

30 | 云计算时代，我们所说的弹性伸缩，弹的到底是什么？

2018-02-25 赵成

赵成的运维体系管理课

[进入课程 >](#)



讲述：黄洲君

时长 06:09 大小 2.82M



现在，我们经常听到的一些高大上的词汇，比如弹性伸缩、水平扩展和自动化扩缩容等等，你能否说一说，这些技术手段的主体是谁，也就是谁的水平扩展？弹性伸缩的是什么？同时，这些名词之间又有什么关系呢？

下面我们就从弹性伸缩入手一起来分析讨论。

弹性伸缩的主体是谁？

弹性伸缩，一说到这个词，我们可能很自然地会联想到资源的弹性伸缩，服务器的弹性伸缩，容量的弹性伸缩，应用的弹性伸缩以及业务的弹性伸缩等等。我想这些理解都没有错误，但是可以发现，当弹性伸缩这个动词前面增加了这么多不同的主语之后，我们一下子就不知道到底该做什么了。

其实有这样的困惑很正常。我们在讲标准化的时候就提到，**做运维和做架构的思路是相通的，我们碰到问题后，一定要找到问题的主体是什么，通过问题找主体，通过主体的特性制定问题的解决方案。**

对于运维，一定要准确识别出日常运维过程中不同的运维对象，然后再进一步去分析这个对象所对应的运维场景是什么，进而才是针对运维场景的分解和开发。

这里可以看到，弹性伸缩其实是一个运维场景，但是我们并没有定义这个场景的主体是谁，所以就会出现上面这些不同的主体。我们来假设以下几种主体。

服务器的弹性伸缩。针对这个场景，假设业务是运行在私有云或公有云上，那我们只要能够通过云平台的 API 申请和释放资源，申请时初始化我们的操作系统，释放时销毁资源就可以。不过，在私有云环境下，为了能够保障弹性，我们必须自己提前采购、上架、装机、配置然后交付资源，需要大量的准备工作，公有云上就可以省去这些步骤，可以即拿即用。

应用的弹性伸缩。这个场景下其实是默认包含第一步的，就是我们首先必须要拿到应用运行的服务器资源才可以，这一步做到了，下面就是应用的部署、启动以及服务上线接入流量。

业务的弹性伸缩。我们可以再进一步思考，通常一个业务可能会包括多个应用，所以为了保障整个业务容量充足，这个时候扩容单个的应用是没有意义的，所以这时要做的就是扩容多个应用，但是这里面就会有一个顺序问题，先扩哪个，后扩哪个，哪些又是可以同时扩容而不会影响业务正常运行的，再进一步，业务承载的服务能力提升后，那网络带宽、缓存、DB 等等这些基础设施需不需要也同时扩容呢？

到这里，这个问题就已经变得非常复杂了，而且已经不仅仅是弹性伸缩这个词的表面含义所能覆盖的了，因为这个问题确实很复杂，所以这里暂时不做详述。

好了，分析到这里，我们就可以看到针对不同的运维对象，弹性伸缩这个概念背后的含义是不一样的，所执行的动作以及制定的方案也是不一样的。通过上面的分析过程，我们在日常思考和工作开展中应该注意以下两点。

第一，一定是从实际问题出发，找到问题的主体，然后才是针对问题的解决方案。要反复问自己和团队，我们解决的问题是什么？解决的是谁的问题？切记，一定不要拿着解决方案来找问题，甚至是制造问题。

比如弹性伸缩这个概念，它就是解决方案，而不是问题本身，问题应该是：业务服务能力不足时，如何快速扩容？业务服务能力冗余时，如何释放资源，节省成本？按照这个思路来，我们自然就提炼出业务服务能力这个主体，面对的场景是快速的扩缩容，然后针对场景进一步细化和分解。

第二，如果问题处于初期，且是发散状态时，主体可能表现出很多个，这时我们一定要找到最本质的那一个，往往这个主体所涉及的运维场景就包括了其它主体的场景。

比如上面我们看到的业务的弹性伸缩，就包含了应用和服务器的弹性伸缩场景，它们只不过是子场景而已。从这个角度出发，我们的思路才能打开，方案才会全面。不然如果只是聚焦在服务器、应用、DB、网络等等这些非本质的主体上，提供出来的解决方案肯定也是片面的，而且经常会出现这样的情况，每个人都只从自己的角度出发说问题，始终无法达成一致。问题的根本还是我们没有一个共同讨论的基础，结果就是工作没少做，却始终没有解决问题。

弹性伸缩、水平扩展和自动化扩缩容，它们之间的关系是什么？

这个问题，我想已经不言自明了，找到它们的主体和要解决的问题，我们就会发现，其实它们之间的含义是一样的。

总结

今天我们以弹性伸缩为例，讨论了如何思考问题和分析问题。我们的讨论和分析归结到一点就是：**独立思考和分析的能力很重要，意识也很重要，切忌不可人云亦云随大流，反而迷失了工作的方向。**

现在业界各种技术上的 Buzzword（时髦词）层出不穷，让人目不暇接，但是仔细观察和思考，你会发现它们背后常常隐藏着很多共性的特点，一定要抓住它们背后所要解决的问题和本质，这样也就不会乱花渐欲迷人眼了。

而且通过这样的分析，我们会更容易发现工作中还需完善的地方，从而引导我们聚焦到实际问题中来，而不是浮于表面，把一些高大上的词汇挂在嘴边，却不见效果。

关于今天我们讨论的主题，你还有哪些想法和心得，欢迎你留言与我讨论。

如果今天的内容对你有帮助，也欢迎你分享给身边的朋友，我们下期见！

赵成的运维体系管理课

带你直击运维的本质

赵成

美丽联合集团技术
服务经理



新版升级：点击「 请朋友读」，10位好友免费读，邀请订阅更有**现金**奖励。

© 版权归极客邦科技所有，未经许可不得传播售卖。页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇 29 | 量体裁衣方得最优解：聊聊页面静态化架构和二级CDN建设

下一篇 31 | 极端业务场景下，我们应该如何做好稳定性保障？

精选留言 (3)

 写留言



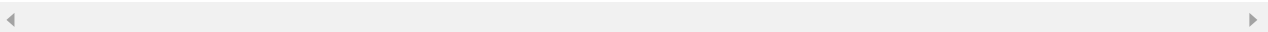
given

2018-03-04

 1

您好，用容器做缩容，如果要减少的容器刚好在处理业务请求，这种情况如何处理？

作者回复：应用层面，服务要先下线，比如从lvs或nginx负载均衡器上下线，如果是服务化应用，则要从配置中心下线。下线完成，然后再做容器缩容





人生海海

2018-09-26



目前尴尬点是扩容速度，自动扩容感知突发事件，但扩容需要10多分钟。当扩容完成后，

突发热点事件流量已经回落、有种没用到刀刃上的感觉。



james

2018-03-29



那另一个问题来了，下线的时候有业务咋办，客户请求了就没有响应了啊，不过得刷新和重试

作者回复: 先下服务，从负载均衡上或者配置中心摘掉，没有流量进来了，且默认静默一段时间，再停服务，再做后面的操作。

