

《Hadoop集群与安全》

图书基本信息

书名：《Hadoop集群与安全》

13位ISBN编号：9787111480627

出版时间：2014-10-1

作者：Danil Zburivsky,Sudheesh Narayanan

页数：184

译者：刘杰,沈鑫

版权说明：本站所提供下载的PDF图书仅提供预览和简介以及在线试读，请支持正版图书。

更多资源请访问：www.tushu111.com

《Hadoop集群与安全》

内容概要

本书手把手教你手动配置高效的Hadoop集群，以便充分利用Hadoop平台的优势，并为Hadoop生态系统实现强健的端到端的安全保障。

本书分为两部分，共13章：第1章概述主要的Hadoop组件以及选择规划；第2章讲解安装和配置主要Hadoop组件的详细步骤；第3章介绍Sqoop、Hive和Impala的配置步骤；第4章讲解确保各种Hadoop组件安全的方法；第5章指导读者逐步将集群开发至实际应用阶段；第6章介绍如何在虚拟环境中使用Hadoop；第7章详细介绍大数据安全参考框架；第8章详细介绍Hadoop安全保障系统的内部设计细节以及关键概念；第9章提供配置Kerberos并建立安全Hadoop集群的详细步骤；第10章介绍Hadoop生态系统组件之间的交互和通信协议；第11章关注如何集成Hadoop安全模型与企业已有的安全系统；第12章提供保护Hadoop生态系统中敏感数据的详细实现方法；第13章深入介绍大数据平台中安全事件监控系统，提供实现安全流程和策略的最新实践。

《Hadoop集群与安全》

作者简介

作者简介

Danil Zburivsky 资深数据库管理员，目前是全球数据基础构架管理公司Pythian的咨询师，其客户涉及金融、娱乐以及通信领域，主要方向是创建各种Hadoop集群。研究兴趣包括Python编程、机器学习等。

Sudheesh Narayanan 在大数据解决方案咨询与实施领域具有丰富经验的实践者和技术战略家。他在IT领域的经验超过15年，涉及信息管理、商务智能、大数据分析以及云应用和J2EE应用开发等。

译者简介

刘杰 中国科学院软件所副研究员，具有多年Hadoop平台研发与实施经验，与团队一起研发基于Hadoop的可视化大数据分析工具Haflow，该工具应用于医疗、交通等多个领域。研究方向包括企业数据集成、面向大数据的系统软件、数据挖掘等。

