

《Hadoop YARN权威指南》

图书基本信息

书名：《Hadoop YARN权威指南》

13位ISBN编号：9787111491815

出版时间：2015-4-13

作者：Arun C. Murthy, Vinod Kumar Vavilapalli, Doug Eadline, Joseph Niemiec, Jeff Markham

页数：242

译者：罗韩梅, 洪志国, 杨旭

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.tushu000.com

《Hadoop YARN权威指南》

内容概要

《Hadoop YARN权威指南》由YARN的创建和开发团队亲笔撰写，Altiscale的CEO作序鼎力推荐，是使用Hadoop YARN建立分布式、大数据应用的权威指南。书中利用多个实例，详细介绍Hadoop YARN的安装和管理，以帮助用户使用YARN进行应用开发，并在YARN上运行除了MapReduce之外的新框架。

《Hadoop YARN权威指南》共12章，第1章讲述Apache Hadoop YARN产生和发展的历史；第2章讲解在单台机器（工作站、服务器或笔记本电脑）上快速安装Hadoop 2.0；第3章介绍Apache Hadoop YARN资源管理器；第4章简要介绍YARN组件的功能，帮助读者开始深入了解YARN；第5章详细讲解YARN的安装方法，包括一个基于脚本的手动安装，以及使用Apache Ambari基于GUI的安装；第6章讲述对YARN集群的管理，涉及一些基本的YARN管理场景，介绍如何利用Nagios和Ganglia监控集群，论述对JVM的监视，并介绍Ambari的管理界面；第7章深入探究YARN的架构，向读者展示YARN的内部工作原因；第8章深入讨论Capacity调度器；第9章描述基于现有MapReduce的应用程序如何继续工作以及利用YARN的优势；第10章通过创建一个JBoss Application Server集群的过程，讲述如何构建一个YARN应用程序；第11章描述建立在YARN上的典型示例程序distributed shell的使用和内部情况；第12章总结运行在YARN上的新兴开源框架。最后提供6个附录，包括补充内容和代码下载、YARN的安装脚本、YARN管理脚本、Nagios模块、资源及其他信息、HDFS快速参考。

《Hadoop YARN权威指南》

作者简介

Arun C. Murthy，自从Apache Hadoop启动以来就是一个全职的贡献者，并创立了Apache YARN项目。他作为雅虎Hadoop MapReduce开发团队的架构师和领导者，负责为整个雅虎公司提供MapReduce技术服务。他是Hortonworks公司的创始人和架构师，Hortonworks公司由雅虎Hadoop团队的核心成员组成，并加速了Hadoop的发展和普及。

Vinod Kumar Vavilapalli 是Hortonworks公司的首席开发者和Apache Hadoop YARN的项目负责人。他之前参与的项目有Hadoop On Demand、Hadoop-0.20、Capacity调度器、Hadoop安全性和MapReduce。Doug Eadline博士作为Linux的HPC集群革命的实践者和见证者开始了他的职业生涯，目前在为大数据分析撰写文档。

