新浪云计算
sinacloud.com

Alibaba Developer Conference

云计算PaaS平台的Key-Value服务

Sina App Engine

陈磊 @simpcl

sina新浪

分享内容

新浪云计算
sinacloud.com

- 背景介绍
- 关于SAE
- SAE的整体架构
- SAE的KVDB服务
- Key-Value服务的改进

sina新浪

背景介绍

- 什么PaaS？

  PaaS是Platform-as-a-Service的缩写，意思是平台即服务。

- Web开发者认为

  PaaS = Web运行环境 + 一堆分布式服务

- 从PaaS实现角度

  隔离、统计、安全、扩展

关于SAE

新浪云计算
sinacloud.com

一、什么是SAE？

- Sina App Engine，一个公有云PaaS平台；
- SAE选择国内流行最广的Web开发语言PHP作为首选的支持语言；
- 现在同时支持Java和Python语言；
- SAE提供了一系列分布式服务，包括了多种计算类服务和存储类服务；

SAE就是简单高效的Web应用开发、运行平台

sina新浪

二、SAE不仅仅是PaaS

新浪云计算
sinacloud.com

Sina Web Services

新浪云计算 sinacloud.com

三、SAE的发展历史

新浪及合作第三方支持

微游戏、微盘、校园微博、 Q微博、互联网的那点事...

