

# 英特尔® Apache Hadoop\* 软件发行版安装手册

---

版本 2.3

2013 年 2 月



#### 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 <http://www.intel.com/design/literature.htm>。

英特尔处理器标号不是性能的指标。处理器标号仅用于区分同属一个系列的处理器的特性，而不能够用于区分不同系列的处理器。**详情请登陆：**  
[http://www.intel.com/products/processor\\_number](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。

Intel, 英特尔® Apache Hadoop® 软件发行版, Intel® Distribution, Intel® Manager for Apache Hadoop® software, 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 年版权所有。所有权保留。



## 文档修订记录

---

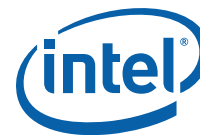
日期	修订	描述
2012 年 7 月	001	英特尔® Apache Hadoop* 软件发行版 v2.0 第一版
2012 年 9 月	002	英特尔® Apache Hadoop* 软件发行版 v2.1 文档更新
2012 年 11 月	003	英特尔® Apache Hadoop* 软件发行版 v2.1.1 文档更新
2012 年 12 月	004	英特尔® Apache Hadoop* 软件发行版 v2.2 文档更新
2013 年 2 月	005	英特尔® Apache Hadoop* 软件发行版 v2.3 文档更新
2013 年 2 月	006	英特尔® Apache Hadoop* 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



## 目录

---

<b>1.0</b>	<b>简介</b>	5
1.1	什么是管理节点?	5
<b>2.0</b>	<b>支持的 Apache Hadoop* 组件列表</b>	6
<b>3.0</b>	<b>设计 Hadoop 集群的要求和推荐</b>	8
3.1	支持的操作系统	8
3.2	硬件要求	8
3.3	软件要求	8
3.4	内存要求	9
3.5	磁盘分区要求	10
3.6	集群和网络拓扑要求	10
3.7	端口要求	11
3.8	推荐配置	13
<b>4.0</b>	<b>支持的 Apache Hadoop* 服务版本</b>	15
<b>5.0</b>	<b>安装 Intel® Manager for Apache Hadoop* software</b>	17
5.1	先决条件	17
5.2	安装 Intel® Manager	17
5.3	第一次登录 Intel® Manager	19
<b>6.0</b>	<b>了解 Hadoop 服务的角色</b>	21
6.1	了解 HDFS 角色	21
6.2	了解 MapReduce 角色	21
6.3	了解 HBase 角色	22
6.4	了解各种服务角色	22
<b>7.0</b>	<b>创建 Apache Hadoop* 集群</b>	23
7.1	创建 Apache Hadoop* 集群的要求	23
7.2	使用 Intel® Manager 向导来创建一个集群	23
7.3	在 Apache Hadoop* 集群中分配角色给每个节点	30
7.4	指定存储 HDFS 数据的路径	32
7.5	指定存储 MapReduce 数据的路径	34
<b>8.0</b>	<b>升级 Apache Hadoop* 英特尔® Apache Hadoop* 软件发行版</b>	36
8.1	先决条件	36
8.2	升级 Intel® Manager for Apache Hadoop* software	36
8.3	升级 Apache Hadoop* 组件	37
<b>9.0</b>	<b>卸载英特尔® Apache Hadoop* 软件发行版</b>	40
9.1	先决条件	40
9.2	卸载英特尔® Apache Hadoop* 软件发行版	40
9.3	卸载示例	41



## 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 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。

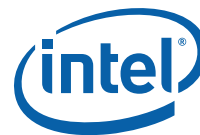


## 2.0 支持的 Apache Hadoop\* 组件列表

组件	版本	安装包
Intel® Manager	2.3	<ul style="list-style-type: none"> <li>intelcloudui-230.12961-1.el6.x86_64</li> </ul>
Hadoop	1.0.3	<ul style="list-style-type: none"> <li>hadoop-tasktracker-1.0.3+230.12957-2.el6.noarch</li> <li>hadoop-namenode-1.0.3+230.12957-2.el6.noarch</li> <li>hadoop-libhdfs-1.0.3+230.12957-2.el6.x86_64</li> <li>hadoop-datanode-1.0.3+230.12957-2.el6.noarch</li> <li>hadoop-pipes-1.0.3+230.12957-2.el6.x86_64</li> <li>hadoop-1.0.3+230.12957-2.el6.x86_64</li> <li>hadoop-native-1.0.3+230.12957-2.el6.x86_64</li> <li>hadoop-doc-1.0.3+230.12957-2.el6.noarch</li> <li>hadoop-secondarynamenode-1.0.3+230.12957-2.el6.noarch</li> <li>hadoop-conf-pseudo-1.0.3+230.12957-2.el6.noarch</li> <li>hadoop-fuse-1.0.3+230.12957-2.el6.x86_64</li> <li>hadoop-sbin-1.0.3+230.12957-2.el6.x86_64</li> <li>hadoop-jobtracker-1.0.3+230.12957-2.el6.noarch</li> </ul>
HBase	0.94.1	<ul style="list-style-type: none"> <li>hbase-regionserver-0.94.1+230.12957-1.el6.noarch</li> <li>hbase-rest-0.94.1+230.12957-1.el6.noarch</li> <li>hbase-master-0.94.1+230.12957-1.el6.noarch</li> <li>hbase-thrift-0.94.1+230.12957-1.el6.noarch</li> <li>hbase-0.94.1+230.12957-1.el6.noarch</li> <li>hbase-doc-0.94.1+230.12957-1.el6.noarch</li> </ul>
Hive	0.9.0	<ul style="list-style-type: none"> <li>hive-server-0.9.0+230.12957-1.el6.noarch</li> <li>hive-metastore-0.9.0+230.12957-1.el6.noarch</li> <li>hive-0.9.0+230.12957-1.el6.noarch</li> </ul>
ZooKeeper	3.4.5	<ul style="list-style-type: none"> <li>zookeeper-server-3.4.5+230.12957-1.el6.noarch</li> <li>zookeeper-3.4.5+230.12957-1.el6.noarch</li> </ul>
Flume	1.3.0	<ul style="list-style-type: none"> <li>flume-node-1.3.0+230.12957-1.el6.noarch</li> <li>flume-1.3.0+230.12957-1.el6.noarch</li> </ul>



