Containerd 上手实践

@张晋涛

关于我

- 知乎/微博： @张晋涛
- 网易有道
- Container/Docker/Kubernetes

目录

- 从 K8S 宣布弃用 dockershim 说起
- containerd 概览
- 主要功能及上手实践
- 用作 K8S 的 runtime

K8S 宣布弃用 dockershim

dockershim 是什么

- Kubelet 中的独立组件
- 以 CRI 操作 Docker
- 2016 年独立为 dockershim 组件
- Kubernetes 创建之初便采用 Docker 作为其容器运行时

CRI 是什么

- Container Runtime Interface
- 2016 年底开始应用
- 目标
  - 增强 Kubelet 的可扩展性
  - 可插拔容器运行时
  - 提高代码可维护性

为何弃用 dockershim

- 维护成本过高
  - 功能/特性变更
- 容器运行时多样
  - containerd 已经成熟
  - cri-o
- 减少调用链
  - containerd 是 Docker 的容器运行时
  - 可直接通过 CRI 与 containerd 交互

有什么影响

- 终端用户无任何影响
- 在 Kubernetes 之上的开发者无影响
- 集群管理员
  - 可考虑是否切换容器运行时
  - Mirantis 和 Docker 公司合作维护 dockershim 组件
  - 可通过树外 dockershim 组件继续使用 Docker 作为容器运行时
- Docker 仍然是容器构建/运行的最佳选择！

相关信息

- #94624 Deprecate Dockershim
- Don't Panic: Kubernetes and Docker
- Dockershim Deprecation FAQ
- What developers need to know about Docker, Docker Engine, and Kubernetes v1.20

containerd 概览

containerd 是什么

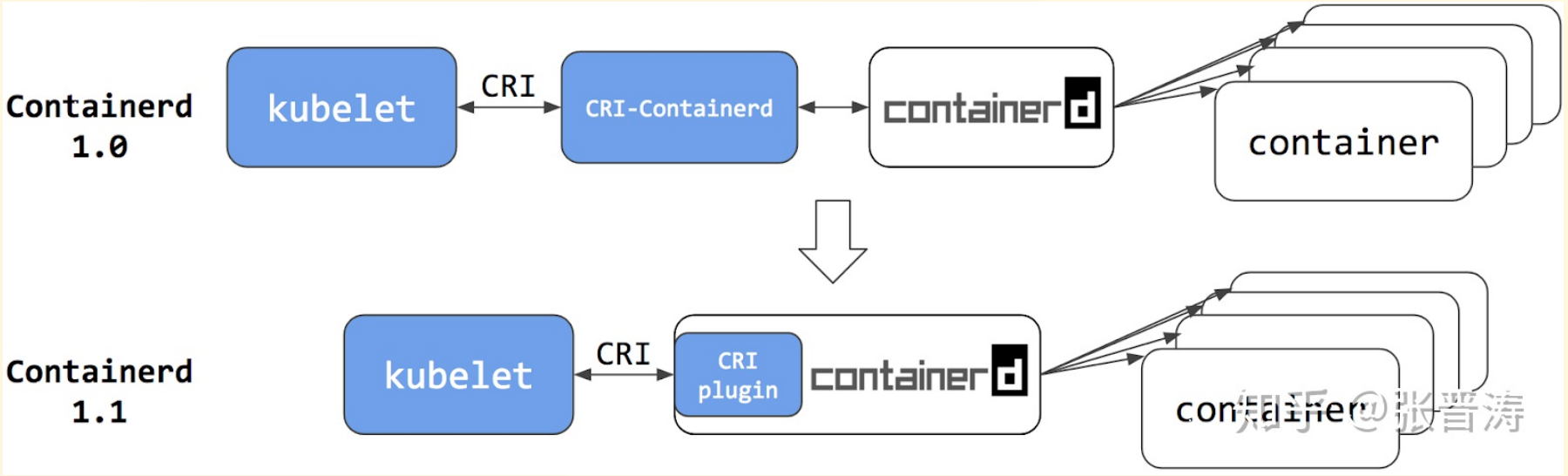
- 中间层容器运行时
  - 平台之下（Docker/Kubernetes ）
  - 底层运行时之上（runc，gVisor）
- 资源管理器
  - 容器进程
  - 容器镜像
  - 文件系统快照
  - 元数据和依赖管理

containerd 相关历史

- 由 Docker 创建
- 从容器管理员变更为完整的容器运行时
  - 作为 Docker 的容器运行时
  - 独立工作
- Docker 于 2017 年捐给 CNCF
- 2019 年从 CNCF 毕业

