



---

# Intel® High Availability 操作手册

1. 简介.....	3
1.1. 文档目的.....	3
1.2. 软件要求.....	3
1.3. 名词解释: .....	4
2. 集群配置.....	4
2.1. 配置高可用性集群.....	4
2.2. 高可用性配置.....	7
2.3. 集群许可证更新.....	9
2.4. 配置更新节点.....	11
2.5. 集群或单一组件的启动停止.....	12
3. 高可用性维护.....	13
3.1. 故障处理后的主节点恢复.....	13
3.2. 主从节点的同步恢复.....	14

## 1. 简介

---

### 1.1. 文档目的

本文档用于描述如何配置，检查和维护 Intel® Expressway 分布式系统套件的高可用性。

### 1.2. 软件要求

集群中所有机器都需要安装 Intel® Expressway 分布式系统套件。并且确认集群组件至少满足以下最低配置要求：

**HDFS 最低配置：**

一台 Namenode，

一台或一台以上 Secondary Namenode，

一台 Standby Namenode，

至少大于文件备份数目的 Datanode；

**MapReduce 最低配置：**

一台 Jobtracker，

一台 Backup Jobtracker，

至少一台 Tasktracker；

**HBase 最低配置：**

至少一台 HBase Master，

三台或三台以上（需要为奇数）数量的 Zookeeper，

至少一台 HBase Region Server，

如需 Thrift 功能，则还需要配置至少一台的 HBase Thrift 服务器；

Hive 最低配置:

至少一台 Hive Server;

管理节点配置:

确保该节点设置为 Ganglia Master,

安装有 Intel® Software Cluster Management Suite 套件。

### 1.3. 名词解释:

管理节点: 安装有 Ganglia Master 以及 Intel® Software Cluster Management Suite 套件, 并确保 <http://【管理节点】/ganglia> 以及 <https://【管理节点】:9443> 能够访问。

Primary 节点: 作为集群 Namenode 和 Jobtracker 的服务器, DRBD 系统中的主节点。

Standby 节点: 作为 Primary 节点的备份节点, DRBD 系统中的从节点。

## 2. 集群配置

### 2.1. 配置高可用性集群

首先需要对集群进行配置, 打开 <https://【管理节点】:9443>, 登陆管理界面后, 点击右上方的配置向导:



注意, 在配置前, 在左侧的下拉列表中选择“控制面板”选项, 请确保集群处于关闭状态, 如图所示:



第一步，设置集群的名称，并且搜索确认集群的节点，选择安装集群组件，这里采用默认配置，然后点击下一步：

第1步

配置新的集群

集群名称：

ClusterStability

请设置集群的名称。

集群DNS服务器：

☐ 集群中包含DNS服务器

如果集群中的节点可以通过已经配置好的DNS服务器使用域名互相访问，请勾选此选项；否则请不要选择此选项，管理软件将自动更新/etc/hosts文件实现域名访问。

集群节点：

节点名称	节点IP
ditest02	192.168.1.107
xmlqa-clv9	192.168.1.106
xmlcore-perf2	192.168.1.104

搜索节点

集群组件：

☒ HDFS：集群中将会安装使用HDFS文件系统。
 ☒ MapReduce：集群中将会部署和运行MapReduce任务。
 ☒ HBase：集群中将会安装使用HBase。
 ☒ Hive：集群中将会安装使用Hive。

下一步

取消

第二步，输入 Namenode 名称，并且输入 Namenode 的虚拟 IP 地址；输入 Jobtracker 的名称，并且输入 Jobtracker 的虚拟 IP 地址，然后点击下一步：

## 第2步

### 集群ClusterStability的Hadoop基本配置

一个HDFS集群包含一个**Namenode**，它是一个中心服务器，负责管理文件系统命名空间和客户端对文件的访问。Namenode负责管理执行文件系统命名空间的操作，包括打开、关闭和重命名文件和目录，而且它决定了数据块到具体Datanode的映射。请在下面输入Namenode的地址，**如果需要配置高可用性，请输入虚拟IP地址。**

