磁盘设备多路径原理分析

|  |  |  |  |
| --- | --- | --- | --- |
| 日期 | 版本 | 修订者 | 修订描述 |
| 2017.11.07 | v1 | 林 清 | 初稿 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

# 前言

目前ISCSI的LUN多路径应用场景如下：

通过iscsiadm对一个LUN在一个target和一个initiator创建多个链路。每个链路可以表示成“target\_IP+initiator\_iface”。iscsiadm创建的每个链路在initiator看来都是一个独立设备。iscsi协议或者iscsi工具本身不能将这些设备融合成一个可用的设备。

多​路​径​设​备​映​射​器​（DM-Multipath） 可以​将​服​务​器​节​点​和​存​储​阵​列​间​的​多​个​ I/O 路​径​配​置​为​一​个​单​一​设​备。

DM-Multipath使多路径设备有以下能力：

冗​余：

DM-Multipath 可​在​主​动​/被​动模式下工作​。在​主​动​/被​动​配​置​中​,只​有​一​半​的​路​径​在​每​次​ I/O 时​使​用​。如​果I/O 路​径​的​任​意​元​素出​现​故​障​，就​会​将​ DM-Multipath 切​换​到​备​用​路​径​。

提​高​的​性​能​：

DM-Multipath可​配​置​为​主​动​/主​动​模​式，其​中​将​ I/O 以​调​度​算​法​方​式​分​布​到​所​有​路​径​中​。DM-Multipath 可​在​ I/O 路​径​中​检​测​负​载​并​动​态​对​负​载​进​行​再​平​衡​。

目前OpenStack中Nova使用多路径的策略有一定限制，多路径提供的冗余、性能提升两个特性都无法完全发挥。主要表现在：Nova只能在宿主机的一张网卡上建立ISCSI链路，当然这可以通过修改代码来解决。

# multipath工具相关命令

## 查看多路径设备的链路状态

multipath –l <WWN>

## 删除多路径设备

#删除所有多路径设备

multipath –F

#删除某个多路设备

multipath –f <wwn>

## 打印所有多路径设备

multipathd show maps

## 重新映射多路径设备

multipathd reload map <map\_name/wwn>

## 切换设备的活动组

multipathd switch map <map\_name> group <group\_id>

map\_name指的是设备wwn或者别名，group\_id指的是group的id从编号1开始。

## 打印当前multipath配置

multipathd show config

# 配置DM-Multipath

配置文件位置/etc/multipath.conf（配置文件模板：/usr/share/doc/device-mapper-multipath-0.4.9/multipath.conf），相关信息可以参看man multipath.conf：

多​路​径​配​置​文​件​可​分​为​以​下​几​个​部​分​：

|  |  |  |
| --- | --- | --- |
| ***配置*** | ***说明*** | ***备注*** |
| **blacklist** | 黑名单，multipath不会将其中的设备映射成多路径设备 |  |
| **blacklist\_exceptions** | 黑名单例外，出现在其中的值即使配置为黑名单，也会映射成多路径设备 |  |
| **defaults** | 所有多路径设备的全局配置 | 覆盖默认值 |
| **multipaths** | 某个多路径设备的配置 | 覆盖defaults、devices |
| **devices** | 某种存储设备（如NetApp、华为）的多路径配置 | 覆盖defaults |

多路径设备的几个关键配置项：

## path\_grouping\_policy

**这个配置项决定如何对链路进行分组。multipath工具可以将链路分成若干组，一个多路径设备往往只有一个组在工作。常用的配置有下面两种：**

**multibus模式下，所以链路都在一个组，这时候所有链路都工作。此时设备工作在Active/Active状态下。**

failover模式下，一个路径一个组，一旦出现读写问题，立刻切换。这是multipath的主备工作模式。

## path\_selector

路径调度器配置，及当有多个工作路径时，IO在这些路径中如何分配。

round-robin 0：每个路径平均分配IO；

queue-length 0：每个IO都分配给负载最小的链路；

service-time 0：根据链路的吞吐量分配链路；

## failback

如果链路出现故障如何切换分组。

immediate：一旦工作链路出现问题，立即切换链路；

manual：手工切换；

followover：一旦优先级最高的组可用，立刻切换到该组；

values > 0：延时切换，延时时间可以指定如values > 5，表示延时5s

## 查看多路径设备的链路状态

multipath –l <WWN>

## 删除多路径设备

#删除所有多路径设备

multipath –F

#删除某个多路设备

multipath –f <wwn>

## 打印所有多路径设备

multipathd show maps

## 重新映射多路径设备

multipathd reload map <map\_name/wwn>

## 切换设备的活动组

multipathd switch map <map\_name> group <group\_id>

map\_name指的是设备wwn或者别名，group\_id指的是group的id从编号1开始。

## 打印当前multipath配置

multipathd show config

# 配置DM-Multipath

配置文件位置/etc/multipath.conf（配置文件模板：/usr/share/doc/device-mapper-multipath-0.4.9/multipath.conf），相关信息可以参看man multipath.conf：

多​路​径​配​置​文​件​可​分​为​以​下​几​个​部​分​：

|  |  |  |
| --- | --- | --- |
| ***配置*** | ***说明*** | ***备注*** |
| **blacklist** | 黑名单，multipath不会将其中的设备映射成多路径设备 |  |
| **blacklist\_exceptions** | 黑名单例外，出现在其中的值即使配置为黑名单，也会映射成多路径设备 |  |
| **defaults** | 所有多路径设备的全局配置 | 覆盖默认值 |
| **multipaths** | 某个多路径设备的配置 | 覆盖defaults、devices |
| **devices** | 某种存储设备（如NetApp、华为）的多路径配置 | 覆盖defaults |

多路径设备的几个关键配置项：

## path\_grouping\_policy

**这个配置项决定如何对链路进行分组。multipath工具可以将链路分成若干组，一个多路径设备往往只有一个组在工作。常用的配置有下面两种：**

**multibus模式下，所以链路都在一个组，这时候所有链路都工作。此时设备工作在Active/Active状态下。**

failover模式下，一个路径一个组，一旦出现读写问题，立刻切换。这是multipath的主备工作模式。

## path\_selector

路径调度器配置，及当有多个工作路径时，IO在这些路径中如何分配。

round-robin 0：每个路径平均分配IO；

queue-length 0：每个IO都分配给负载最小的链路；

service-time 0：根据链路的吞吐量分配链路；

## failback

如果链路出现故障如何切换分组。

immediate：一旦工作链路出现问题，立即切换链路；

manual：手工切换；

followover：一旦优先级最高的组可用，立刻切换到该组；

values > 0：延时切换，延时时间可以指定如values > 5，表示延时5s