SACC

2012中国系统架构师大会

SYSTEM ARCHITECT CONFERENCE CHINA 2012

架构设计 · 自动化运维 · 云计算

云端的数据库

盛大云计算 郭理靖

MYSQL和MONGODB的云服务

Agenda

- **Database As A Service (数据库云)介绍**
- 数据库云整体架构
- MySQL云的实现
- MongoDB云的实现
- 未来的趋势

数据库云是什么？

- 数据库云就是提供数据库服务的云
  - 自助式申请，
    - “所见即所得”
    - 可以按业务需求申请不同配置的数据库
    - 界面管理
  - 完善的服务，
    - 安全控制
    - 数据自动备份
    - 数据库灾难恢复
    - 数据库镜像
    - 监控报表
  - 按时收钱

数据库云的宗旨

Setup without hardware

- 省事

Operator without DBA

- 省力

Start Business with a little money

- 省钱

业界产品

- Google CloudSQL (MySQL)
- Amazon RDS (MySQL/MS SQL Service/Orcale)
- Microsoft Azure SQL (MS SQL Service)
- Enterprisedb （postersql)
- 盛大云 数据库云(MySQL/MongoDB)
- 阿里云 RDS (MySQL/MS SQL Server)
- 新浪SAE Mysql服务
- MongoHQ/MongoLab (MongoDB)
- Garantia/Redis To Go/Redis4you (Redis)

申 请 | 修改密码 | 升级数据库 | 删除 | 重启 | 制作镜像 | 恢复数据库 | 控制台 | 创建只读从节点

数据库名称查询：

	数据库名称	Instance类型	状态	空间	安全組	引擎
☑	ccreport	超微	available	5 GB	cc	mysql
☐	report	超微	available	5 GB	default	mysql
☐	transcodingdb	大型	available	100 GB	TranscodingDB	shared-mongodb

详细信息

详细信息 | 使用空间信息 | 监控 | 最近日志

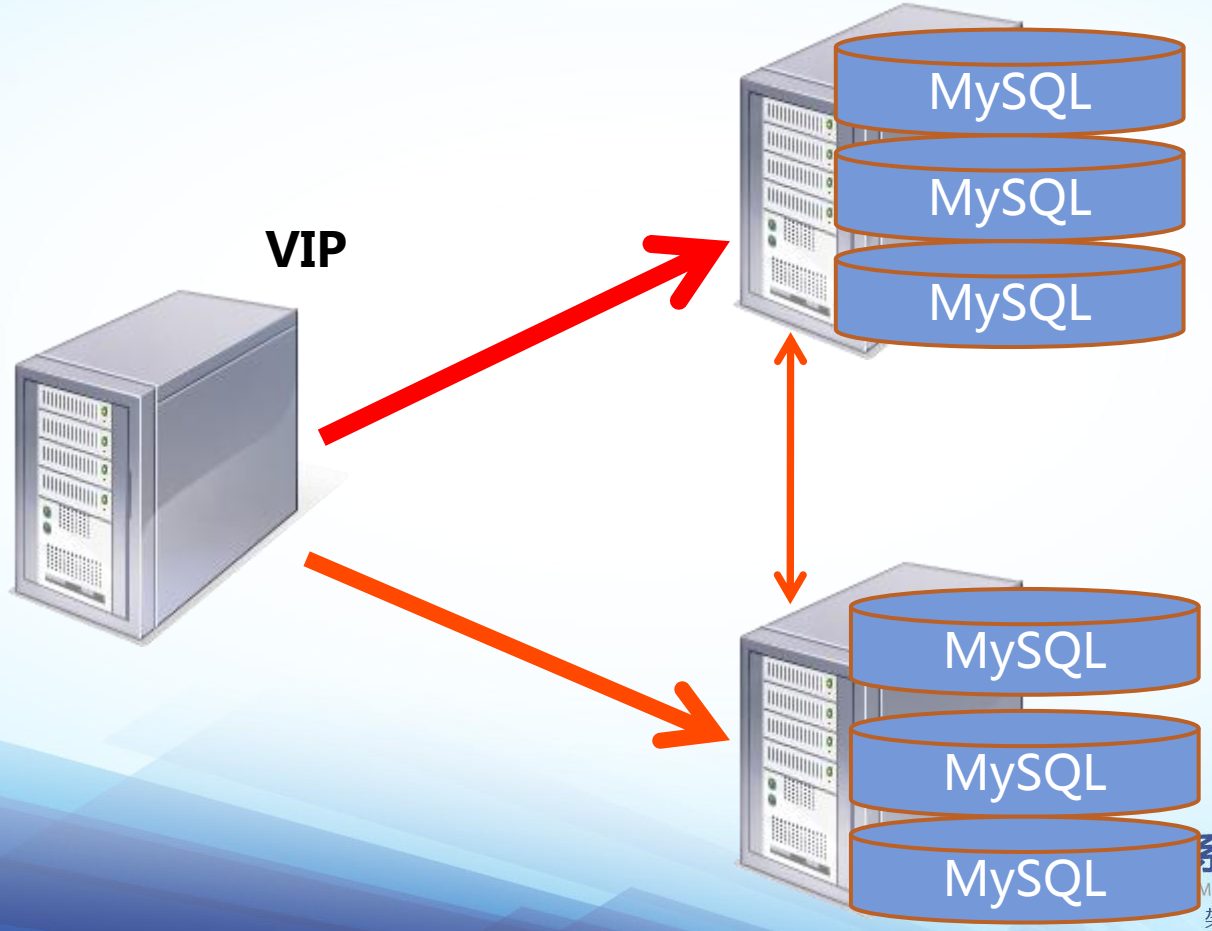
数据库版本： 5.5
数据库状态： available
部署区域： 华东区
IP： 58.215.172.22
备份时间： 00:00-03:00
数据库空间： 5 GB
维护时间： Mon:00:00-Mon:03:00

