



网易云 beta
cloud.netease.com

开源云端数据库架构

网易 杭州研究院

后台技术中心 郭忆

新浪微博：@郭忆_宝

目录

- 网易云数据库
- 系统架构
- 高可用设计
- 监控运维
- 在线ScaleUp和ScaleOut
- 未来与展望

产品方使用MySQL遇到的问题？

- 硬件采购周期长，沟通协调成本高，数据库部署的需求难以得到快速响应。
- 硬件资源利用率低，难以做到按需使用，弹性扩容。
- 服务可用性差，数据可靠性难以保证。
- 运维自动化程度低，人力成本高。
- 监控报警不够完善，出现问题缺少系统诊断方法。



高额的成本+差的用户体验

云计算

■ 云计算：



■ 云平台的三个层次：



DBaaS

DBaaS(DataBase as a Service) , 数据库即服务 , 云环境下的数据库托管平台 , 将数据库本身作为一种云端资源 , 以服务的方式提供给应用开发人员。

主要优势 :

云的特性

- 可伸缩
- 低成本
- 资源利用率高
- 向用户屏蔽软硬件升级

托管平台

- 一键部署
- 自助服务式数

可编程性

- 通过API定义和控制数据库资源
- 支持应用参与数据库的自动化管理

云计算

- 云计算不是天生为数据库设计的，数据库“云”化过程中，存在以下挑战：

云托管的服务器可靠性下降，导致的服务可用性和数据可靠性难以保证

如何利用云的弹性资源分配，提供在线的数据库扩展服务

资源规划难以精准

故障排查难度增加

网易私有云环境



网易私有云环境



■ 系统特色

重视服务质量

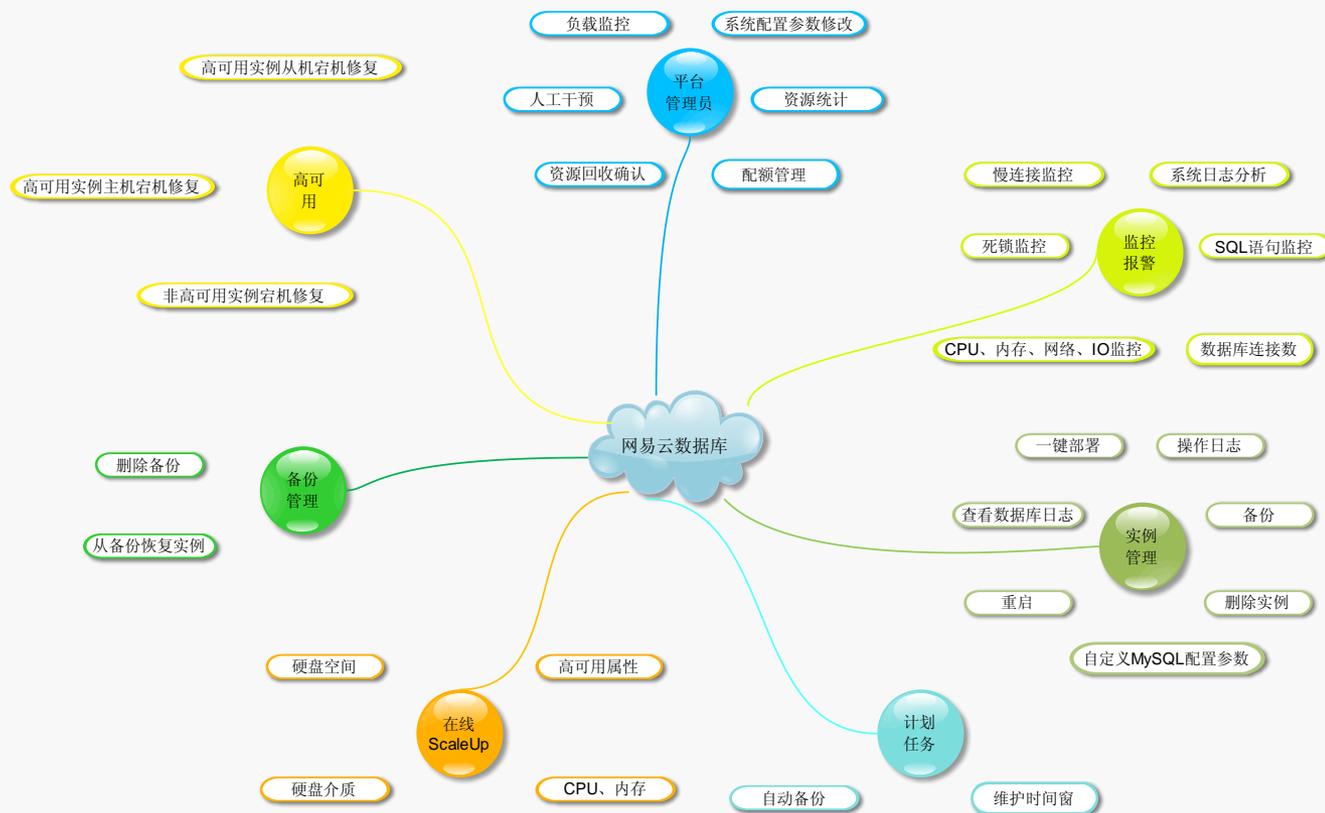
- 拒绝超售
- 高可用&高可靠

重视选择和控制能力

- 自定义CPU和内存组合
- 开放配置参数修改

网易私有云环境

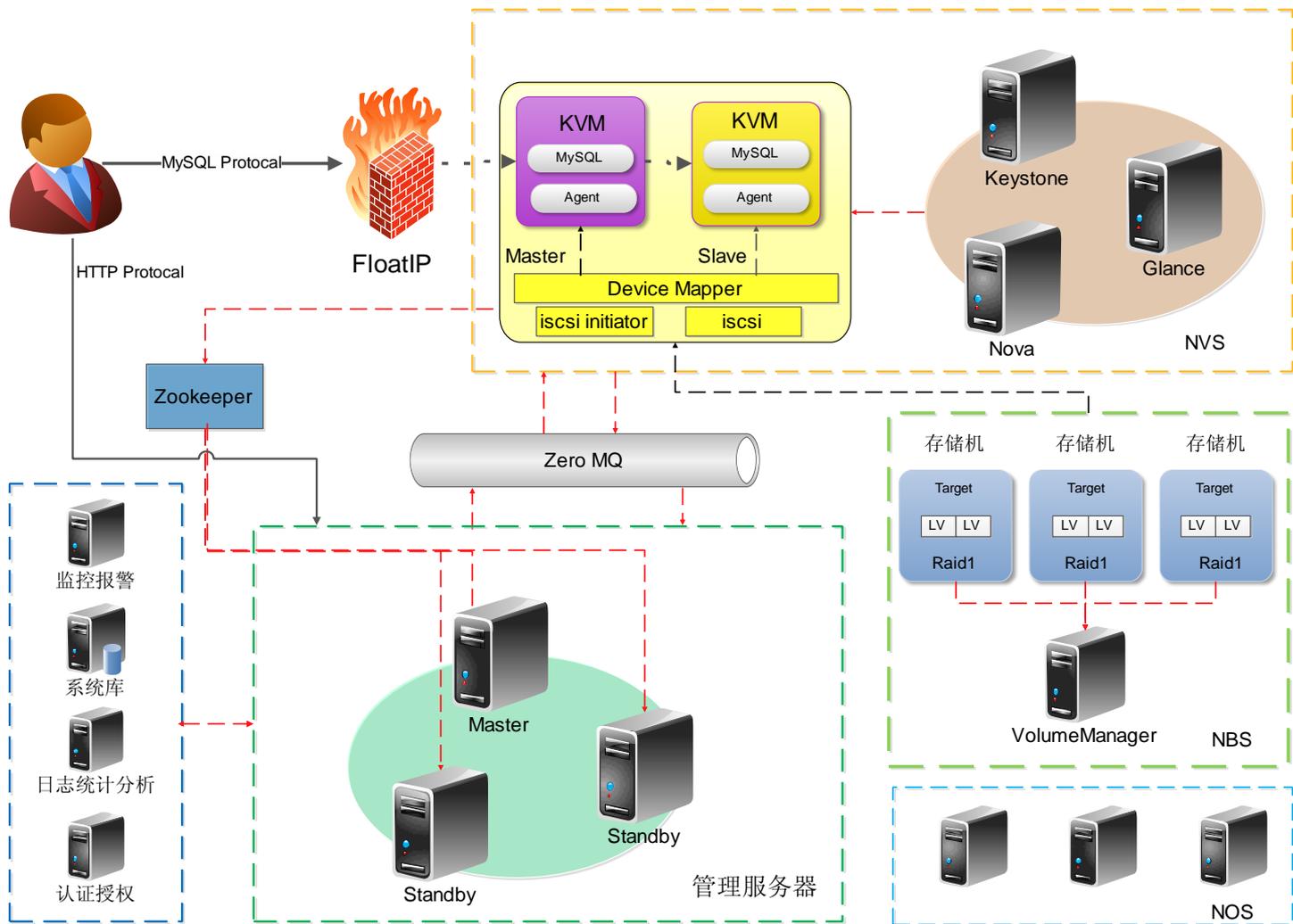
- Netease RDS是由网易数据库技术团队基于开源的MySQL打造的关系数据库云托管平台，构建于网易私有云IaaS服务之上，面向网易众多的互联网和移动终端产品，旨在解决当前数据库管理遇到的诸多问题。



