精通 Elastic Stack

用于各种数据
复杂分析的
综合性实用指南

面向大数据的
搜索与挖掘及
可视化管理方案

[印]Y.古普塔（Yuvraj Gupta）著
[印]R.K.古普塔（Ravi Kumar Gupta）
高凯 岳重阳 苗雪立 张思琪 译

清华大学出版社

精通 Elastic Stack

[印]Y.古普塔（Yuvraj Gupta）
[印]R.K.古普塔（Ravi Kumar Gupta） 著
高凯 岳重阳 苗雪立 张思琪 译

清华大学出版社
北京

北京市版权局著作权合同登记号　图字：01-2017-6579
Yuvraj Gupta and Ravi Kumar Gupta
Mastering Elastic Stack
ISBN-13：978-1786460011

图书在版编目(CIP)数据

精通 Elastic Stack/(印)Y. 古普塔(Yuvraj Gupta)，(印)R. K. 古普塔(Ravi Kumar Gupta)著；高凯，等译. —北京：清华大学出版社，2018(2018.10重印)
书名原文：Mastering Elastic Stack
ISBN 978-7-302-49243-6

Ⅰ. ①精… Ⅱ. ①Y… ②R… ③高… Ⅲ. ①数据处理 Ⅳ. ①TP274

中国版本图书馆 CIP 数据核字(2017)第 329710 号

责任编辑：焦　虹
封面设计：傅瑞学
责任校对：时翠兰
责任印制：沈　露

出版发行：清华大学出版社
网　　址：http://www.tup.com.cn，http://www.wqbook.com
地　　址：北京清华大学学研大厦 A 座　　**邮　　编**：100084
社 总 机：010-62770175　　**邮　　购**：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
课件下载：http://www.tup.com.cn，010-62795954
印 装 者：三河市铭诚印务有限公司
经　　销：全国新华书店
开　　本：185mm×240mm　　**印　　张**：26.75　　**字　　数**：544 千字
版　　次：2018 年 8 月第 1 版　　**印　　次**：2018 年10月第3次印刷
定　　价：79.90 元

产品编号：076709-01

2017 Packt 出版社版权声明

首次出版：2017 年 2 月

产品基准码：1240217

出版商：Packt Publishing Ltd.

地址：Livery Place

35 Livery Street

Birmingham

B3 2PB，UK.

www. packtpub. com

荣誉及致谢

作者

Ravi Kumar Gupta

Yuvraj Gupta

审校人

Israel Farfan

Marcelo Ochoa

组稿编辑

Veena Pagare

采稿编辑

Tushar Gupta

内容编辑

Narendrakumar Tripathi

技术编辑

Sachit Bedi

文字编辑

Safis Editing

项目协调人

Devanshi Doshi

校对

Safis Editing

索引

Aishwarya Gangawane

图表

Disha Haria

产品协调

Deepika Naik

译 者 序

在大数据时代，建立一个网站或应用程序，搜索、挖掘与分析功能是必备的。本书从分布式大数据搜索、日志挖掘、可视化、数据监控与管理等多个角度出发，在 Elastic Stack 5 的基础上，介绍了 Elasticsearch、Logstash、Kibana、Beats、X-Pack 等诸多相关组件。原著作者 Ravi Kumar Gupta 除作为计算机专业技术书籍的审稿人外，也是开源软件社区的维护者；原著的另一名作者 Yuvraj Gupta 是大数据实践领域的技术顾问，其研究领域涉及大数据、数据分析、数据可视化和云计算等。二人合作完成的这部著作，从面向实践的角度出发，比较全面地介绍了 Elastic Stack 5 的实际应用；并结合一些项目实例，介绍了大数据分析的部分关键技术。我们认为，无论对初学者还是有经验的开发人员，原著都是很有参考价值的。它不仅内容全面，而且表达比较通俗易懂，实践指导性较强。原著强调实践、面向初学者，通过实战讲解的方式，可让读者更好地了解相关组件的应用。通过翻译这本书，我们也从中收获很多、受益颇丰。