云计算产品研发

计算类服务 、存储类服务、云应用商店、云服务商店、CDN、IaaS平台

Sina App Engine 两周年 http://sae.sina.com.cn

新浪云计算 SinaCloud.com

node JS python Java

2009.11 Sina App Engine alpha版上线 SAE诞生

2010.2 Sina App Engine alpha2版上线 TmpFS 支持

2010.9 Sina App Engine Beta版上线 首个公测版本发布

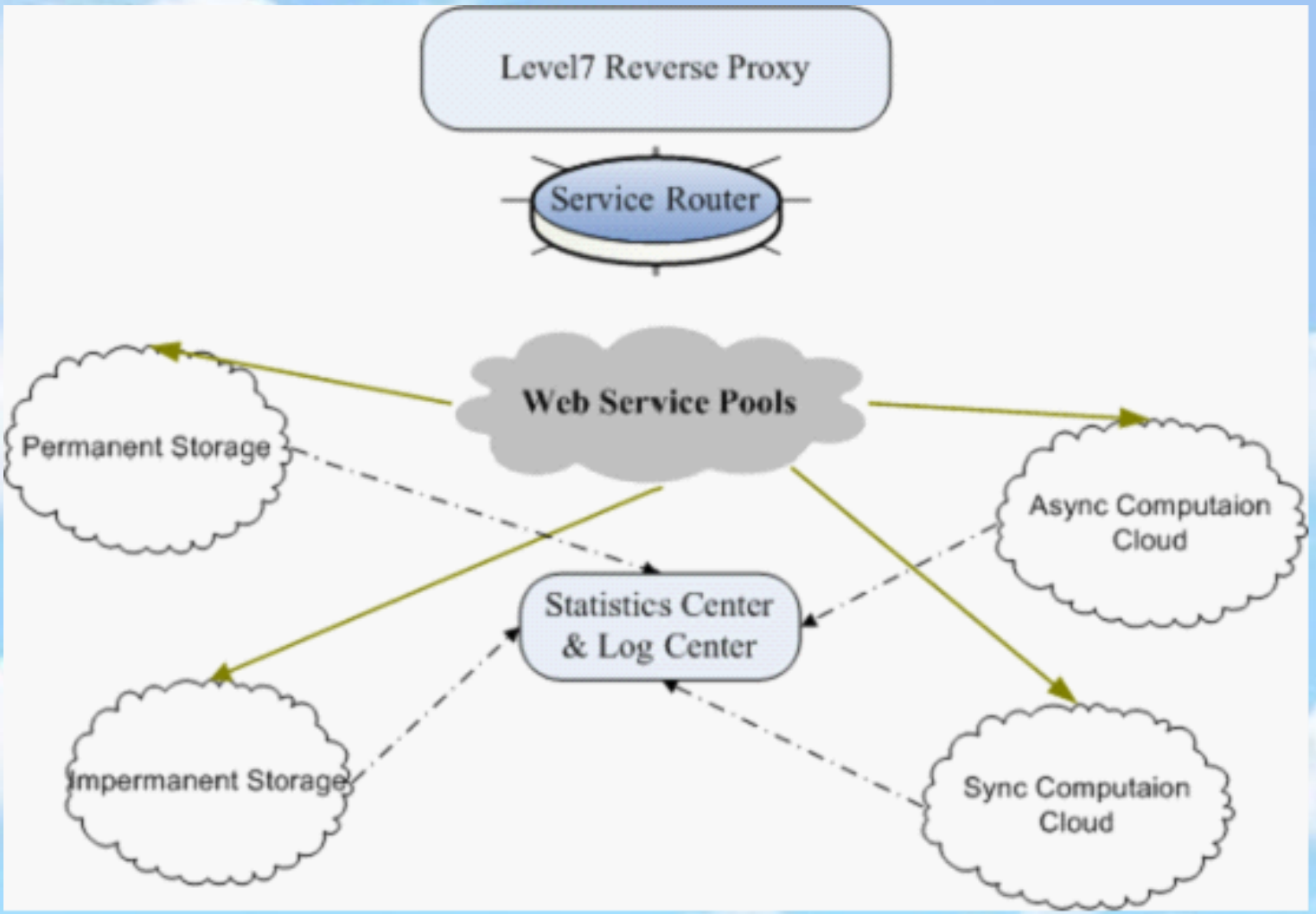
2010.10.10 微盘上线

2011.5.18 Sina App Engine Beta2版上线 正式开放注册

2011.7 开通支付 云应用商店上线

sina 新浪

SAE的整体架构



SAE从架构上采用分层设计，自上而下分别为：反向代理层、路由逻辑层、Web计算服务池、日志和统计中心以及各个分布式服务。