Joseph Niemiec是Hortonworks大数据解决方案工程师，致力于为许多财富1000强公司设计Hadoop解决方案。

Jeff Markham是Hortonworks解决方案工程师。此前，他为VMware、Red Hat和IBM开发过分布式数据应用。

书籍目录

- 译者序
- 推荐序一
- 推荐序二
- 前言
- 第1章 Apache Hadoop YARN：简明历史及基本原理 1
 - 1.1 引言 1
 - 1.2 Apache Hadoop 2
 - 1.3 阶段0：Ad Hoc集群时期 3
 - 1.4 阶段1：Hadoop on Demand 3
 - 1.4.1 HOD世界中的HDFS 5
 - 1.4.2 HOD的特色及优势 5
 - 1.4.3 HOD的缺点 6
 - 1.5 阶段2：共享计算集群的黎明 8
 - 1.5.1 共享集群的演进 8
 - 1.5.2 使用共享MapReduce集群的问题 13
 - 1.6 阶段3：YARN的出现 15
 - 1.7 小结 16
- 第2章 Apache Hadoop YARN安装快速入门 17
 - 2.1 准备开始 18
 - 2.2 配置单节点YARN集群的步骤 18
 - 2.2.1 第1步：下载Apache Hadoop 18
 - 2.2.2 第2步：设置JAVA_HOME 19
 - 2.2.3 第3步：创建用户和用户组 19
 - 2.2.4 第4步：创建数据和日志目录 19
 - 2.2.5 第5步：配置core-site.xml 19
 - 2.2.6 第6步：配置hdfs-site.xml 20
 - 2.2.7 第7步：配置mapred-site.xml 21
 - 2.2.8 第8步：配置yarn-site.xml 21
 - 2.2.9 第9步：调整Java堆大小 21
 - 2.2.10 第10步：格式化HDFS 22
 - 2.2.11 第11步：启动HDFS服务 22
 - 2.2.12 第12步：启动YARN服务 23
 - 2.2.13 第13步：通过Web接口验证正在运行的服务 24
 - 2.3 运行MapReduce示例程序 25
 - 2.4 小结 26
- 第3章 Apache Hadoop YARN的核心概念 27
 - 3.1 不只是MapReduce 27
 - 3.2 Apache Hadoop MapReduce 29
 - 3.2.1 支持非MapReduce应用的需求 30
 - 3.2.2 解决可扩展性 30
 - 3.2.3 提高资源使用率 30
 - 3.2.4 用户敏捷性 30
 - 3.3 Apache Hadoop YARN 31
 - 3.4 YARN组件 32
 - 3.4.1 ResourceManager 32
 - 3.4.2 ApplicationMaster 32
 - 3.4.3 资源模型 33