Namenode :

Namenode IP地址 :

MapReduce框架包含了一个**Jobtracker**，它是一个中心服务器，负责调度组成工作的部分任务，将这些部分任务分配到不同的Tasktracker让其执行。请在下面输入Jobtracker的地址，**如果需要配置高可用性，请输入虚拟IP地址。**

Jobtracker :

Jobtracker IP地址 :

上一步

下一步

取消

第三步，输入文件块复制数和文件块大小，这里采用默认配置，然后点击下一步：

## 第3步

### 集群ClusterStability的HDFS配置

HDFS通过保存多个文件块的副本来保证HDFS的高可靠性和容灾能力。您可以根据自身的业务需求以及环境设置和调整，**默认的文件块复制数为3。**

文件块复制数 :

HDFS中每个文件都被分成一系列的文件块存储在不同的Datanode中，其中除了最后一个文件块，其他的文件块都是一样的。您可以根据情况设置文件块的大小，**默认大小为67108864字节。**

文件块大小(字节) :

上一步

下一步

取消

第四步，输入 MapReduce 集群中 Tasktracker 的估算数量，然后点击下一步：

第4步

集群ClusterStability的MapReduce配置

您已经选择的Jobtracker: **ClusterStabilityHA**。

请输入MapReduce集群中Tasktracker的估算数量。

Tasktracker的数量:

上一步

下一步

取消

第五步，点击完成：

第5步

完成

集群安装配置已经完成，点击 **完成** 按钮完成配置。您可以在配置页面进行更详细的配置。

点击 **上一步** 按钮重新进行配置，点击 **取消** 按钮取消本次集群的安装配置。

上一步

完成

取消

## 2.2. 高可用性配置

点击左侧工具栏“集群节点”菜单项，右边点击右上方的角色配置，选择要配置的组件为 HDFS 配置。选择 1 个节点作为 Primary Namenode，1 个节点作为 Standby

Namenode。这里选择了 xmlqa-clv8 作为 Primary Namenode，xmlqa-clv9 作为 Standby Namenode，如图所示：



节点	IP地址	机架	Primary Namenode	DataNode	Secondary Namenode	Standby Namenode
barcelona	192.168.1.102	Default		✓		
ditest02	192.168.1.107	Default		✓	✓	
xmlcore-perf2	192.168.1.104	Default		✓		
xmlqa-clv8	192.168.1.103	Default	✓			
xmlqa-clv9	192.168.1.106	Default				✓
xmlqa-hpt2	192.168.1.101	Default		✓		
xrt-portal	192.168.1.100	Default		✓		

选择要配置的组件为 MapReduce 设置, 选择 JobTracker 节点和 Backup JobTracker 节点。这里选择 xmlqa-clv8 作为 JobTracker，xmlqa-clv9 作为 Backup JobTracker：



节点	IP地址	机架	JobTracker	Backup Job Tracker	TaskTracker
barcelona	192.168.1.102	Default			✓
ditest02	192.168.1.107	Default			✓
xmlcore-perf2	192.168.1.104	Default			✓
xmlqa-clv8	192.168.1.103	Default	✓		
xmlqa-clv9	192.168.1.106	Default		✓	
xmlqa-hpt2	192.168.1.101	Default			✓
xrt-portal	192.168.1.100	Default			✓

点击左侧菜单栏“集群配置”目录下的 Hadoop，右侧的高可用性配置里，可以设置 VRRP VID 和 VRRP 网络端口，VRRP VID 范围是 0-255：





