MySQL High Availability Deep Dive

Florian Haas (hastexo)
Yves Trudeau (Percona)

HA Sprint: Get Prepared!

- Tutorial instructions at: http://goo.gl/8ZLta
- Follow along -- or just watch! (up to you)

If you need help: we're the guys with the blue SPEAKER tabs, just holler!

About Us

Our virtual environment

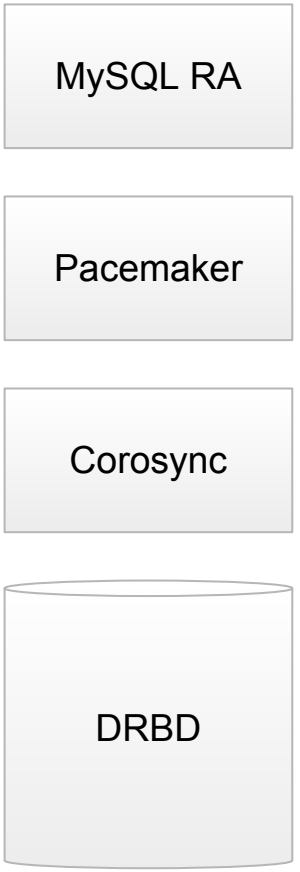
An Overview of the Linux HA stack

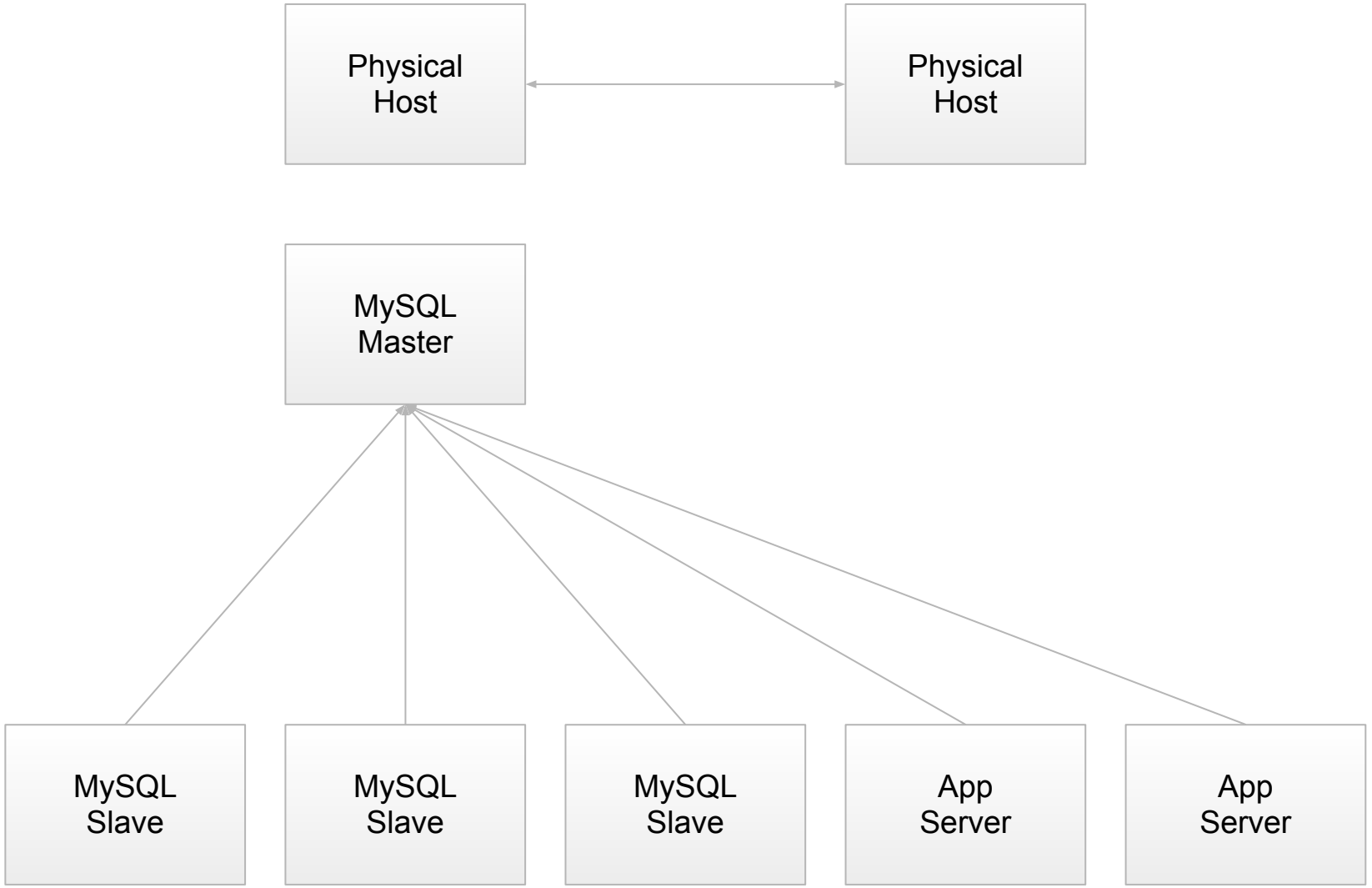
Application Interface

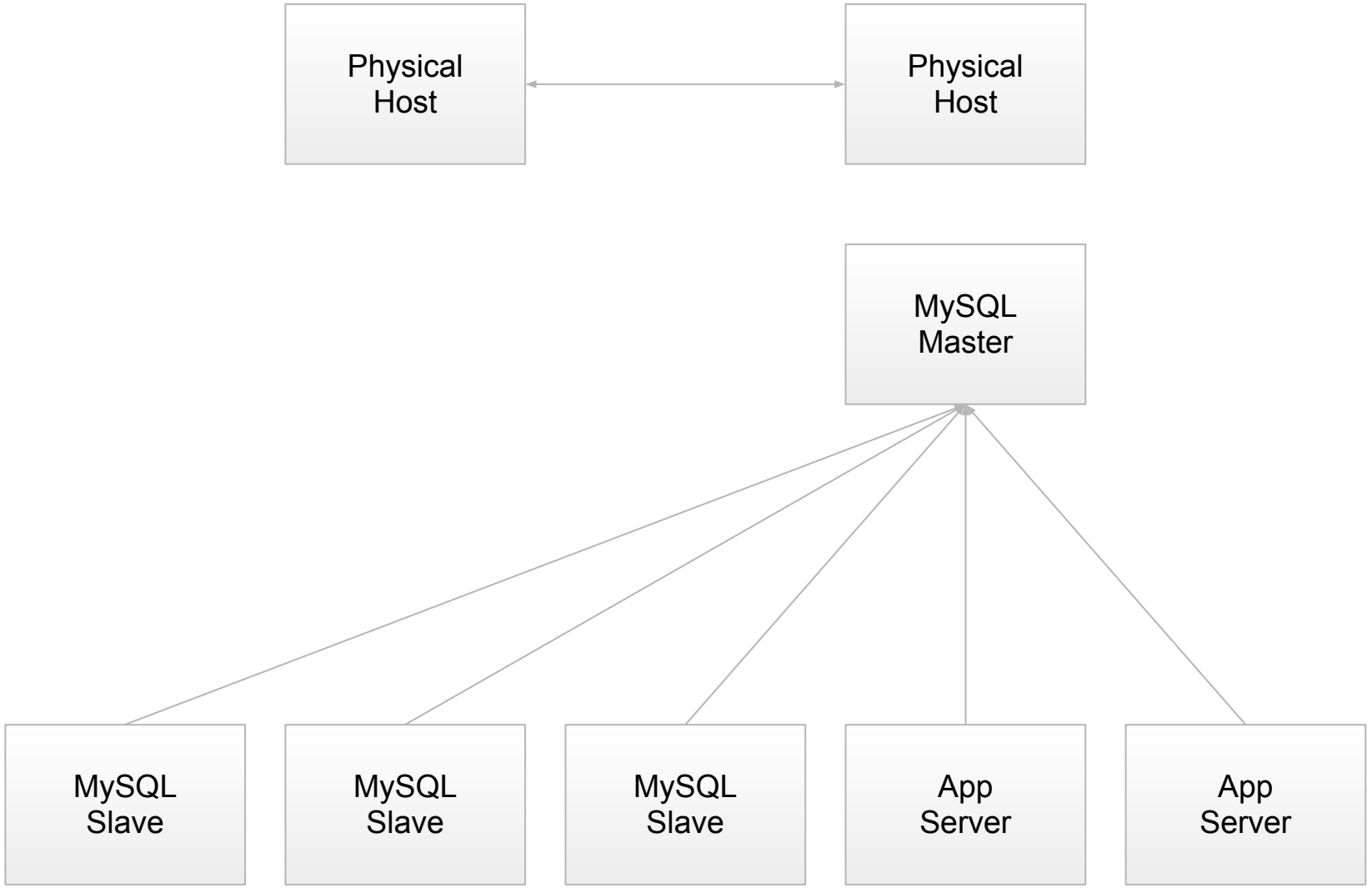
Cluster Resource Mgmt

Cluster Messaging

Storage Replication (optional)




Physical Host
Physical Host
MySQL Master
MySQL Slave
MySQL Slave
MySQL Slave
App Server
App Server


Physical Host
Physical Host
MySQL Master
MySQL Slave
MySQL Slave
MySQL Slave
App Server
App Server

DRBD

