

# k8s 部署Prometheus

---

## 简介

### 1 部署Prometheus-server

- 1.1 创建名称空间和ROLE
- 1.2 创建Prometheus外部配置文件
- 1.3 创建prometheus deployment
- 1.4 通过Endpoint 对外提供服务

### 2 部署exporter

- 2.1 创建node-exporter配置文件
- 2.2 查看应用状态
- 2.3 查看node-exporter 监控指标
- 2.4 部署kube-state-metrics

### 3 部署alertmanager

- 3.1 修改配置告警收件人
- 3.2 配置deployment
- 3.3 配置钉钉webhook-deployment(可选)
- 3.4 prometheus.rules 告警规则

## 结尾

## 简介

本文档会指导你如何在kubernetes集群上部署prometheus并使用其监控你的集群状态，prometheus是高度可扩展的开源架构，它提供了对kubernetes开箱即用的监控功能，解析prometheus被分为了几个部分，如果你想了解更多内容，可以参考<https://prometheus.io/>

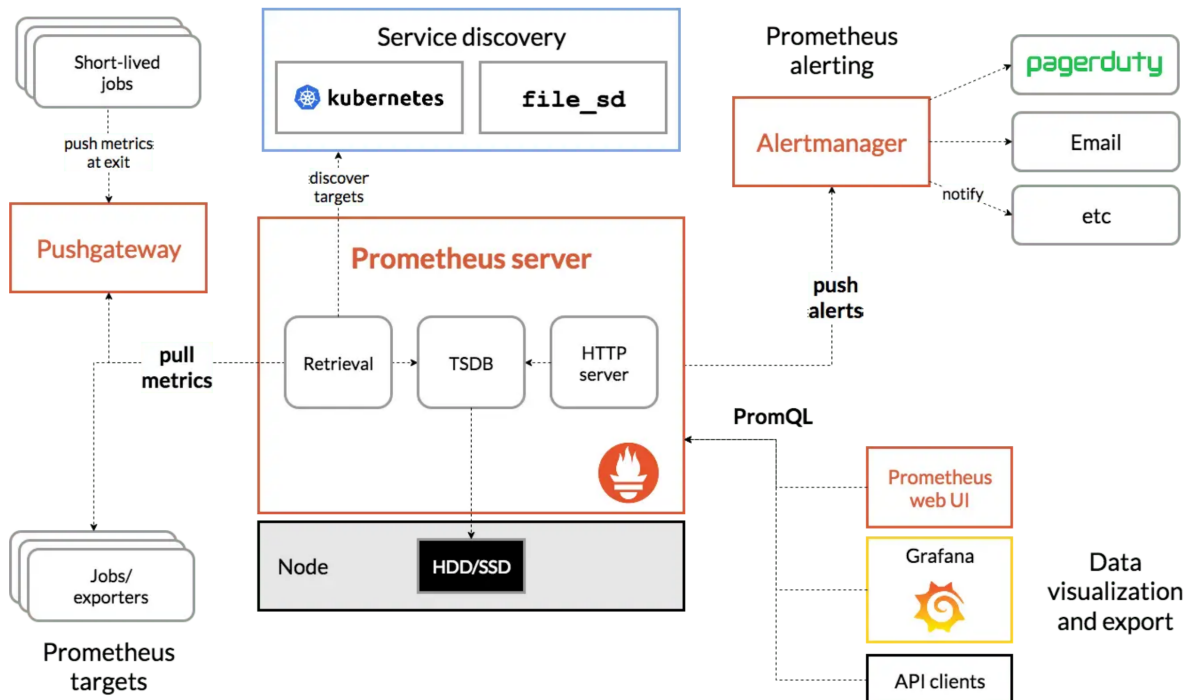
**1 Metric collection :** Prometheus 使用http协议pull 模式拉取监控指标，对于无法直接拉取的指标，prometheus采用push-gateway 直接推送监控数据到prometheus监控端

