数据分析与决策技术丛书
HZ BOOKS
华章IT

The Definitive Guide to Elasticsearch+Logstash+Kibana, Second Edition

ELK Stack权威指南

第2版

饶琛琳◎编著

Elasticsearch+Lo　　析解决方案，基于ELK

5.0版本全面更新

从基础部署到千亿级扩展方案，从性能优化到插件开发，从数据模型到源码解析，全方位解析ELK，融入了作者多年日志分析、数据挖掘的实战经验

机械工业出版社
China Machine Press

数据分析与决策
技术丛书

The Definitive Guide to Elasticsearch+Logstash+Kibana, Second Edition

ELK Stack权威指南

第2版

饶琛琳◎编著

机械工业出版社
China Machine Press

图书在版编目（CIP）数据

ELK Stack 权威指南 / 饶琛琳编著 . —2 版 . —北京：机械工业出版社，2017.4
（数据分析与决策技术丛书）

ISBN 978-7-111-56329-7

I. E… II. 饶… III. 数据处理软件 - 指南 IV. TP274-62

中国版本图书馆 CIP 数据核字（2017）第 050727 号

ELK Stack 权威指南（第 2 版）

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：吴 怡 责任校对：殷 虹
印 刷：北京市荣盛彩色印刷有限公司 版 次：2017 年 5 月第 2 版第 1 次印刷
开 本：186mm × 240mm 1/16 印 张：26
书 号：ISBN 978-7-111-56329-7 定 价：79.00 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

Preface 前　言

《ELK Stack 权威指南》第 1 版面世之后的这一年多时间里，ELK Stack 在 Elastic.co 公司以及社区的共同努力下飞速发展。国内外都出现了不少基于 ELK Stack 实现的日志分析产品和创业公司。ELK Stack 已经成为 DevOps 技术栈中必不可缺少的一个部分，较大型的互联网公司甚至已经配备有专职的 ELK Stack 管理团队。

对于并不精通 ELK Stack 技术及其发展历史的人来说，过去复杂的版本对应是新手的第一道门槛。最近全新更新的 ELK Stack 各组件，统一使用 5.x 系列版本号，大大方便了新手入门。而 5.x 系列同样携带了大量崭新的特性，在日志分析、监控告警等场景，带来性能提升、管理简化、功能丰富等诸多好处。推荐广大读者积极尝试和升级。

IT 运维模式正在向数据驱动、精细化、智能化发展。这个过程中，ELK Stack 恰好是运维人员达成这个目的最方便的工具和平台。基于 ELK Stack 平台，越来越多的周边开源项目在涌现。这次再版，也进一步丰富了这些周边项目的介绍。

与第 1 版相比，第 2 版修订、删补了 180 多页内容，接近全书的一半。修改期间，怀孕的妻子一直默默陪伴左右，时不时叮嘱我注意保存。谨以此书献给她和刚出生的启舟宝贝，我爱你们！

本书章节内容

本书包括三大部分共 19 章，各部分可以独立阅读。但对于还没有大规模应用经验的新手，建议按顺序阅读全文。

第一部分　Logstash

第 1 章：入门示例。该章介绍 Logstash 及其插件的配置安装方法，自定义配置语言的设

计用途，并为不熟悉 Linux 系统管理的开发人员介绍了多种后台运行方式。

第 2 章：插件配置。该章列举 Logstash 最常用的几十种插件，通过实际示例和效果，讲解各插件的配置细节和用途。

第 3 章：场景示例。该章以最常见的运维、网络、开发和数据库场景，介绍 Logstash 处理 Nginx、Postfix、Ossec、Log4J、MySQL、Docker 等日志的最佳实践。

第 4 章：性能与监控。了解 Logstash 的性能情况一直是个难题，该章从 Logstash 设计原理和 JVM 平台本质出发，介绍几种行之有效的检测和监控方案。

第 5 章：扩展方案。该章介绍采用 Redis 和 Kafka 完成 Logstash 水平扩展的方案，同时也介绍其他几种日志收集系统与 Logstash 的配合方式。

第 6 章：Logstash 源码解析。该章解析 Logstash 源码中最重要的 Pipeline 设计，以及 Logstash::Event 的来龙去脉。

第 7 章：插件开发。该章以最常见的用户登录记录和地址库解析、Consul 数据更新等需求，实际演示 Logstash 的自定义 Filter、Input 和 Output 插件的编写，同时还涉及了插件打包的 RubyGems 规范共有 HttpClient 功能项等细节。

第 8 章：Beats。该章讲述 ELK Stack 家族新成员 Beats 生态圈各组件的使用，包括 Filebeat、packetbeat、metricbeat、winlogbeat 等内容。

