

人人网如何使用MySQL

@周彦伟

2012-08-04

大纲

- 现在和未来
- 对软件的摸索
- 在硬件上的尝试

现在和未来





- 监控
- 备份
- 流程
- 架构
- 优化
- 沟通



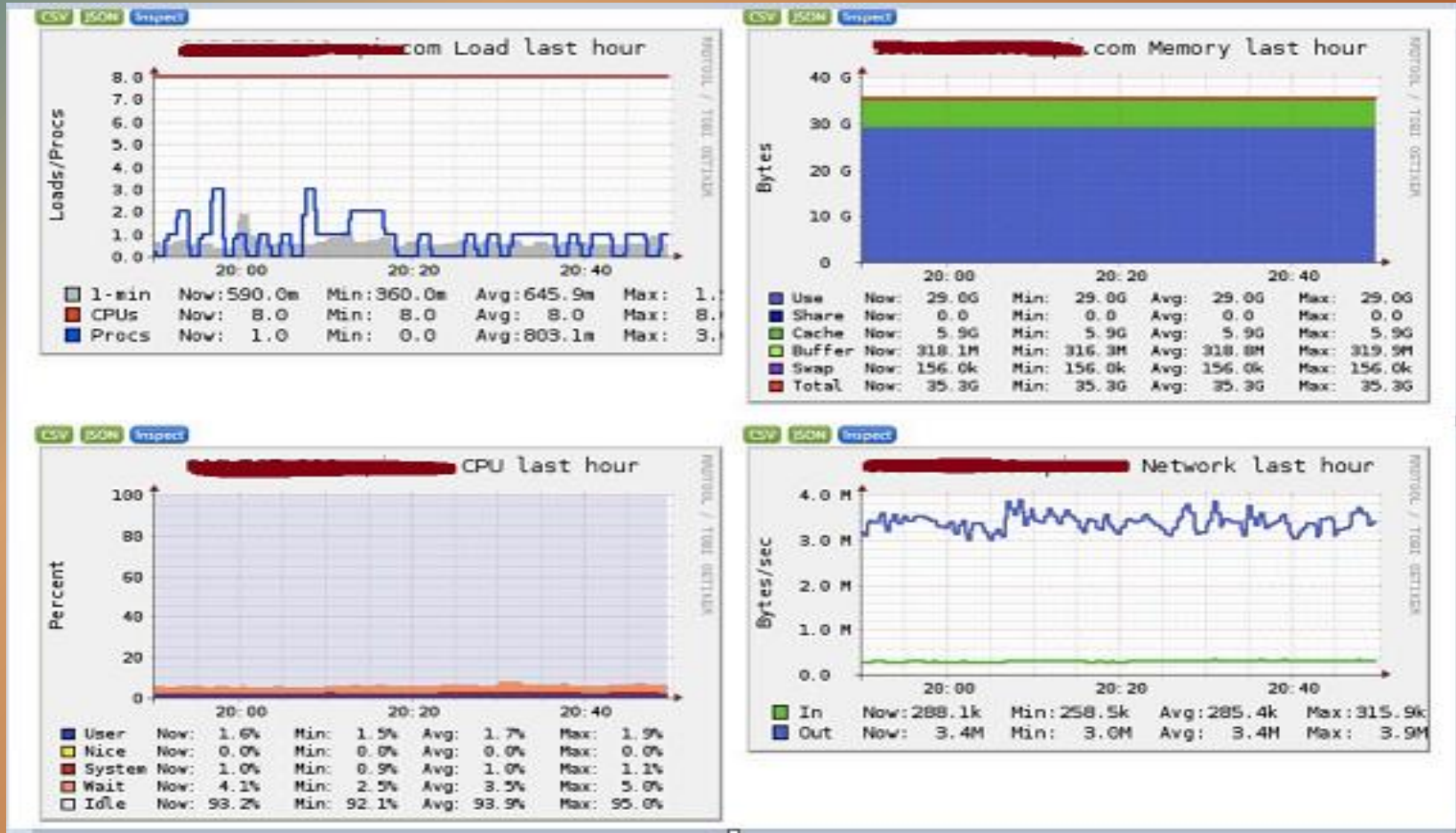
监控

- Nagios
- Ganglia
- Snipers监控系统

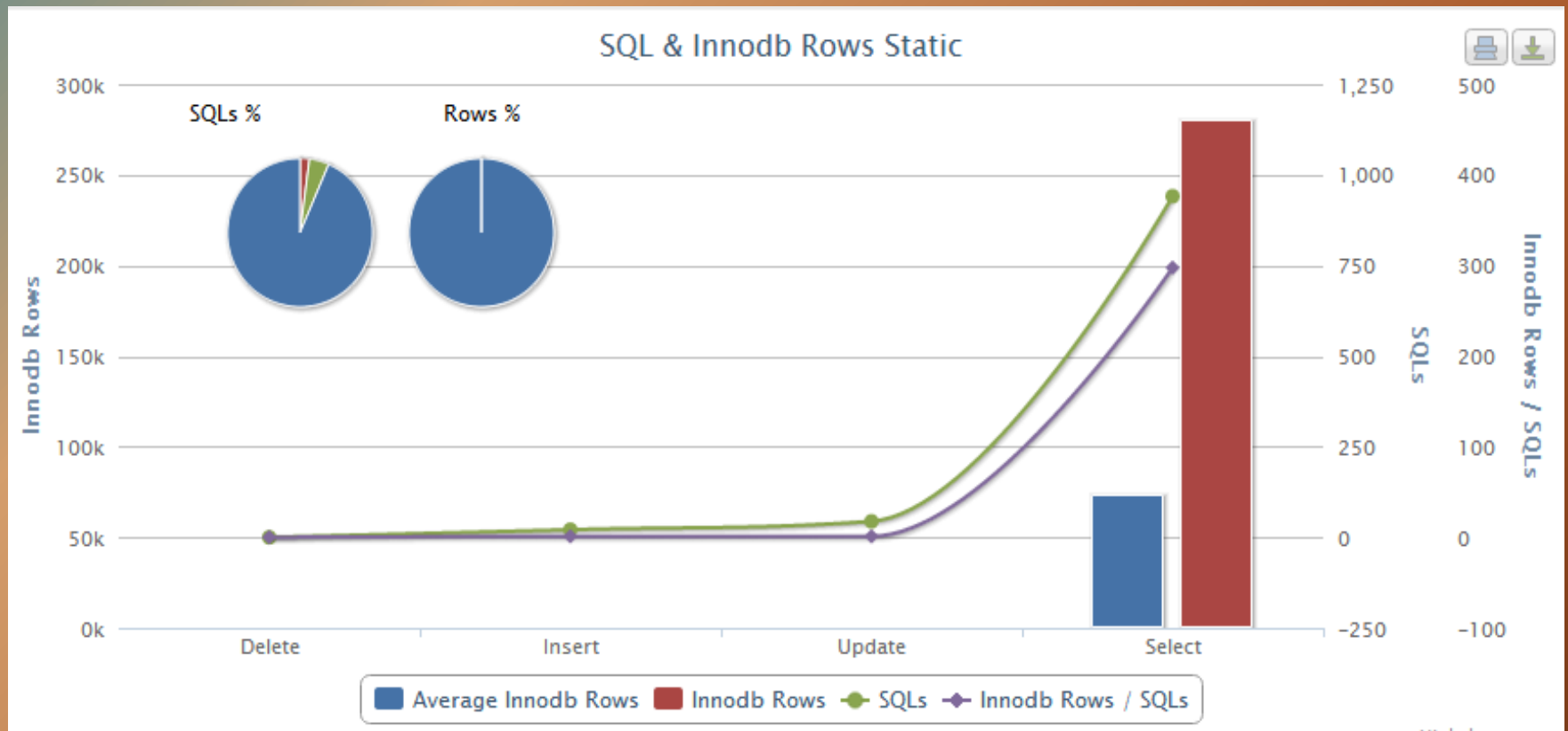
Nagios

		DB_PORT_3306		OK	07-31-2012 16:14:32	64d 0h 20m 9s	1/6
		DB_PORT_3306		OK	07-31-2012 16:14:50	7d 4h 53m 36s	1/6
		DB_PORT_3306		OK	07-31-2012 16:14:32	19d 16h 36m 21s	1/6
		DB_PORT_3307		OK	07-31-2012 16:14:10	13d 22h 32m 12s	1/6
		DB_PORT_3306		OK	07-31-2012 16:14:39	110d 22h 38m 1s	1/6
		DB_PORT_3307		OK	07-31-2012 16:14:10	110d 22h 36m 30s	1/6
		DB_PORT_3306		OK	07-31-2012 16:14:39	110d 22h 39m 58s	1/6
		DB_PORT_3307		OK	07-31-2012 16:14:10	110d 22h 38m 1s	1/6
		DB_PORT_3306		OK	07-31-2012 16:14:40	110d 22h 39m 58s	1/6
		DB_PORT_3307		OK	07-31-2012 16:14:10	110d 22h 38m 1s	1/6
		DB_PORT_3306		OK	07-31-2012 16:14:41	19d 15h 0m 29s	1/6
		DB_PORT_3308	 	CRITICAL	07-31-2012 16:14:08	130d 2h 52m 20s	6/6