**2 Metric Endpoint:** 对于提供监控数据的服务，通常使用/metrics 暴露监控数据给server端定期拉取

**3 PromQL:** Prometheus通过PromQL查询数据，你可以直接在Prometheus的监控面板上直接执行语句查询指标，或者通过grafana调整PromQL图形化展示数据

**4 Prometheus Exporters:** Exporter库转换来自第三方的服务指标，使其可以通过PromQL语句抓取，官方提供了很多的Exporter库，比较常见的Exporter使用方式就是，node-Exporter，提取os层的服务指标

**5 TSDB:** Prometheus 使用可以有效的使用TSDB存储监控数据，默认的方式是直接存储在本地的文件系统内，但是你也可以修改配置，将监控指标存储在Prometheus TSDB中



本文档假设你已经部署了kubernetes集群，并且熟悉kubectl，如果你还没有配置过kubernetes集群，你可以通过这个链接快速部署一套minikube <https://devopscube.com/kubernetes-minikube-tutorial/>

本教程提供的所有yaml文件都已经放置在github，你可以通过clone的方式获取

```
1 git clone https://github.com/lvtujingji/kubernetes-prometheus.git
```

## 1 部署Prometheus-server

### 1.1 创建名称空间和ROLE

```
1 kubectl create namespace monitoring
```

Prometheus使用kubernetes API获取node、pod、deployment的Metric，因此我们必须对通过RBAC的方式对Prometheus进行授权，需要注意的是这个role绑定在monitoring下的default

ServiceAccount, 如果你使用的是其他SA, 需要调整你的绑定用户

Plain Text

```
1  apiVersion: rbac.authorization.k8s.io/v1
2  kind: ClusterRole
3  metadata:
4    name: prometheus
5  rules:
6    - apiGroups: ["" ]
7      resources:
8        - nodes
9        - nodes/proxy
10     - services
11     - endpoints
12     - pods
13     verbs: ["get", "list", "watch"]
14   - apiGroups:
15     - extensions
16     resources:
17     - ingresses
18     verbs: ["get", "list", "watch"]
19   - nonResourceURLs: ["/metrics"]
20     verbs: ["get"]
21 ---
22 apiVersion: rbac.authorization.k8s.io/v1
23 kind: ClusterRoleBinding
24 metadata:
25   name: prometheus
26 roleRef:
27   apiGroup: rbac.authorization.k8s.io
28   kind: ClusterRole
29   name: prometheus
30 subjects:
31 - kind: ServiceAccount
32   name: default
33   namespace: monitoring
```

## 1.2 创建Prometheus外部配置文件

通过configmap去调整服务的配置文件, 你不需要每次修改配置文件的时候都重新构建镜像, 只需要重新应用下configmap, 然后重启下pod就可以生效, 这个configmap将配置文件挂载在/etc/prometheus路径下, 分别被命名为 prometheus.yaml 和 prometheus.rules

```
1 kubectl create -f config-map.yaml
```