目录

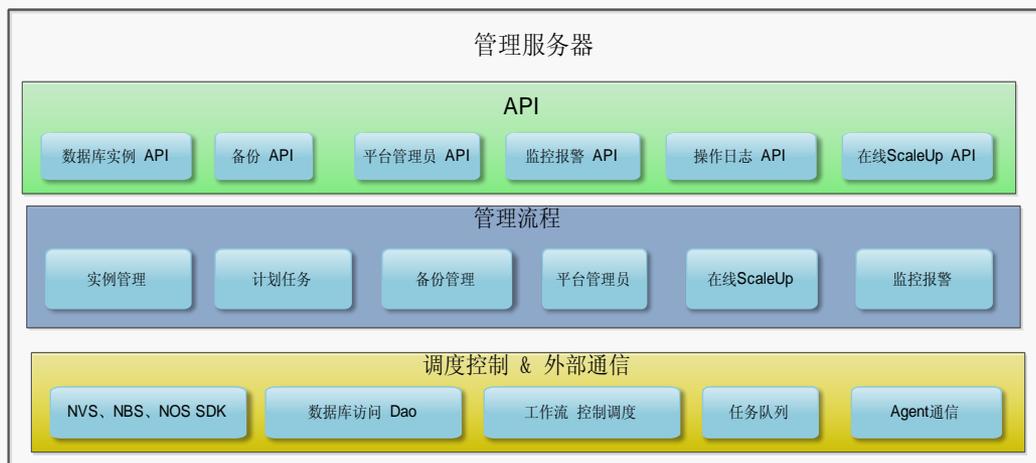
- 网易云数据库
- 系统架构
- 高可用设计
- 监控运维
- 在线ScaleUp和ScaleOut
- 未来与展望

