# lab9 - 应用程序-自动缩放功能

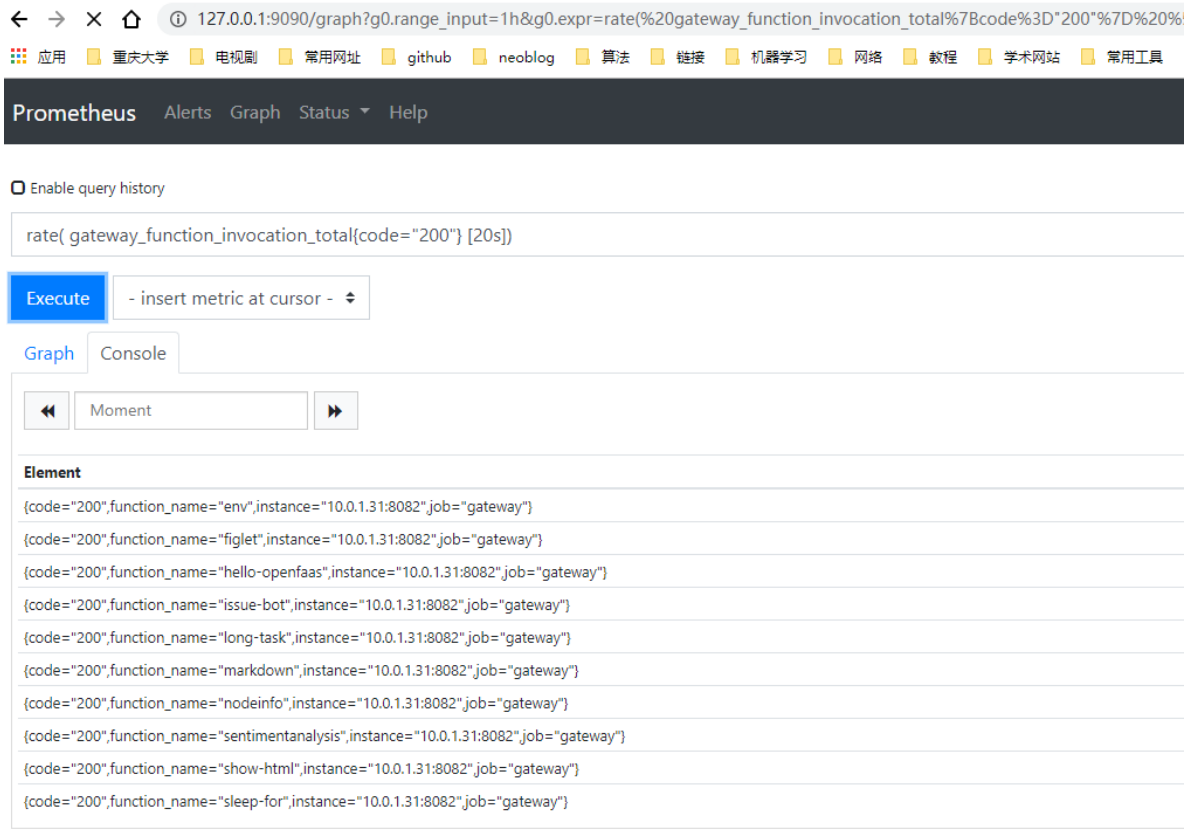
在实验9中，我学习了如何自动缩放OpenFaaS。具体来说，它具有与Prometheus的Alertmanager链接的机制。

开箱即用地配置OpenFaaS，使其将根据 `request per seconds` 通过Prometheus测量的度量标准进行自动缩放。当流量通过API网关时会捕获此度量。如果 `request per seconds` 超过了定义的阈值，则AlertManager将触发。应将此阈值重新配置为适合生产使用的级别。