组件	版本	安装包
Sqoop	1.4.1	<ul style="list-style-type: none"><li>sqoop-1.4.1+230.12957-1.el6.noarch</li><li>sqoop-metastore-1.4.1+230.12957-1.el6.noarch</li></ul>
Pig	0.9.2	<ul style="list-style-type: none"><li>pig-0.9.2+230.12957-1.el6.noarch</li></ul>
Oozie	3.3.0	<ul style="list-style-type: none"><li>oozie-3.3.0+230.12957-1.el6.noarch</li><li>oozie-client-3.3.0+230.12957-1.el6.noarch</li></ul>



## 3.0 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。这包括集群中的每个节点的网络及软硬件配置的决定。

### 3.1 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"><li>• 6.1</li><li>• 6.2</li><li>• 6.3</li></ul>
CentOS*	<ul style="list-style-type: none"><li>• 6.1</li><li>• 6.2</li><li>• 6.3</li></ul>
Oracle Enterprise Linux* with Unbreakable Enterprise Kernel	<ul style="list-style-type: none"><li>• 6.1</li><li>• 6.2</li><li>• 6.3</li></ul>
SUSE Linux Enterprise*	<ul style="list-style-type: none"><li>• 11 SP1</li><li>• 11 SP2</li></ul>

只支持 64 位结构。

### 3.2 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。也就是说，DataNodes 使用的逻辑盘分区不能用于磁盘冗余阵列。
- 确认每个节点至少有二个逻辑盘分区。一个分区是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。

### 3.3 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。
- 集群中的所有节点之间必须允许 SSH，HTTP 和 FTP 通讯。
- 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。
- 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包（repository）。集群中的所有节点必须使用相同的操作系统软件包。  
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包安装 RPM 包。对于 OEL 和 RHEL 操作系统，这即为 yum 软件包。对于 SLES 操作系统，这即为 zypper 软件包。

### 3.4 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：

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

### 3.5 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：

- 必须存在三个逻辑分区：
  - swap
  - 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 的数据，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。
- 每个逻辑分区应该使用 *ext4* 文件系统。
- 用于存放操作系统的根分区，必须至少有 30GB。
- 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，分区的挂载名称应该为： */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。

### 3.6 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 所在的节点成为 Hadoop 集群中的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 Red Ha、Oracle Enterprise Linux 或 CentOS 操作系统已安装在节点上，则在每个节点上确认 FQDN 已在以下文件中注明。



- /proc/sys/kernel/hostname
- /etc/sysconfig/network
- 如果 SUSE Linux Enterprise Server 操作系统已安装在节点上，则在每个节点上确认 FQDN 已在以下文件中注明。
  - /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

### 3.7 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 [6.0 了解 Hadoop 服务的角色](#)。

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. 管理节点的端口要求](#) 以下列出这些要求。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。



表 2. 管理节点的端口要求

服务名称	端口
resourcemonitor	7070
intel-manager	9443
gmond	8649
gmetad	<ul style="list-style-type: none"> <li>8561</li> <li>8562</li> </ul>
httpd	<ul style="list-style-type: none"> <li>80</li> <li>443</li> </ul>
nagios	<ul style="list-style-type: none"> <li>80</li> <li>443</li> </ul>
puppetmasterd	<ul style="list-style-type: none"> <li>18140</li> <li>18141</li> <li>18142</li> <li>18143</li> </ul>
nginx	8140
DNS	默认端口号为 53。端口号也有可能不同，这取决于网络管理员如何配置 DNS。
vsftpd	21 你需要打开一个非特定端口范围，用于处理被动模式的端口。端口范围需要在 <code>/etc/vsftpd/vsftpd.conf</code> 中注明。

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。

表 3. Hadoop 集群的端口要求

服务名称	端口	要求的角色
gmond	8649	这一服务运行于所有节点，无论该节点被分配了哪些角色。
DNS	默认端口号为 53。端口号也有可能不同，这取决于网络管理员如何配置 DNS。	不可用。
vsftpd	21 你需要打开一个非特定端口范围，用于处理被动模式的端口。端口范围需要在 <code>/etc/vsftpd/vsftpd.conf</code> 中注明。	不可用。
HDFS		

