KYLINSOFT
麒麟软件

pacemaker + corosync 与容器技术的结合实践

打造中国操作系统核心力量

kylinos.cn

邹志敏 2022.04.14

目 录

Bundle介绍

KYLINSOFT 麒麟软件

什么是bundle?

- pacemaker 用于支持使用任何所需的基础架构来启动容器的一种特殊语法
- bundle用于管理容器镜像的多个实例，以及容器所需的网络和存储

Pacemaker参数介绍

KYLINSOFT 麒麟软件

bundles——隔离环境

pacemaker bundle支持Docker和rkt容器技术。

例：一个用于容器化web服务器的包

```
<bundle id="httpd-bundle">
   <docker image="pcmk:http" replicas="3"/>
   <network ip-range-start="192.168.122.131"
            host-netmask="24"
            host-interface="eth0">
      <port-mapping id="httpd-port" port="80"/>
   </network>
   <storage>
      <storage-mapping id="httpd-syslog"
                       source-dir="/dev/log"
                       target-dir="/dev/log"
                       options="rw"/>
      <storage-mapping id="httpd-root"
                       source-dir="/srv/html"
                       target-dir="/var/www/html"
                       options="rw"/>
      <storage-mapping id="httpd-logs"
                       source-dir-root="/var/log/pacemaker/bundles"
                       target-dir="/etc/httpd/logs"
                       options="rw"/>
   </storage>
   <primitive class="ocf" id="httpd" provider="heartbeat" type="apache"/>
</bundle>
```

Bundle属性

一个bundle资源必须包含一个<docker>或<rkt>元素

字段	描述
id	Bundle名字（必要）
description	任意描述文本（非必要）

命令示例：pcs resource bundle create **bundle_id** container **docker** [<container options>] [network <network options>] [port-map <port options>]... [storage-map <storage options>]... [meta <meta options>] [--disabled] [--wait[=n]]

注意：在 Pacemaker 中配置 Docker bundle 前，必须安装 Docker，并在允许运行Bundle的每个节点上提供完好配置的 Docker 镜像

docker容器属性

KYLINSOFT 麒麟软件

Pacemaker将创建一个隐式的ocf:heartbeat:docker资源来管理一个bundle的Docker容器。

用户必须确保resource agent安装在允许运行bundle的每个节点上。

字段	默认值	描述
image		Docker 镜像tag（必需）
replicas	promoted-max or 1	要启动的容器实例数
replicas-per-host	1	指定允许在一个节点上运行的容器实例数
promoted-max	0	非负整数，如果为正则表示容器化服务应被视为多状态服务，且此replicas允许在 master 角色中运行该服务
network		执行**docker run** 命令，作为 Docker 容器的网络设置
run-command	如果包含primitive资源，则/usr/sbin/pacemaker_remoted	启动后将在容器内运行（“PID 1”），若bundle包含资源，此命令必须启动 pacemaker_remoted 守护进程
options		传递给 **docker run** 命令的额外命令行选项

replicas通过bundle ID 加上破折号和以零开头的整数计数命名。例如如果名为 httpd-bundle 的资源配置了 replicas=2，则其容器将命名为 httpd-bundle-0 和 httpd-bundle-1

Bundle Network 属性

KYLINSOFT 麒麟软件

一个bundle可以选择包含一个<network>元素

字段	默认值	描述
add-host	TRUE	如果使用TRUE和ip-range-start，Pacemaker将自动确保容器内的/etc/hosts中有每个replicas名及其分配的IP条目。
ip-range-start		从这个IP地址开始，使用指定为 Docker的replicas的连续地址
host-netmask	32	若指定了 ip-range-start，则为该ip子网掩码
host-interface		若指定了 ip-range-start，在此主机interface上创建 IP 地址
control-port	3121	如果bundle包含一个primitive，集群将使用这个TCP端口与容器内的Pacemaker Remote通信。当容器无法侦听默认端口时，例如容器使用主机的网络而不是ip-range-start(在这种情况下，replicas-per-host必须1)，或者当包可能运行在一个pacemaker远程节点已经监听默认端口时可以修改默认值。

