CoreOS

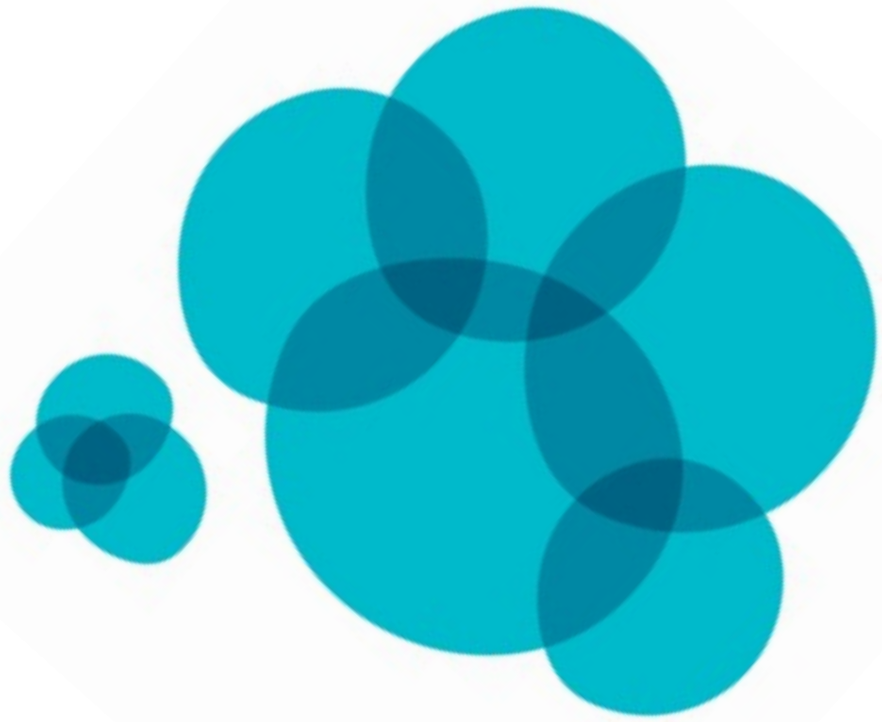
ThoughtWorks® LinFan

行走在云端的系统

CoreOS

- 云·集群
- 集群的那些棘手事儿
- CoreOS 系统级解决方案
- CoreOS 是完美的吗
- 案例和总结

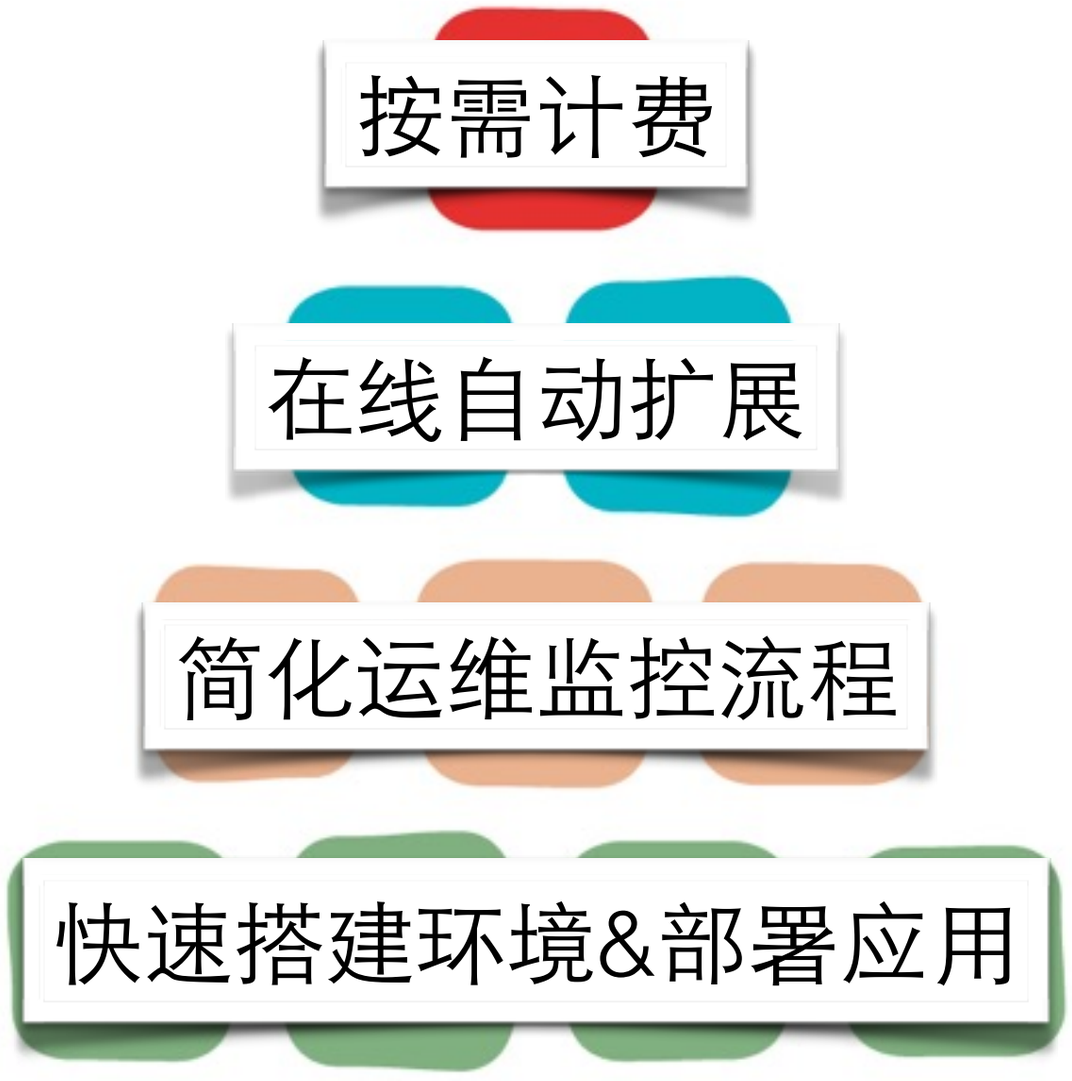
云·集群



云·集群

云带来的变革

云带来的变革



大规模集群带来的问题

大规模集群带来的问题

成员管理

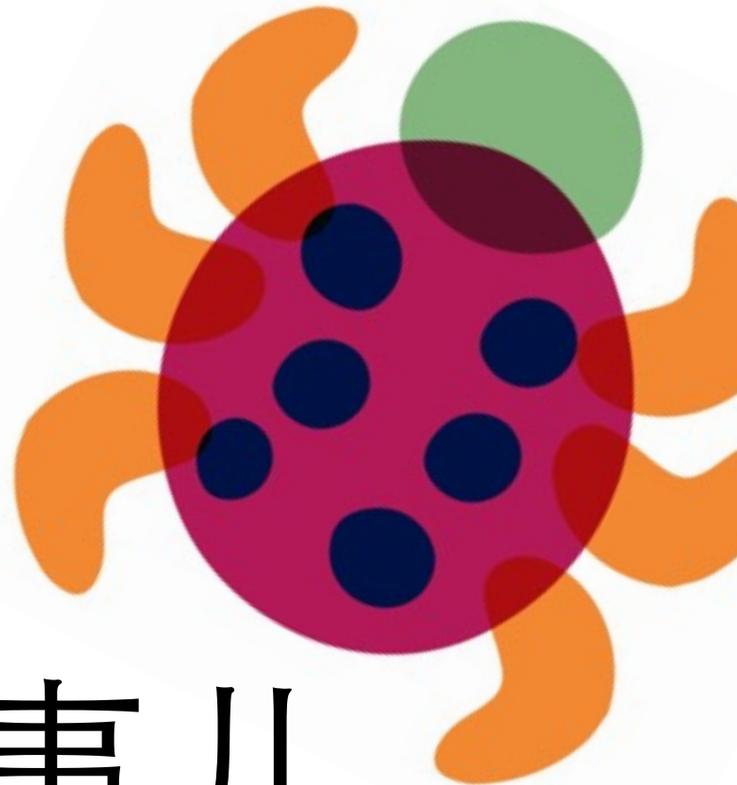
状态信息同步

操作系统补丁和升级

集群内部的服务容灾迁移


CoreOS

CoreOS



集群的那些棘手事儿

集群成员的管理

成员管理

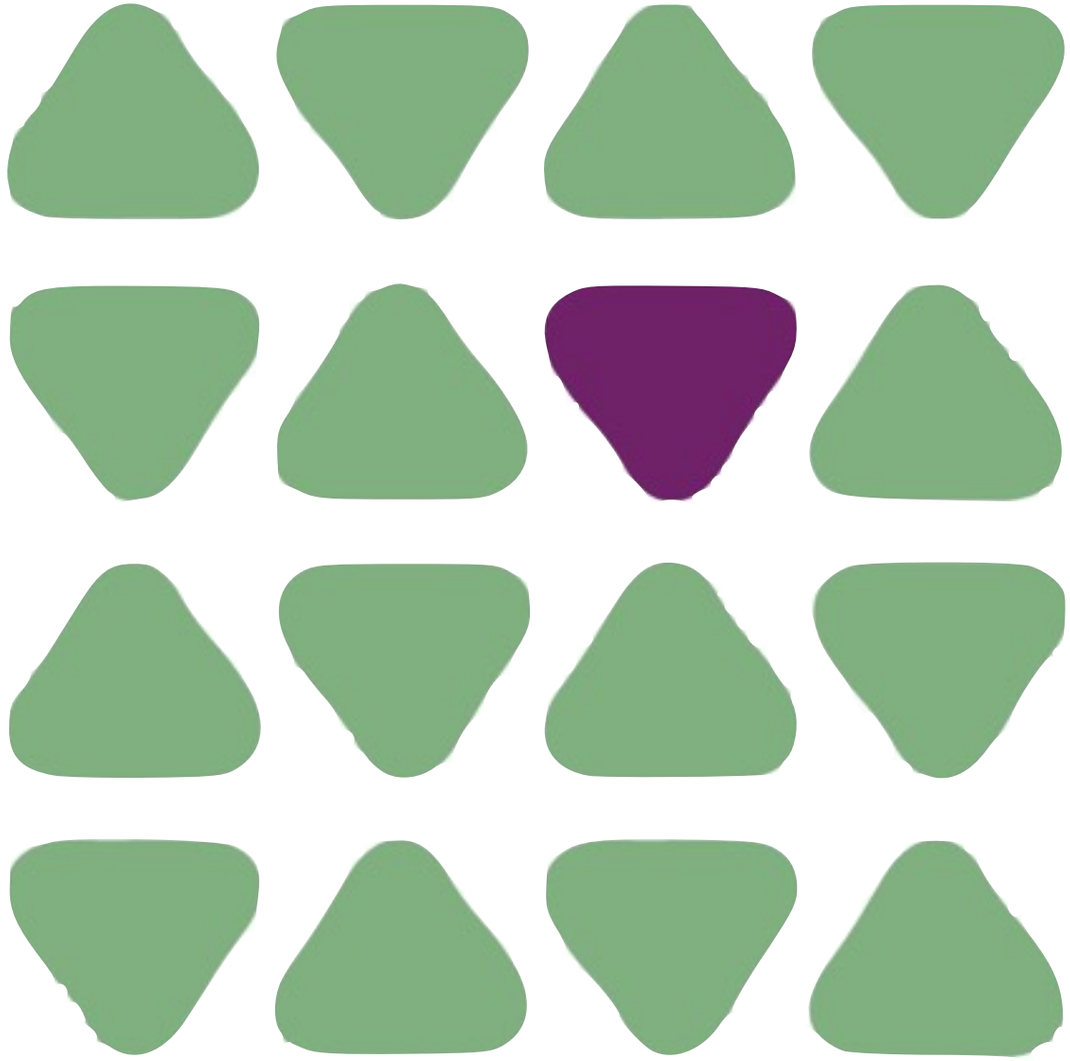
集群成员的管理

集群成员的管理

我们希望:

- 自动注册新加入集群的服务节点
- 自动将不可用的节点从集群剔除
- 节点对所依赖的服务节点变化能够自动适应

集群内节点信息共享



集群内节点信息共享

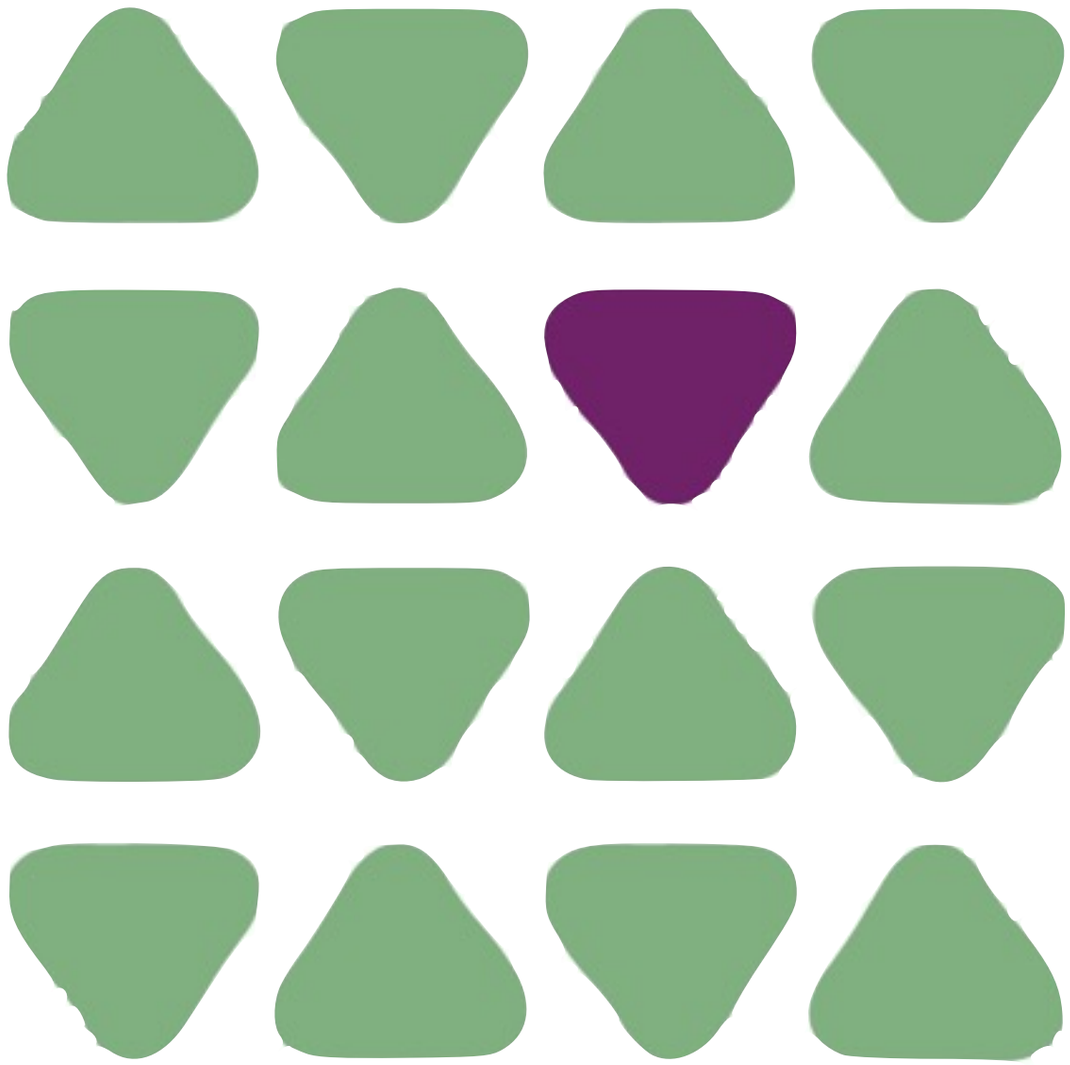
我们计划调整一下集群结构。

由于节点之间互不可知，目前许多节点服务的 IP 地址是固定写在配置里的。

首先需要增加两个 DB 节点，修改所有 DB 节点配置，再修改 WEB001 到 WEB020 的每个节点配置，然后...

等等...呃..感觉好像还忘了什么地方...

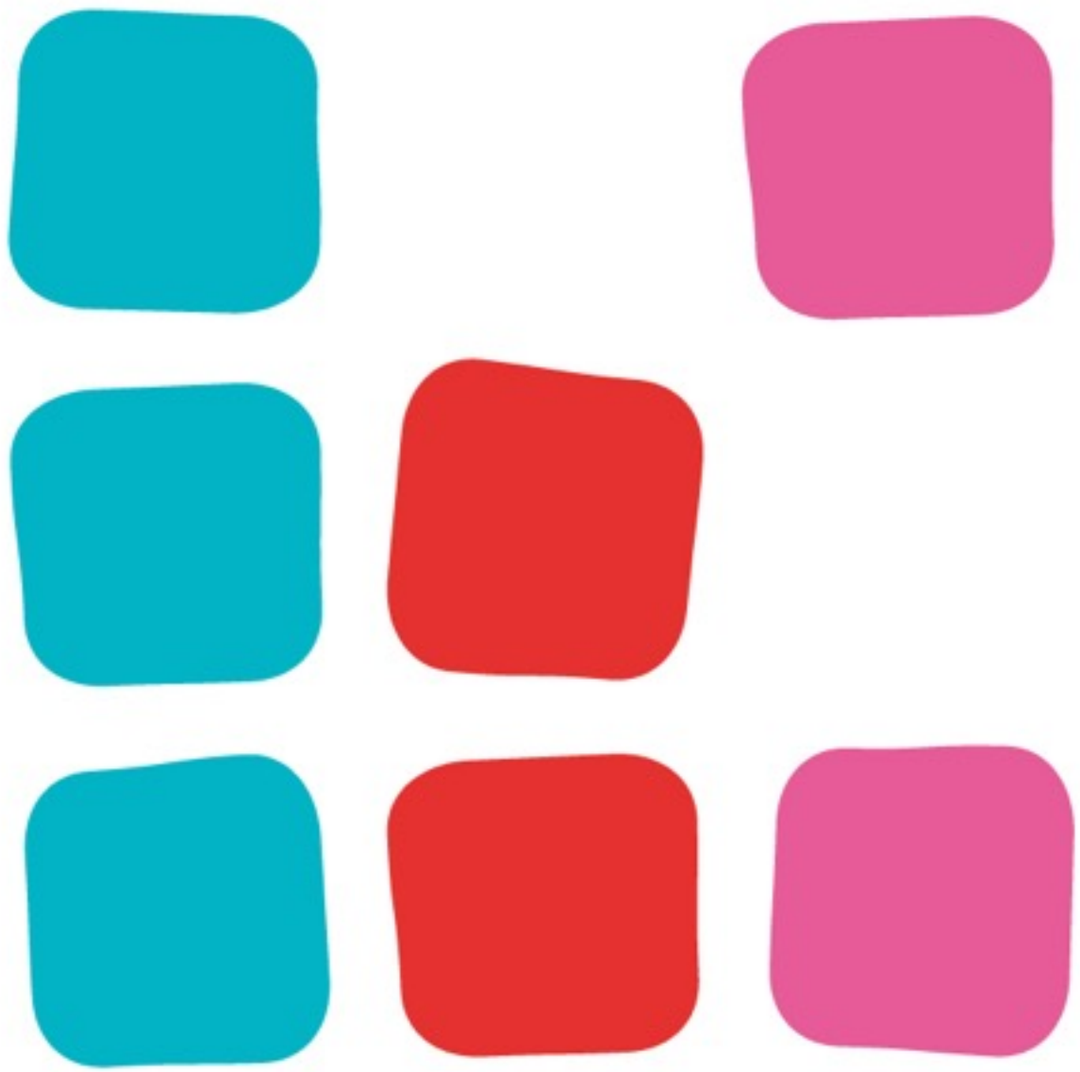
集群内节点信息共享



我们希望:

- 每个节点都能实时获得整个集群的最新状态信息
- 信息的格式是可自定义和扩展的
- 节点的信息同步是实时且透明的

操作系统补丁和升级



操作系统补丁和升级

2014年11月9日美国 BrowserStack 公司服务器遭黑客入侵，入侵者盗取用户信息后以 BrowserStack 公司身份给部分用户发出一封通知邮件称 BrowserStack 公司泄露用户数据并即将关门。

http://techcrunch.com/2014/11/10/hacker-emails-testing-service-browserstacks-customers-says-company-lied-about-security/

操作系统补丁和升级

2014年11月13日，BrowserStack 官方通过Gist向所有用户发出事件分析结果。
由于在几千台服务器集群中仅有一个节点没有及时安装 ShellShock 漏洞补丁，加上秘钥管理上的安全失误，最终导致入侵者获得数据库访问权限和用户信息，并发出了伪造的邮件。

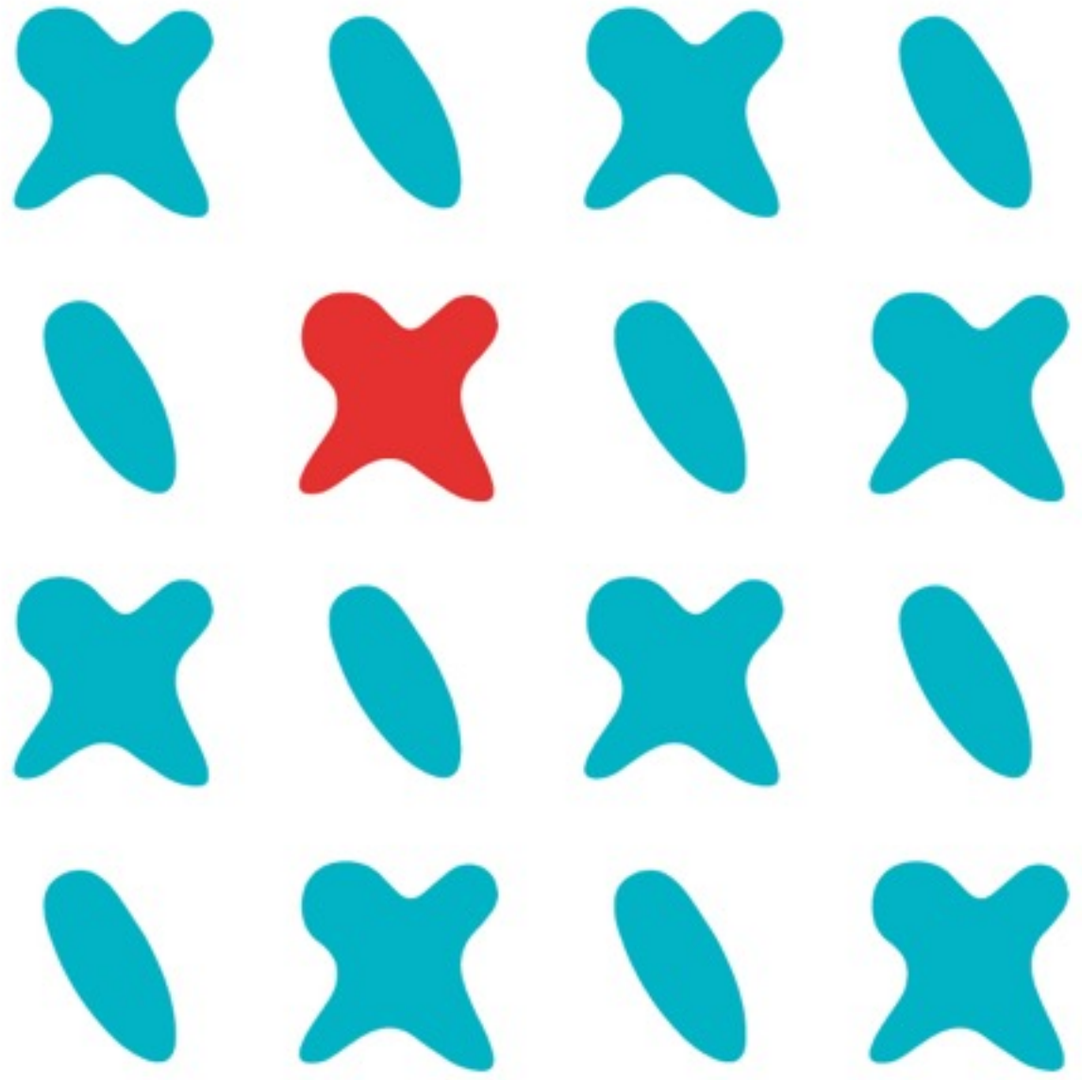
https://gist.github.com/simonsarris/9b16e436e035f90ec35f