本书由高凯、岳重阳、苗雪立、张思琪合作翻译。其中，高凯完成了第 1、3、7 章，岳重阳完成了第 5、9、10、11、12 章，苗雪立完成了第 4、6 章，张思琪完成了第 2、8 章。最后，由高凯统稿。本书翻译过程中得到了多方面的支持与帮助；何晓艺、张姗姗、孟天宏、刘多星、高成亮、毛雨欣、聂颖杰、韩佳、谢宇翔、李明奇、侯雪飞、杨聪聪、江跃华等均提供了协助。

尽管本书的译者在大数据搜索与挖掘及可视化管理方面有一定的经验，也出版过相关的著作、教材等，但毕竟水平有限，译文中难免有不足和有待商榷之处，敬请读者批评指正。

译 者

关于本书作者

本书作者 Ravi Kumar Gupta 是计算机专业技术书籍审稿人、开源软件社区维护者。他于印度彼拉尼邦的伯拉科技学院(Birla Institute of Technology and Science, BITS)获得软件系统硕士学位,于印度拉贾斯坦邦斋浦尔(Jaipur)的 LNMIIT(The LNM Institute of Information Technology)获学士学位,在技术上擅长门户网站研发。

他目前就职于 Azilen Technologies 公司,任技术架构师和项目经理,曾担任 CIGNEX Datamatics 的首席技术顾问[①]。他曾是开源组织 TCS[②] 的核心成员,从事开源软件社区管理和其他的用户界面技术研发。在其职业生涯中,他致力于使用最新技术构建企业级解决方案,并注重用户界面和开源工具的使用。

他喜爱写作、学习,热衷于讨论 IT 新技术。大学期间他的研究领域涉及基于爬虫的搜索引擎的研发,是技术爱好者。他也是由 Packt 出版社出版的 *Test-Driven JavaScript Development* 的作者之一,是软件社区论坛的活跃成员。他目前维护其博客 http://techdc.blogspot.in,并经常在上面发表计算机相关技术的系列文章。

他还维护着 TCS 和 CIGNEX 软件社区(Liferay 5.x 和 6.x 版本),同时也是 Packt 出版社出版的 *Learning Bootstrap* 的审校人。其联系方式如下:

Skype: kravigupta; Twitter: @kravigupta; LinkedIn: https://in.linkedin.com/in/kravigupta

"感谢我的妻子 Kriti。正是她的鼓励和支持,伴我度过了本书写作的艰辛时光。感谢我的妻子、我的家庭,特别是我的岳父母,他们为我提供了很多帮助。更要感谢本书合著者 Yuvraj。作为好朋友,他为我提供了很好的支持和理解,没有他,本书是不可能完成的。我还要感谢 Packt 出版社、审校人、编辑团队的同事们。感谢你们!谢谢!"

本书的另一位作者 Yuvraj Gupta 的研究领域涉及大数据、数据分析、数据可视化和云计算。他是大数据实践领域的技术顾问,喜爱各种社交平台,是小工具插件的开发与爱好者。他喜爱美食、运动、各种影视剧,也一直跟踪最近技术发展动态。他在 Packt 出版社出版了 *Kibana Essentials*。其联系方式如下:

E-mail: gupta.yuvraj@gmail.com; LinkedIn: www.linkedin.com/in/guptayuvraj

---

① 译者注:http://www.cignex.com/。

② 译者注:https://sourceforge.net/projects/opentcs/files/?source=navbar。

关于本书审校人

Marcelo Ochoa 就职于 Universidad Nacional del Centro de la Provincia de Buenos Aires 系统实验室，是 Scotas.com（这是一家使用 Apache Solr 和 Oracle 的实时搜索公司）的 CTO。平时，除完成大学中的工作，他也从事和 Oracle 及大数据技术相关的工作，做过一些和 Oracle 相关的工作（如 Oracle 手册的翻译、CBT 多媒体等），其技术背景为数据库、网络、互联网、Java 技术等。在 XML 领域，他以 Apache Cocoon 工程的 DB Generator 的研发者而著称。他也在开源项目 DBPrism 和 DBPrism CMS（这是一个由 Oracle JVM Directory 实现的基于 Lucene-Oracle 的集成项目）中提供服务，在网站 https://restlet.com/project 有工作经历。他在项目中主要从事 Oracle XDB Restlet Adapter 的开发工作，这是 Oracle JVM 中专注于 REST 的 Web 服务的一个替代品。从 2006 年起，他成为 Oracle ACE program 中的一员。Oracle ACE program 在 Oracle 社区享有崇高声望，拥有众多热情的支持者和倡导者。他也是 ACES 在 Oracle 技术和应用社区的倡导者。作为合著者，他参与编写了由 Digital 出版社出版的 *Oracle Database Programming using Java and Web Services* 和由 Wrox 出版社出版的 *Professional XML Databases*。他还是由 Packt 出版社出版的几部技术著作，如 *Apache Solr 4 Cookbook*，*ElasticSearch Server* 等的审校人。