http://www.hastexo.com/knowledge/drbd

http://www.drbd.org

Creating a DRBD resource

- Name: mysql
- Device name: /dev/drbd0
- Backing disk: /dev/pacemaker-1/lvMySQL
- Meta data: internal
- Replication: 172.30.222.1 ↔ 2

- Filesystem: ext3

Corosync

http://www.hastexo.com/knowledge/corosync

http://www.corosync.org

Creating a Corosync Cluster

- Enable secauth
- 1 ring
- No logging to files, syslog only
- Start Corosync
- Check connectivity
- Check membership

Pacemaker

http://www.hastexo.com/knowledge/pacemaker

http://www.clusterlabs.org

Create a Pacemaker Cluster

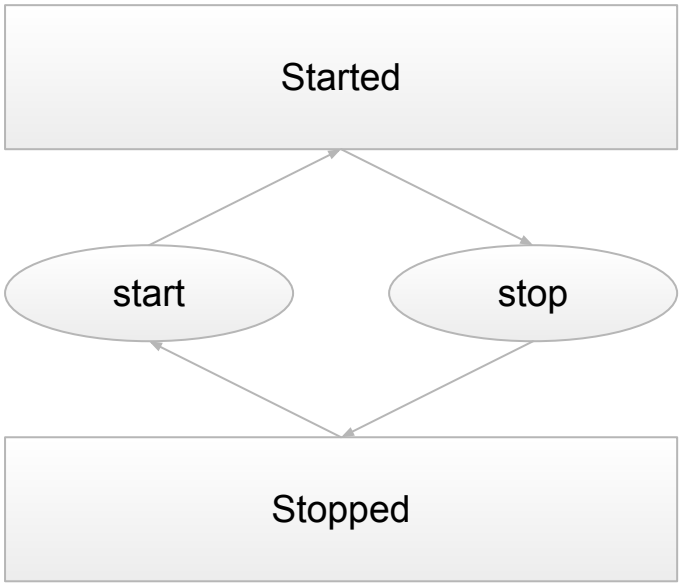
- Add the “pacemaker” service to `/etc/corosync.d`
- Use ver: 1
- Restart corosync
- Start pacemakerd
- Verify both nodes have joined the cluster (`crm_mon`)