系统架构



管理服务器

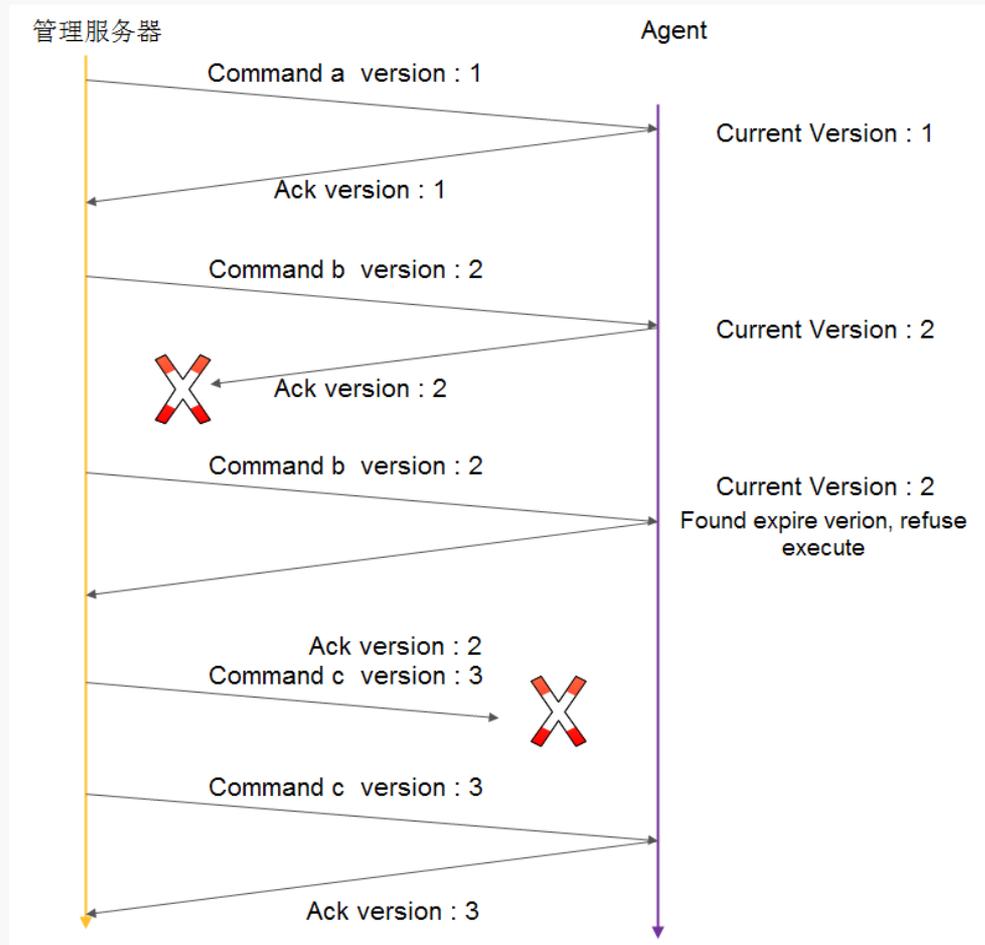
- 高可用
- workflow
- 任务队列



其他开源组件

■ ZeroMQ

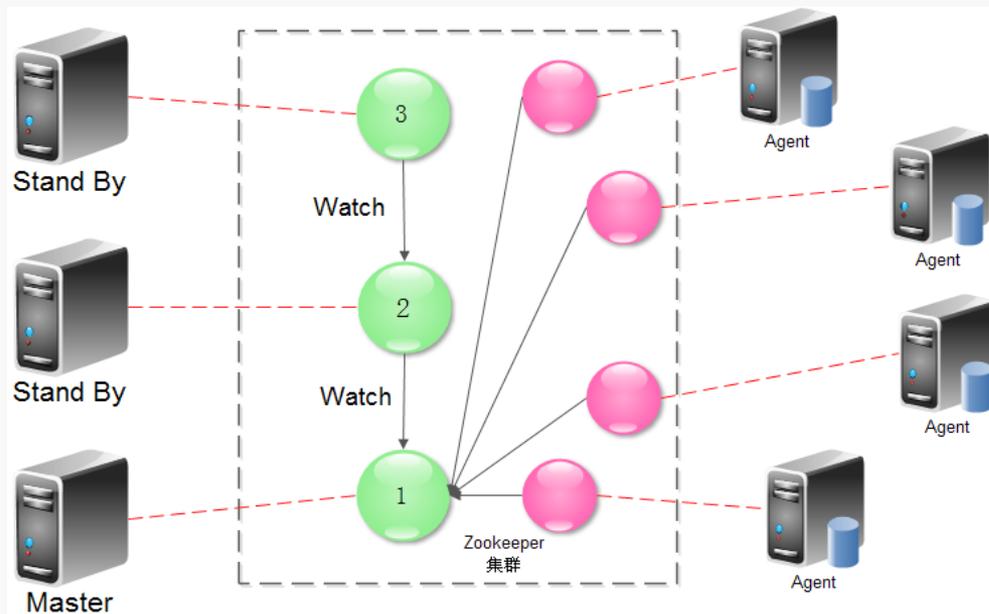
- 轻量级
- 异步传输
- 消息序列化
- 消息持久化



其他开源组件

■ Zookeeper

- 选主
- 分布式锁
- 配置管理



目录

- 网易云数据库
- 架构架构
- 高可用设计