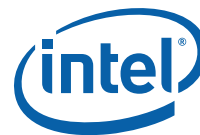


表 3. Hadoop 集群的端口要求

服务名称	端口	要求的角色
hadoop-namenode	<ul style="list-style-type: none"> <li>50070</li> <li>8020</li> </ul>	Primary NameNode
hadoop-datanode	50075	DataNode
hadoop-secondarynamenode	50090	Secondary NameNode
MapReduce		
hadoop-jobtracker	<ul style="list-style-type: none"> <li>50030</li> <li>44436</li> </ul>	JobTracker
hadoop-tasktracker	50010	TaskTracker
HBase		
zookeeper-server	<ul style="list-style-type: none"> <li>3888</li> <li>39123</li> </ul>	ZooKeeper
hbase-master	<ul style="list-style-type: none"> <li>60010</li> <li>60000</li> </ul>	HMaster
hbase-regionserver	60030	RegionServer
hbase-thrift	8080	HBase Thrift
Hive		
mysqld	3306	Primary NameNode
hive-metastore	9083	Primary NameNode
hive-server	10000	Hive Server

### 3.8 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。

- 为避免灾难性的失败，JobTracker 和 PrimaryNameNode 节点应安装在磁盘冗余阵列的节点上。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 1 T 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。但是，最好每个节点有至少 6 T 的可用磁盘空间用于 HDFS。
- 集群中应至少有 5 个 DataNodes。
- 每个节点应配置一块 10GB 的网卡，用于节点间的通讯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。
- 每个节点应该至少有 32GB 的内存。
- 所有节点上的热交换必须关闭。



## 4.0 支持的 Apache Hadoop\* 服务版本

组件	组件	版本
Hadoop	hadoop-conf-pseudo	1.0.3
Hadoop	hadoop-datanode	1.0.3
Hadoop	hadoop-doc	1.0.3
Hadoop	hadoop-fuse	1.0.3
Hadoop	hadoop-jobtracker	1.0.3
Hadoop	hadoop-libhdfs	1.0.3
Hadoop	hadoop-namenode	1.0.3
Hadoop	hadoop-native	1.0.3
Hadoop	hadoop-pipes	1.0.3
Hadoop	hadoop-sbin	1.0.3
Hadoop	hadoop-secondarynamenode	1.0.3
Hadoop	hadoop-tasktracker	1.0.3
Hadoop	hadoop-debuginfo	1.0.3
HBase	hbase	0.94
HBase	hbase-doc	0.94
HBase	hbase-master	0.94
HBase	hbase-regionserver	0.94
HBase	hbase-rest	0.94
HBase	hbase-thrift	0.94
Hive	hive	0.9.0
Hive	hive-metastore	0.9.0
Hive	hive-server	0.9.0
ZooKeeper	zookeeper	3.4.5
ZooKeeper	zookeeper-server	3.4.5
Flume	flume	1.3.0
Flume	flume-node	1.3.0
Sqoop	sqoop	1.4.1
Sqoop	sqoop-metastore	1.4.1
Pig	pig	0.9.2



组件	组件	版本
Oozie	oozie	3.3.0
Oozie	oozie-client	3.3.0





## 5.0 安装 Intel® Manager for Apache Hadoop\* software

---

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

### 5.1 先决条件

在安装 Intel® Manager 前, 你必须了解或操作以下事项:

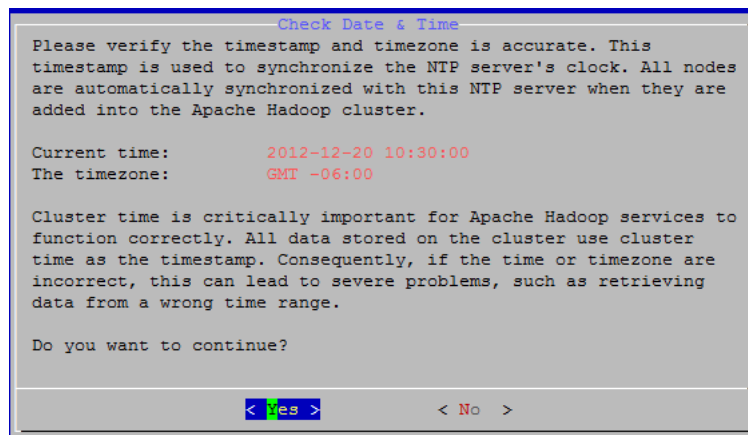
- 确保 Intel® Manager 成为 Hadoop 集群中的节点。这一节点称为管理节点。
- 确认管理节点上没有安装 gnuplot\*。
- 决定管理节点是否通过 DNS 服务器访问。推荐机器可通过 DNS 服务器访问。
- 确认你知道管理节点的 root 密码。
- 管理节点必须对加入集群的所有机器有网络和 SSH 访问权限。
- 管理节点必须和集群中的其他节点属于同一子网。
- 英特尔® Apache Hadoop\* 软件发行版的 tarball 必须被复制到管理节点。
- 管理节点被正确的 NTP 服务器同步。
- 确认管理节点可通过网络访问集群中的所有其他节点。
- 你必须能够登录 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
  - 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包。
- 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, 建议你将在 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