www.packtpub.com

要获取本书相关程序文件，可登录 www.packtpub.com。

你知道吗？Packt 出版社为出版的每一部书籍提供 PDF、ePub 格式的电子书。可以登录 www.packtpub.com 更新、下载电子书，并可享受电子书折扣。如果需要，请联系我们 service@packtpub.com。

在 www.packtpub.com，也能阅读一些免费的技术资料。注册成为自由的信息提供者，将会获得更高折扣并获得由 Packt 出版社提供的纸质书和电子书。

https://www.packtpub.com/mapt

访问 Mapt 网站，可获取更多相关内容。该网站提供所有 Packt 出版社的书籍、视频课程和在工业界领先的工具软件等。这些能帮助你规划未来的个人发展，并获得职业提升的机会。

可以通过浏览器访问该网站，复制、粘贴、打印、收藏由 Packt 出版社发行的电子书。

致消费者

感谢购买由 Packt 出版社出版的图书。在 Packt 出版社，保证图书质量是一切编辑活动的核心。为了帮助我们提高图书质量，诚邀您访问本书在亚马逊的网站：https://www.amazon.com/dp/1786460017，并留下您宝贵的意见及建议。

如果有意成为审稿人，可发送电子邮件到 customerreviews@packtpub.com。为表示感谢，将会赠送您免费的电子书和视频资料。

让我们携手努力，共同提高图书质量！

前言

Foreword

如果数据不能有助于做出决定、提升目前的系统性能，即使是结构化的数据，那也是无用的（更何况非结构化的数据了）。如果对数据感兴趣，或者需要处理用户各种类型的日志数据，或者需要设计高可扩展性的分析系统，或者需要管理日志并进行实时的数据分析，本书可提供“一站式”解决方案。通过集成 Elasticsearch、Logstash、Beats、Kibana 等多个流行软件，ELK Stack 已经进化为 Elastic Stack，它能以近乎实时的高效处理方式，处理几乎各种类型的结构化和非结构化的数据。

本书首先介绍有关 Elastic Stack 的基础知识，之后会涉及一些更复杂和高级的内容。我们将帮助你借助 Elastic Stack，应对数据分析的挑战，并以内网应用环境为例，带你从实战角度理解 Elastic Stack 组件的使用。通过学习，你将会了解日志分析和可视化的高级技术。另外，也将以实例的方式介绍一些新特性——如 Beats 和 X-Pack。

最终，你将会看到如何使用 Elastic Stack 解决现实世界中的实际问题。同时，本书也将会介绍一些在应用 Elastic Stack 中需要注意的问题。

本书包含哪些内容？

第 1 章，Elastic Stack 概述。通过搭建 Elastic Stack 的各种组件，介绍从 ELK 到 Elastic Stack 的转化。

第 2 章，使用 Elasticsearch。介绍如何在工程项目中开始使用 Elasticsearch，介绍 Elasticsearch 工作机制，并介绍各种 Elasticsearch API 和聚合 Aggregations 的用法。

第 3 章，Logstash 及其插件的使用。内容涉及 Logstash 简介、Logstash 结构、各种插件的用法示例。最后，介绍一个有关 Logstash 配置文件以及日志解析的实例。

