案。它可以收集、索引、关联和监控数据，把看似枯燥无味的机器数据转化为价值非凡的智能运维。Splunk 生成的直观且生动的可视化图表可以帮助用户获取商业洞见，更好地做出决策。

本书详细介绍了 Splunk 数据引擎的使用方法，包括如何索引和搜索数据，如何制作可视化图表，如何创建仪表盘和应用程序，如何创建警报和数据模型，等等。本书深入浅出、循序渐进，每一章都以解决特定问题为目的，通过实战来帮助读者迅速掌握 Splunk 的方方面面。主要步骤都配有截图，生动直观地讲解操作方法。俗话说“知其然，更要知其所以然”，所以操作步骤后面紧跟着理论讲解，让读者了解操作背后的原理。同时每节还配有拓展内容，教授读者更多实用技巧。

本书内容丰富，实战性强，讲解细致，任何想通过 Splunk Enterprise 平台获取智能运维的读者都可以阅读本书。书中有些部分会用到正则表达式并介绍一些利用 Python 和 XML 语言的技巧。读者若具有这些方面的经验会更有益处，但不了解相关知识也不会构成任何阅读障碍。

本书的初译稿由刘法宗完成，康宁副教授（青岛科技大学 MTI 中心主任）和宫鑫先生（射手学院创始人）承担了本书的审校工作。本书得以顺利完成，要感谢杨志芳女士和裴淑娟老师给予的大力支持与帮助。

宫鑫

2015 年 7 月

前　言 Preface

在这个以科技为中心的世界里，各式各样的机器产生了大量的数据。Splunk 因此推出了业内领先的大数据智能运维平台——Splunk Enterprise。这个强大的平台能让用户将机器数据转化为可操作的、非常重要的运维智能。

本书融合了各种实用方法，旨在提供指导和实用知识，以便使读者掌握 Splunk Enterprise 6 的各种功能，从数据中提取出强大而重要的运维智能。

本书通过简单易学、循序渐进的操作技巧，教授读者如何有效地收集、分析并创建所在环境的运营数据报表。这些技巧将展示如何加快智能报表的交付，并教授读者通过仪表盘和应用 Splunk Enterprise 中的各种可视化手段来恰当地展示数据。读完本书，读者将能建立一个强大的智能运维应用程序，并学会使用 Splunk Enterprise 平台的多项关键特性。

当读者向他人介绍 Splunk Enterprise 平台和自己新掌握的获取智能运维的能力时，可将书中简单易学的诀窍用作教学工具。

本书的主要内容

第 1 章介绍将数据导入 Splunk 的多种方法，包括从本地文件和目录收集数据，通过 TCP/UDP 端口输入，直接从通用转发器导入或使用脚本化和模块化输入。该章还会介绍一个数据集，在之后的章节都会用到它。我们还将学习如何生成样本数据，以便与本书的每项技巧配套使用。

第 2 章介绍了本书的第一组技巧。该章所介绍的信息和技巧将利用第 1 章获取的数据，教授使用 Splunk 的 SPL（搜索处理语言）来搜索事件数据；应用字段抽取；按字段值将同类事件分组；使用 table、top、chart 和 stats 命令来构建基本的报表。

第 3 章指导读者在前一章创建的报表基础上进行可视化构建。该章将教授如何通过 Splunk 提供的强大的可视化手段将数据和报表生动地呈现出来。该章将要介绍的可视化手段

包括单值、图表（条形图、饼图、折线图和面积图）、散点图和计量器。

第 4 章在前一章的可视化图表的基础上，讲授仪表盘的概念。该章提供的信息和技巧将简要介绍仪表盘的用途，并教授读者如何恰当地使用仪表盘，如何使用仪表板编辑器来建立仪表盘，如何建立表单来搜索事件数据等。

第 5 章让读者更深入地研究数据，介绍事务、次级搜索、并发、关联和更高级的搜索命令。通过该章提供的信息和技巧，读者将学会从不同的来源聚合信息，并了解如何在不同的事件数据之间建立关联。

第 6 章将介绍查找和工作流程操作的概念，为的是增加分析数据。该章介绍的技巧可帮助读者应用这个核心功能来进一步加深对所分析数据的理解。

第 7 章解释为什么预定警报或实时警报是完整的智能运维和运营认知的关键。该章将介绍主动警报的概念和益处，并说明何时应用这些警报最好。该章提供的技巧将指导读者在前面章节所学知识的基础上创建警报。