STONITH/Fencing

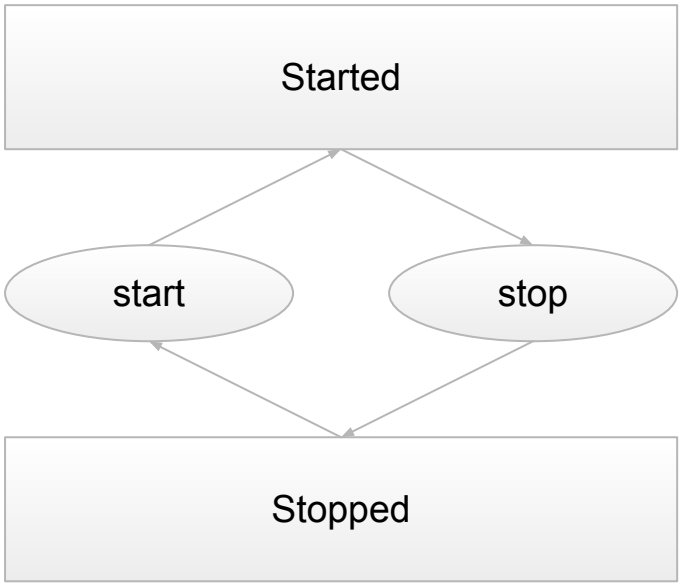
Setting up the cluster

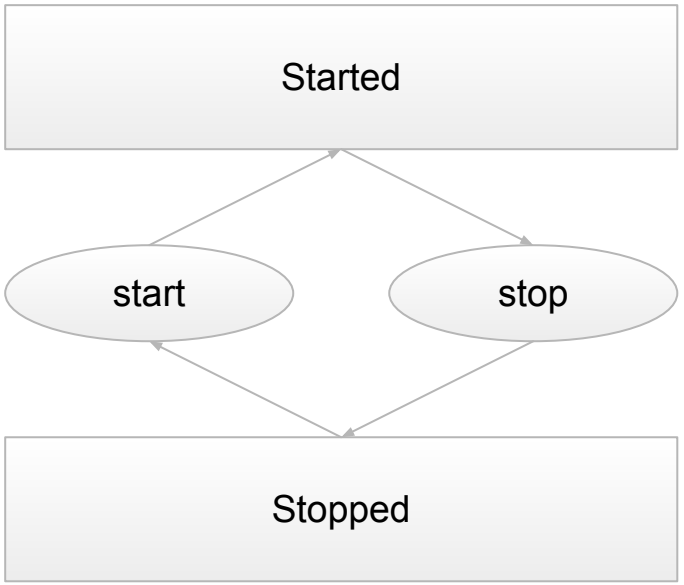
- Open the shell with crm configure
- Set no-quorum-policy to ignore
- Disable STONITH
  - Kids, don't do this at home!
- Commit the configuration

