

APISEVEN 和 Kong EE 的性能评测 -- GigaOm

高性能API管理测试

产品评估：API7和Kong企业版



[1 - 摘要](#)3

[2 - 云上的API 管理](#)5

[API7](#)6

[图1. API7技术架构](#)7

[Kong 企业版](#)7

[3 - GigaOm API负载测试设置](#)9

[API 压力测试](#)9

[测试环境](#)10

[单节点](#)10

[环境清单](#)10

[软件版本信息](#)11

[4 - 测试结果](#)12

[图2. 空转时的压力测试 API 的基线延迟](#)12

[图3. API7与Kong EE在20,000rps时的对比](#)13

[图4. API7与Kong EE在10,000 rps时的 JWT 对比。](#) 13

[图5. API7与Kong EE在10,000 rps时的1,000条路由的对比](#)14

[5 - 结论](#)15

[6 - 附录](#)16

[7 - 关于GigaOm](#)19

[8 - 关于API7](#)20

1 - 摘要

本报告重点介绍了部署在云上的API管理平台。云让企业通过微服务快速地构建差异和创新，在几分钟内就能完成API节点的克隆和扩展。与本地部署相比，云有良好的扩展性，能更快地进行服务器部署和应用程序开发，且能降低计算成本的开销。

更重要的是，许多组织也依赖API和微服务来实现高性能和可用性。在本文中，我们将“**高性能**”定义为**每秒负载超过1000个交易且在整个API环境中最大延迟小于30毫秒**。对公司而言，对性能的需求和对管理的需求一样，因为公司依靠API交易速率来跟上业务发展速度。

API管理解决方案**不能**成为性能瓶颈。许多公司都在寻找跨多个API端点的负载均衡和高交易量吞吐的解决方案。如果业务每秒有1000个交易，一个月内就会有30亿次 API 调用。拥有大流量的公司通常每月API调用次数超过100亿次。因此，在选择API管理解决方案时，性能是一个关键因素。

在本文中，我们展示了使用2个全生命周期API管理平台完成的性能测试结果：**API7 和 Kong 企业版 (Kong EE)**。

在我们的单节点设置中，API7所有的压力测试结果都优于Kong EE。在每秒10,000个请求的情况下，99.99 % 的情况API7的延迟比Kong EE低14倍。API7和Kong EE二者百分比越高延迟差异越明显。在我们所有的测试中，最大延迟差异体现得最明显的是达到 99.9% 和99.99% 的请求时。

云上测试软硬件是非常具有挑战性的。在可用性、虚拟机处理器、内存、最佳输入/输出的存储、网络延迟、软件和操作系统版本以及负载这些方面的配置可能会有利于其中一方。更具挑战性的是测试完全托管的服务产品，我们不知道这些产品的底层配置（处理能力、内存、网络等）是什么样的。我们的测试展示了一小部分潜在配置和负载。

作为报告的发起人，API7选择了默认的API网关——该解决方案没有根据性能进行调整或更改。Kong 企业版选择了在CPU和内存配置方面最接近的配置。

我们把公平性的问题留给读者来决定。作为读者，我们强烈建议你抛开产品的营销信息，自己去辨别什么是有价值的。我们希望这份报告内容丰富，有助于你做选择，了解不同产品的细微差别。

我们在报告中提供了足够的信息去复现测试。我们建议你建立自己的测试环境。

2 - 云上的API 管理

应用编程接口，或称API，是现代信息技术中无处不在的方法和通信标准。大公司已经使用 API 来传递数据，将各个系统连接起来，并把数据变成一种服务。API已经开始用可复用、解耦的微服务来取代传统冗余的方式。这使企业能在不同的系统和应用程序之间共享数据，而不会产生技术债务。

API 和微服务也给公司提供了一个支持应用程序模块化和治理标准的机会。此外，也扩大了数据交换范围，从移动技术和智能设备到物联网，因为企业可以在任何设备上共享数据。

由于 API 的激增，公司需要管理内外部所依赖的众多服务。API在协议、方法、授权/认证和使用方面差异很大。此外，IT部门需要对 API 进行精细化控制，限制速率、调用次数，制定策略和用户身份识别来确保高可用性，防止滥用和安全漏洞。公开 API 为许多合作伙伴打开了大门，他们可以在不了解底层技术的情况下共同创建和扩展核心平台。

