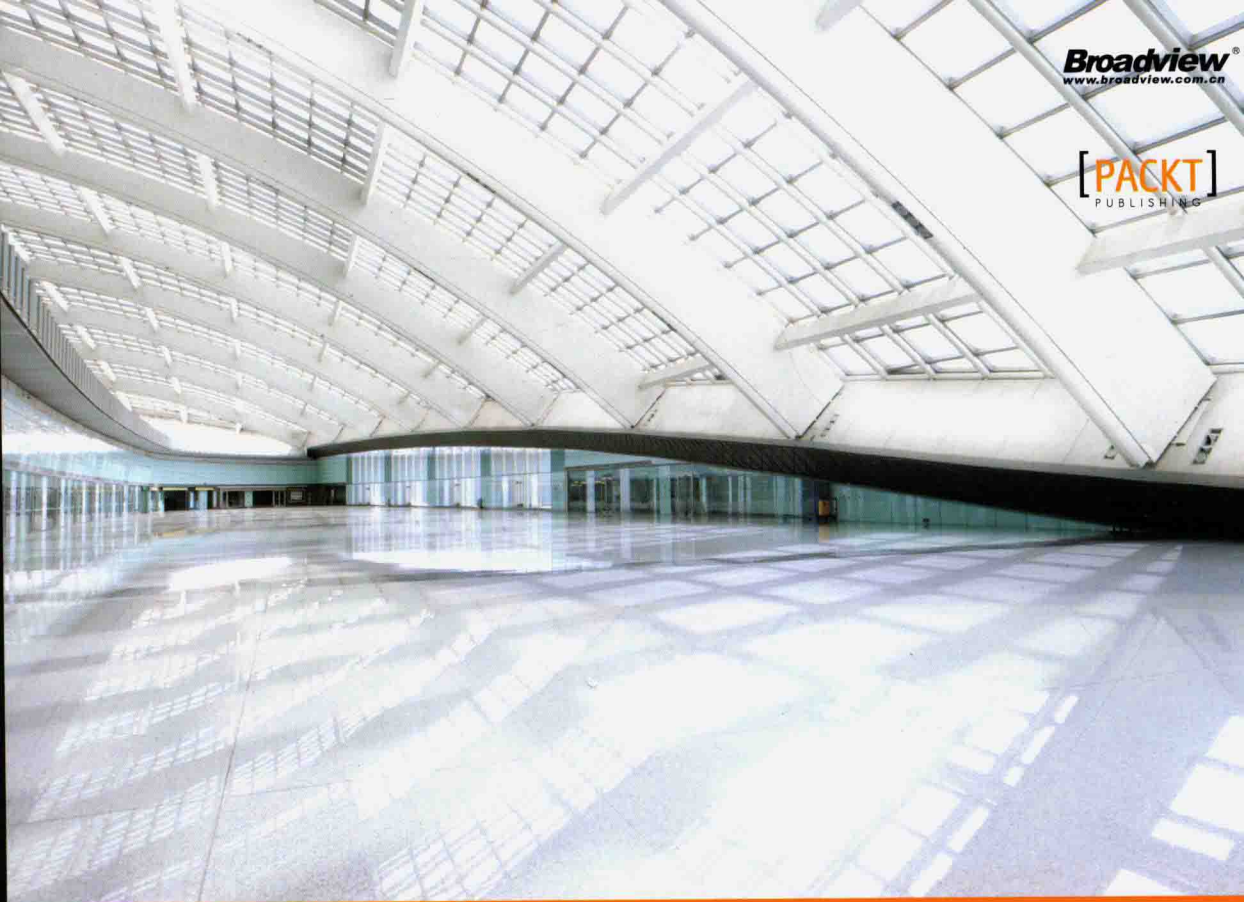


Broadview
www.broadview.com.cn

[PACKT]
PUBLISHING

Learning ELK Stack

Learning ELK Stack 中文版

使用Elasticsearch、Logstash和Kibana分析日志数据并构建迷人的可视化

[美] Saurabh Chhajed 著
宁海元 张新铭 阚海明 林杰 译

中国工信出版集团
電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

Learning ELK Stack

Learning ELK Stack 中文版

[美] Saurabh Chhajed 著
宁海元 张新铭 阚海明 林杰 译

電子工業出版社
Publishing House of Electronics Industry
北京•BEIJING

内 容 简 介

ELK 技术栈是一套新兴的日志处理开源系统，ELK 分别代表 Elasticsearch、Logstash 和 Kibana。Elasticsearch 是基于 Lucene 构建的一套分布式搜索引擎，提供了实时搜索和聚合分析的能力，是整个 ELK 技术栈的核心。Logstash 构建了数据采集、解析和输出的框架，通过插件可以支持不同的输入、过滤和输出的场景。Kibaba 实现了对 Elasticsearch 的搜索查询和数据可视化。三者组合起来是一套经过很多生产环境验证的日志解决方案。本书针对 ELK 技术栈由浅入深地对每个组件做了介绍，并且通过实际的部署和应用案例，真正地讲清楚了如何利用 ELK 技术栈来实现日志数据的完整管道。

版权贸易合同登记号　图字：01-2016-1369

图书在版编目（CIP）数据

Learning ELK Stack 中文版 /（美）苏库拉·塞哈特（Saurabh Chhajed）著；宁海元等译. —北京：电子工业出版社，2016.6
书名原文：Learning ELK Stack
ISBN 978-7-121-28884-5

Ⅰ. ①L… Ⅱ. ①苏… ②宁… Ⅲ. ①数据处理－研究 Ⅳ. ①TP274

中国版本图书馆 CIP 数据核字(2016)第 112566 号

责任编辑：张春雨