## 2.3. 集群许可证更新

打开【管理节点】管理页面，地址是：<https://【管理节点】:9443>。

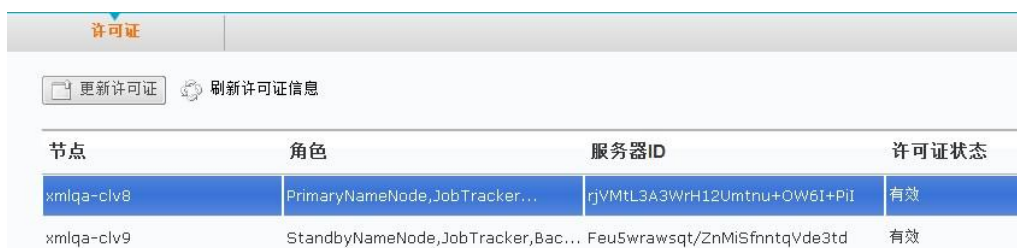
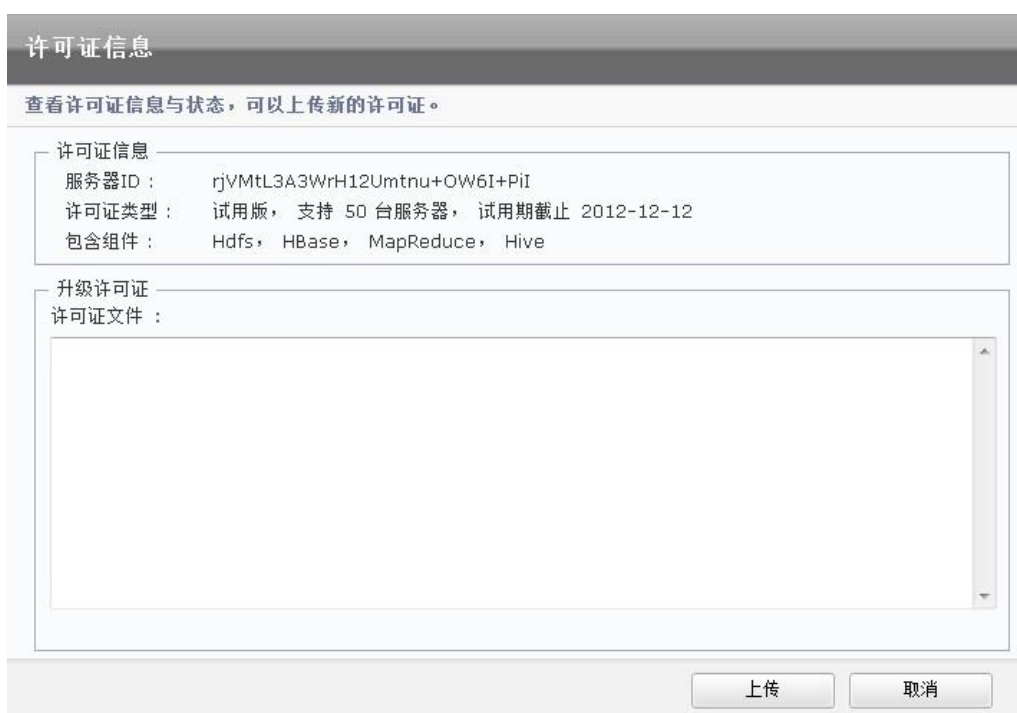
在配置更新前，在左侧的下拉列表中选择“控制面板”选项，请确保集群处于关闭状态，如图所示：