所有的监控报警规则被配置在prometheus.rules文件内这个会在altermanager部分介绍

所有的发现信息和存储信息被配置在prometheus.yaml文件内,包含了pod和service的动态发现

```
1  apiVersion: v1
2  kind: ConfigMap
3  metadata:
4    name: prometheus-config
5    labels:
6      name: prometheus-config
7    namespace: monitoring
8  data:
9    prometheus.rules: |-
10     groups:
11     - name: devopscube demo alert
12       rules:
13       - alert: PrometheusJobMissing
14         expr: absent(up{kubernetes_name="prometheus-service"})
15         for: 0m
16         labels:
17           severity: warning
18         annotations:
19           summary: Prometheus job missing (instance {{ $labels.instance
20 }})
21           description: "A Prometheus job has disappeared\n  VALUE = {{ $v
22 alue }}\n  LABELS = {{ $labels }}"
23
24       - alert: PrometheusTargetMissing
25         expr: up == 0
26         for: 0m
27         labels:
28           severity: critical
29         annotations:
30           summary: Prometheus target missing (instance {{ $labels.instanc
31 e }})
32           description: "A Prometheus target has disappeared. An exporter
33 might be crashed.\n  VALUE = {{ $value }}\n  LABELS = {{ $labels }}"
34
35       - alert: PrometheusTooManyRestarts
36         expr: changes({kubernetes_name=~"prometheus|alertmanager|node-exp
37 orter|grafana"}[15m]) > 2
38         for: 0m
39         labels:
40           severity: warning
41         annotations:
42           summary: Prometheus too many restarts (instance {{ $labels.inst
43 ance }})
44           description: "Prometheus has restarted more than twice in the l
45 ast 15 minutes. It might be crashlooping.\n  VALUE = {{ $value }}\n  LABE
```

```

39 LS = {{ $labels }}"
40
41     - alert: HostOutOfMemory
42       expr: (node_memory_MemAvailable_bytes / node_memory_MemTotal_byte
43 s * 100 < 10) * on(instance) group_left (nodename) node_uname_info{nodena
44 me=~".+"}
45       for: 2m
46       labels:
47         severity: warning
48         annotations:
49           summary: Host out of memory (instance {{ $labels.instance }})
50           description: "Node memory is filling up (< 10% left)\n  VALUE
51 = {{ $value }}\n  LABELS = {{ $labels }}"
52
53     - alert: HostMemoryUnderMemoryPressure
54       expr: (rate(node_vmstat_pgmajfault[1m]) > 1000) * on(instance) gr
55 oup_left (nodename) node_uname_info{nodename=~".+"}
56       for: 2m
57       labels:
58         severity: warning
59         annotations:
60           summary: Host memory under memory pressure (instance {{ $label
61 s.instance }})
62           description: "The node is under heavy memory pressure. High rat
63 e of major page faults\n  VALUE = {{ $value }}\n  LABELS = {{ $labels }}"
64
65     - alert: HostUnusualNetworkThroughputIn
66       expr: (sum by (instance) (rate(node_network_receive_bytes_total[2
67 m])) / 1024 / 1024 > 100) * on(instance) group_left (nodename) node_uname
68 _info{nodename=~".+"}
69       for: 5m
70       labels:
71         severity: warning
72         annotations:
73           summary: Host unusual network throughput in (instance {{ $label
74 s.instance }})
75           description: "Host network interfaces are probably receiving too
76 much data (> 100 MB/s)\n  VALUE = {{ $value }}\n  LABELS = {{ $labels
77 }}"
78
79     - alert: HostUnusualNetworkThroughputOut
80       expr: (sum by (instance) (rate(node_network_transmit_bytes_total
81 [2m])) / 1024 / 1024 > 100) * on(instance) group_left (nodename) node_una
82 me_info{nodename=~".+"}
83       for: 5m
84       labels:
85         severity: warning
86         annotations:

```

```

73         summary: Host unusual network throughput out (instance {{ $label
74 ls.instance }})
75         description: "Host network interfaces are probably sending too
76 much data (> 100 MB/s)\n  VALUE = {{ $value }}\n  LABELS = {{ $labels }}"
77         - alert: HostUnusualDiskReadRate
78           expr: (sum by (instance) (rate(node_disk_read_bytes_total[2m]))
79 / 1024 / 1024 > 50) * on(instance) group_left (nodename) node_uname_info
80 {nodename=~".+"}
81           for: 5m
82           labels:
83             severity: warning
84             annotations:
85               summary: Host unusual disk read rate (instance {{ $labels.insta
86 nce }})
87               description: "Disk is probably reading too much data (> 50 MB/
88 s)\n  VALUE = {{ $value }}\n  LABELS = {{ $labels }}"
89             - alert: HostUnusualDiskWriteRate
90               expr: (sum by (instance) (rate(node_disk_written_bytes_total[2
91 m])) / 1024 / 1024 > 50) * on(instance) group_left (nodename) node_uname_
92 info{nodename=~".+"}
93               for: 2m
94               labels:
95                 severity: warning
96                 annotations:
97                   summary: Host unusual disk write rate (instance {{ $labels.inst
98 ance }})
99                   description: "Disk is probably writing too much data (> 50 MB/
100 s)\n  VALUE = {{ $value }}\n  LABELS = {{ $labels }}"
101                 - alert: HostOutOfDiskSpace
102                   expr: ((node_filesystem_avail_bytes * 100) / node_filesystem_size
103 _bytes < 10 and ON (instance, device, mountpoint) node_filesystem_readonly
104 == 0) * on(instance) group_left (nodename) node_uname_info{nodename=~".
105 +"}
106                   for: 2m
107                   labels:
108                     severity: warning
109                     annotations:
110                       summary: Host out of disk space (instance {{ $labels.instance
111 }})
112                       description: "Disk is almost full (< 10% left)\n  VALUE = {{ $v
113 alue }}\n  LABELS = {{ $labels }}"
114                   - alert: HostHighCpuLoad
115                     expr: (sum by (instance) (avg by (mode, instance) (rate(node_cpu_
116 seconds_total{mode!="idle"}[2m]))) > 0.85) * on(instance) group_left (nod

