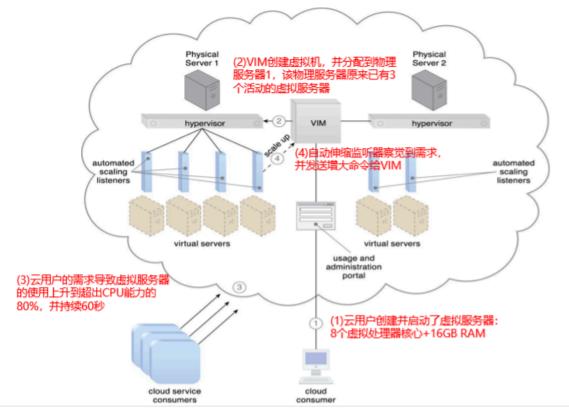
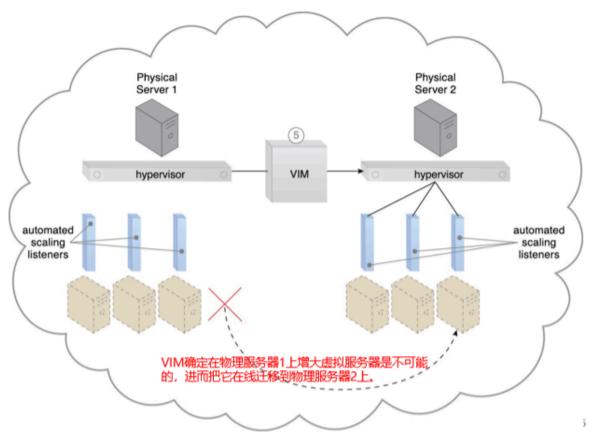
第八章 特殊云机制

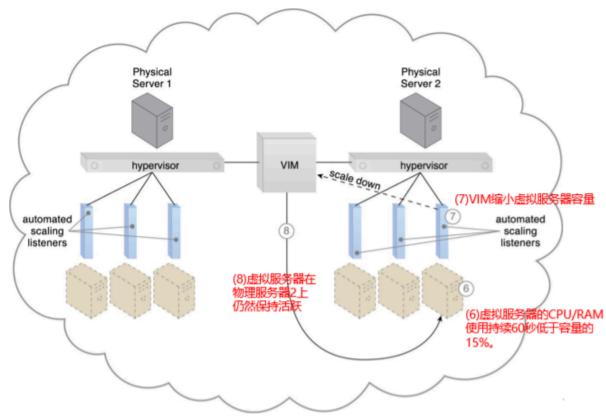
8.1 自动伸缩监听器

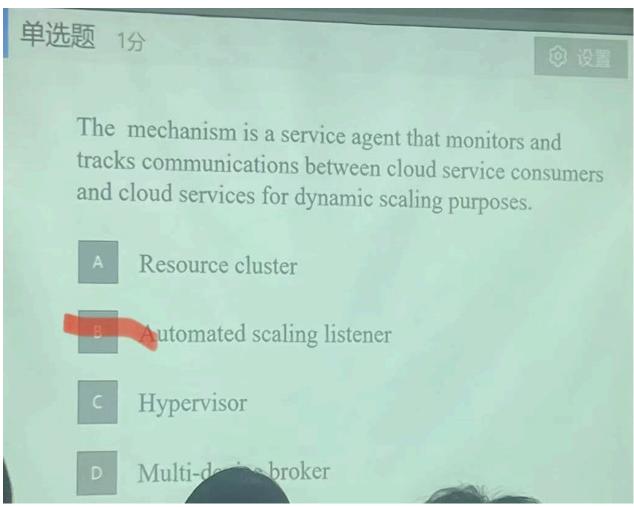
- 1. 自动伸缩监听器 (Automated scaling listener)是一个**服务代理**,它监控和追踪**云服务用 户和云服务**之间的**通信**,用以**动态伸缩**。
- 2. 通常部署在**靠近防火墙的位置**,自动追踪负载状态信息;负载量可以由**前端请求量**,也可以由请求引发的**后端处理需求量**决定。
- 3. 提供不同类型的响应:如通过预定义参数**自动伸缩IT资源**,或者负载超过或低于阈值时**自** 动通知云用户。



4.