port-mapping

<network>元素可以选择包含一个或多个<port-mapping>元素

字段	默认值	描述
id		端口映射的唯一名称(必需的)
port		如果指定，则主机网络上此 TCP 端口号的连接(如果指定了ip-range-start，则为容器分配的IP地址)将转发到容器网络。在端口映射中必须指定一个端口或范围
internal-port	port的值	如果指定了port和internal-port，则到主机网络上的端口的连接将转发到容器网络上的此端口
range		如果指定了这个参数，连接到主机网络(如果指定了ip-range-start，则为容器分配的IP地址)上的这些TCP端口号(表示为first_port-last_port)的连接将被转发到容器网络中的相同端口

如果bundle包含资源，Pacemaker 将自动映射 control-port，因此不需要在端口映射中指定该端口。

Bundle Storage 属性

一个bundle可以有选择地包含一个<storage>元素,storage本身没有属性，但可以包含一个或多个存储映射(bundle-order-partial.xml)

字段	默认值	描述
id		存储映射的唯一名称(必需的)
source-dir		将映射到容器中的主机文件系统的绝对路径。在存储映射中必须指定source-dir和source-dir-root中的一个
source-dir-root		主机文件系统中一个路径的开头，该路径将被映射到容器中，为每个容器实例使用主机上不同的子目录。子目录的命名将与replicas名称相同。
target-dir		映射主机存储的容器内的路径名称（必需）
options		映射存储时使用的文件系统挂载选项

如何使用 source-dir-root 参数命名主机上的子目录?
如果 source-dir-root=/path/to/my/directory, target-dir=/srv/appdata，bundle id为mybundle 且 replicas=2，集群将创建两个容器主机名为 mybundle-0和mybundle-1的实例，并在运行容器的主机上创建两个目录：/path/to/my/directory/mybundle-0 和 /path/to/my/directory/mybundle-1。每个容器将获得其中一个目录，容器内运行的任何应用程序都将该目录视为源目录

Bundle Storage 属性

KYLINSOFT 麒麟软件

注意:

- 如果主机上还没有source目录，Pacemaker 不会定义，但是容器或其资源代理应该会创建source目录。
- 如果bundle包含一个primitive（普通资源），Pacemaker会自动将
  source-dir=/etc/pacemaker/authkey
  target-dir=/etc/pacemaker/authkey
  source-dir-root=/var/log/pacemaker/bundles
  target-dir=/var/log映射到容器中，所以在存储映射中不需要指定这些路径
- PCMK_authkey_location环境变量只能设置为集群中任何节点的/etc/pacemaker/authkey的默认值

Bundle Primitive

KYLINSOFT 麒麟软件

- bundle可以选择性地包含一个 Pacemaker 集群资源。资源可以像往常一样定义操作属性、实例属性和元属性。

- 如果bundle包含资源，**容器镜像必须包含** Pacemaker Remote 守护进程，并且必须在bundle中配置 ip-range -start 或 control-port。Pacemaker 会为连接创建一个隐式 **ocf:pacemaker:remote** 资源，在容器内启动 Pacemaker Remote，并通过 Pacemaker Remote 监控和管理资源。如果bundle有多个容器实例（replicas），Pacemaker 资源将充当隐式克隆，如果bundle的master>0将设置为promotable clone

- 包含资源的bundle中的容器必须具有可访问的网络环境，以便集群节点上的 Pacemaker 可以与容器内的 Pacemaker 远程联系。 Docker选项--net=none**不应该与**primitive共同使用。如果使用 docker 选项 --net=host（使容器共享主机的网络空间），则应为每个bundle指定一个唯一的 control-port 参数。任何防火墙都必须允许访问 control-port。

Bundle Node 属性

如果bundle有一个普通资源，普通资源的资源代理可能希望设置节点属性，比如promotion scores (**Promotable Clone限制资源启动优先级**)。