### 5.2 安装 Intel® Manager

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

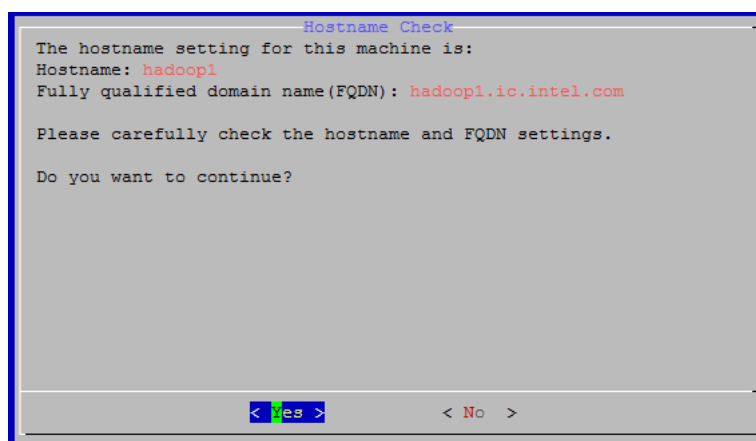


1. 确认管理节点符合章节 5.1 先决条件中的要求。
2. 使用 root 帐户登录管理节点。
3. 通过 `untar` 命令将英特尔® Apache Hadoop\* 软件发行版解压，系统将生成一个 `intelhadoop` 文件夹。
4. 通过 `cd` 命令进入 `intelhadoop` 文件夹。
5. 执行 `install` 二进制文件。
6. 安装文件将显示 Apache Hadoop\* 集群使用的日期，时间和时区，并询问这些数据是否正确。

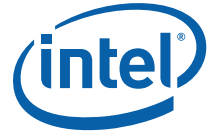


如果日期信息正确，按下键盘上的 **Y** 键。然后，Accept Java License 页面出现。

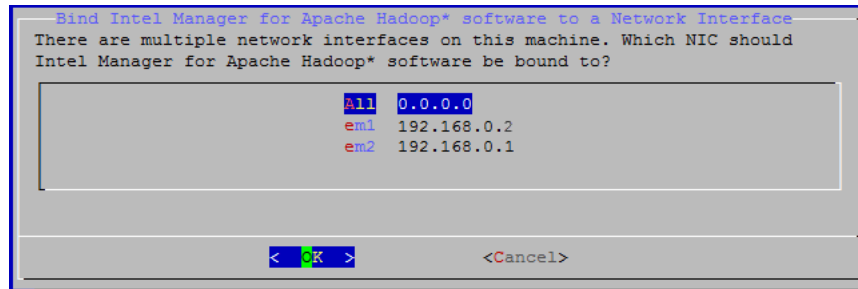
7. 如果你接受 Java 许可协议，则按下键盘上的 **A** 键。
8. 安装向导将显示管理节点的主机名和全称域名。



9. 确认此信息是否正确，如果是，则按下键盘上的 **Y** 键。
10. 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。
11. 当安装过程中显示 *Files are successfully copied*，按下键盘上的 **C** 键。

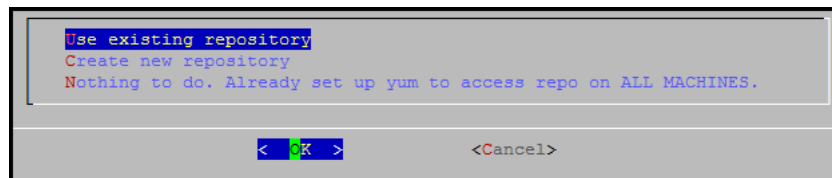


12. 如果安装了 Intel® Manager 的机器被分配了多个 IP 地址，则以下信息会显示：



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

13. 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。
14. 安装文件将出现一个页面，显示 Intel® Manager 被绑定的 IP 地址。在此页面，确认 IP 地址为正确，然后按下键盘上的 **Y** 键。
15. 安装向导将显示信息，询问你是否要移除已存在的软件包文件。在此页面中，按下键盘上的 **N** 键。按下后，设置软件包访问的页面出现。



16. 在 设置软件包页面，按下上下箭头键直至 **Nothing to do** 选项被选中。
17. 按下键盘上的左右箭头直至 **OK** 选项被选中并显示绿色。
18. 在 *Clean Repository Cache* 页面，等待 yum 软件包缓存被清除，然后按下 Enter。
19. 在 *RPM 包开始安装* 的页面，等待软件安装，然后按下 Enter。
20. 在 *Install Apache Hadoop RPMs and Configure Management Software* 页面，等待软件安装，然后按下 Enter。
21. 在 *Set Management Suite* 页面，按下 Enter。
22. 在 *Finish* 页面，当看到以下提示时，你可确认安装已成功完成。  
You have successfully installed Intel Manager for Apache Hadoop!
23. 在 *Finish* 页面，按下 Enter。

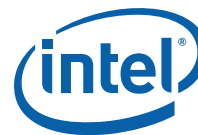
### 5.3 第一次登录 Intel® Manager

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. 点击登录按钮。



## 6.0 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。

### 6.1 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。

### 6.2 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。



6.3 了解 HBase 角色

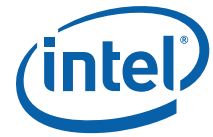
表 6. HBase 角色清单

??	??
HMaster	<p>节点实现 HBase 数据库以下功能：</p> <ul style="list-style-type: none"><li>• 分配区域给 RegionServers，平衡整个集群。</li><li>• 确认 RegionServer 在运行中</li><li>• 如果没找到参考值，删除 parents</li><li>• 管理数据库表</li><li>• 在 RegionServers 中分发消息</li><li>• 监控 RegionServers 以决定是否有必要执行恢复，如果是，则执行恢复</li></ul> <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。