自动伸缩监听器的响应类型包括:

- 负载超过阈值时自动通知云用户
- 5 负载低于阈值时自动通知云用户
- 负载超过阈值时宕机以免给用户产生额外费用
- 预定义参数自动伸缩IT资源

6.

多选题 1分

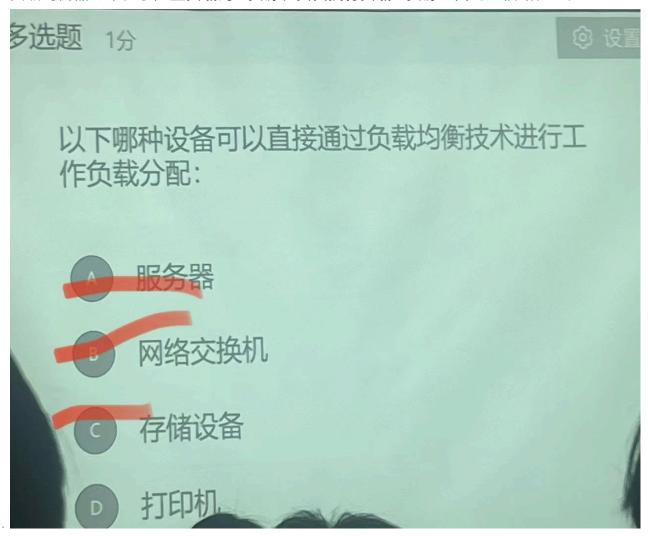
◎ 设置

以下对自动伸缩监听器表述正确的是:

- 自动追踪负载状态信息
- 负载量可以由前端请求量决定
 - 载量可以由请求引发的后端处理需求量决定
 - p 自动伸缩IT资源

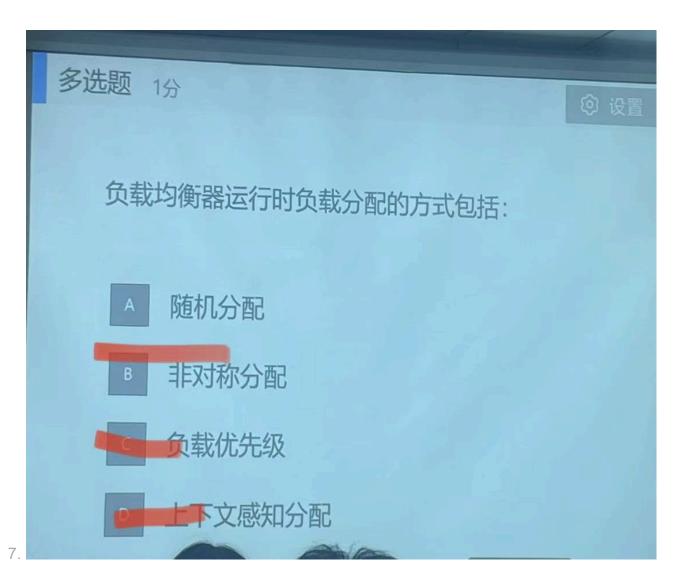
8.2 负载均衡器

- 1. 负载均衡器(Load balancer)机制是一个运行时代理,基本思想是把负载在更多的IT资源上做负载均衡,与单一IT资源相比,提升了性能和容量。
 - 1. 基于水平拓展的思想。
 - 2. 可以负载均衡的资源或者设备有服务器、网络设备、存储设备等。
- 2. 负载均衡器可以执行一组特殊的运行时负载分配功能,包括:
 - 1. **非对称分配**:较大的工作负载被送到具有较强处理能力的IT资源;
 - 2. 负载优先级:根据优先等级进行调度、排队、丢弃和分配;
 - 3. **上下文感知分配**:根据请求内容的指示把请求分配到不同IT资源。
- 3. 负载均衡的机制:
 - 1. 多层网络交换机
 - 2. 专门的硬件设备
 - 3. 专门的基于软件的系统
 - 4. 服务代理
- 4. 负载均衡器通常位于**产生负载的IT资源**和**负责执行负载IT资源**之间的通信路径上。



