Docker

> 说明：本文是基于博文Docker — 从入门到实践进行学习总结； 更多文档请移步官方文档

1. 相关名词

1.1 LXC

> LXC is a userspace interface for the Linux kernel containment features. Through a powerful API and simple tools, it lets Linux users easily create and manage system or application containers.

`LXC`是一组使用linux内核容器功能的用户接口；

1.2 cgroups

wiki：

> cgroups，其名称源自控制组群（control groups）的简写，是Linux内核的一个功能，用来限制、控制与分离一个进程组群的资源（如CPU、内存、磁盘输入输出等）

1.3 aufs

> aufs (short for advanced multi-layered unification filesystem) implements a `union mount`(操作系统中管理文件夹的方式) for Linux file systems.

1.4 UnionFs

中文：统一文件系统 作用：将多个文件系统（同行叫作`分支`）合并成一个统一的fs，具有相同路径的目录会全部展示在合并后的fs的对应目录位置，分支拥有优先级，如果不同分支中包含同名文件，分支合并（组合成UnionFs）后，高优先级分支中的文件会覆盖低优先级的；

> Unionfs is a filesystem service for Linux, FreeBSD and NetBSD which implements a union mount for other file systems. It allows files and directories of separate file systems, known as branches, to be transparently overlaid, forming a single coherent file system. `UnionFs`是`union mount`的实现； aufs is a complete rewrite of the earlier UnionFS.

1.5 runc

`runc`是一个创建和运行容器的客户端工具；

> runc is a CLI tool for spawning and running containers according to the OCI specification.

1.6 OCI

OCI是一个专门制定系统级虚拟化技术开源标准的组织；

> The *Open Container Initiative* (OCI) is a Linux Foundation project to design open standards for operating-system-level virtualization, most importantly Linux containers.

1.7 containerd

`containerd`是个生产级的容器运行时，主打简便性、稳定性和可移植性等特点；

> containerd is an industry-standard container runtime with an emphasis on simplicity, robustness and portability. It is available as a daemon for Linux and Windows, which can manage the complete container lifecycle of its host system: image transfer and storage, container execution and supervision, low-level storage and network attachments, etc.

2. 什么是Docker

2.1 开发语言

> Docker 使用 Google 公司推出的 Go 语言 进行开发实现

2.2 低层技术

> 最初实现是基于LXC，从 0.7 版本以后开始去除LXC，转而使用自行开发的`libcontainer`(容器管理工具，现已弃用，转为用runc)，从 1.11 开始，则进一步演进为使用runc和 containerd。

2.3 虚拟技术与容器

虚拟机技术: 虚拟出一套硬件后，在其上运行一个完整操作系统，在该系统上再运行所需应用进程；

容器： 应用进程直接运行于宿主的内核，容器内没有自己的内核，而且也没有进行硬件虚拟。因此容器要比传统虚拟机更为轻便。

2.3.1 容器的优点

1. 更高效

> 不需要进行硬件虚拟以及运行完整操作系统等额外开销，Docker 对系统资源的利用率更高;

2. 启动更快

> 传统的虚拟机技术启动应用服务往往需要数分钟，而 Docker 容器应用，由于直接运行于宿主内核，无需启动完整的操作系统，因此可以做到秒级、甚至毫秒级的启动时间。

3. 一致的运行时环境
4. 方便持续交付和部署
5. 方便迁移
6. 轻松维护和扩展

> Docker 使用的分层存储以及镜像的技术，使得应用重复部分的复用更为容易，基于基础镜像进一步扩展镜像也变得非常简单。

3. 概念

镜像、容器、仓库、数据卷

4. 安装

4.1 设置加速器

镜像加速器

5. 镜像

5.1 镜像操作

5.1.1 简单示例

以获取并运行Ubuntu镜像为例,参考：

1. 获取镜像

```
# 不添加标签（版本），默认获取latest标签的镜像
docker pull ubuntu

# 如果要添加标签，可以如下
docker pull ubuntu:18.04
```

命令格式：

```
$ docker pull [选项] [Docker Registry 地址[:端口]/]仓库名[:标签]
```

> 仓库名：这里的仓库名是两段式名称，即 <用户名>/<软件名>。对于 Docker Hub，如果不给出用户名，则默认为 library，也就是官方镜像

2. 运行镜像

```
# -i 表示交互式执行命令
# -t 表示终端，因为要执行bash命令，所以需要交互式终端
# -rm 表示容器退出后随之将其删除；默认情况下，为了排障需求，退出的容器并不会立即删除，除非手动 docker rm
docker run -it --rm ubuntu:18.04  bash
```

