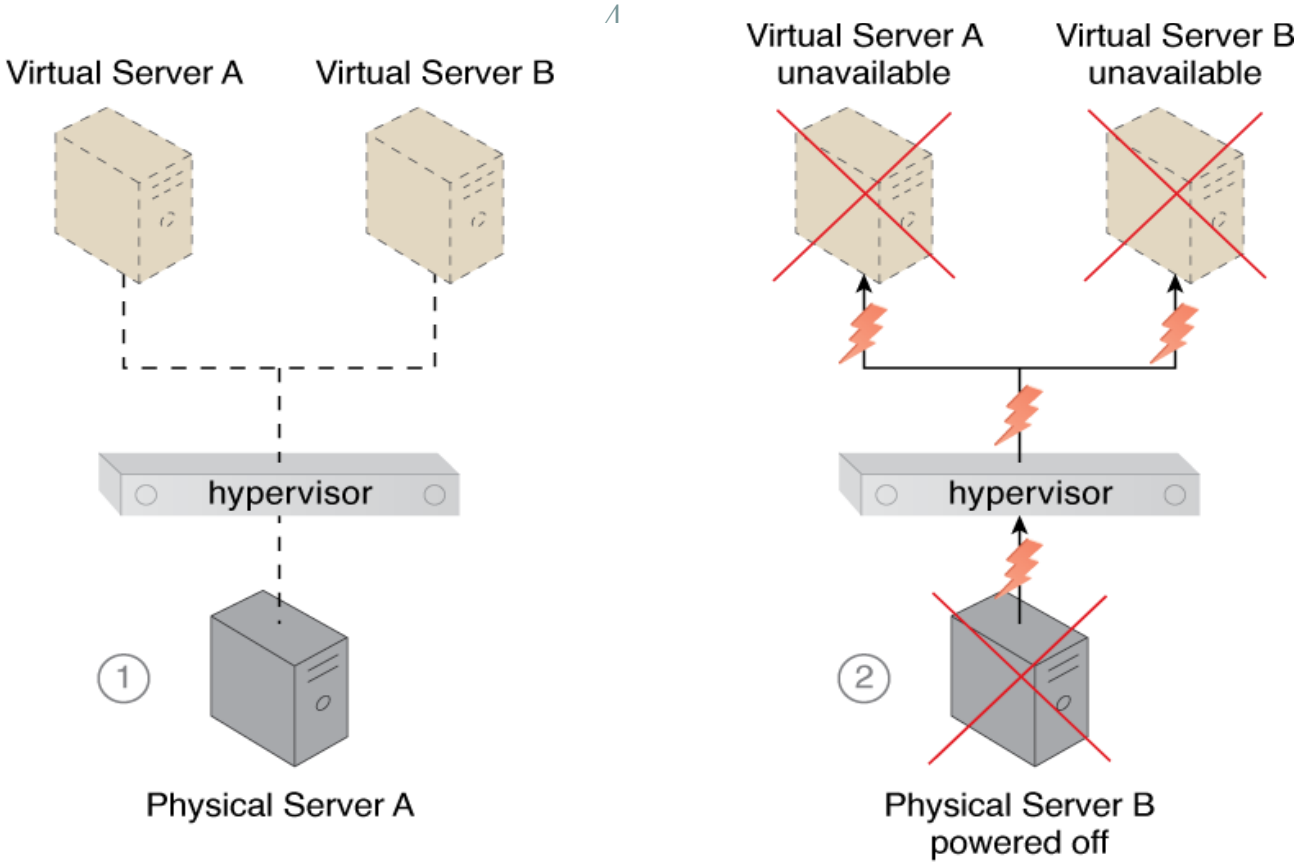
# 第12章 高级云架构

§12.1 虚拟机监控器集群架构
§12.2 粗粒度负载均衡与资源预留
§12.3 云服务容错架构
§12.4 裸机供给与快速供给架构
§12.5 存储负载管理架构

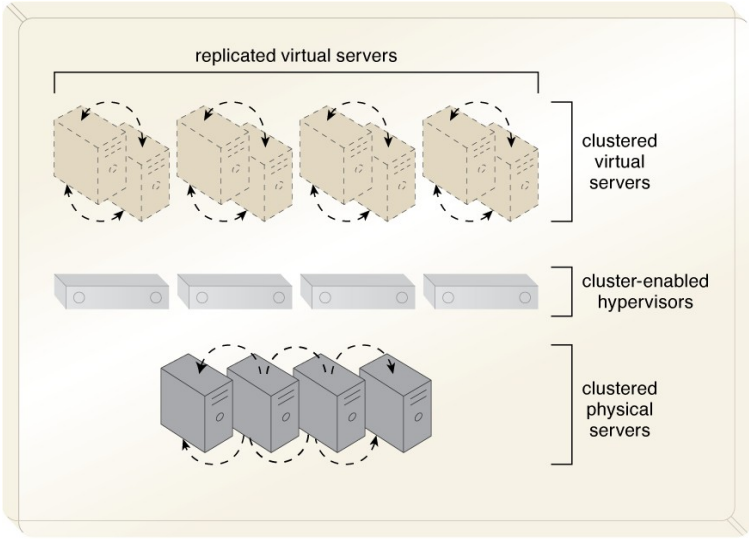
# §12.1虚拟机监控器集群




- *Physical Server **A** is hosting a hypervisor that hosts Virtual Servers **A** and **B** (1).*
- *When Physical Server **A** fails, the hypervisor and two virtual servers consequently fail as well (2).*

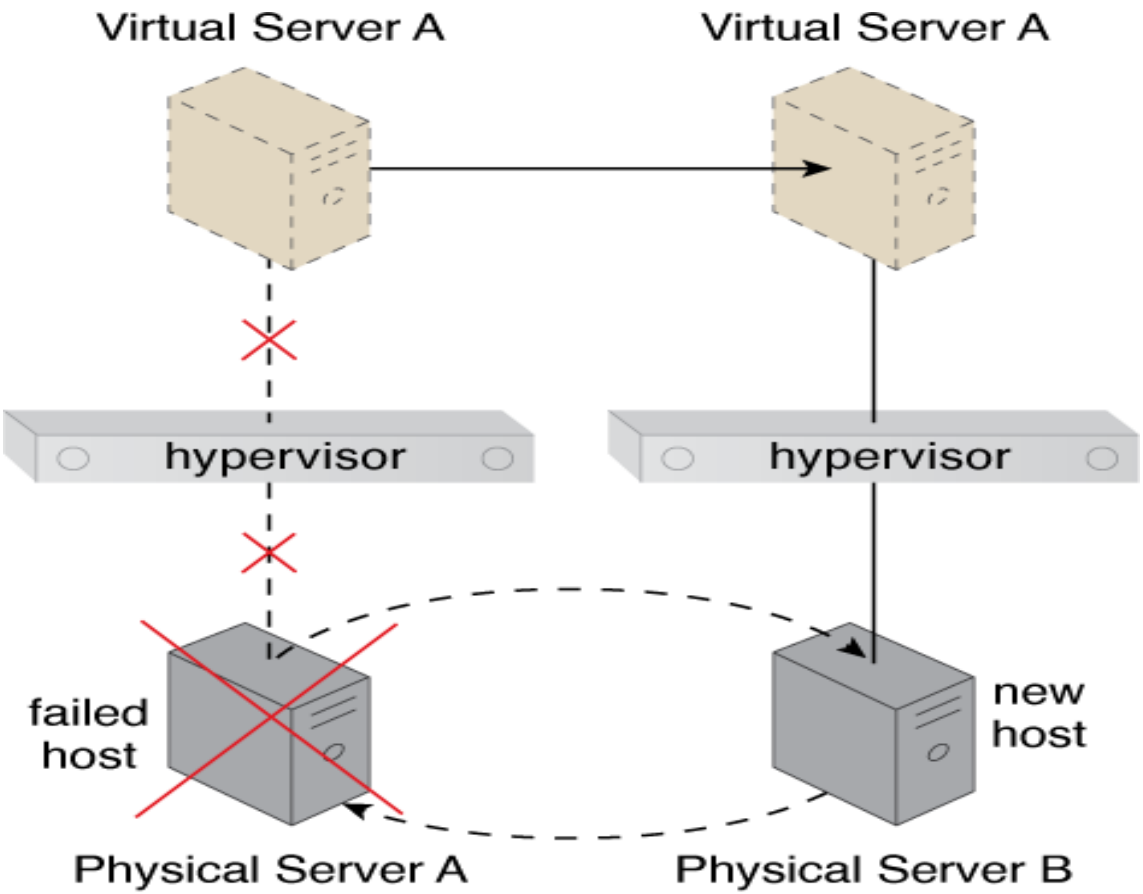
# 虚拟机监控器集群

- Hypervisor Clustering
  - 多个虚拟机监控器构成集群
  - 跨越多个物理服务器
  - 实现高可用
- 虚拟机监控器集群由中心VIM控制
  - 通过常规心跳消息来监测虚拟机监控器的状态
  - 当物理机或者Hypervisor失效时进行在线迁移
  - 使用共享云存储设备实现在线迁移