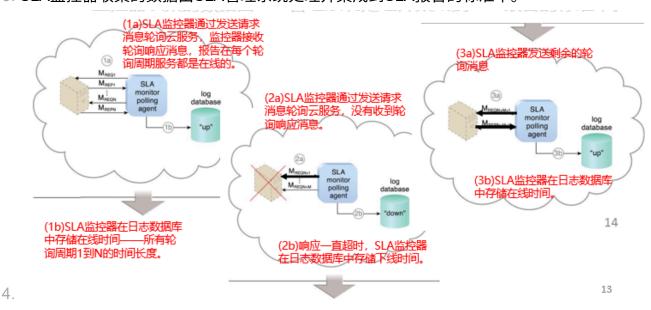
负载均衡器的机制包括:

- 多层网络交换机
- 8 专门的硬件设备
- 专门的基于软件的系统
- D 服务代理



8.3 SLA监控器

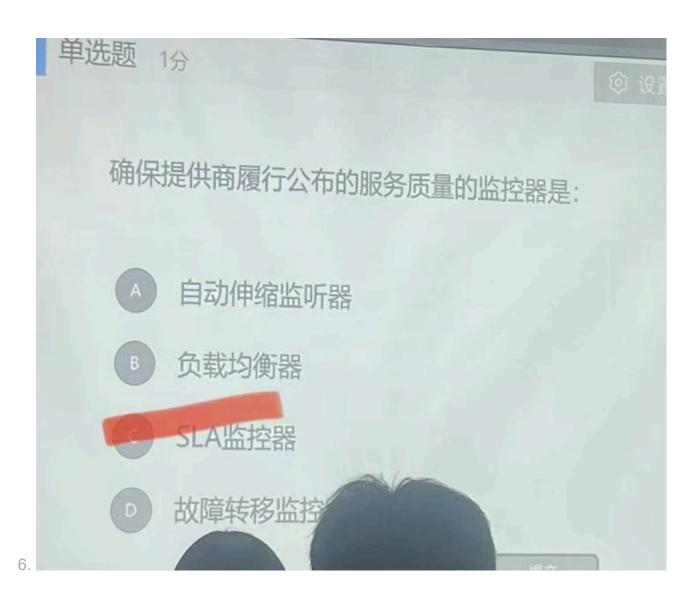
- 1. SLA: Service-Level Agreement,服务等级协议。
- 2. SLA监控器(SLA monitor)被用来专门观察云服务的**运行时性能**,确保他们履行SLA公布的QoS需求。
- 3. SLA监控器收集的数据由SLA管理系统处理并集成到SLA报告的标准中。



mechanism is used to specifically observe the runtime performance of cloud services to ensure that they are fulfilling the contractual QoS requirements that are published in SLAs.