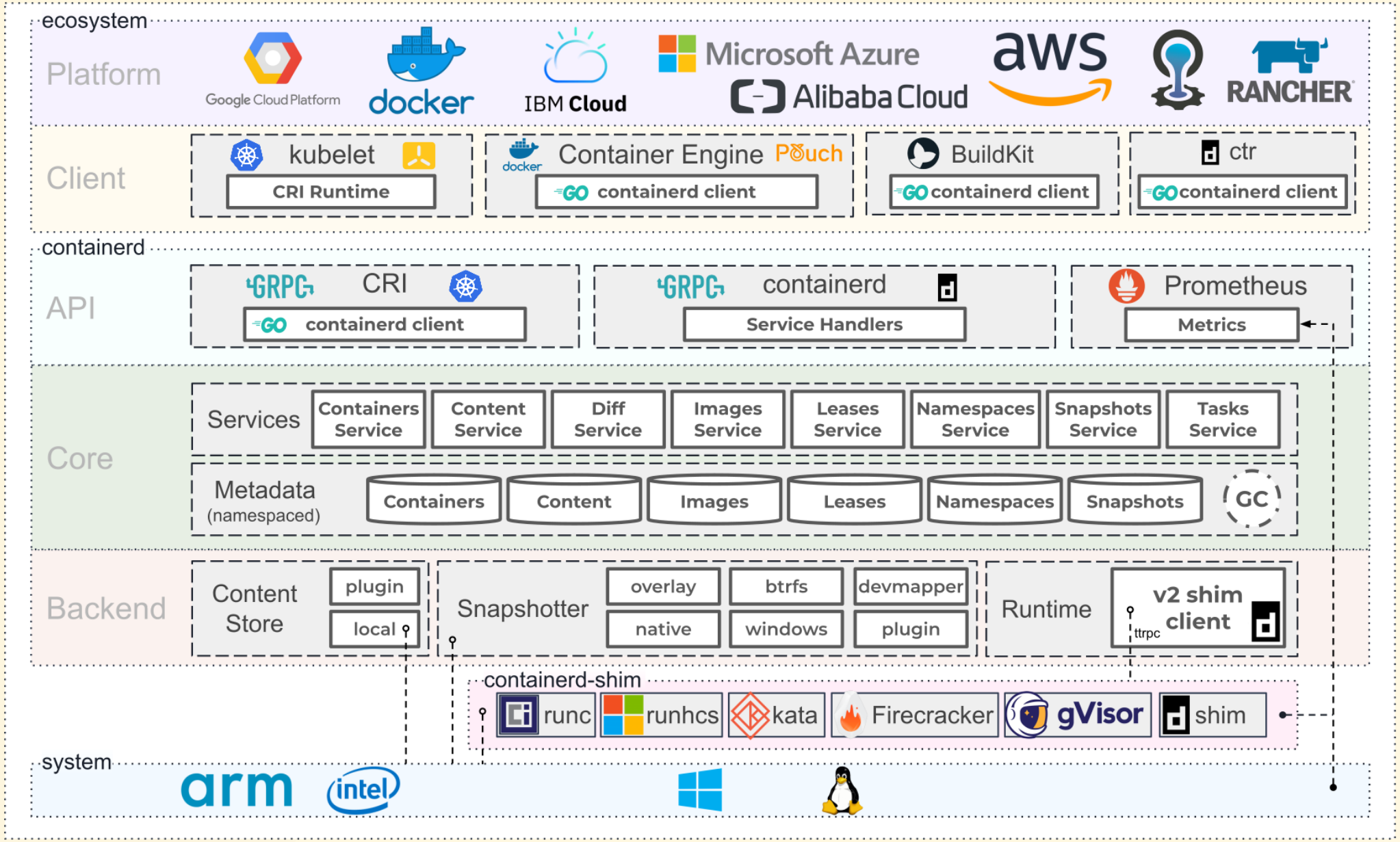
containerd 与 CRI

CRI 最初是作为独立进程支持的



主要特性

- 支持 OCI 镜像规范
- 支持 OCI 运行时规范（aka runC）
- 支持镜像的 push/pull
- 容器网络管理
- 存储支持多租户
- 容器运行时和生命周期支持
- 管理网络名称空间容器以加入现有名称空间


ecosystem
Platform
Google Cloud Platform
docker
IBM Cloud
Microsoft Azure
Alibaba Cloud
aws
RANCHER
Client
kubelet
CRI Runtime
Container Engine
containerd client
BuildKit
containerd client
ctr
containerd client
containerd
API
GRPC
CRI
containerd client
GRPC
containerd
Service Handlers
Prometheus
Metrics
Core
Services
Containers Service
Content Service
Diff Service
Images Service
Leases Service
Namespaces Service
Snapshots Service
Tasks Service
Metadata
(namespaced)
Containers
Content
Images
Leases
Namespaces
Snapshots
GC
Backend
Content Store
plugin
local
Snapshotter
overlay
btrfs
devmapper
native
windows
plugin
Runtime
v2 shim client
ttrpc
containerd-shim
runc
runhcs
kata
Firecracker
gVisor
shim
system
arm
intel