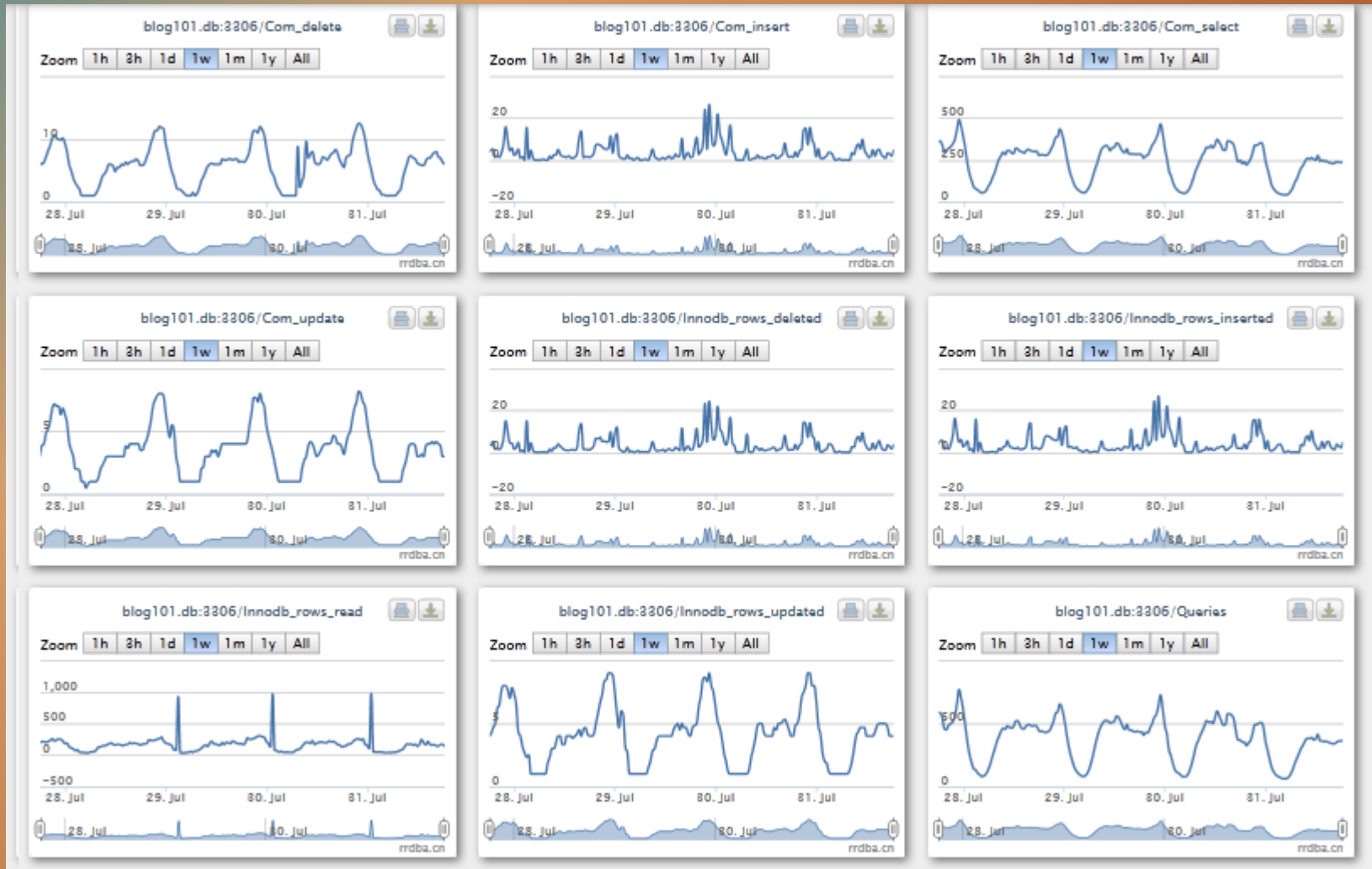
ganglia



Snipers



Snipers

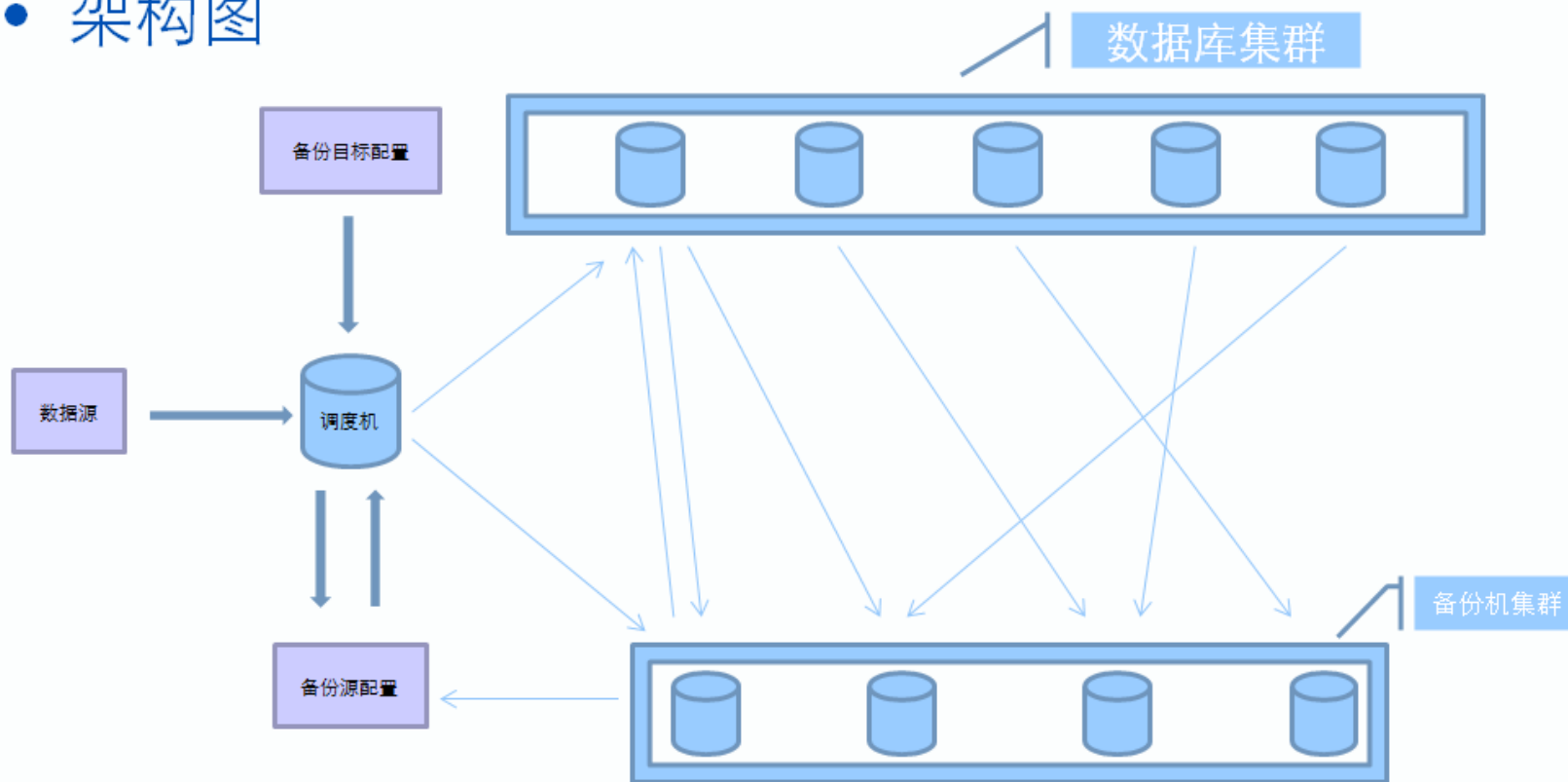


备份

- ~~InnoDB Hot backup~~
- Percona Xtrabackup
- 玄德智能备份系统

玄德备份

- 架构图





流程



- ~~DB需求邮件~~
- ~~DB需求规范邮件~~
- Mantis DB需求todo系统
- DB需求web处理系统

邮件规范



■ 表单创建、结构变更等邮件

1. 标 题: [DDL需求][部门名称]该需求概要描述。例: [DDL需求][安全中心]举报功能新加表 `user_report_log`。
2. 收件人: renren.db@renren-inc.com 
3. 内 容: 填写[数据库表添加/改动申请单](#) , 将内容粘贴至正文, 若涉及sql太多, 可以将sql以附件发送



■ 数据统计需求等邮件



