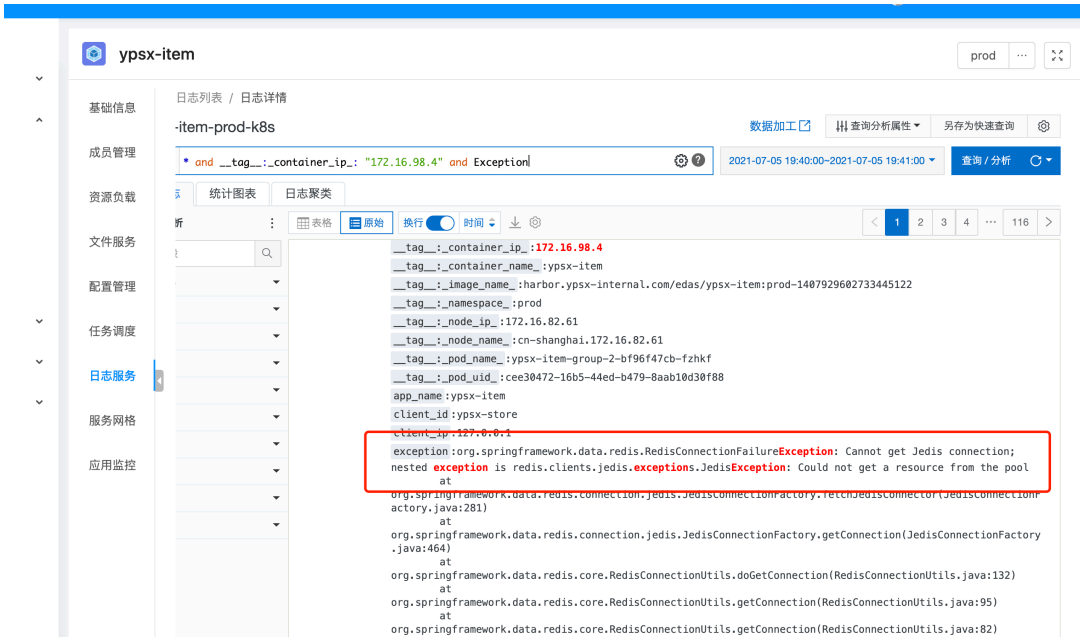
7月5号晚上 19点40分 接到电话反馈商品服务不可用，销售端相关依赖商品服务的页面展示空白页，后台ERP系统商品选品，类目接口页面提示系统异常。

以下是问题分析定位的过程：

第一步 workspace 查看监控，以及杨航反馈的截图，有一台商品的Pod的CPU使用率和内存使用率飙高，单台POD服务不可用，并定位这台异常的pod的容器Ip为 192.168.98.4

但从反馈的频率来看，商品有26个pod实例，线上每次刷新都是空白页；从分析上来看GRPC的负载策略是随机 +轮训，不可能每次都是空白页，页面的报错几乎是每次刷新必现。带着这个疑点去对日志进行了分析，从19点40 开始，按分钟的粒度逐步查询异常的日志，排查的过程中发现出现了如下的日志：