主要功能和上手实践

镜像管理

```
➜  ~ ctr -a /run/containerd/containerd.sock i pull docker.io/library/redis:alpine
docker.io/library/redis:alpine:                                                   resolved       |++++++++++++++++++++++++++++++++++++++|
index-sha256:68d4030e07912c418332ba6fdab4ac69f0293d9b1daaed4f1f77bdeb0a5eb048:    done           |++++++++++++++++++++++++++++++++++++++|
manifest-sha256:aa31e6d2afc72d2222ed3953587197c324f615861771637a64053f9d99ba4b74: done           |++++++++++++++++++++++++++++++++++++++|
layer-sha256:9a8d0188e48174d9e60f943c4e463c23268b864ed4f146041bee8d79710cc359:    done           |++++++++++++++++++++++++++++++++++++++|
layer-sha256:adc9264cf1338a3d0f3cd930237322d1ca9e664ed444d9e9bd3bdc087fa667ec:    done           |++++++++++++++++++++++++++++++++++++++|
config-sha256:c678242f9116148617f225c247be7fb10490d772b3a666d68b2930590d3d25a8:   done           |++++++++++++++++++++++++++++++++++++++|
layer-sha256:8a3f5c4e0176ac6c0887cb55795f7ffd12376b33d546a11bb4f5c306133e7606:    done           |++++++++++++++++++++++++++++++++++++++|
layer-sha256:51a20dbe2f6a64d3e18fe4d93b5d721cfc0819ed7cbbbd7b20f04de82fb52cfb:    done           |++++++++++++++++++++++++++++++++++++++|
layer-sha256:801bfaa63ef2094d770c809815b9e2b9c1194728e5e754ef7bc764030e140cea:    done           |++++++++++++++++++++++++++++++++++++++|
layer-sha256:0aacff13b8d72c393a20ccf266db96a471de3572871fe273ececd1607175b0a4:    done           |++++++++++++++++++++++++++++++++++++++|
elapsed: 21.9s                                                                    total:  9.0 Mi (422.9 KiB/s)
unpacking linux/amd64 sha256:68d4030e07912c418332ba6fdab4ac69f0293d9b1daaed4f1f77bdeb0a5eb048...
done
➜  ~ ctr -a /run/containerd/containerd.sock i ls
REF                            TYPE                                                       DIGEST
           SIZE     PLATFORMS                                                                                 LABELS
docker.io/library/redis:alpine application/vnd.docker.distribution.manifest.list.v2+json sha256:68d4030e07912c418332ba6fdab4ac69f0293d9b1daaed4f1f77bd
eb0a5eb048 10.1 MiB linux/386,linux/amd64,linux/arm/v6,linux/arm/v7,linux/arm64/v8,linux/ppc64le,linux/s390x -
➜  ~
```