SAE的整体架构

1. 反向代理层

a. 基于HTTP的反向代理，工作在最外层

b. 与后端的Web服务池相连，负责接收、分析、转发和响应用户的HTTP请求

c. 同时提供负载均衡、健康检查等功能

2. 路由逻辑层

根据请求的唯一标识，快速的映射（O(1)时间复杂度）到相应的Web服务池，如果发现映射关系不存在或者错误，则给出相应的错误提示；

该层对用户隐藏了很多具体地址信息，使开发者无需关心服务的内部实际分配情况。

SAE的整体架构

3. Web计算服务池

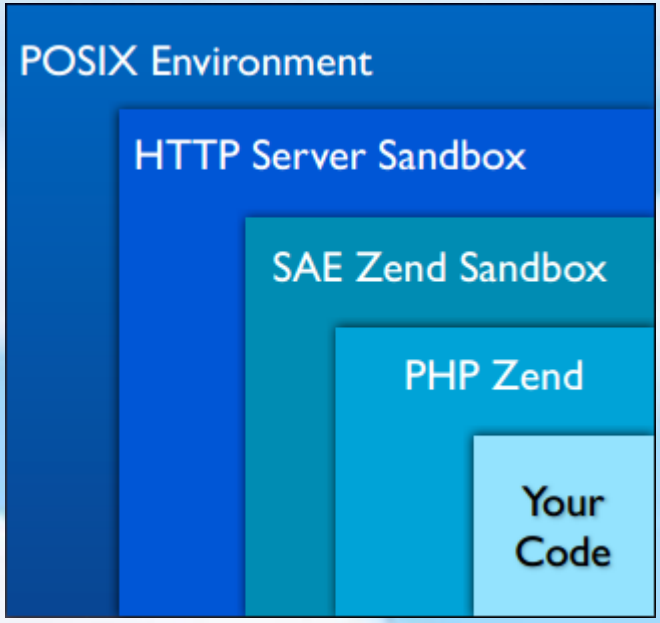
a. 由一些不同特性的Web服务池组成，按照不同的SLA提供不同级别的服务；

b. 一个Web服务池由一些相同属性的Web服务器组成，通过前端的反向代理扩展服务能力；

c. 每台Web服务器上运行相应的Web运行时环境，其嵌入了相应的SAE沙盒。

d. 用户的代码最终通过相应Web运行时环境的API调用各种服务。