第 8 章介绍汇总索引的概念，为的是加快报表速度并节约获取商业情报的时间。该章会简单介绍一些将汇总索引用于加快报表速度或长时间保存聚集统计数据的常见情况。

第 9 章介绍 Splunk Enterprise 6 推出的两个最新、最强大的特性：数据模型和透视工具。该章提供的技巧将指导读者学习建造数据模型并使用透视工具迅速设计基于已构建模型的智能报表。

第 10 章是本书的最后一章，介绍 Splunk 的 4 个非常强大的特性。这些特性允许读者使用 Splunk 来创造丰富和强大的交互体验。该章给出的技巧让读者能超越 Splunk Enterprise 的核心功能并制作自己的带有强大 D3 可视化的智能运维应用程序。此外，它也将讲述查询 Splunk 的 REST API 的技巧，并给出一个基础的 Python 应用来使用 Splunk 的 SDK 执行搜索。

阅读前的准备工作

要学习本书提供的技巧，读者需要安装 Splunk Enterprise 6 并拥有本书附带的样本数据。这些技巧适用于所有 Splunk Enterprise 环境，但为了得到最佳结果，我们建议大家使用本书提供的样本。

Splunk Enterprise 6 可以免费下载并在主流平台上运行，下载地址是 http://www.splunk.com/download。

本书提供的样本会附带 Splunk 事件生成器工具，这样当你学习这些技巧时，事件数据就能刷新，事件会重放。

本书的读者对象

本书面向所有用户。不管是初学者还是高级人员，任何想将 Splunk Enterprise 平台用作智能运维工具的人都可以阅读本书。书中包括的技巧对 IT、安全、产品、营销或任何其他领域的人都有帮助。

尽管本书和书中的技巧任何人都可以学习，但它介绍的概念和特性会越来越复杂，对初学者来说可能较难理解。如果你需要更多地了解某个特性，Splunk 制作了大量文档来介绍 Splunk Enterprise 的所有特性，可以访问 http://docs.splunk.com/Documentation/Splunk 来查找。

书中有些部分会用到正则表达式并介绍一些利用 Python 和 XML 语言的技巧。读者若具有这些方面的经验会更有益处，但这些知识不是必需的。

下载示例代码

你可以登录 http://www.hzbook.com 下载本书示例代码。

Contents 目　录

第 1 章 Chapter 1

游戏时间——导入数据

1.1 简介

加快运维智能的机器数据有很多不同的形式，来源也各不相同。Splunk 可从多种来源收集并索引数据，其中包括 Web 服务器或商业应用程序创建的日志文件，网络设备生成的系统日志数据，及自定义开发脚本输出的数据。即便数据一开始看上去很复杂，我们也可以借助 Splunk 轻松地实时收集、索引、转化和呈现数据。

本章将学习一些基本的技巧，掌握如何将所需的数据导入 Splunk，介绍如何使用样本数据集来构建自己的 Splunk 智能运维应用程序。该数据集是由一个虚拟的三层式电子商务 Web 应用程序生成的，包含 Web 服务器日志、应用程序日志和数据库日志。

Splunk Enterprise 可以索引任何类型的数据，不过，它最适合于索引时间序列数据（带时间戳的数据）。Splunk Enterprise 索引数据是按时间戳和 / 或事件大小将数据分解为事件并编制索引。索引是 Splunk 制作的数据存储区，存取速度很快，可以检索并根据分布式服务器环境进行扩展，常被称为索引器。这也是把导入数据到 Splunk 的过程称为索引的原因。

所有经 Splunk 索引的数据都会被分配一个源类型。数据源类型有助于标识事件的数据格式类型以及事件的来源。Splunk 有多种预置的源类型，但也可以自己指定。示例的源类型包括：access_combined、cisco_syslog 和 linux_secure。当索引器将数据索引至 Splunk 时，源类型就会添加到数据上。用户执行字段提取或在各种搜索中过滤搜索数据时，数据源类型是一项关键字段。

Splunk 社区能帮助用户更轻松地在 Splunk 中导入数据。Splunk 的扩展性让我们有机会

开发输入、命令和应用程序，这些都可以轻松与他人分享。如果我们想索引来自特定系统或应用程序的数据，可能有人已经开发并发布了相关的配置和工具，我们可以轻松地将其用于自己的 Splunk Enterprise 部署上。

