

# 分解式数据中心中的可靠性感知资源分配方法

## 通信和信息

计算机/人工智能/数据处理和信息技术  
数字广播、电信和光电

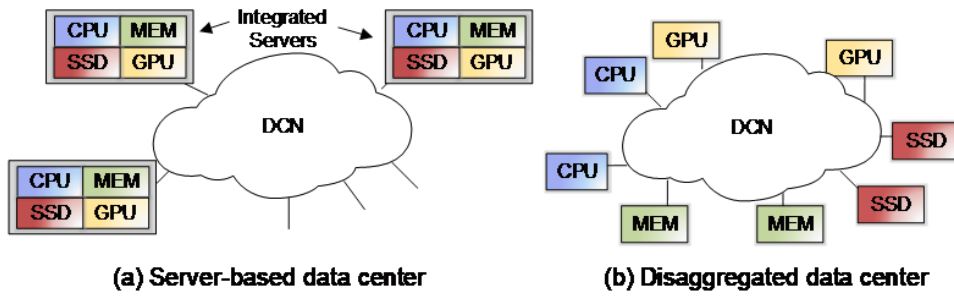


图 1. 基于服务器的分解式 DC。

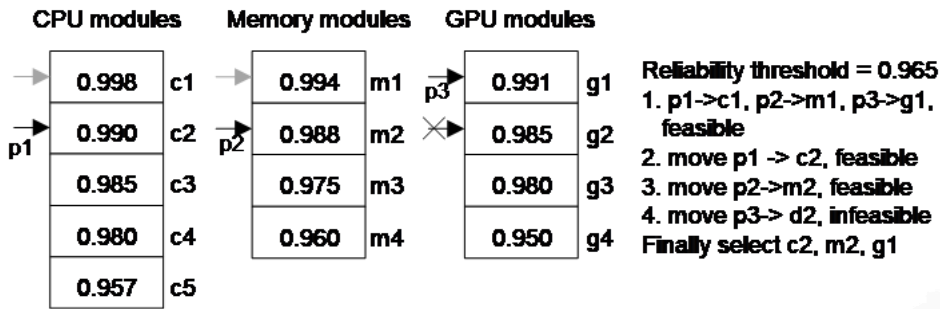


图 2. 根据本发明的一个实施例，使用启发式过程对 DDC 执行服务请求的示例设置和条件的说明图。

## 机会

数据中心（DC）在提供互联网服务方面扮演着越来越重要的角色。全球数据中心市场在2020年的估值为444.2亿美元，预计从2021年到2028年将以13.3%的速度增长。大规模数据中心由众多服务器组成，每个服务器都紧密集成了各种资源。这些服务器通过专用网络相互连接。

然而，当工作负载多样化时，这种架构效率低下，因为每台服务器中不同类型的资源紧密耦合。当服务器中的一种资源耗尽时，其他资源也会变得不可用，从而造成浪费。这也使得服务器升级变得困难。尽管服务器中的组件通常有不同的生命周期，但硬件升级和资源扩展通常仅在服务器级别进行。

随着互联网服务和用户需求的快速发展，硬件和软件正在迅速变化，需要越来越频繁的硬件更新。为了减少更新成本、提高资源利用率和可升级性，资源解耦逐渐成为一个突出的解决方案。通过硬件解耦，不同类型的资源可以独立且灵活地升级或扩展，从而显著降低成本。为了支持这一新兴技术，迫切需要新的资源分配方法。

**IP状态**  
专利已存档

技术成熟度等级 (TRL) ?

4

发明人  
Prof. ZUKERMAN Moshe  
郭朝  
询问: kto@cityu.edu.hk



## 技术

本发明提出了一种面向未来数据中心（DC）的创新性可靠性感知资源分配方法。首先，研究人员对基于服务器或分解式数据中心中的资源分配请求的可靠性进行了建模。然后，他们考虑了一个资源分配问题，以最大化在保证可靠性的前提下接受的请求数量。这被表述为一个整数线性规划问题。此外，他们还提出了一种更简便的启发式方法。通过大量的模拟研究，数值结果表明，使用他们的新颖资源解耦方法可能显著提高数据中心的可靠服务性。

## 优势

- 更好地保证数据中心的可靠服务性
- 提高数据中心的资源效率

## 应用

- 云计算提供商在建设大规模数据中心时使用
- 实现数据中心所有者（如微软Azure和亚马逊Web服务）在保证可靠性的同时最大化利润或最小化成本的资源分配