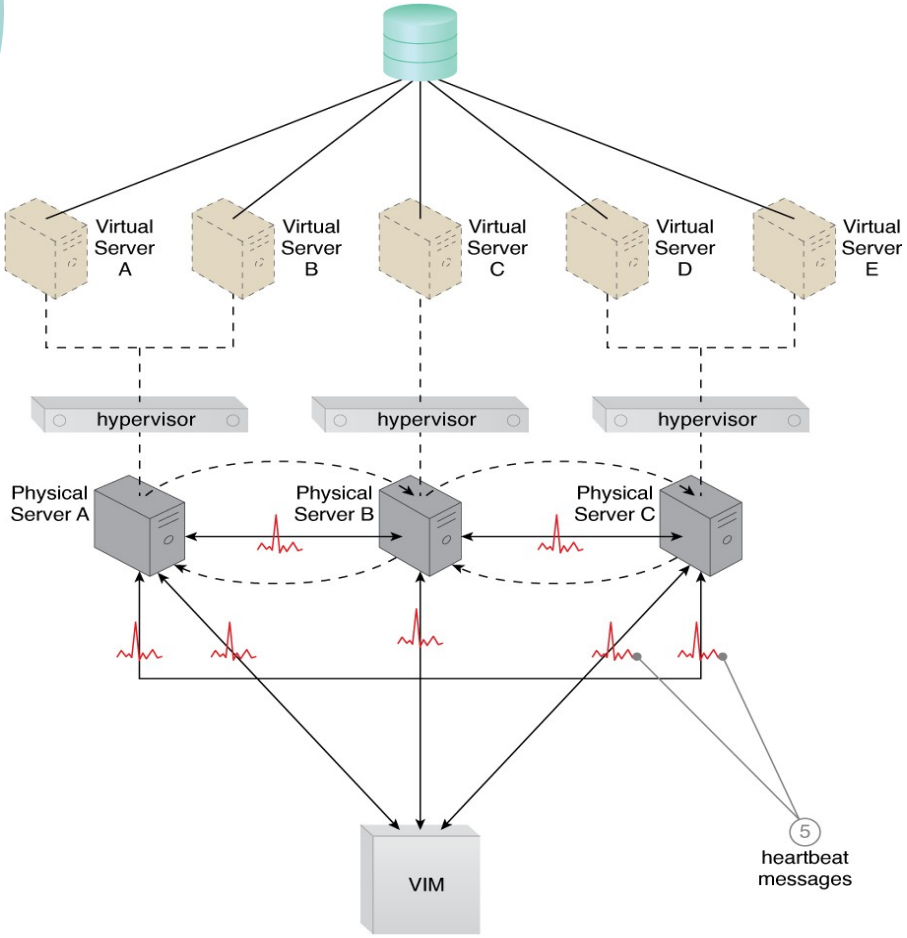



HA virtualization cluster

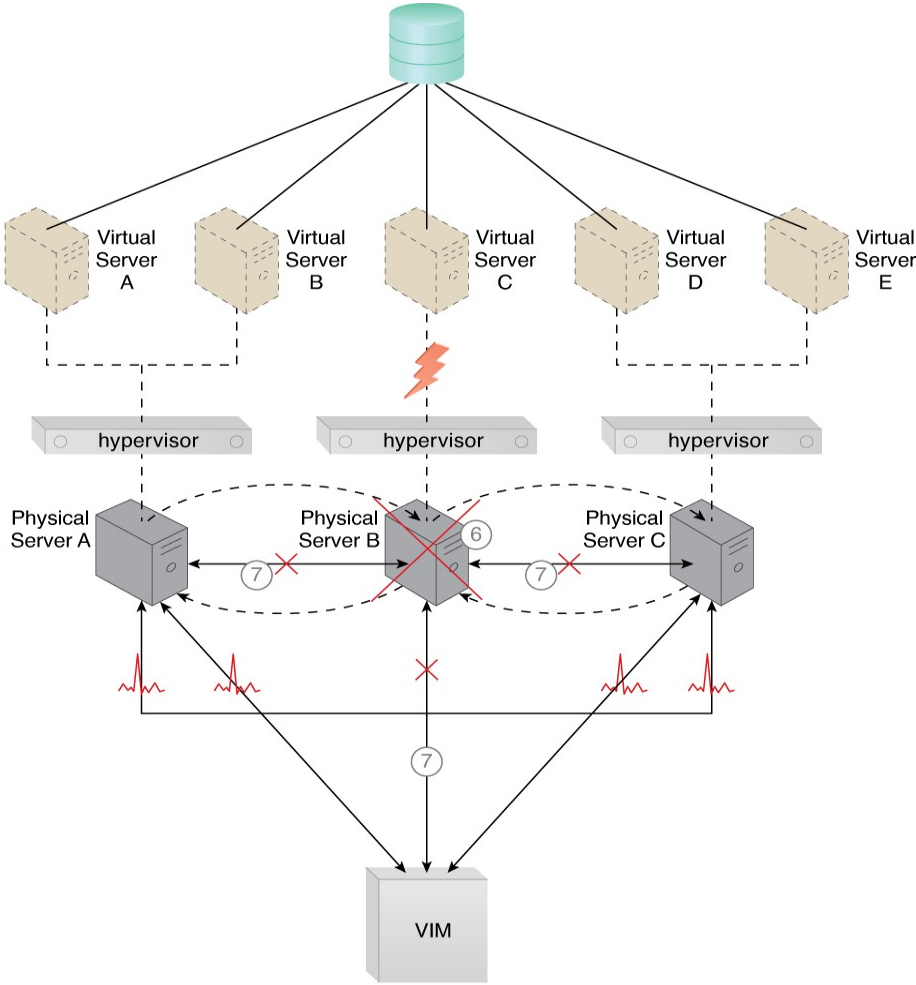
# Hypervisor/Virtual Server在线迁移



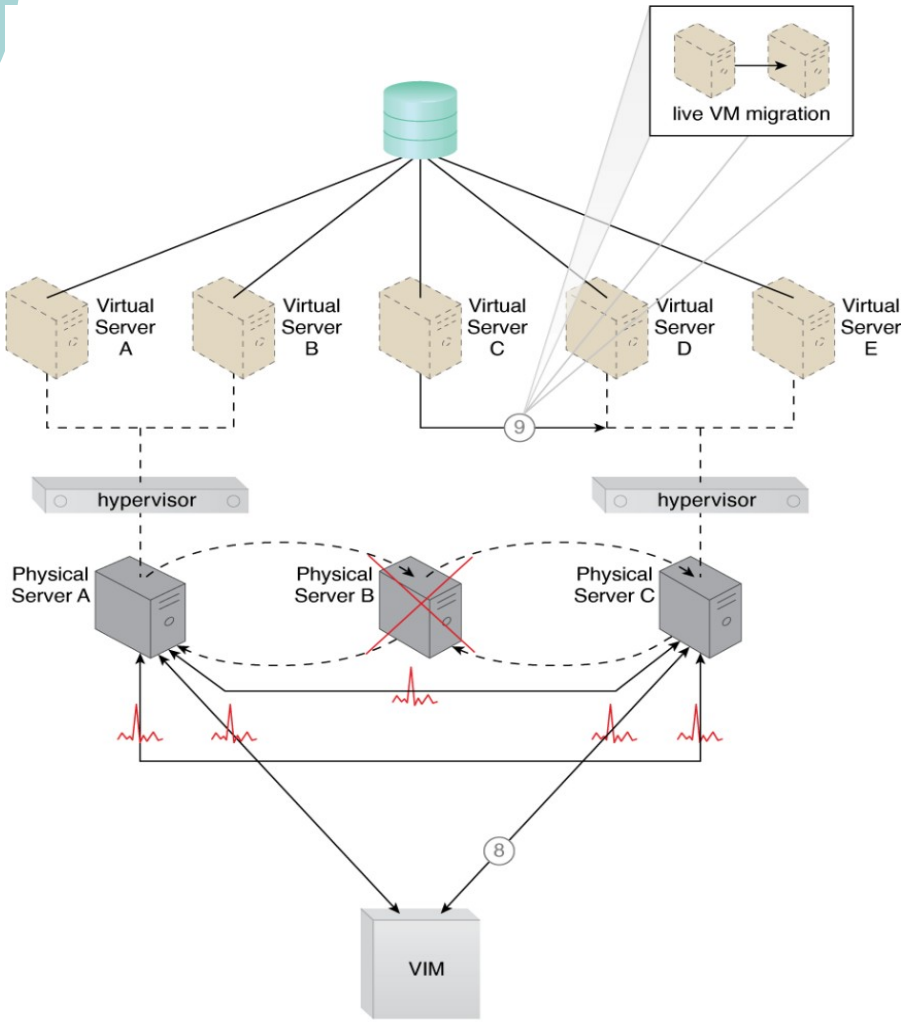
# Hypervisor/Virtual Server在线迁移





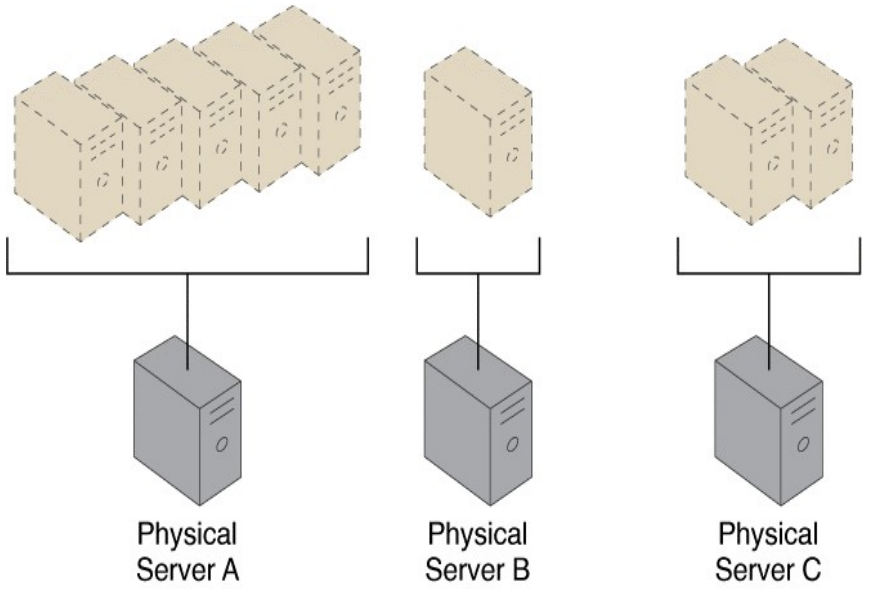
# Hypervisor/Virtual Server 在线迁移



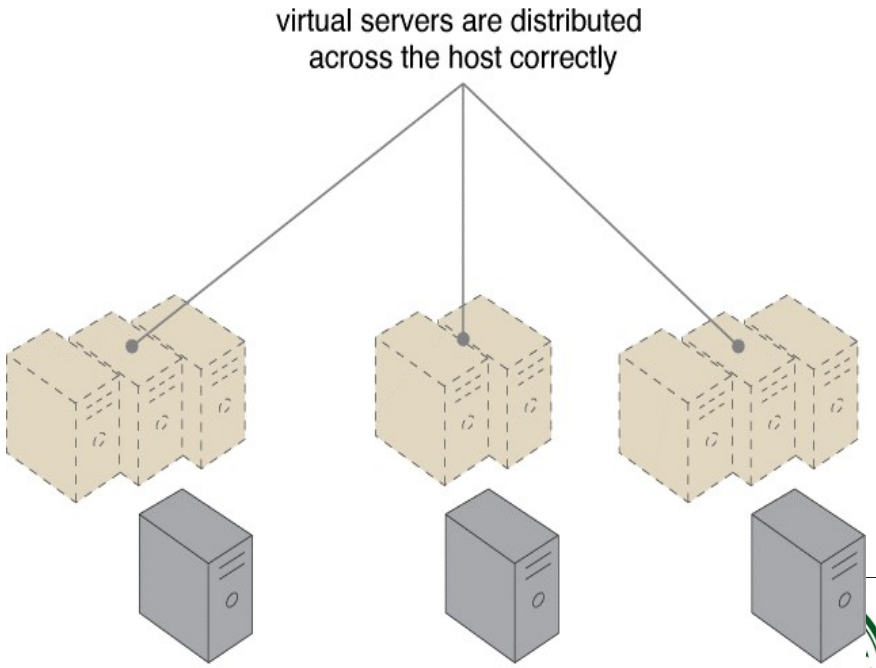



- The VIM chooses Physical Server 3 as the new host after assessing the available capacity.
- Virtual Server C is live-migrated to the hypervisor running on Physical Server 3
- Restarting may be necessary

# 虚拟服务器实例负载均衡





# §12.2 粗粒度负载均衡与资源预留

- 负载均衡的粒度有很多
- 用户请求
  - 前端负载均衡器把用户请求分发给不同的云服务实例
- 虚拟服务器实例
  - 把虚拟服务器放置到不同的物理机
- 云负载均衡
  - 数据中心间的负载均衡

# 虚拟服务器实例负载均衡

- 物理服务器之间的负载均衡是很难的
  - 物理的隔离性
- 负载均衡的虚拟机实例架构
  - 合理分布虚拟机实例，以均衡物理服务器的负载
  - 基于虚拟机监控器集群
  - 核心是容量看门狗（capacity watchdog）系统
    - 处理任务分配到物理服务器之前动态地计算虚拟服务器实例及其相关的工作负载

# 容量看门狗系统

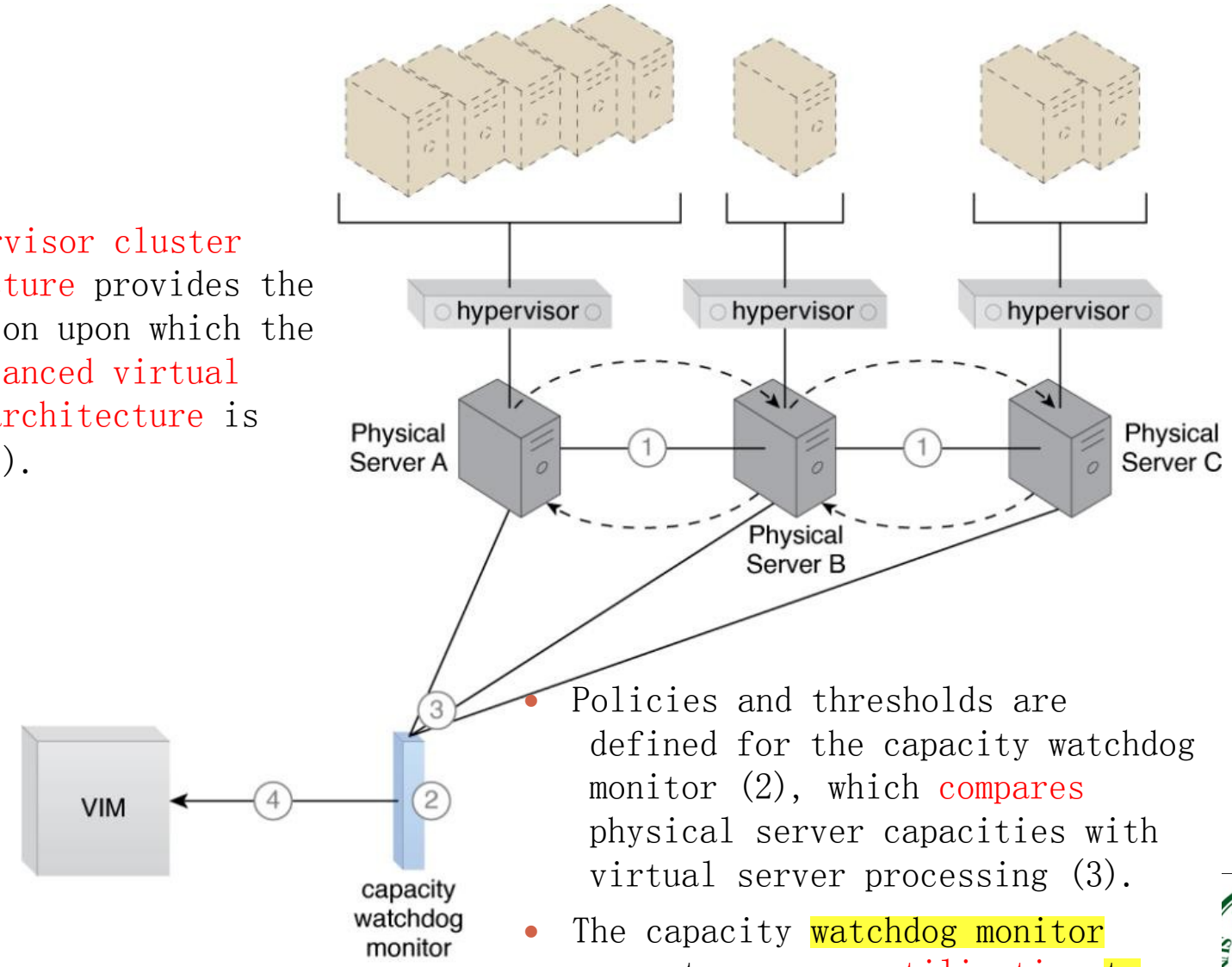
- 主要包括三部分
- 一个容量看门狗云使用监控器
  - 追踪物理和虚拟服务器使用情况
  - 向容量计划器报告波动
- 一个容量计划器
  - 动态计算物理服务器的可用能力和虚拟服务器的容量要求
- VM在线迁移程序