SAE的整体架构



SAE PHP SandBox

HTTP Server Sandbox

✓连接保护

✓请求统计

✓请求控制

✓libc函数保护（DLL注入）

✓… …

SAE Zend Sandbox

✓运行环境隔离

✓CPU控制

✓本地I/O限制

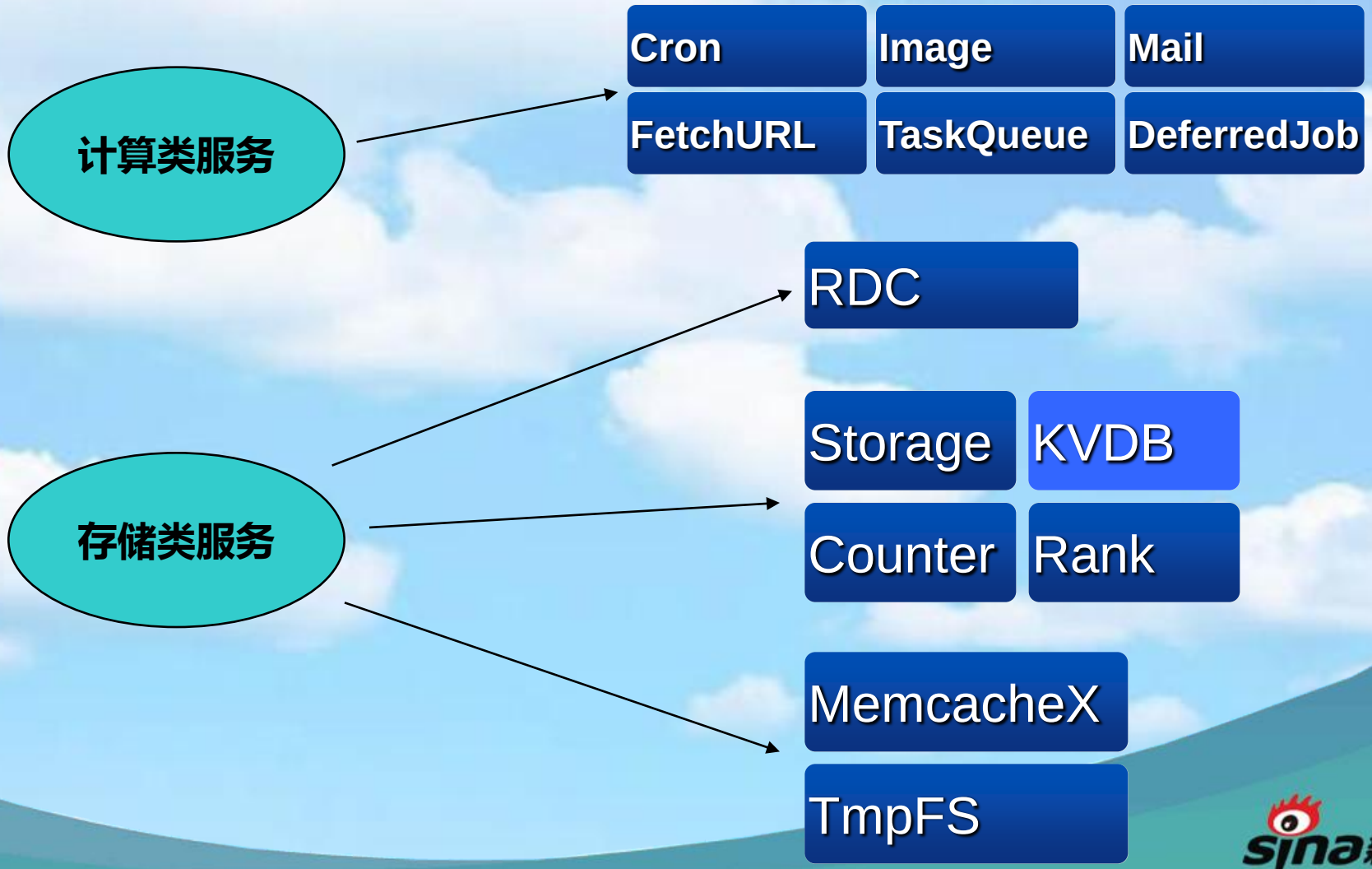
✓网络I/O改造

✓系统级API禁用及修改

✓… …

SAE的整体架构

4. 各种分布式服务



SAE的整体架构

5. 日志和统计中心

负责对用户所使用的所有服务进行统计和资源计费，并设定的分钟配额，来判定是否有非正常的使用。分钟配额描述了资源消耗的速度，当资源消耗的速度到达一个预警阈值时，SAE通知系统会提前向用户发出一个警告，提醒用户应用在某个服务上的使用可能存在问题，需要介入关注或处理，配额系统是SAE用来保证整个平台稳定的措施之一；日志中心负责将用户所有服务的日志汇总并备份，并提供检索查询服务。