3.4.4 ResourceRequest和Container 33

3.4.5 Container规范 34

3.5 小结 34

第4章 YARN组件的功能概述 35

4.1 体系架构概述 35

4.2 ResourceManager 37

4.3 YARN调度组件 38

4.3.1 FIFO调度器 38

4.3.2 Capacity调度器 38

4.3.3 Fair调度器 39

4.4 Container 40

4.5 NodeManager 40

4.6 ApplicationMaster 41

4.7 YARN资源模型 41

4.7.1 客户端资源请求 42

4.7.2 ApplicationMaster Container的分配 42

4.7.3 ApplicationMaster与Container管理器的通信 44

4.8 管理应用程序的依赖文件 44

4.8.1 LocalResource的定义 44

4.8.2 LocalResource时间戳 45

4.8.3 LocalResource类型 46

4.8.4 LocalResource的可见性 46

4.8.5 LocalResource的生命周期 47

4.9 小结 47

第5章 安装Apache Hadoop YARN 49

5.1 基础知识 49

5.2 系统准备 50

5.2.1 第1步：安装EPEL和pdsh 50

5.2.2 第2步：生成和分发ssh密钥 51

5.3 基于脚本安装Hadoop 2 51

5.3.1 JDK选项 52

5.3.2 第1步：下载并解压脚本 52

5.3.3 第2步：设置脚本里的变量 52

5.3.4 第3步：提供节点名字 53

5.3.5 第4步：运行脚本 54

5.3.6 第5步：验证安装 54

5.4 基于脚本的卸载 57

5.5 配置文件处理 57

5.6 配置文件设置 57

5.6.1 core-site.xml 57

5.6.2 hdfs-site.xml 58

5.6.3 mapred-site.xml 58

5.6.4 yarn-site.xml 59

5.7 启动脚本 59

5.8 用Apache Ambari安装Hadoop 60

5.8.1 基于Ambari安装Hadoop 61

5.8.2 第1步：检查要求 61

5.8.3 第2步：安装Ambari服务器 62

5.8.4 第3步：安装和启动Ambari代理 62

- 5.8.5 第4步：启动Ambari服务器 62
- 5.8.6 第5步：安装HDP2.X集群 63
- 5.9 小结 70
- 第6章 Apache Hadoop YARN的管理 71
 - 6.1 基于脚本的配置 71
 - 6.2 监控集群健康：Nagios 76
 - 6.2.1 监控基本的Hadoop服务 77
 - 6.2.2 监控JVM 80
 - 6.3 实时监控系统：Ganglia 82
 - 6.4 使用Ambari管理 83
 - 6.5 JVM分析 88
 - 6.6 基本的YARN管理 90
 - 6.6.1 YARN的管理工具 91
 - 6.6.2 增加或关闭YARN节点 92
 - 6.6.3 Capacity调度器的配置 92
 - 6.6.4 YARN的Web代理 92
 - 6.6.5 使用JobHistoryServer 93
 - 6.6.6 更新用户到用户组的映射 93
 - 6.6.7 更新超级用户代理群组映射 93
 - 6.6.8 更新ResourceManager管理的ACL 93
 - 6.6.9 重新加载服务级授权策略文件 94
 - 6.6.10 管理YARN作业 94
 - 6.6.11 设置Container的内存 94
 - 6.6.12 设置Container核数 94
 - 6.6.13 设置MapReduce配置项 95
 - 6.6.14 用户日志管理 95
 - 6.7 小结 97
- 第7章 Apache Hadoop YARN的架构指南 98
 - 7.1 概述 98
 - 7.2 ResourceManager 99
 - 7.2.1 ResourceManager组件概述 100
 - 7.2.2 客户端和ResourceManager交互 100
 - 7.2.3 应用程序和ResourceManager的通信 102
 - 7.2.4 节点和ResourceManager的通信 103
 - 7.2.5 ResourceManager核心组件 104
 - 7.2.6 ResourceManager安全相关的组件 105
 - 7.3 NodeManager 109
 - 7.3.1 NodeManager各组件概述 109
 - 7.3.2 NodeManager组件 110
 - 7.3.3 NodeManager安全组件 116
 - 7.3.4 NodeManager的重要功能 116
 - 7.4 ApplicationMaster 117
 - 7.4.1 概述 117
 - 7.4.2 活跃 119
 - 7.4.3 资源需求 119
 - 7.4.4 调度 120
 - 7.4.5 调度协议和本地性 121
 - 7.4.6 启动Container 123
 - 7.4.7 完成的Container 124

- 7.4.8 ApplicationMaster失败和恢复 124
- 7.4.9 协调和输出提交 124
- 7.4.10 为客户端提供信息 125
- 7.4.11 安全 125
- 7.4.12 ApplicationMaster退出时进行清理 125
- 7.5 YARN Container 125
 - 7.5.1 Container运行环境 126
 - 7.5.2 与ApplicationMaster通信 127
- 7.6 应用程序开发者的摘要 127
- 7.7 小结 128
- 第8章 YARN中的Capacity调度器 129
 - 8.1 Capacity调度器介绍 129
 - 8.1.1 多租户弹性 130
 - 8.1.2 安全 130
 - 8.1.3 资源感知 130
 - 8.1.4 细粒度调度 130
 - 8.1.5 本地化 131
 - 8.1.6 调度策略 131
 - 8.2 Capacity调度器配置 131
 - 8.3 队列 132
 - 8.4 层级队列 132
 - 8.4.1 关键特性 132
 - 8.4.2 队列间的调度 132
 - 8.4.3 定义层级队列 133
 - 8.5 队列访问控制 134
 - 8.6 层级队列Capacity管理 135
 - 8.7 用户级别限制 137
 - 8.8 预订 139
 - 8.9 队列的状态 140
 - 8.10 应用程序的限制 141
 - 8.11 用户接口 141
 - 8.12 小结 142
- 第9章 Apache Hadoop YARN下的MapReduce 143
 - 9.1 运行Hadoop YARN MapReduce实例 143
 - 9.1.1 可利用的实例列表 143
 - 9.1.2 运行Pi实例 144
 - 9.1.3 使用Web GUI监控实例 146
 - 9.1.4 运行terasort测试 151
 - 9.1.5 运行TestDFSIO基准测试 151
 - 9.2 MapReduce兼容性 152
 - 9.3 MapReduce ApplicationMaster 153
 - 9.3.1 启用ApplicationMaster的重启 153
 - 9.3.2 启用已完成任务的恢复 153
 - 9.3.3 JobHistory服务 153
 - 9.4 计算一个节点的容量 154
 - 9.5 Shuffle服务的变动 155
 - 9.6 运行已有的第1版Hadoop的应用程序 155
 - 9.6.1 org.apache.hadoop.mapred API的二进制兼容性 155
 - 9.6.2 org.apache.hadoop.mapreduce API的源码兼容性 155