1. 标 题: [统计需求][部门名称]该需求概要描述。例: [统计需求][渠道高中部]查询2010第一季度每周注册量
2. 收件人: renren.db@renren-inc.com 
3. 内 容: 填写[数据统计申请单](#) , 将内容粘贴至正文



未分派的 [^] (1 - 1 / 1)



[0001824](#) [小站] 人人小站数据需求
  [所有项目] 人人网 - 2012-05-24 19:09



已解决的 [^] (1 - 10 / 689)



[0001877](#) [3G LBS] 商业活动表结构升级
  [所有项目] 人人网 - 2012-05-25 19:19



[0001878](#) [增值] 校园小组竞选组长--建索引
  [所有项目] 人人网 - 2012-05-25 19:18



[0001875](#) [数据平台] 数据平台修改一张表。增加字段
  [所有项目] 人人网 - 2012-05-25 18:02



[0001876](#) [增值] 给小组表加2个字段
  [所有项目] 人人网 - 2012-05-25 18:01



[0001874](#) [糯米网] 添加字段
  [所有项目] 糯米网 - 2012-05-25 17:24

[0001872](#) [公共主页] 状态墙v5版数据库添加表vir_wall_doings
  [所有项目] 人人网 - 2012-05-25 16:29

[0001870](#) [糯米网] 申请访问权限
  [所有项目] 糯米网 - 2012-05-25 14:36



[0001868](#) [安全运营] 六一活动审核
  [所有项目] 人人网 - 2012-05-25 11:42



[0001865](#) [3G 通讯录] 数据转移
  [所有项目] 人人网 - 2012-05-25 11:20



[0001867](#) [糯米网] 新建表renren_yilidaguoli
  [所有项目] 糯米网 - 2012-05-25 11:12



我报告的 [^] (0 - 0 / 0)



