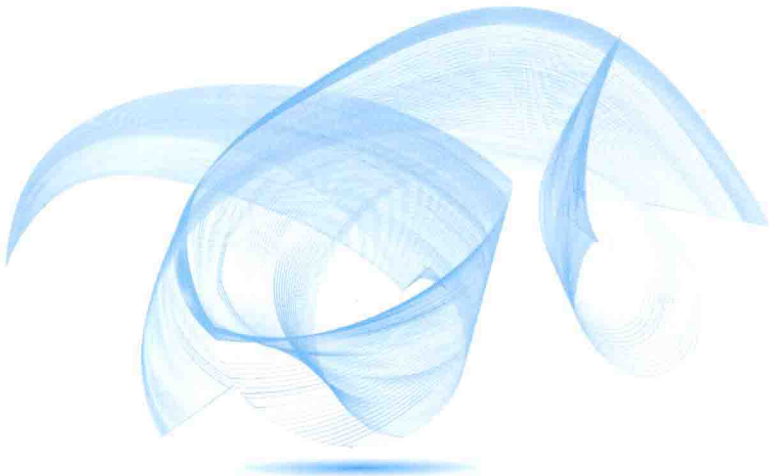
HZ BOOKS
华章科技

[PACKT] PUBLISHING

内容翔实，案例丰富，全方位讲解Splunk大数据开发与应用的各种技术细节、方法和最佳实践

既详细讲解搜索语言的基本知识和表格、图表、指示板的构建，又深入讨论高级搜索案例、应用程序构建和更多的Splunk特性等，为快速掌握并灵活运用Splunk深入挖掘机器数据价值提供最佳指导

大数据技术丛书



Implementing Splunk

Big Data Reporting and Development for Operational Intelligence

Splunk实践指南

（美） Vincent Bumgarner◎著

杨甲东◎译

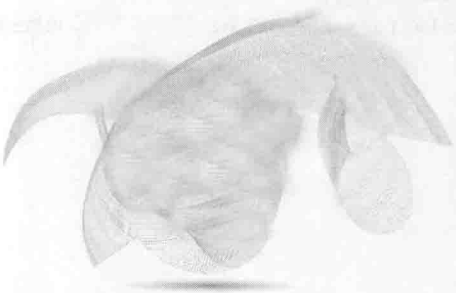
机械工业出版社
China Machine Press

技术丛书

Implementing Splunk
Big Data Reporting and Development for Operational Intelligence

Splunk实践指南

（美） Vincent Bumgarner◎著
杨甲东◎译



机械工业出版社

图书在版编目（CIP）数据

Splunk 实践指南 /（美）布姆加纳（Bumgarner, V.）著；杨甲东译 . —北京：机械工业出版社，2014.12
（大数据技术丛书）
书名原文：Implementing Splunk: Big Data Reporting and Development for Operational Intelligence

ISBN 978-7-111-48767-8

I. S… II. ① 布… ② 杨… III. 数据处理软件 IV. TP274

中国版本图书馆 CIP 数据核字（2014）第 290796 号

本书版权登记号：图字：01-2013-6571

Splunk 实践指南

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：藁城市京瑞印刷有限公司　　版　　次：2015 年 2 月第 1 版第 1 次印刷
开　　本：186mm × 240mm　1/16　　印　　张：21
书　　号：ISBN 978-7-111-48767-8　　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）88378991　88361066　　投稿热线：（010）88379604
购书热线：（010）68326294　88379649　68995259　　读者信箱：hzjsj@hzbook.com

The Translator's Words 译 者 序

随着科学技术的迅猛发展、互联网技术的快速普及和云计算的广泛应用，每天我们身边都会产生浩如烟海的非结构化数据。这些数据渗透到世界的各个行业及各个业务领域，其数量正在以几何级数的方式增长。它们表面看似杂乱无章，毫无规律，但其中蕴含着巨量的有价值的信息。在互联网时代，数据本身就是资产，从海量数据中提纯有用信息的需求不可估量。大数据技术的应用则意味着这些资产能量的有效释放。面对爆炸式增长的数据，企业迫切需要挖掘数据中的潜在价值。采用大数据技术，对海量数据进行开发和应用，将有效提升企业的创新能力并催生新的商业模式，为企业解决传统业务问题带来新的变革机遇。如何盘活已经存在或正在生成的数据资源，使其为科技进步、企业决策及个人服务创造更多的价值，这是大数据技术所要解决的核心问题。

