

Linux 高可用群集体系结构



概述

- ▶ Linux群集发展史
- ▶ Linux HA 体系结构
- ▶ 高可用解决方案概述

◆ Linux群集发展史

- ▶ 百花齐放的开源群集解决方案
- ▶ Pacemaker+Corosync
- ▶ Pacemaker+Corosync+pcs
- ▶ 与其他方案的组合
 - ▶ + CLVM2
 - ▶ + DRBD
 - ▶ + GFS2/OCFS

百花齐放的开源群集解决方案

HAproxy: Web 负载均衡解决方案

Codis: Redis 集群解决方案

LVS: Linux 虚拟服务器

MySQL Cluster: MySQL 集群

DRBD: 文件同步系统

LCMC: Linux 集群管理控制台

Hadoopy: Python 的 Hadoop 扩展

Seesaw: 负载均衡系统

ADSG-LBaaS-Driver: OpenStack LBaaS 驱动

Apache Aurora: Mesos 框架

Apache Ranger: Hadoop 集群权限框架

Redis-Migrate-Tool: Redis 集群迁移工具

Nomad: 集群管理器和调度器

Pound: 反向 HTTP 代理/负载均衡器和 SSL 封装器

Lazy balancer: 基于 Nginx 的负载均衡管理系统

Swiftly: 网络集群化代理服务器

Photon Controller: 集群管理系统

heartbeat: Linux 集群系统

Keepalived: 服务器状态监测

Gearman: 任务调度程序

Terracotta: 开源群集框架

Galera Load Balancer: 负载均衡器

Pen: 负载均衡器

OpenAIS: 集群框架的应用程序接口规范

Apache Mesos: 集群管理器

Pacemaker: 集群管理器

Coherence: 集群计算部件

synctool: 集群配置同步工具

DRBD: 管理控制台

Mesosphere Marathon:

mod_backhand: Apache 负载均衡模块

Apache Ambari: Hadoop 管理监控工具

Corosync: Linux 高可用集群

openMosix: Linux 的群集技术

KTCPVS: TCP 负载均衡器

Zen Load Balancer:

JSession: java web 集群软件

Apache Helix: 集群管理框架

Perlbal:

norbert: CS 模式的集群的 JAVA API

G5: 通讯转发、(负载均衡)通讯分发器

Crossroads: 负载均衡器

mod_cluster:

OpenSSI: Linux 集群解决方案

OAQL: 负载均衡服务器

SessionContainer: java web 集群软件

TrueCL: 集群解决方案

Riverdrums: 负载均衡器

ogslb: DNS 轮询系统

SuperMon: 高速集群监控系统

GreatTurbo Load Balance Server:

HA-Tools:

Plasma: Map/Reduce 框架

Cloudbreak: 基于 Hadoop 的 Docker service API

G6: 负载均衡器 G5 的第二版

nginx-lua-ds-loadbalancer:

ldirectord: 集群服务器管理

izBalancing: 负载均衡脚本

pymesos: Mesos 调度器和执行器

GCMT: Gentoo 集群管理工具

Traefik: HTTP 反向代理、负载均衡软件

High Availability Toolkit: 高可用技术工具包

JBoss Clustering:

L3DSR: 负载均衡服务器

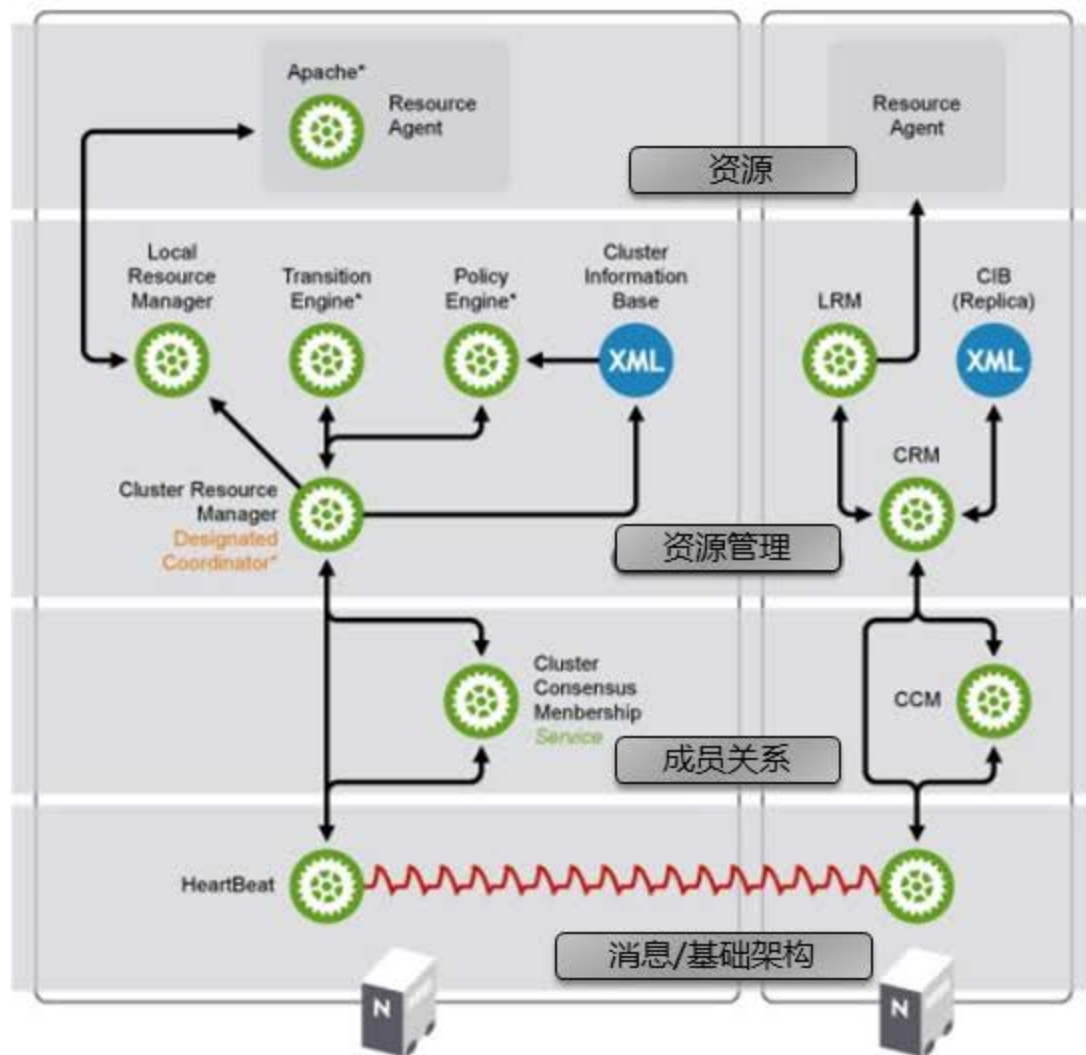
ocelli: 交互负载均衡程序

Redis Cluster Unofficial: 认证模式下的自动化工具

crmsh: 高可用集群管理工具

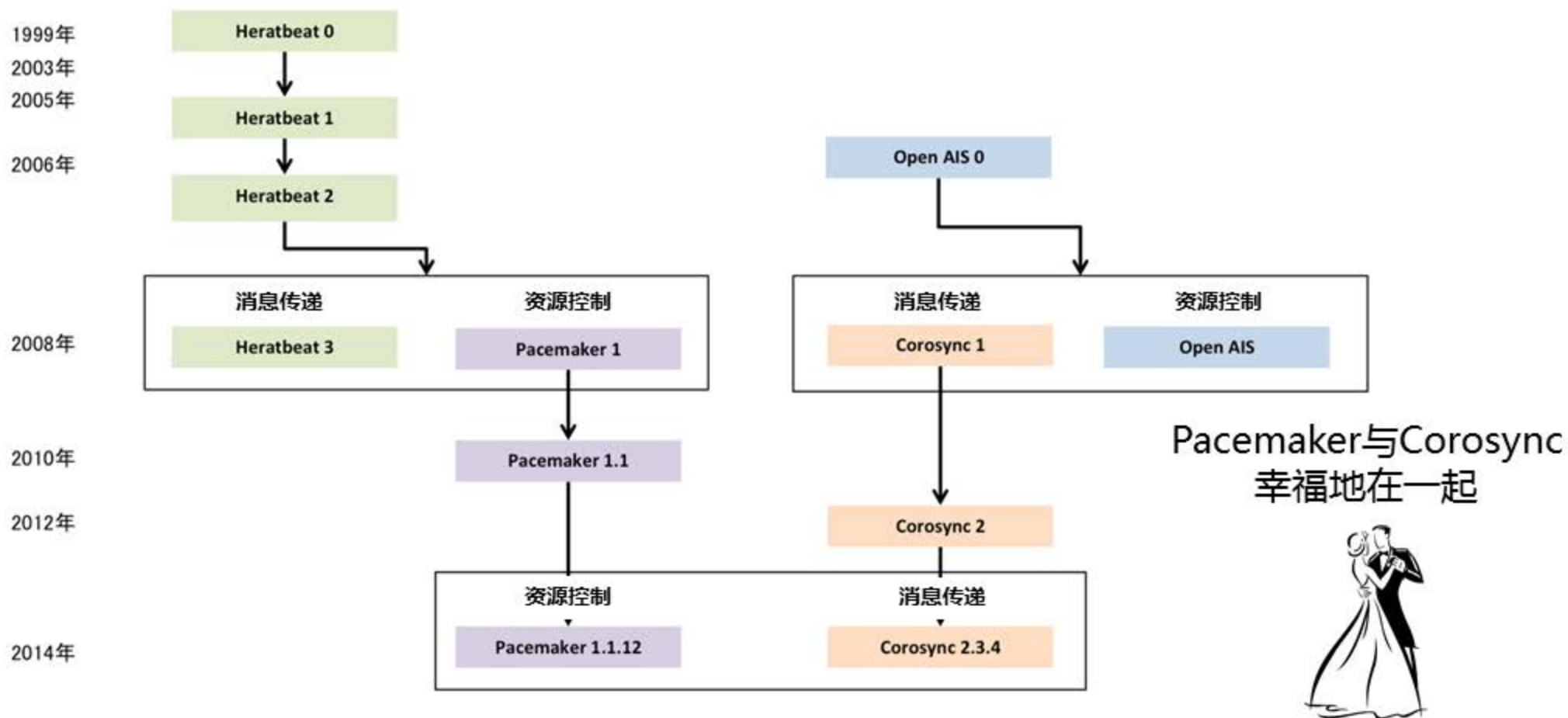
Minuteman: 负载均衡器

高可用集群的层次结构



Pacemaker+Corosync

▶ 出处：http://openstandia.jp/oss_info/corosync/ 日文版



与其他方案的组合

▶ 管理工具

- ▶ crmsh: crm (Cluster Resource Manager) shell
- ▶ pcs: Pacemaker Corosync Server