# 虚拟服务器实例负载均衡

- 需要的其他支撑机制
  - 自动伸缩监听器
  - 负载均衡器
  - 逻辑网络边界
  - 资源复制

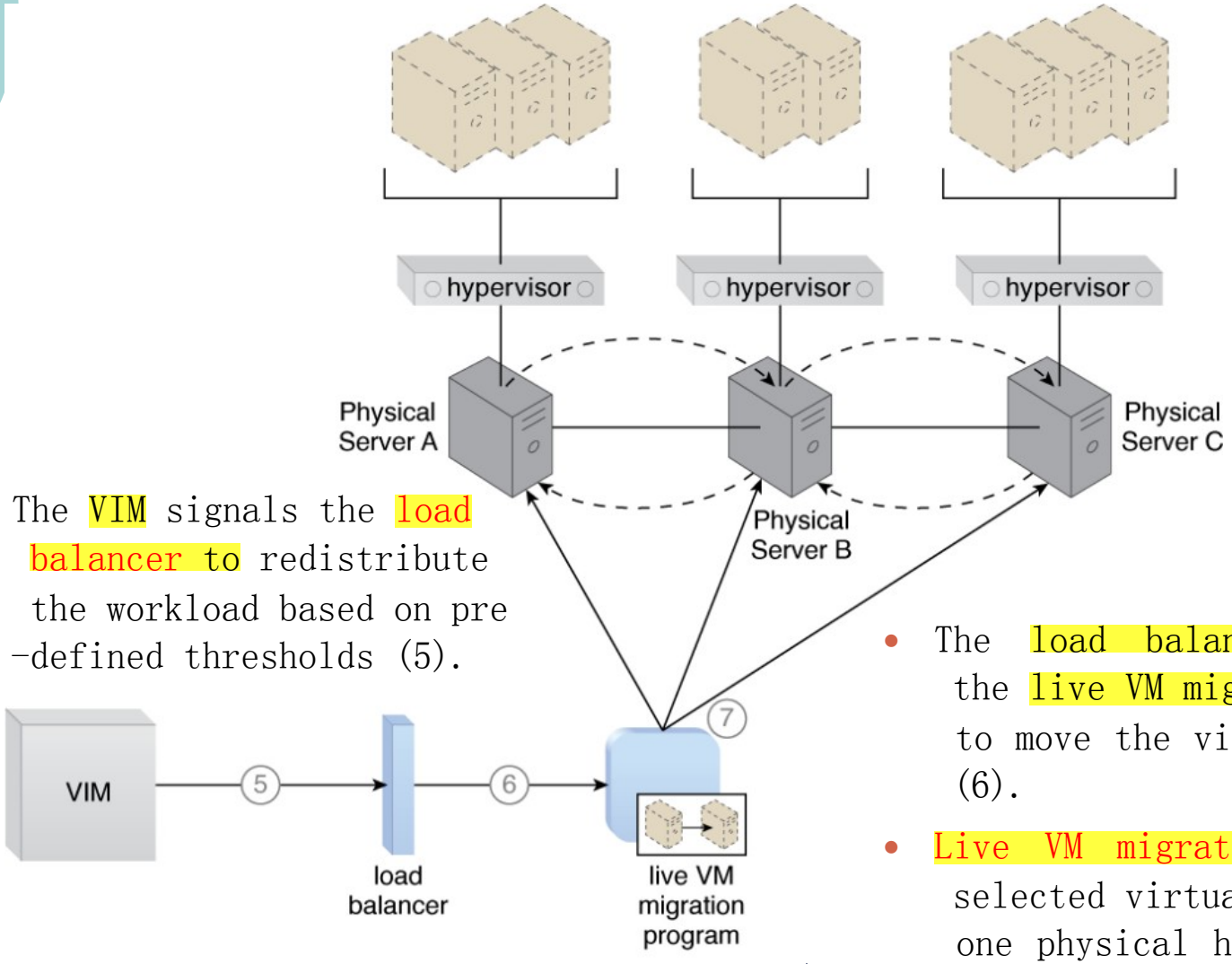
# 虚拟服务器实例均衡

- The hypervisor cluster architecture provides the foundation upon which the load-balanced virtual server architecture is built (1).




- Policies and thresholds are defined for the capacity watchdog monitor (2), which compares physical server capacities with virtual server processing (3).
- The capacity watchdog monitor reports an over-utilization to the VIM (4).

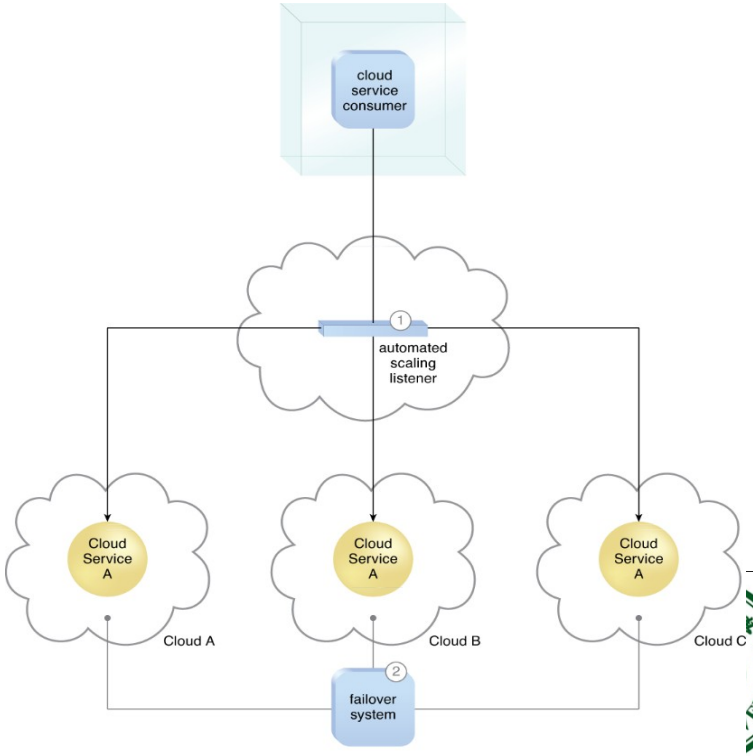
# 虚拟服务器实例均衡




- The VIM signals the load balancer to redistribute the workload based on pre-defined thresholds (5).
- The load balancer initiates the live VM migration program to move the virtual servers (6).
- Live VM migration moves the selected virtual servers from one physical host to another (7).

# 云负载均衡架构

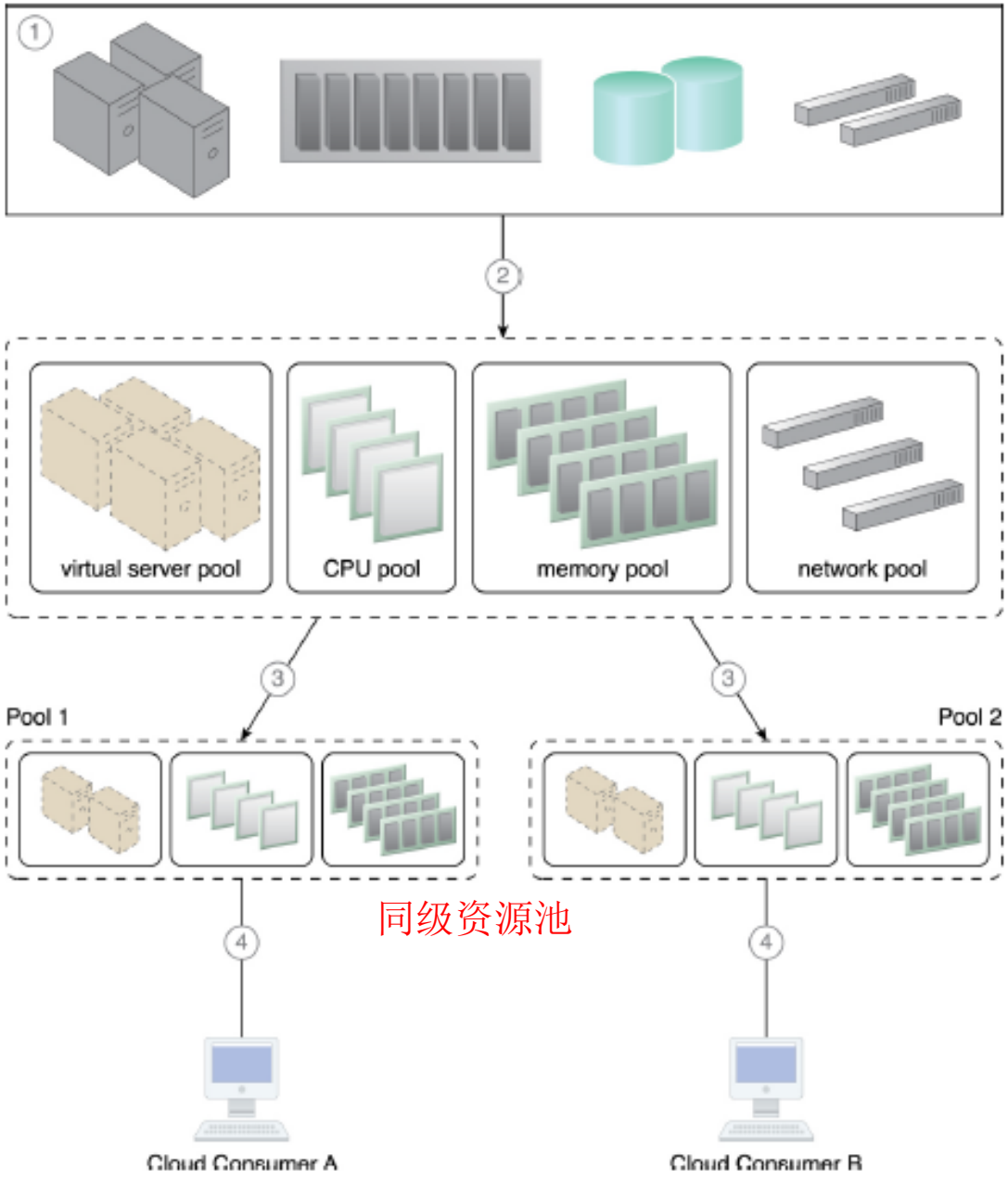
- Cloud balancing architecture
  - 一个特殊的架构模型
  - 在多个云之间进行负载均衡
- 目的
  - 提高服务能力、可扩展性
  - 提高可用性、可靠性
  - 改进负载均衡和资源优化
- 基础
  - 自动伸缩监听器
  - 故障转移系统



