## WORKLOAD CHARACTERIZATION

## D. Workload consolication opportunities across machines

我们对III节C部分的结果进一步进行了统计分析，根据machine cpu利用率和memory利用率的相关系数，我们将machine分为3类：

1. cpu利用率和memory利用率高度相关的节点：当machine cpu利用率和memory利用率的相关系数α>=0.85，我们将该类节点称为高度相关的节点。即该类节点的cpu利用率和memory利用率的变化方向和趋势密切相关。
2. cpu利用率和memory利用率负相关的节点：当machine cpu利用率和memory利用率的相关系数α<0，我们将该类节点称为负相关的节点。即该类节点的cpu利用率和memory利用率的变化趋势相反。
3. cpu利用率和memory利用率低度相关的节点：当machine cpu利用率和memory利用率的相关系数0<α<0.85，我们将该类节点称为低度相关的节点。即该类节点cpu利用率和memory利用率虽然存在联系，但联系的紧密程度比较弱。

为了更加方便的对节点进行分析，我们记录了每个节点的所属类型，并且计算了1313个节点中各类节点的个数，其中a类节点有198个，b类节点有6个，c类节点有1109个。

根据3类节点的各自特点，为了提供工作负载整合机会，我们主要分析a类节点。

从a类节点的统计结果，我们可以发现节点并非是单独工作的，节点88-127，节点261-296，节点830-906为分别为3个集群。集群内的节点运行同类型的job，因为a类节点cpu和memory的利用率是高度相关的，所以在同一个集群内，整合节点后的cpu，memory的使用率是可以计算的，根据计算结果，在资源足够使用的情况下，将节点进行整合，从而最小化节点使用数目。

除此外，根据以上的分类结果，在分配不同类型的job时，我们可以根据节点的特点做出一个更加合理的分配决策。例如，当job类型为计算型或者内存型时，尽可能的不将其分配给a类节点，从而在cpu和memory之间做出一个折中。