根据企业需求和底层架构，云上管理 API 有很大差异。所以为了便于讨论，我们把云上的 API 管理划分成两种部署方式：

混合云 - 混合云 API 管理方案允许企业部署、配置在自己选的云上。例如，可以选择亚马逊云（AWS EC2）或者 Azure 虚拟机来安装 API 组件（管理工具和端点），能自由配置 CPU 内核、内存、存储和网络，能在操作系统层面调整环境。这种方法的优点是能在企业内部、私有云、公共云或三者混合的情况下部署。

公有云 - 完全托管的 API 管理解决方案是指开箱即用的服务产品。部署通常又快又方便，但缺乏精细配置和控制的能力。通常情况下，底层架构是模糊的，用户不知道内部的计算能力。与 "自建 " 的一个主要区别是，完全托管的解决方案通常要固定在一个特定的公有云上，如AWS、Azure 或 Google。在本次测试中，我们没有测试完全托管的云计算产品。

API 管理供应商提供自建或完全托管的云部署，少数的供应商同时提供这两种方式。虽然有很多平台可以提供管理 API 的功能，但我们对高性能场景更感兴趣。同样，为了本报告的测试目标，我们将**"高性能"** 定义为**每秒负载超过 1000 个交易，并在后端 API 和微服务中的最大延迟小于 30 毫秒**。

本文的目的是探讨高性能要求下的 API 管理产品。

API7

API7 是建立在 Apache APISIX 的基础上，由深圳支流科技有限公司维护。Apache APISIX是一个动态、实时、高性能的 API 网关。APISIX 提供丰富的流量管理功能，如负载均衡、动态上游、灰度发布、熔断、认证、可观测等。你可以使用 Apache APISIX 来处理传统的南北流量，以及服务之间的东西向流量。它也可以作为一个 k8s ingress controller。

API7 是 APISIX 的企业版，其功能包括多集群管理、多分区、权限管理、版本管理、审计、统计报告等，满足企业用户的核心需求。

以下是API7基于Apache APISIX的技术架构（**图1**）。

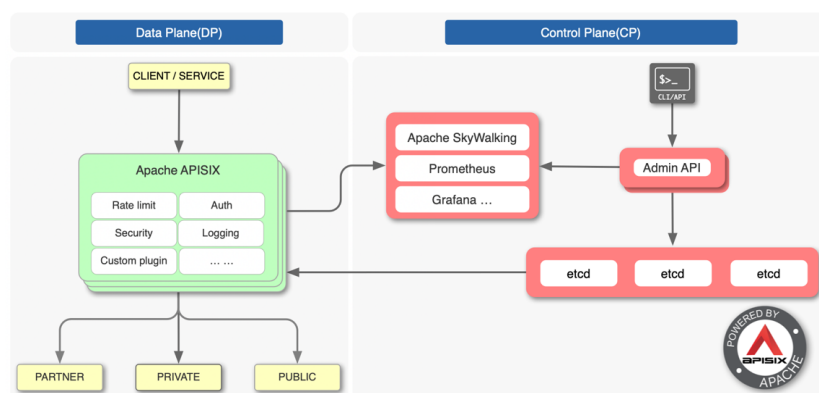


图1. API7技术架构

API7是混合云的部署模式，可用于多云、企业内部和混合环境。

Kong 企业版

Kong 的原名是 Mashape，在 API 平台发布的时候改名为 Kong。Kong 是在 2015 年开源的，功能丰富。Kong 利用 NGINX 的底层 API 网关，并增加了开源插件支持、负载平衡和服务发现等功能。Kong 企业版具有扩展的功能，如管理看板、可定制的开发者门户、安全插件、metrics 和 7×24h 的服务支持。2019 年，Kong 发布了一个完全托管的云产品，Kong Cloud。我们没有测试 Kong Cloud。

Kong 企业版既可以部署在云上，也可以部署在企业内。基于Debian和RedHat 的包管理器（Yum 和 Apt-Get），在他们的仓库里有Kong，而且还有Docker 和 CloudFormation 选项。