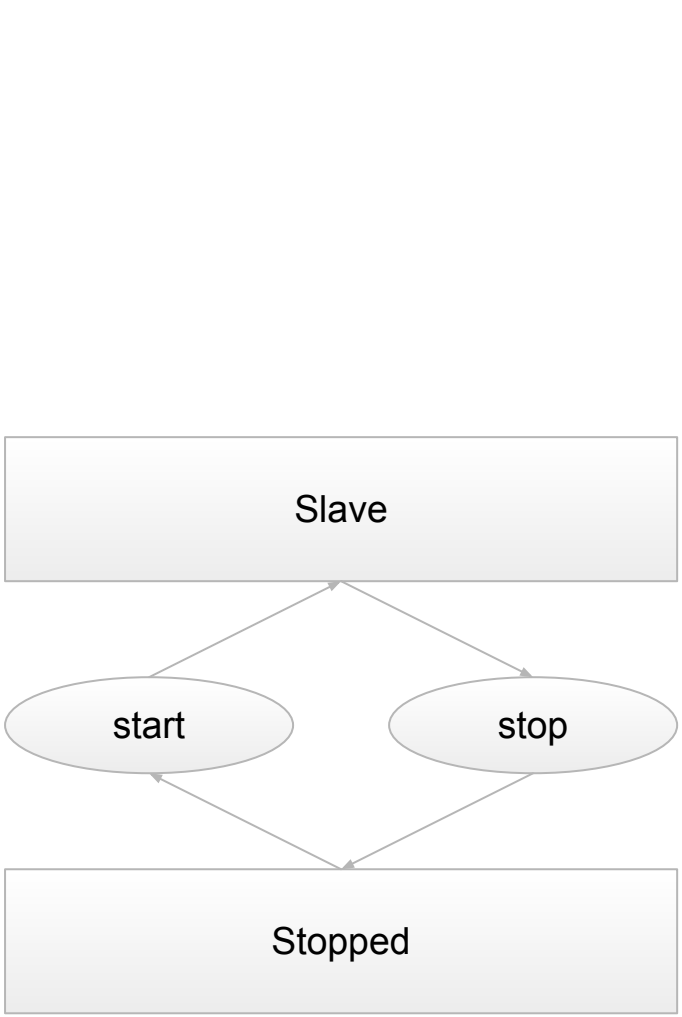
Pacemaker Resources

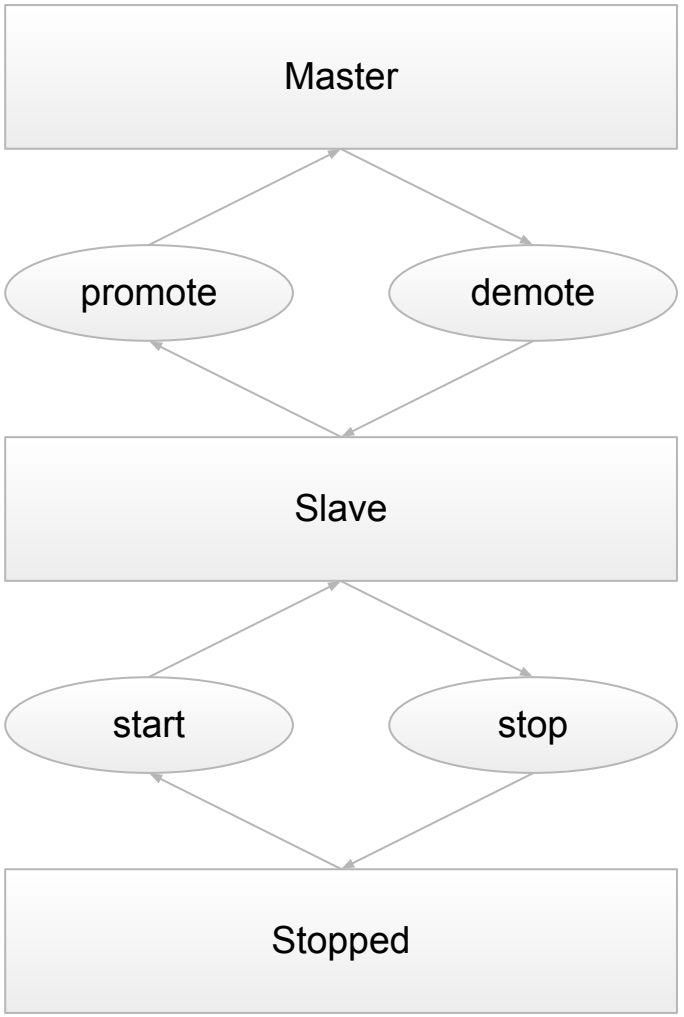

Started
start
stop
Stopped

Master/Slave Sets


Started
start
stop
Stopped


Started
start
stop
Stopped


Slave
start
stop
Stopped


Master
promote
demote
Slave
start
stop
Stopped

Adding a DRBD Master/Slave Set

- Create the DRBD primitive p_drbd_mysql
  - Use the ocf:linbit:drbd RA
  - Add 2 monitor operations (role=Master, role=Slave) with non-equal intervals
- Wrap it in a master/slave set
  - notify=true, clone-max=2, master-max=1
- Commit the configuration
- Watch the cluster promote

Constraints

Creating a Filesystem

- Add the Filesystem resource
  - Name: p_fs_mysql
  - Device: /dev/drbd0
  - Directory: /var/lib/mysql
  - File system type: ext3
- Add required order & colocation constraints

