



Ubuntu Server 从入门到精通

第20章：集群



# 集群

三个臭皮匠，顶个诸葛亮

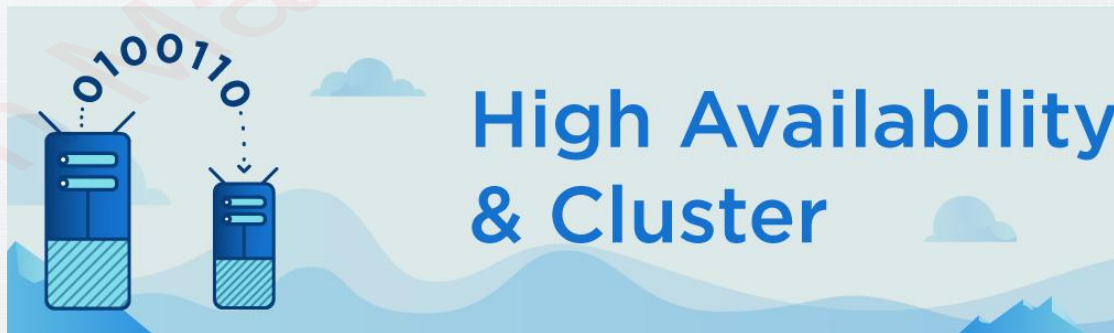
- 保证业务可用性是一切工作的基础
- 丧失可用性的安全毫无意义
- 消灭单点故障隐患是提高可用性的主要手段
- 从单机、多机再到网络所有节点的冗余
- 多机实现容错通常被称为集群 / 群集 (Cluster)
- 负载均衡群集 / HA / 服务器群集 (关键在于是否同步数据)
  - DNS、Keepalive



# 集群

三个臭皮匠，顶个诸葛亮

- 服务器集群
  - 共享存储（存储设备价格较贵）
  - 基于块设备复制（趋势、廉价）
- 群集资源（群集中所有服务器共同享有的一组资源）
  - 域名
  - 地址
  - 存储
  - 硬件
  - 路径
  - 名称
  - 在主机间漂移