# 资源预留

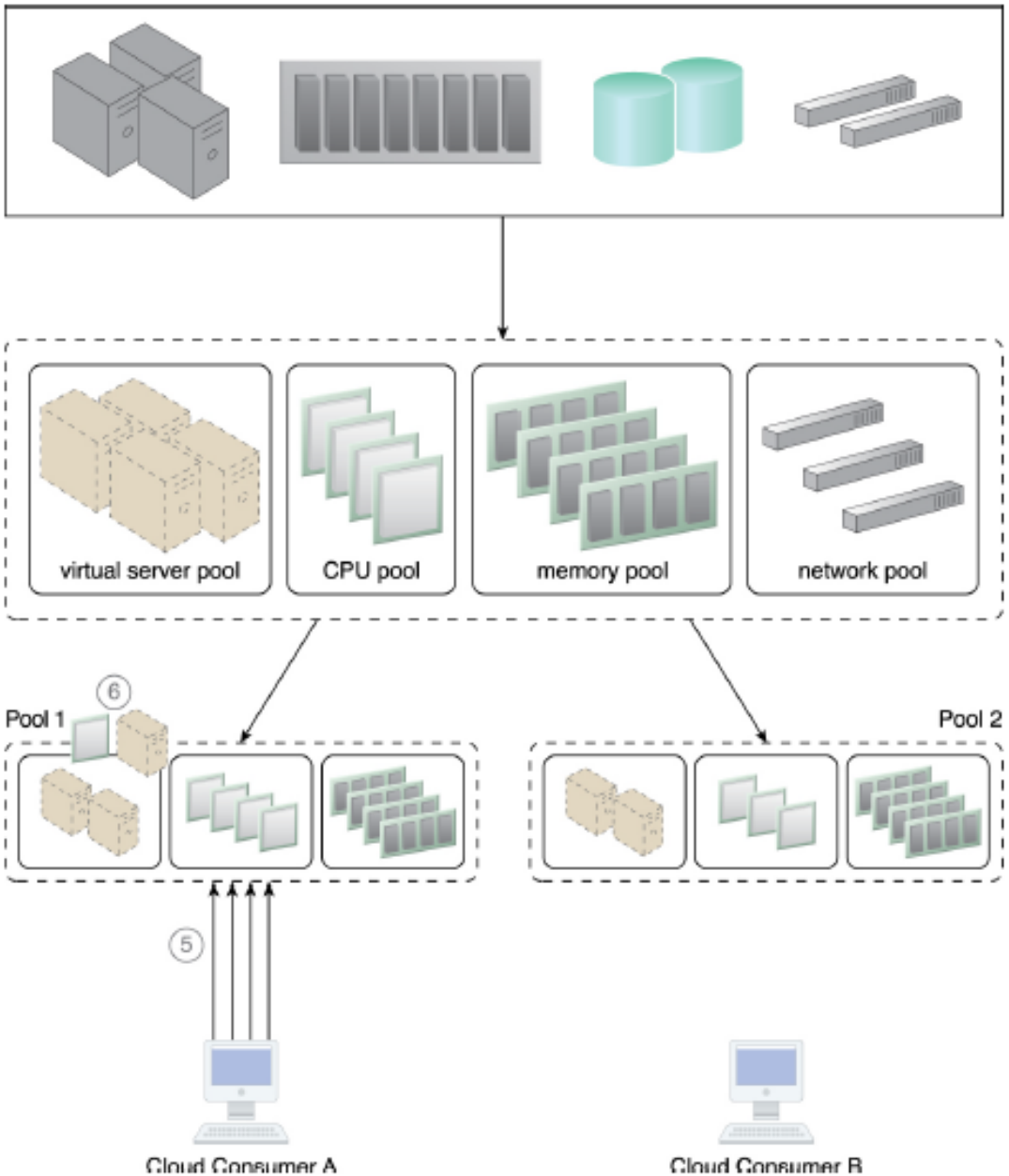
- 资源受限(resource constraint)
  - 太多并发访问可能会导致运行时异常
  - IT资源没有足够的容量
- 资源预留(resource reservation)
  - 专门为给定的云用户保留
    - 单个IT资源
    - 一个IT资源的一部分
    - 多个IT资源

# 资源预留示例---1



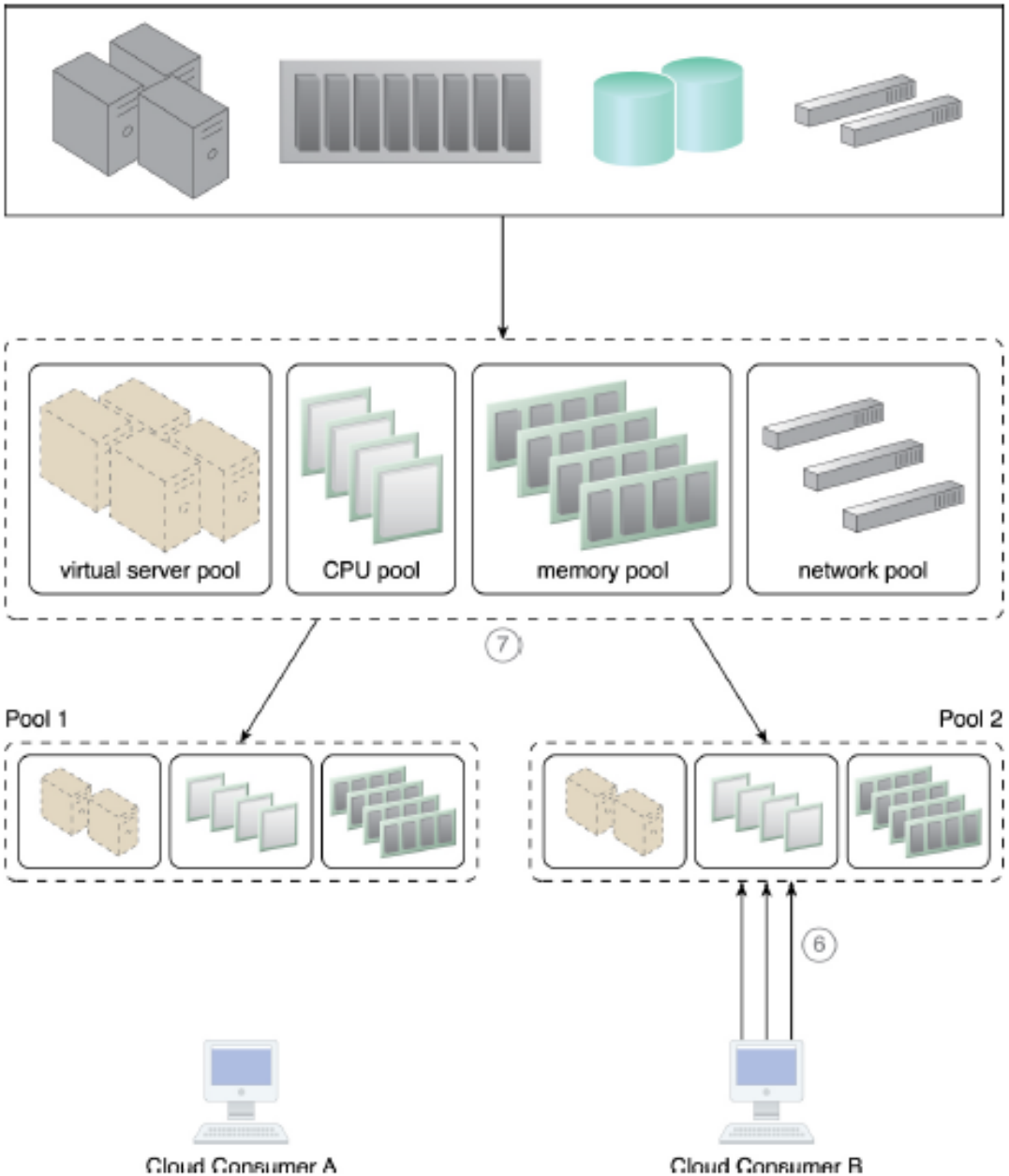

A physical resource group is created (1), from which a parent resource pool is created as per the resource pooling architecture (2). Two smaller child pools are created from the parent resource pool, and resource limits are defined using the resource management system (3). Cloud consumers are provided with access to their own exclusive resource pools (4).

# 资源预留示例---2




An increase in requests from Cloud Consumer A results in more IT resources being allocated to that cloud consumer (5), meaning some IT resources need to be borrowed from Pool 2. The amount of borrowed IT resources is confined by the resource limit that was defined in Step 3, to ensure that Cloud Consumer B will not face any resource constraints (6).

# 资源预留示例---3




Cloud Consumer B now imposes more requests and usage demands and may soon need to utilize all available IT resources in the pool (6). The resource management system forces Pool 1 to release the IT resources and move them back to Pool 2 to become available for Cloud Consumer B (7).

# 资源预留架构

- 依赖多个部件和机制
  - 审计监控器
  - 云使用监控器
  - 虚拟机监控器
  - 逻辑网络边界
  - 资源复制

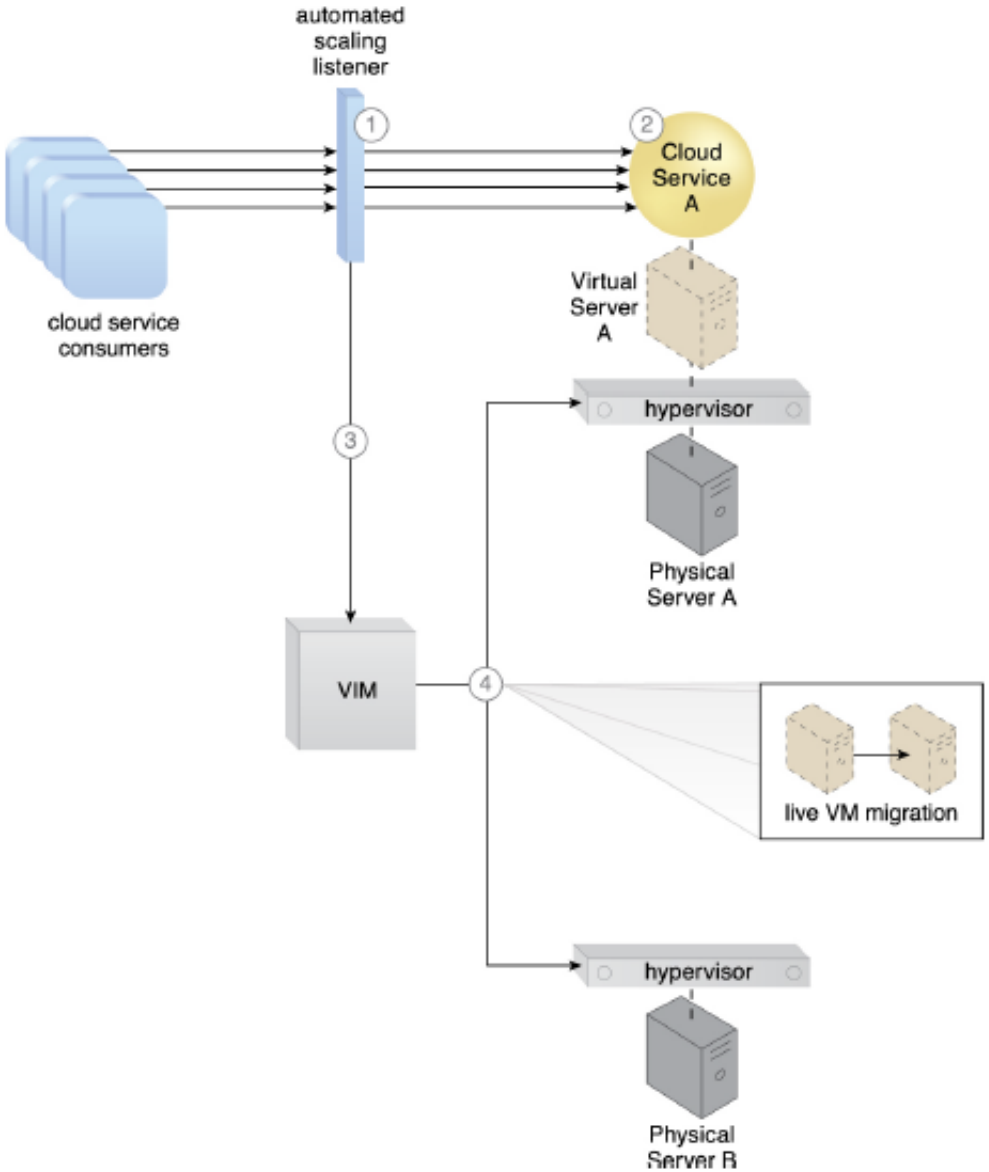
# §12.3 云服务容错架构

- 云服务不可用的原因很多：
  - 运行时需求超出处理能力
  - 维护更新导致的暂时中断
  - 云服务迁移
  - 物理机失效、宕机
- 应对机制
  - 服务迁移
  - 服务故障检测与恢复
  - 物理机容错