3. 检验

```
# 查看系统信息
cat /etc/os-release
```

4. 退出

```
# 退出容器
exit
```

5.1.2 列出镜像

```
# 显示镜像列表
$ docker image ls
# 仓库名过滤，如库名Ubuntu
$ docker image ls ubuntu
# 仓库名+标签过滤，如：Ubuntu:18.04
$ docker image ls ubuntu:18.04
# 更加复杂的过滤，可以使用--filter或-f选项，如：mongo:3.2之后建立的镜像
$ docker image ls -f since=mongo:3.2
# 只列出镜像的ID
$ docker image ls -q
# 镜像列表的结构也可自定义，用到Go的模板语法
$ docker image ls --format "{{.ID}}: {{.Repository}}"
```

```
# 以表格等距显示，并且有标题行，和默认一样，不过自己定义列
$ docker image ls --format "table {{.ID}}\t{{.Repository}}\t{{.Tag}}"

# 由于docker镜像是分层的，可能多个镜像共用同一个基础镜像，所以上面命令得到的镜像列表中所有镜像大小的总和并不是所有镜像实际占用的存储大小，可使用下面命令查看实际大小
$ docker system df
```

注：go-template-syntax

5.1.3 虚悬镜像

查看虚悬镜像列表：

```
$ docker image ls -f dangling=true
```

一般来说，虚悬镜像已经失去了存在的价值，是可以随意删除的，可以用下面的命令删除

```
$ docker image prune
```

5.1.4 中间层镜像

```
$ docker image ls -a
```

5.1.5 删除镜像

命令

```
$ docker image rm [OPTIONS] IMAGE [IMAGE...]
# 更多帮助信息
$ docker image rm --help
```

删除所有仓库名为 redis 的镜像:

```
$ docker image rm $(docker image ls -q redis)
```

5.2 commit镜像

> 当我们运行一个容器的时候（如果不使用卷的话），我们做的任何文件修改都会被记录于容器存储层里。 Docker 提供了一个 docker commit 命令，可以将容器的存储层保存下来成为镜像。

> 注意： docker commit 命令除了学习之外，还有一些特殊的应用场合，比如被入侵后保存现场等。但是，不要使用 docker commit 定制镜像，定制镜像应该使用Dockerfile来完成。

5.2.1 命令

docker commit 的语法格式，请查看help信息：

```
$ docker commit --help
```

5.2.2 慎用commit

commit生成的镜像会越来越臃肿，详细参考

5.3 定制镜像

一般使用`dockerfile`脚本记录构建镜像的每一个步骤；

5.3.1 docker指令

5.3.1.1 FROM

必须以`FROM`开头，如下：以Ubuntu为基础镜像

```
FROM ubuntu
RUN echo '<h1>Hello, Docker!</h1>' > /usr/share/nginx/html/index.html
```

除了选择现有镜像为基础镜像外，Docker 还存在一个特殊的镜像，名为 scratch。这个镜像是虚拟的概念，并不实际存在，它表示一个空白的镜像，以 scratch 为基础镜像的话，意味着你不以任何镜像为基础，接下来所写的指令将作为镜像第一层开始存在：

```
FROM scratch
...
```

> 不以任何系统为基础，直接将可执行文件复制进镜像的做法并不罕见，比如 swarm、coreos/etcd。对于 Linux 下静态编译的程序来说，并不需要有操作系统提供运行时支持，所需的一切库都已经在可执行文件里了，因此直接 FROM scratch 会让镜像体积更加小巧。使用 Go 语言开发的应用很多会使用这种方式来制作镜像，这也是为什么有人认为 Go 是特别适合容器微服务架构的语言的原因之一。

5.3.1.2 RUN

> Union FS 是有最大层数限制的，比如 AUFS，曾经是最大不得超过 42 层，现在是不得超过 127 层

> 每一条RUN指令，就代表创建一层镜像；

所以，不要每条shell命令都添加RUN指令前缀，如下，创建了7层镜像：

```
FROM debian:stretch

RUN apt-get update
RUN apt-get install -y gcc libc6-dev make wget
RUN wget -O redis.tar.gz "http://download.redis.io/releases/redis-5.0.3.tar.gz"
RUN mkdir -p /usr/src/redis
RUN tar -xzf redis.tar.gz -C /usr/src/redis --strip-components=1
RUN make -C /usr/src/redis
RUN make -C /usr/src/redis install
```