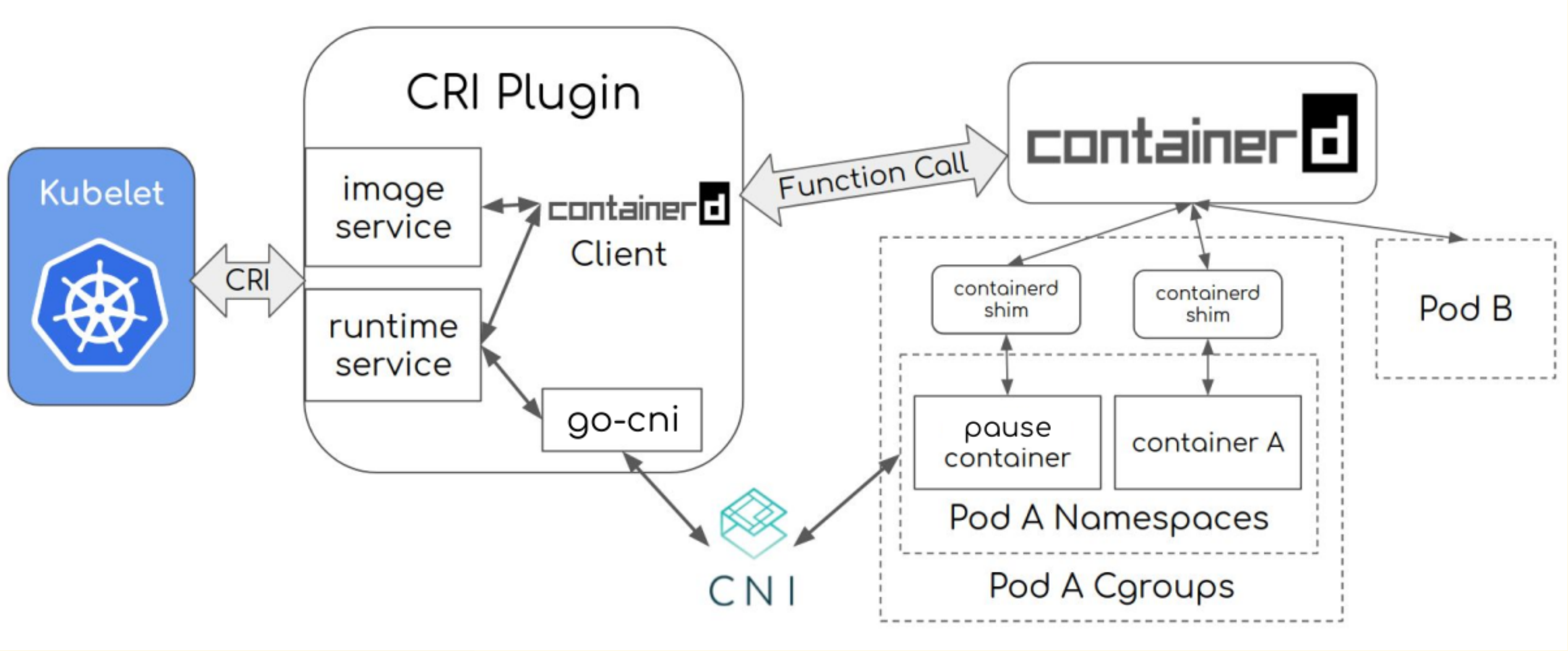
容器管理

```
➜  ~ sudo ctr -a /run/containerd/containerd.sock container create docker.io/library/redis:alpine redis
➜  ~ sudo ctr -a /run/containerd/containerd.sock container ls
CONTAINER    IMAGE                              RUNTIME
redis        docker.io/library/redis:alpine     io.containerd.runc.v2
➜  ~ sudo ctr -a /run/containerd/containerd.sock task ls
TASK    PID    STATUS
➜  ~ sudo ctr -a /run/containerd/containerd.sock task start -d redis
➜  ~ sudo ctr -a /run/containerd/containerd.sock task ls
TASK     PID       STATUS
redis    353643    RUNNING
```

命名空间

```
➜  ~ sudo ctr -n default -a /run/containerd/containerd.sock task ls
TASK     PID       STATUS
redis    353643    STOPPED
➜  ~ sudo ctr -n moby -a /run/containerd/containerd.sock task ls
TASK                                                    PID     STATUS
1017c68b807cd762a17ecff09712a94d8c550550c88e149d086b76d0602ed921    2113    RUNNING
➜  ~ docker ps --no-trunc --format '{{.ID}} - {{.Status}}' |grep 1017c68b807cd762a17ecff09712a94d8c550550c88e149d086b76d0602ed921
1017c68b807cd762a17ecff09712a94d8c550550c88e149d086b76d0602ed921 - Up 25 hours
➜  ~
```

用作 K8S 的 runtime


CRI Plugin
Kubelet
CRI
image service
runtime service
go-cni
Client
Function Call
containerd
containerd shim
containerd shim
Pod B
pause container
container A
Pod A Namespaces
Pod A Cgroups
CNI

检查配置文件

```
# cat /etc/containerd/config.toml
version = 2

[plugins."io.containerd.grpc.v1.cri".containerd]
  default_runtime_name = "runc"
[plugins."io.containerd.grpc.v1.cri".containerd.runtimes.runc]
  runtime_type = "io.containerd.runc.v2"

[plugins."io.containerd.grpc.v1.cri".containerd.runtimes.runc.options]
    [plugins."io.containerd.grpc.v1.cri".cni]
      bin_dir = "/opt/cni/bin"
      conf_dir = "/etc/cni/net.d"
      max_conf_num = 1
      conf_template = ""

[plugins."io.containerd.grpc.v1.cri"]
  sandbox_image = "k8s.gcr.io/pause:3.3"
```

更多资源

- kind
- k3c/k3s

Thanks!