数据库证书： general-public-license
安全組： cc
Instance类型： 超微
端口： 3306
参数組： default.mysql5.5
用户名： ccreport
创建时间： 2012-06-21 16:59:31

Agenda

- Database As A Service (数据库云)介绍
- **数据库云整体架构**
- MySQL云的实现
- MongoDB云的实现
- 未来的趋势

最初的想法



面对的问题

数据

- 数据库访问安全
- 数据库备份与恢复
- 容量规划

公平

- CPU使用公平
- 内存使用公平
- 磁盘IO使用公平

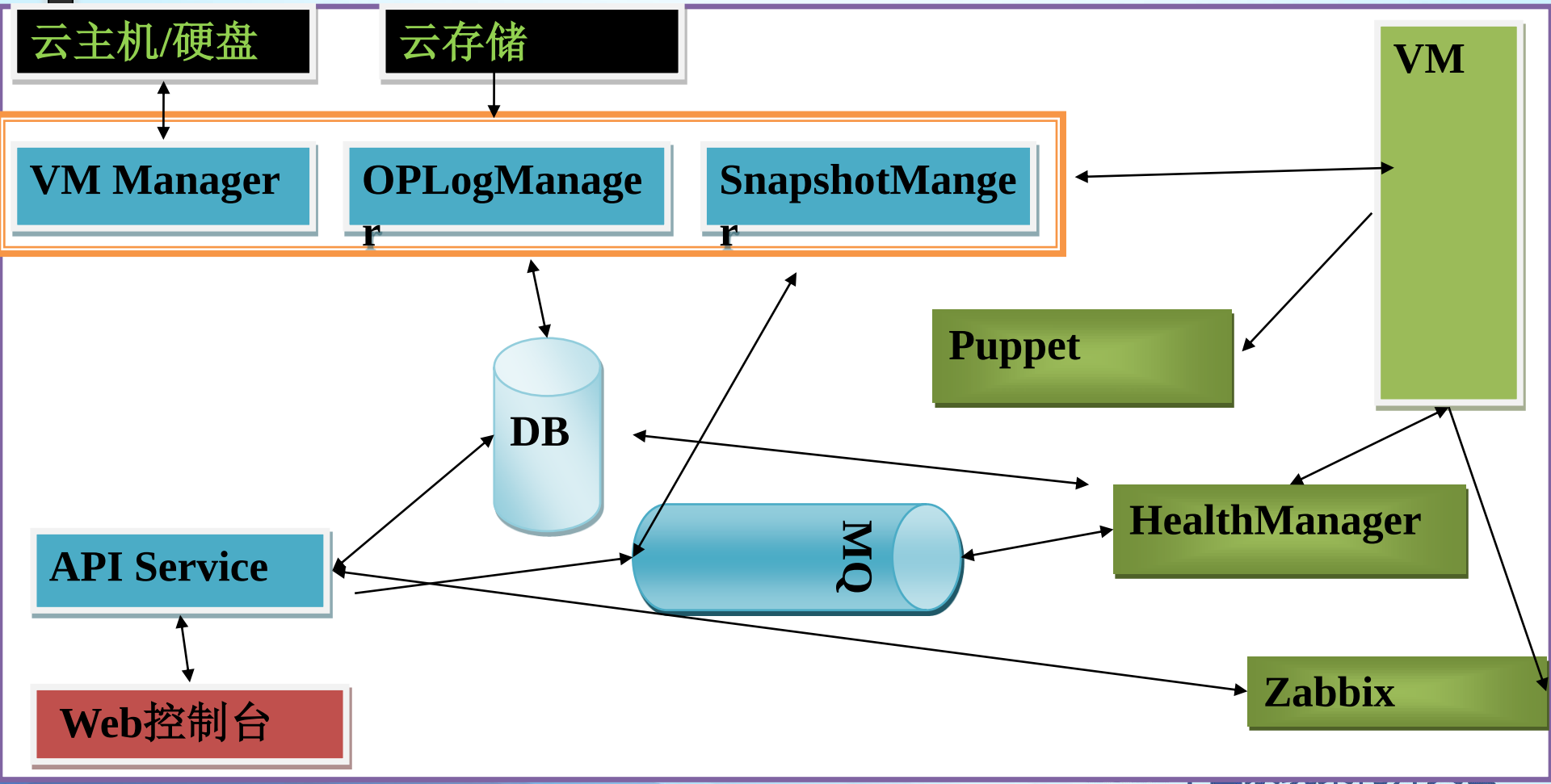
最后的架构

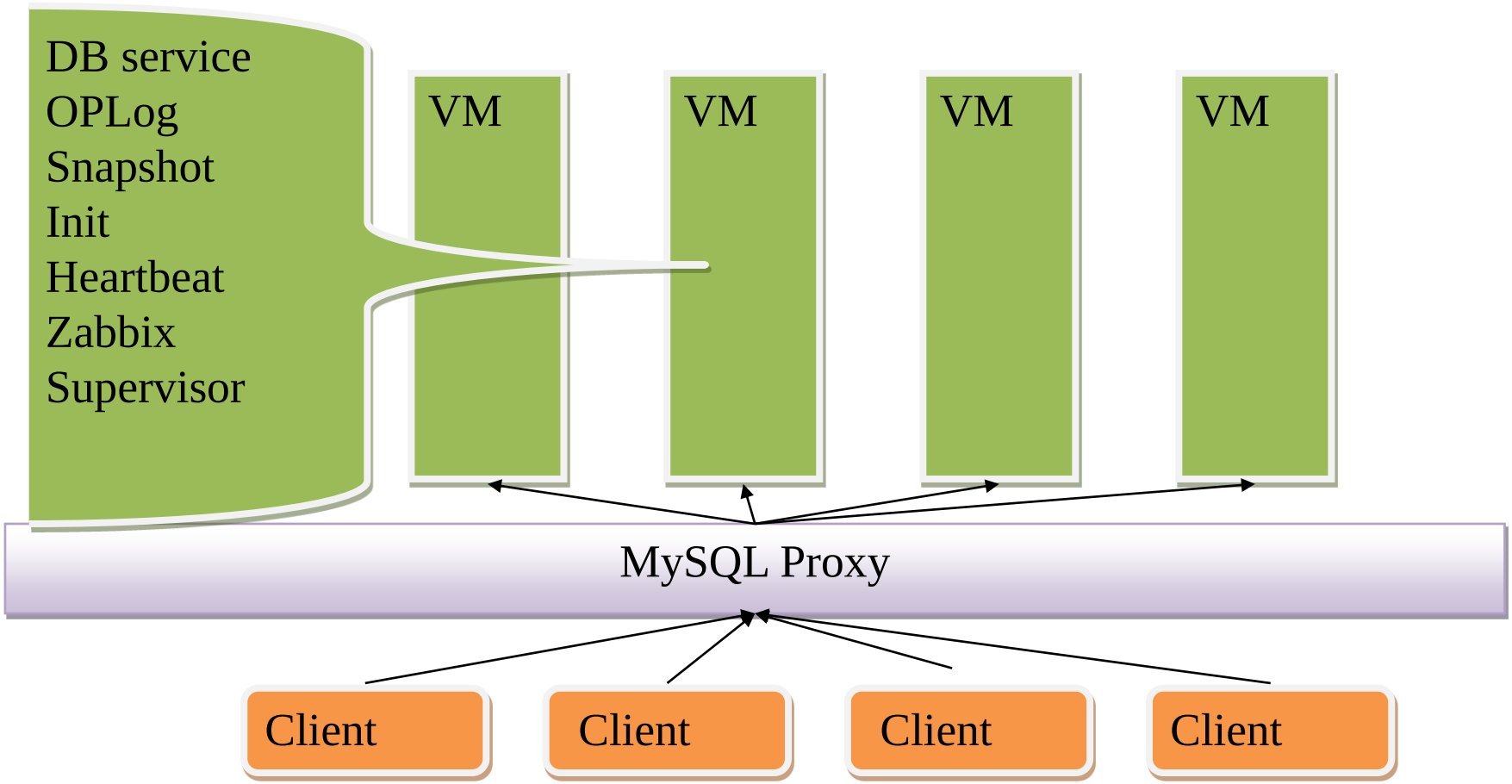
云主机

云硬盘

什么是云硬盘

- **弹性扩展**
  - 用户可独立于云主机申请云硬盘，每块云硬盘空间大小从1G到1T，单台云主机上即可挂载多块云硬盘，从而可以使其空间容量最大扩展到几十T。
- **数据高可靠**
  - 每份云硬盘数据在后台都存有多份冗余，并实时同步，保证不受单机故障影响。
- **独立持久化**
  - 每块云硬盘可以挂载到任意一台云主机上，两者隶属于不同的生命周期，当云主机被删除时，云硬盘数据仍然存在，并可以挂载到其它的云主机上进行访问


云主机/硬盘
云存储
VM
VM Manager
OPLogManager
SnapshotManger
Puppet
DB
MQ
HealthManager
API Service
Zabbix
Web控制台


DB service
OPLog
Snapshot
Init
Heartbeat
Zabbix
Supervisor
VM
VM
VM
VM
MySQL Proxy
Client
Client
Client
Client

Agenda

- Database As A Service (数据库云)介绍
- 数据库云整体架构
- **MySQL云的实现**
- MongoDB云的实现
- 未来的趋势