# 12.3.1服务迁移

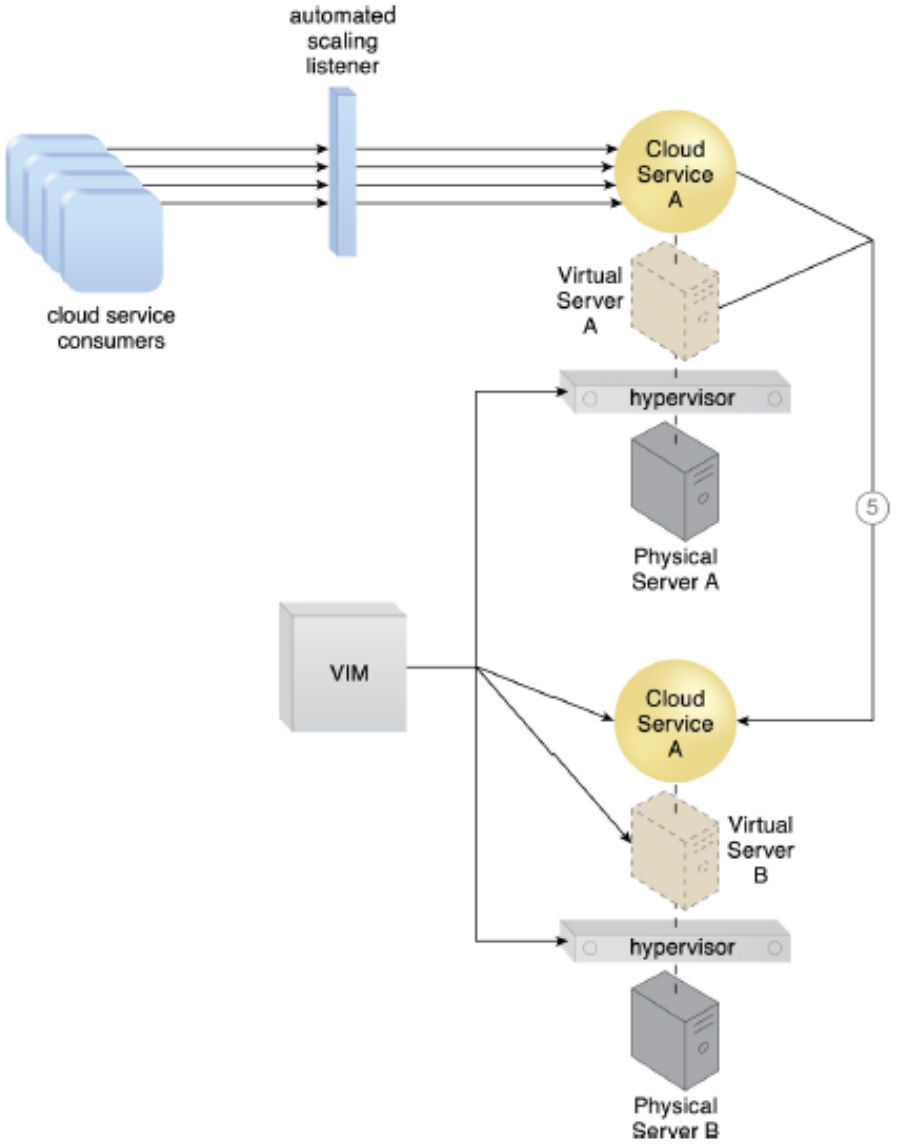
- 不中断服务重定位架构
- Non-disruptive service relocation architecture
  - 预先定义事件，触发云服务实现运行时复制或迁移

# 不中断服务重定位架构---1



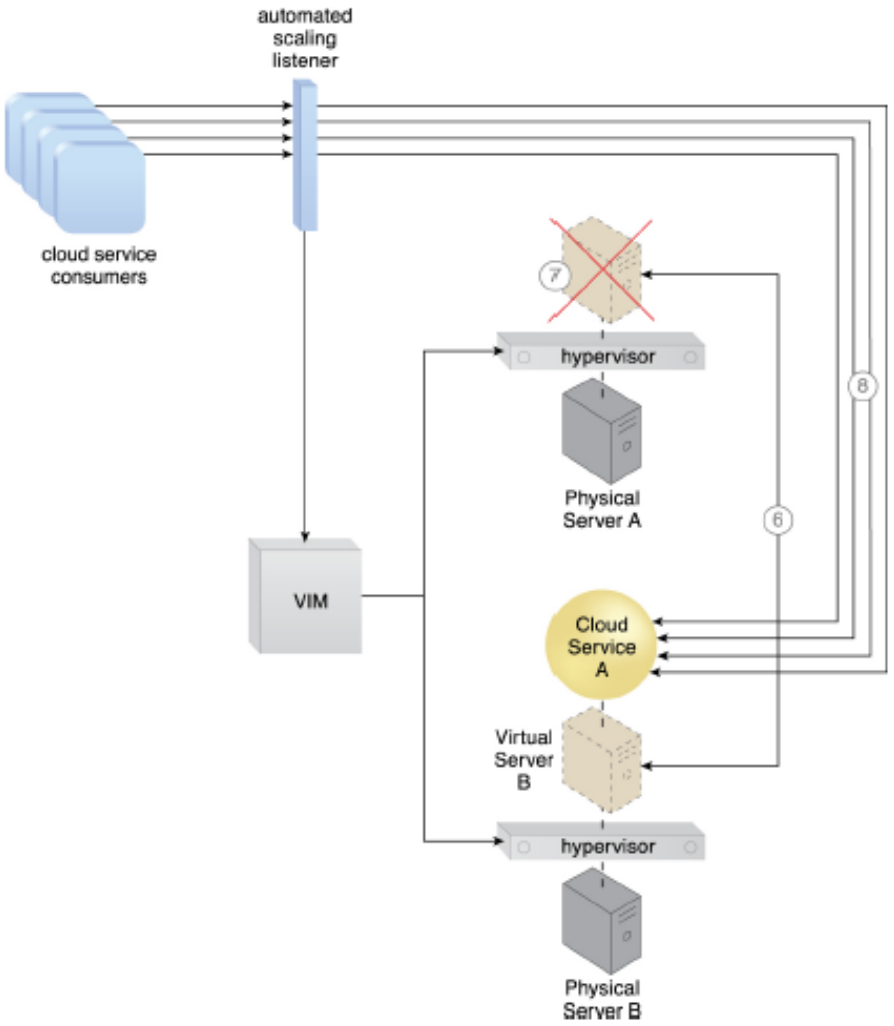

The automated scaling listener monitors the workload for a cloud service (1). The cloud service's predefined threshold is reached as the workload increases (2), causing the automated scaling listener to signal the VIM to initiate relocation (3). The VIM uses the live VM migration program to instruct both the origin and destination hypervisors to carry out runtime relocation (4).

# 不中断服务重定位架构---2




A second copy of the virtual server and its hosted cloud service are created via the destination hypervisor on Physical Server B (5).

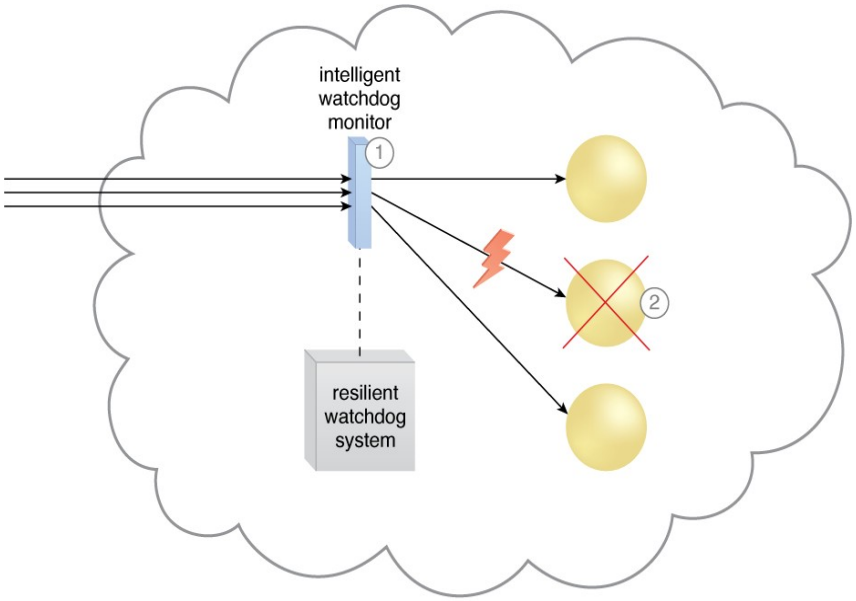
# 不中断服务重定位架构---3



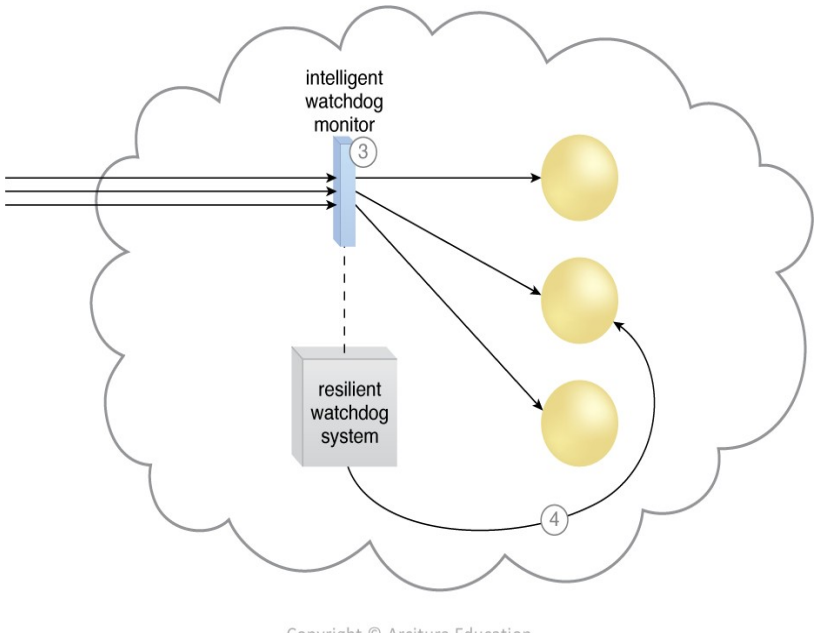

The state of both virtual server instances is synchronized (6). The first virtual server instance is removed from Physical Server A after cloud service consumer requests are confirmed to be successfully exchanged with the cloud service on Physical Server B (7). Cloud service consumer requests are now only sent to the cloud service on Physical Server B (8).

# 12.3.2动态故障检测与恢复

- Dynamic failure detection and recovery
  - 根据预先定义的故障场景
  - 通常基于弹性（Resilient）看门狗系统



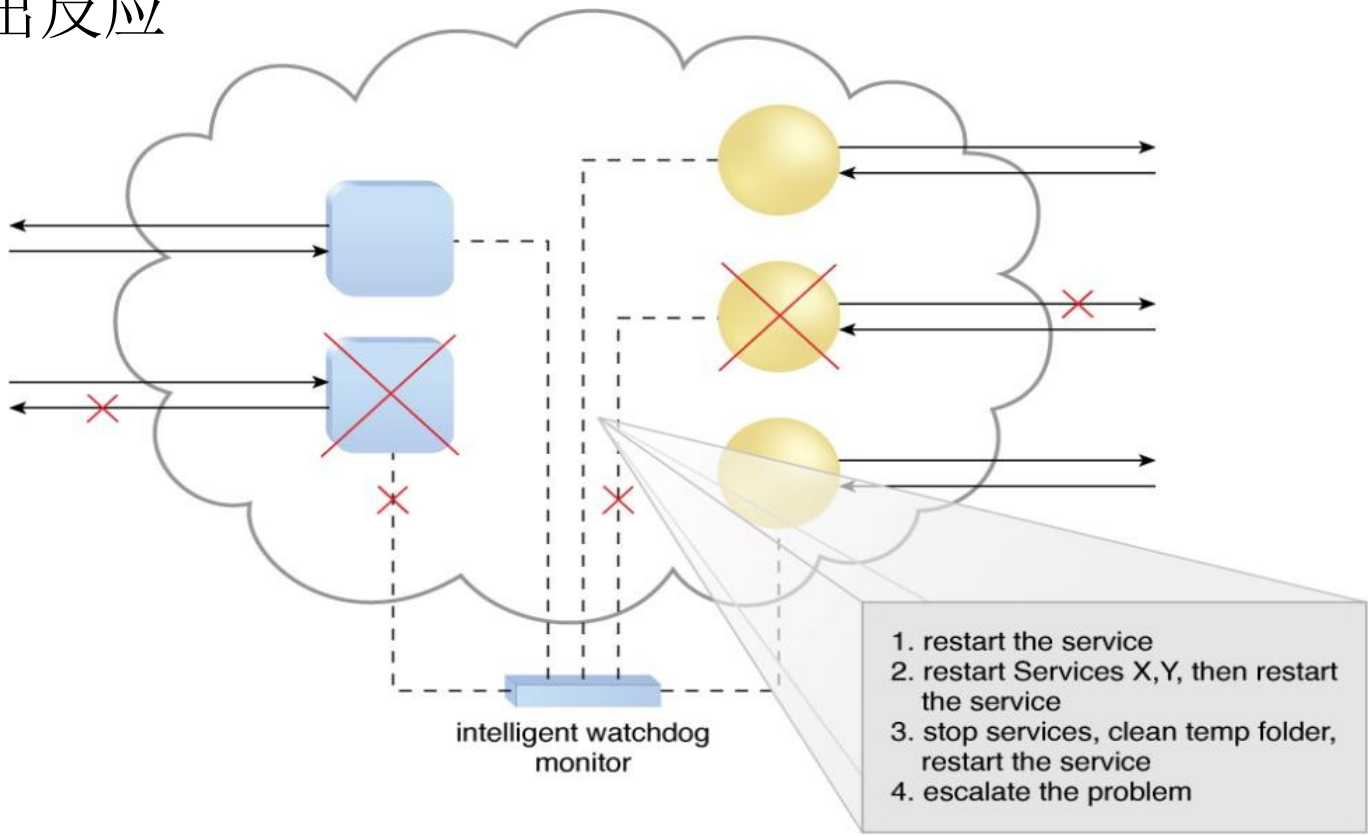

The intelligent watchdog monitor keeps track of cloud consumer requests (1) and detects that a cloud service has failed (2).