Groups

Add a Group

- Create the `g_mysql` resource group
- Add `p_fs_mysql` and `p_ip_mysql`

Requirements for MySQL/DRBD

No MyISAM

log-bin

sync_binlog

Bind or No bind

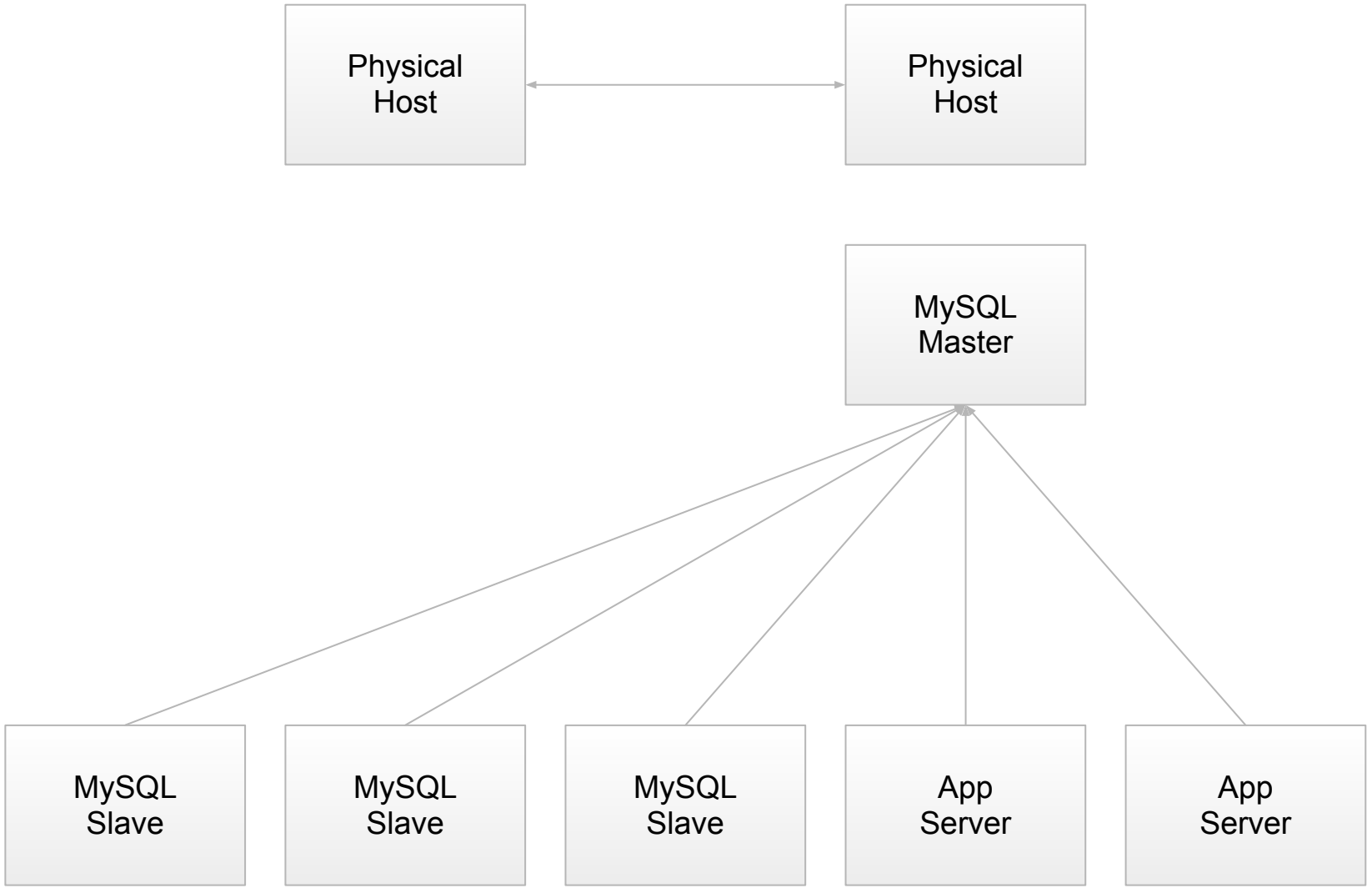
innodb_log_file_size

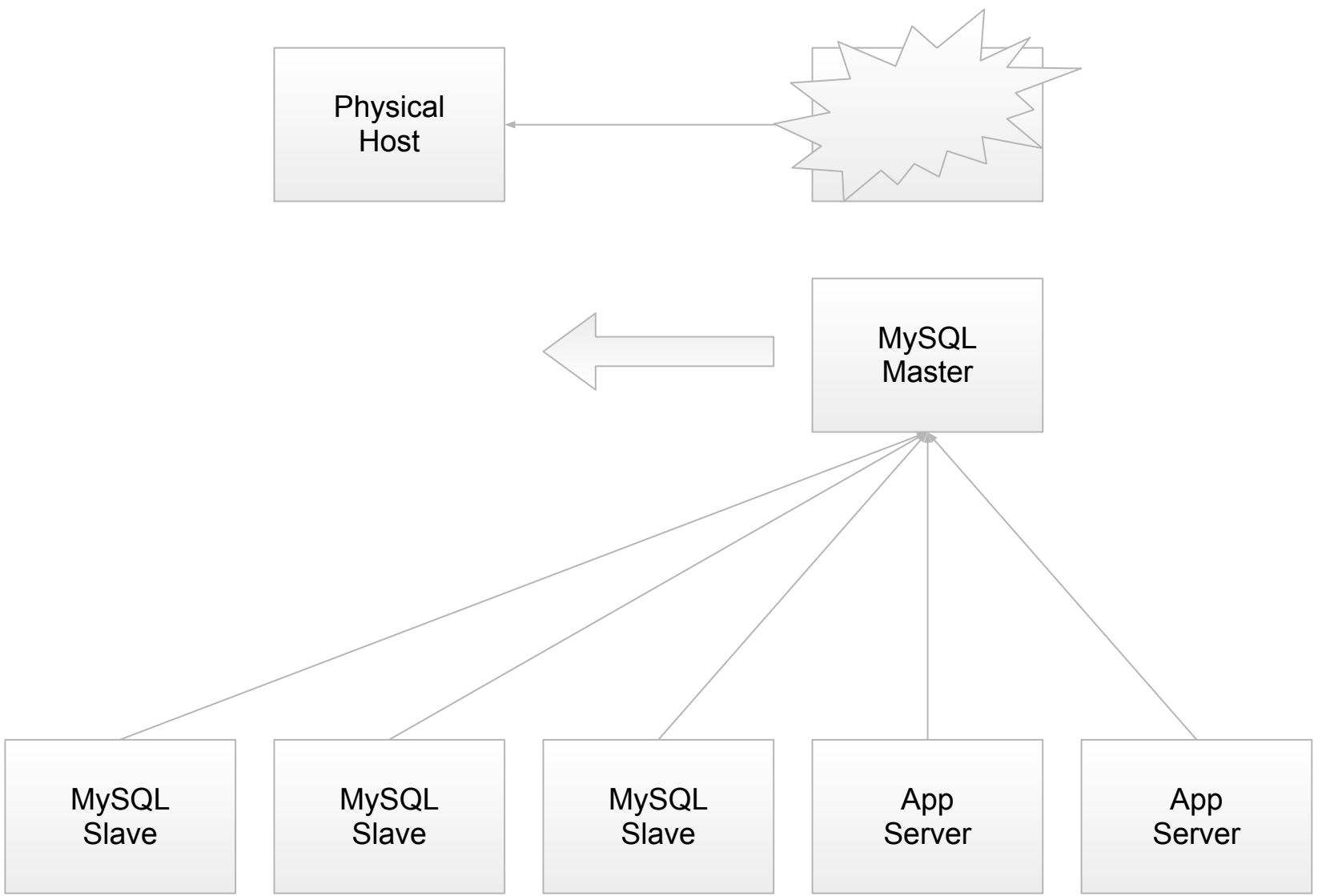
innodb_auto_lru_dump /
innodb_buffer_pool_restore_at_startup

