

前言

这个世界已然被数据淹没。多年来，我系统地收集和产生的大量数据已让我不知所措。有的技术都集中在如何解决数据存储以及如何优化这些数据。有些看上去都挺美好，直到需要基于这些数据做决策分析的候才根本不是那一回事。

Elasticsearch 是一个分布式、可扩展的搜索与数据分析引擎。它能从一开始就给予数据以搜索、分析和探索的能力，这是通常没有料到的。它存在的原因原始数据如果只是存储在磁盘里面根本就无用。

无论是需要全文搜索，是优化数据的索引，或者两者合，本指南都能帮助你了解其中最基本的概念，从最基本的操作开始学习 Elasticsearch。之后，我会逐渐开始探索更加高级的搜索技术，不断提升搜索体系来满足你的需求。

Elasticsearch 不只是全文搜索，我将介绍优化搜索、数据分析、索引的索引、地理位置和索引等。我将探索如何数据建模来充分利用 Elasticsearch 的水平伸缩性，以及在生产环境中如何配置和管理集群。

本书

本书是写任何想要把他的数据拿来干活做点事情的人。不管是新起一从开始是了留系统改造血，Elasticsearch 都能帮助解决有和新的功能，有些可能是之前没有想到的功能。

本书既合初学者也合有经验的用。我希望有一定的编程基础，虽然不是必须的，但有用 SQL 和系统数据会更好。我会从原理解和基本概念出发，帮助新手在你的搜索世界里得一个定的基础。

具有搜索背景的人也会受益于本书。有经验的用将得其所熟悉搜索的概念在 Elasticsearch 是如何和具体的。即使是高用，前面几个章节所包含的信息也是非常有用的。

最后，我也是一名 DevOps，其他部门一直尽可能快的往 Elasticsearch 里面灌数据，而那个阻止 Elasticsearch 服务器起火的消防。只要用在内行事，Elasticsearch 集群容量相当松。不需要知道如何在生产环境前搭建一个定的集群，能要在凌晨三点能出警告信号，以防止生产。前面几章可能不太感兴趣，但本书的最后部分是非常重要的，包含所有需要知道的用以避免系统崩溃的知。

为什么要写本书

我写本书，因为 Elasticsearch 需要更好的描述。有的参考文献是秀的一前提是知道在什么。它假定已熟悉信息索引的概念、分布式系统原理、Query DSL 和许多其他相关的概念。

本书没有假设。它的目的是写一本即便是一个完全不的初学者（不管是搜索是分布式系统）也能拿起它看完几章，就能开始搭建一个原型。

我采取一种基于求解的方式：是一个什么问题，我如何解决？如何时候方案行衡取？我从基础知识开始，循序渐进，一章都建立在前一章之上，同时提供必要的用案例和理解。

有的参考文献解决了如何使用些功能，我希望本书解决的是什和什时候使用些功能。

Elasticsearch 版本

本的初始印刷版 的是 Elasticsearch 1.4.0, 不 我 一直在不断更新内容和完善示例 本 的 上版本 的是 Elasticsearch 2.x。

可以 本 的 [GitHub](#) 来追踪最新 化。

如何 本

Elasticsearch 做了很多努力和 来 的事情 得 , 很大程度上来 Elasticsearch 的成功来源于此。 句 , 搜索以及分布式系 是非常 的, 不 早 也需要掌握一些来充分利用 Elasticsearch。

恩, 是有点 , 但不是魔法。我 向于 系 如同神奇的 子, 能 外部的咒 , 但是通常 里面的工作 很 。 理解了 些 程 就能 散魔法, 理解内在能 更加明 和清晰, 而不是寄托于 子做 想要做的。

本 威指南不 助 学 Elasticsearch, 而且 接触更深入、更有趣的 , 如 [\[distributed-cluster\]](#)、[\[distributed-docs\]](#)、[\[distributed-search\]](#) 和 [\[inside-a-shard\]](#) , 些 然不是必要的 却能 深入理解其内在机制。

本 的第一部分 是在按章 序 , 因 一章建立在上一章的基 上(尽管 也可以 才提到的章)。 后 各章 如 [\[proximity-matching\]](#) 和 [\[partial-matching\]](#) 相 独立, 可以根据需要 性参 。

本 航

本 分 七个部分:

- 章 1 到 11 主要是介 Elasticsearch。介 了 Elasticsearch 的数据 入 出以及 Elasticsearch 如何 理 的文 数据。 如何 行基本的搜索操作和管理 的索引。 本章 束 将学会如何将 Elasticsearch 与 的 用程序 行集成。 章 : 2、 4、 9 和 11 附加章 , 目的是 了解分布式 理的 程, 不是必 的
- 章 12 到 17 深入了解搜索, 如何索引和 的数据, 并借助一些更高 的特性, 如 近 (word proximity) 和部分匹配 (partial matching) 。 将了解相 度 分是如何工作的以及如何控制它来 保第一 是返回最佳的搜索 果。
- 章 18 到 24 解决如何有效使用分析器和 来 理 言的 痛 。我 会从一个 的 言分析下手, 然后逐 深入, 如字母表和排序, 会 及到 干提取、停用 、同 和模糊匹配。
- 章 25 到 35 聚合 (aggregations) 和分析, 的数据 行摘要化和分 来呈 体 。
- 章 36 到 39 介 Elasticsearch 支持的 地理位置 索方式: 坐 点和 的地理形状 (geo-shapes) 。
- 章 40 到 43 到了如何 的数据建模来高效使用 Elasticsearch。在 系数据 里表 体 的 系很容易, 因 系数据 就是 此而 的, 但是在搜索引 里表 体 的 系就没有那 容易 了。 些章 会 述如何 索引来匹配 系 中的数据流。
- 最后, 章 44 到 46 将 生 境上 的重要配置、 控点和如何 断来避免出 。

在 源

因 本 重如何在 Elasticsearch 里解决 , 而不是 法介 , 所以有 候 需要 Elasticsearch 参考手 来 取 明。 可以 以下 址 取最新的 Elasticsearch 参考手 和相 文 : <https://www.elastic.co/guide/>

如果 遇到本 或者参考手 没有收 到的 , 我 建 Elasticsearch 社区来提 , 学 人是如何使用 Elasticsearch 的或者分享 自己的 :

- [英文社区](#)
- [中文社区](#)

本 定

以下是本 中使用的印刷 :

Italic

表示重点、新的 或概念。

Constant width

用于程序列表以及在段落中引用 量或程序元素如: 函数名称、数据 、数据 型、 境 量、 句和 字。

TIP 个 代表小 士, 建 。

NOTE 个 代表一般注意事 。

WARNING 个 代表警告。

使用代 示例

本 的目的是 了 尽快能完成工作。一般来 , 本 提供的示例代 都可以用于 的程序或文 。 不需要 系我 来 可, 除非 打算 用相当大一部分代 。比如, 写一个程序用了一段本 的代 不需要 可, 但是 或者是 行一 包含所有 O'Reilly 的示例代 的 CD 个就需要 可。引用本 、引用示例代 来回答 不需要 可, 将大量的示例代 从 本 中包含到 的 品的文 中, 个 需要 可。

于署名出 , 我 欣 但不是必 。一个出 通常包含: 名、作者、出版商和 ISBN。如: *Elasticsearch: The Definitive Guide* by Clinton Gormley and Zachary Tong (O'Reilly). Copyright 2015 Elasticsearch BV, 978-1-449-35854-9。

如果 得 的示例代 使用超出合理使用或上面 出的 可, 可随 与我 系 permissions@oreilly.com。

Safari® 在

NOTE

Safari Books Online is an on-demand digital library that delivers expert [content](#) in both book and video form from the world's leading authors in technology and business.

技术家、件人、和商意家使用Safari在作他的主要研究源，用于解决、学和培。

系我

有本的反、和寄本的出版商：

我有本的，我会列出勘表、例子和任何外的信息。可以个接：[http://oreil.ly/1ylQuK6\[\]](http://oreil.ly/1ylQuK6[])。

有本的反和技，件bookquestions@oreilly.com。

了解我的更多信息、程、会和新我的站：[http://www.oreilly.com\[\]](http://www.oreilly.com[])。

我的Facebook主：<http://facebook.com/oreilly>

注我的Twitter: <http://twitter.com/oreillymedia>

看我的YouTube: <http://www.youtube.com/oreillymedia>

什配偶是被放到最后一个？但并非是最不重要！在我心中无疑，有个最得我感的人，他是 Clinton 期受苦的老婆和 Zach 的未婚妻。他照着我和我，不怠，忍受我的席和我没完没了的抱怨本要多久完成，最重要的是，依然在我身。

感 Shay Banon 在最初建了 Elasticsearch，感 Elastic 公司支持本的工作。也非常感 Elastic 所有的同事，他助我透的了解 Elasticsearch 内部如何工作并且一直添加完善和修与他相的部分。

其中位同事特得一提：

- Robert Muir 耐心地分享了他的真知灼，特是 Lucene 搜索方面。有几章段落就是直接出自其智慧珠。
- Adrien Grand 深入到代中回答，并我的解，以保他合理。

感 O'Reilly 承担个目和我一起工作使本免在，有一直温柔哄我的 Brian Anderson 和善良而温柔的者 Benjamin Devèze、Ivan Brusic 和 Leo Lapworth。的鼓励，我充希望。