印　　刷：三河市双峰印刷装订有限公司
装　　订：三河市双峰印刷装订有限公司
出版发行：电子工业出版社
　　北京市海淀区万寿路 173 信箱　　邮编：100036
开　　本：787×980　1/16　　印张：12　　字数：232 千字
版　　次：2016 年 6 月第 1 版
印　　次：2016 年 6 月第 1 次印刷
定　　价：65.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888，88258888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式：010-51260888-819　faq@phei.com.cn。

译者序

随着云计算和大数据的发展，分布式架构已经成为常态。分布式架构解决了系统随着业务发展的可扩展性问题，也随之引入了一些新的问题。比如日志，在分布式系统中，日志也随之分布到多台服务器上。这时候，要借助日志来排查系统问题，或者分析业务数据等，成本就比传统的单机系统要高出很多了。

从大数据的角度来看，大数据的来源，主要包括：

1. 数据库
2. 日志文件
3. 爬虫

其中日志文件是最常见也是量最大的数据来源。爬虫也经常会将初步处理的数据以文件的形式存放，也可以归结到日志文件一类。解决日志文件的采集、解析和分析，也是大数据时代的普遍需求。

所以，在云计算和大数据时代，将分布在多台服务器上的日志集中起来，统一地进行存储、解析、搜索和分析展现，成为一种普遍需求。在开源领域，ELK 技术栈是解决日志问题的一种流行方案，这两年来得到了越来越多公司的青睐。ELK 技术栈中的 E，代表 Elasticsearch，是基于 Lucene 构建的一套分布式搜索引擎，并且在传统的基于倒排索引的搜索功能之外，通过引入列式存储 DocValue，具备了不错的分析能力。

我在创立袋鼠云之初，就将日志做为公司大数据产品的主攻方向，基于 Elasticsearch 和 Logstash 开发了袋鼠云日志。在产品的开发过程中，对 Elasticsearch、Logstash 和 Kibana 这套 ELK 技术栈有了更深的认识。在日志场景下，这确实是非常匹配的一套开源的日志解决方案。但要用好这套开源产品，也需要深入理解每个组件的细节，需要有一支技术能力较强的团队来维护和开发。

