



全球软件案例研究  
峰会

# 100

TOP 100 CASE STUDIES OF THE YEAR

全球软件案例研究峰会



TOP 100 CASE STUDIES  
OF THE YEAR

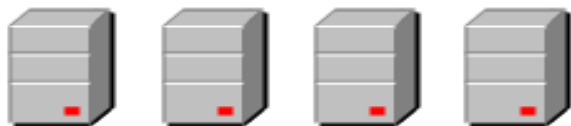
[www.top100summit.com](http://www.top100summit.com)

# 阿里线上压测与容量规划实践

- 线上压测与容量规划概念
- 线上压测方法与场景介绍
- 系统实现
- 线上压测与容量规划作用



——游骥



服务器



## 线上压测与容量规划概念

如何给系统分配合理数量的服务器，既能保障系统的正常运行，又能最大限度的节约成本？

如何知道系统什么时候需要扩容、缩容、扩容多少及缩容多少服务器？

如何最真实地验证系统的性能指标？

## ➤ 系统容量概念定义

单机能力：单台服务器每秒能承受的极限调用量（Qps）

集群能力：系统集群每秒能承受的极限调用量（Qps）

水位系数：系统运行在极限qps乘以该系数的情况下都认为服务状态是健康的

集群负荷：系统当前的调用量（Qps）



## ➤ 系统容量公式

集群水位 = 集群负荷 / 集群能力 \* 100%

理论机器数 = ( 实际机器数 \* 集群负荷 \* 集群水位 ) / ( 集群能力 \* 水位系数 )

机器增减 = 理论机器数 - 实际机器数

唯一的未知变量：**集群能力 <-- 压力测试**



# 线上压测方法与场景介绍

## ➤ 压测环境选择

### 线下

- 自身环境不真实
- 依赖服务的环境不真实
- 简单，可操作性好
- 不用数据和业务安全性

### 线上

- 自身环境真实
- 依赖服务的环境真实；
- 复杂性高，不易于操作
- 需考虑数据和业务安全性

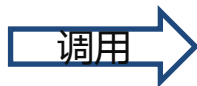
# 线上压测方法与场景介绍

## ➤线上压测方式

压测方式	主要特点
模拟请求	脏数据如何处理；测试流程如何不影响正常流程；如何构造贴近真实的用户场景，否则难保障压测准确性；请求如何构造；可压测单机和集群；
复制请求	写请求如何处理；响应如何处理；可压测单机和集群；
引流	完全真实的请求；无法压测出集群能力，非常适合用来做单机压力测试；
负载均衡权重调整	完全真实的请求；无法压测出集群能力，非常适合用来做单机压力测试；