- 监控运维
- 在线ScaleUp和ScaleOut
- 未来与展望

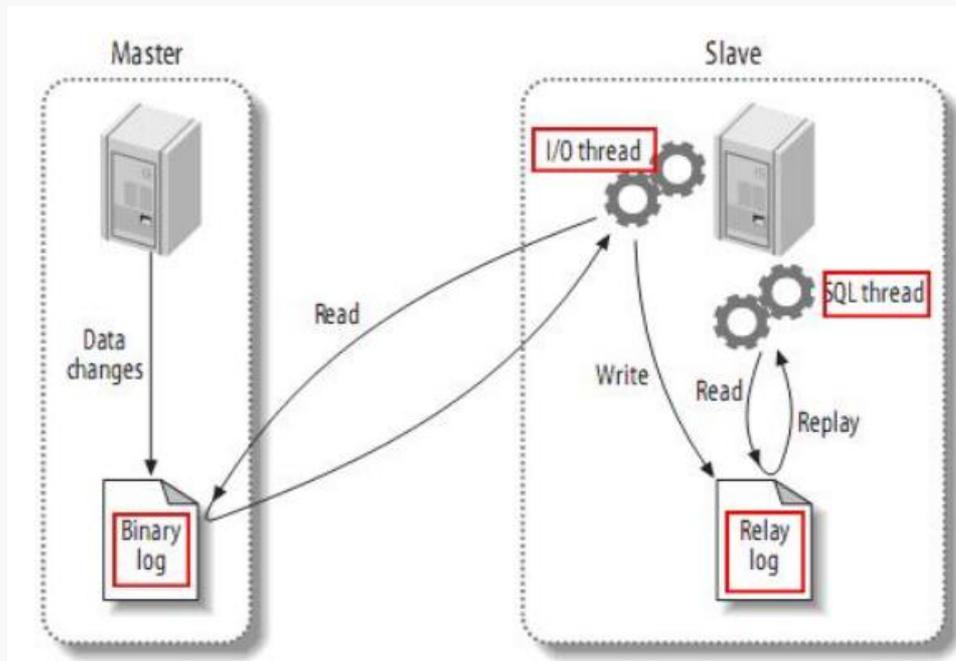
现有业界高可用方案

	可用性 (停服时间)	可靠性 (数据一致性)	对性能损耗	对应用透明	其他限制
DRBD	低	高	有	是	配置管理复杂
共享磁盘	低	高	无	是	无
binlog高可靠	高	高	有	是	需额外开发binlog监听工具
MMM	高	低	无	是	无
MHA	高	高	无	是	主机binlog是高可靠的