操作系统补丁和升级

我们希望:

- 系统具备自动升级和安装补丁的能力
- 操作系统升级全过程，整个集群对外提供的服务不中断
- 更新在必要的情况下能够快速回滚

集群服务的容灾迁移



集群服务的容灾迁移

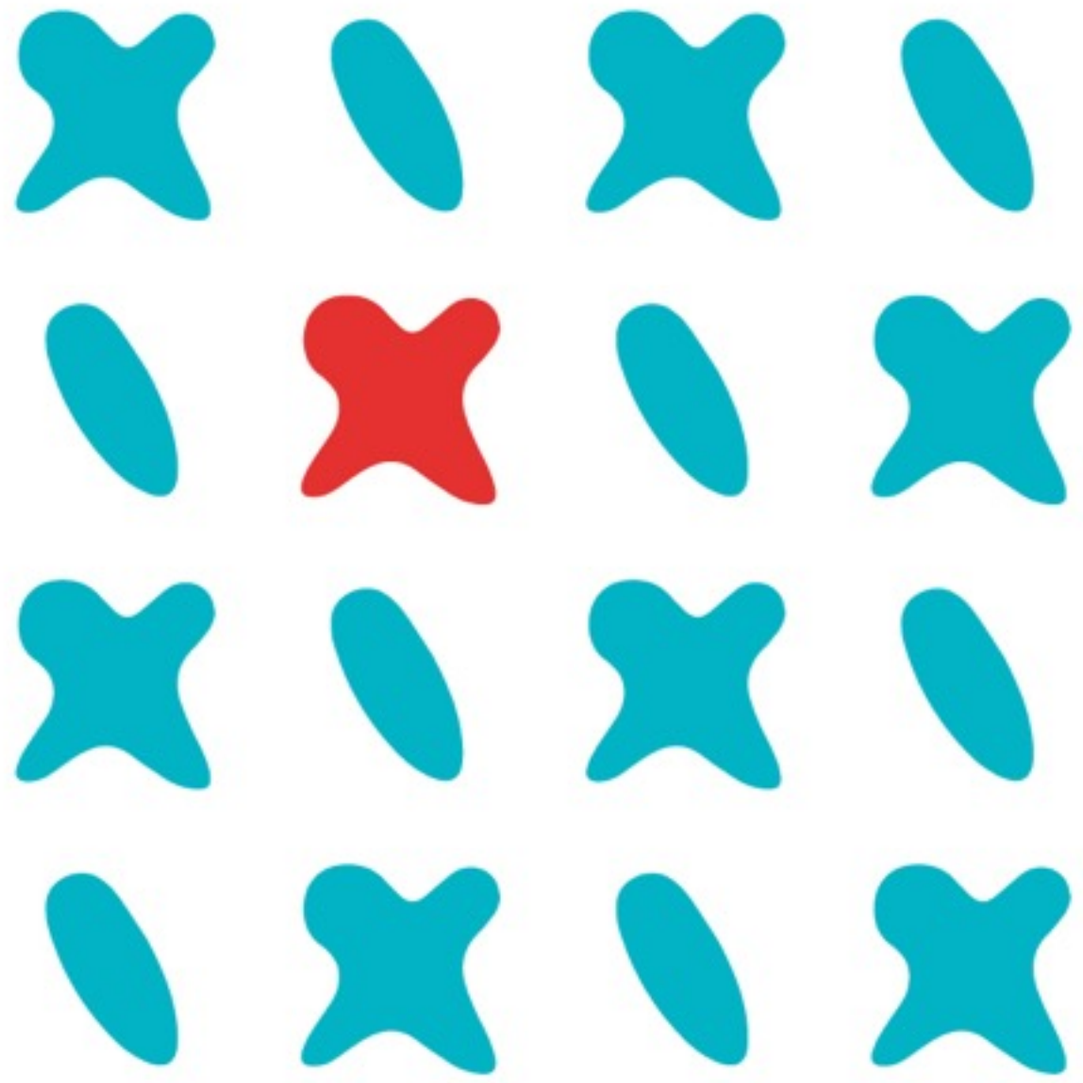
这是一个大量使用 Micro Service 的集群，每一个节点都提供了系统的一个或多个服务。

将一个服务在另一个节点启动时常遇到依赖版本的各种问题...
确保集群一致性是很困难的事情...

集群服务的容灾迁移

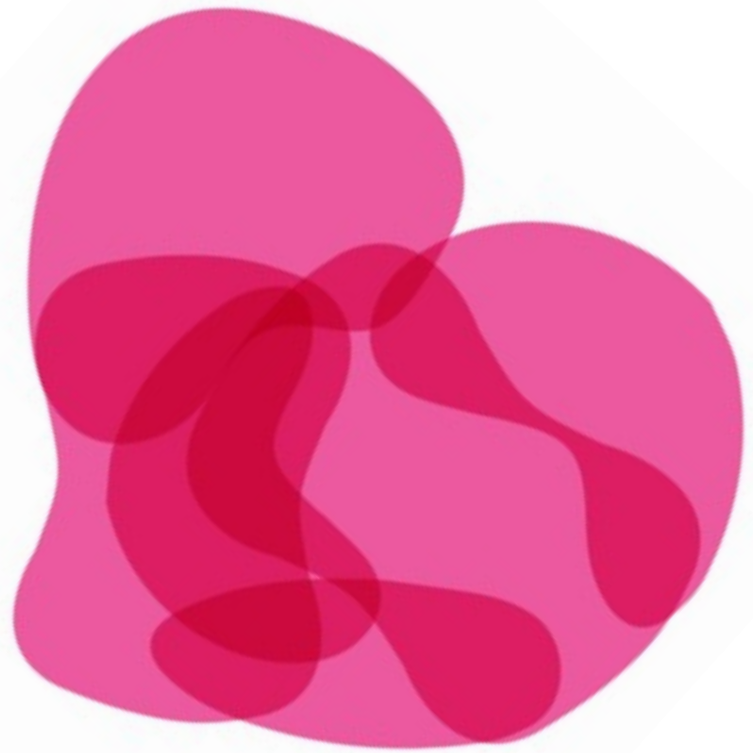
我们希望:

- 当故障发生时，服务能够自动转移到健康的节点
- 配置特定服务允许迁移的节点名单
- 解决服务器之间的运行环境不一致



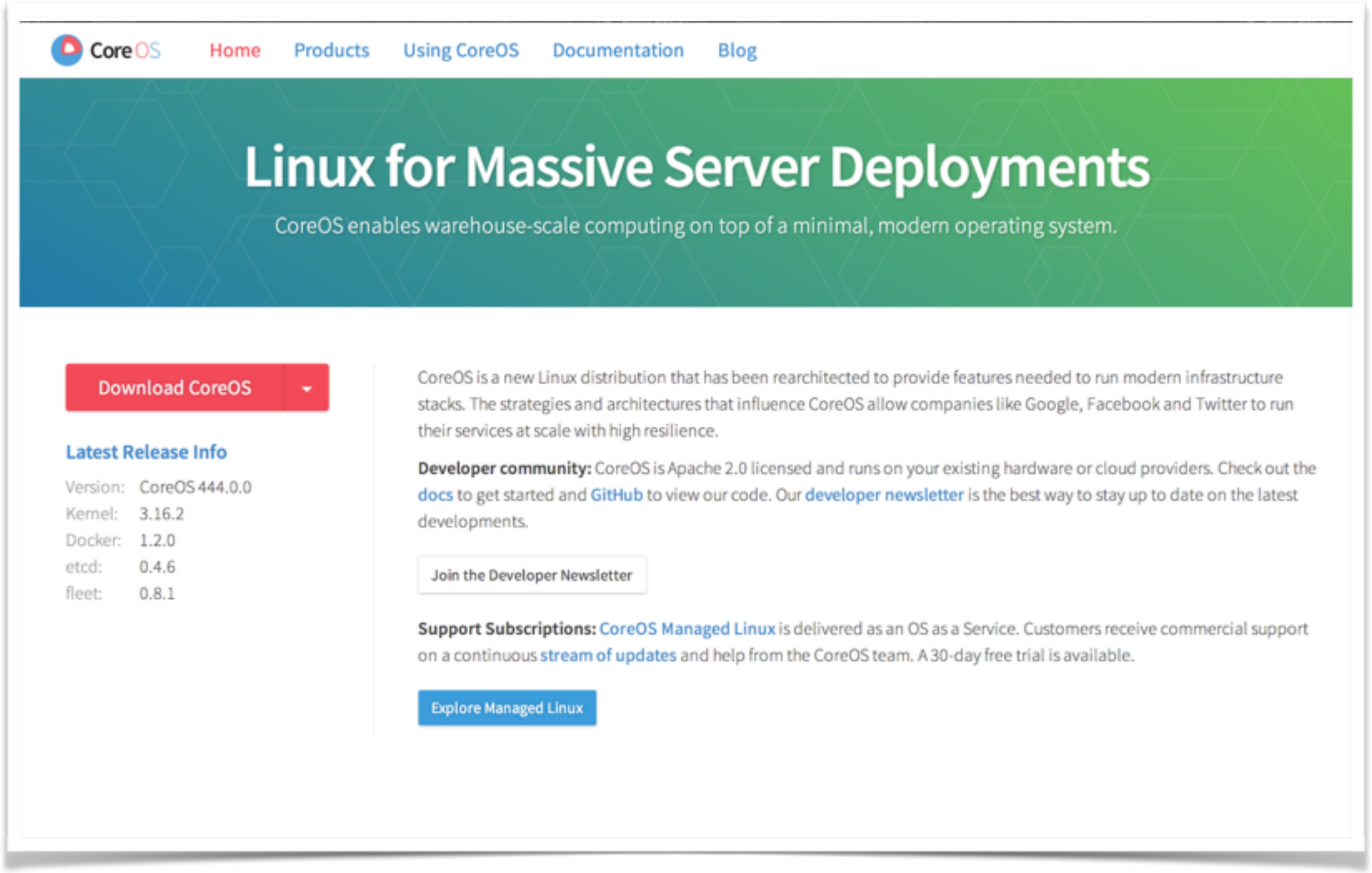
CoreOS 系统级解决方案

CoreOS

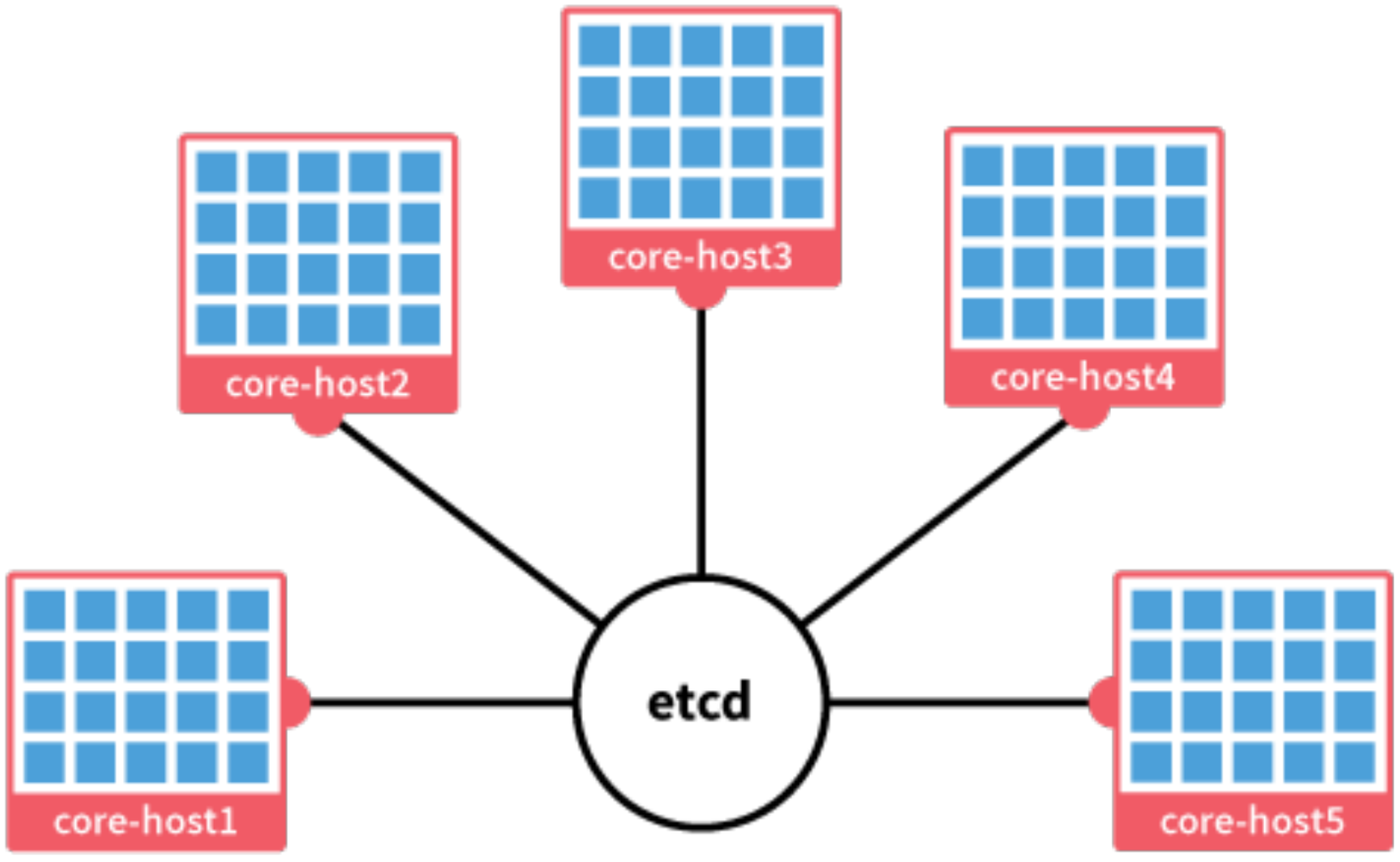


CoreOS 系统级解决方案

“操作系统级”的解决方案

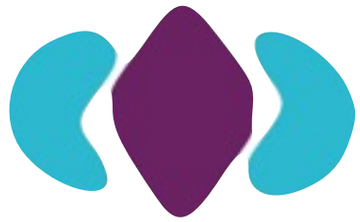

CoreOS
Home
Products
Using CoreOS
Documentation
Blog
Linux for Massive Server Deployments
CoreOS enables warehouse-scale computing on top of a minimal, modern operating system.
Download CoreOS
Latest Release Info
Version: CoreOS 444.0.0
Kernel: 3.16.2
Docker: 1.2.0
etcd: 0.4.6
fleet: 0.8.1
CoreOS is a new Linux distribution that has been rearchitected to provide features needed to run modern infrastructure stacks. The strategies and architectures that influence CoreOS allow companies like Google, Facebook and Twitter to run their services at scale with high resilience.
Developer community: CoreOS is Apache 2.0 licensed and runs on your existing hardware or cloud providers. Check out the docs to get started and GitHub to view our code. Our developer newsletter is the best way to stay up to date on the latest developments.
Join the Developer Newsletter
Support Subscriptions: CoreOS Managed Linux is delivered as an OS as a Service. Customers receive commercial support on a continuous stream of updates and help from the CoreOS team. A 30-day free trial is available.
Explore Managed Linux