模拟调用者

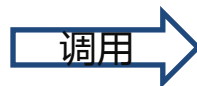


单机或者  
系统集群

1



机器A



机器B



机器C

3

## 线上压测方法与场景介绍

线上正常服务机器

复制真实请求



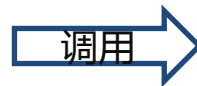
单机或者  
系统集群

2

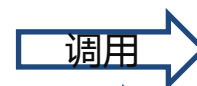
调用者



负载均衡



机器A



机器B



机器C

4



## 线上压测方法与场景介绍

### ➤模拟请求

相关工具：http\_load, webbench, apache ab, jmeter, Siege ,  
Seaplane , 自定义工具；

运用场景及注意事项：

- 1、系统是只读的或者不需要压测写请求；
- 2、已做好对写请求脏数据的特殊处理（拦截或者隔离存储）；

---

### 3、新系统上线；



## 线上压测方法与场景介绍

### ➤复制请求

相关工具：tcpcopy, btrace, nginx post\_action , 自定义 agent等；

运用场景及注意事项：

- 1、系统是只读的或者不需要压测写请求；
- 2、对写请求不存在唯一性约束，允许同一个请求多次写；
- 3、请求的响应已被特殊处理（不能返回给调用方）；



## 线上压测方法与场景介绍

### ➤请求引流

相关工具： apache mod\_jk , mod\_proxy, nginx proxy等；

运用场景及注意事项：

- 1、单机压测；
- 2、系统的集群调用量足以压测到单机的极限；



## 线上压测方法与场景介绍

### ➤ 负载均衡权重调整

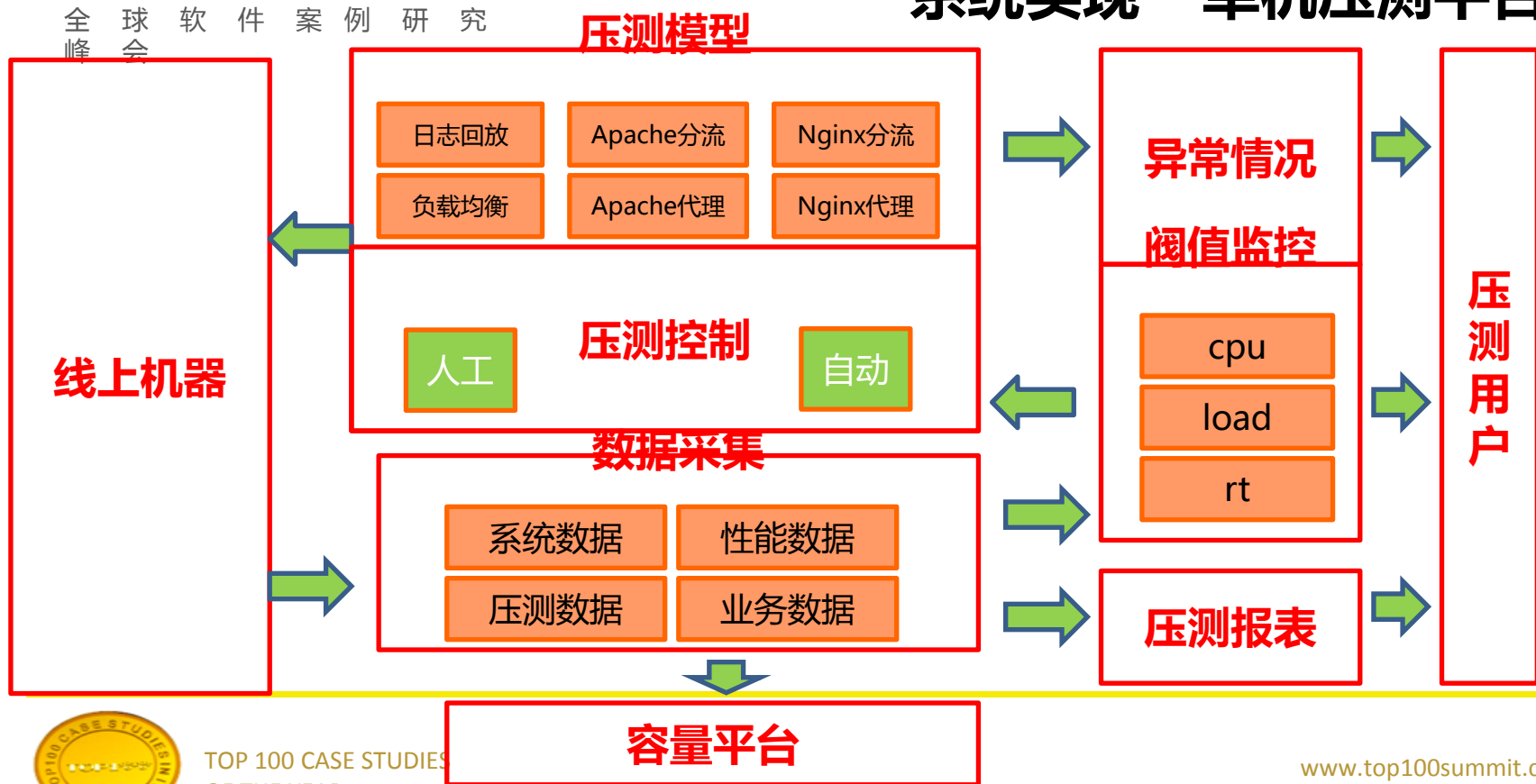
相关工具：F5, LVS, SOA service registration等负载均衡控制器；