- A Audit Monitor
- B Load Balancer
- c Hypervisor

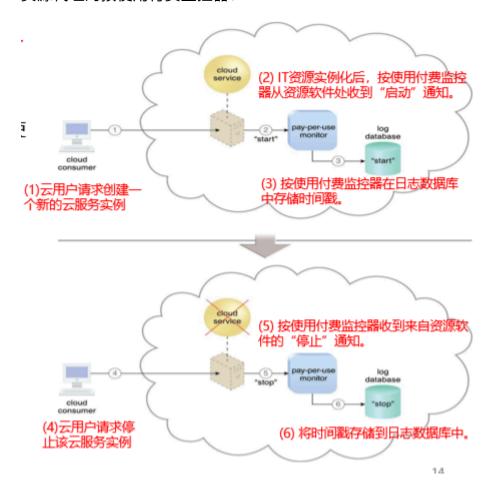




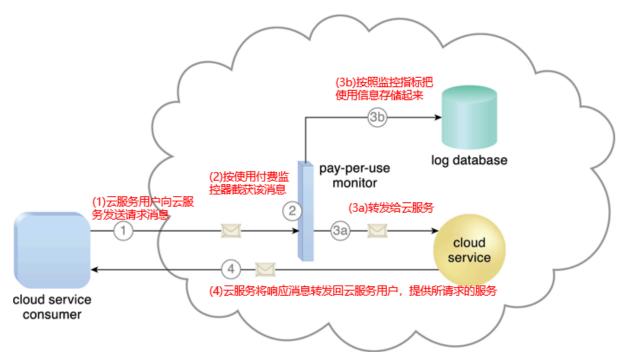
8.4 按使用付费监控器

- 1. **按使用付费监控器**(Pay-per-use monitor)机制按照预先定义好的定价参数测量基于云的 IT资源使用,并生成使用日志用于计算费用。
 - 1. 请求/响应消息数量
 - 2. 传输的数据量
 - 3. 带宽消耗

2. 资源代理的按使用付费监控器:

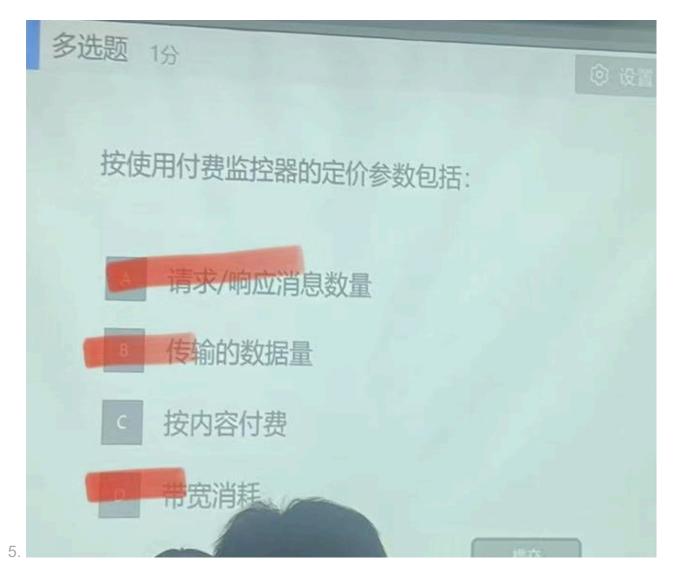


3 监控代理的按使用付费监控器:



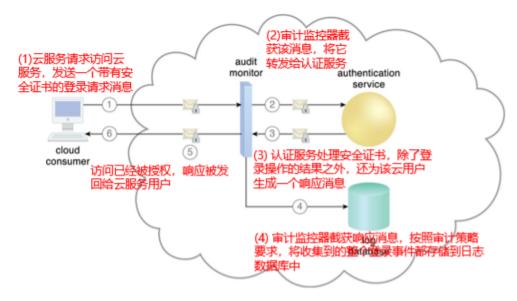
mechanism measures cloud-based IT resource usage in accordance with predefined pricing parameters and generates usage logs for fee calculations and billing purposes.

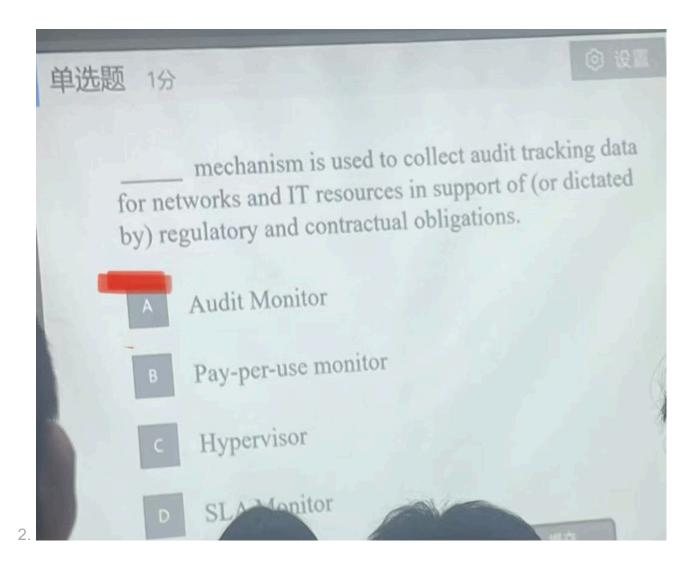
- A Audit Monitor
- Pay-per-use monitor
- c Hypervisor
- D SLAM nitor