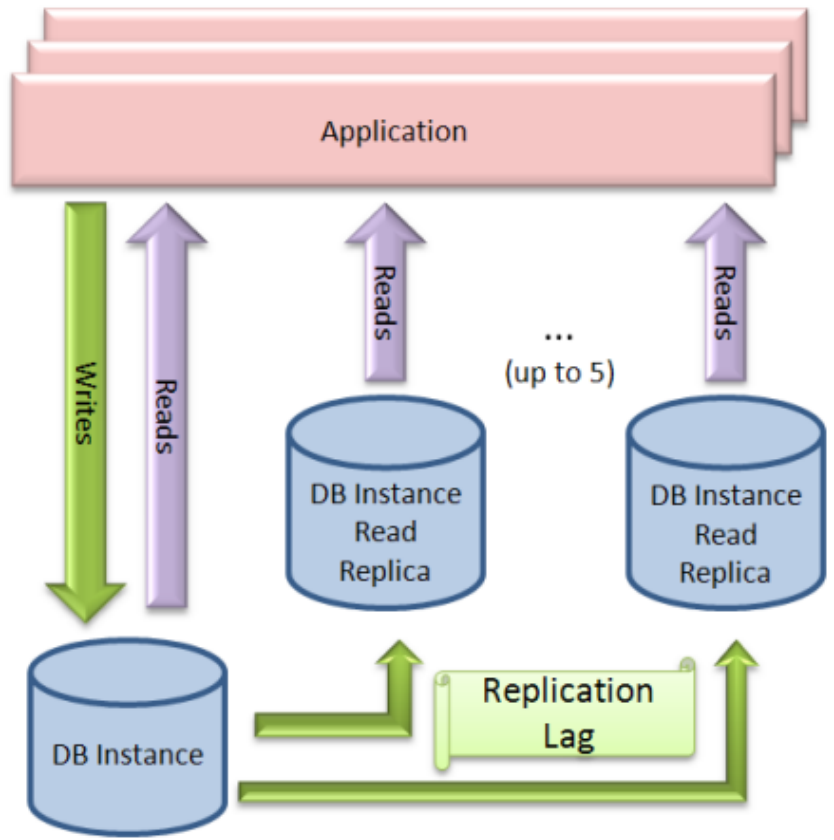
MySQL Snapshot

- EBS(云硬盘)+LVM
- Mylvmbackup
- Snapshot保存到云存储
- 使用go写的脚本边tar边上传
  - 本地可能没有空间可以存放tar包
  - 节约时间
  - 控制资源使用

MySQL Binlog上传

- 每5分钟进行一次flush logs
- Binlog上传到云存储
- 监控binlog的连续性