【集群的成员管理和信息共享】
使用 cloud-init 实现集群服务器自组网
使用 etcd 服务实现分布式数据共享



现有集群

现有集群



新的节点

现有集群

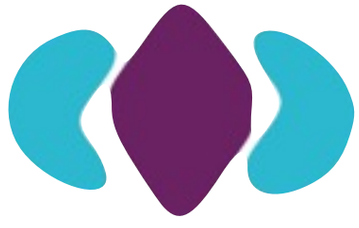
新的节点

新的节点

现有集群

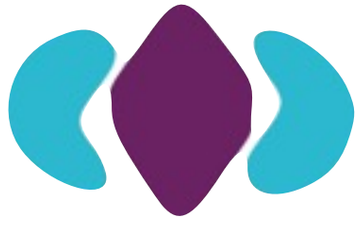
现有集群

现有集群



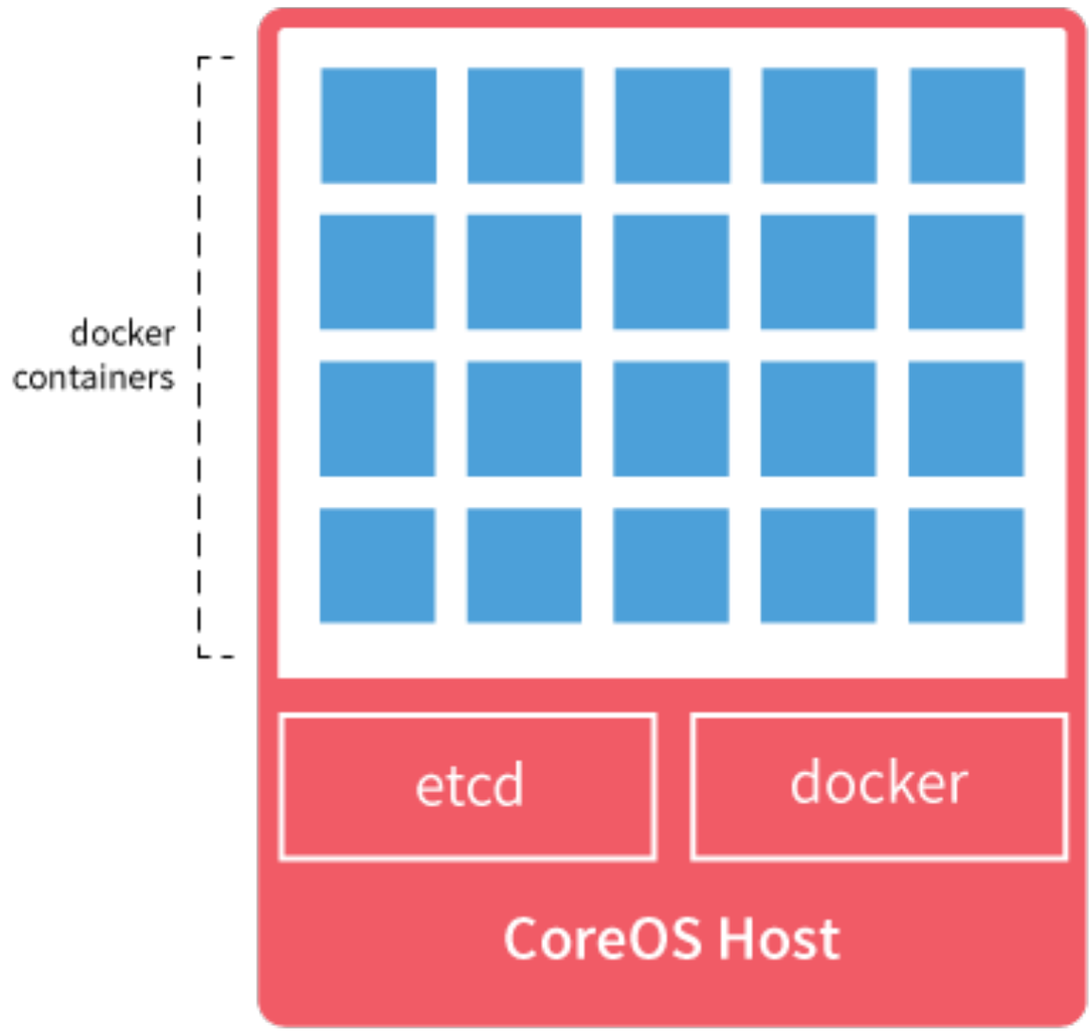
节点失联

现有集群



节点失联

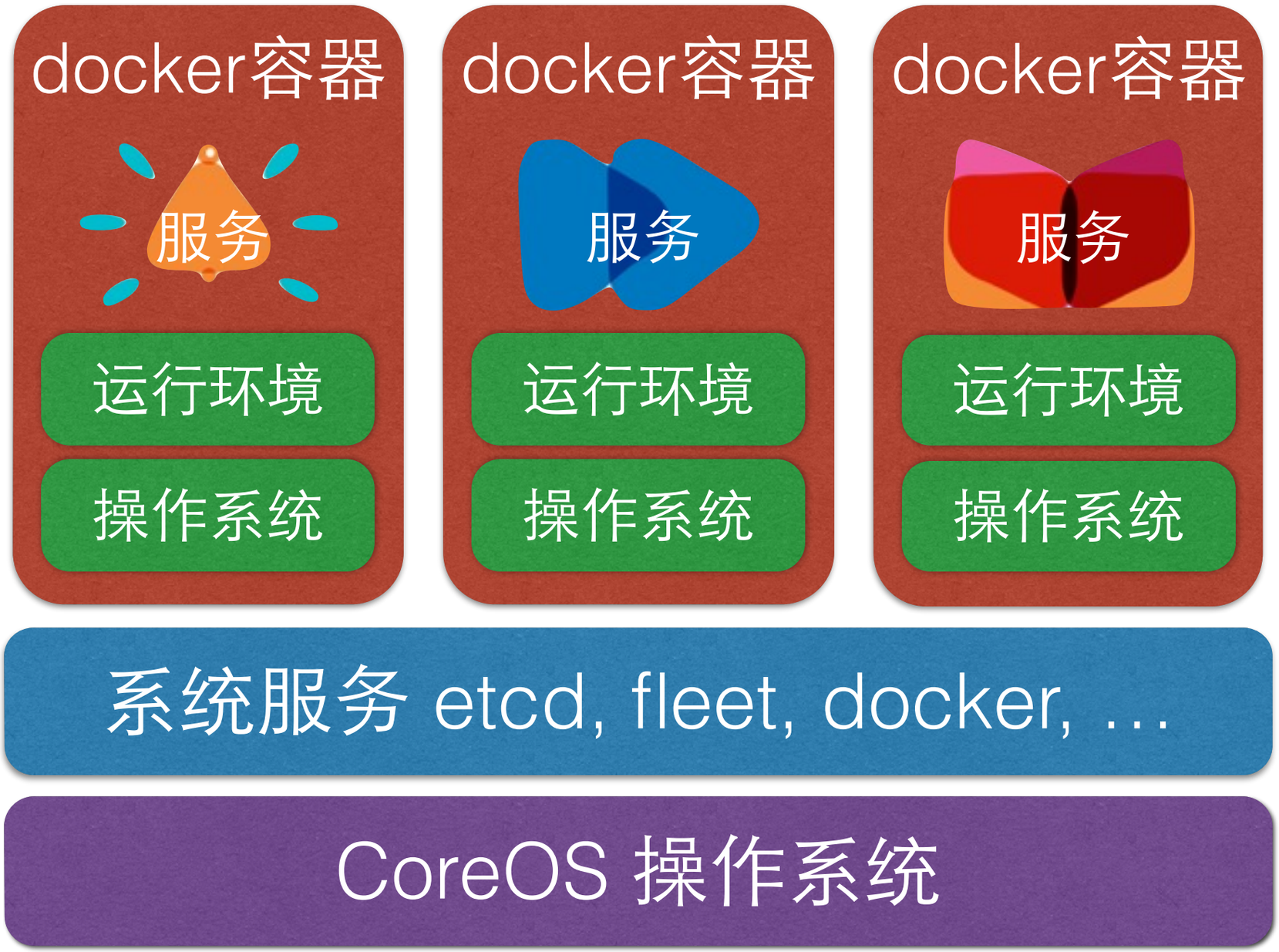
【操作系统与运行环境隔离解决服务依赖冲突】
只读的操作系统分区
原生支持 docker 容器隔离服务环境



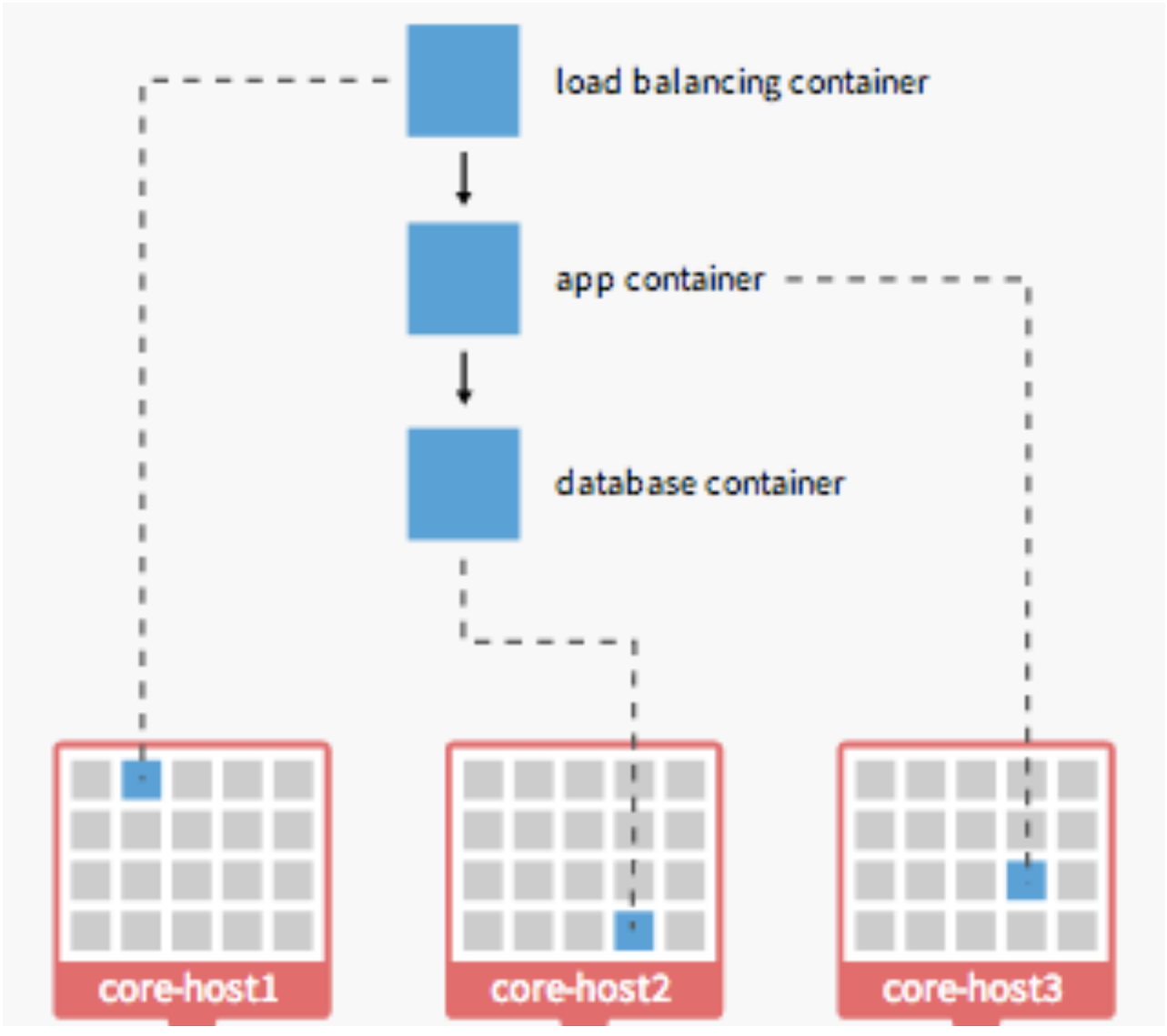
CoreOS 操作系统

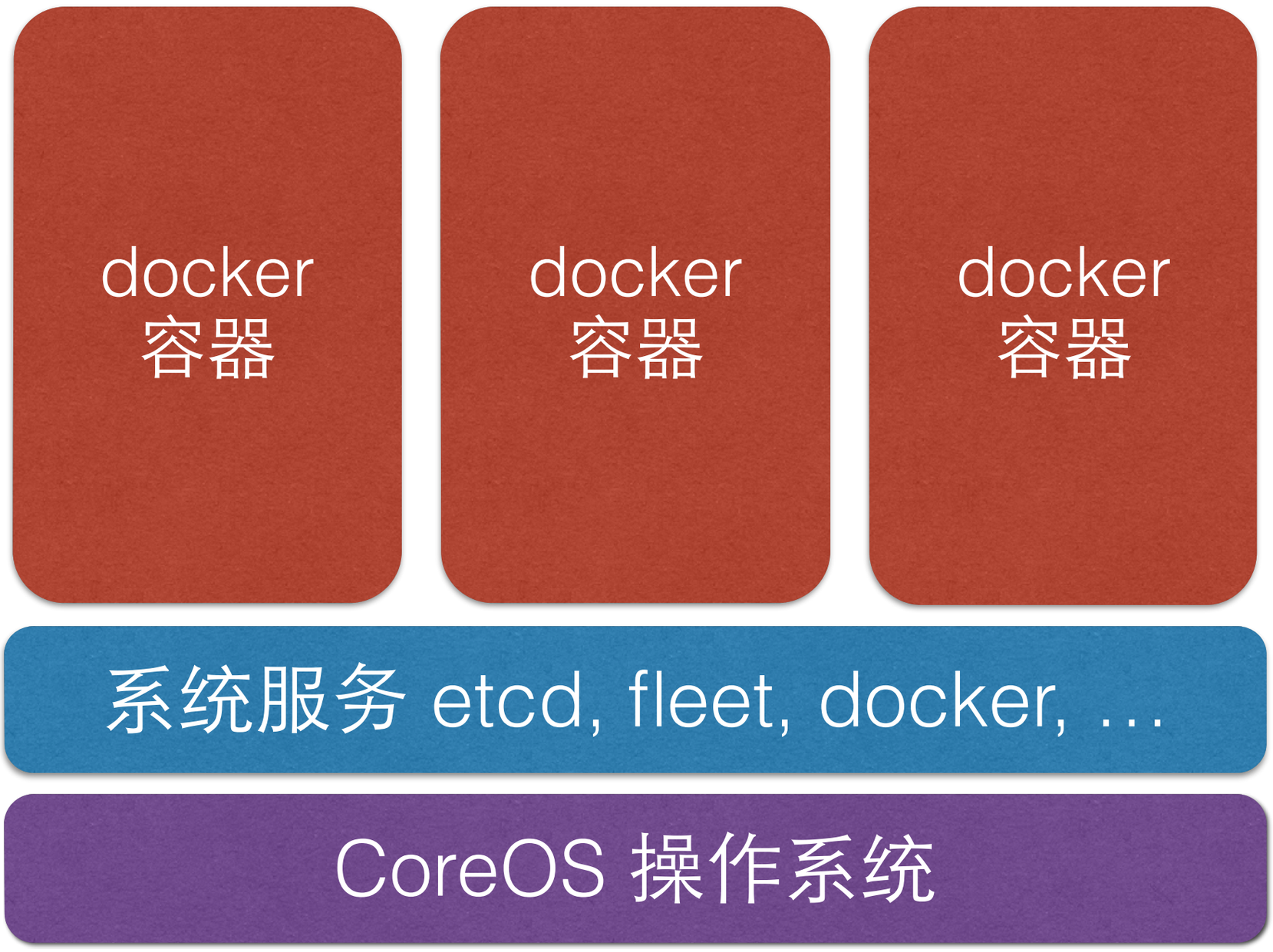
系统服务 etcd, fleet, docker, …

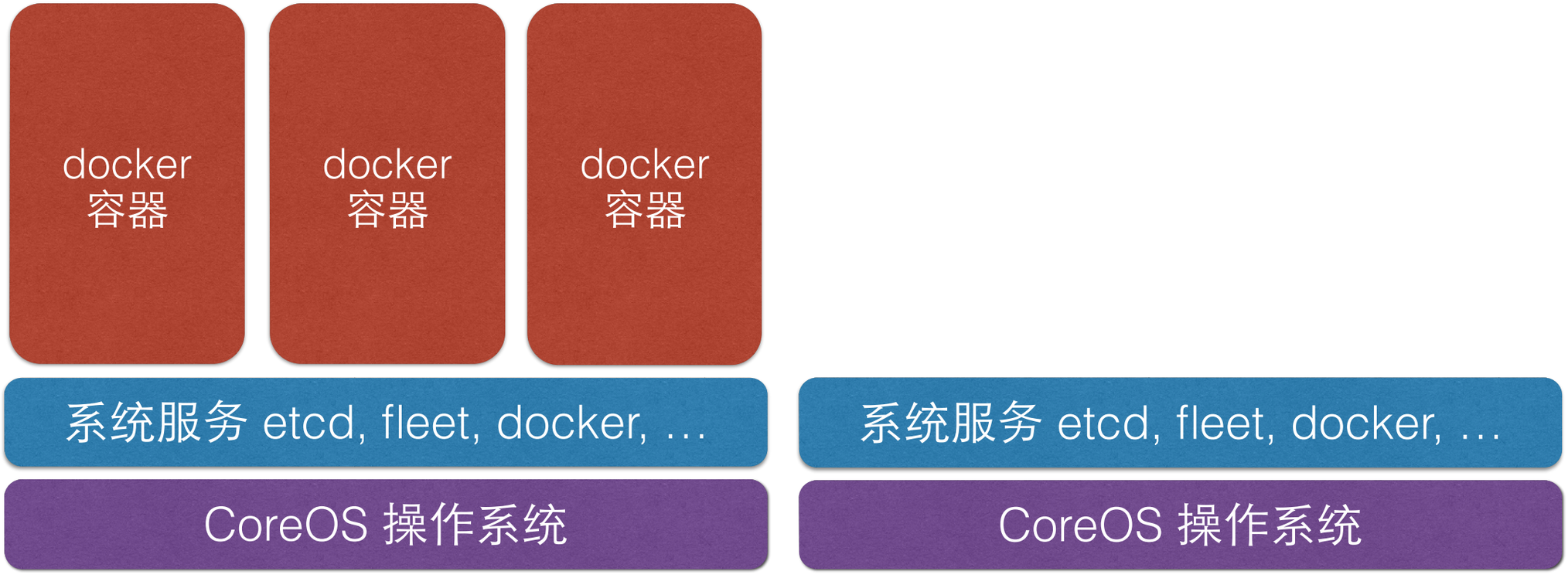
CoreOS 操作系统


docker容器
服务
运行环境
操作系统
docker容器
服务
运行环境
操作系统
docker容器
服务
运行环境
操作系统
系统服务 etcd, fleet, docker, ...
CoreOS 操作系统

