

英特尔® Apache Hadoop* 软件发行版安装手册

版本 2.4

2013 年 5 月

文档编号: 327778-009CN



免责声明和法律信息

本文件中包含关于英特尔产品的信息。本文件不构成对任何知识产权的授权，包括明示的、暗示的，也无无论是基于禁止反言的原则或其他。除英特尔产品销售的条款和条件规定的责任外，英特尔不承担任何其他责任。英特尔在此作出免责声明：本文件不构成英特尔关于其产品的使用和 / 或销售的任何明示或暗示的保证，包括不就其产品的 (i) 对某一特定用途的适用性、(ii) 适销性以及 (iii) 对任何专利、版权或其他知识产权的侵害的承担任何责任或作出任何担保。

除非经过英特尔的书面同意认可，英特尔的产品无意被设计用于或被用于以下应用：即在这样的应用中可因英特尔产品的故障而导致人身伤亡。

英特尔有权随时更改产品的规格和描述而无需发出通知。设计者不应信赖任何英特尔产品所不具有的特性，设计者亦不应信赖任何标有“保留权利”或“未定义”说明或特性描述。对此，英特尔保留将来对其进行定义的权利，同时，英特尔不应为其日后更改该等说明或特性描述而产生的冲突和不相容承担任何责任。此处提供的信息可随时改变而无需通知。请勿根据本文件提供的信息完成一项产品设计。

本文件所描述的产品可能包含使其与宣称的规格不符的设计缺陷或失误。这些缺陷或失误已收录于勘误表中，可索取获得。

在发出订单之前，请联系当地的英特尔营业部或分销商以获取最新的产品规格。

索取本文件中或英特尔的其他材料中提的、包含订单号的文件的复印件，可拨打 1-800-548-4725，或登陆[英特尔网站](#)。

英特尔处理器标号不是性能的指标。http://www.intel.com/products/processor_number 处理器标号仅用于区分同属一个系列的处理器的特性，而不能用于区分不同系列的处理器。详情敬请登陆：http://www.intel.com/products/processor_number

Performance tests and ratings are measured using specific computer systems and/or components and reflect the approximate performance of Intel products as measured by those tests. Any difference in system hardware or software design or configuration may affect actual performance. Buyers should consult other sources of information to evaluate the performance of systems or components they are considering purchasing. For more information on performance tests and on the performance of Intel products, visit [Intel Performance Benchmark Limitations](#).

结果基于模拟测算得出，仅作参考之用。结果通过系统模拟器或模型测算得出。任何系统硬件、软件的设计或配置的不同均可能影响实际性能。

英特尔，英特尔® Apache Hadoop* 软件发行版，英特尔® 发行版，Intel® Manager for Apache Hadoop* software，和 Intel® Manager 是英特尔在美国和 / 或其他国家的商标。

* 其他的名称和品牌可能是其他所有者的资产。

英特尔公司 2013 版权所有。所有权保留。



文档修订记录

日期	修订	描述
2012 年 6 月	001	英特尔® Apache Hadoop* 软件发行版 v2.0 第一版
2012 年 9 月	002	英特尔® Apache Hadoop* 软件发行版 v2.1 文档更新
2012 年 11 月	003	英特尔® Apache Hadoop* 软件发行版 v2.1.1 文档更新
2012 年 12 月	004	英特尔® Apache Hadoop* 软件发行版 v2.2 文档更新
2013 年 2 月	005	英特尔® Apache Hadoop* 软件发行版 v2.3 文档更新
2013 年 2 月	006	英特尔® Apache Hadoop* 软件发行版 v2.3 文档更新
2013 年 2 月	007	英特尔® Apache Hadoop* 软件发行版 v2.3 文档更新
2013 年 2 月	008	英特尔® Apache Hadoop* 软件发行版 v2.3 文档更新
2013 年 5 月	009	英特尔® Apache Hadoop* 软件发行版 v2.4 文档更新



目录

1.0	简介	6
1.1	什么是管理节点?	6
2.0	设计 Hadoop 集群的要求和推荐	7
2.1	支持的操作系统	7
2.2	硬件要求	7
2.3	软件要求	7
2.4	内存要求	8
2.5	磁盘分区要求	9
2.6	集群和网络拓扑要求	9
2.7	端口和防火墙要求	10
2.7.1	防火墙总体要求	10
2.7.2	管理节点的端口要求	11
2.7.3	Apache Hadoop* 服务的端口要求	12
2.8	推荐配置	14
3.0	设置对英特尔® 发行版软件库的访问	15
3.1	先决条件	15
3.1.1	安装前准备	15
3.1.2	在 RHEL 上创建英特尔® 发行版软件库	15
3.1.3	在 SLES 上创建英特尔® 发行版软件库	17
4.0	安装 Intel® Manager for Apache Hadoop* software	19
4.1	先决条件	19
4.2	安装 Intel® Manager	19
4.2.1	执行 Intel® Manager 快速安装	20
4.2.2	执行 Intel® Manager 交互式安装	20
4.3	第一次登录 Intel® Manager	22
5.0	了解 Hadoop 服务的角色	23
5.1	了解 HDFS 角色	23
5.2	了解 MapReduce 角色	23
5.3	了解 HBase 角色	24
5.4	了解各种服务角色	24
6.0	创建 Apache Hadoop* 集群	25
6.1	创建 Apache Hadoop* 集群的要求	25
6.2	使用 Intel® Manager 向导来创建一个集群	25
6.3	在 Apache Hadoop* 集群中分配角色给每个节点	32
6.4	指定存储 HDFS 数据的路径	34
6.5	指定存储 MapReduce 数据的路径	35
7.0	升级英特尔® Apache Hadoop* 软件发行版	37
7.1	先决条件	37
7.2	在 Red Hat Enterprise Linux 上升级英特尔® Apache Hadoop* 软件发行版	37
7.3	在 SUSE Enterprise Linux 上升级英特尔® Apache Hadoop* 软件发行版	38
7.4	升级 Apache Hadoop* 组件	39
8.0	卸载英特尔® Apache Hadoop* 软件发行版	42
8.1	先决条件	42
8.2	卸载英特尔® Apache Hadoop* 软件发行版	42
8.3	卸载示例	43
9.0	管理英特尔® Apache Hadoop* 软件发行版许可证	45
9.1	查看许可证	45



9.2	升级许可证.....	46
A:	支持的 Apache Hadoop* 组件列表.....	48



1.0 简介

Apache Hadoop* 是一个开源软件框架，用于在大型集群中运行海量数据的、分布式的应用程序。Intel® Manager for Apache Hadoop* software 是一个 Hadoop 的中央管理控制平台，它能处理集群的安装设置、Hadoop 服务的配置变更、集群监控、事件和警报的发送、资源优化、以及安全访问。

本文档介绍如何安装 Intel® Manager 并使用 Intel® Manager 来创建 Apache Hadoop* 的 Hadoop 集群。

1.1 什么是管理节点？

管理节点是安装了 Intel® Manager for Apache Hadoop* software 的节点，同时也是 Intel® Manager 运行的网络服务器。



2.0 设计 Hadoop 集群的要求和推荐

在创建 Hadoop 集群前，你必须阅读并理解集群创建和管理的要求和推荐。这包括对集群中的每个节点的网络连接及软硬件配置作出决定。

2.1 支持的操作系统

所有集群中的节点，包括 Intel® Manager，必须运行同一操作系统。英特尔® Apache Hadoop* 软件发行版支持以下操作系统：

操作系统	版本
Red Hat Enterprise Linux*	<ul style="list-style-type: none">• 6.1• 6.2• 6.3• 6.4
CentOS*	<ul style="list-style-type: none">• 6.1• 6.2• 6.3
Oracle Enterprise Linux* with Unbreakable Enterprise Kernel	<ul style="list-style-type: none">• 6.1• 6.2• 6.3
SUSE Linux Enterprise*	<ul style="list-style-type: none">• 11 SP1• 11 SP2

只支持 64 位架构。

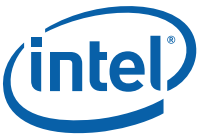
2.2 硬件要求

集群中的每个节点必须符合以下硬件要求：

- 每个节点必须配备至少 1GB 的网卡。
- 每个节点必须至少有 2 个支持超线程技术的四核芯片（8 个核，2 个接口）。
- 除了 JobTracker 和 Primary NameNode，其他所有节点都不能安装在磁盘冗余阵列上。这表示 DataNode 使用的逻辑盘分区不能被用于阵列。
- 确认每个节点至少有二个逻辑盘分区。一个分区是用于安装程序的系统分区。另一个分区则用于存储 HDFS 数据。你可以在一个或多个物理或逻辑分区上存储 HDFS 数据。然而，存储 HDFS 数据的逻辑盘分区只能包含 HDFS 数据，不能包含其他数据。

2.3 软件要求

在安装 Intel® Manager 或创建 Hadoop 集群之前，确认所有被加入到集群的节点都符合以下软件要求。



- 在每个节点上，OpenSSH* 5.3 或更高版本必须已安装并正常运行。sshd daemon 服务器的侦听端口必须为 22。
- Java Runtime Environment* 版本 1.6.0_31 必须已安装。
- sudo 必须已安装并正常运行，这样 root 用户可以通过 su 命令切换到一个或多个无窗口（faceless）帐户。
- yum 或 zypper 必须已安装。
- 确认集群的任何节点都没有安装 MySQL。
- 集群中的所有节点必须可访问用于安装 RPM 包的某个操作系统软件库（repository）。集群中的所有节点必须使用相同的操作系统软件库。
在英特尔® Apache Hadoop* 软件发行版的安装配置过程中，脚本必须能从此软件库安装 RPM 包。对于 OEL 和 RHEL 操作系统，这即为 yum 软件库。对于 SLES 操作系统，这即为 zypper 软件库。

2.4 内存要求

每个节点必须至少有 16GB 的 RAM。根据节点所安装的 Hadoop 服务，节点可能需要超过 16GB 的 RAM。下表列出在节点上运行不同 Hadoop 服务时，该节点所需的额外内存。

表 1. Hadoop 服务的内存要求

服务	内存要求
管理服务器	8GB
MapReduce JobTracker	2GB
MapReduce TaskTracker	2GB
TaskTracker 上的 MapReduce Slots 总数（包括 map 和 reduce 插槽）	512MB * 插槽数
HDFS NameNode	16GB
HDFS Secondary NameNode	16GB
HDFS DataNode	2GB
ZooKeeper	4GB
HBase Master Server	2GB
HBase Region Server	16GB
Hive Server	2GB
Clients	8GB

按以下步骤计算节点需要的内存：

1. 确认所有会在节点上运行的 Hadoop 服务。
2. 确认每个服务要求的内存容量。
3. 将所有内存要求相加。
4. 如果相加后内存要求小于 16GB，则最低内存要求为 16GB。如果相加后内存要求大于 16GB，则最低内存要求为相加后的和。

比如，如果节点上运行以下服务：

- HDFS DataNode (2GB)
- MapReduce TaskTracker (2GB)
- HBase Region Server (16GB)



- MapReduce Slots 在 TaskTracker 上的总数是 16 (512x16=8GB)

那么，节点的内存容量要求如下：

$$2\text{GB} + 2\text{GB} + 16\text{GB} + 8\text{GB} = 28\text{GB}$$

2.5 磁盘分区要求

对 Hadoop 集群中的每个节点，磁盘必须按以下方法分区：

- 以下逻辑分区必须已存在：
 - 系统或根分区，挂载在 / 下。
 - 数据分区，用于存储所有 HDFS 和 MapReduce 数据。可以有多个数据分区。
- 除了 HDFS 和 MapReduce 数据，其他数据不能存储在数据分区。
- 如果磁盘分区之前被 Apache Hadoop* 集群使用过，则在其他集群使用该磁盘前，磁盘必须被格式化。
- 每个逻辑分区应使用 *ext4* 文件系统。
- 用于存放操作系统的根分区，必须至少有 30GB。
- 如果 */var/log* 不属于根分区，则必须至少分配 10GB 给 */var/log* 所属的磁盘分区。
- 每个物理磁盘必须为不同的挂载点。
- 你不可将物理磁盘分成多个挂载点。
- 对每个数据分区，逻辑磁盘分区应有一个符合以下格式的挂载名称：*/mnt/disk[number]*。比如，如果你有二个逻辑磁盘分区，分区的挂载名称应为：*/mnt/disk1* 和 */mnt/disk2*。
- HDFS DataNode 的目录不能放在系统分区上，只能放在数据分区上。