- 上传信息上报数据库
- 容错/可任意时刻kill、重启
- 可接收远程命令

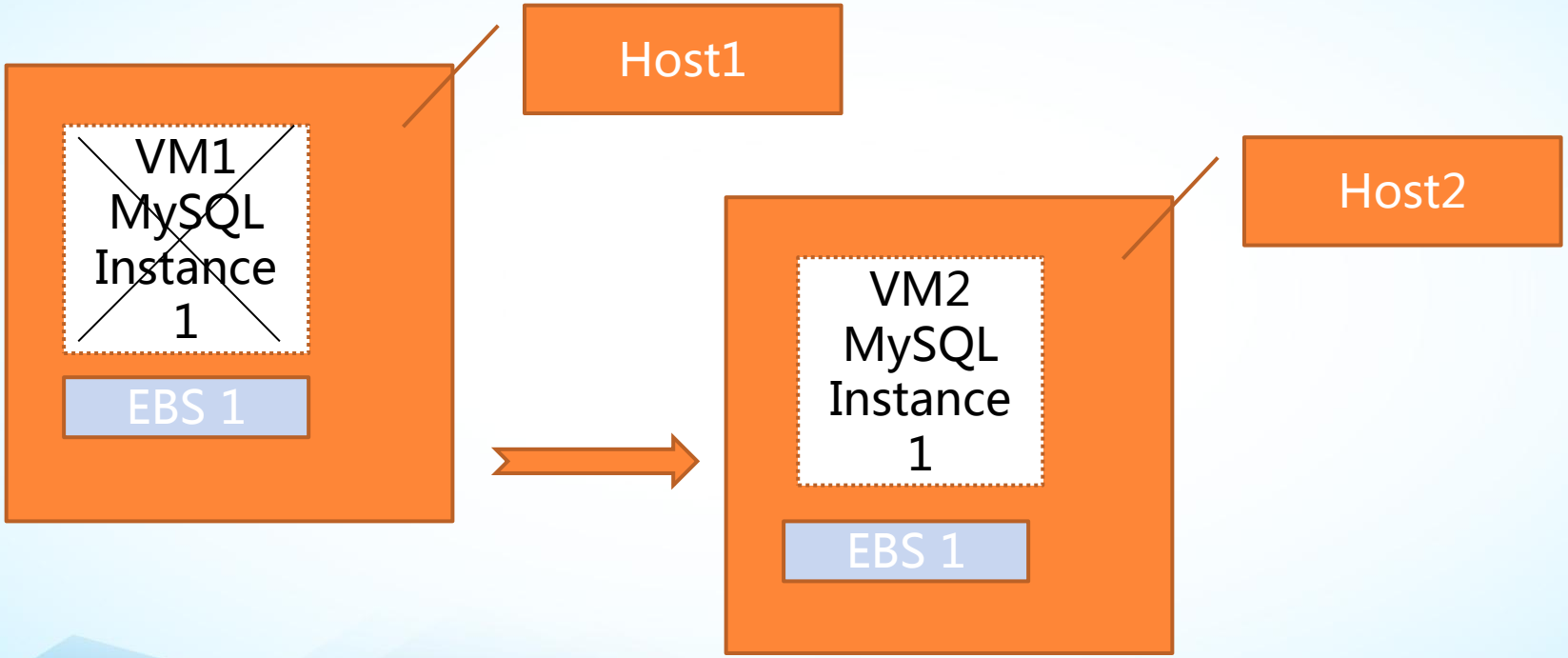
从节点支持


Application
Writes
Reads
Reads
...
(up to 5)
Reads
DB Instance
Read
Replica
DB Instance
Read
Replica
DB Instance
Replication
Lag

