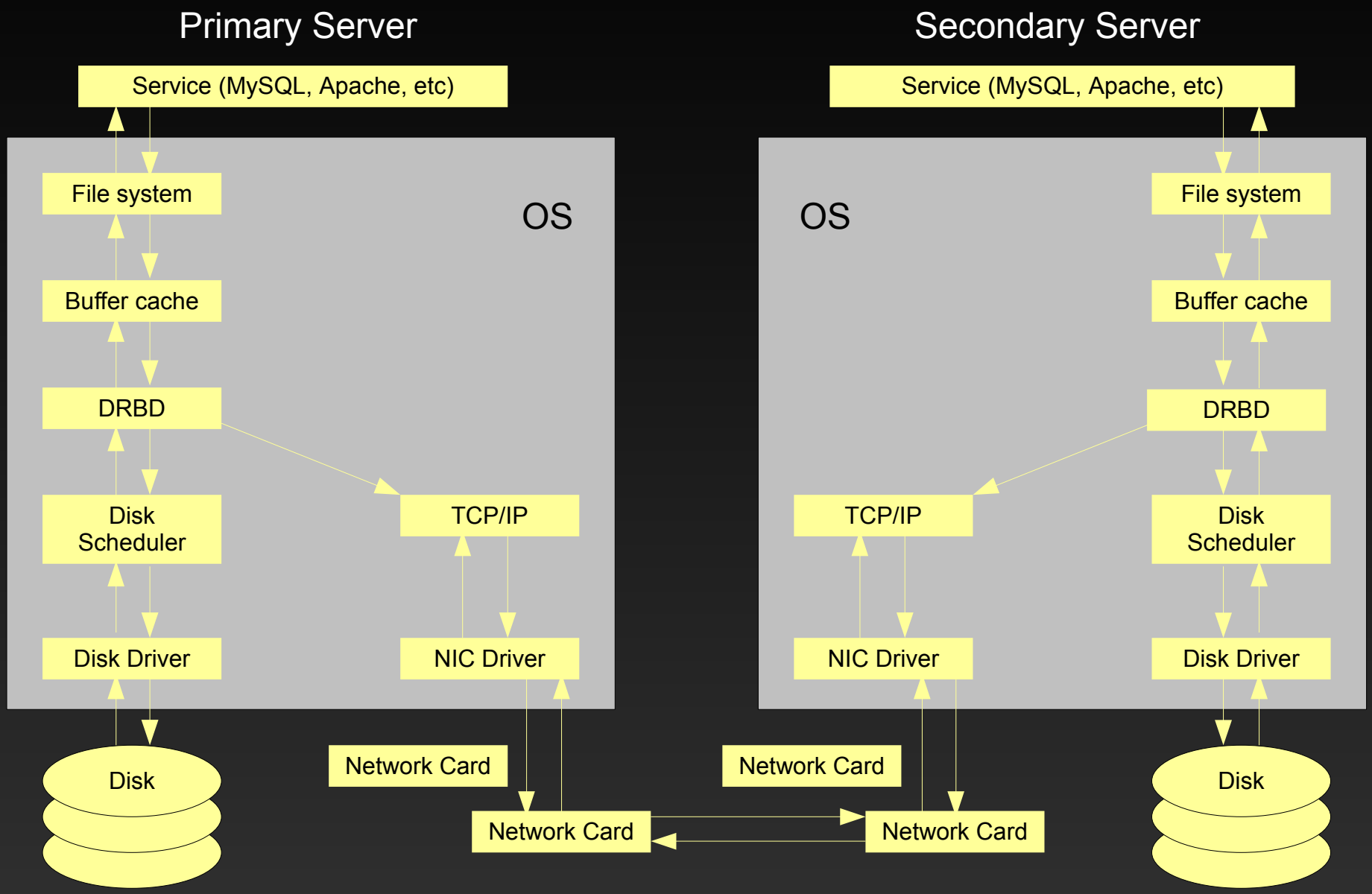
High Availability with DRBD & Heartbeat

Chris Barber
http://www.cb1inc.com

What is DRBD?

- Distributed Replicated Block Device
- RAID 1 mirror across network
- Realtime replications
- Linux-only kernel module
- 2 nodes – Primary/Secondary
- Single/Dual primary mode
- Open source, free