2.6 集群和网络拓扑要求

- 决定集群中的节点数目。
- 决定集群中的机柜数目及每个机柜的名称。
- 决定每个机柜中的节点数目。
- 决定每个节点所在的子网。
- 决定每个节点的主机名和 IP 地址。
- Primary NameNode 会被自动配置为 Hadoop 集群的 NTP 服务器，它将同步集群中所有机器的时间。在 NameNode 加入到 Hadoop 集群之前，你必须确保 Primary NameNode 的时钟和正确的 NTP 服务器同步。
- 决定节点是否可通过 DNS 服务器访问。如果使用 DNS 服务器，服务器必须在 Hadoop 集群创建前，分配主机名给集群中所有机器。
- 在集群的网络环境中，不能存在只有部分机器可通过 DNS 访问的情况。
- 决定哪个机器是管理节点。
- 决定哪个机器是 JobTracker 和 PrimaryNameNode。
- 决定哪些机器是客户端，哪些机器运行 Hadoop 服务，或二者兼是。
- 一旦主机名分配给 Primary NameNode，主机名需永不再更改。
- 确定你知道要加入到 Hadoop 集群中的每个节点的 root 密码。
- 确保 Intel® Manager 所在机器成为 Hadoop 集群中的节点。这一节点称为管理节点。
- 管理节点必须和集群中的其他节点属于同一子网。
- 如果节点上已安装 Red Hat、Oracle Enterprise Linux 或 CentOS 操作系统，则在每个节点上确认以下文件都已指定同样的值。



- /proc/sys/kernel/hostname
- /etc/sysconfig/network
- 如果节点上安装了 SUSE Linux Enterprise Server 操作系统，则在每个节点上确认以下文件都已指定同样的值。
 - /proc/sys/kernel/hostname
 - /etc/HOSTNAME
- 决定你会在集群中使用下列哪些服务：
 - HDFS
 - MapReduce
 - HBase
 - Hive
 - Sqoop
 - Pig
 - Flume

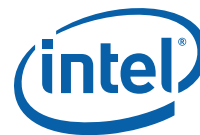
2.7 端口和防火墙要求

以下章节列出了当 Apache Hadoop* 集群中的机器启用了防火墙后必须符合的要求：

2.7.1 防火墙总体要求

Apache Hadoop* 中的每个机器都必须符合以下要求：

- 机器的防火墙必须没有阻挡环回口（loopback network interface）的通信。
- 机器的防火墙必须没有阻挡 ICMP type 0 的通信。
- 取决于服务类型以及它在机群中是如何配置的，你可能需要打开 TCP 或 UDP 端口，或者二者都打开。
- 如果对外的 HTTP 通信必须经过代理服务器，则你必须配置机器使之能使用 HTTP 代理服务器。
- 如果高可用性被启用，且高可用性采用多播模式，则机器的防火墙不能阻挡多播通信。
- 必须允许集群中所有节点之间的 SSH 和 HTTP 通信。
- 如果 Apache Hadoop* 的 RPM 包是从远程 HTTP 软件库中安装，则防火墙不能阻挡对以下 URL 的 HTTP 通信：
 - http://repo1.intelhadoop.com:3424
 - http://repo2.intelhadoop.com:80
 - http://repo3.intelhadoop.com:80
- 在管理节点上，防火墙不能阻挡对以下 URL 的 HTTP 通信：
 - https://registration.intelhadoop.com
 - https://registration2.intelhadoop.com



2.7.2 管理节点的端口要求

管理节点对端口有一些特定的要求，而其他节点则没有。表 2. 列出了这些要求。表中的端口号码必须为可用状态，相关服务可使用这些端口侦听，而且端口不能被机器的防火墙阻止。

注释： 除非特别注明，所有端口为 TCP 端口。

表 2. 管理节点的端口要求

服务名称	端口
intel-manager	9443
	50050
sshd	22
gmond	8649 (TCP)
	8649 (UDP)
gmetad	8651
	8652
httpd	80
	443
puppetmasterd	18140
	18141
	18142
	18143
corosync	5404 (UDP 或 TCP)
注释： 服务仅运行于高可用性配置中。 注释： 如果采用组播通信，则端口为 UDP。如果采用单播通信，则端口为 TCP。	5405 (UDP 或 TCP)
smart-tuner	10503
nginx	8140
DNS	53 (UDP)
ntpd	123 (UDP)

2.7.3 Apache Hadoop* 服务的端口要求

在一个 Hadoop 集群中，各种服务可能会在许多不同的节点上运行。要正常工作，这些服务必须能侦听一个或多个端口。某些服务只能在管理节点上运行，其他服务则只能当该节点被分配正确角色时运行。关于服务角色的更多信息，参见 5.0 了解 Hadoop 服务的角色。

根据你要分配给 Hadoop 集群中节点的服务角色，确认每个节点的所需端口为可用状态。如果管理节点被分配了一个或多个服务角色，该管理节点不仅需要符合这些服务的端口要求，而且需要符合管理节点的端口要求。下表列出了当你添加节点到 Hadoop 集群时你必须考虑的端口要求。表中的端口号码必须为可用状态，相关服务可使用这些端口侦听，而且端口不能被机器的防火墙阻止。

注释：

除非特别注明，所有端口为 TCP 端口。

表 3. Hadoop 集群的端口要求

服务名称	端口	角色
所有		
gmond	8649 (UDP)	没有分配角色。服务运行在所有节点上。
	8649 (UDP)	
DNS	53 (UDP)	没有分配角色。服务运行在所有节点上。
sshd	22	没有分配角色。服务运行在所有节点上。
ntpd	123 (UDP)	没有分配角色。服务运行在所有节点上。
httpd	80	没有分配角色。服务运行在所有节点上。
	443	
HDFS		
hadoop-namenode	50070	Primary NameNode
	8020	
	50470	
hadoop-datanode	50075	DataNode
	50010	
	50020	
	50475	
hadoop-secondarynamenode	50090	Secondary NameNode
ftptoverhdfs	2222	在安装和运行服务的节点上，相应的服务端口必须已打开。
	2220	
	990 (SSL)	
	889 (SSL)	

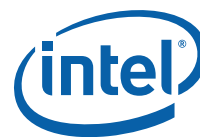
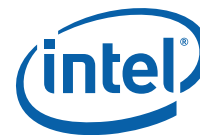


表 3. Hadoop 集群的端口要求

服务名称	端口	角色
MapReduce		
hadoop-jobtracker	50030	JobTracker
	10503	
	54311	
	9290	
hadoop-tasktracker	50060	TaskTracker
HBase		
zookeeper-server	3888	ZooKeeper
	2181	
	2888	
hbase-master	60010	HMaster
	60000	
	10101	
hbase-regionserver	60030	RegionServer
	60020	
	10102	
hbase-thrift	8080	HBase Thrift
Hive		
mysqld	3306	Primary NameNode
hive-metastore	9083	Primary NameNode
hive-server	10000	Hive Server
Oozie		
oozie	11000	Oozie server
High Availability		
DRBD	7789	<ul style="list-style-type: none"> Primary NameNode, Standby NameNode
corosync 注释: 服务仅运行于高可用性配置中。 注释: 如果采用组播通信, 则端口为 UDP。如果采用单播通信, 则端口为 TCP。	5404 (TCP 或 UDP)	<ul style="list-style-type: none"> Primary NameNode Standby NameNode JobTracker Backup JobTracker
	5405 (TCP 或 UDP)	



2.8 推荐配置

以下推荐配置可帮助确保 Hadoop 集群的性能优化和可管理性。

- 节点的主机名解析必须通过 DNS 执行。
- 为避免灾难性的故障，JobTracker 和 Primary NameNode 必须安装在磁盘阵列上。阵列必须为 1 或 5。
- 要同时添加一组节点到集群中，你需要为该组中的每个节点设置同样的 root 密码。这可明显加快在集群中添加节点的进程。
- 要减少网络延迟，集群中的所有节点都必须属于同一子网。
- 如果节点没有使用 10GB 的网卡，则可使用网卡绑定以便将多个网卡组合在一起以提升网络流量。绑定的网卡必须使用工作模式 6。
- 每个节点的推荐最小磁盘空间是二个 1T 的物理硬盘。第一个磁盘必须划分为二个逻辑分区。一个分区必须是系统分区，并至少有 300GB 的磁盘空间。
- 每个节点应该有至少 6T 的可用磁盘空间用于 HDFS。
- 集群中应至少有 5 个 DataNode。
- 如果可能，避免将物理磁盘分为多个逻辑分区。除了系统分区外，每个物理磁盘应当仅有一个分区，且该分区包含整个物理磁盘。
- 每个节点应配置一块 10GB 的网卡，用于节点间的通信和执行集群中需要网络连接的任务。
- 仅使用物理机器，不要使用虚拟机器。虚拟机可能会明显导致 HDFS I/O 的缓慢。
- 强烈推荐使用 DNS 服务器来分配主机名给集群中的所有节点。
- Hadoop 节点所在的单个或多个子网不允许有其他机器。
- 如果集群中的所有节点在同一子网中，则安装了 Intel® Manager 的管理节点应当也在同一子网中。
- 集群中不能同时有物理机器和虚拟机器。
- 要确保集群中的机器不成为性能和 I/O 的瓶颈，所有机器必须有相似的硬件和软件配置，包括 RAM、CPU 和磁盘空间。
- 每个节点应至少有 32GB 的内存。
- 所有节点上的热交换必须关闭。



3.0 设置对英特尔® 发行版软件库的访问

要在集群中安装英特尔® Apache Hadoop® 软件发行版，要加入集群的机器必须能够访问 Apache Hadoop® RPM 包。这些文件通过 tarball 提供。

以下章节解释了如何在 Linux 机器上设置对英特尔® 发行版软件库的访问。

3.1 先决条件

要在机器上设置对英特尔® 发行版软件库的访问，你必须了解或操作以下事项：

- 如果机器上安装了 SLES 操作系统，则 zypper 必须已安装且正常工作。
- 如果机器上安装了 RHEL 或 CentOS 操作系统，则 yum 必须已安装且正常工作。
- 你必须对管理节点有 root 访问权限。
- 你必须已在机器上安装 Apache HTTP 服务器或 Apache FTP 服务器。
- 机器上必须已安装 *createrepo*。
- 从客户支持或销售人员处获得包含所有英特尔® 发行版 RPM 的 tarball。
- 在每个节点上，英特尔® 发行版 repo 文件的 ID 必须是 *idh*

3.1.1 安装前准备

在安装英特尔® 发行版之前，你必须已创建以下软件库：

- 操作系统软件库：若未创建，你可运行英特尔® 发行版的 *install-tool* 进行创建。运行以下命令：

```
cd tools-createrepo
sh install-tool
```
- 英特尔® 发行版软件库：参见以下章节。

3.1.2 在 RHEL 上创建英特尔® 发行版软件库

要从 tarball 创建英特尔® 发行版软件库，在你希望创建软件库的机器上执行以下步骤。

1. 将 tarball 解压到 */tmp* 目录：

```
tar zxvf <tarball name>
```
2. 进入到 */tmp/intelhadoop* 目录：

```
cd intelhadoop
```
3. 在要存放英特尔® 发行版 RPM 的 HTTP（或 FTP）服务器上创建目录：
如果是 HTTP 服务器，输入命令：

```
mkdir -p /var/www/html/idh
```