The intelligent watchdog monitor notifies the watchdog system (3), which restores the cloud service based on pre-defined policies. The cloud service resumes its runtime operation (4).

# 弹性看门狗系统

- 五个核心功能：
  - 监视
  - 选定事件
  - 对事件作出反应
  - 报告
  - 升级处理

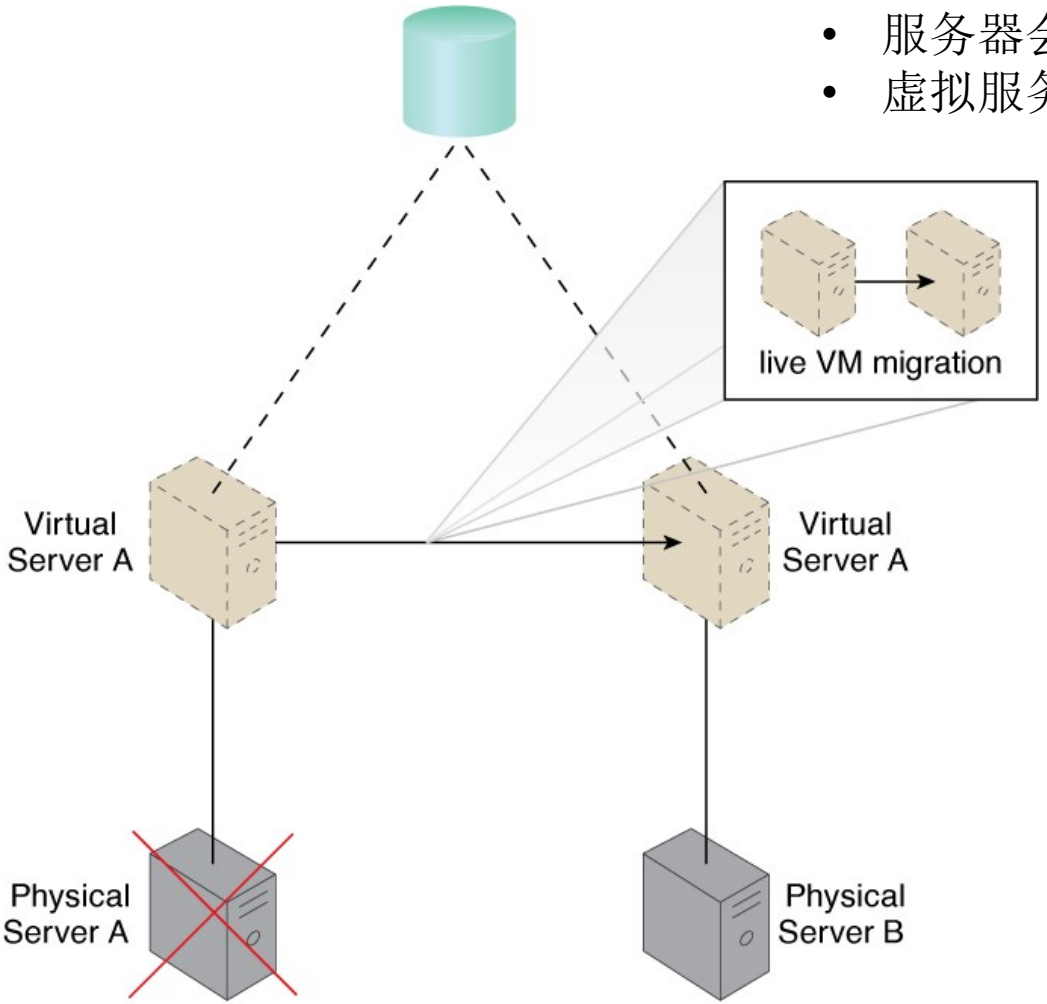


# 动态故障检测与恢复架构

- 智能看门狗监控器升级处理一个问题的最常采用的措施包括：
  - 运行一个批处理文件
  - 发送一个控制台消息
  - 发送一条短信消息
  - 发送一份电子邮件消息
  - 发送一个SNMP陷阱
  - 记录一个通知单

# 12.3.3零宕机架构

- 物理机不中断机制
- 服务器会聚成一组，由容错系统控制
- 虚拟服务器都存储在共享介质



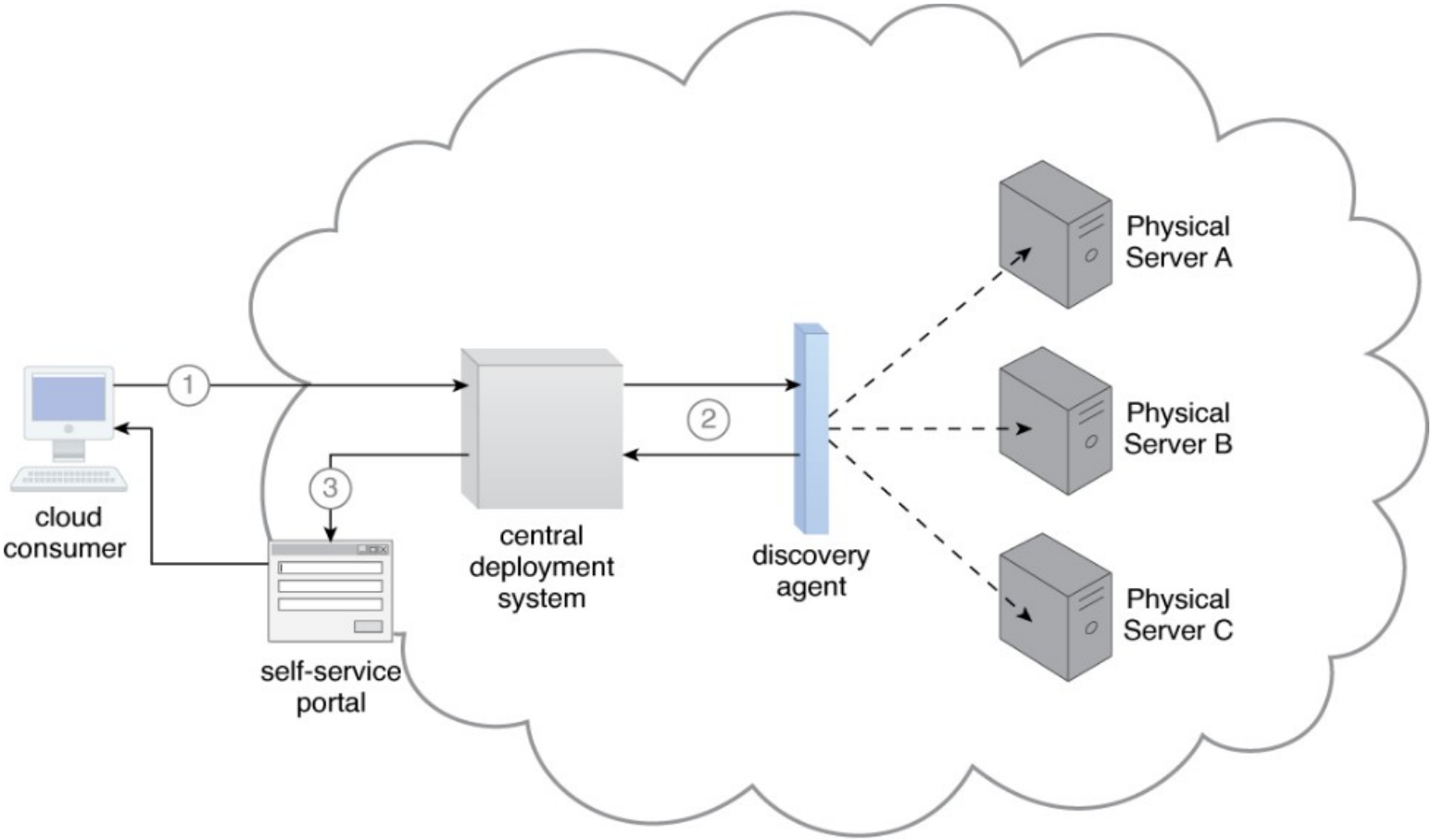
# §12.4 裸机供给与快速供给架构

- 裸机(Bare-metal servers)
  - 指没有预装操作系统或其他任何软件的物理服务器
- 裸机供给架构(Bare-metal provisioning architecture )
  - 弹性增加裸机
  - 基于ROM提供的远程安装支持自动安装系统
    - 连接默认或DHCP配置IP
    - 通过Web或专有接口连接到物理机
    - 自动安装操作系统和软件

# 裸机供给架构的主要构成

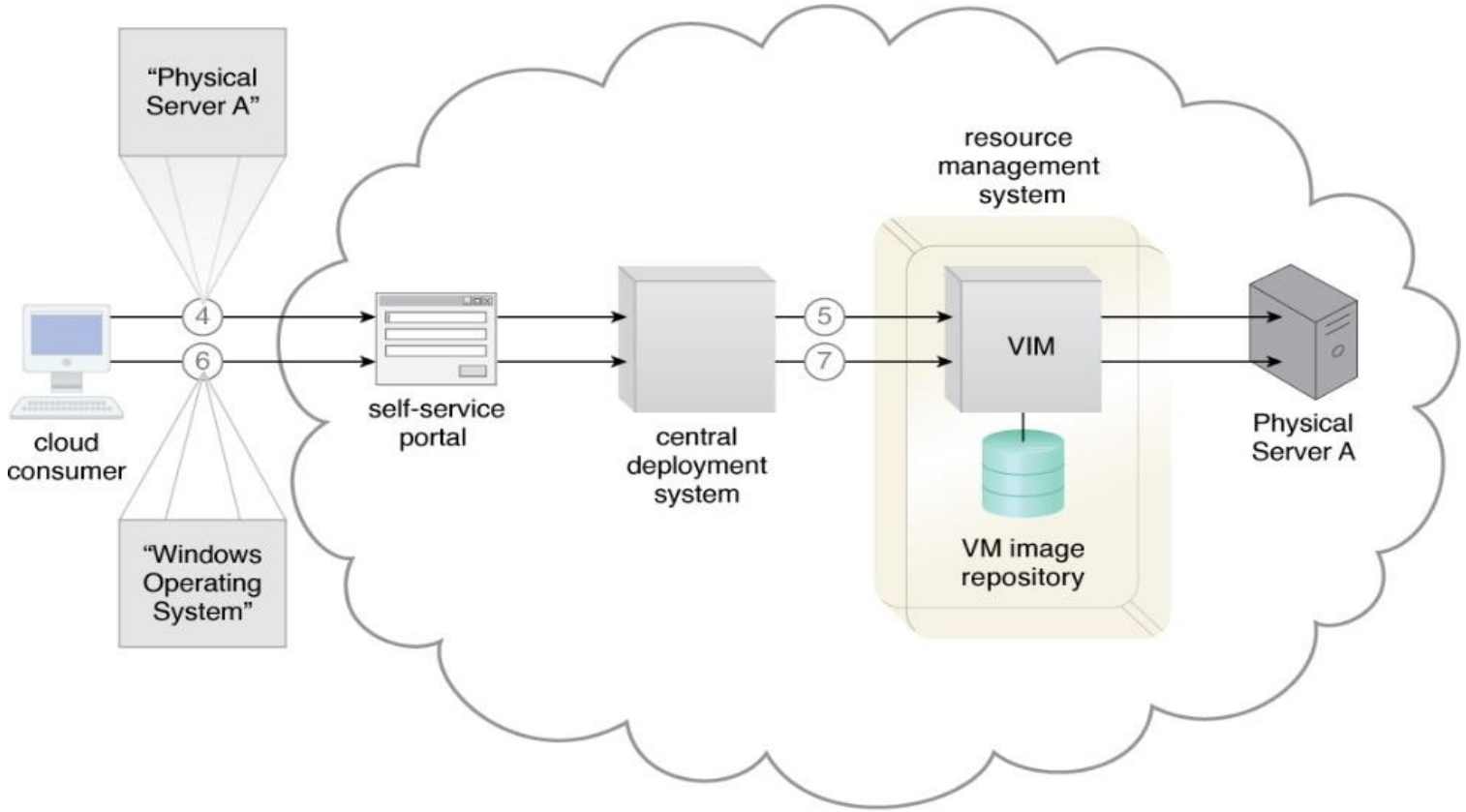
- 发现代理
  - 一种监控代理
  - 搜索并找到可用的物理机
- 部署代理
  - 裸机供给的客户端
  - 加载到物理服务器RAM中
- 部署组件
  - 用于安装操作系统

# 裸机供给架构




- The cloud consumer connects to the deployment solution (1), and uses the deployment solution to perform a search using the discovery agent (2).
- The available physical servers are shown to the cloud consumer, which selects the target server for usage (3).