MySQL复制

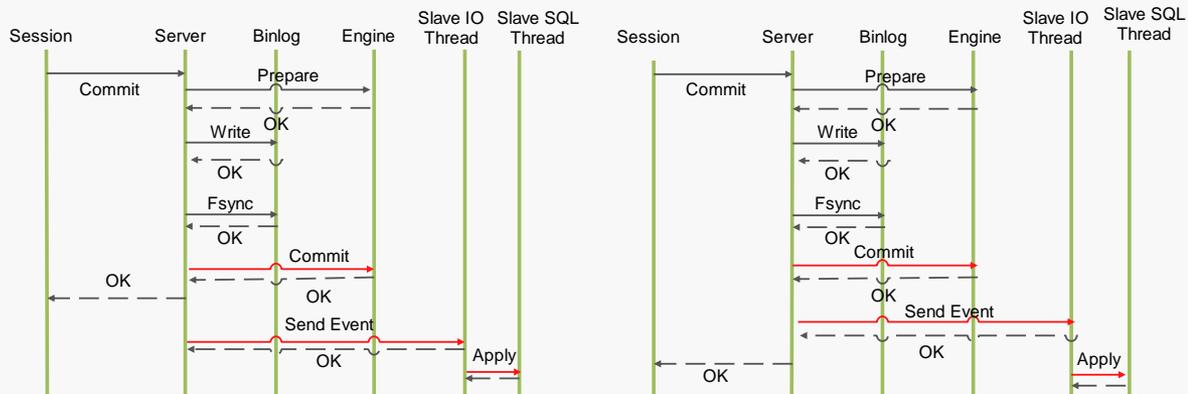
■ 主要挑战：

- 数据一致性
- 实时切换



数据一致性

■ 虚拟同步复制



异步复制

半同步复制



虚拟同步复制

全同步复制

数据一致性

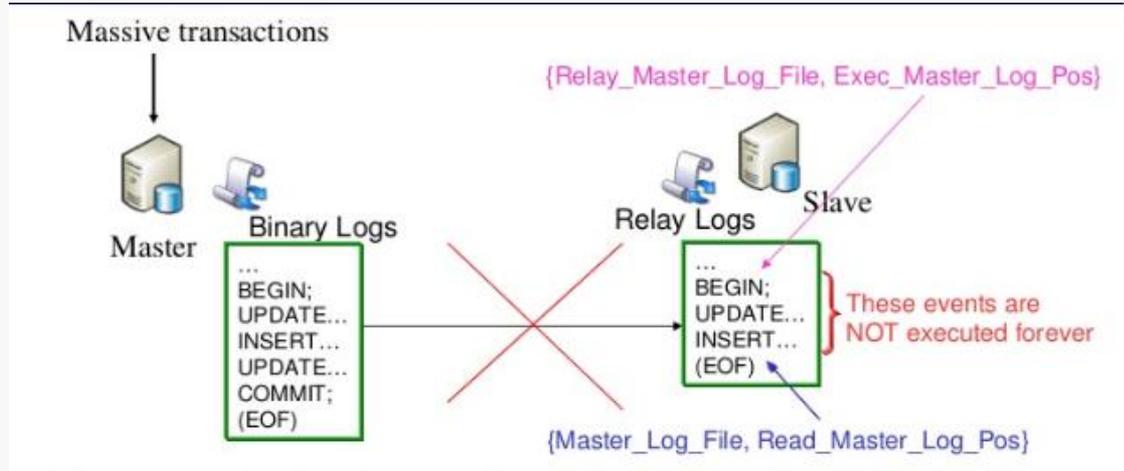
■ IO Thread 隐患

- relay_log_recovery

■ SQL Thread

- Crash safe

■ Partial Transaction



实时切换

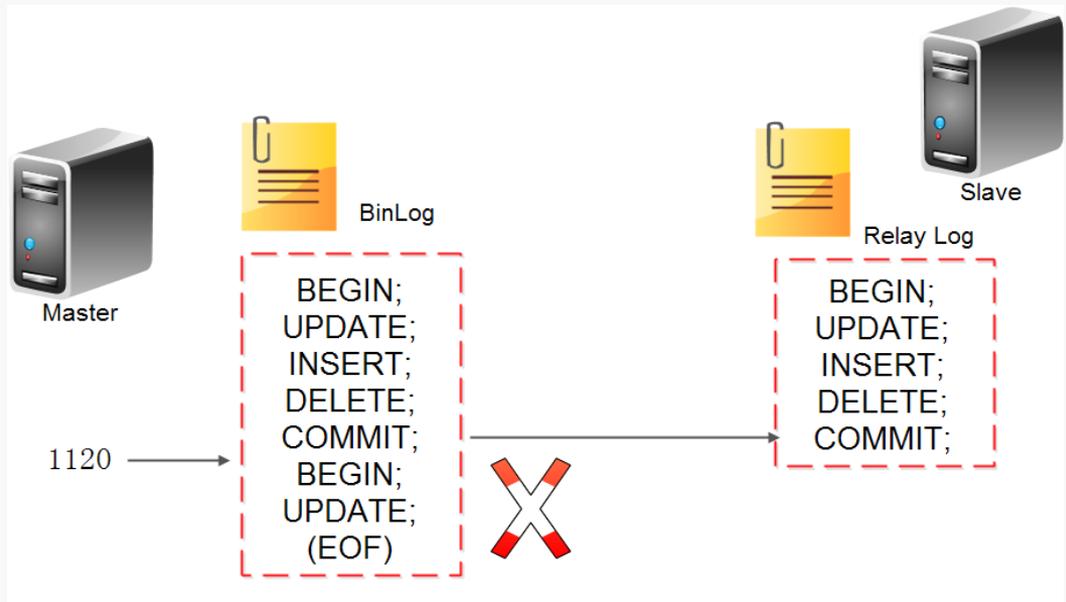
■ Batch Commit

■ 并行复制

- MariaDB
- MySQL 5.7

修复

■ binlog裁剪



目录

- 网易云数据库
- 系统架构
- 高可用设计
- 监控运维
- 在线ScaleUp和ScaleOut
- 未来与展望

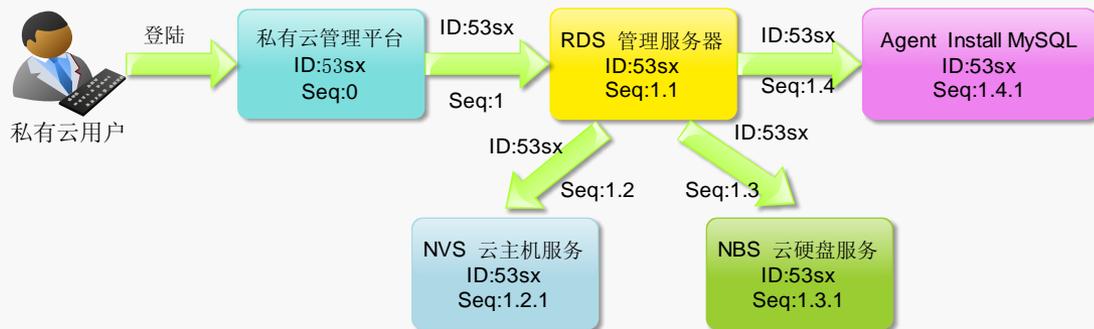
系统监控

■ 监控方式

- 主动
- 被动

■ 统一日志

Key	Value
ID	用户请求唯一标识符
Seq	调用序号，格式为“x.x.x”，同一请求，模块内部每次写入日志或者调用其他模块前加1，跨模块加层，在服务和模块之间传递
Timestamp	时间戳
Level	日志输出级别
Module	模块名称
IP	写入日志的服务器
Identifier	用户标识
Op	操作名称
Object	对象信息，Json格式
Description	详情，异常需包含全部堆栈信息



系统监控

Logtracer

导航菜单 <<
主页 云计算(线上) x

系统管理

- [云计算](#)
- [易信\(vixin_nginx\)](#)
- [推送平台](#)
- [云计算\(线上\)](#)
- [云计算\(联调\)](#)
- [vixin_app](#)
- [nos_nginx](#)
- [datastream_log](#)
- [vixin_apptest](#)

HBASE(cloud_assemble_product_ops)日志查询

Date From: 2013-10-09 15:49:08 To: 2013-10-09 16:49:08 rowkeyPrefix: 6c4a1f2f-b14c-4f1f-b0a6-6de1871defa3 搜索 共 308 条记录