8.5 审计监控器

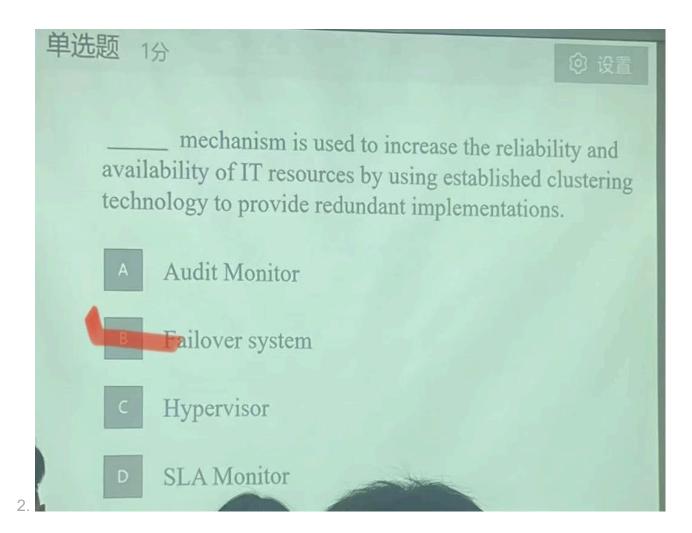
- 1. **审计监控器机制(Audit monitor**)用来收集网络和IT资源的**审计记录数据**,用以满足管理需要或合同义务。
 - 1. 下图审计监控器截获登录请求,在日志数据库中存储请求者的安全证书,以及成功和 失败的登录尝试,以供今后审计报告之用。





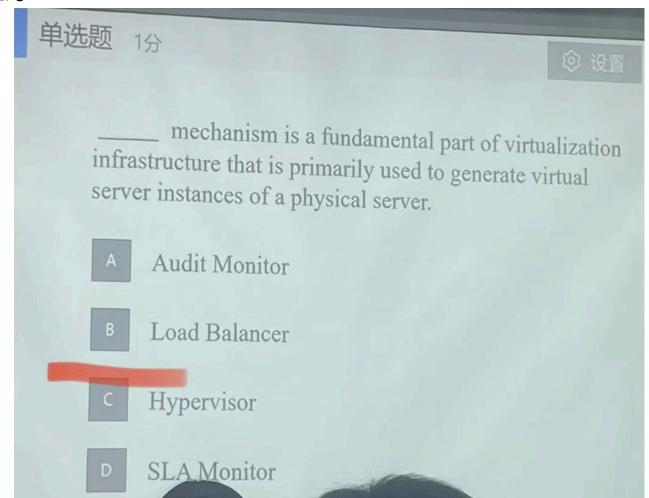
8.6 故障转移系统

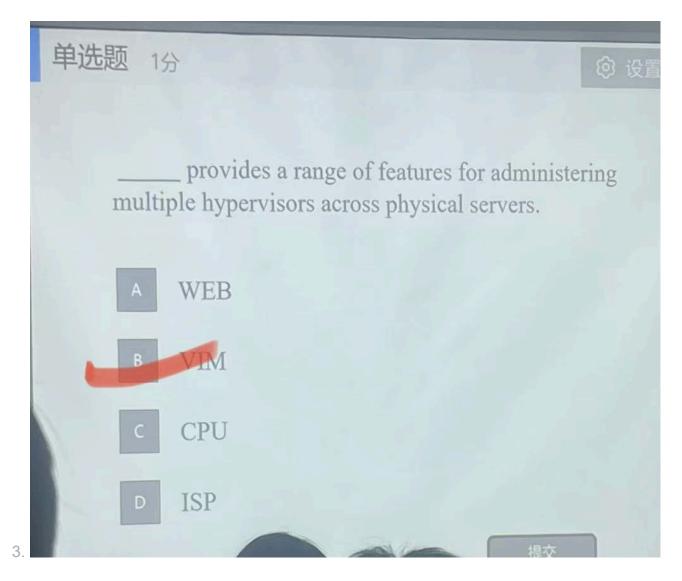
- 1. **故障转移系统**(Failover system)通过使用现有的集群技术提供冗余的实现来增加IT资源的可靠性和可用性。
 - 1. 主动-主动: IT资源的冗余实现会主动地同步服务工作负载; 当发现故障时,将失效 的实例从**负载均衡调度器中剔除**,有效的IT资源就会接管处理工作。
 - 2. 主动-被动: 待机或非活跃的实现会被激活,从不可用的IT资源处接管工作。