Splunk 是一家提供大数据分析服务的智能软件提供商。它成立于 2003 年，总部位于美国旧金山。2012 年 Splunk 在纳斯达克成功上市，进而成为第一家上市的大数据处理公司。其业务范围覆盖北美、亚太、欧洲、非洲及中东等地区。中国市场目前主要集中在电信、保险和银行业。Splunk 软件平台可以实时对任何 APP、服务器和网络设备的机器数据进行索引、监控与分析，并将结果生成图形化报表。这些机器数据既可以来自本地，也可以来自云；既可以是日志、配置文件，也可以是消息和警告等。本书作者 Vincent Bumgarner 先生是一位拥有 20 多年从业经验的资深软件设计师，一直在多种不同语言的平台上从事软件设计与应用，他从 2007 年起使用 Splunk，在长期从事应用培训工作的同时始终关注该技术的进展。本书内容翔实，案例丰富，循序渐进，实用性强，为新老用户提供了学习、掌握大数据开发的基本手段和有效方法。

本书从理解用户交互界面的基本元素入手，介绍了搜索语言的基本知识，对表格、图表、指示板等具体内容进行了细致描述。通过对高级搜索案例和应用程序的逐步分析，向读者展示了构建属于用户自己的应用程序的方法，并用适当的篇幅阐述了配置和扩展 Splunk 等相关内容。本书是学习大数据开发与应用的实用教材。

海量数据是一座金矿，大数据分析软件就是淘金的工具，学习和掌握大数据分析应用

技术，就如同掌握黄金的提纯方法。译者作为在数据挖掘领域工作的一名技术人员，深知海量数据中蕴藏的巨大价值，深知数据分析软件在大数据开发中所起的重要作用，深知大数据开发应用将对各行各业带来的创新和变化。本书的翻译旨在为同行从业人员和国内用户学习大数据处理技术提供具体的、有益的帮助。但由于本人水平有限，如翻译内容有不当之处，恳请各位指正。

杨甲东

2014 年 10 月于北京

Preface 前 言

Splunk 是一个强有力的收集、存储、报警、制表和研究机器数据的工具。机器数据通常来源于服务器的日志，也可以从其他来源收集得到。面对挖掘机器数据价值这一个巨大的问题，Splunk 是目前最灵活且可扩展的解决方案。

本书旨在为 4.3 版的 Splunk 提供系统的辅助性的指南。由于 Splunk 的文档和社区资源非常庞大，从中寻找有价值的线索与信息有时候是一件令人生畏的事情。本书的目的在于以尽可能简洁和有益的形式展示如何有效实现 Splunk。

本书主要内容

第 1 章 Splunk 交互界面，将带领读者了解用户交互界面的基本元素。

第 2 章理解搜索，将涉及搜索语言的基础，同时将特别关注编写高效查询。

第 3 章表格、图表和字段，将展示如何使用字段生成报表，进而介绍如何构建自定义字段。

第 4 章简单的 XML 指示板，将首先介绍使用 Splunk 的 Web 交互界面构建第一个指示板，然后将查看如何构建表格和更高效的指示板。

第 5 章高级搜索案例，将通过若干案例向读者介绍使用 Splunk 搜索语言的有趣方式。

第 6 章扩展搜索，将展示 Splunk 的许多特性，以帮助大家以强有力的方式对事件进行分类和对搜索结果采取行动。

第 7 章使用应用程序，将涉及应用程序的概念，帮助大家安装若干流行的应用程序，同时帮助大家构建属于自己的应用程序。

第 8 章构建高级指示板，将解释高级 XML 指示板的概念，并介绍将简单 XML 转换为高级 XML 指示板的实用方法。

第 9 章摘要索引和 CSV 文件，将介绍摘要索引的概念，以及摘要索引在提高系统性能方面如何发挥作用。该章也将讨论 CSV 文件使用的其他有趣形式。

第 10 章配置 Splunk，将解释 Splunk 一般配置的结构和意义。该章也将详细介绍合

并配置的过程。

第 11 章高级部署，将涉及在多台机器上部署 Splunk 的一般问题，包括数据输入、系统日志、配置管理和系统扩容等问题。

第 12 章扩展 Splunk，将阐述在输入数据、外部查询、渲染、定制命令和定制行为等方面扩展 Splunk 的方法。

阅读本书的准备工作

为了浏览本书所提供的案例，需要安装 Splunk，该 Splunk 最好部署在非生产环境下。如果你已经在使用 Splunk，那么案例中所提及的概念可以直接应用到你自己的数据上。