感我的者，其中一些我只有通各自的 GitHub 才知道他的身，他花告、提供修正或提出改建：

Adam Canady, Adam Gray, Alexander Kahn, Alexander Reelsen, Alaattin Kahramanlar, Ambrose Ludd, Anna Beyer, Andrew Bramble, Baptiste Cabarro, Bart Vandewoestyne, Bertrand Dechoux,

Brian Wong, Brooke Babcock, Charles Mims, Chris Earle, Chris Gilmore, Christian Burgas, Colin Goodheart-Smithe, Corey Wright, Daniel Wiesmann, David Pilato, Duncan Angus Wilkie, Florian Hopf, Gavin Foo, Gilbert Chang, Grégoire Seux, Gustavo Alberola, Igal Sapir, Iskren Ivov Chernev, Itamar Syn-Hershko, Jan Forrest, Jānis Peisenieks, Japheth Thomson, Jeff Myers, Jeff Patti, Jeremy Falling, Jeremy Nguyen, J.R. Heard, Joe Fleming, Jonathan Page, Joshua Gourneau, Josh Schneier, Jun Ohtani, Keiji Yoshida, Kieren Johnstone, Kim Laplume, Kurt Hurtado, Laszlo Balogh, Iondocr, Iosar, Lucian Precup, Lukáš Vlček, Malibu Carl, Margirier Laurent, Martijn Dwars, Matt Ruzicka, Mattias Pfeiffer, Mehdy Amazigh, mhemani, Michael Bonfils, Michael Bruns, Michael Salmon, Michael Scharf, Mitar Milutinović, Mustafa K. Isik, Nathan Peck, Patrick Peschlow, Paul Schwarz, Pieter Coucke, Raphaël Flores, Robert Muir, Ruslan Zavacky, Sanglarsh Boudhh, Santiago Gaviria, Scott Wilkerson, Sebastian Kurfürst, Sergii Golubev, Serkan Kucukbay, Thierry Jossermoz, Thomas Cucchiatti, Tom Christie, Ulf Reimers, Venkat Somula, Wei Zhu, Will Kahn-Greene 和 Yuri Bakumenko。

感谢所有参与本项目的中文作者与校对人，他牺牲了大量宝贵的休息时间，他翻阅内容仔细斟酌，一丝不苟，修改意见真诚以待，各抒己见，不其的行修改与再次校对，这些奉献的可人分是：
， <http://github.com/luotitan>[朗]， <http://github.com/pengqiuyuan>[彭秋源]， <http://github.com/richardwei2008>[魏]， <http://github.com/chenryn>[琳]， <http://github.com/looly>[路小]， <http://github.com/michealzh>[michealzh]， <http://github.com/node>[nodexy]， <http://github.com/sdlyjzh>[sdlyjzh]， <http://github.com/wharstr9027>[落英流]， <http://github.com/sunyonggang>， <http://github.com/zhaochenxiao90>[Singham]， <http://github.com/Josephjin>[]， <http://github.com/lephix>[翔]， <http://github.com/lephix>[思]， <http://github.com/blogsit>[]， <http://github.com/Geolem>[Geolem]， <http://github.com/JessicaWon>[卷]， <http://github.com/kfypmqqw>[kfypmqqw]， <http://github.com/weiqiangyuan>[袁]， <http://github.com/yichao2015>[yichao]， <http://github.com/leo650>[leo]， <http://github.com/tangmisi>[tangmisi]， <http://github.com/cdma>[Alex]， <http://github.com/abia321>[baifan]， <http://github.com/EvanYellow>[Evan]， <http://github.com/fanyer>[fanyer]， <http://github.com/luoruixing>[瑞星]， <http://github.com/Miranda21>[碧琴]， <http://github.com/weikuo0506>[walker]， <http://github.com/javasgl>[songgl]， <http://github.com/kankedong>[]， <http://github.com/smilesfc>[杜]， <http://github.com/qindongliang>[秦亮]， <http://github.com/biyuhao>[biyuhao]， <http://github.com/LiuGangR>[]， <http://github.com/wangxiuwen>[王秀文]， <http://github.com/zcola>[zcola]， <http://github.com/gitqh>[gitqh]， <http://github.com/blackoon>[blackoon]， http://github.com/davidmr_001[David]， <http://github.com/stromdush>[辰]， <http://github.com/echolihao>[echolihao]， <http://github.com/cch123>[Xargin]， <http://github.com/sunzhanya>[abel-sun]， <http://github.com/AlixMu>[]， <http://github.com/donglangdtstack>[冬狼]， <http://github.com/destinyfortune>[王]， <http://github.com/medcl>[Medcl]。