▶ CLVM

- ▶ 集群化的 LVM (Clustered LVM , CLVM)
- ▶ LVM 的一个集群方面的扩展。允许集群中的计算机通过 LVM 管理共享存储。

▶ DRBD

- ▶ Distributed Replicated Block Device(DRBD)
- ▶ 软件实现的、无共享的、服务器之间镜像块设备内容的存储复制解决方案

▶ GFS2/OCFS

- ▶ Global File System 2 / Oracle Cluster File System
- ▶ 集群文件系统使用集群中所有节点并发的通过标准文件系统接口来访问存储设备

◆ Linux HA 体系结构

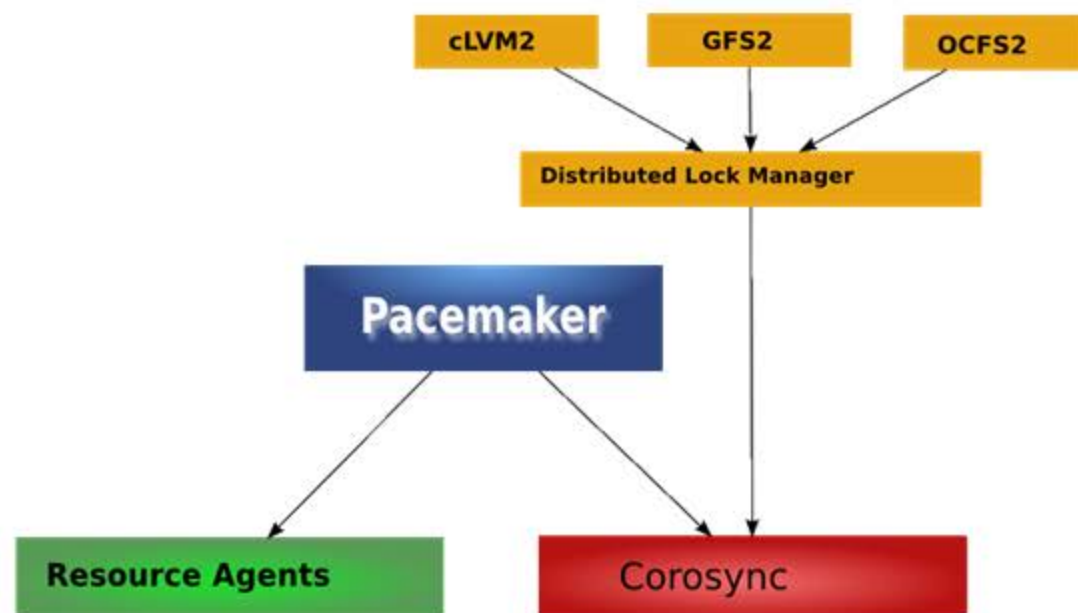
- ▶ Pacemaker体系结构
- ▶ Pacemaker内部组件
- ▶ Corosync
- ▶ Fencing Agent
- ▶ Resource Agent
- ▶ STONITH