6.4 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。



## 7.0 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。

### 7.1 创建 Apache Hadoop\* 集群的要求

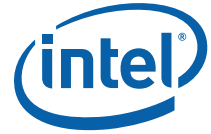
在创建一个 Apache Hadoop\* 集群前，你必须阅读并理解[章节 3.0 设计 Hadoop 集群的要求和推荐](#)中所述的要求和建议。你也必须理解每个 Apache Hadoop\* 角色的含义，参见[章节 6.0 了解 Hadoop 服务的角色](#)。

### 7.2 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。
- 节点主机名为：`hadoop1`，`hadoop2`，和 `hadoop3`。
- 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。
  - `hadoop1` 被分配到 Primary NameNode 和 DataNode 角色。
  - `hadoop2` 和 `hadoop3` 分配到 DataNode 角色。
- MapReduce 角色已分配给以下节点。
  - `hadoop2` 被分配到 JobTracker 和 TaskTracker 角色。
  - `hadoop2` 和 `hadoop3` 分配到 TaskTracker 角色。
- Hive Server 角色已分配给 `hadoop1`、`hadoop2` 和 `hadoop3`。
- Oozie 角色已分配给 `hadoop1`、`hadoop2` 和 `hadoop3`。
- 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 `hadoop2` 和 `hadoop3`。
  - ZooKeeper 角色已分配给所有节点。

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。第一次登录：
  - 如果你的浏览器尺寸不足够大，你将收到提示信息，推荐你使用 1024\*768 的尺寸。点击**确认**继续。
  - 在许可协议页面，阅读许可协议，如果同意，选择**接受**选项。
2. 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\* 发行版操作管理手册。
3. 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：HBase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。

4. 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
5. 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（高可用性）
6. 点击**下一步**。



7. 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
hadoop1	192.168.0.1	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。

8. 点击**添加节点**。



9. 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 `192.168.20.0/32` 并且每个节点的第四个字节为 `30` 和 `40` 之间，则你可在此栏输入 `192.168.20.0`。
    - 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 `192.168.20.0/32` 并且每个节点的第四个字节为 `30` 和 `40` 之间，则你可在此栏输入 `192.168.20.40`。
    - 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**。



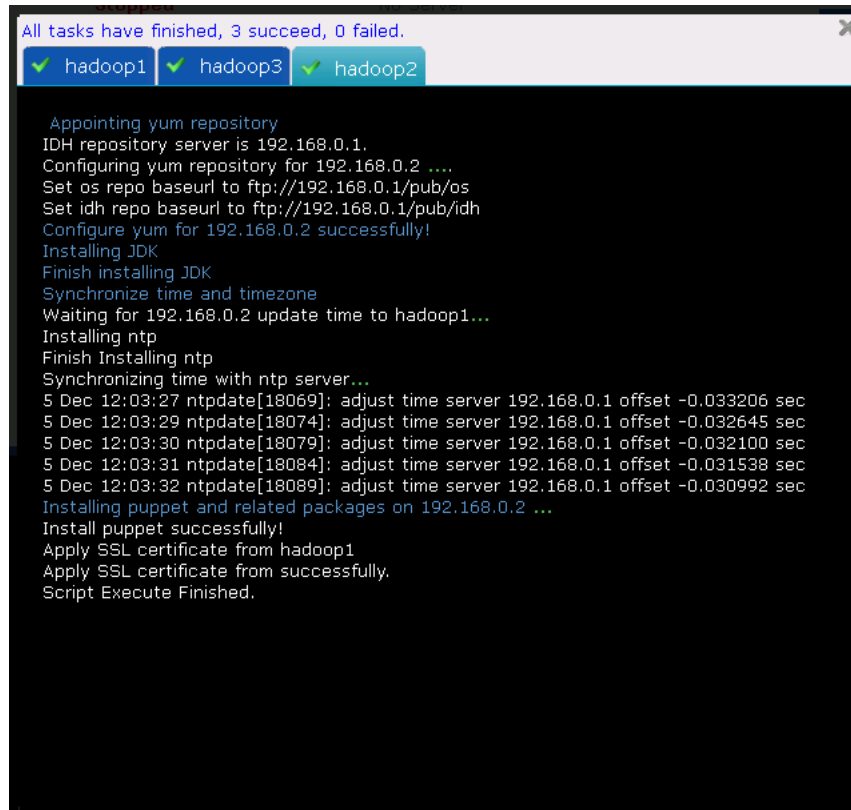
- g. 选择你想要加入到集群的节点的选项框。
- h. 点击**添加**。
- i. 当问你是否想要在节点上安装软件时，点击**确定**。
- j. 在为添加的节点配置无密码登录对话框，等待直至每个节点的状态都变为**已连接**。然后点击**完成**。
10. 如果管理节点尚未被加入到 Apache Hadoop\* 集群，请现在添加。

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。

11. 在第 2 步页面中，确认每个节点的状态为**已连接**。然后，点击**下一步**。
12. 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**。
13. 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**。
14. 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。
15. 在消息对话框，点击**确定**。