如果是 FTP 服务器，输入命令：

```
mkdir -p /var/ftp/pub/idh
```
4. 执行以下命令，将下列英特尔® 发行版 RPM 复制到新创建的目录。
如果是 HTTP 服务器，输入命令：



```
cp idh /var/www/html/idh
cp manager /var/www/html/idh
cp os_related /var/www/html/idh
如果是 FTP 服务器，输入命令：
cp idh /var/ftp/pub/idh
cp manager /var/ftp/pub/idh
cp os_related /var/ftp/pub/idh
```

5. 执行以下命令，在新建目录下创建 repo 数据：

如果是 HTTP 服务器，输入命令：

```
cd /var/www/html/idh
createrepo .
```

如果是 FTP 服务器，输入命令：

```
cd /var/ftp/pub/idh
createrepo .
```

6. 重启 Apache* HTTP（或 FTP）服务器：

如果是 HTTP 服务器，输入命令：

```
service httpd restart
```

如果是 FTP 服务器，输入命令：

```
service vsftpd restart
```

7. 确认 Apache* HTTP（或 FTP）服务器在运行中，且能被要加入到集群的所有节点访问。
8. 在 `/etc/yum.repos.d` 中，创建一个名为 `idh-local.repo` 的文件，用于指向你创建的新 repository。以下为 repo 文件的内容示例：

```
[idh]
name=Intel-Distribution-for-Apache-Hadoop-R2.4
baseurl=http://[HOSTNAME]/idh
gpgcheck=0
enabled=1
```

9. 在 repo 文件中，确保 `baseurl` 属性的值指向 HTTP 服务器的 `idh` 目录。

如果是 HTTP 服务器，`baseurl` 地址如下：

```
http://hostname/idh
```

如果是 FTP 服务器，`baseurl` 地址如下：

```
ftp://hostname/pub/idh
```

比如，如果机器主机名是 `acme.com`，且 HTTP 服务器侦听端口为 9090，则 `baseurl` 的值应为：
`http://acme.com:9090/idh`。

10. 执行以下命令清除 yum 缓存：

```
yum clean all
```

11. 执行以下命令查看 yum 列表：

```
yum repolist
```

12. 确认在运行 `repolist` 命令的过程中没有出错消息。



13. 如果有出错消息，则表明机器和存放 RPM 包的 HTTP（或 FTP）服务器通信受到干扰。这些网络通信问题通常由防火墙或代理服务器引起。
14. 要使另一台机器能访问此软件库，使用 `scp` 命令复制 `/etc/yum.repos.d/idh-local.repo` 到远程机器的 `/etc/yum.repos.d/` 目录。

3.1.3 在 SLES 上创建英特尔® 发行版软件库

要从 tarball 创建英特尔® 发行版软件库，在你希望创建软件库的机器上执行以下步骤。

1. 将 tarball 解压到 `/tmp` 目录：


```
tar zxvf <tarball name>
```
2. 进入到 `/tmp/intelhadoop` 目录：


```
cd intelhadoop
```
3. 在要存放英特尔® 发行版 RPM 的 HTTP 服务器上创建目录：


```
mkdir -p /srv/www/htdocs/idh
```
4. 执行以下命令，将下列英特尔® 发行版 RPM 复制到新创建的目录。
 - `cp manager /var/www/html/idh`
 - `cp os_related /srv/www/htdocs/idh`
 - `cp idh /srv/www/htdocs/idh`
5. 执行以下命令，在新建目录下创建 repo 数据：


```
cd /srv/www/htdocs/idh
createrepo .
```
6. 确认 Apache* HTTP 服务器在运行中，且能被要加入到集群的所有节点访问。
7. 在 `/etc/zypp/repos.d` 中，创建一个名为 `idh-local.repo` 的文件，用于指向你创建的新 repository。以下为 repo 文件的内容示例：

```
[idh]
name=Intel-Distribution-for-Apache-Hadoop-R2.4
baseurl=http://[HOSTNAME]/idh
gpgcheck=0
enabled=1
```

8. 在 repo 文件中，确保 `baseurl` 属性的值指向 HTTP 服务器的 `idh` 目录。如果是 HTTP 服务器，`baseurl` 地址如下：

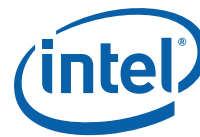

```
http://hostname/idh
```

 比如，如果机器主机名是 `acme.com`，且 HTTP 服务器侦听端口为 9090，则 `baseurl` 的值应为：


```
http://acme.com:9090/idh.
```
9. 执行以下命令清除本地缓存：


```
zypper clean
```
10. 执行以下命令刷新软件库：


```
zypper refresh
```
11. 确认在运行 `refresh` 命令的过程中没有出错消息。
12. 如果有出错消息，则表明机器和存放 RPM 包的 HTTP 服务器通信受到干扰。这些网络通信问题通常由防火墙或代理服务器引起。



13. 要使另一台机器能访问本地镜像软件库，使用 `scp` 命令复制 `/etc/zypp/repos.d/idh-local.repo` 到远程机器的 `/etc/zypp/repos.d` 目录。



4.0 安装 Intel® Manager for Apache Hadoop* software

以下章节描述如何安装 Intel® Manager for Apache Hadoop* software.

4.1 先决条件

在安装 Intel® Manager 前，你必须了解或操作以下事项：

- 确认管理节点已配置且能够从英特尔® 发行版软件库安装 RPM 包。关于访问设置的详细信息，参看章节 3.0 设置对英特尔® 发行版软件库的访问。
- 确保 Intel® Manager 成为 Hadoop 集群中的节点。这一节点称为管理节点。
- 确认管理节点上没有安装 gnuplot*。
- 决定管理节点是否通过 DNS 服务器访问。推荐机器可通过 DNS 服务器访问。
- 确认你知道管理节点的 root 密码。
- 管理节点必须和集群中的其他节点属于同一子网。
- 管理节点被正确的 NTP 服务器同步。
- 确认管理节点可通过 SSH 访问要加入集群的所有其他节点。
- 你必须能够登录 Intel® Manager for Apache Hadoop* software。你使用的登录用户名必须已被分配管理员的角色。默认的管理员登录帐户如下：
 - 用户名: admin
 - 密码: admin
- 集群中的所有节点运行的操作系统必须和 Intel® Manager 所安装的操作系统一样。
- Intel® Manager 支持以下网页浏览器：
 - Firefox 17 或更高版本
 - Chrome 20 或更高版本
- 确认集群中所有节点可访问操作系统软件库。
- 在你运行安装文件的终端窗口，确认客户端的字符集不是 UTF-8，而是 ISO-8859 字符集之一。如果客户端使用 UTF-8 字符集，安装文件可能不会在终端窗口正确呈现。如果你使用 Putty* SSH 进程，建议你 Putty 客户端的字符集设置为 *Use Font Encoding* 选项。
- Intel® Manager for Apache Hadoop* software 对芯片和内存有以下最低要求：
 - 8GB 内存。
 - 2 个四核芯片（8 核，2 个接口）。
 - 30GB 可用磁盘空间。
 - 1GB 网卡。推荐管理节点使用 10GB 网卡。

4.2 安装 Intel® Manager

Intel® Manager 可通过以下方式安装：



- 快速安装 — 在此安装模式下，你只需要通过一些命令行运行安装脚本，脚本将自动完成 Intel® Manager 的安装过程，此模式要求最小程度的用户输入。
- 交互安装 — 根据向导逐步安装 Intel® Manager。此模式要求用户对每一步都进行输入。

以下章节解释了如何执行这二种安装。

4.2.1 执行 Intel® Manager 快速安装

要安装 Intel® Manager for Apache Hadoop* software，执行以下步骤。

1. 确认管理节点符合章节 4.1 先决条件中的要求。
2. 使用 root 帐户登录管理节点。
3. 解压安装包。
4. 确认管理节点已配置且能够从英特尔® 发行版软件库安装 RPM 包。要了解访问设置的详细信息，参见章节 3.0 设置对英特尔® 发行版软件库的访问。
5. 确定你想要绑定 Intel® Manager 的网卡名称。
6. 决定是否在安装过程中接受 JDK 许可协议。
7. 如果你想要接受 JDK 许可协议，则进入 tarball 解压后的目录，执行以下命令：

```
sh install.sh --mode=silent --networkinterface={nic name}  
--acceptjdklicense=accept
```

比如，如果 Intel® Manager 要和 eth0 网卡绑定，则你需要执行以下命令：

```
sh install.sh --mode=silent --networkinterface=eth0  
--acceptjdklicense=accept
```

4.2.2 执行 Intel® Manager 交互式安装

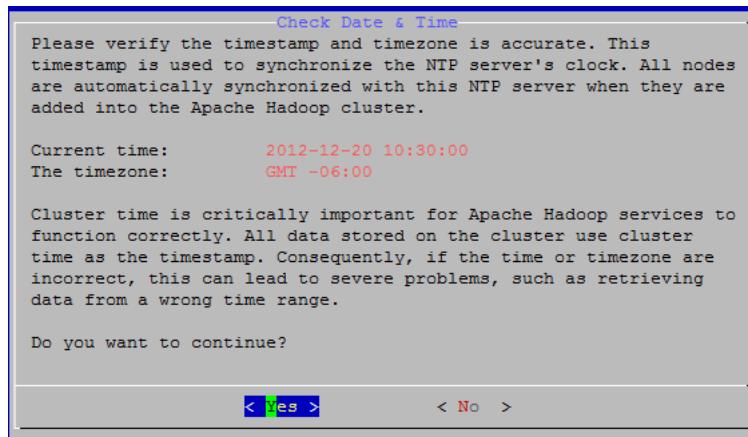
要安装 Intel® Manager for Apache Hadoop* software，执行以下步骤。

1. 确认管理节点符合章节 4.1 先决条件中的要求。
2. 使用 root 帐户登录管理节点。
3. 解压安装包。
4. 确认管理节点已配置且能够从英特尔® 发行版软件库安装 RPM 包。要了解访问设置的详细信息，参见章节 3.0 设置对英特尔® 发行版软件库的访问。
5. 进入 tarball 解压后的目录，执行以下命令：

```
sh install.sh -m=dialog
```

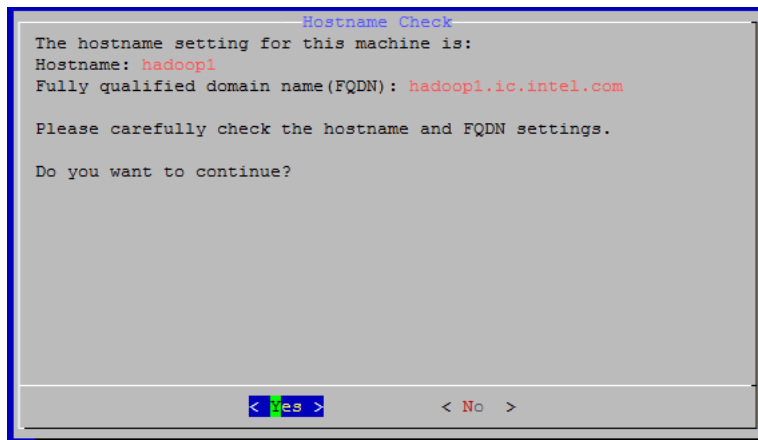


6. 安装文件将显示 Apache Hadoop* 集群使用的日期、时间和时区，并询问这些数据是否正确。

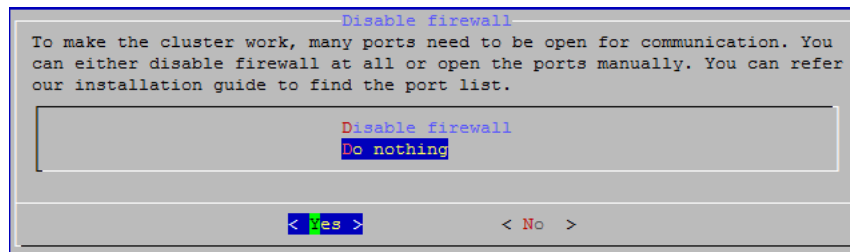


如果日期信息正确，按下键盘上的 Y 键。然后，Accept Java License 页面出现。

7. 如果你接受 Java 许可协议，则按下键盘上的 A 键。
8. 安装向导将显示管理节点的主机名和全称域名。



9. 确认此信息是否正确，如果是，则按下键盘上的 Y 键。然后，Disable Firewall 页面出现。
10. 在 Disable Firewall 页面中，按下箭头键直至 *Do Nothing* 选项被选中，然后按下键盘上的 Y 键。