然后点击左侧工具栏“许可证管理”项，进入许可证管理界面。



许可证管理界面中列出了所有需要许可证的服务器，双击选择需要更新的服务器，在弹出的窗口中输入新的许可证。



输入许可证字符串后，点击“上传”按钮，上传成功后在页面上方的工具栏中选择“更新许可证”项，在集群中部署新的许可证信息。

## 2.4. 配置更新节点

打开【管理节点】管理页面，地址是：<https://【管理节点】:9443>。

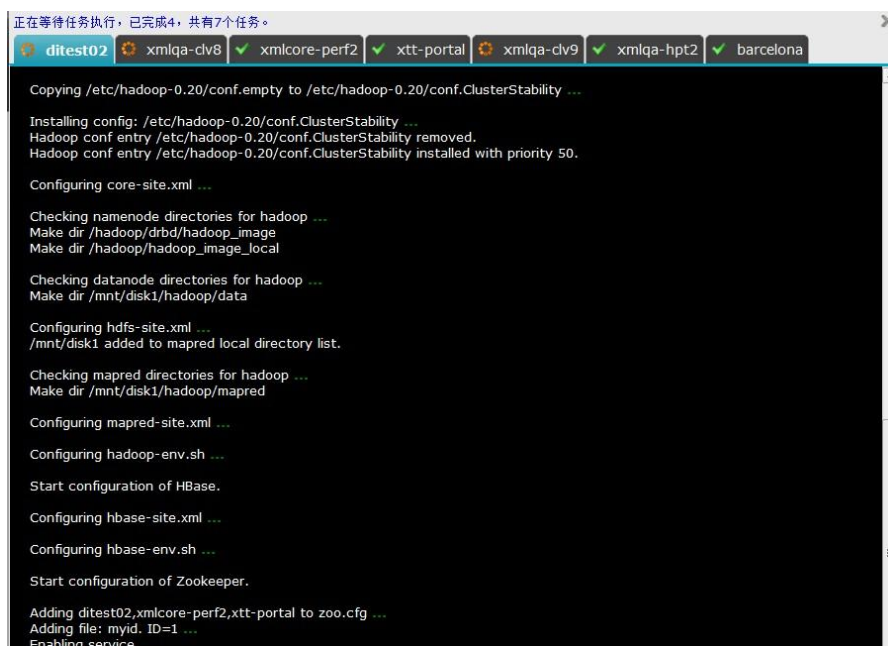
在配置更新前，请确保集群所有服务处于“未运行”状态，如图所示：



然后点击左侧工具栏“集群节点”项，进入集群节点配置界面。



在节点配置上方工具栏中选择“配置所有节点”项，点击后，集群开始重新配置。

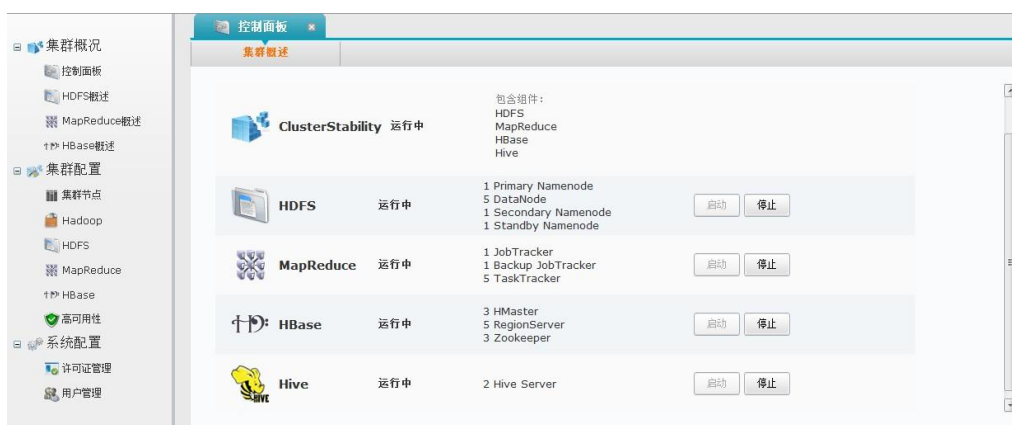


如图所示的配置界面，等到所有等待标志  变成成功标志  后，集群所有配置更新完毕。

如果配置 High Availability，根据 DRBD 分区大小不同等待时间会比较长，可能会在 30 分钟到 1 小时左右。

## 2.5. 集群或单一组件的启动停止

打开 <https://【管理节点】:9443>，在左侧的下拉列表中选择“控制面板”选项，打开控制面板页面：



如图所示，在运行中的四样组件分别为 HDFS, MapReduce, HBase 和 Hive。除 HDFS 外，所有组件都只有两种状态“运行中”和“未运行”，在界面的最右侧有操作列表，可以通过点击按钮来对集群单一组件进行“启动/停止”操作。