第二部分　Elasticsearch

第 9 章：架构原理。该章从更高级的架构层面，介绍 Elasticsearch 分布式设计中涉及稳定性和高性能的部分原理，并由此引发相关的优化配置介绍。另外，还提供了一种针对时序数据索引的读写分离方案，适用于拥有少部分 SSD 设备的用户。

第 10 章：数据接口用例。该章介绍 Elasticsearch 的 RESTful 接口的基础知识，并针对常见的重建索引需求提供两种快速实现方案，为有 Spark 经验的读者介绍通过 Spark Streaming 接口读写 Elasticsearch 的方法。

第 11 章：性能优化。该章介绍 Elasticsearch 在日志处理场景下的读写优化知识和官方推荐的 curator 工具，其中重点介绍了 Elasticsearch 中几种不同 cache 的区别和有效场景。

第 12 章：测试和扩展方案。该章介绍 Elasticsearch 在生产环境中需要的一些周边工具，比如 Puppet 配置管理、Shield 权限管理、版本升级操作、别名切换流程设计等。新增了快照与恢复功能。

第 13 章：映射与模板的定制。该章详细介绍 Elasticsearch 中的核心类型及其对应的常见映射设置，以及如何通过动态模板简化映射定制操作的复杂度。

第 14 章：监控方案。Elasticsearch 作为一个分布式系统，也是有一定的运维难度的，因此其本身的监控也相当重要。该章介绍 Elasticsearch 自带的一系列监控接口，以及由此衍生的

多种实时或长期的监控方案。

第 15 章：Elasticsearch 在运维监控领域的其他应用。该章介绍 Elasticsearch 在运维方面的其他运用方式，包括实时过滤接口、定时报警系统设计、时序数据存储和相关性排序等。

第三部分　Kibana

第 16 章：Kibana 的产品对比。该章介绍 Kibana 3 与 Kibana 5 之间，以及它们与 Hadoop、Splunk 之间的差异，方便读者在不同场景需求下选择更正确的工具。

第 17 章：Kibana 5。该章介绍 Kibana 5 的安装部署和界面操作方式，重点介绍 Kibana 5 提供的几种可视化图表的配置细节和效果，并以几种场景的日志分析需求演示了 Kibana 5 全新的子聚合功能的效果。最后还介绍了一种采用 phantom.js 截图方式记录长期报表数据的方案。

第 18 章：Kibana 5 源码解析。该章介绍 Kibana 4 的界面实现，重点包括其内部 ORM 实现的 Counrier 类、可视化绘图的 Vislib 类等。

第 19 章：Kibana 插件开发示例。该章讲述 Kibana 最常用的插件类型二次开发实例，包括可视化效果、服务器段进程、完整 App 演示等内容。

致谢

我本人虽然接触 ELK 较早，但本身专于 Web 和 App 应用数据方面，动笔以来得到诸多朋友的帮助，在此深表感谢。此外，还要特别感谢 Elastic.co 公司的曾勇（medcl）和吴晓刚（Wood），曾勇完成 Elasticsearch 在国内的启蒙式分享，并主办 Elasticsearch 中国用户大会，吴晓刚积极帮助新用户，并最早分享了携程的 ELK 日亿级规模的实例。

目　录 *Contents*

第一部分 Part 1

Logstash

Logstash is a tool for managing events and logs. You can use it to collect logs, parse them, and store them for later use (like, for searching). ——http://logstash.net

Logstash 项目诞生于 2009 年 8 月 2 日。其作者是世界著名的运维工程师乔丹·西塞 (Jordan Sissel)，乔丹·西塞当时是著名虚拟主机托管商 DreamHost 的员工，还发布过非常棒的软件打包工具 fpm，并主办着一年一度的 Sysadmin Advent Calendar（advent calendar 文化源自基督教氛围浓厚的 Perl 社区，在每年圣诞来临的 12 月举办，从 12 月 1 日起至 12 月 24 日止，每天发布一篇小短文介绍主题相关技术）。

Logstash 动手很早，对比一下，Scribed 诞生于 2008 年，Flume 诞生于 2010 年，Graylog2 诞生于 2010 年，Fluentd 诞生于 2011 年。Scribed 在 2011 年进入半死不活的状态，大大激发了其他各种开源日志收集处理框架的蓬勃发展，Logstash 也从 2011 年开始进入 commit 密集期并延续至今。