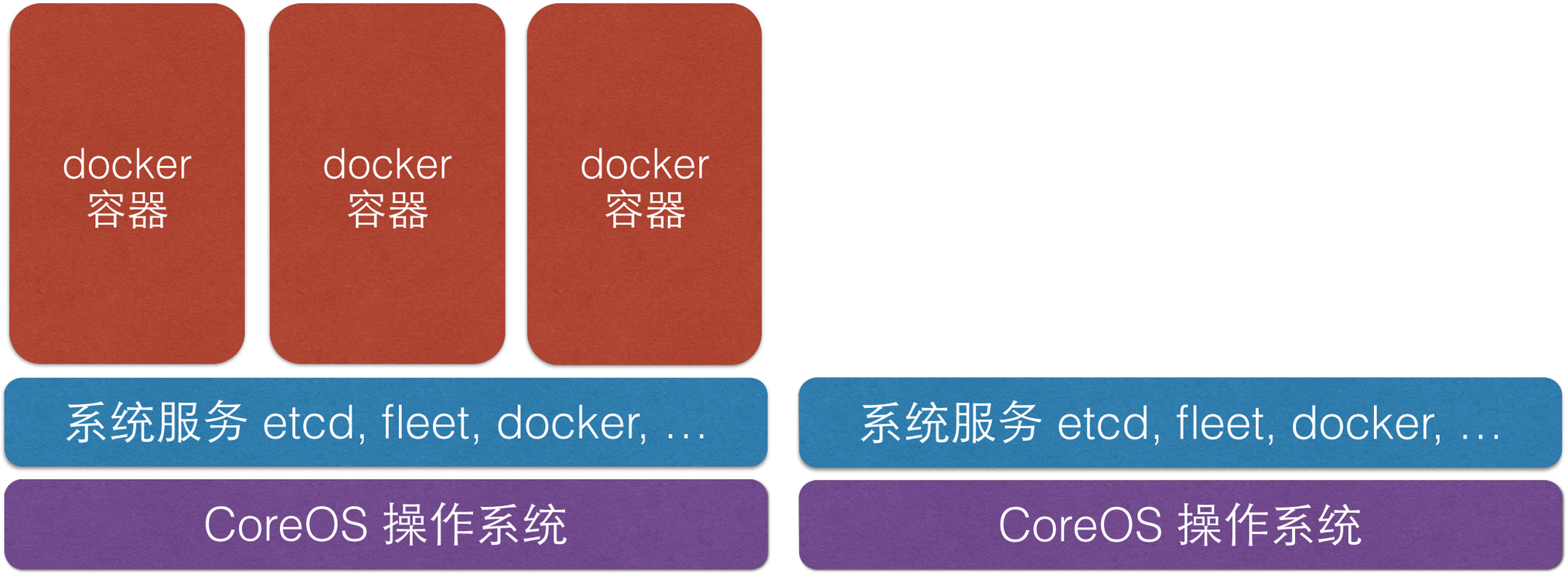
【集群服务的管理和迁移】
使用 docker 容器确保服务快速迁移
使用 fleet 管理协调集群事务

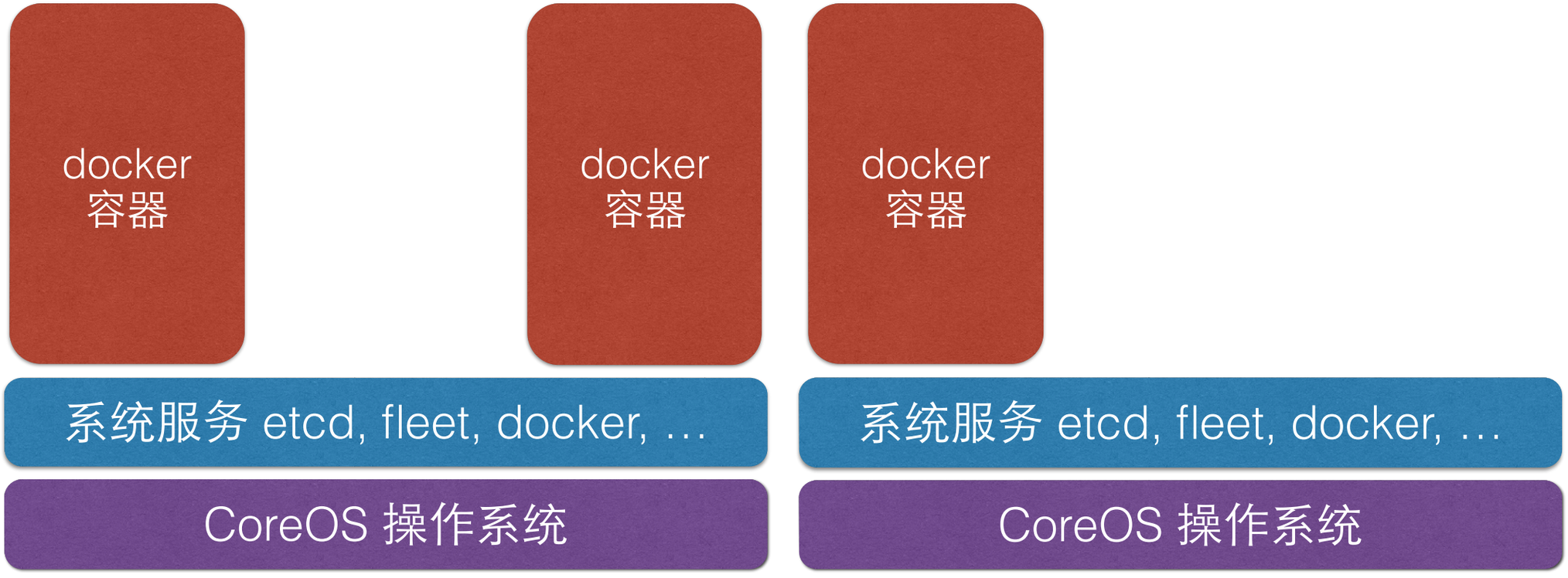



docker
容器
docker
容器
docker
容器
系统服务 etcd, fleet, docker, ...
CoreOS 操作系统

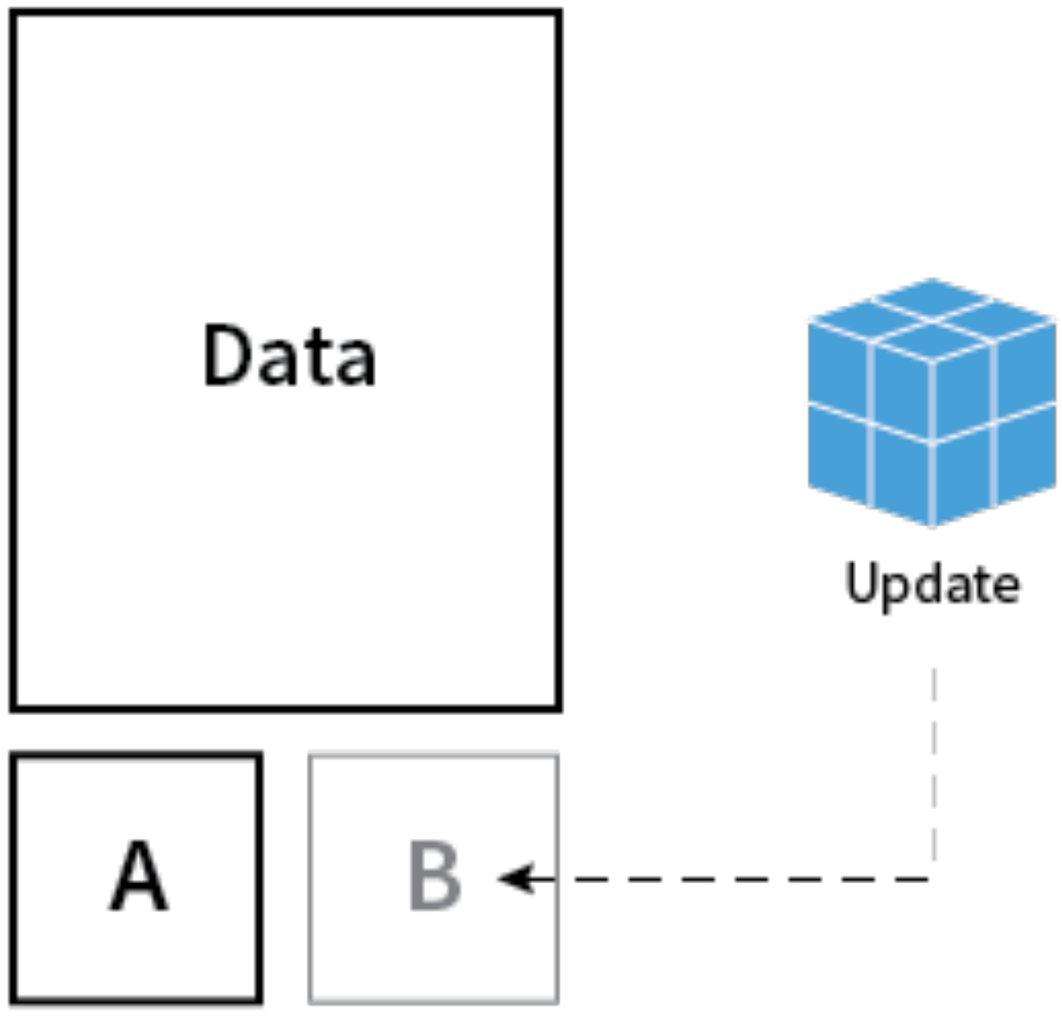

docker
容器
docker
容器
docker
容器
系统服务 etcd, fleet, docker, …
系统服务 etcd, fleet, docker, …
CoreOS 操作系统
CoreOS 操作系统

fleet 服务


docker
容器
docker
容器
docker
容器
系统服务 etcd, fleet, docker, …
CoreOS 操作系统
系统服务 etcd, fleet, docker, …
CoreOS 操作系统


docker
容器
docker
容器
docker
容器
系统服务 etcd, fleet, docker, …
系统服务 etcd, fleet, docker, …
CoreOS 操作系统
CoreOS 操作系统

【操作系统自动升级】
locksmith 服务器升级调度
双系统分区 集群快速升级



根分区
/

运行
系统分区

备用
系统分区

根分区
/

运行
系统分区

备用
系统分区

/usr
/bin
/sbin
/lib
/lib64

CoreOS.com

根分区
/

运行
系统分区

v485.0.0

备用
系统分区

v485.0.0

CoreOS.com

Stable通道的
最新版本是?