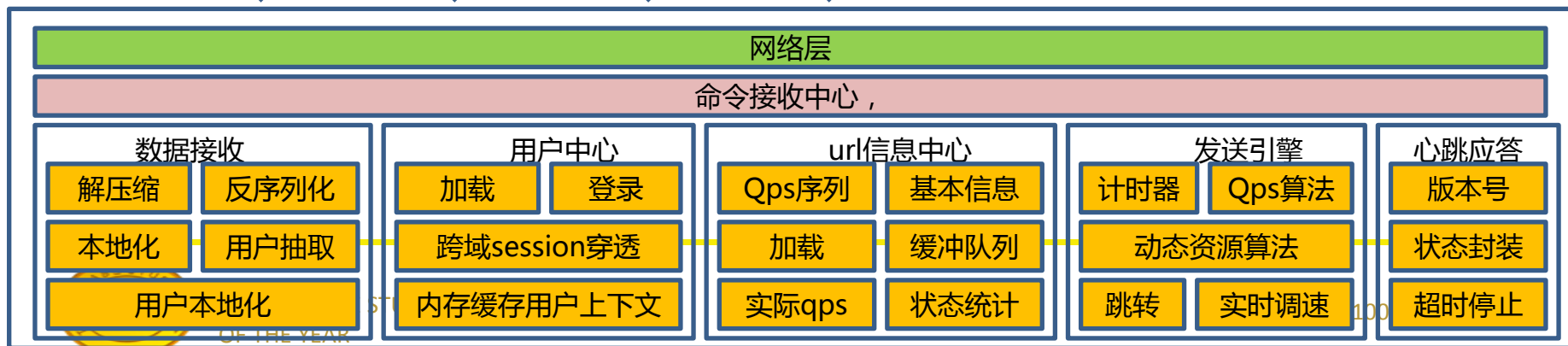
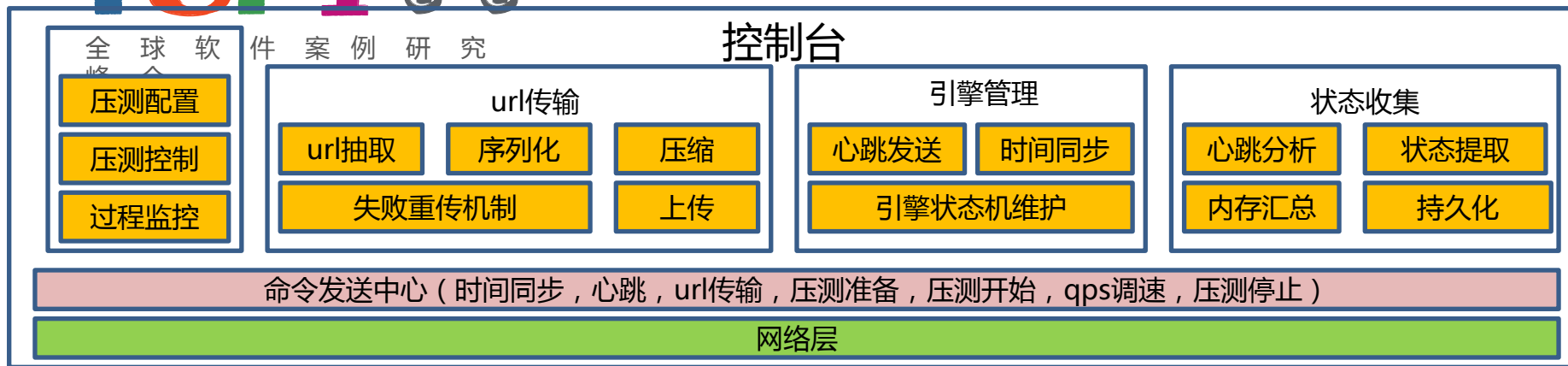
运用场景及注意事项：

- 1、单机压测；
- 2、系统的集群调用量足以压测到单机的极限；



# 系统实现—单机压测平台





## 线上压测与容量规划作用

- 1、阿里系统扩容或者缩容的数据依据；
- 2、阿里大促时期资源评估数据参考；
- 3、帮助系统发现大量性能瓶颈与容量问题；
- 4、阿里稳定性保障与验证利器（有了它，让双11来得更猛烈些吧！）；