# 集群

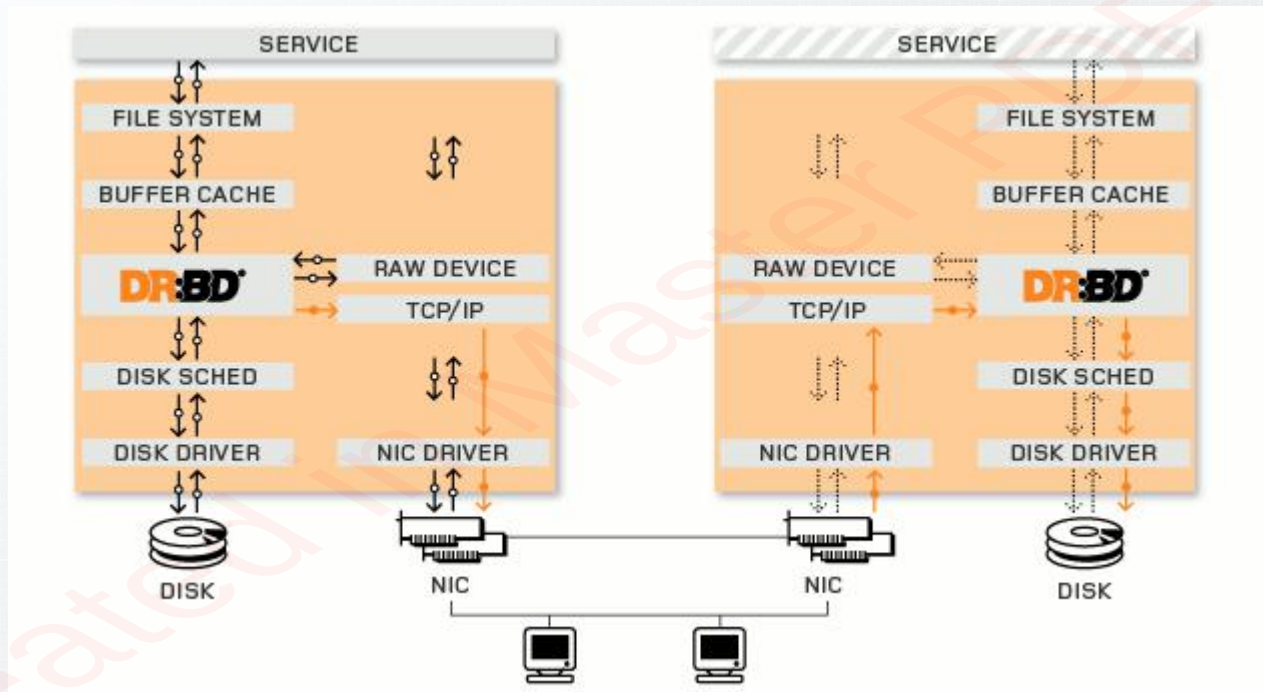
三个臭皮匠，顶个诸葛亮

- 主机之间基于块设备复制
- 保持主机间的数据同步是一切的基础
- DRBD — Distributed Replicated Block Device
  - 分布式块设备复制技术
  - 由 linbit 开源（培训、服务收费）
  - 由 Linux 内核模块 + 用户空间管理工具组成
  - 实现服务器间块设备的同步镜像（跨主机的 RAID 1）
  - 通常采用主 / 备模式部署（Primary / Slave）



# 集群

三个臭皮匠，顶个诸葛亮





# 集群

三个臭皮匠，顶个诸葛亮

- 保证数据一致性（协议 / 方法）
  - A：异步复制，本地写成功即确认，发送buffer中的数据可能丢失
  - B：内存同步，本地写入并发送成功即确认，如掉电，则数据丢失
  - C：同步复制，节点全部写入成功后确认
  - C 可最高确保数据同步，但写入时间长，性能处于劣势
- 清除块设备数据
  - `sudo dd if=/dev/zero of=/dev/sdb1`



# 集群

三个臭皮匠，顶个诸葛亮

- DRBD 安装准备
  - Node 1 : usrv01
  - Node 2 : usrv02
  - 两节点各增加一块硬盘，用于镜像（大小必须一致）
    - 物理磁盘、分区、RAID、LVM、都可作为镜像块设备
  - IP 地址
  - 域名解析（/etc/hosts、hostname）
- 安装（两节点分别执行）
  - `sudo apt install -y ntp`
  - `sudo apt install drbd8-utils`
- 主配置文件
  - `/etc/drbd.conf`



# 集群

三个臭皮匠，顶个诸葛亮

```
global { usage-count no; }
common { syncer { rate 100M; } }
resource r0 {
    protocol C;
    startup {
        wfc-timeout 15;
        degr-wfc-timeout 60;
    }
    net {
        cram-hmac-alg sha1;
        shared-secret "secret";
    }
}
```

```
on usrv01 {
    device /dev/drbd0;
    disk /dev/sdb1;
    address 1.1.1.1:7788;
    meta-disk internal;
}
on usrv02 {
    device /dev/drbd0;
    disk /dev/sdb1;
    address 1.1.1.2:7788;
    meta-disk internal;
}
}
```



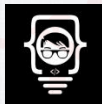




# 集群

三个臭皮匠，顶个诸葛亮

- 复制配置文件
  - `scp /etc/drbd.conf yuanfh@usrv02:~`
- 初始化元数据存储（两节点分别执行）
  - `sudo drbdadm create-md r0`
- 启动服务
  - `sudo systemctl start drbd.service`
  - `sudo systemctl enable drbd`
  - `sudo drbdadm up r0` # 两节点初始都是备用节点
- 指定主节点（在一个节点上执行）
  - `sudo drbdadm -- --overwrite-data-of-peer primary all`
- 观察同步进程（在另一节点上执行）
  - `watch -n1 cat /proc/drbd`



# 集群

三个臭皮匠，顶个诸葛亮

- 查看每个节点的状态
  - lsblk # 147
  - sudo drbd-overview
  - sudo cat /proc/drbd
- 格式化并挂载
  - sudo mkfs.ext4 /dev/drbd0
  - sudo mount /dev/drbd0 /var/www/html
- 测试（主节点）
  - sudo cp index.html /var/www/html
  - sudo umount /var/www/html
  - sudo drbdadm secondary r0