今年春节前，电子工业出版社的张春雨在朋友圈征集这本书的译者时，刚好是我们团队在设计袋鼠云日志的阶段，所以毫不犹豫地点赞抢得了这个机会。完稿之后，才发现这本书写得真是非常扎实，各种细节娓娓道来，真正做到了深入浅出，对于技术写作来说，这相当不容易。所以这次翻译工作也让我们团队获益不小，希望本书的读者同样也能有所收获。

为了让本书尽快和读者见面，我们组织了一个袋鼠云的专家团队。我本人负责第一、二章，张新铭负责第三、四章，阚海明负责第五、六、七章，林杰负责第八、九、十章，最后由我统一审稿。由于时间比较紧张，疏漏之处难免，敬请读者朋友们指正。

宁海元 袋鼠云 CTO

2016/5/6 于西园

关于作者

Saurabh Chhajed

是一名技术专家，在为产品和服务行业构建企业应用领域，有着丰富而专业的经验。他在大数据分析和机器学习领域有很多实际的构建工程应用的经历，并且乐于布道大数据和 NoSQL 技术。Saurabh 利用他丰富的技术经验，帮助美国很多大型的金融和工业企业从零开始构建大型的产品体系和分布式应用。他在个人网站 `http://saurzcode.in` 分享了很多的个人技术经验。

Saurabh 过去还帮助审阅了 Packt 出版社出版的技术书籍，包括 *Apache Camel Essentials* 和 *Java EE 7 Development with NetBeans 8*。

我想谢谢我的家人——Krati，她对于我非常的支持和鼓励，尽管我为写此书花费了很多本该陪伴她的时间。

我也要感谢本书的技术审稿人和内容编辑，没有他们，本书不可能面世。

关于审稿人

Isra Ellsa 于 2014 年 1 月获得约旦大学计算机本科学士学位。毕业后，她在加州 Santa Clara 的 Seclytics Security 公司做了一年的软件工程师，在这里接触到了多种技术工作。Isra 目前在约旦首都安曼一家名叫 iHorizons 的公司工作。

Anthony Lapenna 的职业生涯是从软件开发开始的，后来转到运维相关的工作，目前是 WorkIT 的系统工程师。他是运维自动化和 DEVOPS 文化的忠实粉丝。他非常热心于追踪最新的技术，并通过撰写技术文章和分享自己的软件活跃于开源生态圈。

Blake Praharaj 是一位软件工程师，精于在忙碌的创业环境驰骋。他目前就职于 Core Informatics 公司，依靠实验室测试和有效的数据解释为多个行业的科学家们提供数据管理解决方案。和任何优秀的开发者一样，他也在不断地学习和探索新的技术！

我想在这里感谢我的另一半，她非常支持和理解我花时间来审阅此书。

我也要感谢 Core Informatics 的整个团队的支持，尤其是 Vico，使得我有时间来学习 ELK 这一技术。

前言

本书旨在介绍如何使用开源的 Elasticsearch、Logstash 和 Kibana 技术栈来构建企业自己的 ELK 数据管道。本书也覆盖了每个技术组件的核心概念，以及怎样快速使用这些技术搭建日志分析解决方案。本书一共分为十章，其中第 1 章主要介绍如何快速地部署 ELK 技术栈的各个组件，这样在第 2 章中就可以开始构建第一个数据管道；第 3 章到第 7 章分别详细介绍了技术栈各个组件的详情；第 8 章则使用 ELK 构建了一个完整的数据管道；第 9 章则介绍了 ELK 技术栈的几个实际使用案例；最后，在第 10 章中也涉及了如何利用一些周边工具来提升 ELK 技术栈的能力。

本书的内容

第 1 章，主要说明了 ELK 技术栈的概念，以及它能够解决的问题。本章介绍了 ELK 技术栈中每个组件的功能，也涵盖了如何安装 Elasticsearch、Logstash 和 Kibana 组件的内容，可以帮助用户快速运行起整个技术栈。

第 2 章，在这一章构建了一个基本的 ELK 数据管道，输入的数据源是 CSV 格式的文件。通过这个例子中对 ELK 基本配置项的解读，可以帮助我们完成 ELK 搭建并快速地开始分析日志。