Splunk Enterprise 致力于让数据收集更轻松，不久我们就能为自己或他人向 Splunk 中导入大量数据——至少会用完 Splunk License 许可的索引量。

1.2 索引文件和目录

从文件和目录输入数据是向 Splunk 导入数据最常用的方法。这种类型的输入主要是为了索引日志文件。几乎每个应用程序或系统都会产生日志文件，当中包括了我们想搜索和制作报表的很多数据。

Splunk 能够持续监控写入现有文件的新数据或添加到目录中的新文件，并且能够实时索引这些数据。根据生成日志文件的应用类型不同，可以将 Splunk 设定为监控单一文件（基于其位置）或扫描整个目录并监控其中的所有文件。当生成的日志文件包含唯一的文件名（比如名字中含有时间戳）时，后一种配置更常使用。

本节将学习如何配置 Splunk 来持续监控并索引 Splunk 服务器上的一个日志文件，这个日志文件的内容会不断增加。本节将专门展示如何监控并索引 Linux 系统上的 messages 日志文件（/var/log/messages）。然而，同样的方法也适用于 Windows 系统上的日志文件，本书也提供了一个示例文件。但请不要用这种方法索引 Windows 事件日志，因为 Splunk 有专门的 Windows 事件输入法。

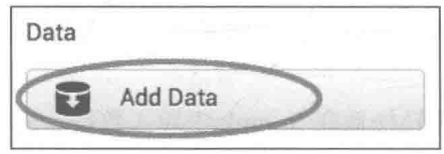
做好准备

要进行本节的操作，需运行 Splunk Enterprise 服务器并且有权限访问读取 Linux 上的 /var/log/messages 文件。没有其他先决条件。如果使用的不是 Linux 系统，并且 / 或者没有权限访问 Splunk 服务器上的 /var/log/messages，要使用本书提供的 cp01_messages.log 文件并将其加载到 Splunk 服务器上一个可访问的目录中。

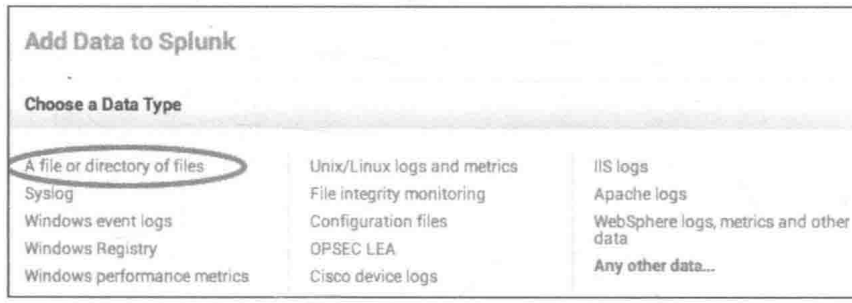
如何操作

按下列步骤监控并索引文件内容。

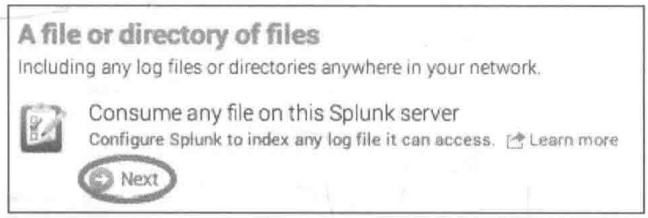
1. 登录 Splunk 服务器。
2. 从右上角的主启动器，单击“添加数据”按钮。