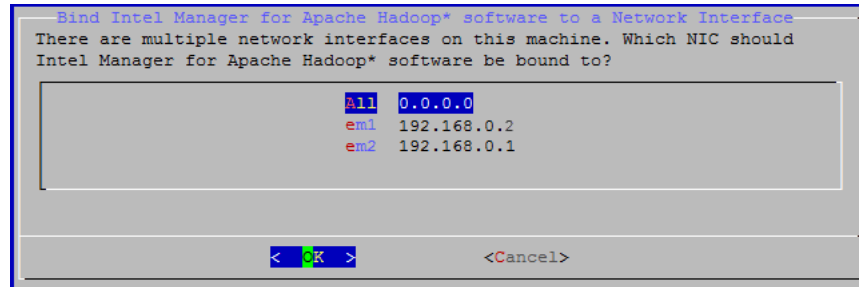


然后，Ports need to be opened 页面出现。

11. 在 Ports need to be opened 页面中，按下键盘上的 O 键。按下后，访问操作系统软件库的设置页面出现。



12. 在访问操作系统软件库的设置页面，按下上下箭头键直至 *Nothing to do* 选项被选中，然后按下键盘上的 Enter 键。
13. 在 Clean Repository Cache 页面中，按下键盘上的 0 键。按下后，访问英特尔® 发行版软件库的设置页面出现。
14. 在访问英特尔® 发行版软件库的设置页面中，确认 *Nothing to do* 选项被选中，然后按下键盘上的 Enter 键。
15. 安装向导将显示安装 RPM 包的进程。如果进程结束，则按下键盘上的 C 键。
16. 如果安装了 Intel® Manager 的机器被分配了多个 IP 地址，则以下信息会显示：



使用上下箭头键来选择可访问 Intel® Manager for Apache Hadoop* software 的 IP 地址。

17. 一旦你将 Intel® Manager 和某个 IP 地址绑定，使用键盘上的左右箭头选择 **OK** 选项，然后按下 Enter。
18. 在 RPM 包开始安装的页面，等待软件安装，然后按下 Enter。
19. 安装过程中将出现数个页面。在每个页面中，按下 Enter 键直至安装完成。

4.3 第一次登录 Intel® Manager

第一次登录 Intel® Manager for Apache Hadoop*，执行以下步骤：

1. 获取 Intel® Manager 的 IP 地址或全限定域名。
2. 打开网页浏览器。
3. 在网页浏览器的地址栏中，输入以下地址：[https://\[Intel® Manager IP 地址或主机名\]:9443](https://[Intel® Manager IP 地址或主机名]:9443)。然后，按下 Enter。
4. 在 Intel® Manager 登录页面，执行以下步骤。
 - a. 在用户名栏内，输入 **admin**。
 - b. 在密码栏内，输入 **admin**。
 - c. 点击**登录**。



5.0 了解 Hadoop 服务的角色

运行集群中的 Hadoop 服务，要求集群中的一个或多个节点执行特定的功能。这些功能被归类为功能性的角色，角色用于定义节点能为集群中的服务提供哪些功能。角色分配是必须的，没有角色集群将无法正常工作。Intel® Manager 提供中央控制点，使你能按需分配角色给任意节点，并可通过点击配置所有节点链接将这些变化部署给所有节点。

在你分配角色前，你需要了解这些角色的含义，以及对应于服务节点所需要的角色。

5.1 了解 HDFS 角色

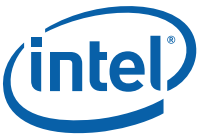
表 4. HDFS 角色列表

角色	描述
Primary NameNode	HDFS 系统中的节点用于维护文件系统中所有文件的目录结构并跟踪文件数据存储在哪些数据节点。当客户端需要从 HDFS 文件系统中获得文件时，它通过和 NameNode 通讯来知道客户端哪个数据节点上有客户端需要的文件。 一个 Hadoop 集群中只能有一个 NameNode。Primary NameNode 不能被赋予其他角色。
DataNode	在 HDFS 中，DataNode 是用来存储数据块的节点。
Secondary NameNode	为 Primary NameNode 上的数据创建周期性检查点的节点。节点将周期性下载当前 NameNode 镜像和日志文件，将日志和镜像文件合并为一个新的镜像文件然后上传到 Primary NameNode。 被分配了 Primary NameNode 角色的机器不应再被分配 Secondary NameNode 角色。

5.2 了解 MapReduce 角色

表 5. MapReduce 角色列表

角色	描述
JobTracker	运行 JobTracker 服务的一个或多个节点。JobTracker 服务负责分配 MapReduce 工作任务给 Hadoop 集群中的节点。一旦任务被分配，JobTracker 将监视执行任务的节点以确认任务完成。
TaskTracker	运行 TaskTracker 服务的一个或多个节点。TaskTracker 服务负责执行 MapReduce 工作任务。JobTracker 服务用于分配 Map，Reduce 或 Shuffle 任务。 推荐不要同时分配 JobTracker 和 TaskTracker 角色给节点。



5.3 了解 HBase 角色

表 6. HBase 角色清单

角色	描述
HMaster	<p>节点实现 HBase 数据库以下功能：</p> <ul style="list-style-type: none">• 分配区域给 RegionServers，平衡整个集群。• 确认 RegionServer 在运行中• 如果没找到参考值，删除 parents• 管理数据库表• 在 Regionservers 中分发消息• 监控 RegionServers 以决定是否有必要执行恢复，如果是，则执行恢复 <p>你可以分配 HMaster 角色给一个或多个节点以进行备份切换。如果你分配角色给多个节点，这将创建一个 active-standby 状态的 HMaster 节点的集群，即一个节点处于 active 状态而集群中的另一个节点处于 standby 状态。如果 active 状态的 HMaster 停止，ZooKeeper 集群将选出一个 inactive 状态的 HMaster 来作为 active 状态的 HMaster。</p>
RegionServer	负责服务和管理 HBase 区域的节点。
ZooKeeper	<p>一个或多个运行 ZooKeeper 服务的节点。ZooKeeper 服务是指包含一个或多个节点的集群提供服务框架用于集群管理。对于集群，ZooKeeper 服务提供的功能包括维护配置信息、命名、提供 HBase 的分布式同步，以及当 HMaster 停止时触发 master 选择。HBase 需要有一个 ZooKeeper 集群才能工作。</p> <p>推荐在 ZooKeeper 集群中至少有 3 个节点。</p>
HBase Thrift	Thrift Client API 开放的节点，客户端可通过 Thrift 和 HBase 通讯。

5.4 了解各种服务角色

表 7. 了解各种服务角色

角色	服务	描述
Hive Server	Hive	Hive 运行的节点。Hive 服务帮助查询和管理 HDFS 的大型数据集。Hive 服务提供工具从文件系统中提取、转换和装载数据， 实施数据格式结构， 以及提供 HDFS 或 HBase 的直接文件访问。这一服务也能通过 MapReduce 执行查询任务。
Oozie Server	Oozie	运行 oozie 服务的节点。Oozie 服务是 Apache Hadoop* 作业的流程调度系统。



6.0 创建 Apache Hadoop* 集群

以下内容解释如何使用 Intel® Manager for Apache Hadoop* software 来创建一个 Apache Hadoop* 集群。

6.1 创建 Apache Hadoop* 集群的要求

在创建 Apache Hadoop* 集群前，你需要了解或操作以下事项：

- 阅读并理解章节 2.0 设计 Hadoop 集群的要求和推荐中所述的要求和建议。
- 理解每个 Apache Hadoop* 角色的含义，参见章节 5.0 了解 Hadoop 服务的角色中的描述。
- 确认将要成为集群中节点所有机器都已配置且能够从英特尔® 发行版软件库安装 RPM 包。要了解设置访问的详细信息，参见章节 3.0 设置对英特尔® 发行版软件库的访问。

6.2 使用 Intel® Manager 向导来创建一个集群

参考以下情形：

- 创建一个 Apache Hadoop* 包含三个节点的集群。
- 每个节点有三个数据分区，并挂载如下：
 - 系统分区是 /
 - HDFS 第一个逻辑磁盘分区是 `/mnt/disk1`
 - HDFS 第二个逻辑磁盘分区是 `/mnt/disk2`
- 节点在同一子网内。子网是 10.302.23.0/24。
- 所有节点都有同样的 root 密码。
- 节点可通过 DNS 访问。
- 节点的主机名为：`ip-10-0-20-17.ec2.internal`，`ip-10-0-20-18.ec2.internal` 和 `ip-10-0-20-19.ec2.internal`。
- 集群上已安装以下 Apache Hadoop* 服务：HDFS、HBase、MapReduce、ZooKeeper、Oozie 和 Hive。
- HDFS 角色已分配给以下节点。
 - `ip-10-0-20-17.ec2.internal` 已被分配 Primary NameNode 角色和 DataNode 角色。
 - `ip-10-0-20-18.ec2.internal` 和 `ip-10-0-20-19.ec2.internal` 已被分配 DataNode 角色。
- MapReduce 角色已分配给以下节点。
 - `ip-10-0-20-18.ec2.internal` 已被分配 JobTracker 和 TaskTracker 角色。
 - `ip-10-0-20-18.ec2.internal` 和 `ip-10-0-20-19.ec2.internal` 已被分配 TaskTracker 角色。
- Hive Server 角色已分配给 `ip-10-0-20-17.ec2.internal`，`ip-10-0-20-18.ec2.internal` 和 `ip-10-0-20-19.ec2.internal`。
- Oozie 角色已分配给 `ip-10-0-20-17.ec2.internal`，`ip-10-0-20-18.ec2.internal` 和 `ip-10-0-20-19.ec2.internal`。



- HBase 角色已分配给以下节点。
 - HMaster 角色已分配给所有节点。
 - RegionServer 角色已分配给 *ip-10-0-20-18.ec2.internal* 和 *ip-10-0-20-19.ec2.internal*。
 - ZooKeeper 角色已分配给所有的三个节点。

以下步骤演示了如何使用 Intel® Manager 在上述情形中创建一个集群。

1. 登录到 Intel® Manager for Apache Hadoop* software。
2. 如果这是你第一次登录，你需要阅读商业或试用许可协议。如果你接受许可协议，则点击**接受**按钮。
3. 如果 Intel® Manager 不能和产品注册服务器通信，屏幕将显示以下信息：

Contact the Registration server

Failed to contact the Registration server. This could be because of incorrect proxy settings. Please fix the settings and try again.