注：本处的Linux HA是指Pacemaker+Corosync组合

Pacemaker体系结构



集群概念层次总览

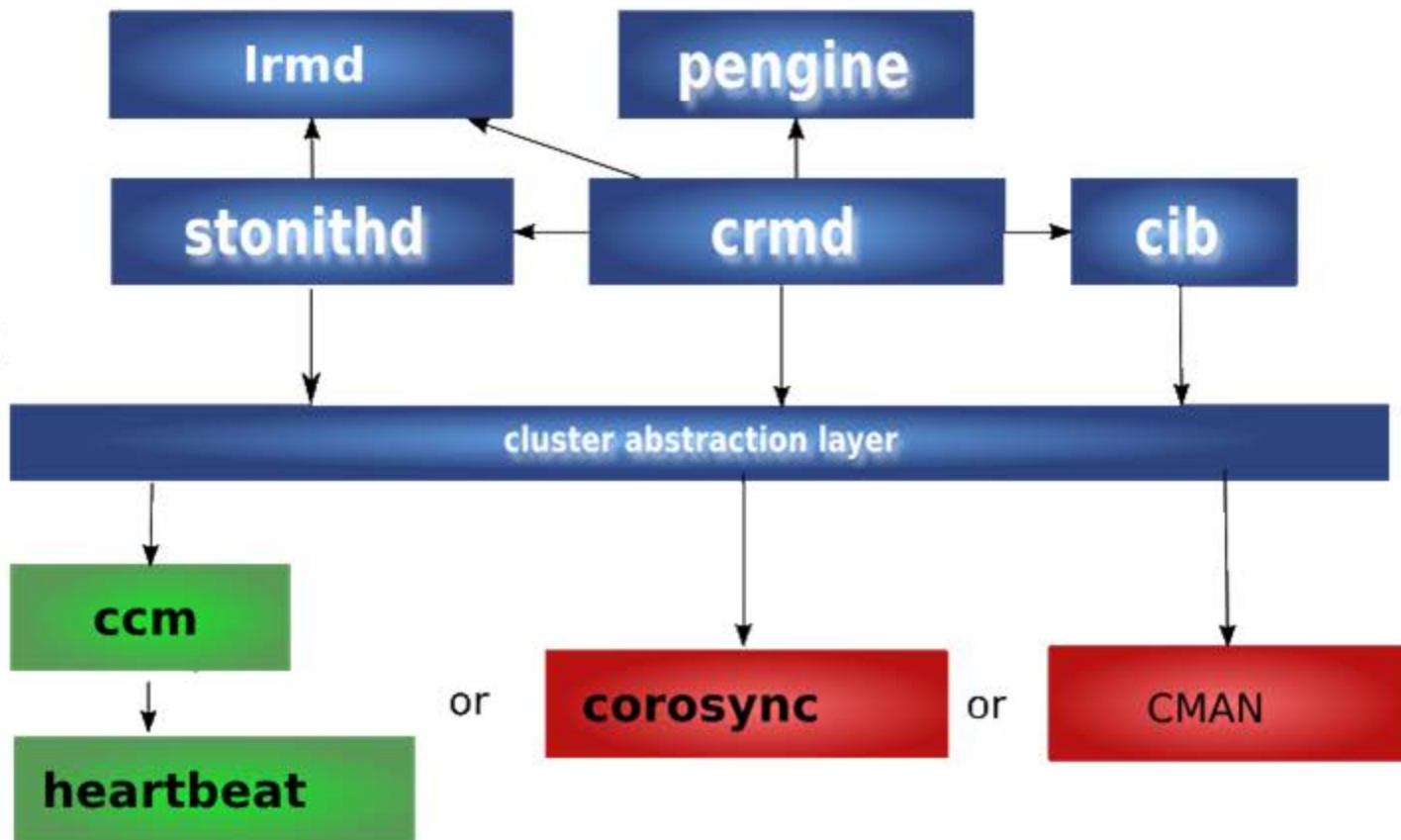


Pacemaker采用Corosync时的架构

Pacemaker内部组件

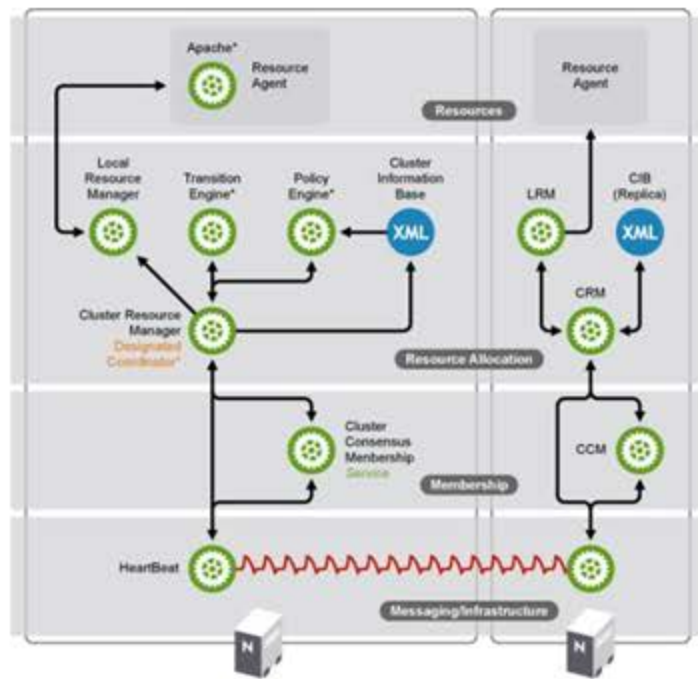
四个关键组件组成:

- ▶ CIB (集群信息基础)
- ▶ CRMd (集群资源管理守护进程)
- ▶ PEngine (PE or 策略引擎)
- ▶ STONITHd
 - ▶ Shoot-The-Other-Node-In-The-Head
 - ▶ 爆其他节点的头



Corosync

- ▶ Corosync 位于高可用集群架构中的底层，为各节点（node）之间提供心跳信息传递这样的角色
- ▶ 心跳：
 - ▶ 就是将多台服务器用网络连接起来，而后每一台服务器都不停的将自己依然在线的信息使用很简短很小的通告给同一个网络中的其它主机，告诉它们自己依然在线，其它服务器收到这个心跳信息就认为它是在线的，尤其是主服务器。