对于当前普遍使用的系统平台，可从 http://www.splunk.com/download 免费下载获得 Splunk。

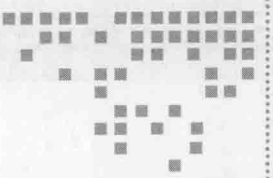
由于样例代码是在 UNIX 系统上开发的，因此使用运行在 UNIX 操作系统的 Splunk 安装版可以更好地理解本书内容。在本书后几章中，需要 Python 相关的背景知识来理解其中所涉及的案例。

本书的目标读者

对于 Splunk 的各级别用户、指示板的设计者和系统管理员等读者，本书都将提供有益的帮助。需要指出的是，本书并非作为 Splunk 官方手册的替代品，而是作为理解相关概念的捷径。

理解和掌握正则表达式，以及拥有阅读 Python 代码的能力，对阅读本书的许多章节都有极大的帮助。

Contents 目 录



第 1 章 Chapter 1

Splunk 交互界面

本章将向你介绍 Splunk 界面中最基本的组成元素，同时将涉及后续章节所要详细介绍的若干概念。或许你希望直奔主题，但对于用户界面元素的概要了解可能会使你事半功倍。在本章我们将要介绍：

- 登录与应用程序的选择。
- 搜索界面组件的细节。
- 管理界面简介。

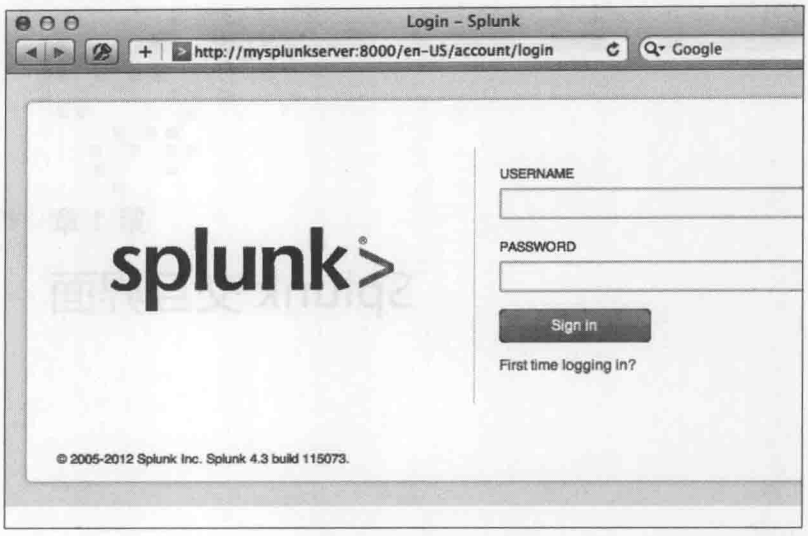
1.1 登录 Splunk

Splunk 交互界面是基于 Web 的，这意味着用户无需安装客户端软件。而新一代的带有快速 JavaScript 引擎的浏览器（例如 Chrome、Firefox 和 Safari 等），将能够更好地支持 Splunk 交互界面。

对于 4.3 版本的 Splunk，浏览器无需安装其他扩展。对于 4.2 以及更早版本的 Splunk，则需要安装 Flash，以渲染图形。有特别说明的早期浏览器和应用程序除外。

Splunk 安装的默认端口是 8000。浏览器地址形式如：http://mysplunkserver:8000 或者 http://mysplunkserver.mycompany.com:8000。如果你已经在本地机器上安装了 Splunk，地址将变化为：http://localhost:8000、http://127.0.0.1:8000、http://machinename:8000 或者 http://machinename.local:8000。