☐ Direct Connection ☒ HTTP Proxy

Proxy :

Port :

Username :

Password :

Cancel Retry

Outbound HTTP communication to the following product registration URLs must be permitted: *https://registration.intelhadoop.com* and *https://registration2.intelhadoop.com*. 如果管理节点必须通过代理服务器和这些外部的 HTTP 服务器通信，则在 Contact Registration Server 页面中执行以下步骤。

- a. 在**代理服务器**栏内，输入 HTTP 代理服务器的 IP 地址或 FQDN。
 - b. 在**端口**栏内，输入代理服务器侦听 HTTP 通信的端口号。
 - c. 如果代理服务器要求客户端验证，则在用户名和密码栏内输入正确的验证信息。
 - d. 点击**重试**按钮。
4. 如果你已经创建一个 Apache Hadoop* 集群，则一个或多个组件可能仍在运行。你必须在创建集群前停止这些组件的运行。要停止某个组件，参见英特尔® Apache Hadoop* 软件发行版操作管理手册。

5. 如果向导没有自动出现，选择页面右上角的[配置向导](#)链接。

第1步

配置新的集群

集群名称：

选择集群中将会使用的组件，包括HDFS，MapReduce，HBase，Hive，Sqoop，Pig，Flume 和 Oozie；另外高可用性组件将会使用两台主备机器来保证集群的高可用性。

集群组件：

☒ HDFS：HDFS是一个分布式的文件系统。

☒ MapReduce：MapReduce是一种用于分布式系统的并行计算框架。

☒ ZooKeeper：ZooKeeper是一个针对大型分布式系统的可靠协调系统。

☒ HBase：HBases是基于HDFS的分布式的，可伸缩的，版本化的数据库系统。

☒ Hive：Hive是基于Hadoop的数据仓库工具。

☐ Sqoop: Sqoop是用于结构化数据存储和Hadoop之间的数据传输的工具。

☐ Pig: Pig是一个基于Hadoop的大规模数据分析平台。

☐ Flume: Flume是一个分布式的、可靠的、和高可用的海量日志聚合的系统。

☒ Oozie: Oozie是一个用于管理和调度Hadoop任务的工作流引擎。

☐ 高可用性：集群中将会有一台备份机器来保证高可用性。

下一步

取消

6. 在第 1 步页面，确认以下选项框已被勾选。
- HDFS
 - MapReduce
 - ZooKeeper
 - HBase（要求已安装 ZooKeeper）
 - Hive
 - Oozie
7. 在第 1 步页面，你可能还希望安装一些其他的组件。要安装某个组件，选中该组件的选项框。你可以安装以下服务：
- Sqoop
 - Pig
 - Flume
 - Oozie
 - High Availability
8. 点击[下一步](#)。

9. 在**网络环境**下拉菜单中，选择**通过配置好的 DNS 服务器**选项。

第2步

指定集群节点以及网络环境

网络环境：集群中的节点能通过主机名互相访问(通过配置好的DNS服务器或/etc/hosts文件)

节点名称	节点IP	状态
ip-10-0-20-17.ec2.internal	10.0.20.17	已连通

添加节点删除节点

上一步下一步取消

警告：

如果 DNS 服务器可用，不要使用“通过 IP 地址来访问”这一选项。

10. 点击**添加节点**。
11. 在添加节点对话框，执行以下步骤。
- a. 在**添加方式**下拉菜单中，选择**批量添加**选项。

添加机器

添加机器。可以添加单台机器，也可以添加指定IP地址范围内的机器，需要提供root用户的凭证。

添加方式：批量添加

起始IP地址：

结束IP地址：

root用户密码：

开始查找取消



- b. 在**起始 IP 地址**栏内，输入你想要 Intel® Manager 搜索的节点的 IP 地址范围起始地址。比如，如果节点位于以下子网 `10.0.20.0/32` 并且每个节点的第四个字节在 `10` 和 `20` 之间，则你可在此栏内输入 `192.168.10.0`。
- c. 在**结束 IP 地址**栏内，输入你想要 Intel® Manager 搜索的节点的 IP 地址范围结束地址。比如，如果节点位于以下子网 `10.0.20.0/32` 并且每个节点的第四个字节在 `10` 和 `20` 之间，则你可在此栏内输入 `192.168.10.20`。
- d. 在**Root 用户密码**栏内，输入你希望加入到 Apache Hadoop* 集群的子网中的机器的 root 密码。
- e. 点击**开始查找**。点击后，查找结果对话框出现，显示 Intel® Manager 正在搜索指定 IP 地址范围内的节点。
- f. 一旦你看到你想要加入到集群的节点出现，点击停止查找。

查找结果

查找到的机器列表。

查找已经完成。使用提供的用户名和密码能访问的机器有2台,已经选择0台。

<input type="checkbox"/> IP地址	机器名
<input type="checkbox"/> 10.0.20.18	ip-10-0-20-18.ec2.internal
<input type="checkbox"/> 10.0.20.19	ip-10-0-20-19.ec2.internal

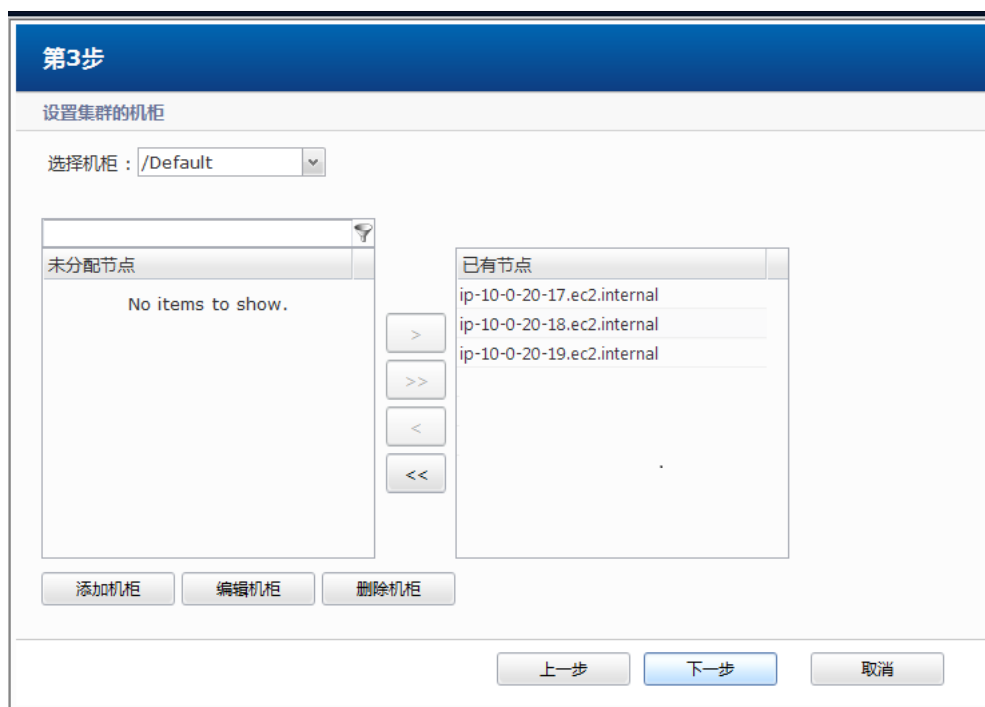
重新查找

添加

关闭

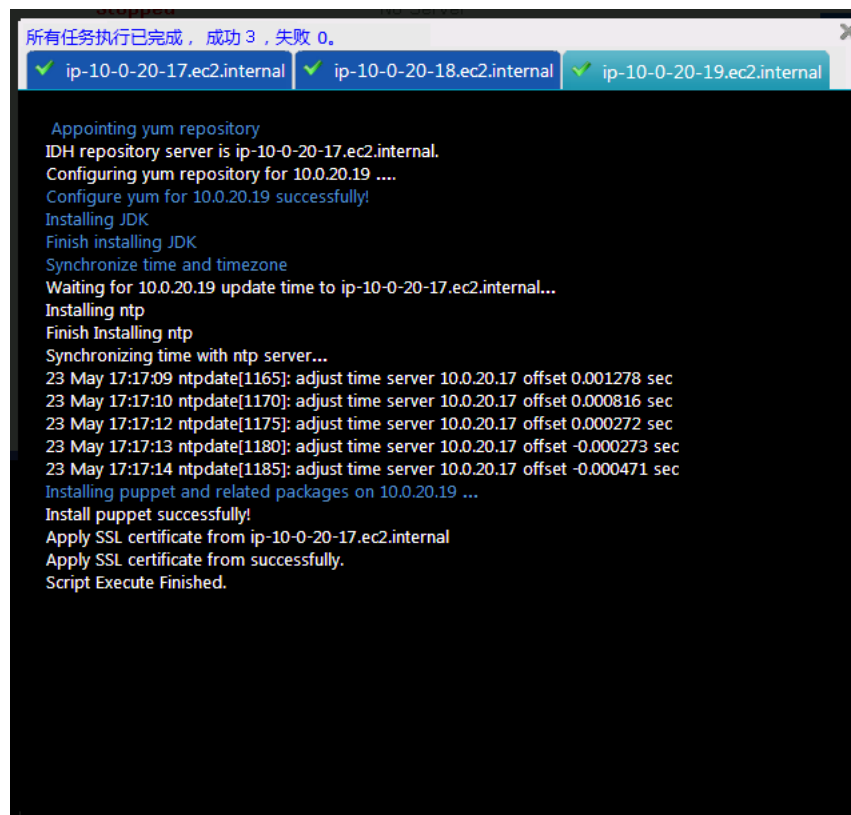
- g. 选择你想要加入到集群的节点的选项框。
 - h. 点击**添加**。
 - i. 当问你是否想要在节点上安装软件时，点击**确定**。
 - j. 在为添加的节点配置无密码登录对话框，等待直至每个节点的状态都变为**已连通**。然后点击**完成**。
12. 如果管理节点尚未被加入到 Apache Hadoop* 集群，请现在添加。
- 注意：** 如果管理节点不属于 APACHE HADOOP* 集群，管理功能将不能正常使用。
13. 在第 2 步页面中，确认每个节点的状态为**已连通**。然后，点击**下一步**。

14. 在第 3 步页面中，所有节点将默认被加入到 /Default 机架。你可在该页面创建一个新的机架，并将节点加入到该机架。然后，点击**下一步**。



15. 在第 4 步页面，指定控制用户访问集群的方法，然后点击**下一步**。

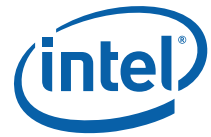
16. 在确认对话框，问你是否想要在所有节点上安装软件，点击**确定**。点击后，将显示节点安装进程的页面。



```
所有任务执行已完成，成功 3，失败 0。
✓ ip-10-0-20-17.ec2.internal ✓ ip-10-0-20-18.ec2.internal ✓ ip-10-0-20-19.ec2.internal

Appointing yum repository
IDH repository server is ip-10-0-20-17.ec2.internal.
Configuring yum repository for 10.0.20.19 ....
Configure yum for 10.0.20.19 successfully!
Installing JDK
Finish installing JDK
Synchronize time and timezone
Waiting for 10.0.20.19 update time to ip-10-0-20-17.ec2.internal...
Installing ntp
Finish Installing ntp
Synchronizing time with ntp server...
23 May 17:17:09 ntpdate[1165]: adjust time server 10.0.20.17 offset 0.001278 sec
23 May 17:17:10 ntpdate[1170]: adjust time server 10.0.20.17 offset 0.000816 sec
23 May 17:17:12 ntpdate[1175]: adjust time server 10.0.20.17 offset 0.000272 sec
23 May 17:17:13 ntpdate[1180]: adjust time server 10.0.20.17 offset -0.000273 sec
23 May 17:17:14 ntpdate[1185]: adjust time server 10.0.20.17 offset -0.000471 sec
Installing puppet and related packages on 10.0.20.19 ...
Install puppet successfully!
Apply SSL certificate from ip-10-0-20-17.ec2.internal
Apply SSL certificate from successfully.
Script Execute Finished.
```

17. 在消息对话框，点击**确定**按钮。如果某个节点的状态是不成功，则你需要重试软件安装直至状态变为成功。
18. 在显示所有节点安装进程的页面，点击右上角的 **X**。
19. 在第 5 步页面，等待直至每个节点的状态都变为**成功**，然后点击**下一步**。
20. 在第 6 步页面，执行以下步骤部署 HDFS 角色。
- 在 **Primary NameNode** 下拉菜单中，选择将会运行 Primary NameNode 服务的节点。
 - 点击**下一步**。
21. 在第 7 步页面，执行以下步骤部署 MapReduce 角色。
- 在 **JobTracker** 下拉菜单中，选择将会运行 JobTracker 服务的节点。
 - 点击**下一步**。
22. 在第 8 步页面，执行以下步骤使至少三个节点成为 ZooKeeper 服务器。
- 在**未分配节点**栏中，双击你希望运行 Zookeeper 服务的节点。双击后，节点将从**未分配节点**栏中消失，出现在**已有节点**栏中。
 - 重复以上步骤，直至三个节点加入到表中。节点的数目必须为奇数。
 - 点击**下一步**。
23. 在第 9 步页面，点击**下一步**。
24. 在第 10 步页面，使至少三个节点成为 Hive 服务器。然后，点击**下一步**。
25. 在第 11 步页面，设置至少三个节点作为 Oozie 服务器。然后，点击**下一步**。