沈鑫 毕业于同济大学计算机科学与技术系，资深网络工程师，从事网络安全、管理信息系统的开发与维护，参与开发了多个相关的项目。兴趣爱好是网络安全技术与嵌入式技术。

书籍目录

目 录

译者序

作者简介

审校者简介

前言

第1章 构建Hadoop集群1

1.1 选择Hadoop集群硬件2

1.1.1 选择DataNode硬件3

1.1.2 低存储密度集群4

1.1.3 高存储密度集群5

1.1.4 NameNode和JobTracker硬件配置6

1.1.5 网关和其他辅助服务8

1.1.6 网络配置8

1.1.7 Hadoop硬件总结9

1.2 Hadoop发行版10

1.2.1 Hadoop版本10

1.2.2 选择Hadoop发行版11

1.2.3 Cloudera Hadoop 发行版11

1.2.4 Hortonworks Hadoop发行版12

1.2.5 MapR12

1.3 为Hadoop集群选择操作系统13

1.4 小结14

第2章 安装和配置Hadoop15

2.1 在Hadoop集群中配置操作系统15

2.1.1 选择和设置文件系统15

2.1.2 设置Java开发包16

2.1.3 其他操作系统设定17

2.1.4 设置CDH存储库18

2.2 设置NameNode18

2.2.1 JournalNode节点、ZooKeeper以及故障转移控制器22

2.2.2 Hadoop配置文件23

2.2.3 NameNode高可用方案配置25

2.2.4 JobTracker配置31

2.2.5 DataNode配置36

2.3 小结47

第3章 配置Hadoop生态系统48

3.1 托管Hadoop生态项目48

3.2 Sqoop49

3.2.1 安装和配置Sqoop49

3.2.2 Sqoop导入示例50

3.2.3 Sqoop导出示例52

3.3 Hive52

3.3.1 Hive架构53

3.3.2 安装Hive Metastore54

3.3.3 安装Hive客户端 56

3.3.4 安装Hive Server57

3.4 Impala59

- 3.4.1 Impala架构59
- 3.4.2 安装Impala state store60
- 3.4.3 安装Impala server60
- 3.5 小结63
- 第4章 Hadoop安全64
 - 4.1 Hadoop安全概述64
 - 4.2 Hadoop分布式文件系统安全65
 - 4.3 MapReduce安全66
 - 4.4 Hadoop服务级别验证68
 - 4.5 Hadoop和Kerberos69
 - 4.5.1 Kerberos概述70
 - 4.5.2 Hadoop中的Kerberos71
 - 4.6 小结76
- 第5章 监控Hadoop集群77
 - 5.1 监控策略介绍77
 - 5.2 Hadoop参数78
 - 5.2.1 JMX参数79
 - 5.2.2 使用Nagios监控Hadoop80
 - 5.2.3 监控Hadoop分布式文件系统81
 - 5.2.4 NameNode校验81
 - 5.2.5 JournalNode检查83
 - 5.2.6 ZooKeeper检查83
 - 5.3 监控MapReduce84
 - 5.4 使用Ganglia监控Hadoop85
 - 5.5 小结86
- 第6章 在云端使用Hadoop87
 - 6.1 Amazon Elastic MapReduce87
 - 6.1.1 安装EMR命令行接口88
 - 6.1.2 选择Hadoop版本89
 - 6.1.3 启动EMR集群89
 - 6.2 使用Whirr93
 - 6.3 小结94
- 第7章 Hadoop平台安全概述95
 - 7.1 为什么需要保障Hadoop生态系统的安全96
 - 7.2 确保Hadoop生态系统安全面临的挑战96
 - 7.3 关键安全因素97
 - 7.4 小结99
- 第8章 Hadoop安全体系设计100
 - 8.1 什么是Kerberos100
 - 8.1.1 Kerberos关键术语101
 - 8.1.2 Kerberos如何工作102
 - 8.1.3 Kerberos的优点103
 - 8.2 不采用Kerberos的Hadoop默认安全模型103
 - 8.3 Hadoop Kerberos安全模型实现105
 - 8.3.1 用户层次的访问控制105
 - 8.3.2 服务层次的访问控制105
 - 8.3.3 用户和服务认证106
 - 8.3.4 授权令牌106

- 8.3.5 作业令牌106
- 8.3.6 数据块访问令牌107
- 8.4 小结108
- 第9章 配置一个安全Hadoop集群109
 - 9.1 前提条件109
 - 9.2 设置Kerberos110
 - 9.3 配置Hadoop使用Kerberos认证117
 - 9.3.1 在所有Hadoop节点设置Kerberos客户端117
 - 9.3.2 配置Hadoop服务标识118
 - 9.4 Hadoop用户设置124
 - 9.5 安全Hadoop自动部署124
 - 9.6 小结125
- 第10章 Hadoop生态系统安全保障126
 - 10.1 为Hadoop生态系统组件配置Kerberos127
 - 10.1.1 Hive安全设置127
 - 10.1.2 Oozie安全设置130
 - 10.1.3 Flume安全设置131
 - 10.1.4 HBase安全设置134
 - 10.1.5 Sqoop安全设置137
 - 10.1.6 Pig安全设置138
 - 10.2 Hadoop生态系统组件安全保障最佳实践138
 - 10.3 小结139
- 第11章 集成Hadoop与企业安全系统140
 - 11.1 集成EIM系统141
 - 11.1.1 配置EIM与Hadoop集成142
 - 11.1.2 集成基于Active Directory的EIM系统与Hadoop生态系统143
 - 11.2 从企业网络访问安全Hadoop集群144
 - 11.2.1 HttpFS145
 - 11.2.2 HUE145
 - 11.2.3 Knox Gateway Server146
 - 11.3 小结147
- 第12章 Hadoop中敏感数据安全保护148
 - 12.1 Hadoop中敏感数据及保护方法148
 - 12.2 小结154
- 第13章 安全事件与审计日志155
 - 13.1 Hadoop集群安全事故和事件监控155
 - 13.2 Hadoop集群审计日志设置158
 - 13.3 小结160
- 附录 Hadoop安全机制解决方案161

《Hadoop集群与安全》

精彩短评

- 1、依赖于linux系统的用户管理与权限管理，hadoop的安全挑战非同一般。kerberos的集成可以解决身份认证权限管理的大部分问题，不过如果缺少统一配置管理的支撑，繁重的工作量和配置管理的复杂性会引入不少新的安全麻烦，rhino等项目的发展带来一丝曙光，在数据安全、审计方面颇有建树。不过生产系统整体安全还是任重道远。
- 2、两本书的合集，部署那本是Hadoop1的，安全那本涉及2，看看了解个大概，用的时候直接看具体官方手册了

《Hadoop集群与安全》

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:www.tushu111.com