最近修改 [^] (1 - 10 / 788)



[0001877](#) [3G LBS] 商业活动表结构升级
  [所有项目] 人人网 - 2012-05-25 19:19



[0001878](#) [增值] 校园小组竞选组长--建索引
  [所有项目] 人人网 - 2012-05-25 19:18



[0001875](#) [数据平台] 数据平台修改一张表。增加字段
  [所有项目] 人人网 - 2012-05-25 18:02



[0001876](#) [增值] 给小组表加2个字段
  [所有项目] 人人网 - 2012-05-25 18:01



[0001874](#) [糯米网] 添加字段
  [所有项目] 糯米网 - 2012-05-25 17:24

[0001872](#) [公共主页] 状态墙v5版数据库添加表vir_wall_doings
  [所有项目] 人人网 - 2012-05-25 16:29

[0001870](#) [糯米网] 申请访问权限
  [所有项目] 糯米网 - 2012-05-25 14:36

[0001868](#) [安全运营] 六一活动审核
  [所有项目] 人人网 - 2012-05-25 11:42

[0001865](#) [3G 通讯录] 数据转移
  [所有项目] 人人网 - 2012-05-25 11:20

[0001867](#) [糯米网] 新建表renren_yilidaguoli
  [所有项目] 糯米网 - 2012-05-25 11:12

架构

- 数据切分
- DbDescriptor
- MMM
- DHEA

mod

blog_0

blog_1

blog_2

blog_3

blog_99

log_201201

log_201202

log_201203

log_201204

log_201212

div