Kong 可以作为一个单节点运行，也可以作为多节点集群。在集群中，负载均衡器（如开源的 NGINX）被用在边缘节点为客户端提供单个地址，并使用选定的策略（例如循环或加权）在 Kong 节点之间分发请求。横向扩展 Kong 很简单。Kong 是无状态的，所以向集群添加新节点需要将新节点指向外部数据库（PostgreSQL 或 Cassandra），它就可以处理 API 请求和响应所需的配置、安全、服务、路由和消费者信息，包括前面负载均衡器的 IP 地址或 fully qualified domain name（FQDN）。

Kong 有一个插件生态系统（Kong Hub），支持开源和企业版的插件，如 LDAP 认证、CORS、动态 SSL、AWS Lambda、Syslog 等。因为它是基于 NGINX 的，所以 Kong 允许用户使用 LuaJIT 创建自己的插件。

3 - GigaOm API负载测试设置

API 压力测试

GigaOm API Workload Field Test 是一个简单的压力测试，用一连串相同的 GET 请求 API 或 API 管理节点（或管理节点集群前的负载均衡器），每秒的请求数不变。

为了进行压力测试，我们使用了 HTTP 压力测试工具 WRK2¹，这是 Github 上一个免费的压力测试套件。WRK2 工具返回每个测试的延迟分布和状态代码摘要。压测工具将延迟测量发出单个 API 请求和收到API 响应的时间间隔。因此，如果我们在 60 秒内每秒发出 10,000 个请求，压测工具就会记录 600,000 个延迟值。我们用这些数据来统计测试结果。

该测试还需要一个能够监听和响应请求的后端 API（部署在NGINX OSS上），我们的后端 API 监听一个GET 请求，例如：

```
http://localhost:8000/test_uri
```

然后，API 会响应一个由 1024 个伪随机 Unicode 字符组成的字符串，如：

```
taZ3psgHkQojalo...
```

我们在附录中详细地记录了使用的后端API。

Kong Enterprise 2.2 是从 AWS 市场²获得并安装的。API7 提供了一个API7 v.1.7 tarball。

对于 API7 和 Kong，我们用了速率限制插件，并在给定的时间内把可以发出的 HTTP 请求的数设置为 2,999,999。

1 <https://github.com/giltene/wrk2>

2 <https://aws.amazon.com/marketplace/pp/B08P51PKC1>

每轮测试中，我们对平台、配置和请求率都重复进行了五次。我们从每秒10,000 个请求（rps）的压测开始，扩大到20,000 rps。我们每个测试都进行了60秒。我们捕获了50%、90%、95%、99%、99.9%和99.99% 时的请求响应延迟，以及在测试期间看到的最大延迟。我们记录了导致中位数最大值的测试。错误状态码包括 HTTP 状态码 429 Too Many Requests 或任何5xx 码，最常见的是 500

Internal Server Error。100% 的成功率（如果图表上没有错误标记）意味着所有请求都返回了200 OK 状态码。

此外，我们测试了在相同配置下启用 JSON Web令牌（JWT）身份验证，并使用 JWT 凭据对每个请求进行身份验证。我们使用10,000 rps的压测条件进行测试。

在第四个测试中，我们生成了1000个相同的路由，并以10,000 rps 在所有路由上平均分配请求。

测试结果记录在报告的第四节。

测试环境

单节点

名称	数量	EC2
攻击节点	1	c4.4xlarge
API网关数据平面	1	c4.2xlarge
API网关控制平面 + 存储: Kong: PostgreSQL API7: etcd	1	c4.2xlarge
上游服务器	1	c5n.2xlarge

环境清单

名称	值
Ping延迟(两个服务器之间)	小于 0.150 ms
操作系统 (API7)	CentOS 7.8
操作系统 (Kong EE)	Amazon Linux 2
速率限制(API7)	每秒2,999,999个请求
速率限制(Kong)	每秒2,999,999个请求
每个进程的文件描述符 (ulimit -n)	1,024,000

软件版本信息

名称	版本
API 网关	API7 1.7
API 网关	Kong Enterprise 2.2.0.0-beta1
上游	NGINX 1.14.0
测试工具	WRK2 4.0.0