根分区
/

运行
系统分区

v485.0.0

备用
系统分区

v485.0.0

CoreOS.com

根分区
/

运行
系统分区

备用
系统分区

v485.0.0

v485.0.0

版本号 v494.0.0

CoreOS.com

根分区
/

运行
系统分区

备用
系统分区

v485.0.0

版本升级

CoreOS.com

根分区
/

运行
系统分区

备用
系统分区

v485.0.0

v494.0.0

CoreOS.com

根分区
/

备用
系统分区

运行
系统分区

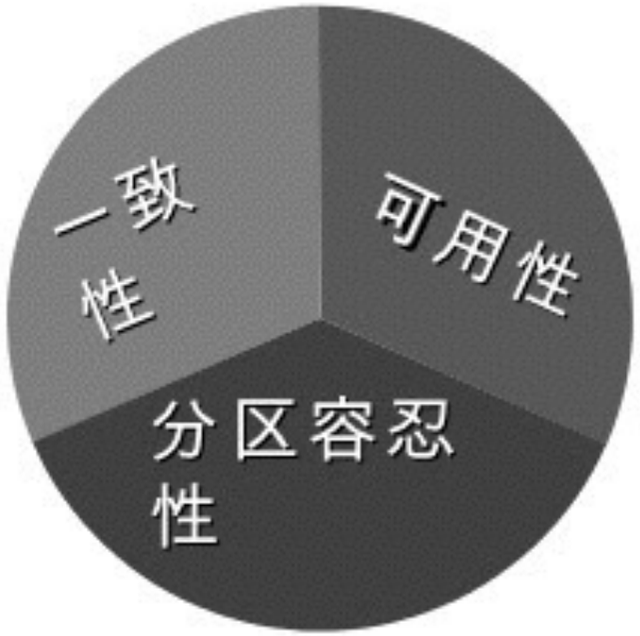
v485.0.0

v494.0.0

CoreOS 是完美的吗

CoreOS 是完美的吗

分布式系统的 CAP 理论



C consistency 一致性

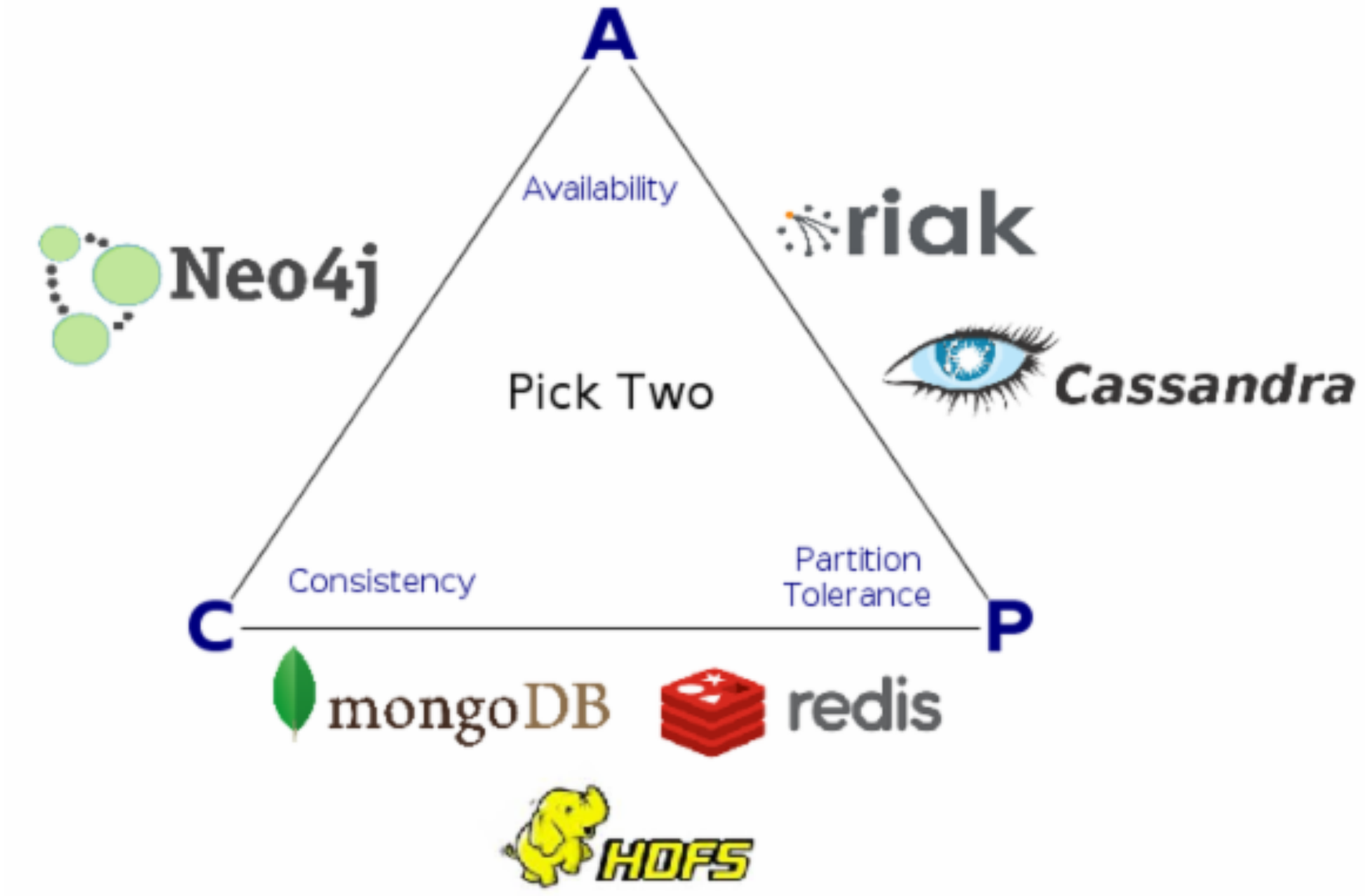
A availability 可用性

P tolerance of network partition 分区容错性

一个分布式系统不可能满足一致性、可用性、分区容错性这三个需求；最多只能同时满足两个。

主流NoSQL数据库实现

Where is CoreOS' ectd?

