

碎片化指标V2.0

- 问题
- 范畴
 - 对象：虚拟化资源池（Pool）
 - 粒度：资源池（而非业务）
- 指标
- 指标数据

问题

指标1.0(wiki)在数据获取上有很大的限制导致无法获取，无法真正体现碎片情况。原因是碎片化率指标的计算需要获取“业务”的全链路“极限”压测时的“CPU”利用率数据。

- “业务”维度的极限压测执行上有困难，如大搜，还未能实现全链路压测，只能单独压测BS（BCBS是独立的资源池）
- “业务”维度的碎片化数据不可获取，如大商业、凤巢、网盟、原生为统一资源池，各业务压测并不进行统一压测（高峰期不同），之前只拿凤巢一个业务的极限压测时的峰值利用率来计算碎片率，结果是偏大。
- “极限”压测，各业务的“极限”大原则一致，即压测到break KPI，但实际判断标准上会有一些diff，比如大商业会在观察单机CPU IDLE达到一定阈值时提前停止压测，结果也会略微偏大。
- 只能体现“CPU”碎片，因为压测时内存利用率变化不大。
- 碎片化率的整体指标与分拆指标没有直接换算关系，不好量化分拆指标对于整体碎片化率指标的影响。

范畴

- 对象：虚拟化资源池（Pool）
 - 每个业务都有多种资源池，包括虚拟化资源池、物理资源池、次级资源池。从技术上来讲，虚拟化资源池更方便进行资源精细化管理，因此是虚拟资源池中的碎片是我们的主要优化对象。而物理资源池，会随着虚拟化的推进逐步缩小，因此这部分资源碎片由虚拟化改造来优化（Paas化机器资源利用率高出物理机5%，2018.4）。次级资源池，尚未大量应用，暂不考虑优化。
- 粒度：资源池（而非业务）
 - 虚拟化资源池在使用上会有以下三种场景，单个业务多个资源池（大搜），多个业务共用资源池（大商业）、一个业务专用资源池（feed）。由于Paas的各种资源调度策略均以资源池生效，且随着合池推进，搜索公司会逐步合成大商业、大搜、Feed、用户产品等几个大资源池，并最终合成一个，因此碎片化也以资源池（Pool）为粒度进行优化。
- 结论：

资源池	所属业务	所属Paas
www_beehive_bcbs	大搜	Beehive
SSG_SAG-uniform	大商业（凤巢&原生&网盟）	Opera
	Feed	Sofacloud

指标

碎片场景	指标1
外碎片	<p>外碎片率</p> <p>=（未分配的资源量 / 资源总量）* 100%</p> <p>说明：虚拟资源池中的空机器、故障机器，会放在在线率指标中考虑。需要从外碎片中去掉。</p>
内碎片	<p>内碎片率 = 内碎片率A + 内碎片率B</p> <p>= Σ（ 单个APP的峰值利用率 - 单元内所有APP峰值利用率的均值 * 单个APP分配资源量） / 单元内的资源总量 * 100%</p> <p>内碎片率A = Σ（超均值使用的APP的峰值利用率 - 单元内所有APP峰值利用率的均值）* 单个APP分配资源量 / 单元内的资源总量</p> <p>内碎片率B = Σ（单元内所有APP峰值利用率的均值 - 低于均值使用的APP的峰值利用率）* 单个APP分配资源量 / 单元内的资源总量</p> <p>说明1：理想的“单元内所有APP峰值利用率的均值”根据冗余度来计算。例如3+1冗余的业务，APP理想利用率峰值为66%（利用率上限按照100%计算）</p> <p>说明2：内碎片率A = 外碎片率B</p>

碎片	碎片率 = 外碎片率 + 内碎片率
	与CPU利用率优化的关系 : 外碎片率优化 + 内碎片率 B 优化 - 内碎片率 B 优化

说明：

外碎片体现资源分配的不充分性；

内碎片率体现的是资源使用的不均衡性，因此也会将[超过利用率均值的部分](#)作为碎片考虑，但是这部分优化不会带来资源利用率的提升而是下降，[内碎片量≠内碎片优化带来的资源收益](#)。

指标数据

	外碎片率（CPU/MEM）	内碎片率（CPU/MEM）	利用率均值（CPU/MEM）
大商业	23.23%/12.05%	36.92%/19.48%	73.39%/63.02%
大搜			
Feed			