在浏览器中输入地址并确定后，你将首先看到登录界面：



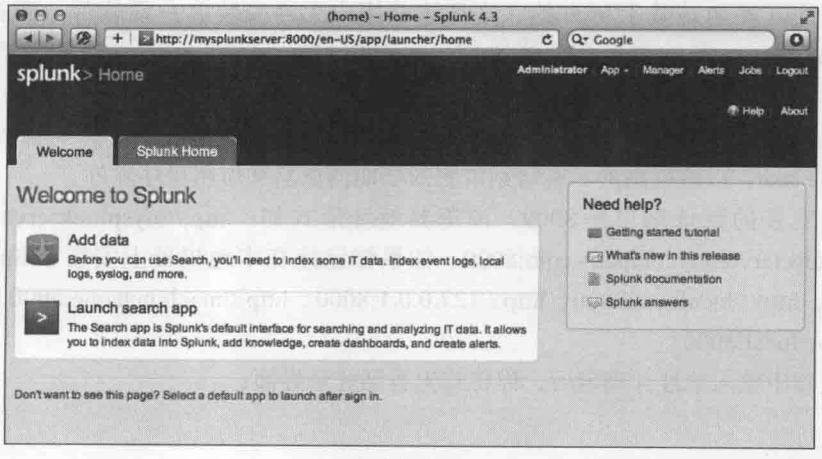
登录界面

默认用户名及密码分别为“admin”和“changeme”。当你第一次登录时，将被提示更改管理员用户（admin）的密码。为避免意外更改系统，修改管理员密码是一个好办法。

默认情况下，用户信息将被配置并存储于 Splunk 中。对授权进行配置后，用户可以使用其他系统，例如 LDAP 等。

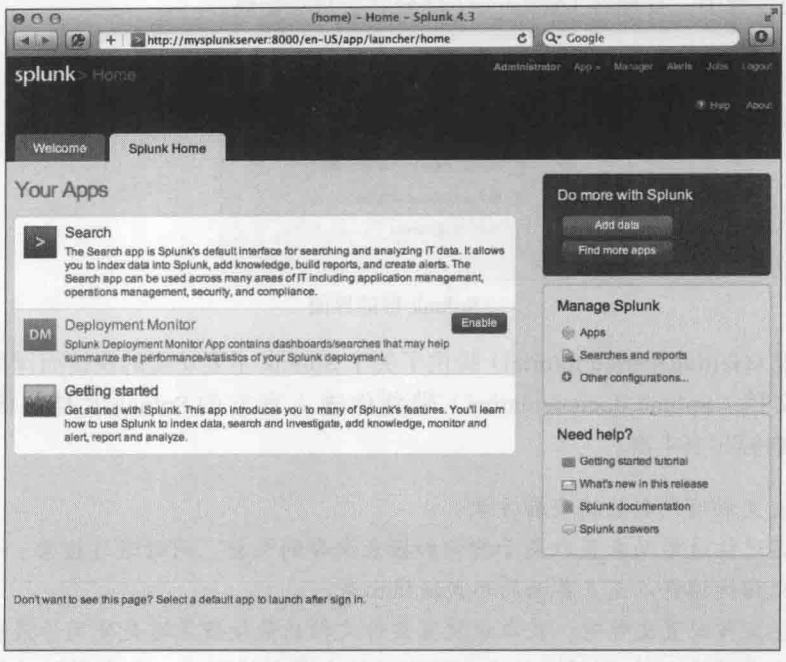
1.2 首页应用程序

在登录后，默认的应用程序是首页（Home）。首页应用程序以面板的形式展示其功能及相关教程。



欢迎标签页

欢迎（Welcome）标签页提供了两项重要的快捷链接：添加数据（Add data）和启动搜索应用程序（Launch search app）。在名为 Splunk 首页（Splunk Home）的第二项标签页中，同样含有上述两项链接。



Splunk Home 标签页

你的应用程序（Your Apps）部分展示了你所安装的 Splunk 中含有图形用户界面的应用程序。

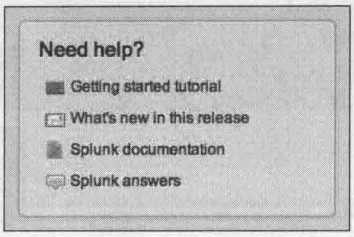
在 Splunk 中，应用程序（App）是一个重载（overloaded）的术语。一个应用程序不一定需要有图形用户界面。它可以仅仅是一个配置集合，这些配置被封装在对 Splunk 有某种意义的目录结构中。我们将在第 7 章对其进行更为详细地讨论。