26. 在第 12 步页面，点击**完成**。
27. 在确认对话框，点击**确定**。
28. 在确认对话框，点击**确定**。
29. 在显示所有节点进程的页面，等待直至配置完成，然后点击右上角的 **X**。
30. 执行章节 6.4 指定存储 HDFS 数据的路径中的步骤。
31. 执行章节 6.5 指定存储 MapReduce 数据的路径中的步骤。
32. 在集群节点页面，点击**节点配置**子页面。
33. 在节点配置子页面，点击**格式化集群**链接。

警告：

一旦集群被格式化，所有格式化之前保存在 HDFS 的数据将被删除。

34. 当被问到是否要重新配置集群，点击**确定**。根据网络速度、机器电源以及集群中的机器数目的情形而定，格式化过程可能需要较长时间。请耐心等待至少 10 到 15 分钟，直至格式化成功。

在集群格式化过程中，页面将显示每个节点的格式化进度。如果所有节点的格式化过程成功完成，则 Apache Hadoop* 集群创建成功。

6.3 在 Apache Hadoop* 集群中分配角色给每个节点

要让 Apache Hadoop* 集群正常工作，每个 Apache Hadoop* 服务要求集群中的一个或多个节点履行一个或多个角色。以下步骤演示了如何分配角色给这些节点。以下步骤假设你使用章节 6.2 使用 Intel® Manager 向导来创建一个集群情形中的节点。

1. 确认你已执行章节 6.2 使用 Intel® Manager 向导来创建一个集群中的步骤。
2. 确认你理解每个服务的角色含义。关于这些角色的信息，参见章节 5.0 了解 Hadoop 服务的角色。
3. 在集群配置菜单中，选择**集群节点**选项。
4. 在集群节点页面，选择**角色配置**子页面。
5. 在角色配置子页面，选择**推荐配置**链接。
6. 在确认对话框，点击**确定**。
7. 在介绍页面，点击**下一步**。

第 1 步

HDFS 组件控制节点配置

Primary NameNode :

(*)必填，集群中必须包含一个 Primary NameNode。

Secondary NameNode :

选填，Secondary Namenode 可以备份 Primary NameNode 的元数据。

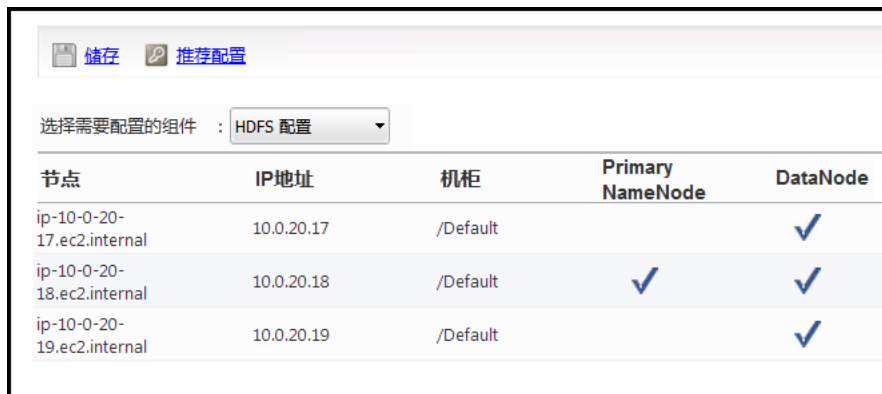
上一步

下一步

取消

8. 在第 1 步页面，执行以下步骤部署 HDFS 角色。
 - a. 在 **Primary NameNode** 下拉菜单中，选择将会运行 Primary NameNode 服务的节点。
 - b. 点击**下一步**。
9. 在第 2 步页面，执行以下步骤部署 MapReduce 角色。

- a. 在 **JobTracker** 下拉菜单中，选择将会运行 JobTracker 服务的节点。
 - b. 点击**下一步**。
10. 在第 3 步页面，执行以下步骤使至少三个节点成为 ZooKeeper 服务器。
 - a. 在表格中，点击带 + 号的绿色按钮。点击后，一个带 - 号的红色按钮出现在新的一行里。
 - b. 双击包含红色按钮的行，直至出现一个显示节点清单的下拉菜单。
 - c. 从下拉菜单中，选择一个节点。
 - d. 重复以上步骤，直至三个节点加入到表中。
 - e. 点击**下一步**。
11. 在第 4 步页面，分配 HBase master 角色给三个或更多节点。然后点击**完成**。
12. 在第 5 步页面，执行以下步骤使一个或多个节点成为 Hive 服务器。
 - a. 在表格中，点击带 + 号的绿色按钮。点击后，一个带 - 号的红色按钮出现在新的一行里。
 - b. 双击包含红色按钮的行，直至出现一个显示节点清单的下拉菜单。
 - c. 从下拉菜单中，选择一个节点。
 - d. 重复以上步骤，直至三个节点加入到表中。
 - e. 点击**下一步**。
13. 在第 6 步页面，分配 Oozie Server 角色给三个或更多节点。然后点击**完成**。
14. 在第 7 步页面，点击**完成**。
15. 在角色配置子页面，执行以下页面分配 DataNode 角色。
 - a. 在**需要配置的节点**下拉菜单中，选择 **HDFS 配置**。



节点	IP地址	机柜	Primary NameNode	DataNode
ip-10-0-20-17.ec2.internal	10.0.20.17	/Default		✓
ip-10-0-20-18.ec2.internal	10.0.20.18	/Default	✓	✓
ip-10-0-20-19.ec2.internal	10.0.20.19	/Default		✓

- b. 在 DataNode 栏内，确认该节点是否显示打勾。打勾则表示该节点是一个 DataNode。
 - c. 如果 DataNode 栏内没有打勾，而你希望这一节点成为 DataNode，在此节点的行上点击。点击后，该节点将显示打勾。
 - d. 如果你不希望该节点成为 DataNode，而该节点当前显示打勾，在此节点的行上点击。点击后，该节点显示的打勾将消失。
16. 在角色配置子页面，执行以下步骤分配 HBase 角色。
 - a. 在**需要配置的节点**下拉菜单中，选择 **HBase 配置**。
 - b. 在 HMaster 和 RegionServer 栏内，确认该节点是否显示打勾。HMaster 栏打勾则表示该节点是一个 HBase master 节点。RegionServer 栏打勾则表示该节点是一个 RegionServer 节点。
 - c. 如果 HMaster 或 RegionServer 栏内没有打勾，而你希望这一节点成为 Hbase Master 节点或 RegionServer，在对应栏的此节点的行上点击。点击后，该节点将显示打勾。

- d. 如果 HMaster 或 RegionServer 栏内没有打勾，而你希望这一节点成为 Hbase Master 节点或 RegionServer，在对应栏的此节点的行上点击。点击后，该节点显示的打勾将消失。
17. 在角色配置页面，点击右上角的**储存**链接。

6.4 指定存储 HDFS 数据的路径

DataNodes 用来存储 HDFS 数据。DataNodes 用来存储 HDFS 数据。默认情况下，Intel® Manager 分配节点上所有挂载的磁盘分区来存储 HDFS 数据。如果你需要改变用于存储数据的磁盘分区，你可以按照以下步骤执行，但注意你必须将数据和系统分区完全隔离。

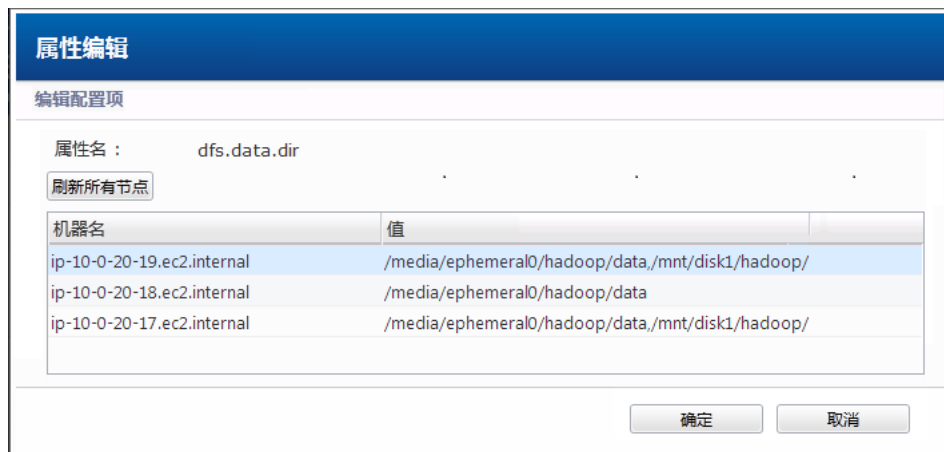
要为每个节点分配用于存储 HDFS 数据的磁盘分区，执行 Intel® Manager 以下步骤。

1. 对每个 DataNode，确认你知道用来存储 HDFS 数据的每个逻辑磁盘分区的挂载名称。
2. 展开集群配置菜单。
3. 在集群配置菜单中，选择 **HDFS** 选项。
4. 在 HDFS 页面中，选择**高级配置**子页面。
5. 在高级配置子页面，在**关键字过滤**栏内输入 **dfs.data.dir**。



属性	值
Datanode 在本地文件系统存储块的位置	dfs.data.dir

6. 在值栏内，双击 (...) 省略号。



机器名	值
ip-10-0-20-19.ec2.internal	/media/ephemeral0/hadoop/data,/mnt/disk1/hadoop/
ip-10-0-20-18.ec2.internal	/media/ephemeral0/hadoop/data
ip-10-0-20-17.ec2.internal	/media/ephemeral0/hadoop/data,/mnt/disk1/hadoop/



7. 为每个节点输入存储 HDFS 数据的逻辑磁盘分区列表，分区间以逗号区隔。每个路径必须以逻辑挂载名称开始，以路径 `/hadoop/data` 结尾。比如，如果一个节点有二个逻辑磁盘分区，`/mnt/disk1` 和 `/mnt/disk2`，你需要为该节点输入以下路径：`/mnt/disk1/hadoop/data`，`/mnt/disk2/hadoop/data`。

属性编辑

编辑配置项

属性名：

dfs.data.dir

刷新所有节点

机器名	值
ip-10-0-20-19.ec2.internal	/mnt/disk1/hadoop/data, /mnt/disk2/hadoop/data
ip-10-0-20-18.ec2.internal	/mnt/disk1/hadoop/data, /mnt/disk2/hadoop/data
ip-10-0-20-17.ec2.internal	/mnt/disk1/hadoop/data, /mnt/disk2/hadoop/data

确定

取消

- 警告：
- 确保在逗号前后没有空格。
8. 在属性编辑对话框，点击**确定**。

9. 在高级配置页面，点击右上角的**储存**。

10. 在提交修改对话框，输入简短修改描述，然后点击**提交**。

11. 在消息对话框，点击**确定**。
- 6.5 指定存储 MapReduce 数据的路径
- JobTracker 和 TaskTracker 节点用来存储 MapReduce 数据。默认情况下，Intel® Manager 分配节点上所有挂载的磁盘分区来存储 MapReduce 数据。如果你需要改变用于存储数据的磁盘分区，你可以按照以下步骤执行，但注意你必须将数据和系统分区完全隔离。
- 要为每个节点分配用于存储 MapReduce 数据的磁盘分区，执行 Intel® Manager 以下步骤。
1. 对每个 JobTracker 和 TaskTracker 节点，确认你知道用来存储 MapReduce 数据的每个逻辑磁盘分区的挂载名称。
- 注释：

这些分区必须和存储 HDFS 数据的分区位置相同。关于更多 HDFS 数据的存储，参见章节 [6.4 指定存储 HDFS 数据的路径](#)。

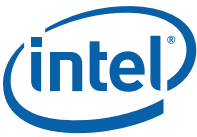
2. 展开集群配置菜单。

3. 在集群配置菜单，选择 **MapReduce** 选项。

4. 在 MapReduce 页面中，选择**高级配置**子页面。

英特尔® Apache Hadoop* 软件发行版安装手册

35



5. 在高级配置子页面，在关键字过滤栏内输入 `mapred.local.dir`。

简要配置高级配置Intel® 智能优化

保存重置编辑添加删除

选择配置类：全部配置关键字过滤：mapred.local.dir

属性	值
执行任务所需的 mapred.local.dir 最小空间 mapred.local.dir.minspacestart	0
mapred.local.dir 最小空间，低于这个值杀死任务 mapred.local.dir.minspacekill	0
MapReduce 存储中间结果的本地文件夹 mapred.local.dir	(...)