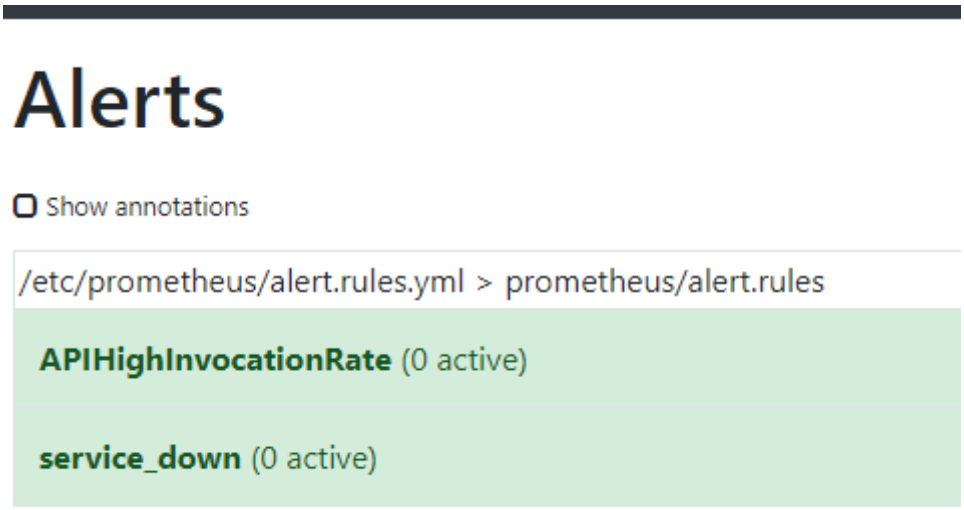
其实就是自动调整负载。

打开普罗米修斯并添加一个图：

```
在普罗米修斯上执行这个，rate( gateway_function_invocation_total{code="200"} [20s])
```




http://127.0.0.1:9090/alerts 可以查看何时超过每秒请求的阈值。

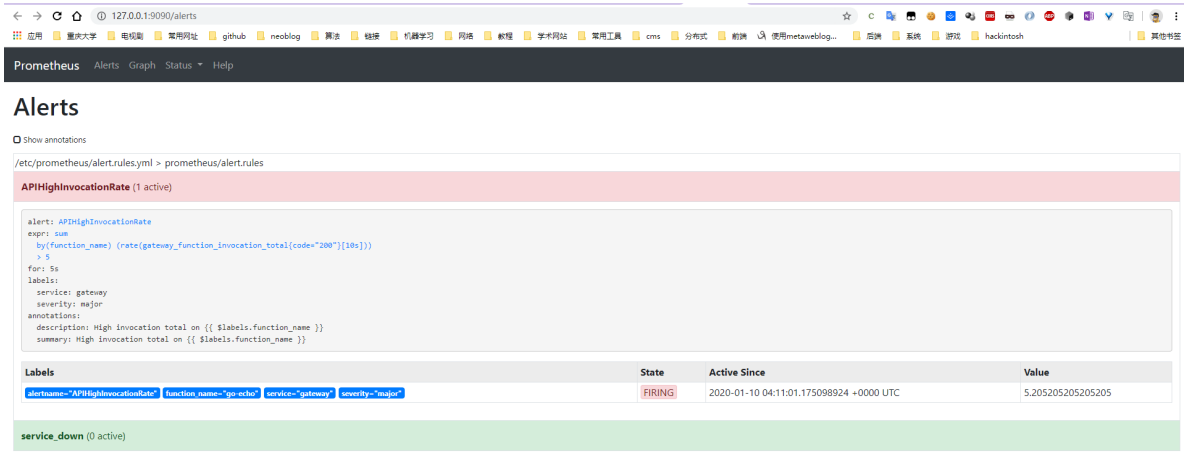



## 触发警告测试

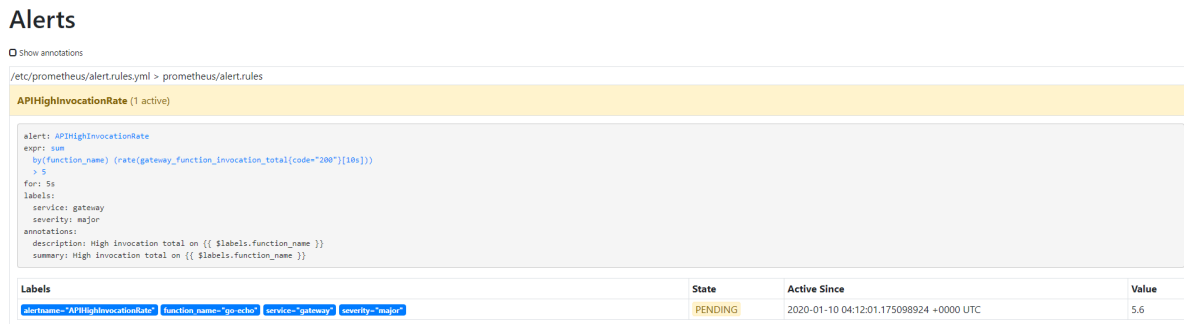
```
$ git clone https://github.com/alexellis/echo-fn \
 && cd echo-fn \
 && faas-cli template store pull golang-http \
 && faas-cli deploy \
  --label com.openfaas.scale.max=10 \
  --label com.openfaas.scale.min=1
```


```
$ for i in {0..10000};
do
   echo -n "Post $i" | faas-cli invoke go-echo && echo;
done;
```


可以看到go-echo的调用次数一直增加。




一段时间后出现了高负载警告




自动缩放后，状态更改为pending，然后警告消失。