16. 在显示所有节点安装进程的页面，点击右上角的 X。



```

All tasks have finished, 3 succeed, 0 failed.
✓ hadoop1 ✓ hadoop3 ✓ hadoop2

Appointing yum repository
IDH repository server is 192.168.0.1.
Configuring yum repository for 192.168.0.2 ....
Set os repo baseurl to ftp://192.168.0.1/pub/os
Set idh repo baseurl to ftp://192.168.0.1/pub/idh
Configure yum for 192.168.0.2 successfully!
Installing JDK
Finish installing JDK
Synchronize time and timezone
Waiting for 192.168.0.2 update time to hadoop1...
Installing ntp
Finish installing ntp
Synchronizing time with ntp server...
5 Dec 12:03:27 ntpdate[18069]: adjust time server 192.168.0.1 offset -0.033206 sec
5 Dec 12:03:29 ntpdate[18074]: adjust time server 192.168.0.1 offset -0.032645 sec
5 Dec 12:03:30 ntpdate[18079]: adjust time server 192.168.0.1 offset -0.032100 sec
5 Dec 12:03:31 ntpdate[18084]: adjust time server 192.168.0.1 offset -0.031538 sec
5 Dec 12:03:32 ntpdate[18089]: adjust time server 192.168.0.1 offset -0.030992 sec
Installing puppet and related packages on 192.168.0.2 ...
Install puppet successfully!
Apply SSL certificate from hadoop1
Apply SSL certificate from successfully.
Script Execute Finished.

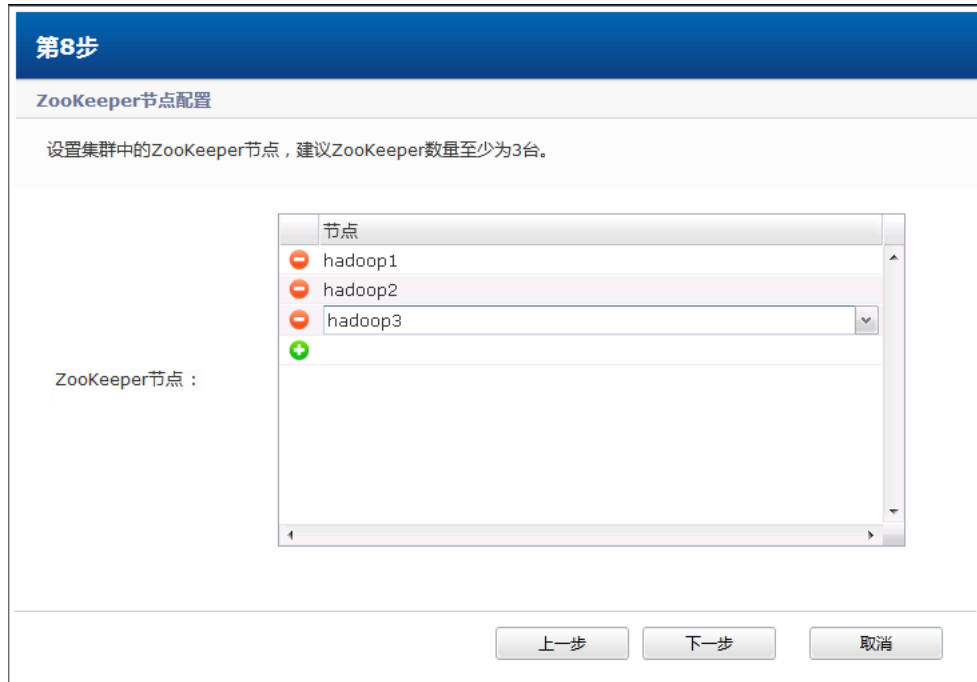
```

17. 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。

**警告：** 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。
19. 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。
20. 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。
  - 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。
  - 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。节点的数目必须为奇数。



- e. 点击下一步。

21. 在第 9 步页面，点击**下一步**。
22. 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**。
23. 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**。
24. 在第 12 步页面，点击**完成**。
25. 在确认对话框，点击**确定**。
26. 在确认对话框，点击**取消**。
27. 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**。
28. 执行章节 7.4 指定存储 HDFS 数据的路径中的步骤。
29. 执行章节 7.5 指定存储 MapReduce 数据的路径中的步骤。
30. 在节点页面，点击**节点子**页面。
31. 在节点子页面，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。

32. 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。



### 7.3 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 7.2 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 7.2 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 6.0 了解 Hadoop 服务的角色](#)。
3. 在配置菜单中，选择**节点**选项。
4. 在节点页面，选择**角色**子页面。
5. 在角色配置子页面，选择**推荐配置**链接。
6. 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。