在 Splunk 相关应用（Do more with Splunk）菜单条的下方，我们可以看到：

- 添加数据（Add data）：它链接到名为向 Splunk 添加数据（Add Data to Splunk）的界面。这一界面为利用 Splunk 获得本地数据流提供了一个良好的开端。新的名为预览数据（Preview data）的交互界面将极大地帮助你减少日期配置和断行等操作的复杂性。对这些交互界面，这里不再做深入讲解。我们将在第 10 章中详细介绍由这些组件所产生的配置文件。
- 寻找更多应用程序（Find more apps）：它将帮助你从 Splunk 库（Splunkbase）中寻找并安装更多的应用程序。Splunk 库（http://splunk-base.splunk.com/）是一个非常有用的社区驱动的资源。Splunk 用户和 Splunk 员工在此提出并解答问题、讨论代码片段

和发布应用程序。

管理 Splunk（Manage Splunk）带领用户进入 Splunk 的管理器（Manager）部分。Splunk 管理器部分被用来对 Splunk 进行更多方面的配置，其中的选项随着用户的能力而发生变化。在学习本书的过程中，管理器（Manager）部分将会贯穿始终。



Splunk 帮助界面

入门教程（Getting started tutorial）提供了关于 Splunk 主要功能的快捷而详尽的综述。

Splunk 文档（Splunk documentation）带领你进入官方的 Splunk 文档。该文档（位于 splunk.com）内容相当丰富。

Splunk 文档有两种快速查看方式：

可以通过快速帮助来获取关于搜索和报表命令的文档。同时通过搜索，一个关于该命令的文档链接将以交互界面的形式提供出来。

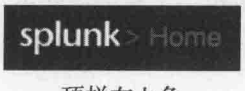
当直接操作配置文件时，获取该配置文件文档的最便捷渠道是使用你最喜欢的搜索引擎，搜索“Splunk 配置文件名 .conf”。绝大多数情况下，搜索得到的第一条链接将是所需要的文档。

Splunk 问答（Splunk answer）链接到先前提到的 Splunk 库的站点。早期版本中的 Splunk 库和 Splunk 问答是不同的两个站点，但是现在已经合并到一起了。

1.3 顶栏

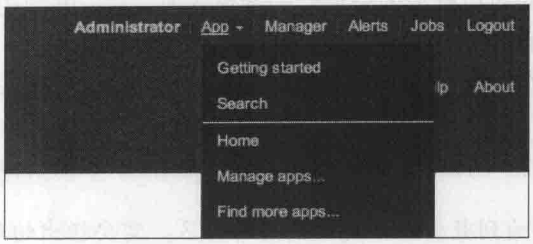
窗口顶部的文字栏包含了关于你的位置以及快速链接信息，这个链接指向首选项、其他应用程序及应用程序管理。

当前的应用程序显示在窗口左上角处。



顶栏左上角

点击 Splunk 的标志或文本，将转到该应用程序的默认页面。大多数应用程序仅仅可以改变标志旁边的文本。对于整个区域而言，用户可以通过修改应用程序的 CSS 实现自定义标志和替代文本。关于这一部分内容，我们将在第 7 章中详细介绍。



顶栏

窗口右上方包含了绝大多数情况下可用的链接：

- 首先显示的是当前登录用户的用户名。在本例中，用户为管理员（Administrator）。点击用户名将转到你的账号（Your account）界面。
- 应用程序（App）菜单提供了已安装应用程序和应用程序管理的快速链接。只有那些含有图形用户界面组件，并同时当前用户具有访问权限的应用程序，才会显示在这个菜单中。
- 窗口顶部的管理器（Manager）链接总是可用的。管理器页面中的可用选项由当前用户的角色决定。
- 点击任务（Jobs）链接将弹出任务（Jobs）窗口。任务窗口提供了当前和过去在该Splunk上运行的搜索任务的列表。它有助于检索过去任务的结果，同时也有助于确定哪些搜索任务正在使用系统资源。我们将在第2章中详细讨论这一界面。
- 注销（Logout）链接将结束当前的会话，并强制用户重新登录。

下面的屏幕截图展示了你的账号页面的形态。

你的账号页面