作为一个系出名门的产品，Logstash 的身影多次出现在 Sysadmin Weekly 上，它和小伙伴们 Elasticsearch、Kibana 直接成为了和商业产品 Splunk 做比较的开源项目（乔丹·西塞曾经在博客上承认设计想法来自 AWS 平台上最大的第三方日志服务商 Loggly，而 Loggly 两位创始人都曾是 Splunk 员工）。

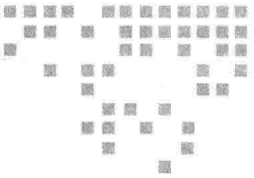
2013 年，Logstash 被 Elasticsearch 公司收购，ELK Stack 正式成为官方用语。Elasticsearch 本身也是近两年最受关注的大数据项目之一，三次融资已经超过一亿美元。在 Elasticsearch 开发人员的共同努力下，Logstash 的发布机制、插件架构也愈发科学和合理。

社区文化

日志收集处理框架很多，如 Scribe 是 Facebook 出品，Flume 是 Apache 基金会项目，都算声名赫赫。但 Logstash 因乔丹·西塞的个人性格，形成了一套独特的社区文化。每一个在 Google Groups 的 Logstash-users 组里问答的人都会看到这么一句话：

Remember: if a new user has a bad time, it's a bug in Logstash.

所以，Logstash 是一个开放的、极其互助和友好的大家庭。如有问题，仅管在 Github Issue、Google Groups、Freenode#logstash Channel 上发问就好！



第 1 章 Chapter 1

入门示例

什么是 Logstash？为什么要用 Logstash？怎么用 Logstash？这是本章将要介绍的内容。本章从最基础的知识着手，从以下几步介绍 Logstash 的必备知识。1）下载安装。介绍 Logstash 软件的多种安装部署方式，并给出推荐的方式。2）初次运行。通过 Hello World 示例，演示 Logstash 最简单的运用，解释其逻辑上的基础原理。3）配置语法。介绍 Logstash 的 DSL 设计，Logstash 命令的运行参数。4）插件安装。灵活和丰富的插件是 Logstash 最重要的优势。本节会介绍 Logstash 插件的安装方式。5）长期运行方式。从初次终端测试到长期后台稳定运行，本节会介绍几种不同方案，供读者根据实际场景选择。

1.1 下载安装

1. 下载

Logstash 从 1.5 版本开始，将核心代码和插件代码完全剥离，并重构了插件架构逻辑，所有插件都以标准的 Ruby Gem 包形式发布。

下载官方软件包的方式有以下几种：

- 压缩包方式

https://artifacts.elastic.co/downloads/logstash/logstash-5.1.1.tar.gz

- Debian 平台

https://artifacts.elastic.co/downloads/logstash/logstash-5.1.1.deb

- Redhat 平台

https://artifacts.elastic.co/downloads/logstash/logstash-5.1.1.rpm

2. 安装

在上面这些包中，你可能更偏向使用 rpm、dpkg 等软件包管理工具来安装 Logstash，开发者在软件包里预定义了一些依赖。比如，logstash-5.0.2 就依赖于 jre 包。

另外，软件包里还包含有一些很有用的脚本程序，比如 /etc/init.d/logstash。

如果你必须在一些很老的操作系统上运行 Logstash，那你只能用源代码包部署了，记住要自己提前安装好 Java：

```
yum install openjdk-jre
export JAVA_HOME=/usr/java
tar zxvf logstash-5.0.2.tar.gz
```

3. 最佳实践

但是真正的建议是：如果可以，请用 Elasticsearch 官方仓库来直接安装 Logstash！

❑ Debian 平台

```
wget -qO - https://artifacts.elastic.co/GPG-KEY-elasticsearch | sudo apt-key add -
sudo apt-get install apt-transport-https
echo "deb https://artifacts.elastic.co/packages/5.x/apt stable main" | sudo tee
    -a /etc/apt/sources.list.d/elastic-5.x.list
sudo apt-get update && sudo apt-get install logstash
```

❑ Redhat 平台

```
sudo rpm --import
https://artifacts.elastic.co/GPG-KEY-elasticsearch
sudo cat > /etc/yum.repos.d/elk.repo <<EOF
[logstash-5.x]
name=Elastic repository for 5.x packages
baseurl=https://artifacts.elastic.co/packages/5.x/yum
gpgcheck=1
gpgkey=https://artifacts.elastic.co/GPG-KEY-elasticsearch
enabled=1
autorefresh=1
type=rpm-md
EOF
sudo yum install -y logstash
enabled=1
EOF
yum clean all
yum install logstash
```

1.2 Hello World

与绝大多数 IT 技术介绍一样，我们也以一个输出“Hello World”的形式开始学习 Logstash。