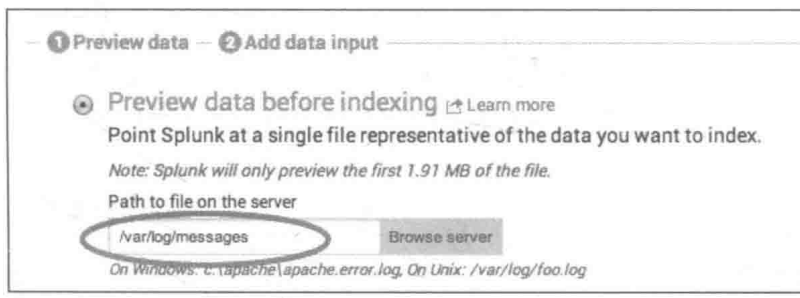
3. 在“选择数据类型”列表中，单击“文件或文件目录”。



4. 在“索引此 Splunk 服务器上的所有文件”选项中单击“下一步”按钮。



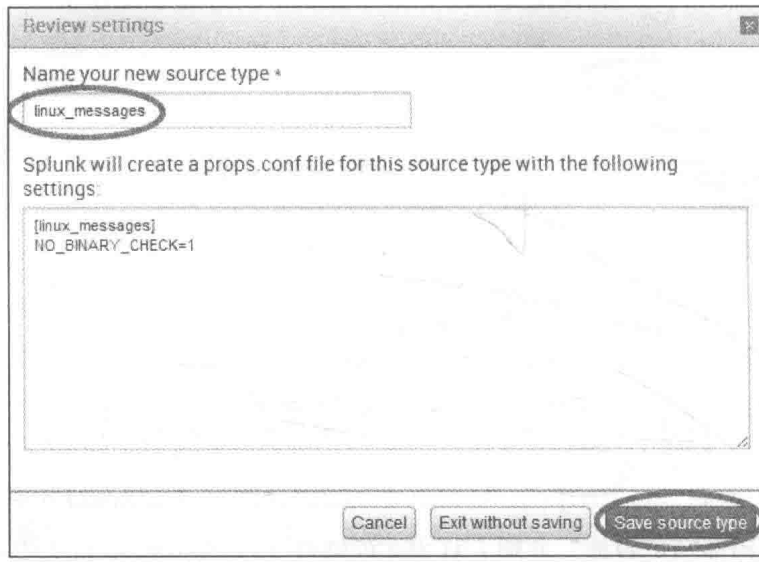
5. 选择“索引前预览数据”并输入日志文件路径（/var/log/messages 或 cp01_messages.log 文件的位置）并单击“继续”按钮。



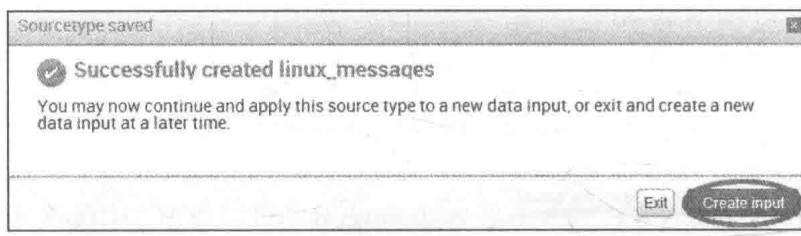
6. 选择“开始一个新的源类型”并单击“继续”按钮。

7. 假如你使用本书提供的文件或本地的 /var/log/messages 文件，数据预览将展示正确分行后的事件及时间戳标志。单击“继续”按钮。

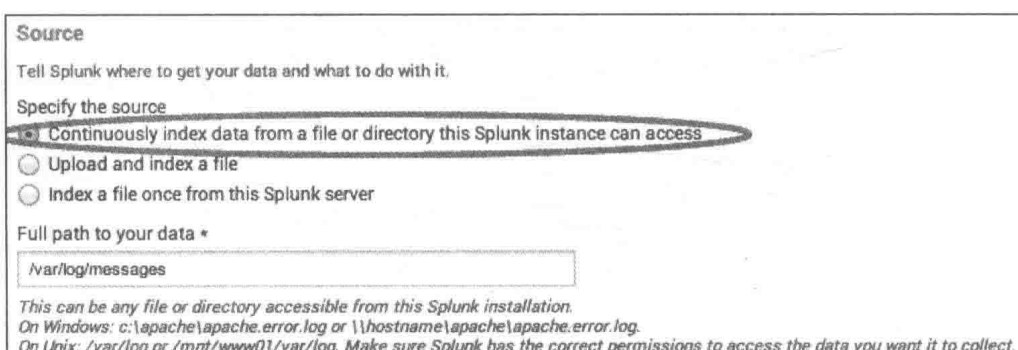
8. 会弹出“预览设置”框。在源类型输入 linux_messages，单击“保存源类型”按钮。