SAE的KVDB服务

SAE的KV存储服务需求：

A.持久存储Key-Value数据

B.存储服务而非存储引擎

C.支持数据隔离、认证和统计

D.支持读写分离

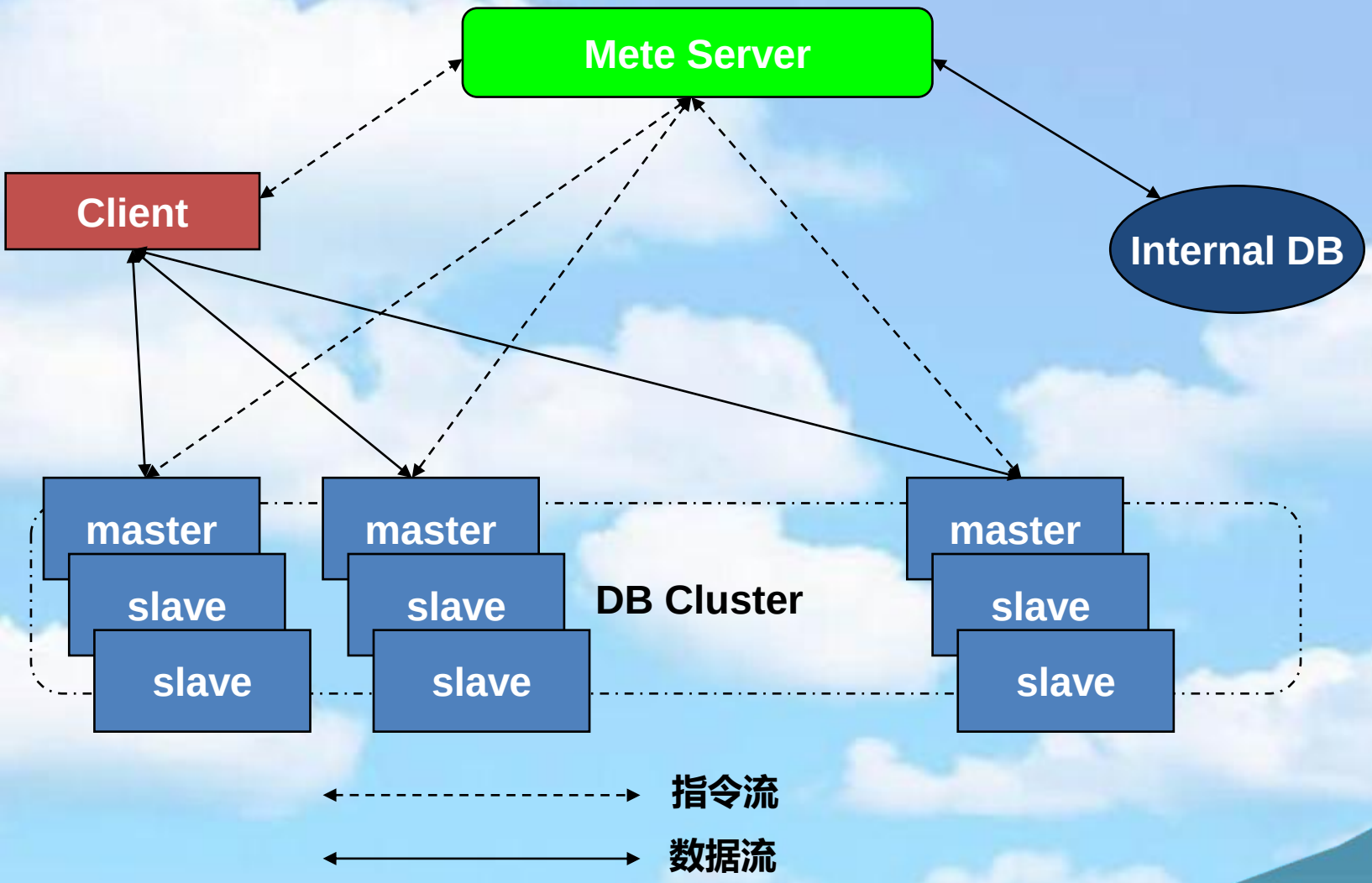
E.服务器宕机自动切换

F.服务可以任意水平扩展

G.支持重平衡、无缝迁移

H.服务API功能丰富、简单易用

SAE的KVDB服务



SAE KVDB 服务架构图

SAE的KVDB服务

SAE KV 服务的如何工作？

1.客户端从Web Runtime中获取当前请求所属的appkey；

2.客户端向Meta Server发出请求，获取该appkey相应的appname-key到实际存储节点的映射关系；

3.客户端根据取得的映射信息访问相应的存储节点；

SAE的KVDB服务

新浪云计算 sinacloud.com

客户端API简单易用、功能丰富、支持前缀查找

```php
$kv = new SaeKV();
$ret = $kv->init();                          // 初始化SaeKV对象
$ret = $kv->add( 'abc', 'aaa' );             // 增加key-value
$ret = $kv->set( 'abc', 'bbb' );             // 更新key-value
$ret = $kv->replace( 'abc', 'ccc' );         // 替换key-value
$ret = $kv->get( 'abc' );                    // 获得key-value
$ret = $kv->delete( 'abc' );                 // 删除key-value
$keys = array( 'abc1', 'abc2' );
$ret = $kv->mget( $keys );                   // 获取多个key-values
$ret = $kv->pkrget( 'abc', 3 );              // 前缀范围查找

$file = 'saekv://config.php';
$content = "<?php\n\$site_name = 'Hello'; \n?>";
$ret = file_put_contents( $file, $content );
require_once( $file );
var_dump( $site_name );
```

sina新浪

SAE的KVDB服务

客户端与Meta Server

1. 客户端可以缓存从Meta Server获取的信息；
2. 会话超时机制以及不可用服务器标记功能；
3. 长连接支持，有效的减少到服务端的连接数量；
4. 多机房服务器列表，防止机房故障；