- 容器使用的共享存储，设置bundle节点的promotion score；
- 容器使用本地存储，设置集群节点的promotion score。

相当于定义约束规则限制资源的启动位置

container 资源的meta-attribute属性允许用户指定要使用的方法。如果设置为 host，则在集群节点上使用用户定义的节点属性；否则使用bundle节点的属性。这个行为只适用于用户定义的属性，集群总是检查本地节点是否有集群定义的属性，例如集群名称uname。

Bundle Meta-Attributes

bundle上设置的任何元数据属性都将由bundle中包含的资源继承，以及 Pacemaker 为bundle创建的任何资源。这包括 priority、target-role 和 is-managed 等选项。

集群使用元属性来决定资源应该如何运行，可以使用crm_resource命令设置元属性。

Limitations of Bundles

Pacemaker bundle在以下限制下运行：

- 当bundle处于非管理状态或集群处于维护模式时重新启动起pacemaker可能会导致bundle失败。
- bundle不能被克隆或包含在组中。这包括bundle的primitive和Pacemaker为bundle隐式创建的任何资源。
- bundle没有实例属性或者操作属性，尽管bundle的primitive可能有这些属性。
- 只有bundle使用不同的 control-port 时，包含普通资源的bundle才能在Pacemaker Remote节点上运行。

配置实例

在使用集群的bundle功能时，集群节点需要有容器相关软件包及容器中的镜像。

配置实例以docker为例，镜像名称为： pcmktest:http2

```
[root@node2 ~]# docker images
REPOSITORY          TAG                 IMAGE ID            CREATED             SIZE
httpd               latest              1132a4fc88fa        5 months ago        143MB
pcmktest            http2               6de12f4bfa53        9 months ago        672MB
```

bundle绑定普通资源httpd，资源类型为ocf:heartbeat:apache。

Bundle资源中的端口及存储映射均为此资源运行而设置。

在pcmktest:http2镜像中需要具备的软件包如下：

httpd bind-utils lsof wget resource-agents openssh-clients pacemaker pacemaker-remote

配置实例

KYLINSOFT 麒麟软件

PCS命令创建bundle

pcs resource bundle create httpd-bundle container docker image="pcmktest:http2" replicas="2" run-command="/usr/sbin/pacemaker_remoted"

network ip-range-start="172.17.127.185" host-interface="ens33" host-netmask="24" port-map id="httpd-port" port="80"

storage-map id="httpd-root" source-dir-root="/var/local/containers" target-dir="/var/www/html" options="rw" storage-map id="httpd-logs" source-dir-root="/var/log/pacemaker/bundles" target-dir="/etc/httpd/logs" options="rw

注：实际上是一条命令，为了展示方便分开写

replicas表示的是需要启动的容器数量，ip-range-start即表示有序ip的起始位置，例如本次设置中将会占用172.17.127.185-186两个ip，容器、镜像名、网卡掩码请按照实际情况进行设置

配置实例

KYLINSOFT 麒麟软件

查看httpd-bundle资源参数：pcs resource config httpd-bundle

```
[root@node1 ~]# pcs resource config httpd-bundle
 Bundle: httpd-bundle
  Docker: image=pcmktest:http2 replicas=2 run-command=/usr/sbin/pacemaker_remoted
  Network: host-interface=ens33 host-netmask=24 ip-range-start=172.17.127.185
  Port Mapping:
   port=80 (httpd-port)
  Storage Mapping:
   options=rw source-dir-root=/var/local/containers target-dir=/var/www/html (httpd-root)
   options=rw source-dir-root=/var/log/pacemaker/bundles target-dir=/etc/httpd/logs (httpd-logs)
```

配置实例

KYLINSOFT 麒麟软件

update更新bundle配置

如果需要修改某一项参数：
pcs resource bundle update <bundle id> [container <container options>] [network <network options>] [port-map (add <port options>) | (delete | remove <id>...)]... [storage-map (add <storage options>) | (delete | remove <id>...)]... [meta <meta options>][--wait[=n]]

修改容器数量：pcs resource bundle update httpd-bundle container replicas="1"