6. 在值栏内，双击省略号打开 `mapred.local.dir` 属性。

属性编辑

编辑配置项

属性名：mapred.local.dir刷新所有节点

机器名	值
ip-10-0-20-19.ec2.internal	/media/ephemeral0/hadoop/mapred,/mnt/disk1/ha
ip-10-0-20-18.ec2.internal	/media/ephemeral0/hadoop/mapred
ip-10-0-20-17.ec2.internal	/media/ephemeral0/hadoop/mapred,/mnt/disk1/ha

确定取消

7. 为每个节点输入存储 HDFS 数据的逻辑磁盘分区列表，分区间以逗号区隔。每个路径必须以逻辑挂载名称开始，以路径 `/hadoop/mapred` 结尾。比如，如果一个节点有二个逻辑磁盘分区，`/mnt/disk1` 和 `/mnt/disk2`，你需要为该节点输入以下路径：`/mnt/disk1/hadoop/mapred`，`/mnt/disk2/hadoop/mapred`。

警告： 确保在逗号前后没有空格。

- 8. 在属性编辑对话框，点击**确定**。
- 9. 在高级配置页面，点击右上角的**储存**。
- 10. 在提交修改对话框，输入简短修改描述，然后点击**提交**。
- 11. 在消息对话框，点击**确定**。



7.0 升级英特尔® Apache Hadoop* 软件发行版

以下说明解释了如何在英特尔® Apache Hadoop* 软件发行版集群中为整个集群升级所有组件。在升级过程中，你需要上传包含 Intel® Manager 的 RPM 软件包的一个或多个 tarball 文件，Intel® Manager 将自动更新集群中每个节点的 RPM 软件包。

7.1 先决条件

在升级前，你需要执行或了解以下事项：

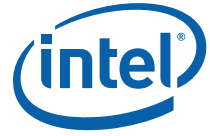
- 你需要知道你即将升级哪个或哪些组件。你可以升级以下组件：
 - Core Hadoop
 - HBase
 - ZooKeeper
 - Flume
 - Sqoop
 - Pig
 - Oozie
 - Intel® Manager
- 除了 Intel® Manager，你需要将要升级的组件停止。
- 确认将要成为集群中节点所有机器都已配置且能够从最新版本的英特尔® Apache Hadoop* 软件发行版软件库安装 RPM 包。要了解访问设置的详细信息，参见[章节 3.0 设置对英特尔® 发行版软件库的访问](#)。
- 如果发生升级，则很可能集群中的节点被配置了和较旧版本的英特尔® Apache Hadoop* 软件发行版软件库通信。确保集群中所有节点用于和较旧版本的英特尔® Apache Hadoop* 软件发行版软件库通信的 repo 文件已被删除或禁用。

注意：要避免安装错误版本的 RPM 包，至关重要的一点是，集群中所有系统对较旧版本的英特尔® APACHE HADOOP* 软件发行版软件库的访问已被永久禁用。

7.2 在 Red Hat Enterprise Linux 上升级英特尔® Apache Hadoop* 软件发行版

要升级英特尔® Apache Hadoop* 软件发行版，执行以下步骤：

1. 执行以下步骤，从英特尔® 发行版 软件库获取 `intelcloudui` 和 `idh-management` RPM 包。
 - a. 登录到能访问当前发布的英特尔® 发行版软件库的机器。
 - b. 确认该机器不是管理节点。
 - c. 安装 `yum-downloadonly` RPM 包。
 - d. 执行以下命令：`yum clean all`
 - e. 执行以下命令：`mkdir -p /tmp/intelmanager-rpms`
 - f. 执行以下命令：
`yum install idh-management --downloadonly --downloadaddir=/tmp/intelmanager-rpms`



- g. 执行以下命令：

```
yum install intelcloudui --downloadonly --downloadaddir=/tmp/
intelmanager-rpms
```
- h. 将 `/tmp/intelmanager-rpms` 下的软件包复制到你想要登录 Intel® Manager 的客户机上。
2. 在客户机上，以管理员身份登录 Intel® Manager。
3. 在系统配置菜单中，点击**组件升级**选项。



4. 在 Intel® Manager 子页面，点击**浏览**按钮。
5. 在打开对话框，浏览并双击 `idh-management` 这一 RPM 包。
6. 点击**上传**。
7. 在打开对话框，浏览并双击 `intelcloudui` 这一 RPM 包。
8. 点击**上传**。
9. 点击**升级**。

7.3 在 SUSE Enterprise Linux 上升级英特尔® Apache Hadoop* 软件发行版

要升级英特尔® Apache Hadoop* 软件发行版，执行以下步骤：

1. 执行以下步骤，从英特尔® Apache Hadoop* 软件发行版软件库获取 `intelcloudui` 和 `idh-management` RPM 包。
 - a. 登录到能访问当前发布的英特尔® Apache Hadoop* 软件发行版软件库的机器。
 - b. 确认该机器不是管理节点。
 - c. 确认 zypper 工具的缓存已启用。
 - d. 执行以下命令：`zypper clean`
 - e. 执行以下命令：

```
zypper install --download-only idh-management
```
 - f. 执行以下命令：

```
zypper install --download-only intelcloudui
```

然后，`intelcloudui` 和 `idh-management` RPM 包将被下载到目录 `/var/cache/zypp` 下。
 - g. 将 `/var/cache/zypp` 下的软件包复制到你想要登录 Intel® Manager 的客户机上。
2. 在客户机上，以管理员身份登录 Intel® Manager。

3. 在系统配置菜单中，点击**组件升级**选项。



4. 在 Intel® Manager 子页面，点击**浏览**按钮。
5. 在打开对话框，浏览并双击 *idh-management* 这一 RPM 包。
6. 点击**上传**。
7. 在打开对话框，浏览并双击 *intelcloudui* 这一 RPM 包。
8. 点击**上传**。
9. 点击**升级**。

7.4 升级 Apache Hadoop* 组件

要升级 Apache Hadoop* 组件，执行以下步骤：

1. 在管理节点上，将新的 RPM 软件包复制到 *idh* 软件包然后进行更新，安装包管理软件可因此探测、检索并升级这些新的软件包。在 RHEL 操作系统中，安装包管理软件是 yum。在 SLES 操作系统中，安装包管理软件是 zypper。
2. 以管理员身份登录 Intel® Manager。
3. 在系统配置菜单中，点击**组件升级**选项。

4. 在组件升级页面，选择 **Hadoop 组件** 子页面。

Intel® Manager

Hadoop组件

刷新

组件名称	版本号	包列表	操作
HADOOP	1.0.3	<div><div>hadoop-1.0.3+230.12865-2.el6.x86_64</div><div>hadoop-conf-pseudo-1.0.3+230.12865-2.el6.noarch</div><div>hadoop-datanode-1.0.3+230.12865-2.el6.noarch</div><div>hadoop-debuginfo-1.0.3+230.12865-2.el6.x86_64</div><div>hadoop-doc-1.0.3+230.12865-2.el6.noarch</div><div>hadoop-jobtracker-1.0.3+230.12865-2.el6.noarch</div><div>hadoop-libhdfs-1.0.3+230.12865-2.el6.x86_64</div><div>hadoop-namenode-1.0.3+230.12865-2.el6.noarch</div><div>hadoop-native-1.0.3+230.12865-2.el6.x86_64</div><div>hadoop-pipes-1.0.3+230.12865-2.el6.x86_64</div><div>hadoop-sbin-1.0.3+230.12865-2.el6.x86_64</div><div>hadoop-secondarynamenode-1.0.3+230.12865-2.el6.noarch</div><div>hadoop-tasktracker-1.0.3+230.12865-2.el6.noarch</div><div>hadoop-fuse-1.0.3+230.12865-2.el6.x86_64</div></div>	<div>上传文件</div>
HBASE	0.94.1	<div><div>hbase-0.94.1+230.12865-1.el6.noarch</div><div>hbase-doc-0.94.1+230.12865-1.el6.noarch</div><div>hbase-master-0.94.1+230.12865-1.el6.noarch</div><div>hbase-regionserver-0.94.1+230.12865-1.el6.noarch</div><div>hbase-rest-0.94.1+230.12865-1.el6.noarch</div><div>hbase-thrift-0.94.1+230.12865-1.el6.noarch</div></div>	<div>上传文件</div>
HIVE	0.9.0	<div><div>hive-0.9.0+230.12865-1.el6.noarch</div><div>hive-metastore-0.9.0+230.12865-1.el6.noarch</div><div>hive-server-0.9.0+230.12865-1.el6.noarch</div></div>	<div>上传文件</div>
ZOOKEEPER	3.4.5	<div><div>zookeeper-server-3.4.5+230.12865-1.el6.noarch</div><div>zookeeper-3.4.5+230.12865-1.el6.noarch</div></div>	<div>上传文件</div>
FLUME	1.3.0	<div><div>flume-1.3.0+230.12865-1.el6.noarch</div><div>flume-node-1.3.0+230.12865-1.el6.noarch</div></div>	<div>上传文件</div>
SQOOP	1.4.1	<div><div>sqoop-1.4.1+230.12865-1.el6.noarch</div><div>sqoop-metastore-1.4.1+230.12865-1.el6.noarch</div></div>	<div>上传文件</div>
PIG	0.9.2	<div><div>pig-0.9.2+230.12865-1.el6.noarch</div></div>	<div>上传文件</div>
OOZIE	3.3.0	<div><div>oozie-3.3.0+230.12865-1.el6.noarch</div><div>oozie-client-3.3.0+230.12865-1.el6.noarch</div></div>	<div>上传文件</div>



5. 在 Hadoop 组件子页面，确认你想要升级的组件。然后，在相应的组件区域，点击**上传文件**。

上传HBASE组件升级包

检测是否有可用的升级包或者上传升级包。

可用的升级包。 [上传新包](#)

Package Name
hbase.noarch-0.94.1+230.12939-1.el6
hbase-doc.noarch-0.94.1+230.12939-1.el6
hbase-master.noarch-0.94.1+230.12939-1.el6
hbase-regionserver.noarch-0.94.1+230.12939-1.el6
hbase-rest.noarch-0.94.1+230.12939-1.el6
hbase-thrift.noarch-0.94.1+230.12939-1.el6

取消下一步

6. 在升级组件对话框，点击**下一步**。在已安装的 RPM 软件包中，所有带新版本的 RPM 软件包已被更新。

Update Nodes

Upgrade all the nodes in cluster .

Node	Status	Operation
hadoop1.ic.intel.com	Upgrade Successfully	N/A
hadoop2.ic.intel.com	Upgrade Successfully	N/A
hadoop3.ic.intel.com	Upgrade Successfully	N/A

All nodes are updated to the latest version.