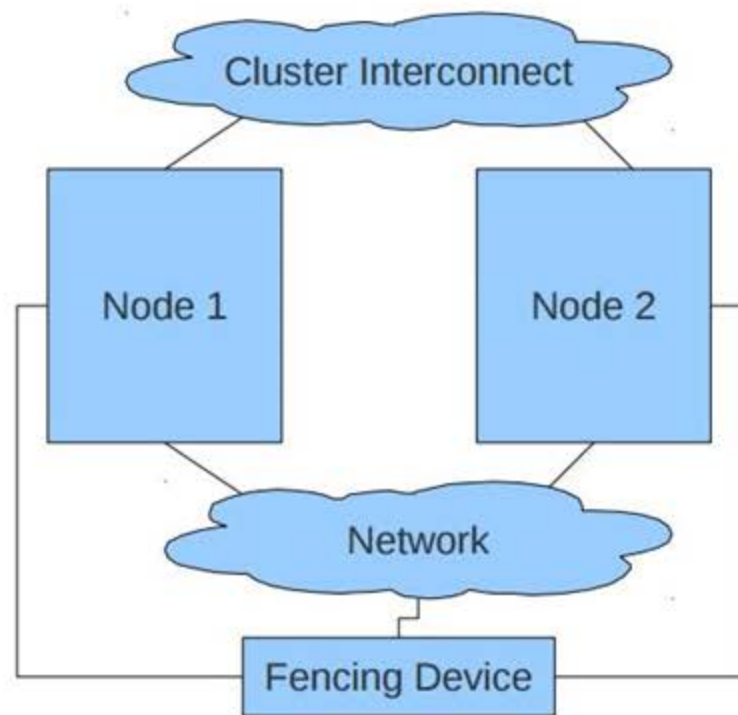


资源代理 (Resource Agent - RA)

- ▶ RA 是管理一个集群资源的可执行程序，没有固定其实现的编程语言，但是大部分RA都是用 shell 脚本实现的。
- ▶ Pacemaker 使用 RA 来和受管理资源进行交互，它既支持它自身实现的70多个RA，也支持第三方RA
- ▶ 要实现一个 RA，需要遵守 OCF 的规范
- ▶ Pacemaker 支持三种类型的 RA：
 - ▶ LSB Resource Agents
 - ▶ OCF Resource Agents
 - ▶ legacy Heartbeat Resource Agents
- ▶ RA 支持的主要操作包括：
start、stop、monitor、validate-all、meta-data...

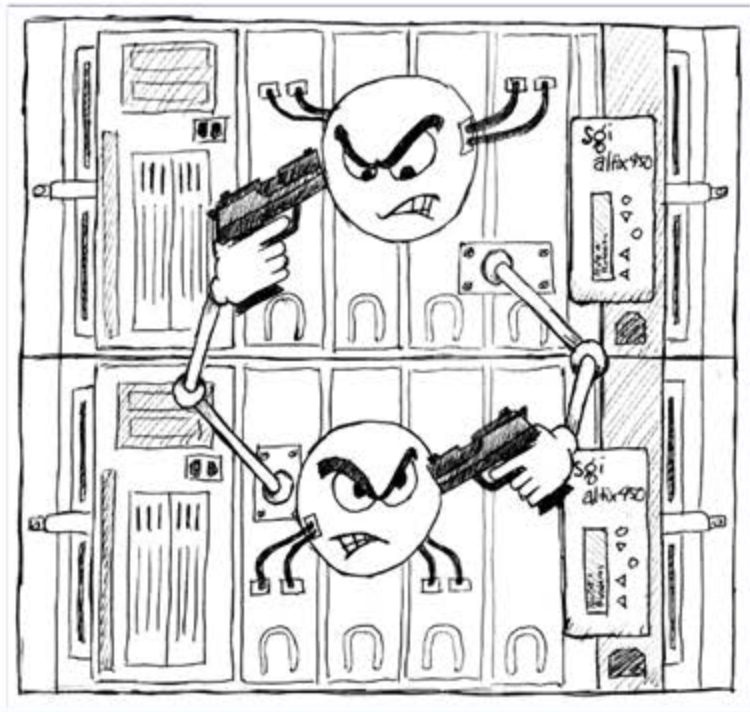
Fencing Agent

- ▶ 集群通常需要使用 Fencing agent
- ▶ Fencing 是在一个节点不稳定或者无答复时将其关闭，使得它不会损坏集群的其它资源，其主要用途是消除脑裂。
- ▶ 常见Fencing agent的类型
 - ▶ Power 类型的 Agent
会将节点的电源断电，它通常连到物理的设备比如 UPS
 - ▶ Storage 类型的Agent
会确保某个时刻只有一个节点会读写共享的存储。