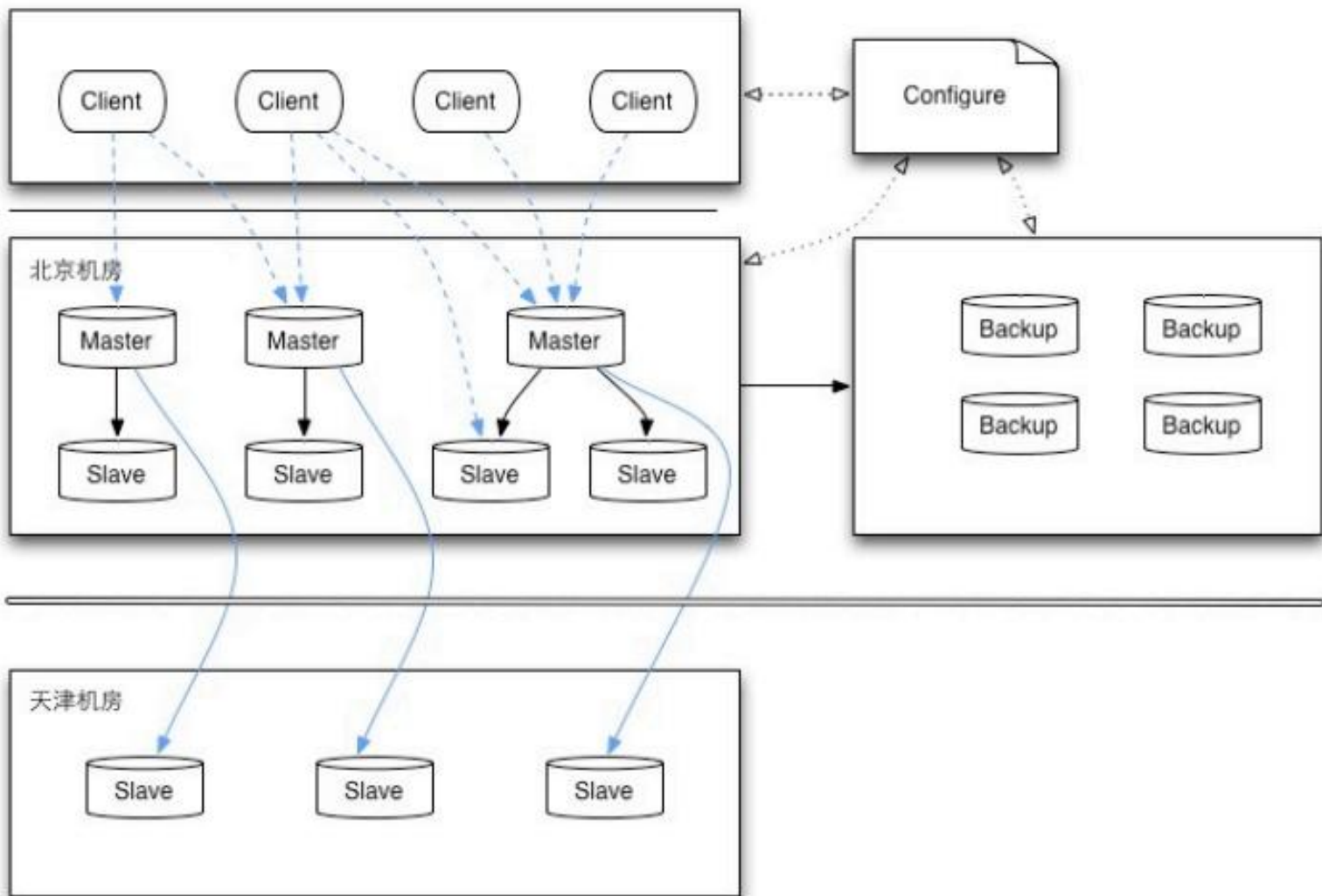
feed_17

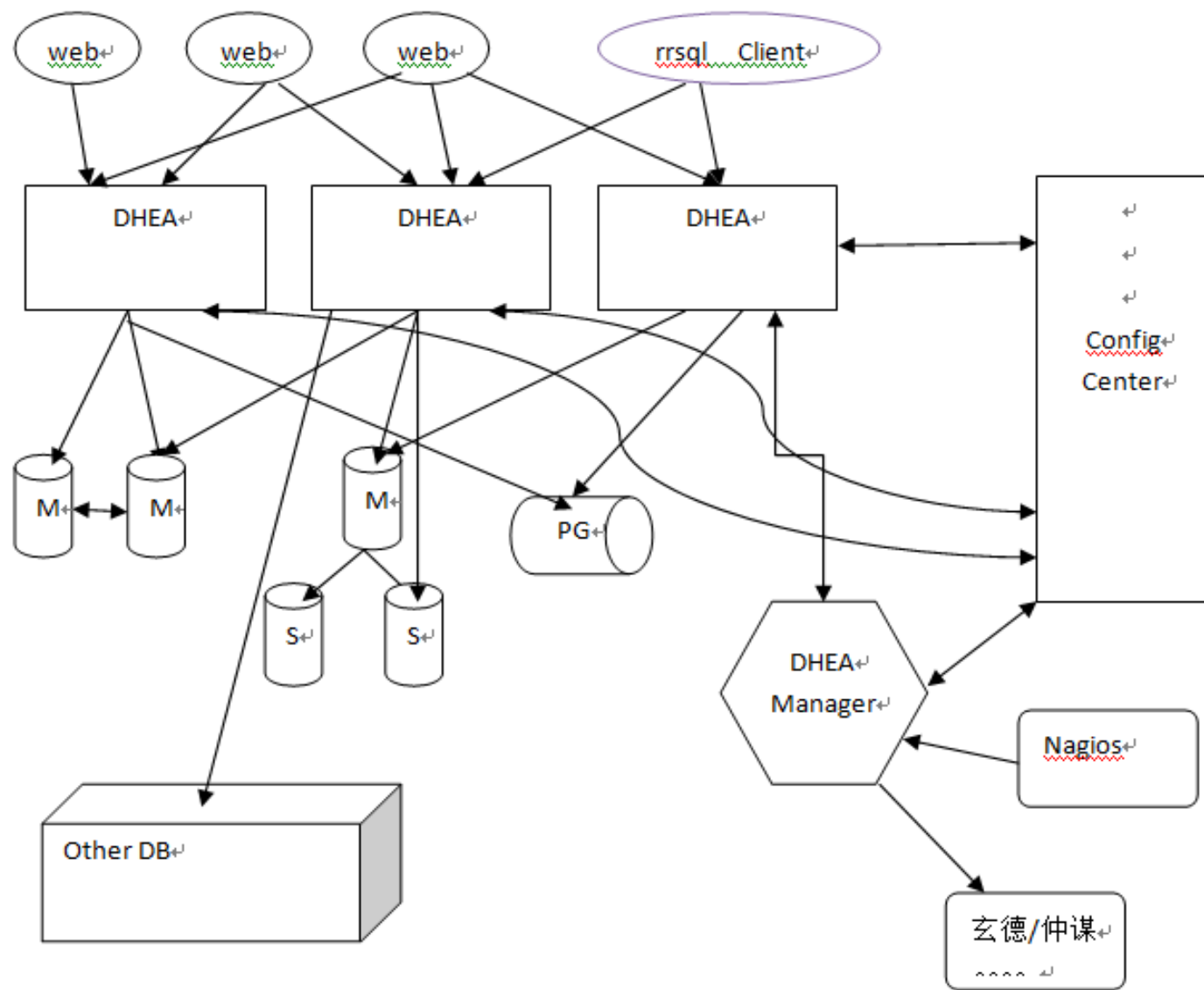
feed_20

feed_21

feed_22

feed_29





优化

- 业务
- 硬件
- 数据库
- 表
- SQL

优化就像，除非刚做过，否则永远也不够

沟通





你這是專程來吵架的？



拼全力为众生，牺牲也值得



曾经有一份真诚的爱情放在我面前

沟通是合格**DBA**的生存技能之一

对软件的摸索

- MySQL Cluster
- InfoBright
- Percona xtrabackup / InnoDB Hot backup
- Percona xtraDB
- Percona xtraDB Cluster(galera)
- Percona toolkit
- Tungsten
- MMM





在硬件上的尝试

- ssd
- stec
- fusionio
- virident
- memblaze
- flashcache

SSD

- x3630 M3
- CPU E5645 @ 2.40GHz 6 x 2
- MEM:96G
- DISK:160G x 6 + 600G x 12 + 300G x 2 System
- SSD Intel 320S 160G OPed 135G
- SSD作为裸设备，用作flashcache
- xfs: ssize=4k bsize=4k

FusionIO

- 直接做存储或者用作Flashcache
- fio: bsize=4K none op
- 驱动加载参数:
 - iomemory-vsl use_workqueue=0
 - iomemory-vsl disable-msi=0
 - iomemory-vsl use_large_pcie_rx_buffer=1