# 裸机供给架构

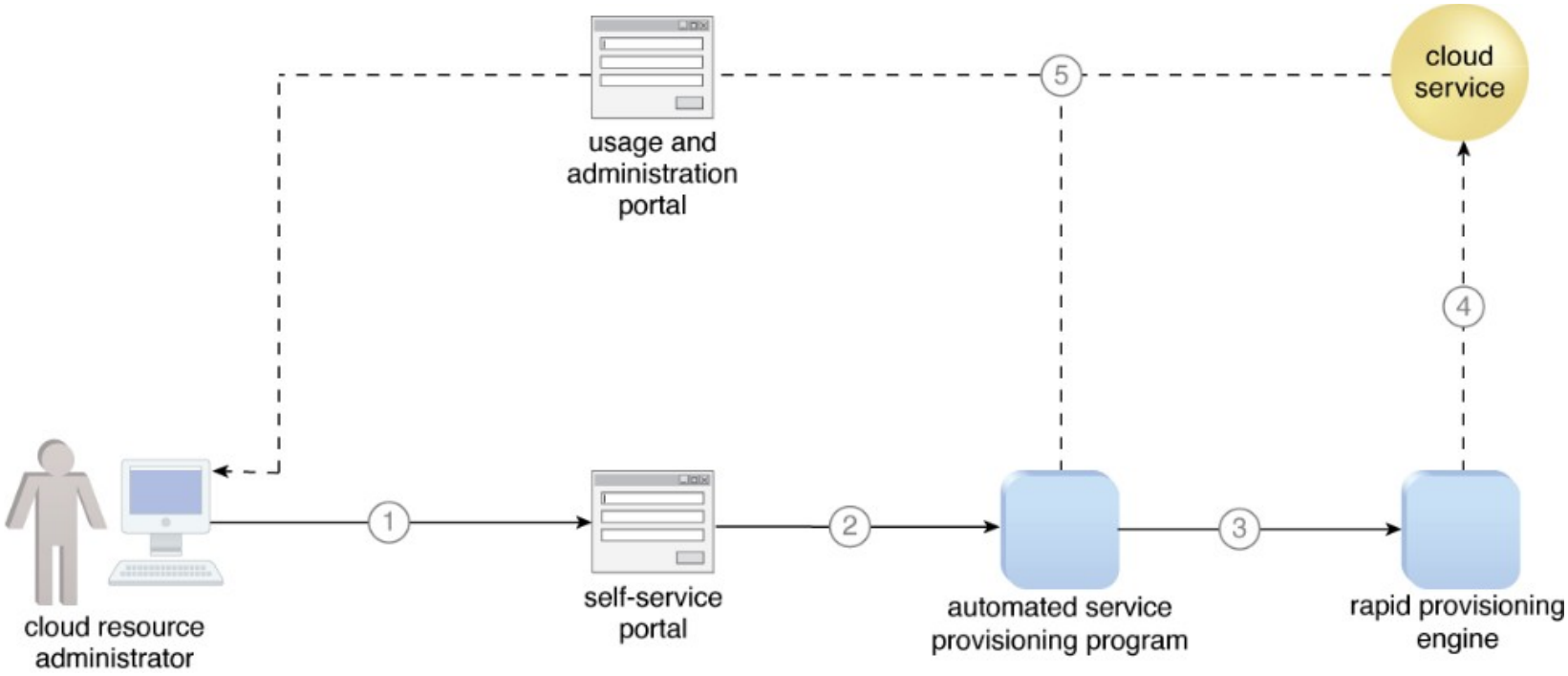



- The deployment agent is loaded to the physical server's RAM via the remote management system (4).
- The cloud consumer selects an operating system and method of configuration via the deployment solution (5).
- The operating system is installed and the server becomes operational (6).

# 快速供给架构

- Rapid provisioning architecture
  - 实现大范围的IT资源供给的自动化
  - 单个IT资源或者复合IT资源
  - 大用户量或者大资源量
- 主要构成部分
  - 顺序管理器（Sequence Manager）
    - 组织自动化供给任务的顺序
  - 顺序日志记录器（Sequence Logger）

# 快速供给架构

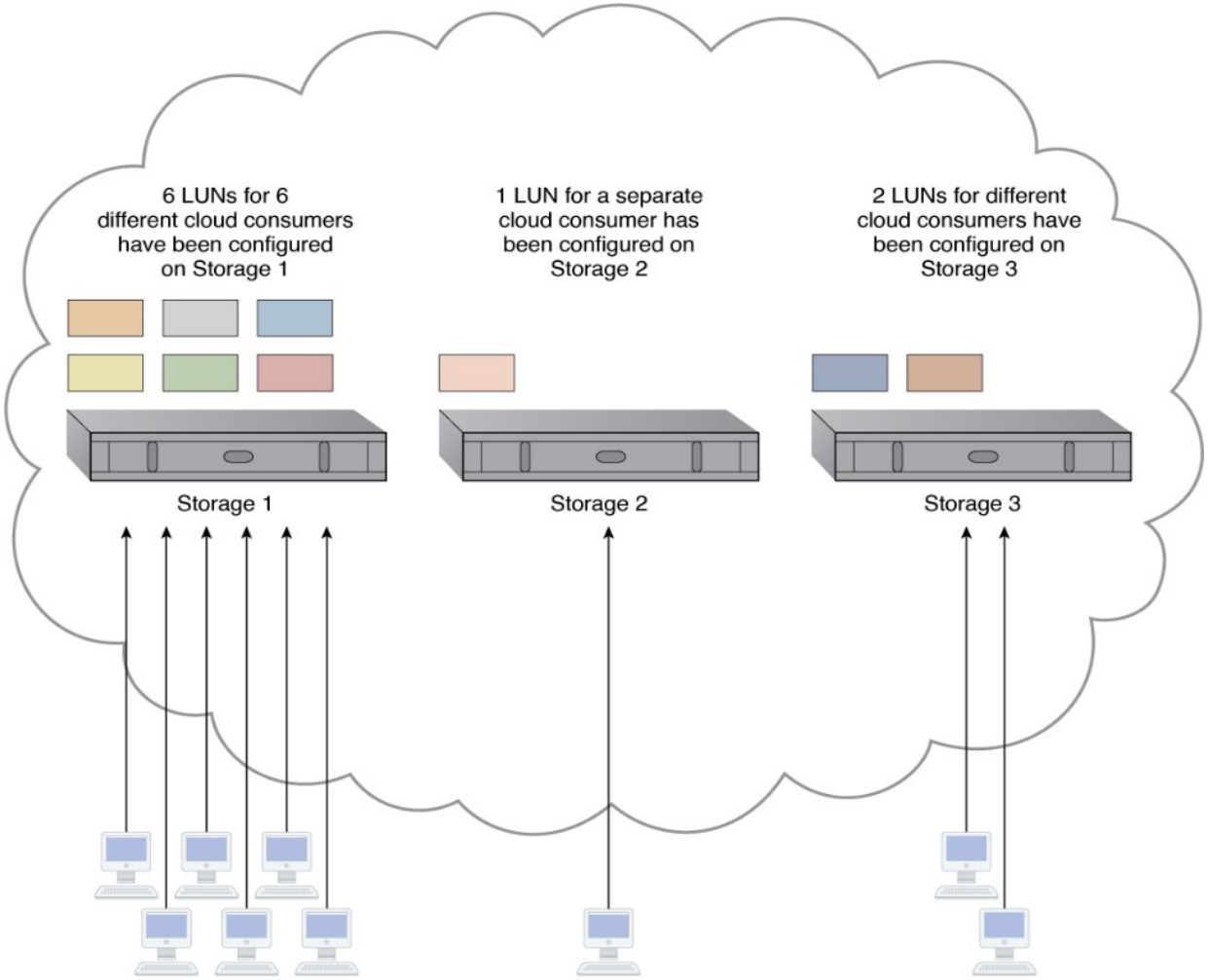



- A cloud consumer requests a new cloud service through the self-service portal (1).
- The self-service portal passes the request to the automated service provisioning program (2), which passes the necessary tasks to be performed to the rapid provisioning engine (3).
- The rapid provisioning engine announces when the new cloud service is ready (4).
- The automated service provisioning program finalizes and publishes the cloud service on the usage and administration portal for cloud consumer access (5).

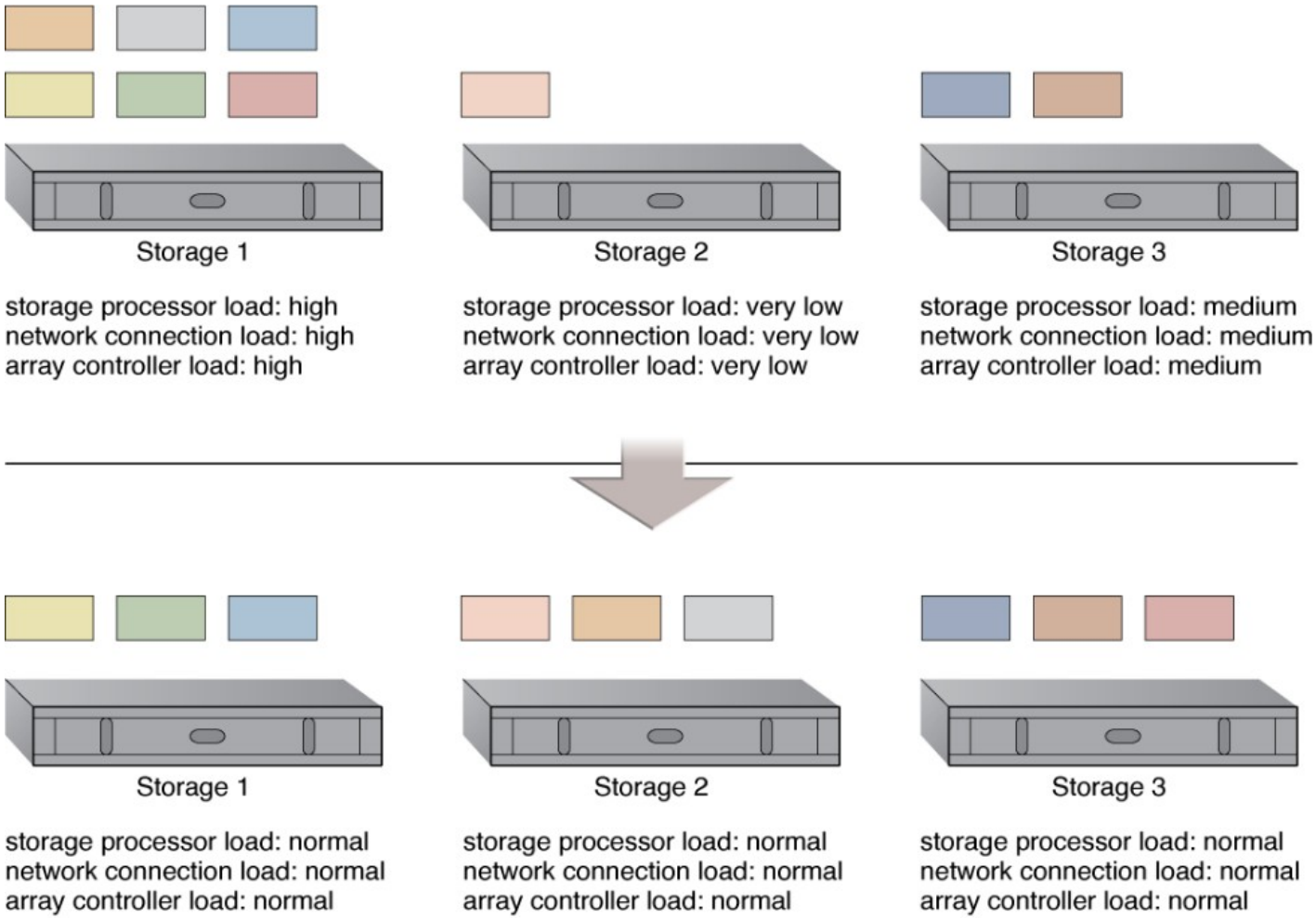
# §12.5 存储负载管理架构

- Storage workload management architecture
  - 使得LUN可以均匀地分布在可用的云存储设备上.
- LUN迁移
  - 把LUN从一个存储设备移动到另一个上而无需中断
  - 同时还对云用户保持透明。
- 存储容量系统
  - 则用来确保运行时工作负载均匀地分布在LUN上。