Visual Overview



Original chart: http://www.drbd.org/users-guide/ch-fundamentals.html

Uses

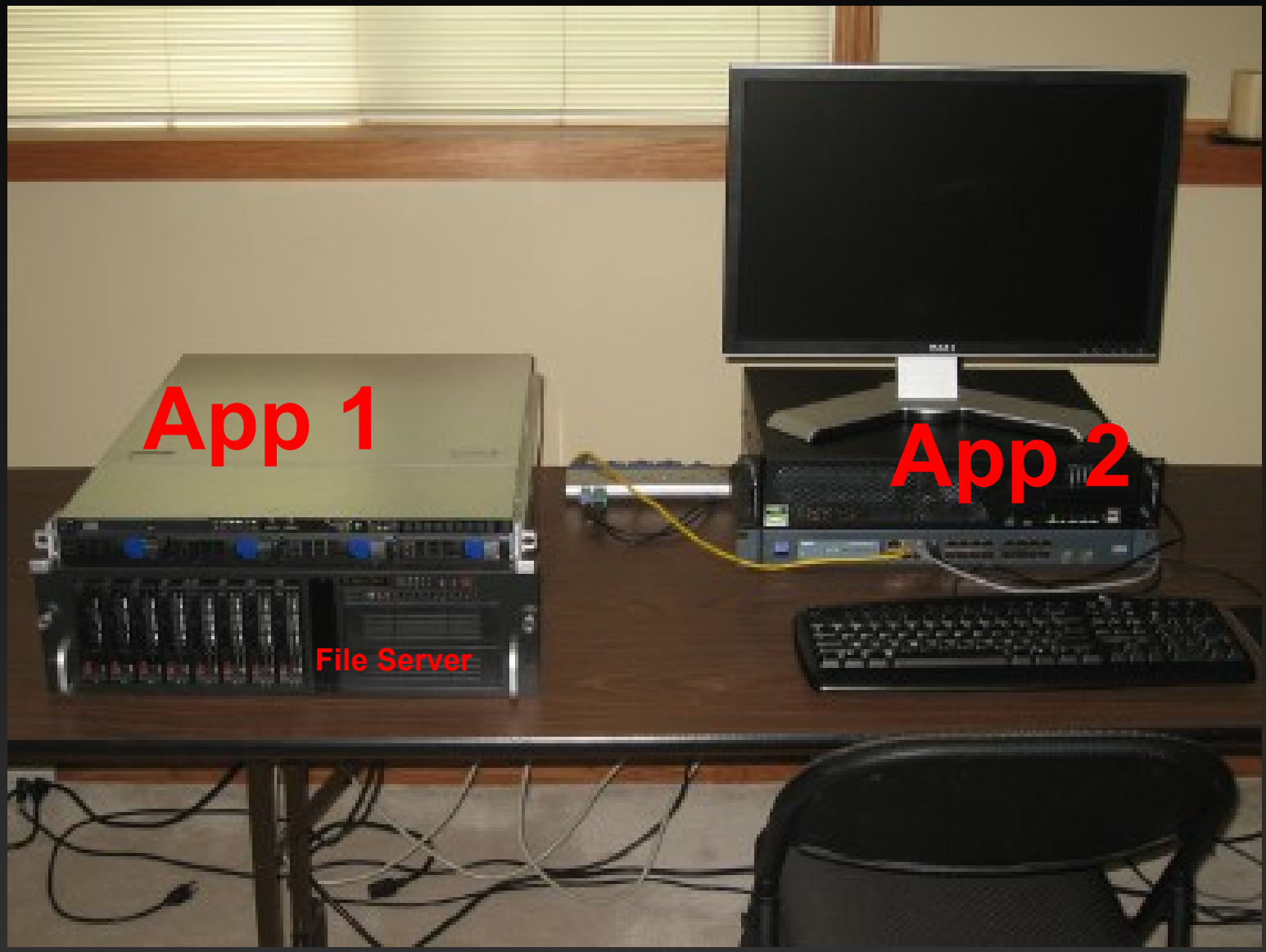
- Database Server(MySQL)
- Web Server (Apache)
- File Server (Samba)
- FTP
- DNS
- Email (Postfix)
- Source Repository (Subversion)