# 集群

三个臭皮匠，顶个诸葛亮

- 测试（备节点）
  - `sudo drbdadm primary r0`
  - `sudo mount /dev/drbd0 /var/www/html`
  - 节点主备不会自动切换
  - 不可同时做主并挂载
  - 需要群集管理器来实现资源的在不同节点间自动切换



# 集群

三个臭皮匠，顶个诸葛亮

- Pacemaker

- 集群资源管理器（CRM - Cluster Resource Manager）
- 开源，适合各种大小（类型）集群管理（主/主、主/备）
- 基于资源级别的监测和恢复保证应用最大可用性
- 基础组件(Corosync、Heartbeat)实现集群各成员间通信和关系管理
- Corosync：消息层组件（心跳传输），管理成员关系、消息和仲裁
- 可使用共享存储或基于块设备的复制（DRBD）



# 集群

三个臭皮匠，顶个诸葛亮

- 安装 ( 2 Node )
  - `sudo systemctl stop / disable drbd`
  - `sudo umount /var/www/html`
  - `sudo drbdadm down r0`
  - `sudo apt-get install -y pacemaker`
- 配置 ( 2 Node )
  - [`sudo vi /etc/corosync/corosync.conf`](#)
- 启动服务
  - `sudo systemctl restart corosync`
  - `sudo systemctl start pacemaker`





# 集群

三个臭皮匠，顶个诸葛亮

- 查看群集状态

- `sudo crm status` # `sudo apt install crmsh`

- 配置和创建群集资源

- `sudo crm configure`

- `property stonith-enabled=false` # 使出现问题的节点从集群环境中脱离

- `property no-quorum-policy=ignore` # 集群选举需超半数（至少3节点）

- `primitive drbd_res ocf:linbit:drbd params drbd_resource=r0 op monitor interval=29s`  
`role=Master op monitor interval=31s role=Slave` # 定义磁盘存储资源

- `ms drbd_master_slave drbd_res meta master-max=1 master-node-max=1 clone-max=2`  
`clone-node-max=1 notify=true` # master/slave群集

- `primitive fs_res ocf:heartbeat:Filesystem params device=/dev/drbd0 directory=/var/www/html`  
`fstype=ext4` # 文件系统挂载点

- `colocation fs_drbd_colo INFINITY: fs_res drbd_master_slave:Master`

- `order fs_after_drbd mandatory: drbd_master_slave:promote fs_res:start`

- `commit / show / quit`





# 集群

三个臭皮匠，顶个诸葛亮

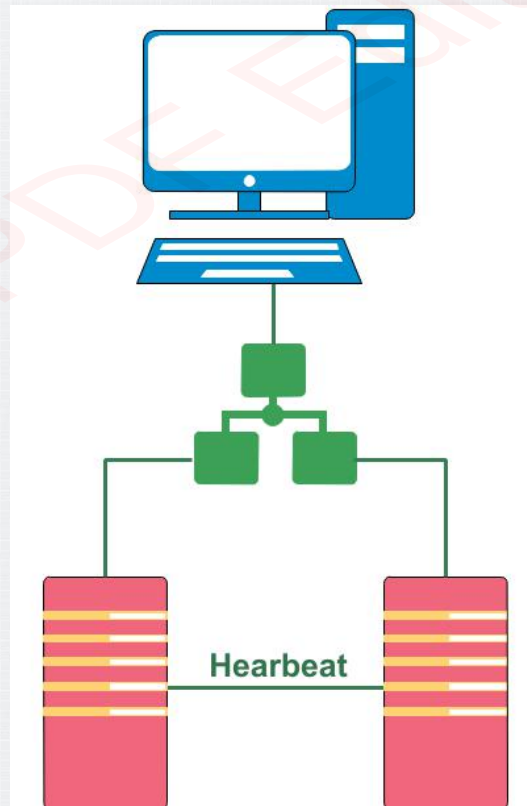
- 重启主节点测试自动切换



# 集群

三个臭皮匠，顶个诸葛亮

- 基于 heartbeat 创建 Mysql 群集
  - 主机添加第二块网卡，用作 DRBD 数据复制
- 安装
  - `sudo apt install drbd8-utils heartbeat -y`
- 启动服务
  - `sudo systemctl start drbd`
  - `sudo systemctl start heartbeat`
  - `sudo systemctl enable drbd`
  - `sudo systemctl enable heartbeat`