如需启动集群，需要严格按照启动顺序：HDFS，MapReduce，HBase，Hive。

单一组件的启动顺序如下：

HDFS 不需要依赖另外组件；

MapReduce 启动之前，需要确保 HDFS 处于运行状态下；

HBase 启动之前，需要确保 HDFS 处于运行状态下；

Hive 启动之前，需要确保 HDFS，MapReduce 以及 HBase 处于运行状态下。

如需停止集群，需要严格按照停止顺序：Hive，HBase，MapReduce，HDFS。

单一组件的停止顺序如下：

HDFS 停止之前，需要确保 MapReduce，HBase，Hive 处于未运行状态下；

MapReduce 停止之前，需要确保 Hive 处于未运行状态下；

HBase 停止之前，需要确保 Hive 处于未运行状态下；

Hive 可以在没有任务执行时直接停止。

## 3. 高可用性维护

---

### 3.1. 故障处理后的主节点恢复

处理故障后，主节点可能发生变化，如果希望恢复原来的 MASTER 节点，需要手动修改：

找到现在的 MASTER 节点，断开 MASTER 节点的网络，在 MASTER 节点上输入命令：

```
service keepalived stop
```

```
vim /etc/keepalived/ keepalived.conf
```

```
global_defs {
    router_id xmlqa-clv8
}

vrrp_instance VI_1 {
    state MASTER
    interface eth0
    virtual_router_id 77
    priority 50
    advert_int 1
    virtual_ipaddress {
        192.168.1.108 dev eth0
    }
    nopreempt
    notify /etc/keepalived/notify.sh
}
```

将配置文件中的 state MASTER 改为 state BACKUP

将节点连上网络，在节点上重新启动 keepalived 服务：

```
service keepalived start
```

### 3.2. 主从节点的同步恢复

如果在/var/log/messages 中出现类似信息：

```
Split-Brain detected, dropping connection!
```

则发生脑裂故障，Primary 节点和 Standby 节点两个节点间数据不再同步。

检查，在 Standby 节点机器上命令行输入命令：

```
service drbd status
```

如果出现以下信息，则为正常状态：

```
m:res cs          ro          ds          p mounted  fstype
0:r0 Connected Secondary/ Primary UpToDate/ UpToDate C
```

Connected 表示已连接状态；Secondary/Primary 表示本节点为 Secondary 节点，连接到 Primary 节点。

如果出现类似信息，出现故障：

```
m:res cs          ro          ds          p mounted  fstype
0:r0 WFconnection Secondary/ Unknown UpToDate/ DUnknown C
```



WFConnection 表示等待连接状态;Secondary/Primary 表示本节点为 Secondary, 连接未知。

去 Primary 节点检查, 在 Primary 节点机器上命令行输入命令:

```
service drbd status
```

如果出现以下信息, 则为正常状态:

```
m:res cs          ro          ds          p mounted  fstype
0:r0 Connected Primary/Secondary UpToDate/ UpToDate C  hadoop/hrbd ext3
```

Connected 表示已连接状态; Primary/Secondary 表示本节点为 Primary 节点, 连接到 Secondary 节点。

如果出现类似信息, 显示处于 StandAlone 状态, 出现故障:

```
m:res cs          ro          ds          p mounted  fstype
0:r0 StandAlone Primary/Unknown UpToDate/DUnknown C  hadoop/hrbd ext3
```

StandAlone 表示独立运行状态; Primary/Unknown 表示本节点为 Primary 节点, 连接未知。

解决方法, 删除不一致数据, 在 Standby 节点机器上命令行输入:

```
drbdadm disconnect r0
```

```
drbdadm secondary r0
```

```
drbdadm -- --discard-my-data r0
```

在 Primary 节点上重连接资源, 在 Primary 节点机器命令行输入:

```
drbdadm connect r0
```

在 Standby 节点上再次启动 drbd 服务:

```
service drbd start
```

再次查看 drbd 服务, 查看状态是否正常:

```
service drbd status
```