不同配置的结果可能有所不同，再次建议创建自己的压测条件。

4 - 测试结果

这一节分析了上文所述 API 压力测试中各种情况下60秒内的延迟，单位是毫秒。延迟越低越好，这意味着API响应速度越快。此外，还列出了50%、90%、95%、99%、99.9% 和 99.99% 时的响应时间以及最大延迟。这些数值是服务水平协议（SLA）的重要数值，也是用户可能经历的最慢响应时间。

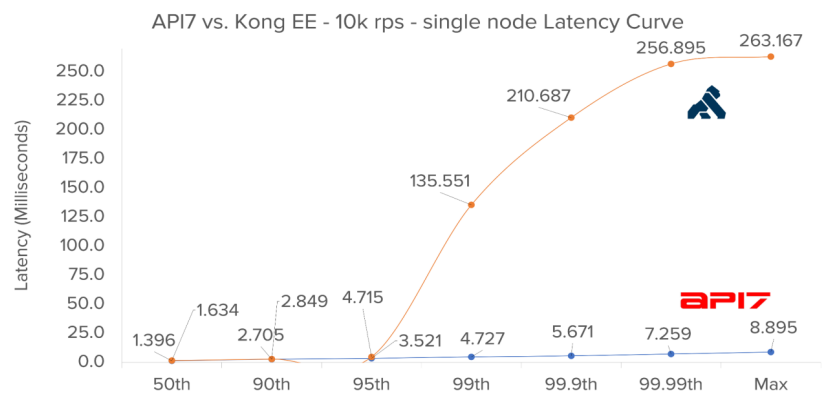


图2. 空转时的压力测试 API 的基线延迟

API7 在单节点情况下的所有压测结果都优于 Kong EE。首先，我们强调了在单节点上以10,000 rps来进行测试，**如图2**所示。尽管在达到99% 之前差异很小，但延迟的差异随后呈指数级增长。对比这张图和后面 20000 rps 时延迟的图（**图3**），你可以看到50%、90%、95%、99%、99.9% 和 99.99% 的延迟以及测试运行期间的最大延迟。在达到99.99 %时，API7 的延迟比 Kong EE 低97%。