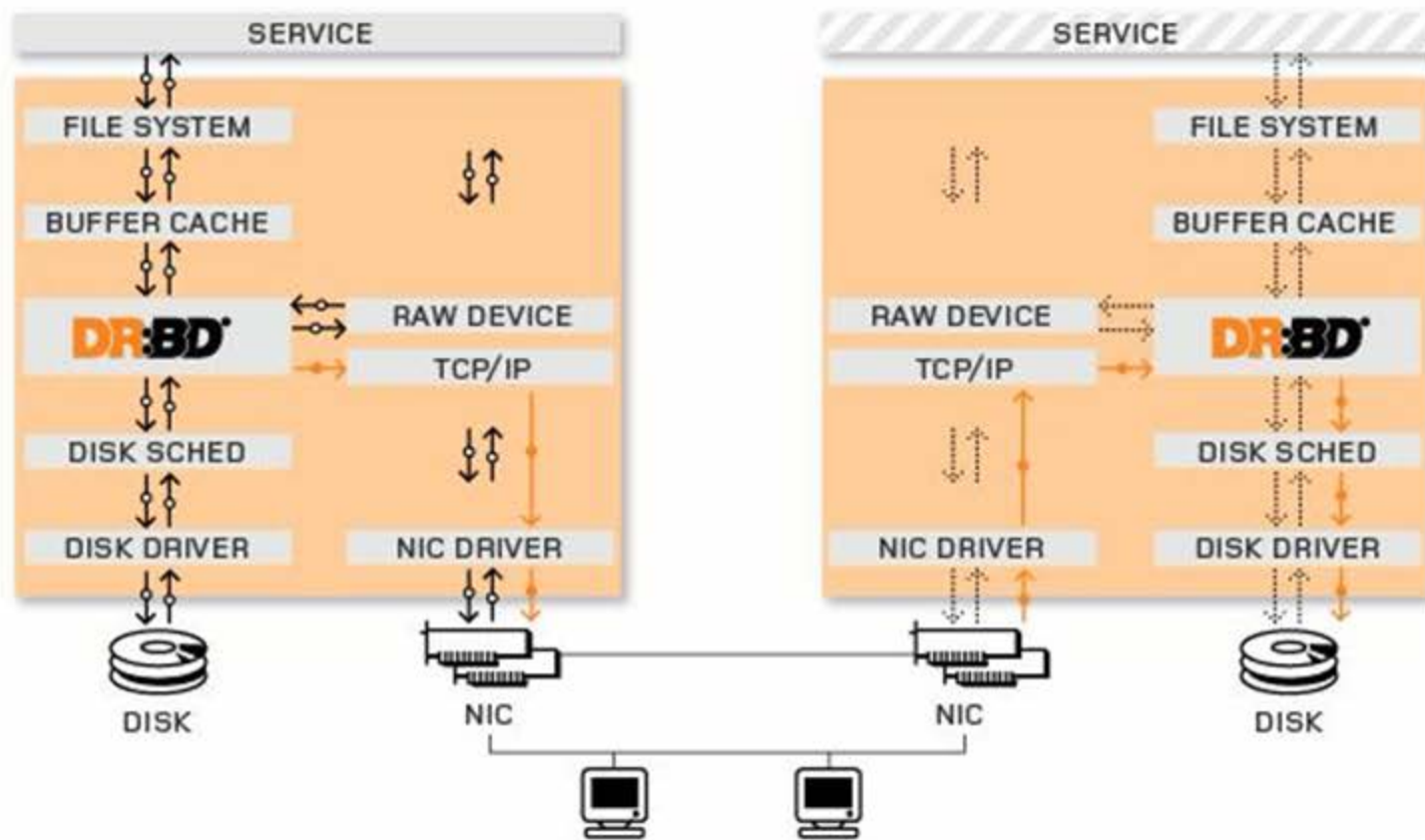
STONITH

- ▶ STONITH 是 “shoot the other node in the head” 的首字母简写
- ▶ 它是 Heartbeat 软件包的一个组件
- ▶ 它允许使用一个远程或 “智能的” 连接到健康服务器的电源设备自动重启失效服务器的电源，
- ▶ stonith设备可以关闭电源并响应软件命令，运行Heartbeat的服务器可以通过串口线或网线向stonith设备发送命令，它控制高可用服务器对中其他服务器的电力供应：
 - ▶ 主服务器可以复位备用服务器的电源
 - ▶ 备用服务器也可以复位主服务器的电源