第 3 章，覆盖了 Logstash 的关键特性，也说明了如何集成各种不同的输入和输出源。本章也比较细致地介绍了 Logstash 的输入、过滤和输出插件，来帮助完成数据的采集、解析、转换和传输。

第 4 章，解释了在标准插件无法满足时，如何为多样化的需求创建自定义的 Logstash 插件。本章描述了 Logstash 插件的生命周期，以及如何开发和发布不同类型的输入插件、过滤插件和输出插件。

第 5 章，描述了 ELK 技术栈中的 Elasticsearch 的作用、基本概念和关键特性，比如索

引、文档、分片、集群，等等。另外也包含了 Elasticsear 中不同的索引和搜索 API，以及查询 DSL。

第 6 章，描述了如何利用 Kibana 来搜索、查看存储在 Elasticsearch 索引中的数据，以及如何对着写数据进行实时的交互操作。本章探索了各种不同的搜索选项，以及如何使用 Kibaba 界面的搜索页功能。

第 7 章，通过一些案例来描述如何使用 Kibana 可视化和仪表盘的细节。同时也对设置页做了一些说明，包括配置索引模式、衍生字段，等等。

第 8 章，演示了如何整合 ELK 技术栈的三个技术组件，来构建完整的数据管道，也再次解释了前面章节中提过的各组件的功能。

第 9 章，解释了在生产环境中使用 ELK 技术栈需要注意的关键点。并且也提供了一些不同生产环境中的 ELK 技术栈实际案例。

第 10 章，描述了通过联合使用 ELK 的周边工具，来扩展技术栈的能力。

阅读本书需要准备什么

- Unix 操作系统（任意喜欢的版本都可以）
- Elasticsearch 1.5.2
- Logstash 1.5.0
- Kibana 4.0.0032

本书面向的读者

本书适合任何希望低成本分析数据的人。读者不一定需要有 ELK 技术栈或者某些组件的背景知识，但熟悉 NoSQL 数据库或者掌握一些编程基础知识会有帮助。

约定

本书中使用了很多格式的文本，以区分各种不同的信息。这里我们举例说明这些格式，并解释它们的含义。

在文本、数据库表名、文件夹名称、文件名、文件扩展名、路径名、伪 URL、用户输入和 Twitter 中，格式是这样的，例如："上面的命令会安装 `rabbitmq` 输入插件。"

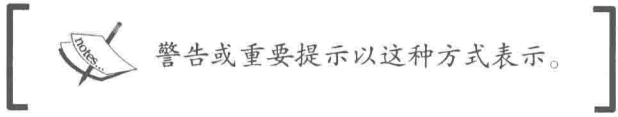
代码格式如下：

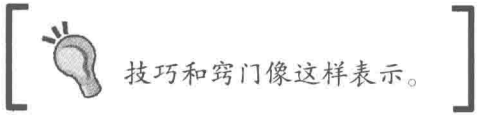
```
filter {
drop {
}
}
```

编写任何命令行输入或输出如下：

```
$bin/plugin install logstash-input-rabbitmq
```

新术语和**重要词汇**以黑体显示。在屏幕上看到的词，例如，在菜单或对话框上，你这样表示："现在点击 Next 按钮，将移动到下一屏幕。"





下载示例代码

你可以从 `http://www.broadview.com.cn` 下载所有已购买的博文视点书籍的示例代码文件。

勘误表

虽然我们已经尽力谨慎地确保内容的准确性，但错误仍然存在。如果你发现了书中的错误，包括正文和代码中的错误，请告诉我们，我们会非常感激。这样，你不仅帮助了其他读者，也帮助我们改进后续的出版。如发现任何勘误，可以在博文视点网站相应图书的页面提交勘误信息。一旦你找到的错误被证实，你提交的信息就会被接受，我们的网站也会发布这些勘误信息。你可以随时浏览图书页面，查看已发布的勘误信息。

目录