- 9.6.3 命令行脚本的兼容性 156
- 9.6.4 MRv1和早期MRv2 (0.23.x) 应用程序兼容性的权衡 156
- 9.7 运行第1版MapReduce现有的代码 157
 - 9.7.1 在YARN上运行Apache Pig脚本 157
 - 9.7.2 在YARN上运行Apache Hive查询 157
 - 9.7.3 在YARN上运行Apache Oozie工作流 157
- 9.8 高级特性 158
 - 9.8.1 Uber作业 158
 - 9.8.2 可插拔的Shuffle和Sort 158
- 9.9 小结 159
- 第10章 Apache Hadoop YARN应用程序范例 160
 - 10.1 YARN客户端 161
 - 10.2 ApplicationMaster 175
 - 10.3 小结 192
- 第11章 使用Apache Hadoop YARN Distributed-Shell 193
 - 11.1 使用YARN Distributed-Shell 193
 - 11.1.1 简单例子 194
 - 11.1.2 使用更多Container 195
 - 11.1.3 带有shell命令参数的Distributed-Shell 195
 - 11.2 Distributed-Shell内部实现 197
 - 11.2.1 应用的常量定义 198
 - 11.2.2 Client 198
 - 11.2.3 ApplicationMaster 201
 - 11.2.4 普通Container 205
 - 11.3 小结 205
- 第12章 Apache Hadoop YARN框架 206
 - 12.1 Distributed-Shell 206
 - 12.2 Hadoop MapReduce 206
 - 12.3 Apache Tez 207
 - 12.4 Apache Giraph 207
 - 12.5 Hoya : HBase on YARN 208
 - 12.6 Dryad on YARN 208
 - 12.7 Apache Spark 208
 - 12.8 Apache Storm 209
 - 12.9 REEF : Retainable Evaluator Execution Framework 209
 - 12.10 Hamster : Hadoop and MPI on the Same Cluster 210
 - 12.11 小结 210
- 附录A 补充内容和代码下载 211
- 附录B YARN的安装脚本 212
- 附录C YARN的管理脚本 224
- 附录D Nagios模块 229
- 附录E 资源及附加资料 235
- 附录F HDFS快速参考 237

《Hadoop YARN权威指南》

精彩短评

- 1、 just so so
- 2、 几天前小组长才买完hadoop1权威指南，为什么yarn权威指南没有人看呢？其实yarn才是大数据框架的未来，本书第四章和第七章介绍架构部分是精华，其他地方可以略过。本书还是很值得一读。
- 3、 对hadoop 2.0 yarn的架构有深入的讲解。
- 4、 翻译减一分

《Hadoop YARN权威指南》

精彩书评

1、1、hadoop1.0，只有mapreduce和hdfs，这个架构设计不利于扩展。yarn横空出世，专门负责资源调度
2、yarn由三部分组成，resourcemanager，nodemanager，和applicationmaster。rs负责资源仲裁，nm负责监控资源使用，app负责资源申请，属于container03、hadoop的jar包自带了一些基准用例。

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.tushu000.com