	id	seq	module	op	level	timestamp
1	6c4a1f2f-b14c-4f1f-b0a6-6de1871defa3	2.1	RDS	CreateDBInstance	debug	1381308245543
2	6c4a1f2f-b14c-4f1f-b0a6-6de1871defa3	2.2	RDS	CreateDBInstance	info	1381308245640
3	6c4a1f2f-b14c-4f1f-b0a6-6de1871defa3	2.3	RDS	CreateDBInstance	info	1381308245643
4	6c4a1f2f-b14c-4f1f-b0a6-6de1871defa3	2.4	RDS	CreateDBInstance	info	1381308245643
5	6c4a1f2f-b14c-4f1f-b0a6-6de1871defa3	2.5	RDS	CreateDBInstance	info	1381308245736
6	6c4a1f2f-b14c-4f1f-b0a6-6de1871defa3	2.6	RDS	CreateDBInstance	info	1381308245737
7	6c4a1f2f-b14c-4f1f-b0a6-6de1871defa3	2.7.1	NVS	__call__	info	1381308252338
8	6c4a1f2f-b14c-4f1f-b0a6-6de1871defa3	2.7.2	NVS	__process_stack	info	1381308252349
9	6c4a1f2f-b14c-4f1f-b0a6-6de1871defa3	2.7.3	NVS	__process_stack	debug	1381308252364
10	6c4a1f2f-b14c-4f1f-b0a6-6de1871defa3	2.7.4	NVS	__process_stack	info	1381308252705
11	6c4a1f2f-b14c-4f1f-b0a6-6de1871defa3	2.7.4	NVS	reserve	debug	1381308252451
12	6c4a1f2f-b14c-4f1f-b0a6-6de1871defa3	2.8	RDS	CreateDBInstance	info	1381308246291
13	6c4a1f2f-b14c-4f1f-b0a6-6de1871defa3	2.9	RDS	CreateDBInstance	info	1381308246291
14	6c4a1f2f-b14c-4f1f-b0a6-6de1871defa3	2.10.1	NVS	__call__	info	1381308252892
15	6c4a1f2f-b14c-4f1f-b0a6-6de1871defa3	2.10.2	NVS	__process_stack	info	1381308252904
16	6c4a1f2f-b14c-4f1f-b0a6-6de1871defa3	2.10.3	NVS	__process_stack	debug	1381308252917
17	6c4a1f2f-b14c-4f1f-b0a6-6de1871defa3	2.10.4	NVS	reserve	debug	1381308252970
18	6c4a1f2f-b14c-4f1f-b0a6-6de1871defa3	2.10.5.1	NVS	process_request	info	1381308253041
19	6c4a1f2f-b14c-4f1f-b0a6-6de1871defa3	2.10.5.1	NVS	process_request	info	1381308253041

10
显示1到10,共308记录

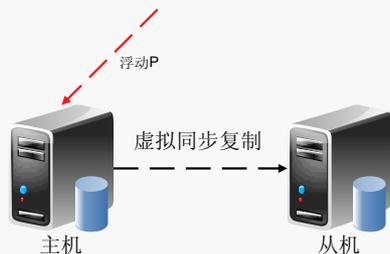
目录

- 网易云数据库
- 架构架构
- 高可用设计
- 监控运维
- 在线ScaleUp和ScaleOut
- 未来与展望

扩容

■ 在线Scale Up

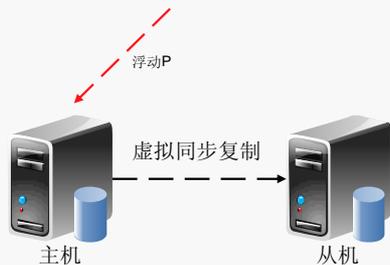
■ 在线Scale Out



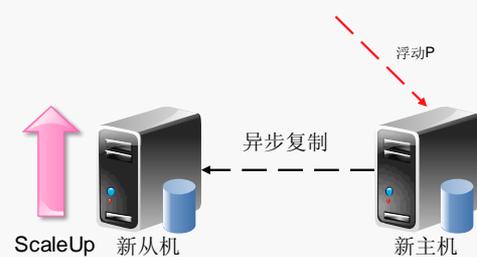
(1) 主备虚拟同步复制



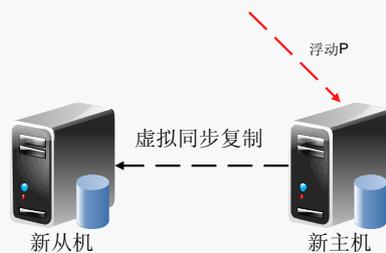
(2) 主机切异步复制、从机ScaleUp



(3) 重新建立虚拟全同步复制



(4) 主从切换、新从机ScaleUp



(5) 建立新从机到新主机的虚拟同步复制

目录

- 网易云数据库
- 架构架构
- 高可用设计
- 监控运维
- 在线ScaleUp和ScaleOut
- 未来与展望

■ 未来工作

- 多版本管理
- LXC (Linux Container)
- 支持Memcached



网易云 **beta**
cloud.netease.com