9. 会出现“源类型已保存”的提示框，选择“创建输入”按钮。



10. 在“源”部分，选择“从 Splunk 本次访问的文件或目录中持续索引数据”，并填写数据路径。



如果只是想一次性加载某个文件，可选择“上传并索引文件”。这个选项适合于索引一组数据到 Splunk，既可用来回填一些缺失或不完整的数据，也可仅仅为了利用其搜索和报表工具。

11. 暂且忽略其他设定，直接单击“保存”。然后，在下一个界面中，点击“开始搜索”。在搜索栏，输入下列搜索，时间范围设定为“全部时间”：

```
sourcetype=linux_messages
```

在本节中，可直接使用常见的系统日志源类型，不过，创建新的源类型往往是更好的选择。根据数据源不同，系统日志格式可能差异很大。因为像提取字段这样的知识对象是建立在源类型之上的，为所有对象使用单一的系统日志源类型可能会较难搜索到所需的数据。

工作原理

当添加一个新的文件或目录来导入数据时，也就是在后台向 inputs.conf 文件添加新的配置节。Splunk 服务器可包含一个或多个 inputs.conf 文件，它们位于 $SPLUNK_HOME/etc/ system/local 或 Splunk 应用程序的 local 目录。

Splunk 使用的输入类型是监控器，并被设定为指向某个文件或目录。如果设置对一个目录进行监控，目录中的所有文件都会被监控。当 Splunk 监控文件时，它会首先从头索引所有能读取的数据。完成后，Splunk 将保留上次读取数据的位置记录，如有任何新的数据写入文件，它将读取这个数据并继续记录。这个过程几乎和在 UNIX 操作系统下使用 tail 命令相同。如需要监控一个目录，Splunk 也会提供很多附加的配置选项，比如将不需要 Splunk 索引的文件列入黑名单。

若想获取 Splunk 配置文件的更多信息，请访问 http://docs.splunk.com/Documentation/Splunk/latest/Admin/ Aboutconfigurationfiles。

更多内容

除了可按照本节的方法通过 Splunk 的 Web 界面添加输入信息来监控文件和目录，还有其他方法来快速地添加多种输入信息。这些方法允许我们自定义 Splunk 提供的多种配置选项。

通过 CLI（命令行界面）添加文件或目录数据输入

除了通过 GUI（图形用户界面）之外，还可通过 Splunk CLI（command-line interface）来添加文件或目录输入。进入 $SPLUNK_HOME/bin 目录并执行下列命令（将需要监控的文件或目录替换成你自己的）。

UNIX 系统：

```
./splunk add monitor /var/log/messages -sourcetype linux_messages
```

Windows 系统：

```
splunk add monitor c:\filelocation\cp01_messages.log -sourcetype
linux_messages
```

许多参数可以随文件位置一起被传送到监控器。参考 Splunk 技术文档来了解更多 CLI 数据输入的用法（http://docs.splunk.com/Documentation/Splunk/latest/Data/ MonitorfilesanddirectoriesusingtheCLI）。

通过 inputs.conf 添加文件或目录输入

另一种添加文件和目录输入的常用方法是手动将其直接添加至 inputs.conf 配置文件中。该方法常用于大环境中或配置 Splunk 转发器来监控终端上的文件或目录。

编辑 $SPLUNK_HOME/etc/system/local/inputs.conf 并添加输入。添加完输入后，需要重启 Splunk 来识别更改。

UNIX 系统：

```
[monitor:///var/log/messages]
sourcetype = linux_messages
```

Windows 系统：

```
[monitor://c:\filelocation\cp01_messages.log]
sourcetype = linux_messages
```

如需进行多个输入，编辑 inputs.conf 通常可以更快地添加新文件和目录来监控。编辑 inputs.conf 时，要确保使用正确的语法，并重启 Splunk 来使修改生效。此外，在 inputs.conf 文件中设定源类型是指定源类型的最佳做法。

通过 Splunk CLI 一次性索引数据文件

除了从 Splunk GUI 中选择“上传并索引文件”进行操作之外，也可使用很多 CLI 功能来执行一次性批量加载数据。

使用 oneshot 命令告知 Splunk 文件位置及所用参数，比如源类型：

```
./splunk add oneshot XXXXXXX
```

另一种方法是将希望索引的文件放入 Splunk spool 目录，$SPLUNK_ HOME/var/spool/splunk，然后使用 spool 命令添加文件：

```
./splunk spool XXXXXXX
```

如使用 Windows 系统，应省略 Splunk 命令前的“./”。