DRBD

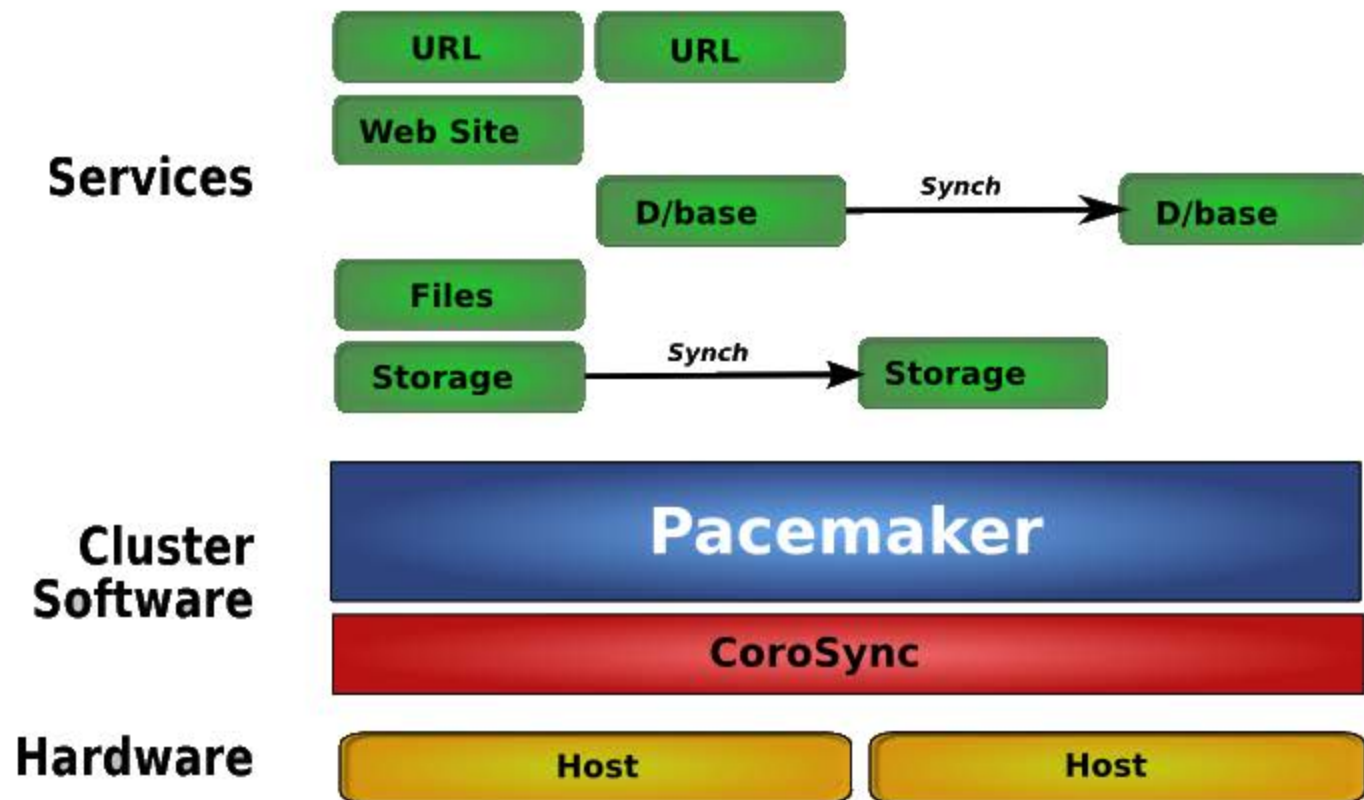
- ▶ DRBD (Distributed Replication Block Device) 即分布式复制块设备



