## 消除热点、冷点与负载均衡的亲和性问题

## 需求

## 冷热检测

定义温度T为资源占用率，按照基本的资源分配中不能过热和过冷的要求，给出分配的温度上限HOT、温度下限COLD。

PM每个维度都有温度，构成一个向量TD=（Td）,d=1…DIM。任意资源维度的温度大于HOT，则说明PM在这个维度上过热，将PM定义为热点PM。

1. **系统整体热、冷点检测**

计算系统所有VM的资源需求总量RS，计算所有PM的资源总量CS，计算总体资源占用率:TS=RS/CS。

其中E为容错修正系数，E<1。

1. **PM热点检测**

任何一台PM任意资源维度的温度>HOT，则该PM为热点机。

1. **PM冷点检测**

一台PM任意资源维度的温度<COLD，则该PM为冷点机；也可定义为各维度温度的最大值<COLD。

1. **负载均衡检测**

在没有冷点和热点的情况下，各PM之间温度的均方差，作为负载均衡度D。如果D>给定的阈值f，则系统处于不均衡状态。

## 热点消除

输入一组VM，PM，以及亲和性信息，调度算法检查出当前PM中存在热点（>hot，如0.8）的PM，然后通过迁移部分VM到其它PM，同时不破坏亲和性，使该PM中的各个资源维度都不存在热点。

疑问：不破坏亲和性，无法解除热点，如何处理？

## 冷点整合

输入一组VM，PM，以及亲和性信息，调度算法检查出当前系统状态下是否存在冷点，如果系统整体比较冷，

当前PM中存在冷点（<cold，如0.2）的PM，然后通过迁移部分VM到其它PM，同时不破坏亲和性，使该PM中的各个资源维度都不存在冷点。

## 负载均衡

对不存在冷热点的且负载均衡度较高的系统，进行负载均衡。

1. 基本定义
2. PM维度：
3. PM综合温度值：TP，定义为PM各维度温度的加权求和值。

计算公式：TP=W.T. W为各维度的权向量。

系统最舒适温度值BT：定义为各PM综合温度的均值。

PM均衡偏离值DT：为PM综合温度与舒适温度值的偏离值，即两者之差。

计算公式：DT=TP-BT.

## 总体处理策略

如何满足亲和性？

对一个给定存在热点或冷点的PM，热点的称为hot\_PM，冷点的称为cold\_PM。以热点消除为例，取出hot\_PM上的VM，根据亲和性进行分组，得到一个VMG列表，对每个VMG，获得可放置的其它PM集合，然后，从可放置PM集合中选取一个最佳放置的PM，将VMG放入，使能消除热点。

对每个VMG基于亲和性确定可放置PM的方法：

1）如果没有亲和性，所有PM都可放置。

2）如果有同框亲和性，与该VMG有同框亲和性的其它VMG所在框的所有PM；

3）如果有同组亲和性，与该VMG有同组亲和性的其它VMG所在组的所有PM；

4）如果有不同框亲和性，与该VMG有不同框亲和性的其它VMG所在框的PM之外的PM；

5）如果有不同组亲和性，与该VMG有不同组亲和性的其它VMG所在组的PM之外的PM。

满足亲和性之后，再考虑基于容量来确定可放置的PM，筛选出能放得下并且能消除热点的PM，如果有多个PM，则选出取**最好**（基于流量与距离加权确定最合适的，基于容量最合适的）的PM。

解决方案：

初始放置之后，记录VM所在流量亲和组。在放置时，依据PM容量被切割后生成亲和小组，保留小组与大组的关系信息。小组之间依据流量关系需要就近放置。

在动态调度时，流量亲和组信息，从外部传入。存储流量亲合组的关系信息，每个VM记录所在组号。

<vm atvmg=”0” atpm=”” >

st\_vmg2vmg亲和性

xml格式:

<affinity id=”1” vmids=”” >

接口：增加规则类，一种亲和性对应一种规则。

get\_diff\_pm\_pms(vmgPtr vmg): 与当前VMG不同板的PM；

get\_sam\_grp\_pms(vmgPtr vmg): 与当前VMG同组的PM；

get\_diff\_grp\_pms(vmgPtr vmg): 与当前VMG不同组的PM；

get\_sam\_frame\_pms(vmgPtr vmg): 与当前VMG同框的PM；

get\_diff\_frame\_pms(vmgPtr vmg): 与当前VMG不同框的PM；

get\_canfit\_pms(vmgPtr vmg): 获取可放置当前VMG的PM；