PreviousFinish

7. 确认每个节点的状态是**升级成功**。如果某个节点不是这个状态，则说明升级失败，你需要重新为该节点升级。
8. 点击**完成**。



8.0 卸载英特尔® Apache Hadoop* 软件发行版

以下章节描述如何卸载英特尔® Apache Hadoop* 软件发行版。卸载表示以下事项将会发生：

- 所有 Apache Hadoop* 服务将停止并从集群中的每个节点删除。
- Apache Hadoop* 服务生成或使用的所有配置文件及日志将从集群中的每个节点删除。
- Intel® Manager for Apache Hadoop* software 将从管理节点删除。

8.1 先决条件

在卸载英特尔® Apache Hadoop* 软件发行版前，你必须了解或执行以下事项：

- 获得 *uninstall.sh* 脚本，这是英特尔® Apache Hadoop* 软件发行版安装包的一部分，复制该脚本到管理节点（如果你还没这么做）。
- 确认你有管理节点的 root 访问权限。
- 确认你有集群中每个节点的 root 密码。
- 确认将要成为集群中节点所有机器都已配置且能够从最新版本的英特尔® Apache Hadoop* 软件发行版软件库安装 RPM 包。要了解访问设置的详细信息，参见[章节 3.0 设置对英特尔® 发行版软件库的访问](#)。
- 确认集群中所有节点上的下列服务已停止：
 - ZooKeeper
 - HDFS
 - MapReduce
 - HBase
 - Hive
 - Oozie

你可以使用 Intel® Manager 来停止这些服务。关于具体步骤，参见[英特尔® Apache Hadoop* 软件发行版操作管理手册](#)。

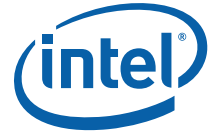
8.2 卸载英特尔® Apache Hadoop* 软件发行版

参考以下情形：

- 你在有三个节点的集群中安装了英特尔® Apache Hadoop* 软件发行版。
- 节点主机名为：*ip-10-0-20-17.ec2.internal*，*hadoop2* 和 *hadoop3*。
- *ip-10-0-20-17.ec2.internal* 是管理节点。
- 管理节点可以访问 DNS 服务，以便将节点主机名解析为 IP 地址。

以下步骤演示了如何在上述情形中卸载英特尔® Apache Hadoop* 软件发行版。

1. 在终端窗口，通过 SSH 命令用 root 证书连接到管理节点。
2. 通过 *cd* 命令进入包含 *uninstall.sh* 脚本的目录。
3. 按照以下方法执行脚本：



```
sh uninstall.sh ALL
```

4. 每个节点会提示以下信息:

```
Type yes to continue or no to exit uninstallation...[yes|no]
```

该提示出现后, 输入 **yes**, 然后按下 Enter。

8.3 卸载示例

```
ip-10-0-20-17.ec2.internal:~ # sh uninstall.sh ALL
```

```
*****
*****
*****

Clean hadoop hbase hive zookeeper sqoop mahout flume pig ganglia puppet
nginx in tel-manager for hadoop2?

*****
*****
*****

Type yes to continue or no to exit uninstallation...[yes|no]: yes
Uninstalling pacemaker ...
Uninstalling corosync ...
Uninstalling hadoop ...
Uninstalling hbase ...
Uninstalling hive ...
Uninstalling mysql ...
Uninstalling zookeeper ...
Uninstalling sqoop ...
Uninstalling mahout ...
Uninstalling flume ...
Uninstalling pig ...
Uninstalling ganglia ...
Uninstalling nagios ...
Uninstalling puppet ...
Uninstalling nginx ...
Uninstalling ftpoverhdfs...
Uninstalling other related packages
Uninstalling Intel Manager for Apache Hadoop
Removing related directories ...
Uninstallation for hadoop2 finished.

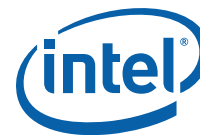
*****
*****

Clean hadoop hbase hive zookeeper sqoop mahout flume pig ganglia puppet
nginx intel-manager for hadoop3?

*****
```



```
*****
Type yes to continue or no to exit uninstallation...[yes|no]: yes
Uninstalling pacemaker ...
Uninstalling corosync ...
Uninstalling hadoop ...
Uninstalling hbase ...
Uninstalling hive ...
Uninstalling mysql ...
Uninstalling zookeeper ...
Uninstalling sqoop ...
Uninstalling mahout ...
Uninstalling flume ...
Uninstalling pig ...
Uninstalling ganglia ...
Uninstalling nagios ...
Uninstalling puppet ...
Uninstalling nginx ...
Uninstalling ftpoverhdfs...
Uninstalling other related packages
Uninstalling Intel Manager for Apache Hadoop
Removing related directories ...
Uninstallation for hadoop3 finished.
```



9.0 管理英特尔® Apache Hadoop* 软件发行版许可证

英特尔® Apache Hadoop* 软件发行版提供以下二种许可证：

- 免费版：支持 50 台服务器，以及 8TB 的 HDFS 容量。
- 商业版：支持的服务器和 HDFS 容量取决于你的商业需求。具体数量根据你购买时的要求而定。

在安装英特尔® Apache Hadoop* 软件发行版后，默认将使用免费版许可证。你也可将免费版升级到商业版以满足你的商业需求。

以下章节解释了如何查看和升级英特尔® Apache Hadoop* 软件发行版许可证。

9.1 查看许可证

要查看当前许可证信息：

1. 以管理员身份登录 Intel® Manager for Apache Hadoop* software。
2. 在集群配置菜单，选择**许可证管理**选项。

节点	角色	服务器ID	许可证状态
ip-10-0-20-80.ec2.internal	Management.Primary NameNode.JobTracker	OusfNZG76498Y/tPizAic4F1qqp1	有效

3. 在许可证管理页面，选中你要查看许可证的节点，然后点击**详细信息**。



许可证信息

查看许可证信息与状态，可以上传新的许可证。

许可证信息

服务器ID： O632mYRRPVUllBtNX7tUmopb28p

许可证类型： 免费版，支持 50 台服务器，支持 8TB HDFS容量

升级许可证

许可证文件：

上传

取消

4. 你可查看以下许可证信息：

- 服务器 ID：Intel® Manager for Apache Hadoop* software 服务器的唯一识别号。你需要提供此服务器 ID 以用来生成进行升级的新许可证。
- 许可证类型：免费版或商业版。
- 包含组件：该许可证允许使用的英特尔® Apache Hadoop* 软件发行版组件。

9.2 升级许可证

要升级许可证，执行以下步骤：

1. 以管理员身份登录 Intel® Manager for Apache Hadoop* software。
2. 在集群配置菜单，选择**许可证管理**选项。
3. 在许可证管理页面，选中你要查看许可证的节点，然后点击**详细信息**。
4. 在升级许可证栏内，输入你获得的新许可证，点击**上传**。



许可证信息

查看许可证信息与状态，可以上传新的许可证。

许可证信息

服务器ID: O632mYRRRePVUllBtNX7tUmopb28p

许可证类型: 免费版, 支持 50 台服务器, 支持 8TB HDFS容量

升级许可证

许可证文件:

```
8pSzFQpd+66gaV4EAJl/Y81Hx4pj31Nri0Lzaxqr0MmFamwHEuKsB1JfE6S8Fwhq
FZ1WU8csPYxd2PyV3GgA68jRSQZL6JBcU/nJsHyo4XI7QdCLYwgYraTZzRzE76dH
71n/B7SVR7WYVSUt4EZ801Zh0tN01uPWajq/wmFnpXpTLlGT229Bqc4rnZX6vVos
/UYE.dkE0XaTjHof3Qht08IRX3Mfg4L6kjCpcjFQQwrrrGz28gAS4RVkelwRWu6Qs
N8s00Wyy/4VM3Gq+IQSQP2JEmAnBq9r1nh2uYXPsb57I08NfDb0n9LCJm3X3ypSY
cB6+hgUW6QM35D8goMzURfVq708vMHUb0+0vg5eoVg6n4BnRAV62YLyYsAoFot/Y
fMT24Nax102phGfdk18+0m3VmMni04fFXDvvGq4/Te+J2kJ4nx5bPd2A.jxCYX18x
3EPvxOTJyECbChIrpeXUVhjFejy06bYYs2927pOmcDuA6SuCIg+VxWKYtBPrSHaS
kbIem/Nj3Iviad5lbnmrUmBlj6RxlCYjxcZQe6CV3QaNwulCvk0QEZVC/zd36gpy
GPoX+4yTDnKXYPp8n0sDeYrLMrU4fH40tp3WhXrjN1QjyrHv4WastHwWlRtGZdrU
```

- 在弹出的消息对话框中，阅读并点击**确认**。

消息

成功提交许可证，请返回许可证管理页面点击“部署许可证”按钮，将新的许可证更新到整个集群！

- 在许可证管理页面中，点击**部署许可证**，将新的许可证部署到整个集群中。
- 部署完成后，系统将弹出消息对话框显示“部署许可证成功”，点击**确认**关闭。
- 点击**刷新许可证信息**以确保系统显示新的许可证信息，然后点击**许可证信息**查看新的许可证信息。



附录 A: 支持的 Apache Hadoop* 组件列表

组件	版本	安装包
Intel® Manager	2.4	<ul style="list-style-type: none"> intelcloudui-240.15530-1.el6.x86_64 idh-management-240.15530-1.el6.x86_64
Hadoop	1.0.3	<ul style="list-style-type: none"> hadoop-doc-1.0.3+240.15529-2.el6.noarch hadoop-libhdfs-1.0.3+240.15529-2.el6.x86_64 hadoop-1.0.3+240.15529-2.el6.x86_64 hadoop-native-1.0.3+240.15529-2.el6.x86_64 hadoop-tasktracker-1.0.3+240.15529-2.el6.noarch hadoop-datanode-1.0.3+240.15529-2.el6.noarch hadoop-namenode-1.0.3+240.15529-2.el6.noarch hadoop-jobtracker-1.0.3+240.15529-2.el6.noarch hadoop-sbin-1.0.3+240.15529-2.el6.x86_64 hadoop-pipes-1.0.3+240.15529-2.el6.x86_64 hadoop-secondarynamenode-1.0.3+240.15529-2.el6.noarch hadoop-conf-pseudo-1.0.3+240.15529-2.el6.noarch hadoop-fuse-1.0.3+240.15529-2.el6.x86_64
HBase	0.94.1	<ul style="list-style-type: none"> hbase-regionserver-0.94.1+240.15529-1.el6.noarch hbase-0.94.1+240.15529-1.el6.noarch hbase-thrift-0.94.1+240.15529-1.el6.noarch hbase-rest-0.94.1+240.15529-1.el6.noarch hbase-doc-0.94.1+240.15529-1.el6.noarch hbase-master-0.94.1+240.15529-1.el6.noarch
Hive	0.9.0	<ul style="list-style-type: none"> hive-metastore-0.9.0+240.15529-1.el6.noarch hive-0.9.0+240.15529-1.el6.noarch hive-server-0.9.0+240.15529-1.el6.noarch
ZooKeeper	3.4.5	<ul style="list-style-type: none"> zookeeper-3.4.5+240.15529-1.el6.noarch zookeeper-server-3.4.5+240.15529-1.el6.noarch
Flume	1.3.0	<ul style="list-style-type: none"> flume-node-1.3.0+240.15529-1.el6.noarch flume-1.3.0+240.15529-1.el6.noarch



组件	版本	安装包
Sqoop	1.4.1	<ul style="list-style-type: none">sqoop-metastore-1.4.3+240.15529-1.el6.noarchsqoop-1.4.3+240.15529-1.el6.noarch
Pig	0.9.2	<ul style="list-style-type: none">pig-0.9.2+240.15529-1.el6.noarch
Oozie	3.3.0	<ul style="list-style-type: none">oozie-3.3.0+240.15529-1.el6.noarchoozie-client-3.3.0+240.15529-1.el6.noarch