Flashcache

- `echo 90 > /proc/sys/dev/flashcache/sdb1+sda3/dirty_thresh_pct`
- `echo 1 > /proc/sys/dev/flashcache/sdb1+sda3/fast_remove`
打开fast remove特性，关闭机器时，无需将cache中的脏块写入磁盘
- `echo 1 > /proc/sys/dev/flashcache/sdb1+sda3/reclaim_policy`
脏块刷出策略，0: FIFO，1: LRU。
- `echo 1 > /proc/sys/dev/flashcache/sdb1+sda3/cache_all`
- `echo 0 > /proc/sys/dev/flashcache/sdb1+sda3/fallow_delay`
- `mount -o rw,noatime,nodiratime,barrier=0 /dev/mapper/cachedev /data`
- `skip_seq_thresh_kb 128`

Raid

- SAS SATA RAID

Adapter 0-VD 1(target id: 1):

- Cache Policy:WriteBack, ReadAdaptive, Cached, No Write Cache if bad BBU
- queue/scheduler: deadline
- queue/read_ahead_kb: 128

- SSD RAID

Adapter 0-VD 2(target id: 2):

- Cache Policy:WriteBack, ReadAheadNone, Direct, No Write Cache if bad BBU
- queue/scheduler: noop
- queue/read_ahead_kb: 0



EOF