第 4 章，Kibana 界面设计。介绍各种 Kibana 界面的用法，并通过一些例子来演示如何将各种界面相结合并设计面板可视化。

第 5 章，使用 Beats。介绍 Beats，讲述 Beats 和 Logstash 的不同之处，

并介绍各种类型的 Beats 的功能以及设置方法，最后介绍在 Elastic Stack 中如何使用 Beats。

第 6 章，Elastic Stack 实战。介绍在局域网环境下实际使用 Elastic Stack 的方法，并通过例子解释如何使用 Elastic Stack 组件解决实际的具体问题。

第 7 章，个性化配置 Elastic Stack。介绍如何扩展 Elastic Stack 中的各种组件，并介绍定制个性化组件的方法。

第 8 章，Elasticsearch API。通过介绍各种 Elasticsearch API 的用法，使读者理解 Elasticsearch 诸模块的工作机制；并介绍节点、组件发现策略及使用 Java 客户端实现对 Elasticsearch 的各种操作。

第 9 章，X-Pack 插件中的 Security 与 Monitoring 组件。内容涉及 X-Pack 简介与安装、安全和监控等，本章还涉及 Shield、Marvel、Profiler 相关功能的使用。

第 10 章，X-Pack 插件中的 Alerting、Graph 和 Reporting 组件。本章还介绍了 Watcher、Graph、Reporting 等组件的使用、功能、特性等。

第 11 章，最佳实践范例。本章将分成多个小节，使读者理解为什么需要遵循最佳实践标准。

第 12 章，案例分析——Meetup。通过扩展 Logstash 的功能、生成新的功能插件等方法，使读者加深对相关问题的理解，学习如何应用 Elastic Stack 组件来分析和处理端到端的 Meetup 数据，展示 Elastic Stack 在数据分析方面的强大功能。

需要下载什么软件工具？

为了能运行书中的示例，下表列出了可能需要的软件和工具。通过表中列出的链接，可下载相应章节需要的软件。

软　件	版本	链　接
Elasticsearch	5.1.1	https://www.elastic.co/downloads/past-releases/elasticsearch-5-1-1
Logstash	5.1.1	https://www.elastic.co/downloads/past-releases/logstash-5-1-1
Kibana	5.1.1	https://www.elastic.co/downloads/past-releases/kibana-5-1-1
Filebeat	5.1.1	https://www.elastic.co/downloads/past-releases/filebeat-5-1-1
Packetbeat	5.1.1	https://www.elastic.co/downloads/past-releases/packetbeat-5-1-1
Winlogbeat	5.1.1	https://www.elastic.co/downloads/past-releases/winlogbeat-5-1-1
Metricbeat	5.1.1	https://www.elastic.co/downloads/past-releases/metricbeat-5-1-1
Elasticsearch	1.4.1	https://www.elastic.co/downloads/past-releases/elasticsearch-1-4-0

续表

软 件	版本	链 接
Liferay	6.2CEGA4	https://sourceforge.net/projects/lportal/files/Liferay Portal/6.2.3 GA4/liferay-portal-tomcat-6.2-ce-ga4-20150416163831865.zip/download①
Java	8.x	http://www.oracle.com/technetwork/java/javase/downloads/index.html
Elasticray	1.2.0	https://web.liferay.com/marketplace/-/mp/application/41044606
Go	1.7.5	https://golang.org/dl
Ruby	2.4.0	https://www.ruby-lang.org/en
NodeJS	6.9.0	https://nodejs.org/en/download/releases/
Gradle	2.13	https://gradle.org/gradle-download
Python	2.7.10	https://www.python.org
Virtualenv		https://virtualenv.pypa.io/en/stable/
cookiecutter		https://github.com/audreyr/cookiecutter

谁适合阅读本书?

如果曾经听说过 ELK Stack,想学习有关它的最新发展,了解它如何演化成为 Elastic Stack,那么这本书就是为你而准备的;如果正在进行数据分析,或计划通过可视化技术来展现数据,本书也适合你。它能帮助你了解 Elastic Stack 中的组件是如何发挥作用的。

惯用法与记号说明

为区分不同种类的信息,本书采用了不同风格的文本。这里列出一些例子并进行说明。

文本中的代码、数据表名字、文件夹名、文件名、文件扩展名、路径名、虚拟的 URL、用户输入、Twitter 句柄等以如下方式展示。例如,下面这些代码读取数据并赋值给 BeautifulSoup 函数。

```
#import packages into the project
from bs4 import BeautifulSoup
from urllib.request import urlopen
import pandas as pd
```