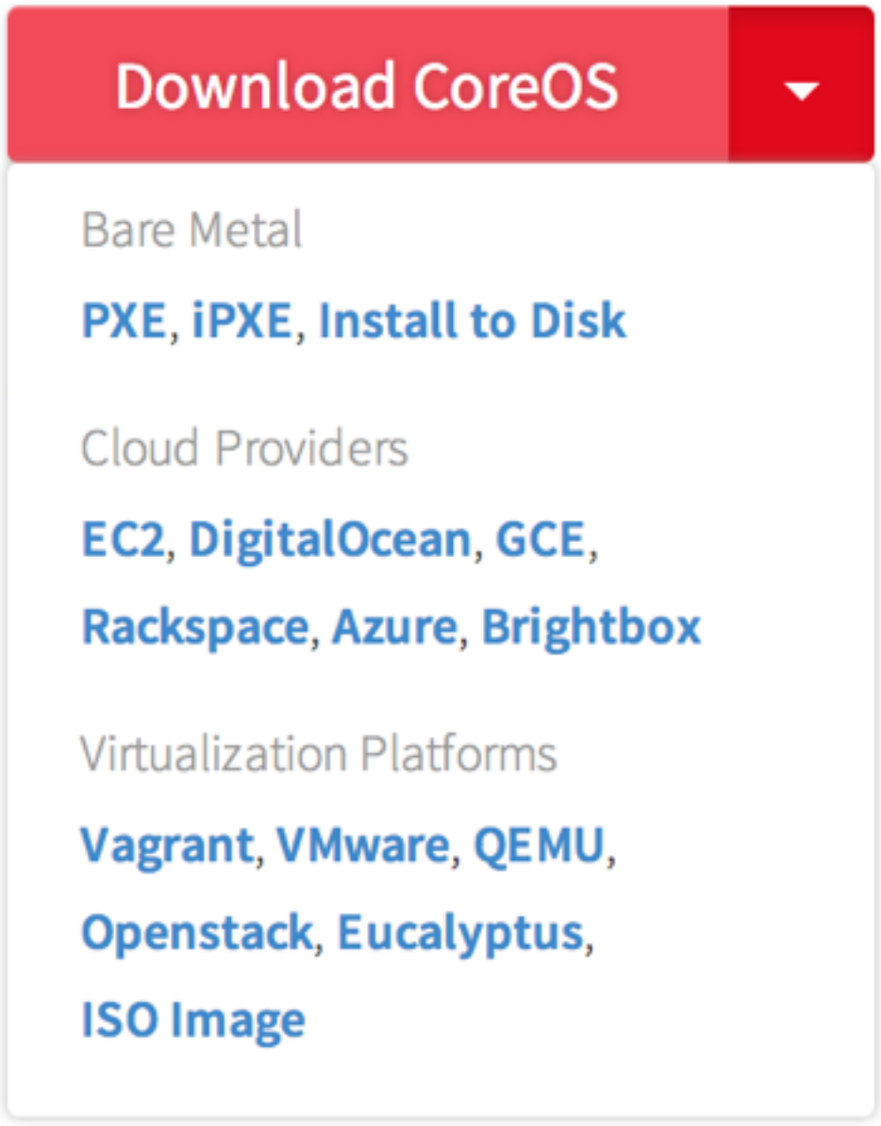


Etcd leader 存在潜在的单点依赖

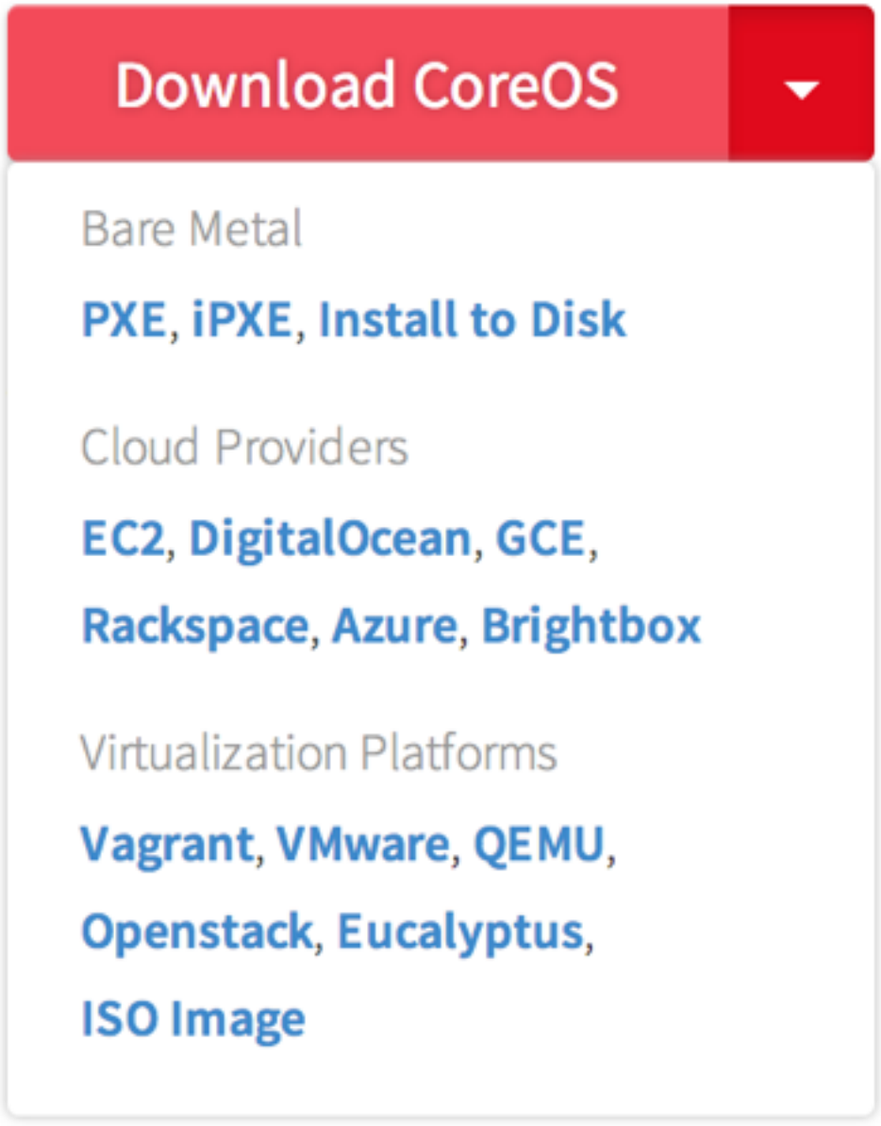
Etcd leader 存在潜在的单点依赖

建议：在设计软件时应考虑访问失败的异常处理

cloud-init 依赖特定平台定制



cloud-init 依赖特定平台定制



建议：尽可能在官方推荐的平台上使用 CoreOS

缺少软件包管理器

缺少软件包管理器

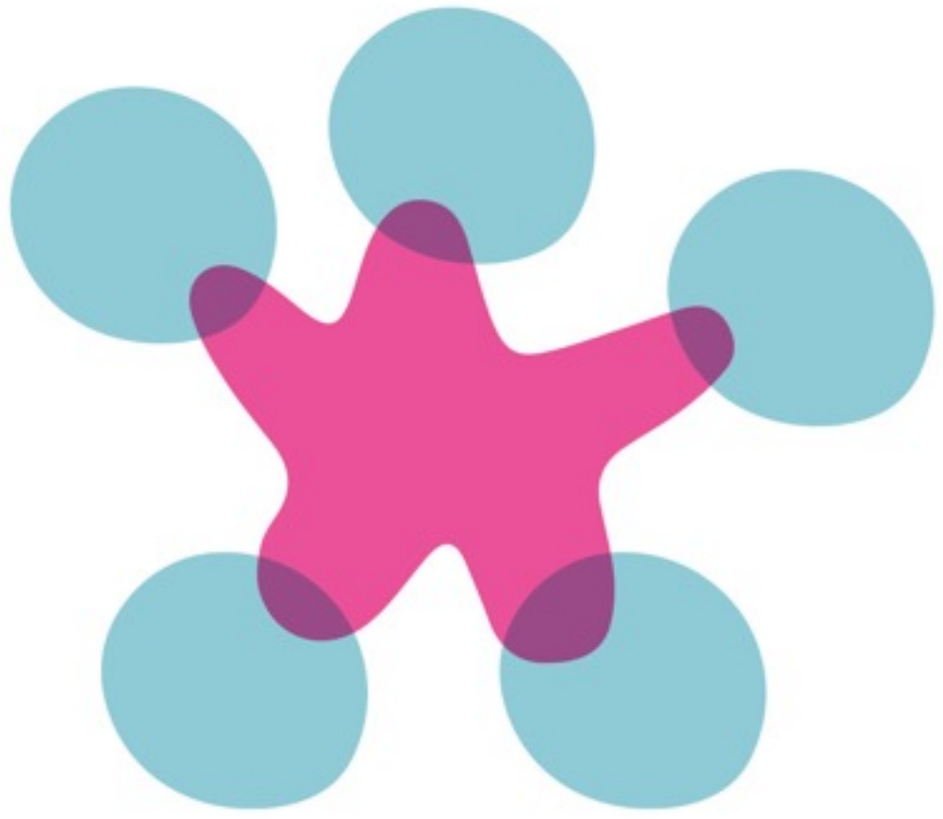
建议：CoreOS 官方的 toolbox 工具能部分解决问题

CoreOS 的 systemd 与 docker 存在兼容问题

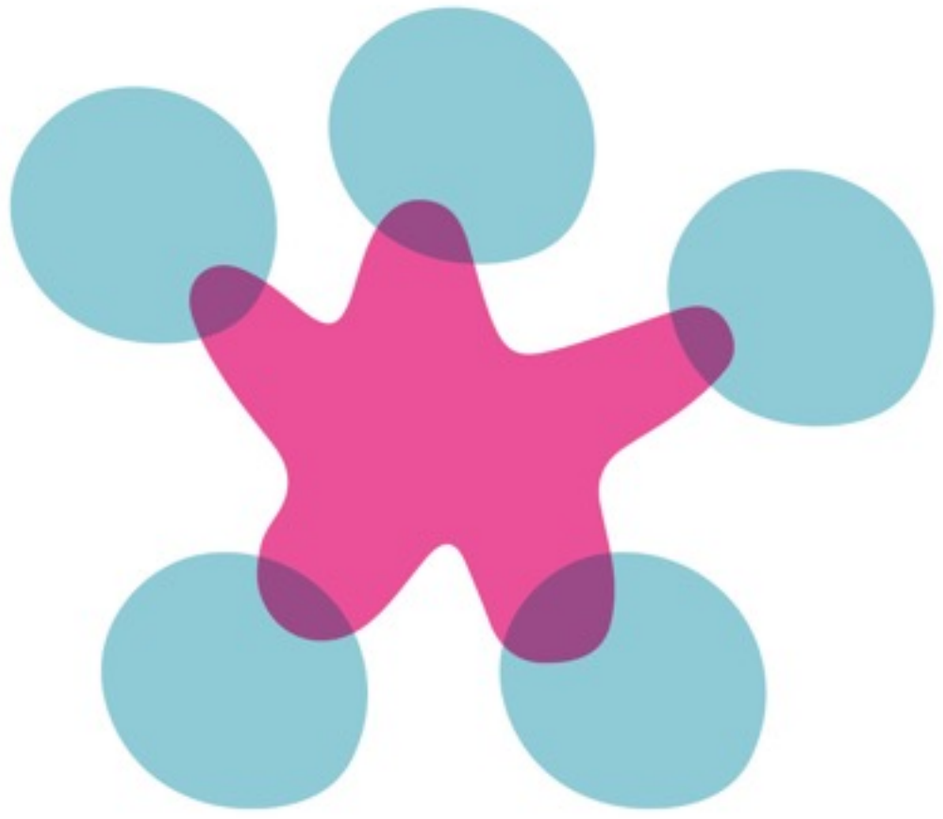
CoreOS 的 systemd 与 docker 存在兼容问题

建议：1) 避免向 systemd 管理的 docker 容器发信号
2) 使用第三方工具 systemd-docker

集群对节点的识别基于 IP 区分
集群内出现重复 IP 将造成不易修复的后果



集群对节点的识别基于 IP 区分
集群内出现重复 IP 将造成不易修复的后果



建议：1) 使用地址数量较多内网 IP 地址段，如 10.x.x.x
2) 配置DHCP，避免新节点复用存在过的 IP地址

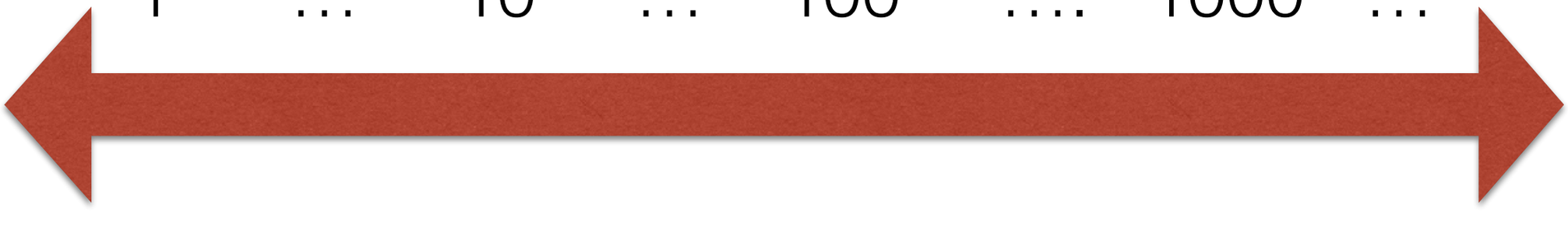
CoreOS

案例和基于当下的总结

我们什么时候应该考虑使用 CoreOS

集群的节点规模

1 … 10 … 100 …. 1000 …



Ubuntu Mint CentOS RedHat CoreOS

SUSE Fedora Arch Debian

运维团队的工具自动化程度

人工管理 … 半自动化 … 高度自动化

Ubuntu

Mint

CentOS

RedHat

CoreOS

SUSE

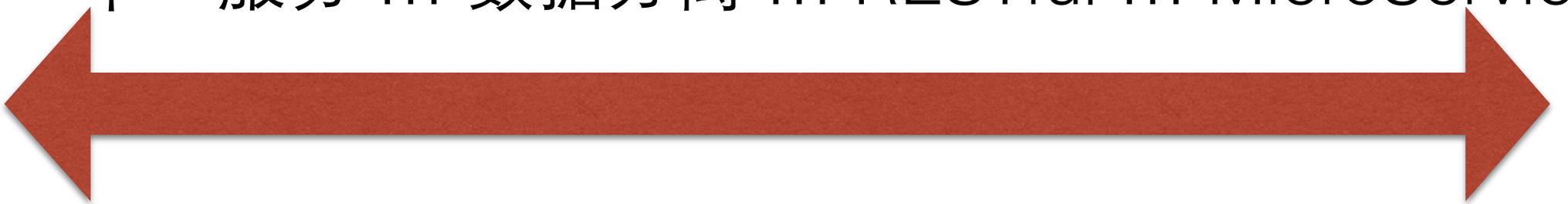
Fedora

Arch

Debian

服务的独立程度

单一服务 … 数据分离 … RESTful … MicroService



Ubuntu

Mint

CentOS

RedHat

CoreOS

SUSE

Fedora

Arch

Debian

数据持久性和安全性的需求

应用服务器 … 运算服务器 … 数据库服务器

CoreOS

Ubuntu

Mint

CentOS

RedHat

SUSE

Fedora

Arch

Debian

2014年9月美国主流云服务商之一的 DigitalOcean 将 CoreOS 作为继 Ubuntu、Debian、CentOS 和 RedHat 系统外的第5款推荐操作系统。

Ubuntu Debian CentOS RedHat CoreOS

Google Cloud、AWS、Azure 和 RackSpace 等云服务商已经宣布提供 CoreOS 镜像。

国内的云服务商步伐比较慢，只有部分服务商，如 UStack (www.ustack.com) 开始提供 CoreOS 镜像支持。

有多少产品已经在使用 CoreOS 系统呢?
这个诞生仅仅一年的年轻系统未来路还很长

运行在 CoreOS 上的 3亿 Docker 实例

http://www.iron.io/

- Etcd 系统级分布式数据存储共享服务
  https://github.com/coreos/etcd

- Fleet 集群管理工具
  https://github.com/coreos/fleet

- 自组网服务端和Cloud-init节点端配置
  https://github.com/coreos/discovery.etcd.io
  https://github.com/coreos/coreos-cloudinit

- Locksmith 基于 etcd 的升级调度服务
  https://github.com/coreos/locksmith

- 解决 CoreOS 安装和使用 tcpdump 等系统工具的镜像
  https://github.com/coreos/toolbox

- 解决 systemd 和 docker 兼容问题的第三方工具
  https://github.com/ibuildthecloud/systemd-docker