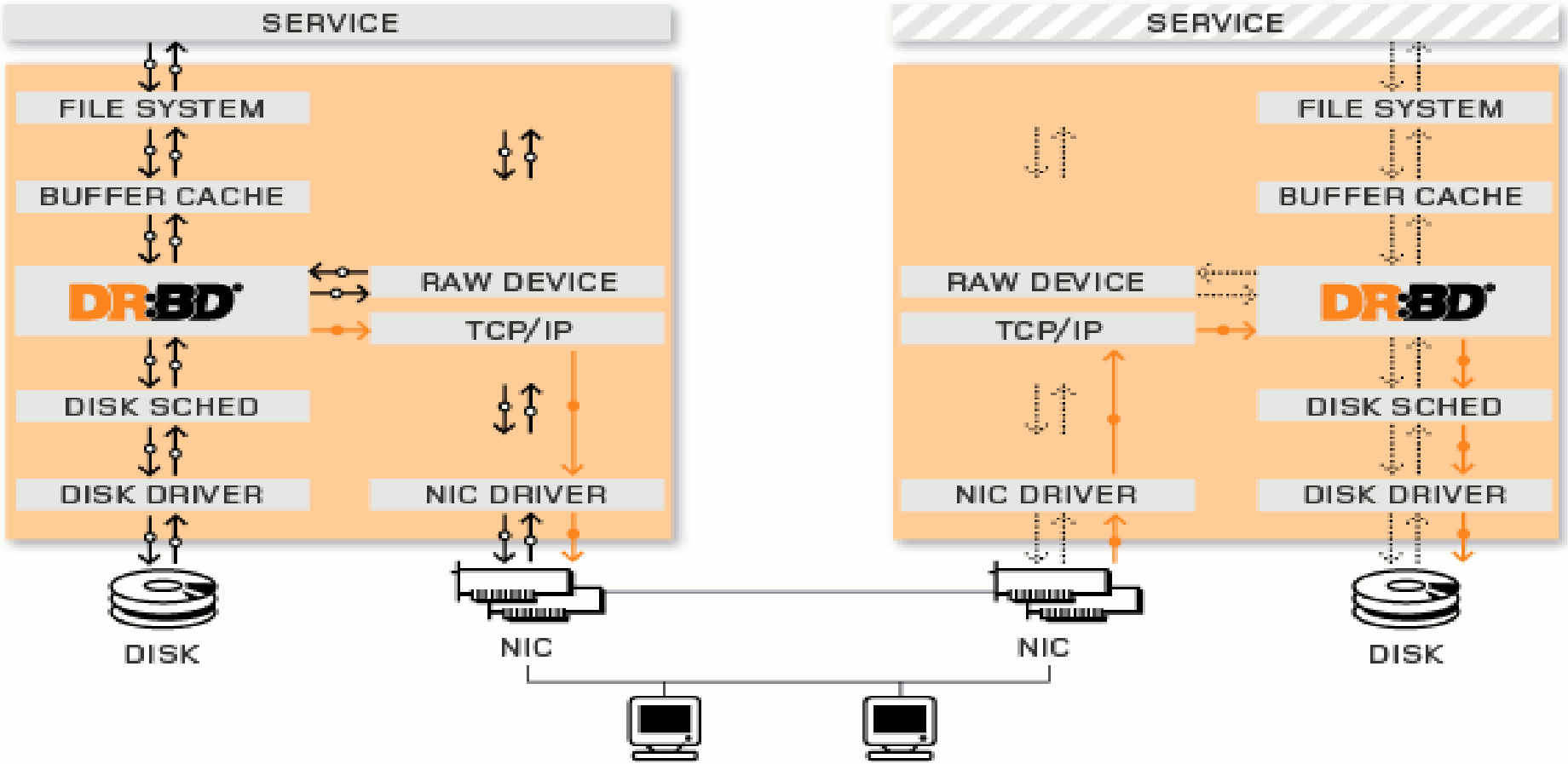
从节点支持

- 拿到最新的snapshot
- Replay最近的oplog
- 再从主进行同步
- 避免增加从节点时对主节点增加压力

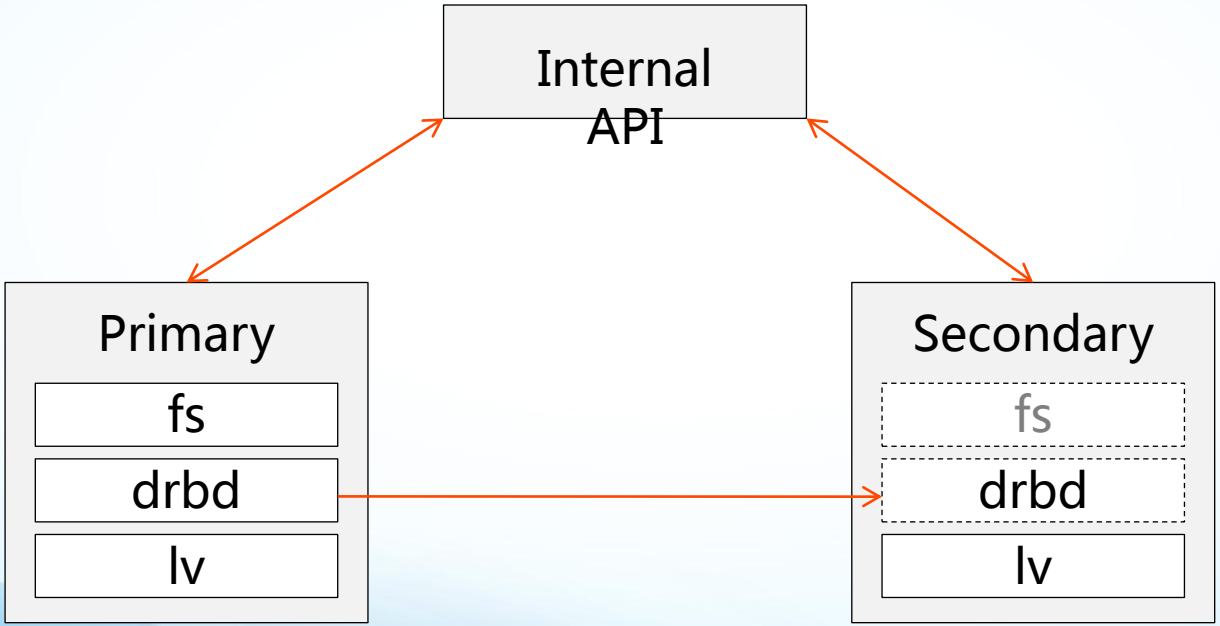
故障迁移的实现


Host1
VM1
MySQL
Instance
1
EBS 1
Host2
VM2
MySQL
Instance
1
EBS 1

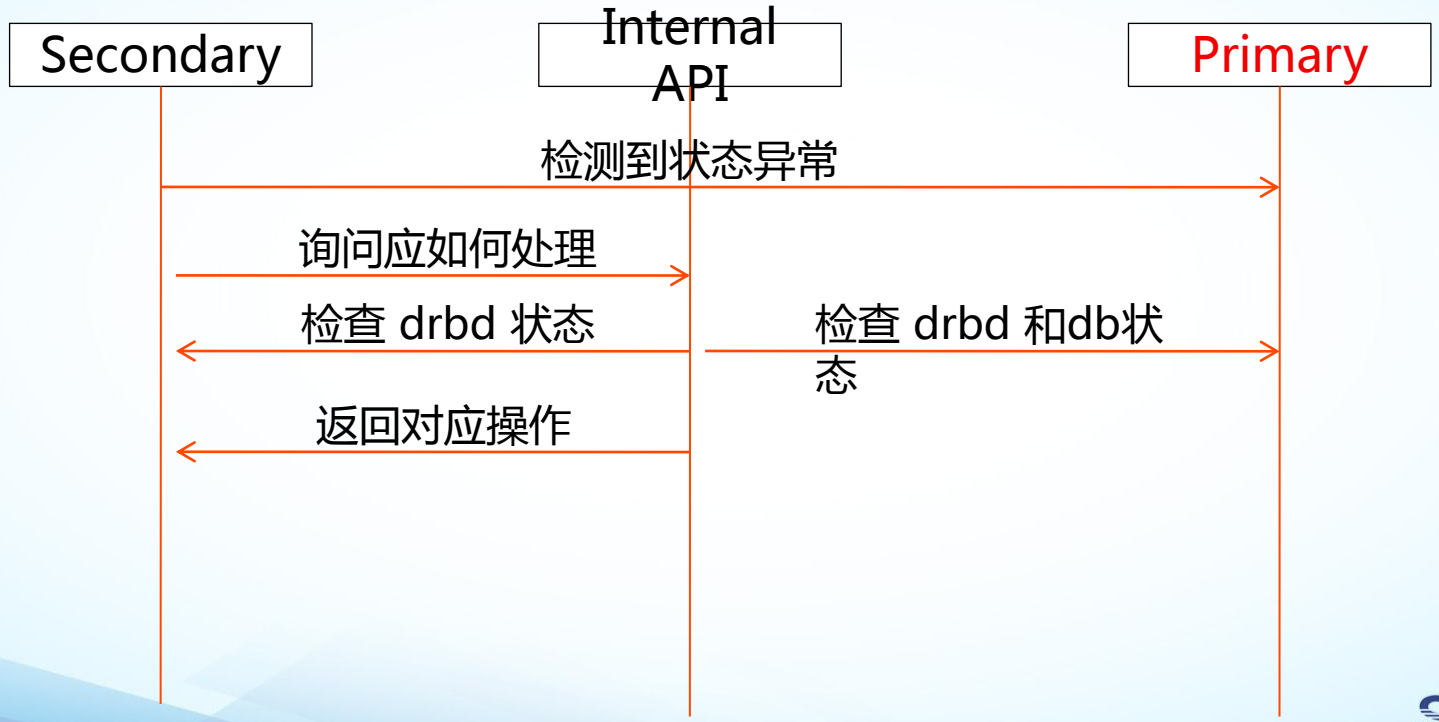
真正高可用的实现-DRBD



实际HA Solution



HA Solution


Secondary
Internal API
Primary
检测到状态异常
询问应如何处理
检查 drbd 状态
检查 drbd 和db状态
返回对应操作

Agenda

