

智能运维：从0搭建大规模分布式AIOps系统



[智能运维：从0搭建大规模分布式AIOps系统_下载链接1](#)

著者:彭冬

出版者:电子工业出版社

出版时间:2018-7

装帧:平装

isbn:9787121346637

《智能运维：从0搭建大规模分布式AIOps系统》结合大企业的智能运维实践，全面完整地介绍智能运维的技术体系，让读者更加了解运维技术的现状和发展。同时，帮助运维工程师在一定程度上了解机器学习的常见算法模型，以及如何将它们应用到运维工作

中。

《智能运维：从0搭建大规模分布式AIOps系统》共分4篇。第1篇运维发展史，重点阐述当前运维的发展现状及面临的技术挑战；第2篇智能运维基础设施，重点讲述大数据场景下的数据存储、大数据处理和分析的方法与经验，以及海量数据多维度多指标的处理分析技术；第3篇智能运维技术详解，重点关注在新时期大数据时代下智能化的运维技术，包括数据聚合与关联、数据异常点检测、故障诊断和分析、趋势预测算法；第4篇技术案例详解，为大家梳理了通过开源框架ELK快速构建智能监控系统的整体方案，还将分享微博平台和微博广告两个不同业务场景下智能监控系统的技术实践。

《智能运维：从0搭建大规模分布式AIOps系统》适合运维、开发、架构、DevOps工程师及广大互联网技术爱好者研读和借鉴。

作者介绍:

彭冬：微博广告基础架构团队负责人、技术专家，商业大数据平台及智能监控平台发起人，目前负责广告核心引擎基础架构、Hubble智能监控系统、商业基础数据平台（D+）等基础设施建设。关注计算广告、大数据、人工智能、高可用系统架构设计、区块链等方向。在加入微博之前，曾就职于百度负责大数据平台建设，曾担任趣点科技联合创始人兼CTO等职位。毕业于西北工业大学，曾在国内外知名期刊发表多篇学术论文，拥有9项发明专利。

朱伟@kimi：微博广告SRE团队技术负责人，高级运维工程师，2016年4月加入微博，目前主要负责微博广告智能监控报警平台和服务治理等项目的建设与研究。

刘俊：微博平台部监控技术负责人，负责微博平台、PC微博大规模监控系统的建设，主要关注实时大数据、运维自动化、智能化方向。2014年加入微博，之前曾在新浪、搜狐等公司从事运维监控方面的工作。

王莉：University of Georgia硕士研究生，主要研究用机器学习方法，识别植物被水淹没的季节性规律，研究成果已发表在SCI高影响因子期刊。2017年加入微博广告团队，致力于用数据分析和机器学习模型，优化广告业务策略，洞悉商业价值。

陆松林：微博广告数据仓库负责人，高级研发工程师，先后就职于搜狐、爱奇艺，主要研究数据仓库、数据治理相关技术。

车亚强：微博广告大数据开发工程师，曾在百度外卖负责实时流、微服务相关研发工作，目前主要研究方向为实时流、微服务架构设计。

目录: 第1篇 开门见山：运维发展史

第1章 运维现状 2

1.1 运维工程 2

1.1.1 认识运维 2

1.1.2 主要职责 4

1.1.3 运维技术 5

1.2 运维发展历程 6

1.2.1 人工阶段 6

1.2.2 工具和自动化阶段 7

1.2.3 平台化阶段 7

1.2.4 智能运维阶段 8

1.3 运维现状 9

1.3.1 故障频发	9
1.3.2 系统复杂性	10
1.3.3 大数据环境	12
1.4 本章小结	14
1.5 参考文献	14
第2章 智能运维	15
2.1 海量数据的存储、分析和处理	16
2.2 多维度、多数据源	18
2.3 信息过载	19
2.4 复杂业务模型下的故障定位	21
2.5 本章小结	22
2.6 参考文献	22
第2篇 站在巨人肩上：智能运维基础设施	
第3章 开源数据采集技术	25
3.1 数据采集工具对比	25
3.2 轻量级采集工具Filebeat	26
3.2.1 Filebeat工作原理	26
3.2.2 Filebeat的安装与配置	28
3.2.3 启动和运行Filebeat	38
3.3 日志采集解析工具	38
3.3.1 Logstash工作原理	39
3.3.2 安装Logstash	40
3.3.3 配置Logstash	41
3.3.4 启动Logstash	49
3.4 本章小结	49
3.5 参考文献	50
第4章 分布式消息队列	51
4.1 开源消息队列对比与分析	51
4.1.1 概述	51
4.1.2 ZeroMQ	51
4.1.3 ActiveMQ	52
4.1.4 RocketMQ	52
4.1.5 Kafka	53
4.2 Kafka的安装与使用	53
4.2.1 组件概念	53
4.2.2 基本特性	53
4.2.3 安装与使用	54
4.2.4 Java API的使用	55
4.3 案例分析	57
4.3.1 日志采集	58
4.3.2 实时结算	58
4.3.3 实时计算	58
4.4 本章小结	58
4.5 参考文献	59
第5章 大数据存储技术	60
5.1 传统数据存储	60
5.1.1 传统应用的架构	60
5.1.2 传统存储的运行机制	61
5.1.3 传统存储带来的问题	62
5.2 基于HDFS的分布式存储	63
5.2.1 分布式存储的定义	63
5.2.2 HDFS的基本原理	64
5.2.3 HDFS架构解析	65
5.2.4 HDFS的优势	66