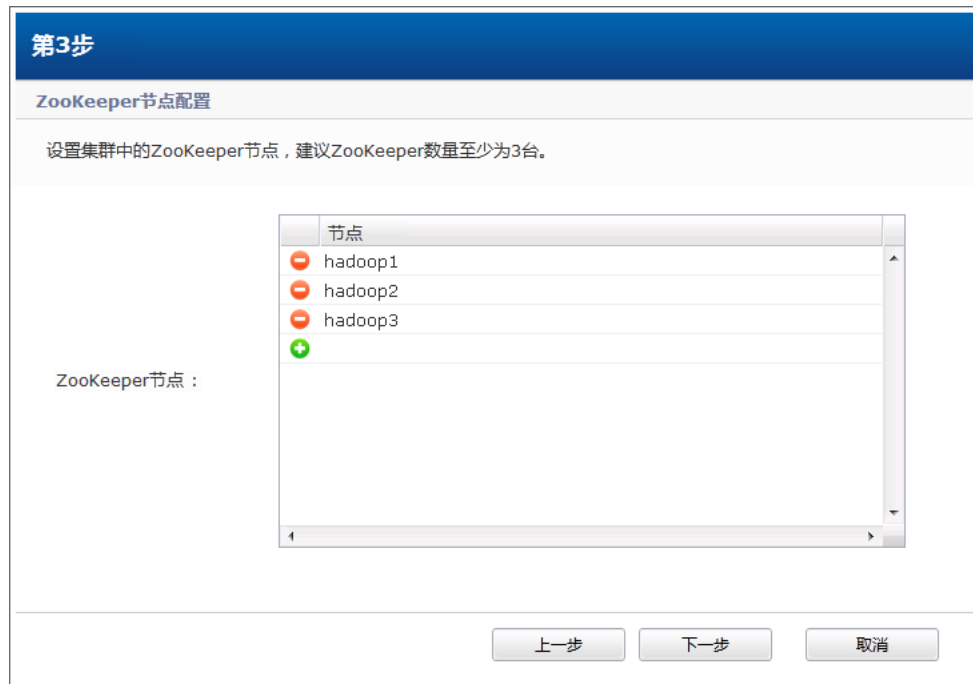
上一步

下一步

取消

7. 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**。
8. 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**。

9. 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。



节点
hadoop1
hadoop2
hadoop3

ZooKeeper节点：

上一步 下一步 取消

- 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**。
10. 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**。
11. 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**。
12. 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**。
13. 在第 7 步页面，点击**完成**。

14. 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
hadoop1	192.168.0.1	/Default	✓	✓
hadoop2	192.168.0.2	/Default		✓
hadoop3	192.168.0.3	/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。
15. 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。
16. 在角色配置页面，点击右上角的**储存**链接。

## 7.4 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 配置**选项。
4. 在 HDFS 页面，点击**全配置**子页面。

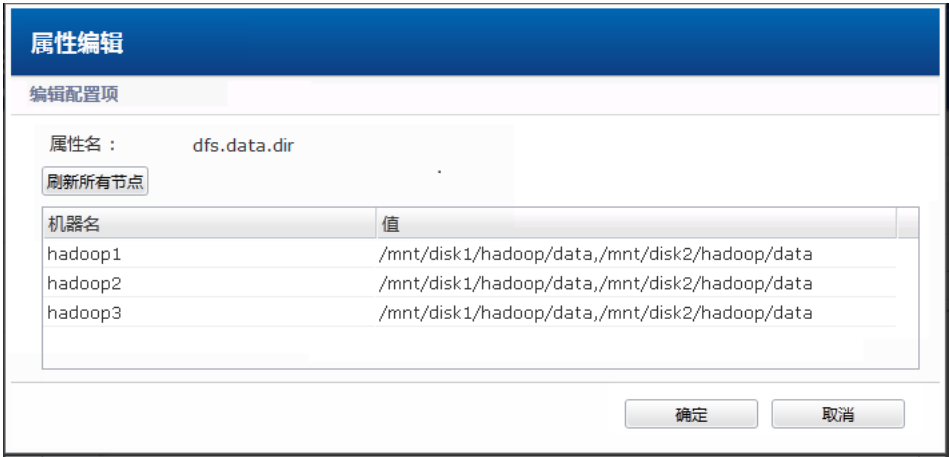




5. 在全配置子页面，在关键字过滤栏内输入 `dfs.data.dir`。



6. 在值栏内，双击 (...) 省略号。

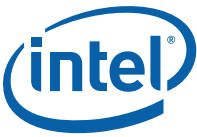


7. 为每个节点输入存储 HDFS 数据的逻辑磁盘分区列表，分区间以逗号区隔。每个路径必须以逻辑挂载名称开始，以路径 `/hadoop/data` 结尾。比如，如果一个节点有二个逻辑磁盘分区，`/mnt/disk1` 和 `/mnt/disk2`，你需要为该节点输入以下路径： `/mnt/disk1/hadoop/data`， `/mnt/disk2/hadoop/data`。

机器名	值
hadoop1	/mnt/disk1/hadoop/data,/mnt/disk2/hadoop/data
hadoop2	/mnt/disk1/hadoop/data,/mnt/disk2/hadoop/data
hadoop3	/mnt/disk1/hadoop/data,/mnt/disk2/hadoop/data