- Database As A Service (数据库云)介绍
- 数据库云整体架构
- MySQL云的实现
- **MongoDB云的实现**
- 未来的趋势

MongoDB云模式

- Master/Slave模式
  - 与MySQL服务形式与API上保持一致
  - 可以只申请一台机器
  - 可以增加从节点
- Replica Set模式
  - 至少申请3台机器 (2大1小)
  - 可以增加从节点

MongoDB Snapshot

- EBS(云硬盘)+LVM
- 方案1：
  - fsync mongodb -> lock mongodb -> fsync mongodb -> lvm snapshot -> unlock mongodb
  - 缺点:lock的时候mongodb不可用，进程不能死，死了就不能unlock
- 方案2
  - 开启journal选项
  - 直接lvm snapshot
  - 缺点: 性能上有些损失, 根据snapshot进行恢复时需要先恢复journal,恢复时间会变长

MongoDB Oplog上传

- 后台进程一直读取local上的Oplog
- 满5分钟或者10M写一次磁盘
- Oplog文件上传到云存储(上传进程和写进程分离)
- 监控Oplog的连续性
- 上传信息上报数据库
- 容错/可任意时刻kill、重启
- 可接收远程命令

Agenda

- Database As A Service (数据库云)介绍
- 数据库云整体架构
- MySQL云的实现
- MongoDB云的实现
- **未来的趋势**

未来的趋势

- 基于SSD的数据库服务
- 基于内存的数据库服务
- 云平台自己定义的数据库服务
  - Amazon DynamoDB/Simple DB
- Scale up 的关系型数据库服务
  - Xeround MySQL

Q&A