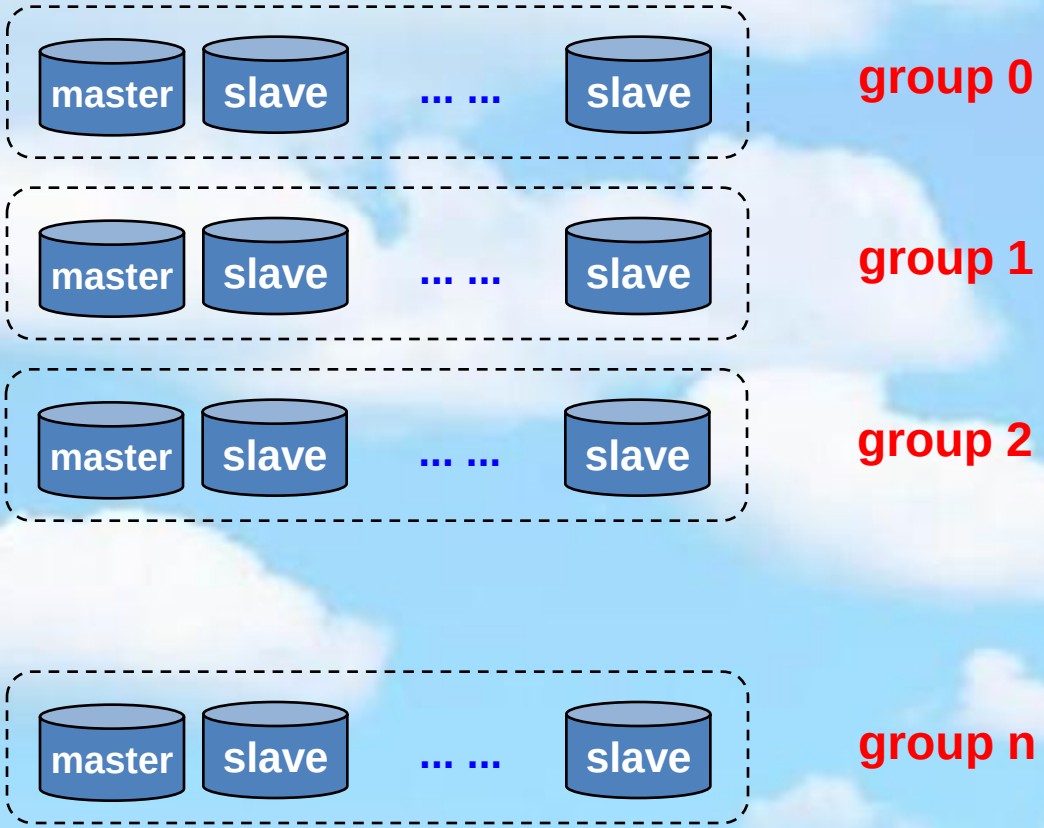
客户端与DB Server

客户端通过AppKey来作为自己的身份认证

SAE的KVDB服务

新浪云计算 sinacloud.com

DB Cluster 示意图



sina新浪

SAE的KVDB服务

DB Cluster 介绍

1. 分成多个组；
2. 每一组服务器一主多从，Master服务器由组内各服务器投票选举产生；
3. 通过复制，组内的每台服务器数据完全相同，实现读写分离和备份；
4. 通过增加服务器组来实现水平扩展；
5. 每一组服务器存储哪些数据由Meta Server决定；

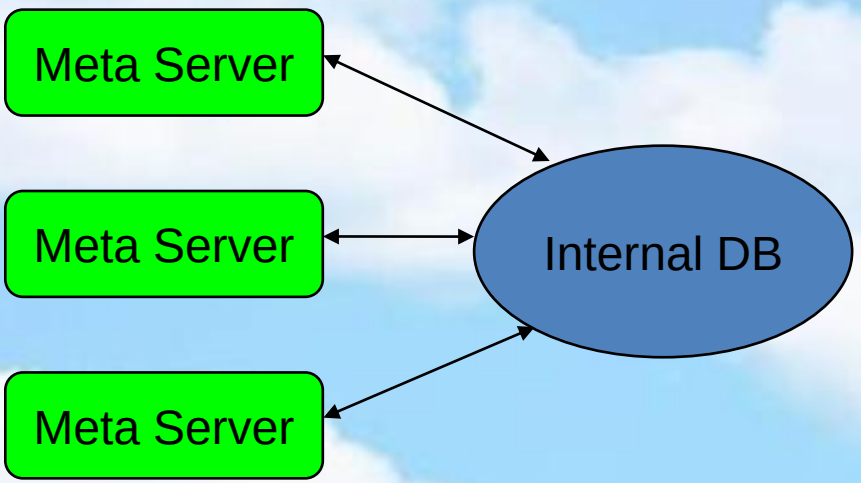
SAE的KVDB服务

关于单个DB节点

1. 每台服务器的服务层与存储层分离；
2. 服务层提供统计功能并定时汇报给统计中心；
3. 通过AppKey实现数据的隔离；

SAE的KVDB服务

Meta Server Cluster



- ✓多台Meta Server
- ✓定时获取Internal DB数据
- ✓缓存信息并提供查询服务
- ✓定时获取主从信息
- ✓定时获取DB节点信息
- ✓发起重平衡，迁移数据

SAE的KVDB服务

关于Meta Server的一些问题

I. 如果保证meta server的一致性？

类paxos算法

II. 如何触发重平衡？

维度： 容量和性能

方法： 数学期望和方差

III. 重平衡如何做到无缝？

双写单读

Key-Value服务的改进

- 服务层支持选举
- 服务层实现主从复制技术
- 更多种类的Key-Value服务

Thank You !

新浪 SAE 陈磊
@simpcl
simpcl2008@gmail.com