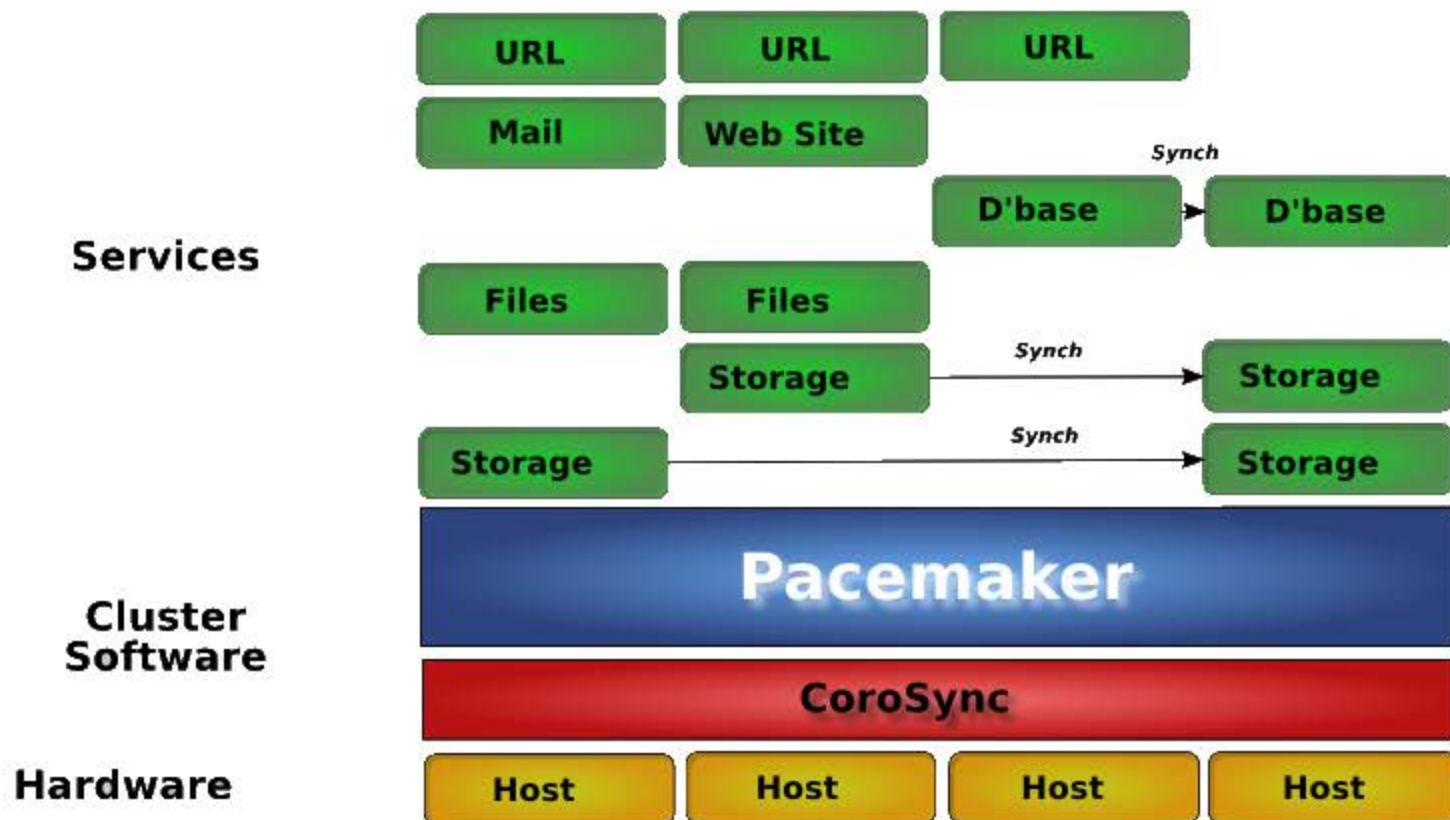
◆ 高可用解决方案概述

- ▶ Active/Passive
- ▶ Shared Failover
- ▶ N to N Redundancy

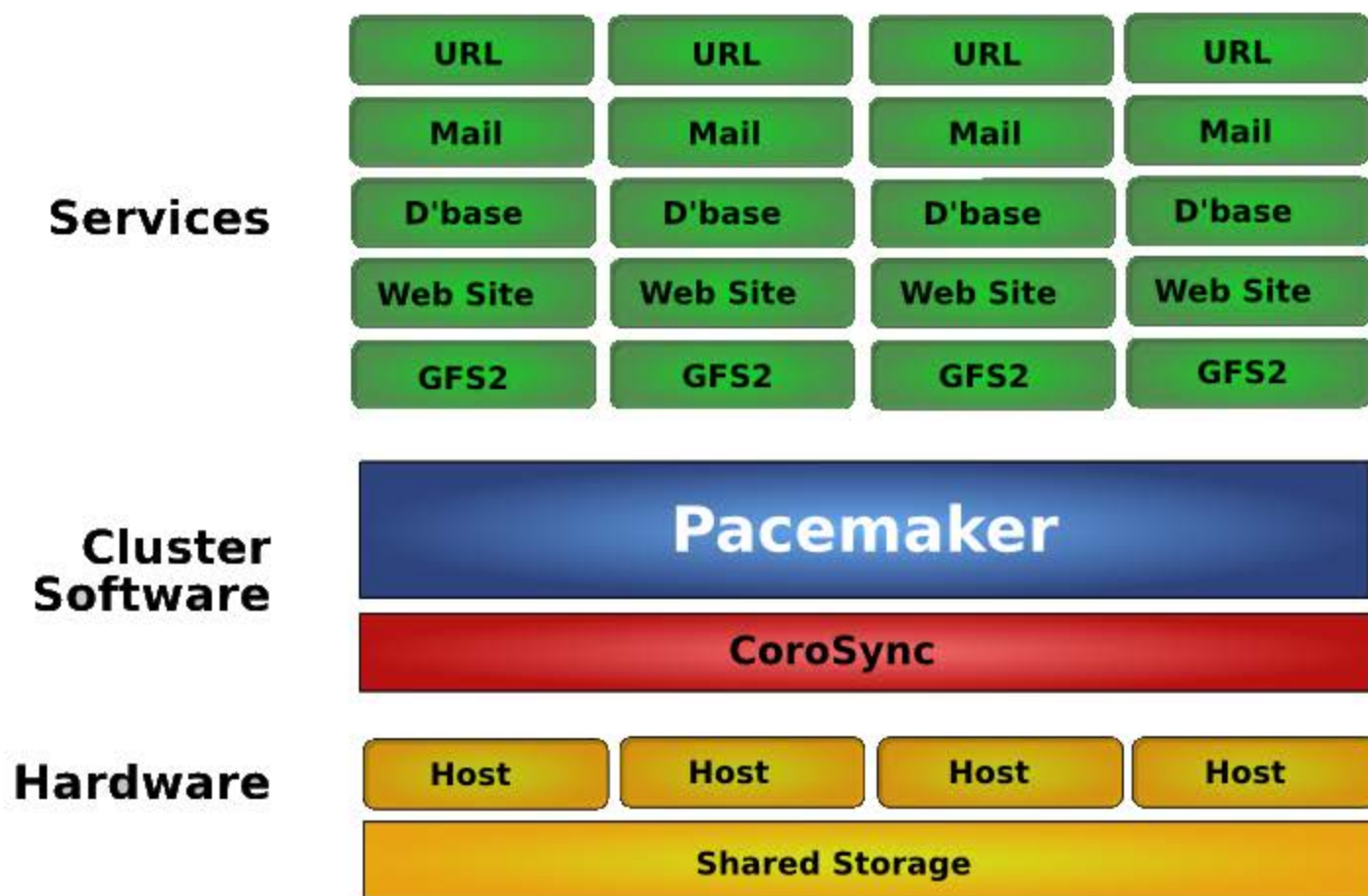
Active/Passive Redundancy



Shared Failover



N to N Redundancy



总结

- ▶ Linux群集发展史
- ▶ Linux HA 体系结构
- ▶ 高可用解决方案概述