My Setup

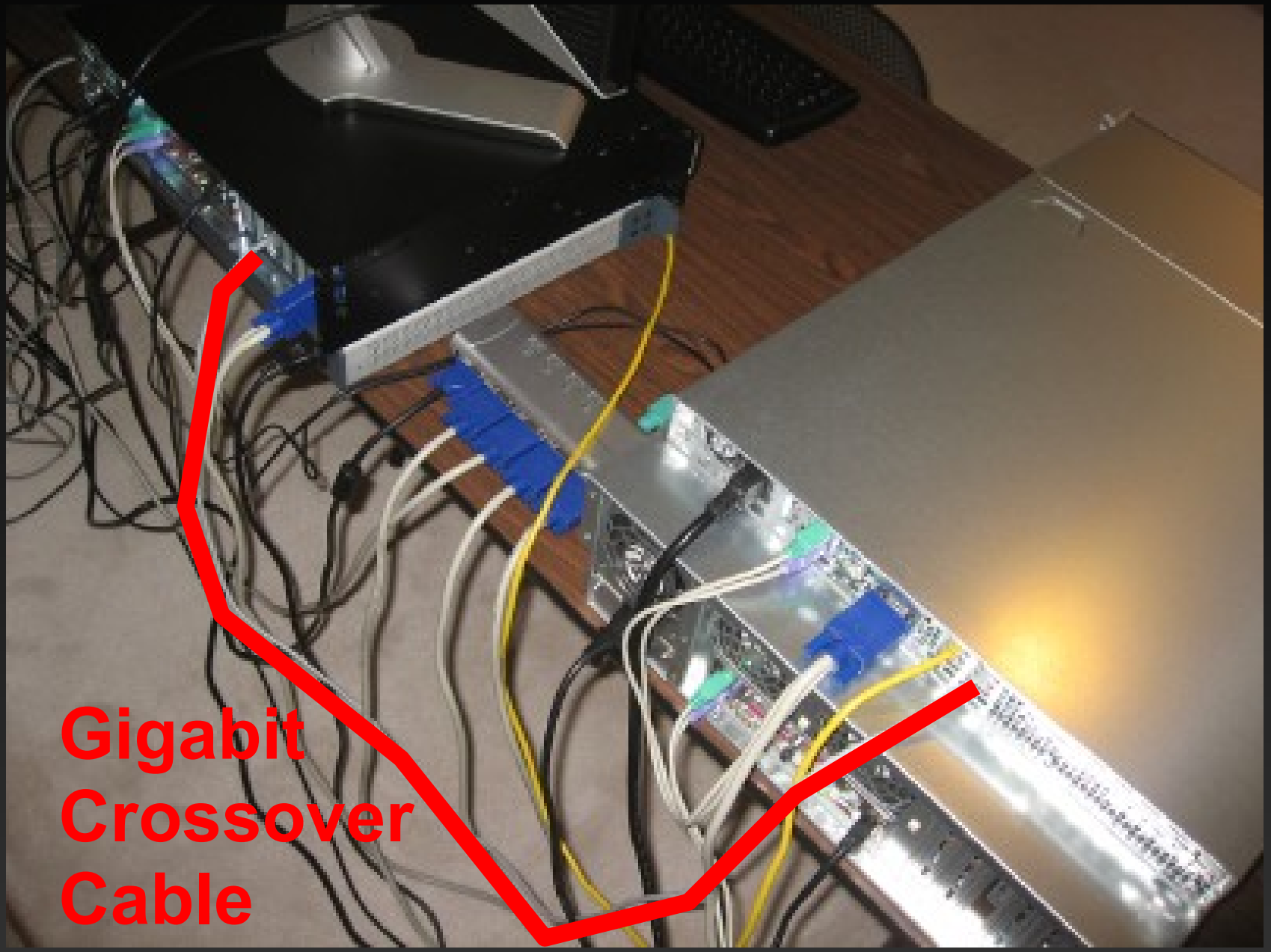
- App 1
- AMD Opteron 180 2.4GHz dual-core
- 4GB RAM
- 2 x 250GB SATA Software RAID 1
- 2 x Gigabit NICs

- App 2
- AMD Sempron 2800+
- 3GB RAM
- 1 x 250 IDE Hard Drive
- 2 x Gigabit NICs

My Setup



My Setup



Partitions

- 250GB total
- 20GB Operating system (/)
- 4GB swap
- 226GB DRBD (/mnt/data)

/mnt/data

- /mnt/data/apache2 – Apache conf, SSL certs, htpasswd
- /mnt/data/mail – Mail directory (Postfix/CourierIMAP)
- /mnt/data/mysql – MySQL conf, database files, binlog
- /mnt/data/svn – Subversion repository
- /mnt/data/www - Websites

DRBD Install Overview

- Download source
- Compile driver
- modprobe drbd
- Create /etc/drbd.conf
- Create meta disk - drbdadm create-md r0
- Bring up device - drbdadm up r0
- Become primary
  - drbdadm -- --overwrite-data-of-peer primary r0
- mkfs.ext3 /dev/drbd0
- mount /dev/drbd0 /mnt/data

/etc/drbd.conf

```
global {
  usage-count no;
}

common {
  protocol C;

  syncer {
    rate 30M;
    al-extents 1801;
  }

  startup {
    wfc-timeout  0;
    degr-wfc-timeout 15;
  }

  disk {
    on-io-error   detach;
    # fencing resource-and-stonith;
  }

  net {
    sndbuf-size 512k;
    timeout        60;    #  6 seconds  (unit = 0.1 seconds)
    connect-int    10;    # 10 seconds  (unit = 1 second)
    ping-int       10;    # 10 seconds  (unit = 1 second)
    ping-timeout   5;     # 500 ms (unit = 0.1 seconds)
    max-buffers    8000;
    max-epoch-size 8000;
    cram-hmac-alg  "sha1";
    shared-secret  "secret";
  }
}

resource r0 {
  on app1 {
    disk       /dev/md2;
    address    10.10.10.21:7788;
    device     /dev/drbd0;
    meta-disk  internal;
  }

  on app2 {
    disk      /dev/sda3;
    address   10.10.10.31:7788;
    device    /dev/drbd0;
    meta-disk  internal;
  }
}
```