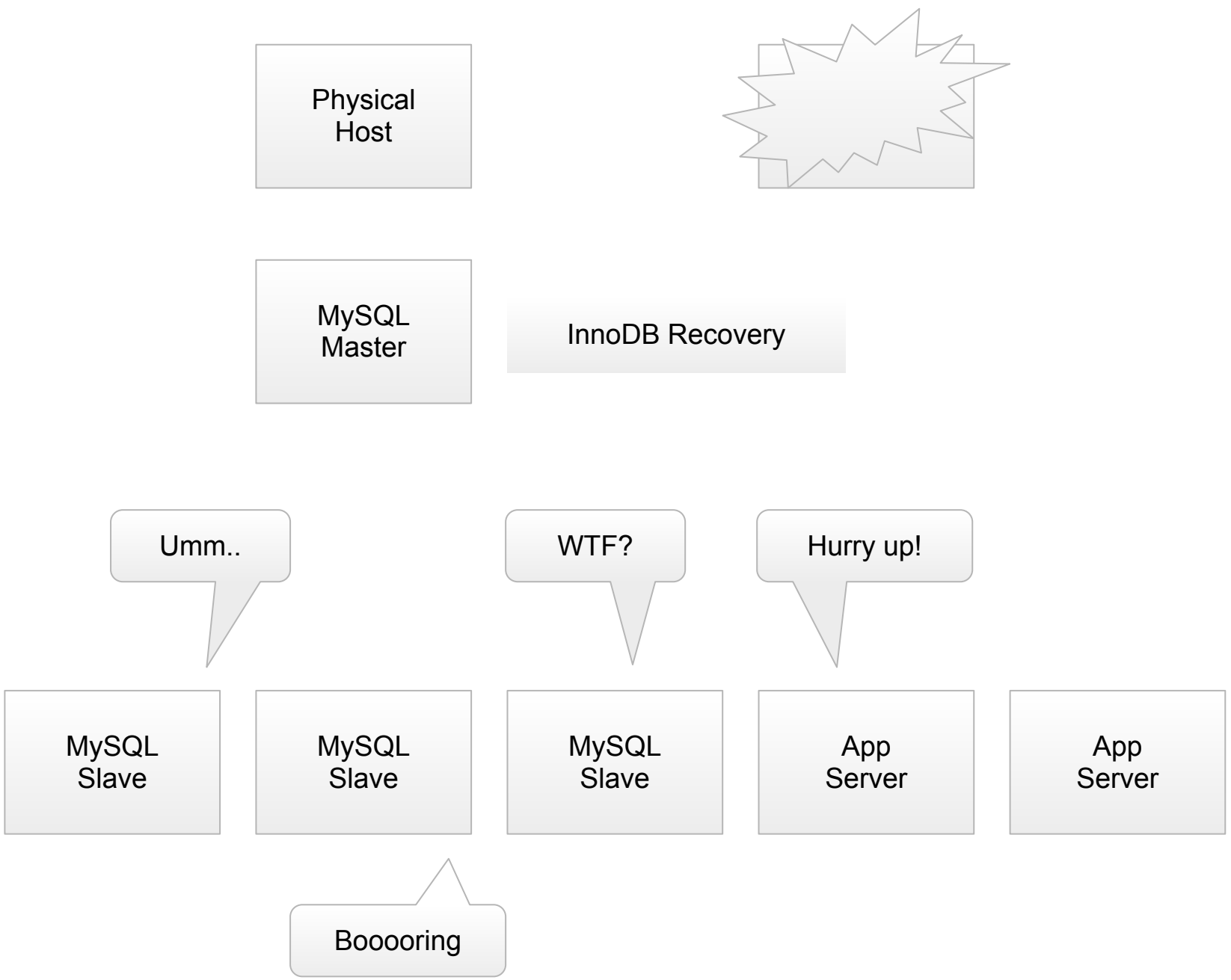
Add the MySQL resource

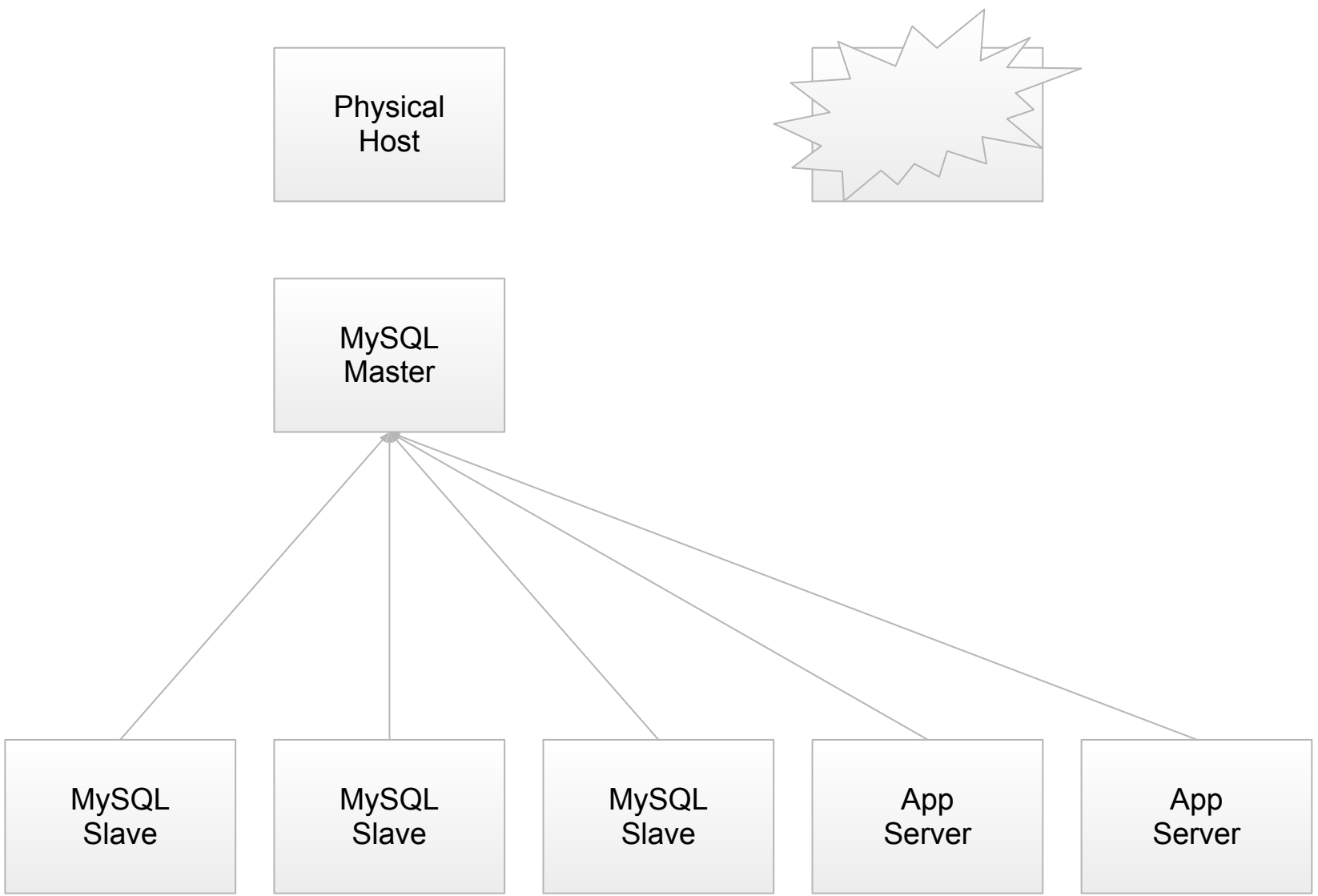
- Run `mysql_install_db`
- Create an `ocf:heartbeat:mysql` primitive `p_mysql`
- Add resource level monitoring
- Add it to the `g_mysql` group
- Watch the constraints update automagically
- Commit, Rejoice!