# 集群

三个臭皮匠，顶个诸葛亮

- 静态IP

network:

version: 2

renderer: networkd

ethernets:

eth0:

dhcp4: no

dhcp6: no

addresses: [192.168.1.70/24]

gateway4: 192.168.1.1

nameservers:

addresses: [192.168.1.1, 8.8.8.8, 8.8.4.4]



# 集群

三个臭皮匠，顶个诸葛亮

- 格式化硬盘
  - `sudo echo -e 'n\np\n1\n\n\nnw' | fdisk /dev/sdb`
- 群集配置文件
  - `/etc/drbd.conf` # P8
  - `sudo vi /etc/ha.d/ha.cf`  
keepalive 1 # Check Interval  
deadtime 10 # Time before server declared dead  
initdead 60 # Secondary wait delay at boot  
auto\_failback off # Auto-failback  
bcast enp0s8 # Heartbeat Interface  
node db1 # Nodes to monitor  
node db2





# 集群

三个臭皮匠，顶个诸葛亮

- 群集资源
  - /etc/ha.d/haresources
  - db1 1.1.1.1/24 drbddisk::r0 Filesystem::/dev/drbd0::/var/lib/mysql::ext4::noatime
- 身份认证
  - sudo vi /etc/ha.d/authkeys
  - auth1
  - 1 sha1 5baa61e4c9b93f3f0682250b6cf8331b7ee68fd8
- 权限
  - sudo chmod 600 /etc/ha.d/authkeys



# 集群

三个臭皮匠，顶个诸葛亮

## • 节点 1

- `sudo drbdadm create-md r0`
- `sudo systemctl restart drbd`
- `sudo drbdadm outdate r0`
- `sudo drbdadm -- --overwrite-data-of-peer primary all`
- `sudo drbdadm primary r0`
- `sudo mkfs.ext4 /dev/drbd0`
- `sudo chmod 600 /etc/ha.d/authkeys`
- `sudo mkdir /var/lib/mysql`



# 集群

三个臭皮匠，顶个诸葛亮

- 节点 2
  - `sudo drbdadm create-md r0`
  - `sudo systemctl restart drbd`
  - `sudo chmod 600 /etc/ha.d/authkeys`
  - `sudo mkdir /var/lib/mysql`
  - 等待硬盘同步完成
- 启动服务（BOTH）
  - `sudo systemctl start heartbeat`
- 验证IP、硬盘资源
  - 挂载与切换测试



# 集群

三个臭皮匠，顶个诸葛亮

- 安装 MariaDB
  - `sudo apt install mariadb-server`
  - `sudo mysql_secure_installation`
- 禁用 mysql 服务自动启动
  - 服务启动控制权交给集群
  - `sudo systemctl disable mysql`
- 停止备用节点服务 ( db2 )
  - `sudo systemctl stop mysql`
  - `sudo rm -rf /var/lib/mysql/*` # 来自主节点复制



# 集群

三个臭皮匠，顶个诸葛亮

- Mysql 远程访问
  - `sudo mysql`
  - `CREATE USER 'root'@'192.168.8.%' IDENTIFIED BY 'password';`
  - `GRANT ALL PRIVILEGES ON *.* TO 'root'@'192.168.123.%' WITH GRANT OPTION;`
  - `FLUSH PRIVILEGES;`
  - `QUIT;`
- 服务侦听地址
  - `sudo sed -i 's/127.0.0.1/0.0.0.0/g' /etc/mysql/mariadb.conf.d/*.cnf`
- 数据库服务资源加入群集
  - `db1 1.1.1.1/24 drbddisk::r0 Filesystem::/dev/drbd0::/var/lib/mysql::ext4::noatime mysql`





# 集群

三个臭皮匠，顶个诸葛亮

- 双节点重启服务（先1后2）
  - `sudo systemctl restart heartbeat`
- 测试
  - `mysql -h 192.168.8.3 -u root -p`  
create database test;  
quit;
  - 重启heartbeat服务模拟故障
  - 再次远程连接查看  
show databases;

# 远程连接、建库



# Questions ?