Checking DRBD

- cat /proc/drbd

```
root@app1:/mnt/data# cat /proc/drbd
version: 8.0.11 (api:86/proto:86)
GIT-hash: b3fe2bdfd3b9f7c2f923186883eb9e2a0d3a5b1b build by phil@mescal, 2008-02-12 11:56:43
 0: cs:Connected st:Primary/Secondary ds:UpToDate/UpToDate C r---
    ns:237740648 nr:7160 dw:19725440 dr:242166751 al:2274 bm:17504 lo:0 pe:0 ua:0 ap:0
        resync: used:0/31 hits:13841287 misses:13628 starving:0 dirty:0 changed:13628
        act_log: used:0/1801 hits:4927296 misses:2427 starving:0 dirty:153 changed:2274
```

Heartbeat

- Exchange heartbeats
- When the heartbeat fails, other begins takeover
- Use 2 NICs (try to prevent split brain)
- Uses LSB init scripts
- Version 1
  - Simple as cake
- Version 2
  - GUI, allows more complex situations

Heartbeat Installation Overview

- Install heartbeat (apt-get install heartbeat)
- 3 conf files

/etc/ha.d/haresources

```
app1 \
        IPaddr::192.168.1.211/24/eth0 \
        IPaddr::192.168.1.212/24/eth0 \
        drbddisk::r0 \
        Filesystem::/dev/drbd0::/mnt/data::ext3 \
        bind9 \
        postfix \
        courier-imap \
        courier-imap-ssl \
        mysql \
        apache2 \
        samba
```

/etc/ha.d/ha.cf

```
keepalive 2
deadtime 15
warntime 5
initdead 30
autojoin none
bcast eth1
mcast eth0 239.0.0.42 694 1 0
auto_failback off
node app1 app2
logfacility local0
```

/etc/ha.d/authkeys

```
auth 1
1 sha1 secret
```

Finalizing Heartbeat Install

- Stop services
  - /etc/init.d/apache2 stop
- Disable services (Ubuntu/Debian)
  - update-rc.d -f apache2 remove
- Disable mounting of DRBD drive (/etc/fstab)
  - /dev/drbd0 /mnt/data ext3 noauto 0 0

Running Heartbeat

- /etc/init.d/heartbeat start
- /usr/lib/heartbeat/hb_takeover
- /usr/lib/heartbeat/hb_standby

Samba

```
[global]
        log file = /var/log/samba/log.%m
        passwd chat = *Enter\snew\sUNIX\spassword:*
%n\n *Retype\snew\sUNIX\spassword:* %n\n
*passwd:*password\supdated\ssuccessfully* .
        socket options = TCP_NODELAY
        obey pam restrictions = yes
        encrypt passwords = true
        passwd program = /usr/bin/passwd %u
        passdb backend = tdbsam
        dns proxy = no
        server string = %h server (Samba, Ubuntu)
        invalid users = root
        workgroup = WORKGROUP
        os level = 20
        auto services = data
        security = user
        syslog = 0
        panic action = /usr/share/samba/panic-action %d
        max log size = 1000
#   guest account = nobody
```

```
[data]
        writeable = yes
        path = /mnt/data
        force group = root
        force user = root
        create mode = 664
        directory mode = 775

[www]
        force user = www-data
        writeable = yes
        create mode = 664
        path = /mnt/data/www
        directory mode = 775
        force group = www-data
```

Split Brain

- Both servers are primary
- Which one is up-to-date?
- How do you prevent split brain?

Stonith

apt-get install stonith

Shoot
The
Other
Node
In
The
Head

- Home made stonith device
- Connects serial port to other server's reset switch
- ~$40 in Radio Shack parts to build 2 of them
- http://www.scl.co.uk/rcd_serial/README.rcd_serial

MySQL

- For maximum performance, use RAID 10 with battery backed cache RAID controller
- Write transactions to disk so they get replicated
  - innodb_flush_log_at_trx_commit=1
  - sync_binlog=1
- More info:
  - http://www.mysqlperformanceblog.com/2008/06/02/how-much-overhead-drdb-could-cause/

Demo

- Reboot app1
- App2 become primary
- App1 comes online, re-syncs
- Turn off app1's network
- DRBD on app2 loses connection with app1
- Run home and enable app1's networking :)

More Info

- http://www.drbd.org/
- http://www.cb1inc.com/taxonomy/term/70

Questions?

Thank you!