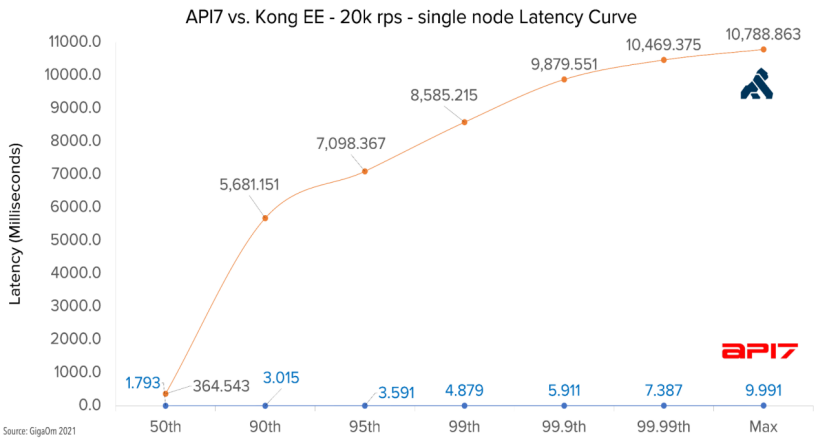


图3. API7与Kong EE在20,000rps时的对比

注：Kong EE有1876个，或9.4% 的非2xx或3xx响应（失败）。

我们使用 JWT 身份验证来测试相同的配置下的延迟，**如图4**所示。

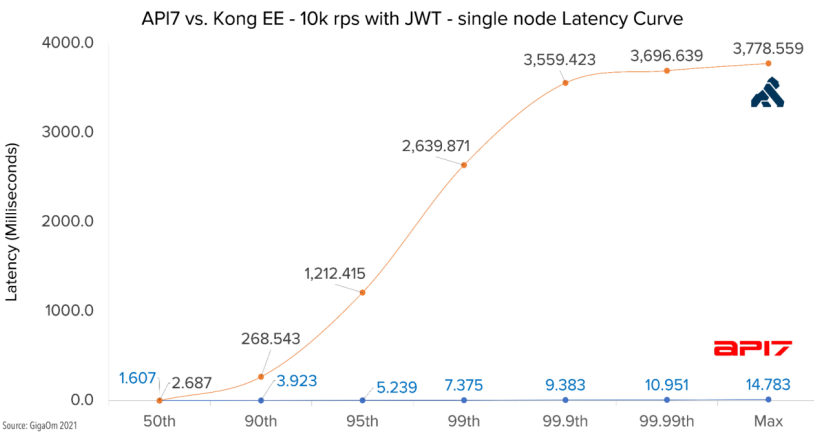


图4. API7 与 Kong EE 在 10,000 rps 时的 JWT 对比。

注：Kong EE有606个，或6% 的非2xx或3xx响应（失败）。

第三，我们测试了1000个相同的路由（没有JWT），如图5所示。

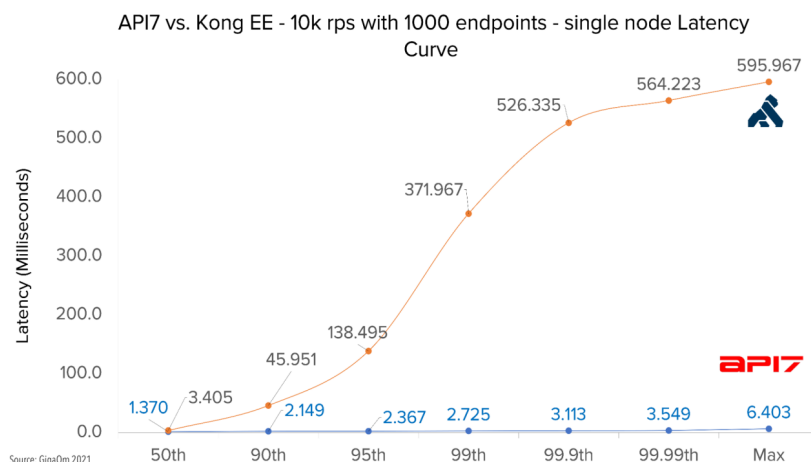


图5. API7与Kong EE在10,000 rps时的1,000条路由的对比

5 - 结论

本报告概述了GigaOm API压力测试结果。

在单节点设置中，API7 在所有压测情况下的延迟都优于 Kong EE。在每秒10,000个请求的情况下，API7在 99.99 % 时的延迟比 Kong EE 低 14 倍。API7 和 Kong EE二者百分比越高延迟差异越明显。在我们所有的测试中，延迟的差异在 99.9% 和 99.99 % 以及最大延迟上都很明显。