**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. 在全配置页面，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**。



7.5 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。

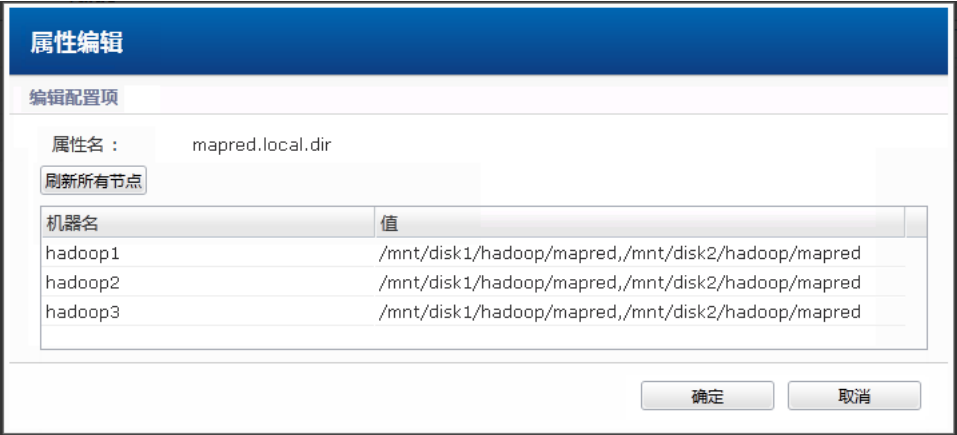
- 1. 对每个 JobTracker 和 TaskTracker 节点，确认你知道用来存储 MapReduce 数据的每个逻辑磁盘分区的挂载名称。

**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 7.4 指定存储 HDFS 数据的路径。

- 2. 展开集群配置菜单。
- 3. 在配置菜单，选择 **MapReduce** 选项。
- 4. 在 MapReduce 页面，点击**全配置**子页面。
- 5. 在全配置子页面，在**关键词过滤**栏内输入 `mapred.local.dir`。



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





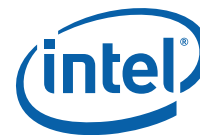
7. 为每个节点输入存储 HDFS 数据的逻辑磁盘分区列表，分区间以逗号区隔。每个路径必须以逻辑挂载名称开始，以路径 `/hadoop/mapred` 结尾。比如，如果一个节点有二个逻辑磁盘分区，`/mnt/disk1` 和 `/mnt/disk2`，你需要为该节点输入以下路径：`/mnt/disk1/hadoop/mapred`，`/mnt/disk2/hadoop/mapred`。

机器名	值
hadoop1	/mnt/disk1/hadoop/mapred,/mnt/disk2/hadoop/mapred
hadoop2	/mnt/disk1/hadoop/mapred,/mnt/disk2/hadoop/mapred
hadoop3	/mnt/disk1/hadoop/mapred,/mnt/disk2/hadoop/mapred

**警告：**

确保在逗号前后没有空格。

- 在属性编辑对话框，点击**确定**。
- 在全配置页面，点击右上角的**储存**。
- 在提交修改对话框，输入简短修改描述，然后点击**提交**。
- 在消息对话框，点击**确定**。



## 8.0 升级 Apache Hadoop\* 英特尔® Apache Hadoop\* 软件发行版

以下说明解释了如何在英特尔® 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。

### 8.1 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
  - Hive
  - ZooKeeper
  - Flume
  - Sqoop
  - Pig
  - Oozie
  - Intel® Manager
- 除了 Intel® Manager，你需要将要升级的组件停止。
- 你需要为每个要升级的组件准备最新的 RPM 软件包。这些 RPM 软件包即为组件需要升级的部分。

### 8.2 升级 Intel® Manager for Apache Hadoop\* software

要升级 Intel® Manager for Apache Hadoop\* software，执行以下步骤：

1. 以管理员身份登录 Intel® Manager。
2. 在系统配置菜单中，点击**组件升级**选项。



3. 在 Intel® Manager 子页面，点击**浏览**按钮。
4. 在 “ 打开 ” 对话框，浏览文件并双击要用于升级 Intel® Manager 的 RPM 软件包。
5. 点击**上传**。
6. 点击**升级**。

### 8.3 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**。



## 9.0 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。

### 9.1 先决条件

在卸载英特尔® 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® 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）。
- 获得 Apache Hadoop\* 集群中所有节点的 IP 地址或主机名，包括管理节点。
- *uninstall.sh* 脚本使用你提供的 IP 地址或主机名，通过 SSH 命令连接每台机器并进行卸载。这表示如果你提供主机名，管理节点必须可访问 DNS 服务器以将主机名解析为 IP 地址。
- 确认你有管理节点的 root 访问权限。
- 管理节点必须和集群中的其他节点分开卸载，而且必须是最后一个被卸载的节点。
- 确认你有集群中每个节点的 root 密码。
- 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\* 发行版操作管理手册。

### 9.2 卸载英特尔® Apache Hadoop\* 软件发行版

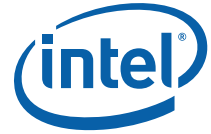
参考以下情形：

- 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：*hadoop1*，*hadoop2*，和 *hadoop3*。
- *hadoop1* 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® 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. 输入 `sh uninstall.sh`, 然后输入除了管理节点外的每个节点的主机名, 名称用空格间隔。在以上情形中, 你需要输入以下命令:

```
sh uninstall.sh hadoop2 hadoop3
```

**注意:** 你不能将管理节点的主机名或地址列入你想要进行卸载的节点列表中。

4. 按下 Enter。

5. 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。

6. 当卸载完成后, 为管理节点执行 `uninstall.sh` 脚本。在以上情形中, 你需要输入以下命令:

```
sh uninstall.sh hadoop1
```

### 9.3 卸载示例

```
hadoop1:~ # sh uninstall.sh hadoop2 hadoop3
```

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

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.
```