One more thing...


Physical Host
Physical Host
MySQL Master
MySQL Slave
MySQL Slave
MySQL Slave
App Server
App Server


Physical
Host
MySQL
Master
MySQL
Slave
MySQL
Slave
MySQL
Slave
App
Server
App
Server


Physical Host
MySQL Master
InnoDB Recovery
Umm..
WTF?
Hurry up!
MySQL Slave
MySQL Slave
MySQL Slave
App Server
App Server
Booooring


Physical Host
MySQL Master
MySQL Slave
MySQL Slave
MySQL Slave
App Server
App Server

Handling MySQL replication

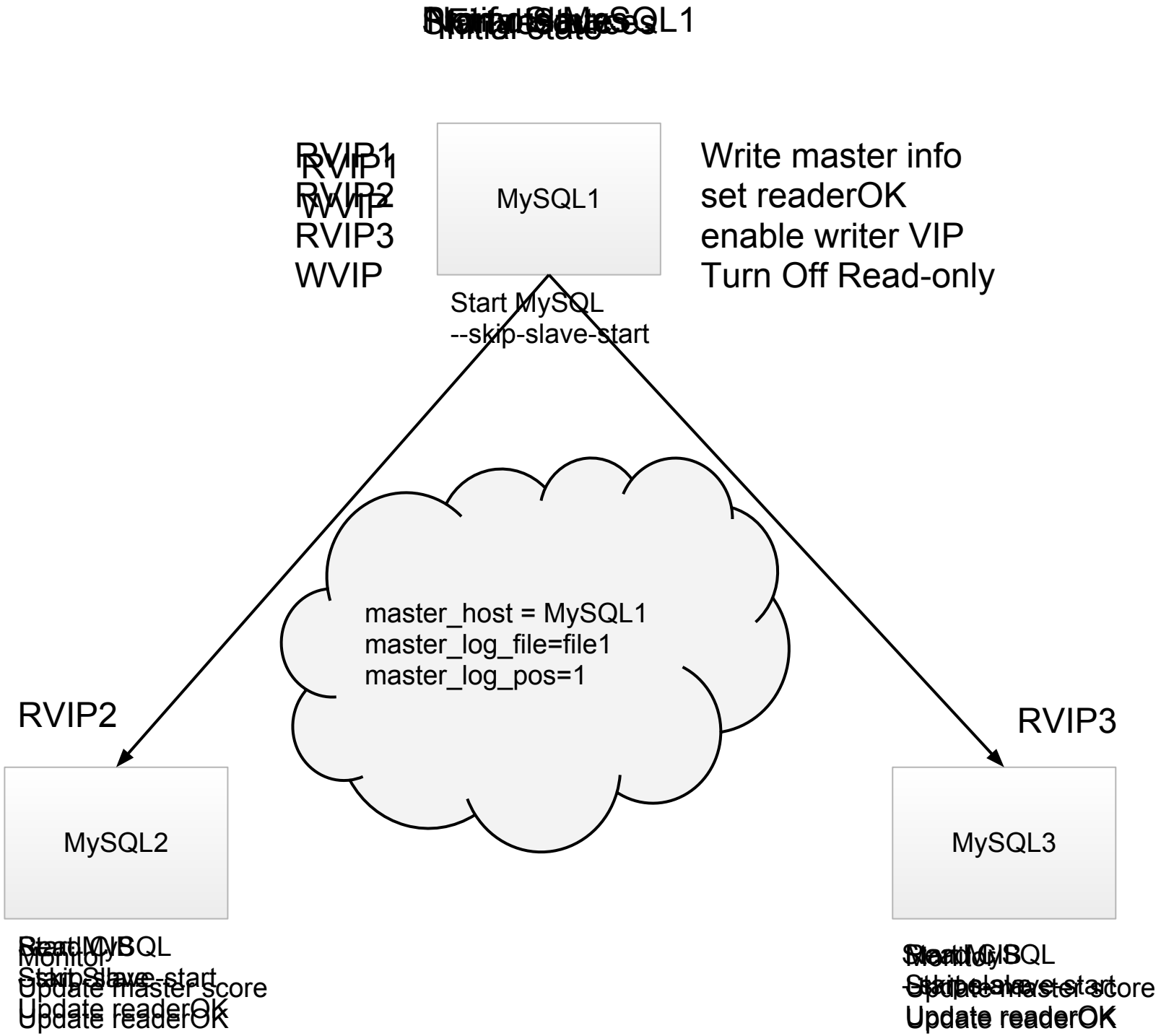
Read-only

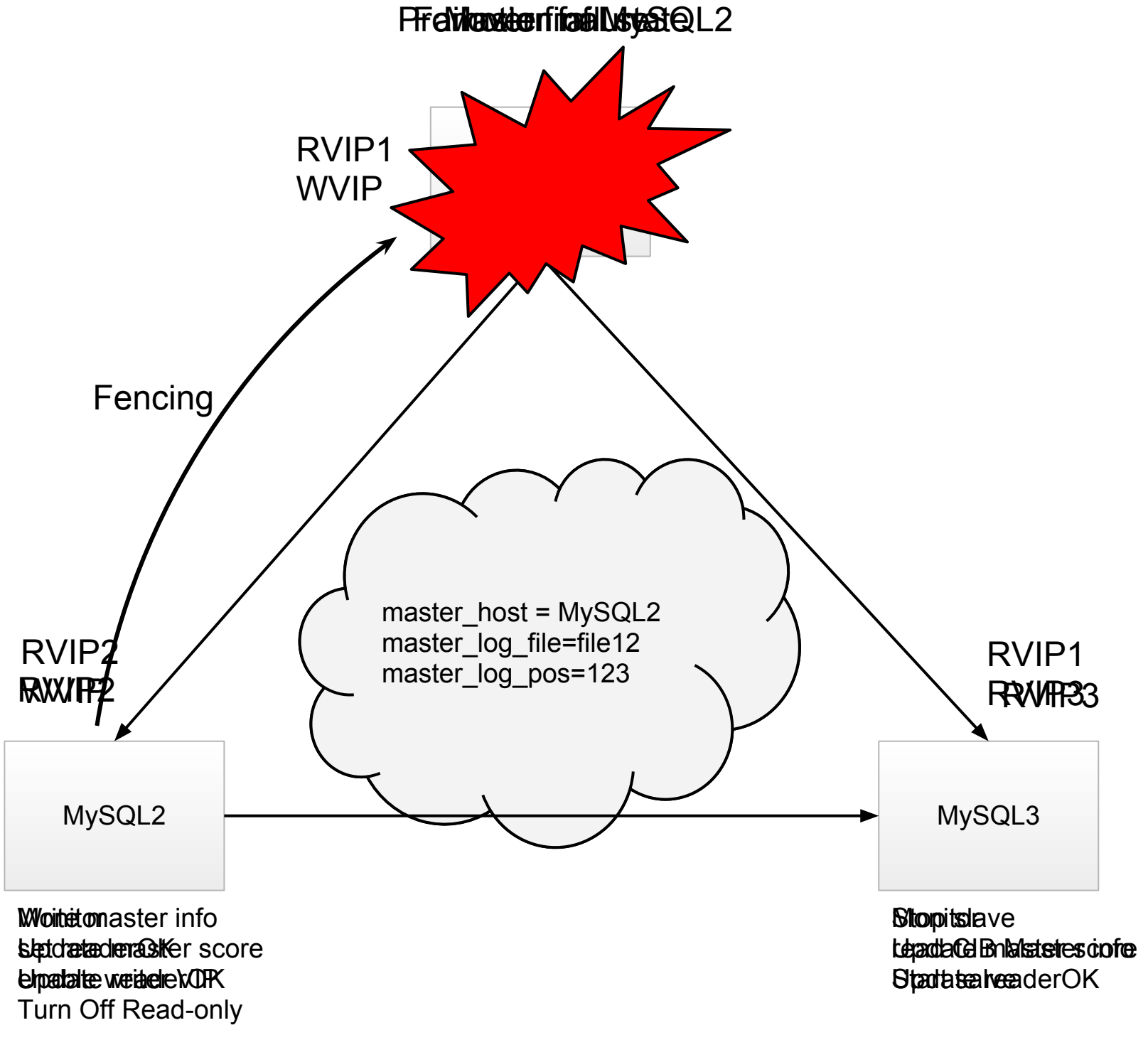
Promotion of the master

Attaching slaves

Monitoring slave state

Demotion


RVIP1
RVIP2
WVIP
RVIP3
WVIP
MySQL1
Write master info
set readerOK
enable writer VIP
Turn Off Read-only
Start MySQL
--skip-slave-start
master_host = MySQL1
master_log_file=file1
master_log_pos=1
RVIP2
RVIP3
MySQL2
MySQL3
Update master score
Update readerOK
Update readerOK


RVIP1
WVIP
Fencing
master_host = MySQL2
master_log_file=file12
master_log_pos=123
RVIP2
RVIP1
MySQL2
MySQL3

Setup

- MySQL Grants
- Node IP attribute
- MySQL primitive
- VIP primitive
- Master-slave ms resource
- Location rules
- Colocation rules
- Order rule

Variation on the theme

- Slave only nodes
- How to add slave nodes
- Setup with no VIP
- Avoid cluster collapse

Further Information

http://www.clusterlabs.org

http://www.drbd.org

http://www.suse.com/products/highavailability

http://www.hastexo.com

Liked this talk?

- Go to www.hastexo.com/shoutbox
- Log in
  (you can just use your OpenID from Wordpress, Blogger, Yahoo, Google Apps, Google Profile)
- Leave us a message!