8.7 虚拟机监控器

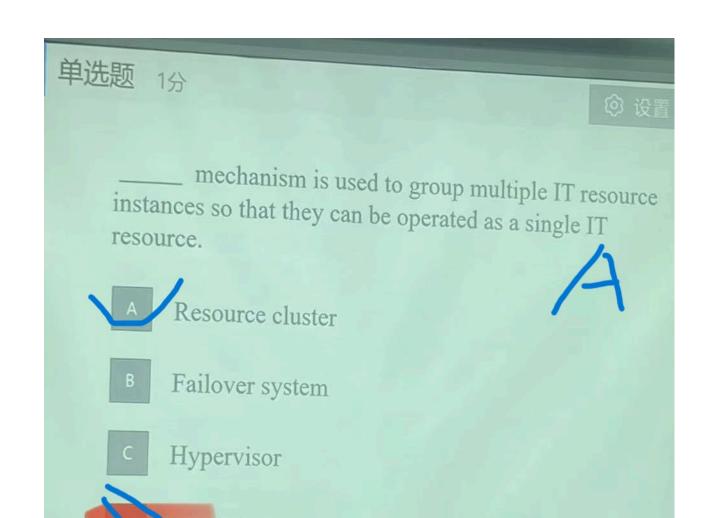
1. **虚拟机监控器**(Hypervisor)机制是虚拟化基础设施中最基础的部分,主要用来在物理服务器上生成虚拟服务器实例。





8.8 资源集群

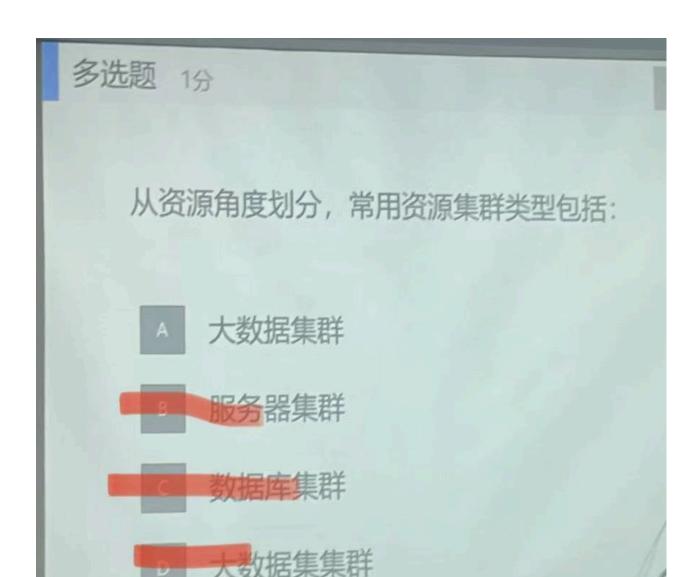
- 1. **资源集群**(Resource cluster)将多个分布的IT资源分为一组,使得他们能**像同一个IT资 源一样进行操作**。这增强了集群化IT资源的组合计算能力、负载均衡能力和可用性。
- 2. 从资源角度划分,常用资源集群类型包括:
 - 1. 服务器集群:物理或虚拟服务器组成集群。
 - 2. 数据库集群:用于改进数据的可用性,具有同步的特性,可以维持集群中各种存储设备上存储数据的一致性。
 - 3. 大数据集集群:实现了数据的分区和分布,目标数据集可以有效的划分区域,而不需要破坏数据的完整性或计算的准确性。
- 3. 从性能角度划分,资源集群的两个基本类型:
 - 1. 负载均衡的集群:这种资源集群在集群节点中分布工作负载,既提高IT资源容量又保持资源的集中管理。
 - 2. 高可用集群:在遇到多节点失效时,仍能维持系统的可用性,而且大多数或者所有集群的IT资源都有冗余实现。它实现一个**故障转移系统机制**,监控失效情况,并自动将工作负载重定向为远离故障节点。