正确的方式：

```
# 这是RUN的正确使用方式
# 并且，最后还进行了rm清理工作，将编译redis编译过程中使用得软件和包清理掉
RUN buildDeps='gcc libc6-dev make wget' \
    && apt-get update \
    && apt-get install -y $buildDeps \
    && wget -O redis.tar.gz "http://download.redis.io/releases/redis-5.0.3.tar.gz" \
    && mkdir -p /usr/src/redis \
    && tar -xzf redis.tar.gz -C /usr/src/redis --strip-components=1 \
    && make -C /usr/src/redis \
    && make -C /usr/src/redis install \
    && rm -rf /var/lib/apt/lists/* \
    && rm redis.tar.gz \
    && rm -r /usr/src/redis \
    && apt-get purge -y --auto-remove $buildDeps
```

5.3.1.3 更多指令

参考 | ONBUILD

5.3.2 构建镜像

构建命令：

```
$ docker build -t name:tag pathOfContext

# 更多help信息
$ docker  build --help
```

5.3.3 构建镜像上下文

参考: docker-image-build-context / 镜像构建上下文

> 一般来说，应该会将 Dockerfile 置于一个空目录下，或者项目根目录下。如果该目录下没有所需文件，那么应该把所需文件复制一份过来。如果目录下有些东西确实不希望构建时传给 Docker 引擎，那么可以用 .gitignore 一样的语法写一个 .dockerignore，该文件是用于剔除不需要作为上下文传递给 Docker 引擎的。

5.3.4 其他构建方法

5.4 镜像传输

5.4.1 regestry

推荐使用 `Docker Registry` 管理镜像；

5.4.2 非Regestry

docker save and docker load 上面博文中有个命令，自行理解：

```
$ docker save <镜像名> | bzip2 | pv | ssh <用户名>@<主机名> 'cat | docker load'
```

5.5 多标签

```
$ docker tag ubuntu:18.04 username/ubuntu:18.04

$ docker image ls
REPOSITORY                                          TAG                 IMAGE ID            CREATED
SIZE
ubuntu                                              18.04               275d79972a86        6 days
ago          94.6MB
username/ubuntu                                     18.04               275d79972a86
```

5.6 镜像推送

对于自定义仓库的地址(如：`127.0.0.1:5000`，对外ip为`192.168.199.100`)，如果跨主机进行镜像推送的话（如：`docker push 192.168.199.100:5000/ubuntu:latest`），是会推送失败的，因为 Docker 默认不允许非 HTTPS 方式推送镜像，解决方式：(对于`Ubuntu 16.04+, Debian 8+, centos 7`)在`/etc/docker/daemon.json`中添加`insecure-registries`配置

```
{
  "registry-mirror": [
```

```
    "https://registry.docker-cn.com"
  ],
  "insecure-registries": [
    "192.168.199.100:5000"
  ]
}
```

参考

5.7 多阶段构建

参考多阶段构建

6. 容器

6.1 启动

主要命令：

```
$ docker run -t -i ubuntu:18.04 /bin/bash
```

查看运行中的容器：

```
$ docker ps
$ docker container ls

# 查看所有容器，包括已关闭的容器
$ docker ps -a
$ docker container ls -a
```

再次启动已经关闭的容器：

```
$ docker container -ia start 容器id
```

6.2 守护态

后台运行容器：

```
$ docker run -d ubuntu:18.04 /bin/sh -c "while true; do echo hello world; sleep 1; done"
```

查看后台运行容器输出：

```
$ docker container logs <容器id>
```

6.3 关闭

```
$ docker container stop
```

6.4 进入

进入正在运行的容器：

6.4.1 attach

使用这种方式进入容器，在通过`exit`退出容器时，容器也会关闭：

```
$ docker attach <容器id>
```

6.4.2 exec

使用此命令进入容器，在通过`exit`退出时，容器不会关闭，所以推荐使用这种方式操作容器：

```
$ docker exec -it <容器id>
```

6.5 导入导出

6.5.1 导出快照

```
$ docker export <容器id> ubuntu.tar
```

6.5.2 导入

```
$  cat ubuntu.tar | docker import - test/ubuntu:v1.0
```

> 用户既可以使用 docker load 来导入镜像存储文件到本地镜像库，也可以使用 docker import 来导入一个容器快照到本地镜像库。这两者的区别在于容器快照文件将丢弃所有的历史记录和元数据信息（即仅保存容器当时的快照状态），而镜像存储文件将保存完整记录，体积也要大。此外，从容器快照文件导入时可以重新指定标签等元数据信息。