# 负载不均衡存储

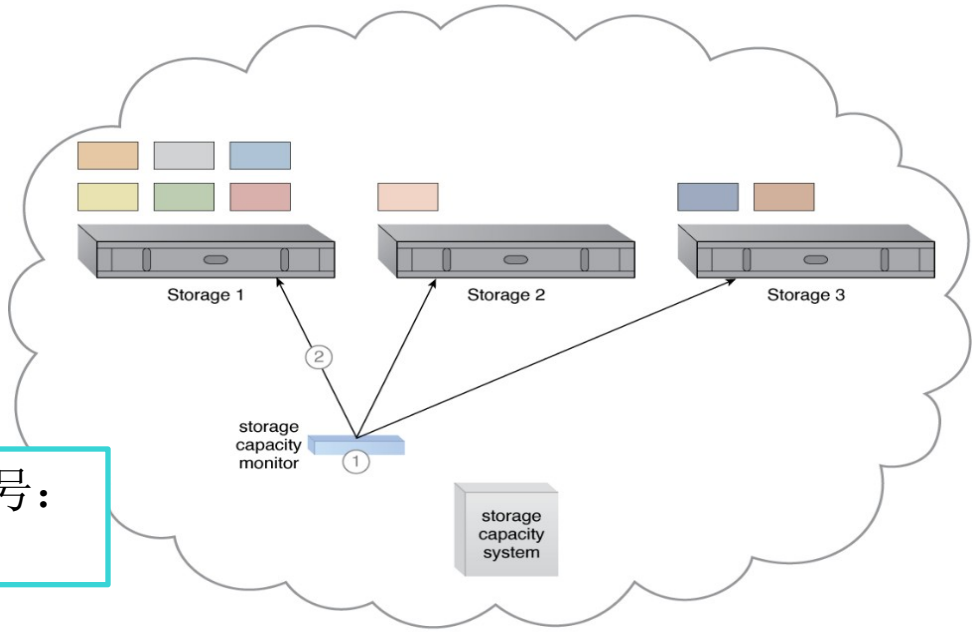


# 存储负载管理示例


Storage 1
storage processor load: high
network connection load: high
array controller load: high
Storage 2
storage processor load: very low
network connection load: very low
array controller load: very low
Storage 3
storage processor load: medium
network connection load: medium
array controller load: medium
Storage 1
storage processor load: normal
network connection load: normal
array controller load: normal
Storage 2
storage processor load: normal
network connection load: normal
array controller load: normal
Storage 3
storage processor load: normal
network connection load: normal
array controller load: normal

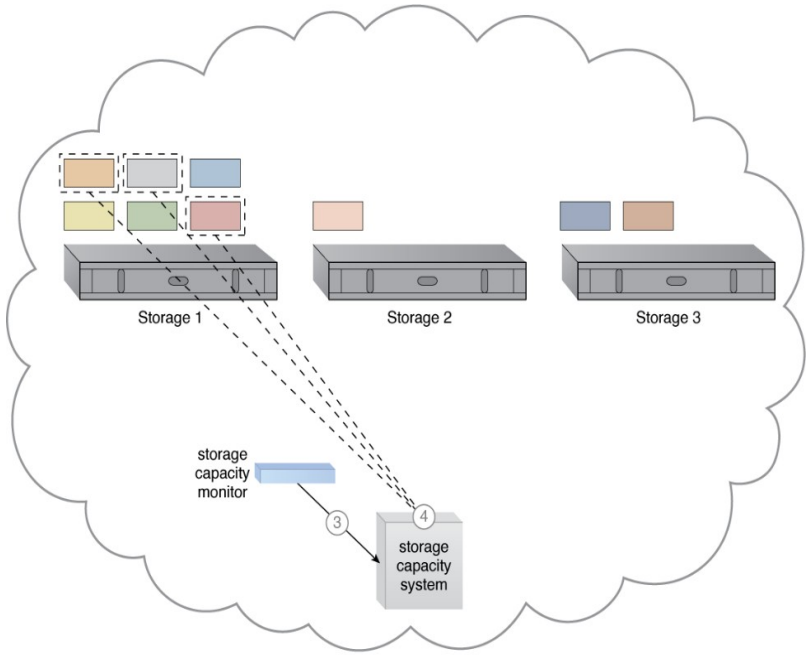
# 存储负载管理架构

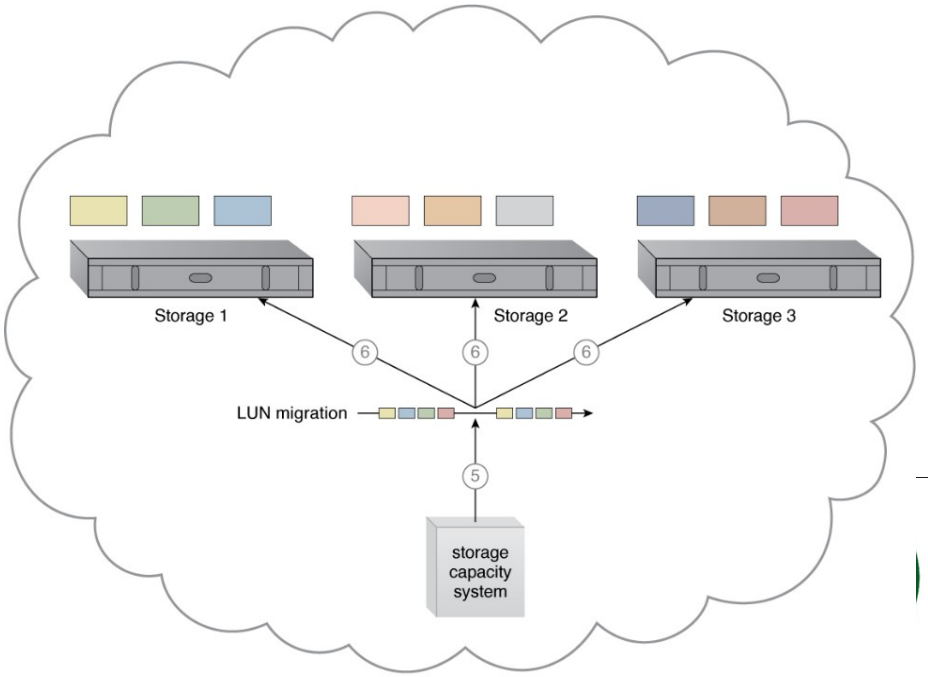
- 云存储设备合并成一个组
  - LUN数据在可用的存储主机上均匀地分布
- 放置一个自动伸缩监听器
  - 监控并且动态分配云存储设备之间的运行时工作负载
- 其他支持机制
  - 审计监控器
  - 负载均衡器
  - …



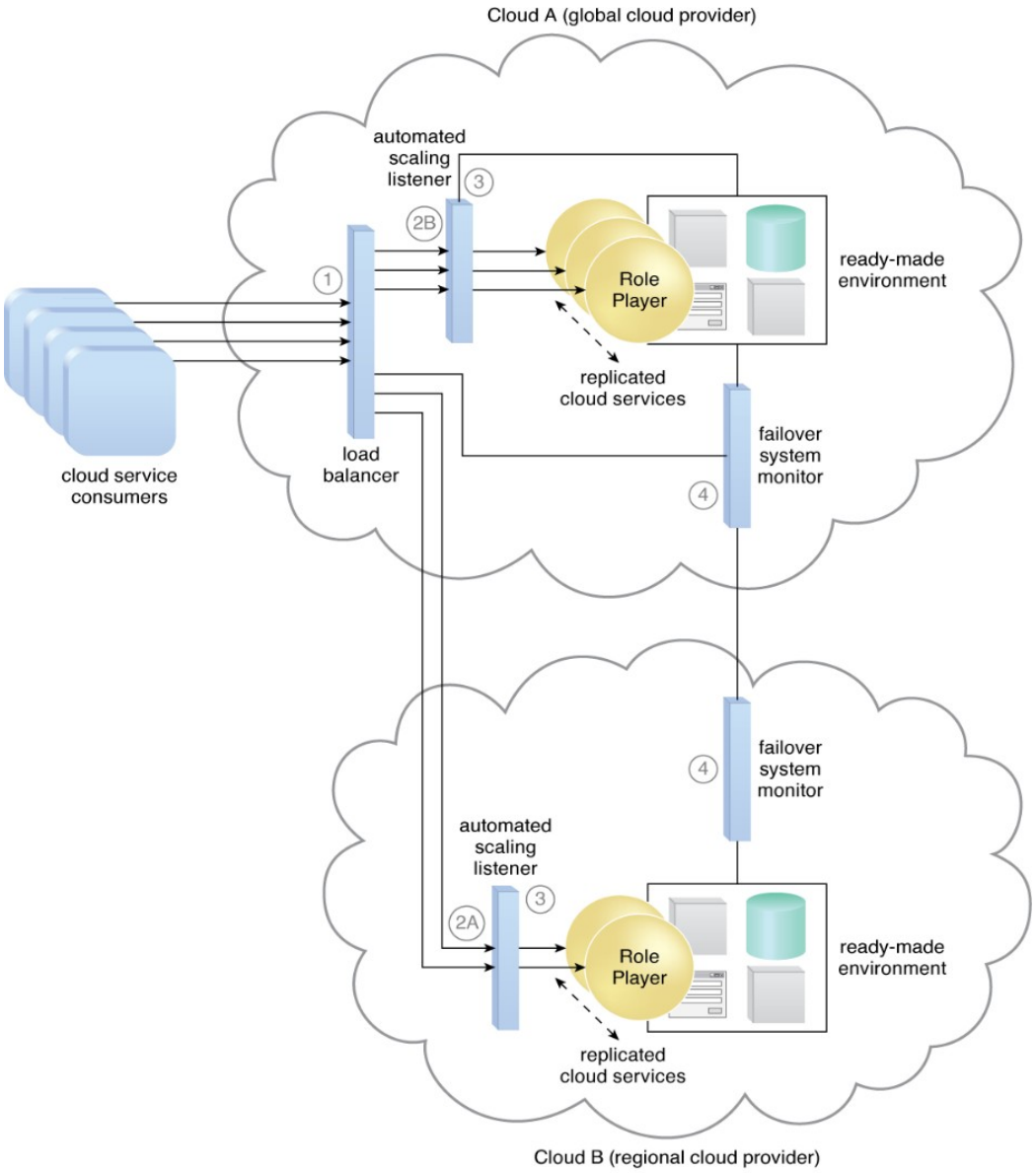

LUN，Logical Unit Number，逻辑单元号：存储管理的逻辑单元。

# 存储负载管理过程


Storage 1
Storage 2
Storage 3
storage capacity monitor
3
4
storage capacity system


Storage 1
Storage 2
Storage 3
6
6
6
LUN migration
5
storage capacity system

# 跨云的负载均衡



# 课后题

1、分析比较云服务容错的几种机制。

2、思考云服务负载均衡和存储负载均衡两种负载均衡需求及对应机制的差别。