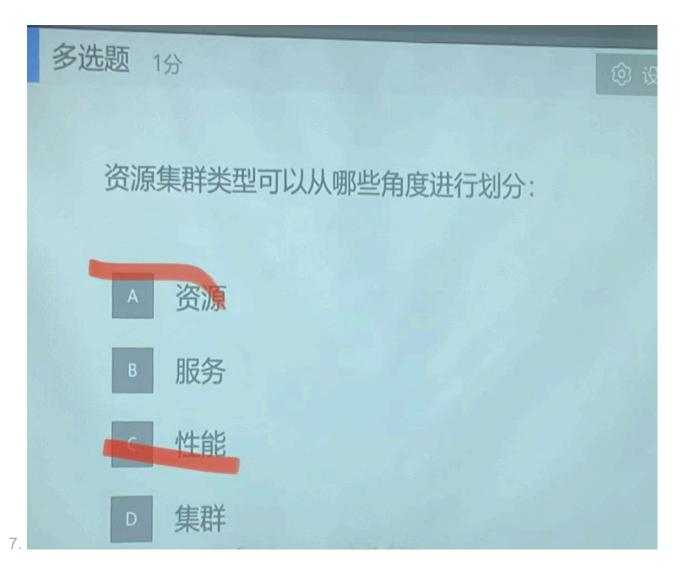


4.

从性能角度划分,资源集群的基本类型包括:

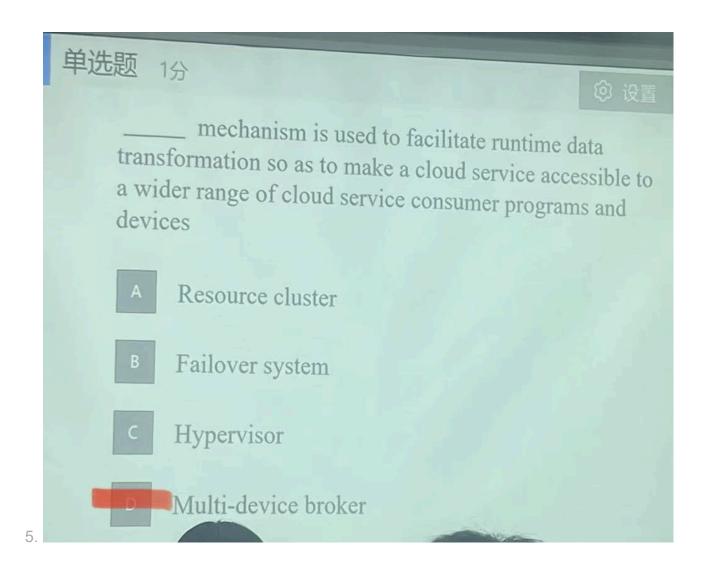
- **负载均衡的集群**
 - B 负载优先的集群
 - 6 负载冗余的集群
- 高可用集群





8.9 多设备代理

- 1. 一个云服务可能会被大量的云用户访问,他们对主机硬件设备和通信需求都不同。为了克服云服务和不同云服务用户之间的不兼容性,需要创建映射逻辑来改变运行时交换的信息。
- 2. **多设备代理(Multi-devise broker**)来帮助运行时的**数据转换**,使得云服务能够被更广泛的云用户程序和设备所使用。
- 3. 多设备代理通常作为网关存在。
- 4. 可创建转化的逻辑层次包括:
 - 1. 传输协议
 - 2. 消息协议
 - 3. 存储设备协议
 - 4. 数据模式/数据模型



8.10 状态管理数据库

- 1. **状态管理数据库**(State management database):一种用来暂时地保存软件程序的**状态数 据**。
- 2. 软件程序可以把状态数据保存到数据库中,用以降低程序占用的运行时内存。