5.2.5 HDFS不适合的场景	67
5.3 分层存储	68
5.3.1 数据仓库	68
5.3.2 数据仓库分层架构	70
5.3.3 分层存储的好处	73
5.4 案例分析	73
5.4.1 数据存储架构	73
5.4.2 数据仓库建模	74
5.4.3 常见的存储问题及解决方案	80
5.5 本章小结	80
5.6 参考文献	80
第6章 大规模数据离线计算分析	82
6.1 经典的离线计算	82
6.1.1 Linux神级工具sed和awk	82
6.1.2 Python数据处理Pandas基础	84
6.1.3 Python的优势和不足	88
6.2 分布式离线计算	89
6.2.1 MapReduce离线计算	89
6.2.2 离线计算的数据倾斜问题	97
6.2.3 分布式离线计算的技术栈	100
6.3 案例分析	101
6.3.1 离线计算管理	102
6.3.2 离线计算原子控制	103
6.3.3 离线计算的数据质量	103
6.4 本章小结	104
6.5 参考文献	105
第7章 实时计算框架	106
7.1 关于实时流计算	106
7.1.1 如何提高实时流计算的实时性	106
7.1.2 如何提高实时流计算结果的准确性	107
7.1.3 如何提高实时流计算结果的响应速度	107
7.2 Spark Streaming计算框架介绍	107
7.2.1 概述	107
7.2.2 基本概念	108
7.2.3 运行原理	108
7.2.4 编程模型	109
7.2.5 Spark Streaming的使用	110
7.2.6 优化运行时间	114
7.3 Flink计算框架	115
7.3.1 基本概念	116
7.3.2 Flink特点	116
7.3.3 运行原理	118
7.3.4 Java API的使用	121
7.4 案例分析	124
7.4.1 背景介绍	125
7.4.2 架构设计	126
7.4.3 效果分析	126
7.5 本章小结	126
7.6 参考文献	126
第8章 时序数据分析框架	127
8.1 时序数据库简介	127
8.1.1 什么是时序数据库	127
8.1.2 时序数据库的特点	128
8.1.3 时序数据库的对比	130

8.2 时序数据库Graphite	131
8.2.1 Graphite简介	131
8.2.2 Graphite在微博广告监控系统中的应用	137
8.3 多维分析利器Druid	139
8.3.1 什么是Druid	139
8.3.2 Druid架构	140
8.3.3 Druid在微博广告监控平台中的应用	144
8.4 性能神器ClickHouse	147
8.4.1 什么是ClickHouse	147
8.4.2 ClickHouse的特性	148
8.4.3 ClickHouse的不足	149
8.4.4 安装配置ClickHouse	149
8.4.5 表引擎	153
8.4.6 函数支持	157
8.5 本章小结	160
8.6 参考文献	160
第9章 机器学习框架	161
9.1 简介	161
9.2 TensorFlow介绍	162
9.2.1 什么是TensorFlow	162
9.2.2 下载安装	162
9.2.3 “Hello TensorFlow” 示例	166
9.3 TensorFlow进阶	166
9.3.1 基础理论	167
9.3.2 模型准备	169
9.3.3 训练数据	169
9.3.4 模型训练	171
9.3.5 生成seq2seq句子	174
9.3.6 运行演示	175
9.4 本章小结	178
9.5 参考文献	179
第3篇 运维新时代：智能运维技术详解	
第10章 数据聚合与关联技术	182
10.1 数据聚合	182
10.1.1 聚合运算	183
10.1.2 多维度聚合	186
10.2 降低维度	188
10.2.1 将告警聚合成关联“事件”	189
10.2.2 减少误报：告警分类	190
10.3 数据关联	192
10.4 实时数据关联案例	193
10.4.1 设计方案	193
10.4.2 效果	195
10.5 本章小结	195
10.6 参考文献	195
第11章 数据异常点检测技术	196
11.1 概述	196
11.2 异常检测方法	198
11.2.1 基于统计模型的异常点检测	199
11.2.2 基于邻近度的异常点检测	202
11.2.3 基于密度的异常点检测	203
11.3 独立森林	204
11.4 本章小结	207
11.5 参考文献	207