当需要提醒注意某部分代码时,以黑体显示。

```
<head>
```

① 译者注:原文中,该 URL 中的空格是用%20 表示的。这里直接以空格表示。

```
<script src="d3.js" charset="utf-8"></script>
<meta charset="utf-8">
<meta name="viewport" content="width=device-width">
<title>JS Bin</title>
</head>
```

基于命令行的输入或输出，将按如下方式展现：

```
C:\Python34\Scripts>pip install -upgrade pip
C:\Python34\Scripts>pip install pandas
```

新词和重要词汇以黑体显示。在屏幕、菜单或对话框中显示的内容，将采用如下方式展示。例如，为了下载新的模块，将在 **Files**|**Settings**|**Project Name**|**Project Interpreter** 菜单选项中完成相应的操作。

> 警告或重要的说明将出现在这种框中。

> 提示和操作技巧将出现在这种框中。

读者意见反馈

欢迎读者提出反馈意见。请告诉我们，针对本书有什么意见？喜欢什么？不喜欢什么？读者的意见反馈对作者来说是很重要的，可便于我们进一步修改、完善内容。可以发送电子邮件到 feedback@packtpub.com。请在邮件主题中注明本书书名。如果您是某一个领域的专家并且愿意撰写相关内容的图书，请参阅 Packt 出版社如下网站中的“作者须知”：www.packtpub.com/authors。

读者服务

作为 Packt 出版社尊贵的读者，从购买本书开始，您将享受 Packt 出版社提供的各种服务。

下载示例代码

可以在 Packt 出版社提供的 http://www.packtpub.com 网站，用您的账号下载本书的示例代码。不论在哪里购书，均可访问 http://www.packtpub.com/support。完成注册后，我们会通过电子邮件给您发送相关文件。

按照如下步骤下载代码文件：

(1) 用您的电子邮件地址和设定的密码访问我们的网站，完成用户登录或新用户注册。

(2) 定位到页面顶部的 SUPPORT 标签页。

(3) 单击 Code Downloads & Errata。

(4) 在搜索框中输入书名。

(5) 选择要下载代码文件的书名。

(6) 从下拉菜单中选择所购买的图书。

(7) 单击 Code Download。

完成下载后，确保计算机中有如下的新版解压缩软件。

- Windows 环境：WinRAR/7-Zip；
- Mac 环境：Zipeg/iZip/UnRarX；
- Linux 环境：7-Zip/PeaZip。

本书提供的代码文件也可从 Github 下载：https://github.com/PacktPublishing/Mastering-Elastic-Stack。其他更丰富的相关资源和视频也可访问：https://github.com/PacktPublishing/。去试试吧！

勘误表

虽然我们已经尽力确保书中内容的准确性，但错误可能仍无法完全避免。如果在书中文字或代码中发现了错误，请告诉我们，我们将不胜感激。这不仅能避免错误给读者带来疑惑，也能帮助我们提高图书再版的质量。如果发现书中有错误，请访问 http://www.packtpub.com/submit-errata，选择该书，单击 Errata Submission Form 链接，输入有关勘误的信息。一旦确认，提交的勘误信息将会更新到我们的网站，或者追加到现有的勘误表中。

要查阅早期提交的勘误信息，可访问 https://www.packtpub.com/books/content/support。在搜索框中输入书名，会出现相关的勘误信息。

著作权

在各种媒体上，通过互联网对著作权侵权是目前经常发生的问题。在 Packt 出版社，我们非常注重对著作版权的保护。如果在互联网上见到对我们作品的非法获取、复制等各类盗版行为，请提供给我们有关的地址、网站等信息，以便进行处理。

请联系我们：copyright@packtpub.com，并请提供疑似盗版的相关链接。

非常感谢您在保护著作权方面为我们提供的帮助，以便我们能为您带来更多有价值的内容。

其他问题

如果您有关于本书的其他任何问题，可以通过如下电子邮件联系我们 questions@packtpub.com。我们将竭诚为您提供帮助。

Contents

目录