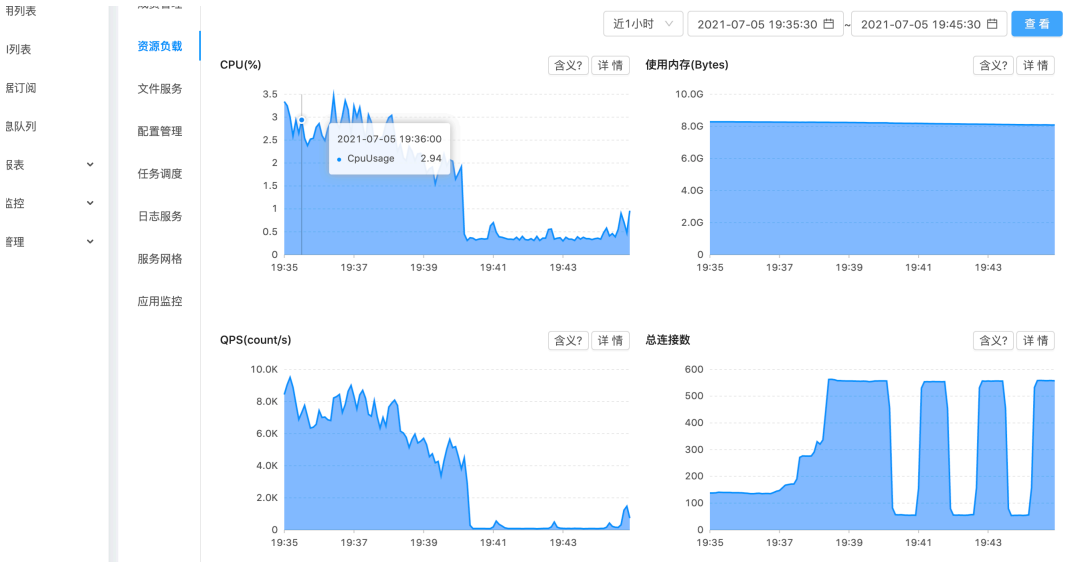


获取不到Jedis链接一般来说有两种可能：

1、Jedis的负载很高，响应很慢或者没有链接可以分配，会导致客户端的连接池获取不到链接

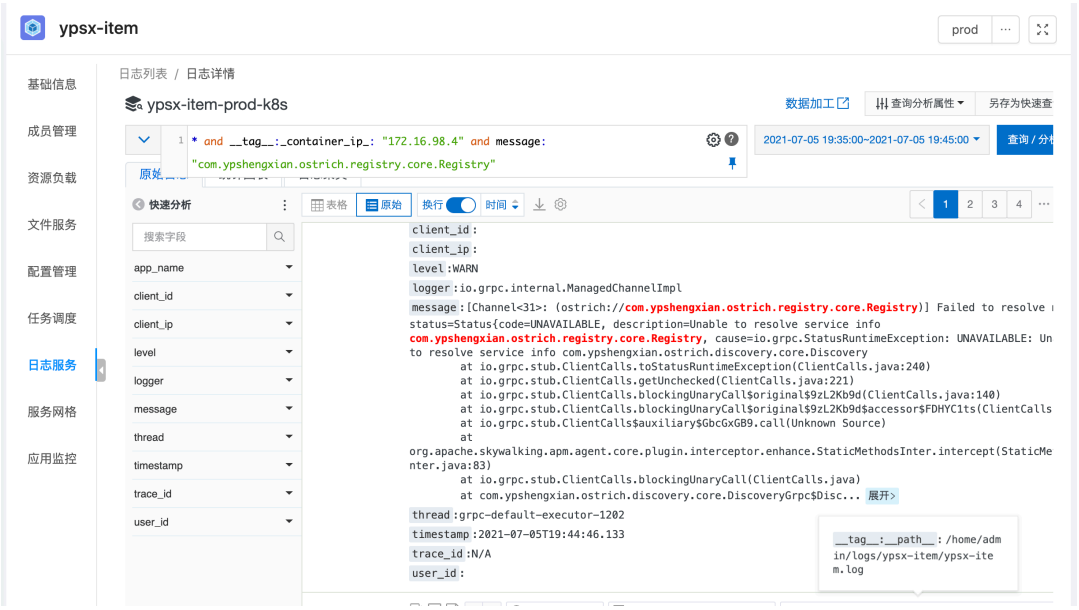
2、应用的负载很高，CPU没有资源来调度应用线程，这个时候也会造成链接Jedis的请求长时间没有响应，导致新来的请求把池中的链接资源分配光但不能及时释放链接的场景，也会导致这样的错误。

从时间上来分析，19点40的节点，应用的CPU使用已经 250%，cpuLoad全被占用，内存使用率也满了，很容易可以分析出这个时间的应用线程是没有资源被CPU调度到的，都是挂起的状态，自然Jedis的链接是长时间没有响应的。推断属于第二种场景，于是又去看了Redis和MySQL的负载情况（MySQL是正常的这里就不截图了）



可以很明显的看到，Redis除了链接数激增，响应时间和负载都是正常的，而QPS骤降，和连接数的上升，说明了在19点40这个点 请求一下变少了，但是连接数增加；验证了我们上述推断，应用疯狂的向连接池申请创建链接，但却很长时间未释放链接（因为被Ying用Hold住了）

回到Jedis异常的日志点，以上可以分析得出Jedis并不是问题的引起方，这是应用负载高了以后的一种报警形式，继续按一分钟的粒度往前推移，发现在3分钟内存在大量的如下日志：



这时又切换了别的Pod所在的容器的IP，并无类似的日志，于是猜测有没有可能是因为服务发现中心在获取服务可用列表的时候只发现了一台provider，导致所有的流量在某一个时间段都路由到了同一个服务实例上？如果是这样的话也

就能解释为什么26个Pod只有一个pod出问题导致整个商品服务都是不可用的了。带着这个问题又重新去看了应用的负载和监控：

可以看到出问题的pod和其他的Pod，CPU使用率，内存，网络IO流量，磁盘在出问题的时间点趋势完全是呈反比的，看上去就像所有的请求 突然从别的实例上都转移到了这个pod上。貌似也复核了我们之前的推断，有没有可能是服务发现的负载出了问题，可用服务实例列表的缺失造成的负载不均；从而导致单台应用无法承载整个集群的请求量?

这个时候带着疑问，又去查看了在出问题的点附近 是否存在大量的接口调用，或者导入导出等耗资源的操作

经过排查我们排除了导入导出操作造成的影响，同时可以较为明显的看到19点到问题发生的点并没有接口调用量的激增。

大体分析下来可能就是 服务发现的负载问题，导致了POD压力瞬间过载 从而影响整个服务。