第12章 故障诊断和分析策略	208
12.1 日志标准化	209
12.2 全链路追踪	210
12.3 SLA的统一	210
12.4 传统的故障定位方法	211
12.4.1 监控告警型	211
12.4.2 日志分析型	212
12.5 人工智能在故障定位领域的应用	213
12.5.1 基于关联规则的相关性分析	213
12.5.2 基于决策树的故障诊断	217
12.6 本章小结	222
12.7 参考文献	222
第13章 趋势预测算法	223
13.1 移动平均法	223
13.2 指数平滑法	224
13.3 ARIMA模型	226
13.3.1 简介	226
13.3.2 重要概念	226
13.3.3 参数解释	228
13.3.4 建模步骤	230
13.3.5 ARIMA模型案例	232
13.4 神经网络模型	236
13.4.1 卷积神经网络	236
13.4.2 循环神经网络	238
13.4.3 长短期记忆网络	239
13.4.4 应用说明	241
13.5 本章小结	241
13.6 参考文献	242
第4篇 智能运维架构实践：技术案例详解	
第14章 快速构建日志监控系统	244
14.1 Elasticsearch分布式搜索引擎	244
14.1.1 基本概念	244
14.1.2 分布式文档存储与读取	248
14.1.3 分布式文档检索	250
14.1.4 分片管理	252
14.1.5 路由策略	254
14.1.6 映射	255
14.2 可视化工具Kibana	258
14.2.1 Management	260
14.2.2 Discover	260
14.2.3 Visualize	262
14.2.4 Dashboard	263
14.2.5 Timelion	263
14.2.6 Dev Tools	264
14.3 ELK搭建实践	265
14.3.1 Logstash安装配置	265
14.3.2 Elasticsearch集群安装配置	266
14.3.3 Kibana安装配置	273
14.4 本章小结	274
14.5 参考文献	274
第15章 微博广告智能监控系统	275
15.1 背景介绍	275
15.1.1 监控指标体系	275
15.1.2 功能设计原则	276

15.2 整体架构	277
15.3 核心功能分析	278
15.3.1 全景监控	278
15.3.2 趋势预测	281
15.3.3 动态阈值	285
15.3.4 服务治理	285
15.4 本章小结	287
第16章 微博平台通用监控系统	288
16.1 背景	289
16.2 整体架构	290
16.3 核心模块	291
16.3.1 数据采集 (Logtail)	291
16.3.2 数据路由 (Statsd-proxy)	293
16.3.3 聚合运算 (Statsd)	294
16.3.4 数据分发 (C-Relay) 和数据存储	294
16.3.5 告警模块	295
16.3.6 API设计	299
16.3.7 数据可视化	300
16.4 第三方应用	301
16.4.1 决策支持系统	301
16.4.2 运维自动化	302
16.4.3 成本分析和容量日报	302
16.4.4 机器学习	302
16.5 本章小节	302
附录A 中国大数据技术大会2017 (BDTC 2017) CSDN专访实录	303
· · · · ·	(收起)

[智能运维：从0搭建大规模分布式AIOps系统_下载链接1](#)

标签

运维

AIOps

计算机

计算机科学

机器学习

ops

ai

computer

评论

罗列一些开源组件的功能原理，罗列的一推算法，真正关于一些智能运维的功能该怎么做就含糊带过，新浪的智能运维系统也是一笔带过，这年头写书都这么随意了吗？从零搭建？能搭起来我跟你姓。写荐序的那几人什么收了多少钱？对得起社区吗？

写的太浅了，比较失望。

国内书籍作者为多个的烂书的概率极大，想了一下也是，多个作者分写各章，没有一个对整书负责的。Alops，概念高大上，落到书中全是一堆水货文章，现在出书的门槛低到这种程度的，开源软件的配置、代码贴贴就成一本书了，搞技术的人这么浮躁，这样出书不仅挣不到资历，相反还暴露了自己的急功近利。最近几年烂技术书籍越来越多了，各种蹭热点，写几篇博文也就算了，出书这不是浪费树木嘛~

知识百科编辑

虽然有些落差，但也算国内在这个领域内少数真枪实战的心得了。重点看 AI 那章就可以了。

对于第八章，第十五章河第十六章第内容很失望

内容涉及到的东西很多，可以概要了解下运维需要的技术栈。书中某些章节有拼凑嫌疑。

一个运维人员的工作总结，也行

简单介绍了一些组件的功能，有一点点借鉴意义。

AI部分可以略过

不过把热门一些技术堆在一起，什么也都没讲透

多人合作，水平参差不齐，适合对运维整体快速了解，深度一般，智能应用场景不够刚需。短期内不看好AIOps，长期看还是有想象空间的。推荐同时看一下《企业级 AIOps 实施建议》白皮书。

参考书，用的时候翻翻就好，web component 选型用

百度内容大汇总？

[智能运维：从0搭建大规模分布式AIOps系统_下载链接1](#)

书评

[智能运维：从0搭建大规模分布式AIOps系统_下载链接1](#)