添加或者删除端口port-map和存储映射 storage-map注意在关键字后加入add或者delete

- pcs resource bundle update httpd-bundle port-map delete httpd-port

- pcs resource bundle update httpd-bundle port-map add id="httpd-port" port="81"

配置实例

KYLINSOFT 麒麟软件

pcs status

```
Full List of Resources:
  * Container bundle set: httpd-bundle [pcmktest:http2]:
    * httpd-bundle-docker-0 (172.17.127.185)    (ocf::heartbeat:docker):         Started node2
    * httpd-bundle-docker-1 (172.17.127.186)    (ocf::heartbeat:docker):         Started node1
```

pcs resource config

```
[root@node1 ~]# pcs resource config
 Bundle: httpd-bundle
  Docker: image=pcmktest:http2 replicas=2 run-command=/usr/sbin/pacemaker_remoted
  Network: host-interface=ens33 host-netmask=24 ip-range-start=172.17.127.185
  Port Mapping:
   port=81 (httpd-port)
  Storage Mapping:
   options=rw source-dir-root=/var/local/containers target-dir=/var/www/html (httpd-root)
   options=rw source-dir-root=/var/log/pacemaker/bundles target-dir=/etc/httpd/logs (httpd-logs)
 Resource: apache (class=ocf provider=heartbeat type=apache)
  Attributes: configfile=/etc/httpd/conf/httpd.conf httpd=/usr/sbin/httpd
  Meta Attrs: target-role=Stopped
  Operations: monitor interval=10s timeout=20s (apache-monitor-interval-10s)
```

配置实例

KYLINSOFT 麒麟软件

reset命令重设bundle

pcs resource bundle reset <bundle id> [container <container options>] [network <network options>] [port-map <port options>]... [storage-map <storage options>]... [meta <meta options>] [--disabled] [--wait[=n]]

与bundle update不同的是，这个命令根据给定的选项重置bundle——不保留之前的选项。包中的资源保持原样。reset指令修改现有的bundle资源需要带上镜像名称，可以更改也可以不更改。

修改镜像pcs resource bundle reset httpd-bundle container image="pcmktest:http" replicas="2" run-command="/usr/sbin/pacemaker_remoted" ...

相当于删除旧的bundle重建

注：必须带上镜像名称

配置实例

KYLINSOFT 麒麟软件

绑定bundle和普通资源

pcs resource create httpd ocf:heartbeat:apache bundle httpd-bundle

```
[root@node1 ~]# pcs resource config httpd-bundle
 Bundle: httpd-bundle
  Docker: image=pcmktest:http2 replicas=2 run-command=/usr/sbin/pacemaker_remoted
  Network: host-interface=ens33 host-netmask=24 ip-range-start=172.17.127.185
  Port Mapping:
   port=80 (httpd-port)
  Storage Mapping:
   options=rw source-dir-root=/var/local/containers target-dir=/var/www/html (httpd-root)
   options=rw source-dir-root=/var/log/pacemaker/bundles target-dir=/etc/httpd/logs (httpd-logs)
  Resource: httpd (class=ocf provider=heartbeat type=apache)
   Operations: monitor interval=10s timeout=20s (httpd-monitor-interval-10s)
               start interval=0s timeout=40s (httpd-start-interval-0s)
               stop interval=0s timeout=60s (httpd-stop-interval-0s)
```

```
Node List:
  * Online: [ node1 node2 ]
  * GuestOnline: [ httpd-bundle-0@node2 httpd-bundle-1@node1 ]

Active Resources:
  * Container bundle set: httpd-bundle [pcmktest:http2]:
    * httpd-bundle-0 (172.17.127.185)   (ocf::heartbeat:apache):        Started node2
    * httpd-bundle-1 (172.17.127.186)   (ocf::heartbeat:apache):        Started node1
```

在集群节点上执行ip a命令，可以看到ip-range-start设置的ip地址。

在浏览器中可以对http://ip:80 进行访问，此服务即为节点上的容器提供的httpd服务。

KYLINSOFT

麒麟软件

打造中国操作系统核心力量

THANKS

官方网站：

kylinos.cn

服务热线：

400-089-1870