在每秒20,000个请求时，延迟差异发生在90 %时，Kong EE延迟5,681毫秒，而 API7仅延迟3毫秒。

JWT认证自然会增加延迟，但并没有改变 Kong 在评测中的表现，例如，Kong EE 的最大延迟是3,778 毫秒，而API7仅14毫秒。

使用1,000个端点也没有改变表现情况。你可以预料到在每秒10,000个请求和 1,000个端点的情况下，API7的延迟只有几毫秒（个位数）。

在这个测试中，使用特定的负载和特定的配置，API的请求在API7上的延迟最低，吞吐量最高，而不是 Kong EE。

注意，随着产品的发展或内部测试指向不同的配置，所有平台上的优化都是可能的。

6 - 附录

本测试中使用的后端API使用以下配置部署在NGINX上：

该程序的工作原理是由NGINX监听GET请求，例如：

```
GET http://fqdn-or-ip-address:8080/
```

API响应来自 /dev/urandom的1024个伪随机Unicode字符，例如：

```
taZ3psgHkQnwohs...
```

以下是后端API的NGINX配置，你可以自由地使用和修改。GigaOm 对超出本测试或报告范围之外的使用不做任何保证或索赔。

```
master_process on;
worker_processes auto;
worker_cpu_affinity auto;
error_log /var/log/nginx/error.log;
pid /run/nginx.pid;
worker_rlimit_nofile 20480;

events {
    accept_mutex off;
    worker_connections 10620;
}

http {
    access_log off;
    server_tokens off;
```

```

keepalive_requests 10000;

tcp_nodelay on;


server {
    listen 1980 reuseport;
    location / {
        return 200 "
jduukjyvpeamuzmyxwoqchstlbcqdrurkvejsmsulrjeztqnvvpaygrtfizfevcxnlrikzrbcjxwlhwmygdhslcauoxoljcs
bnbuefblbaddnlxioxujckovkzrjijgwzkycsolwdasnmanrjiwfenhrxauhzzgllbijpvbnvpquvedmniqlwflgaaiocen
inrlsrkngpsfhhagudbiphxrvbwidxsfwxkzcbdqkdgnwmvtlkeuddebdlrifsgccrocknvypkhepgwhsmskuymkz
gbqiocnrlggdsodzqztzsbcbkuwsdscexovpbwgdmnkwxgfdxltbvcxmvpeyursouqfmqiuuzfzvonikxqqkqdvmp
mnnvzeqencvjqpnmclbnmsiajdwclwglrrjlbzavttnafyhfsbdbtecjrblankkivmqhqcxxfwyqblirxuxhtzrytoanvt
qeliujysfsvjbtptwcrbruclamdtzgpjxpdvyhxckqfsdzbuuyhjwsjpiviyyhjkllzknrywuqogrppxkamifiukuexpsieaxz
vxwcvmpxuohzpmhrjdadrphdkpfvosfhhbskdgemvvzuvvgksxclygwrazjsqfetpsagynfvwkqrqvzdexalnufibftc
raznitxnajvutmxzabzgxhoniicafdlhgmuagbhstwfuxlhtuwkuemsyaxhkqsrxomnsylybxnehpeklqpawfpawflpt
bmsscjaplnzxfimljzufacxwnkqrfgyvpvjoqapmnfleihooqcuoazopsiprflfybvpbzwhxqehnnkuwemijosnclupjaxhj
otmkinqdxaijzvvyvlnwswtyjgxqrzdxhywoshxsaxwcpunyamcsphyptcparakjoeakigwgtncviswvugtdmeyclp
xtrnshzdynsvyamknmmmmohsibafuqnweonfzvwpjjqcmibtmuycgaygiarqqtketpovfxilhsvkouxbe";
    }
}

```

The following is the NGINX configuration for the load balancer:

```

master_process on;

worker_processes auto;

worker_cpu_affinity auto;

error_log /var/log/nginx/error.log;

pid /run/nginx.pid;

worker_rlimit_nofile 20480;


events {
    accept_mutex off;
    worker_connections 10620;
}


http {
    upstream backend {
        server 172.31.7.42:8000;

```

```
server 172.31.9.201:8000;
}
server {
    listen 80;
    location / {
        proxy_pass http://backend;
    }
}
}
```

7 - 关于GigaOm

GigaOm为IT企业、商业提供技术、运营和商业建议。GigaOm与企业领导人、CIO和技术公司合作，为业务现代化转型提供实用、可操作、有战略性、长远的建议。GigaOm的建议使企业能够在日益复杂的商业环境中保持竞争，这需要 GigaOm对不断变化的客户需求有充分的了解。

GigaOm与企业合作，使用成熟的研究方法来避开陷阱和错误，规避风险和力求创新。研究方法包括但不限于采用调查、使用案例、访谈、投资回报率/TCO、市场前景、战略趋势和技术评测。我们的分析员拥有 20 多年的经验，为早期采用者和企业客户提供建议。

GigaOm 保持中立的视角，这种视角下，GigaOm与敬业、忠诚的用户联系在一起。

8 - 关于 API7

API7是由深圳支流科技有限公司的产品。深圳支流科技是一家提供 API 处理和分析的开源基础设施软件公司，拥有API网关、k8s入口控制器和服务网格等微服务和实时流量处理的产品和解决方案。我们致力于为全球企业管理和可视化业务关键流量，如 API 和微服务，通过大数据和人工智能（AI）加速业务决策并推动数字化转型。

API7 的底层是APISIX，它是由深圳支流科技捐赠给Apache软件基金会的。它是新一代的云原生API网关，提供丰富的流量管理功能，如负载均衡、动态上游、灰度发布、服务熔断、身份验证、可观察性等。

API7 包括多集群管理、多工作分区、权限管理、版本管理、审计、统计报告等企业产品，能够满足企业用户的核心需求。

全球已有数百家企业用户在使用Apache APISIX处理核心业务流量，包括金融、互联网、制造、零售、运营商和其他许多行业。了解更多关于API7的信息，请访问：<https://www.apiseven.com/en>。