6.6 删除

删除一个处于终止状态的容器：

```
$ docker container rm [container]
```

> 如果要删除一个运行中的容器，可以添加 -f 参数。Docker 会发送 SIGKILL 信号给容器。

清理所有已经停止运行的容器：

```
$ docker container prune
```

7. 数据管理

7.1 数据卷

相关命令：

```
# 创建数据卷
$ docker volume create my-vol

# 查看数据卷
$ docker volume ls

# 查看指定数据卷的信息
$ docker volume inspect my-vol

# 删除数据卷
$ docker volume rm my-vol
```

7.1.1 挂载

用docker run命令的时候，使用--mount标记来将数据卷挂载到容器里。在一次docker run中可以挂载多个数据卷:

```
# 创建一个名为 web 的容器，并加载一个 数据卷 到容器的 /webapp 目录
$ docker run -d -P \
    --name web \
    # -v my-vol:/wepapp \
    --mount source=my-vol,target=/webapp \
    training/webapp \
    python app.py
```

> 注意：Docker挂载点（上面的`target`参数）不支持相对路径 Docker does not support relative paths for mount points inside the container.

查看数据卷的具体信息:

```
# 使用以下命令可以查看 web 容器的信息
$ docker inspect web
```

数据卷信息在`Mounts`属性里：

```
"Mounts": [
    {
        "Type": "volume",
        "Name": "my-vol",
        "Source": "/var/lib/docker/volumes/my-vol/_data",
        "Destination": "/app",
        "Driver": "local",
        "Mode": "",
        "RW": true,
        "Propagation": ""
    }
]
```

7.1.2 删除

废弃的数据卷需要手动删除，如果需要在删除容器的同时移除数据卷。可以在删除容器的时候使用 docker rm -v 这个命令；可以通过`$ docker volume prune`命令，清理无主的数据卷；

7.1.3 共享

- 在创建容器的时候，如果有数据卷挂载操作，容器内待挂载的目录下又有文件或子目录，那么docker会先将目录下的内容（文件和子目录）拷贝到卷中（宿主机中的目录），然后再挂载卷；如果此时其他容器也使用了相同的卷，那么其他容器也可以看见拷贝的文件

> If you start a container which creates a new volume, as above, and the container has files or directories in the directory to be mounted (such as /app/ above), the directory's contents are copied into the volume. The container then mounts and uses the volume, and other containers which use the volume also have access to the pre-populated content.

8. 网络

Docker的网络模式

8.1 映射端口

有`-P`（大写），`-p`（小写）两个参数：

- 当使用 -P 标记时，Docker 会随机映射一个 49000~49900 的端口到内部容器开放的网络端口
- -p 则可以指定要映射的端口，并且，在一个指定端口上只可以绑定一个容器。支持的格式有 ip:hostPort:containerPort | ip::containerPort | hostPort:containerPort。`-p` 标记可以多次使用来绑定多个端口

```
# 本地的 5000 端口映射到容器的 5000 端口
$ docker run -d -p 5000:5000 training/webapp python app.py

#指定映射使用一个特定地址，比如 localhost 地址 127.0.0.1
$ docker run -d -p 127.0.0.1:5000:5000 training/ebapp python app.py
# 可以使用 udp 标记来指定 udp 端口
$ docker run -d -p 127.0.0.1:5000:5000/udp training/webapp python app.py

# 绑定 localhost 的任意端口到容器的 5000 端口，本地主机会自动分配一个端口
$ docker run -d -p 127.0.0.1::5000 training/webapp python app.py
```

8.1.1 查看映射端口配置

```
$ docker port nostalgic_morse 5000
```

8.2 容器互联

8.2.1 创建docker网络

-d 参数指定 Docker 网络类型，有 bridge overlay。其中 overlay 网络类型用于 Swarm mode，在本小节中你可以忽略它。

```
$ docker network create -d bridge my-net
```

8.2.2 连接网络

参考

8.2.3 compose

如果你有多个容器之间需要互相连接，推荐使用 Docker Compose。

9. 更多命令

`docker --help` *可以帮到你*

```
# 镜像实际占用硬盘空间
$ docker system df

# 启动终止的容器
$ docker container start

# 重启运行中的容器
$ docker container restart

#  查看某个镜像后创建的镜像
$ docker image inspect --format='{{.RepoTags}} {{.Id}} {{.Parent}}' $(docker image ls -q --filter since=
<imageId>)
```

10. 扩展

10.1 自制SSL证书

使用openssl自行签发站点SSL证书请参考

11. 注意

docker最佳实践

11.1 RUN

- 永远将 RUN apt-get update 和 apt-get install 放在一条 RUN 中；

> 固定版本会迫使构建过程检索特定的版本，而不管缓存中有什么。这项技术也可以减少因所需包中未预料到的变化而导致的失败。

- 不要使用 RUN apt-get upgrade 或 dist-upgrade；