```

```

105     ename) node_uname_info{nodename=~".+"}
106         for: 10m
107         labels:
108             severity: warning
109         annotations:
110             summary: Host high CPU load (instance {{ $labels.instance }})
111             description: "CPU load is > 80%\n  VALUE = {{ $value }}\n  LABELS = {{ $labels }}"
112
113     - alert: HostOomKillDetected
114       expr: (increase(node_vmstat_oom_kill[1m]) > 0) * on(instance) group_left (nodename) node_uname_info{nodename=~".+"}
115       for: 0m
116       labels:
117           severity: warning
118       annotations:
119           summary: Host OOM kill detected (instance {{ $labels.instance }})
120           description: "OOM kill detected\n  VALUE = {{ $value }}\n  LABELS = {{ $labels }}"
121
122     - alert: KubernetesNodeNotReady
123       expr: kube_node_status_condition{condition="Ready",status="true"} == 0
124       for: 10m
125       labels:
126           severity: critical
127       annotations:
128           summary: Kubernetes Node not ready (instance {{ $labels.instance }})
129           description: "Node {{ $labels.node }} has been unready for a long time\n  VALUE = {{ $value }}\n  LABELS = {{ $labels }}"
130
131     - alert: KubernetesNodeMemoryPressure
132       expr: kube_node_status_condition{condition="MemoryPressure",status="true"} == 1
133       for: 2m
134       labels:
135           severity: critical
136       annotations:
137           summary: Kubernetes Node memory pressure (instance {{ $labels.instance }})
138           description: "Node {{ $labels.node }} has MemoryPressure condition\n  VALUE = {{ $value }}\n  LABELS = {{ $labels }}"
139
140     - alert: KubernetesNodeNetworkUnavailable
141       expr: kube_node_status_condition{condition="NetworkUnavailable",status="true"} == 1

```



```

141         for: 2m
142         labels:
143             severity: critical
144         annotations:
145             summary: Kubernetes Node network unavailable (instance {{ $labels.instance }})
146             description: "Node {{ $labels.node }} has NetworkUnavailable condition\n VALUE = {{ $value }}\n LABELS = {{ $labels }}"
147
148         - alert: KubernetesContainerOomKiller
149             expr: (kube_pod_container_status_restarts_total - kube_pod_container_status_restarts_total offset 10m >= 1) and ignoring (reason) min_over_time(kube_pod_container_status_last_terminated_reason{reason="OOMKilled"}[10m]) == 1
150             for: 0m
151             labels:
152                 severity: warning
153             annotations:
154                 summary: Kubernetes Container oom killer (instance {{ $labels.instance }})
155                 description: "Container {{ $labels.container }} in pod {{ $labels.namespace }}/{{ $labels.pod }} has been OOMKilled {{ $value }} times in the last 10 minutes.\n VALUE = {{ $value }}\n LABELS = {{ $labels }}"
156
157         - alert: KubernetesJobFailed
158             expr: kube_job_status_failed > 0
159             for: 0m
160             labels:
161                 severity: warning
162             annotations:
163                 summary: Kubernetes Job failed (instance {{ $labels.instance }})
164                 description: "Job {{ $labels.namespace }}/{{ $labels.job_name }} failed to complete\n VALUE = {{ $value }}\n LABELS = {{ $labels }}"
165
166         - alert: KubernetesCronjobSuspended
167             expr: kube_cronjob_spec_suspend != 0
168             for: 0m
169             labels:
170                 severity: warning
171             annotations:
172                 summary: Kubernetes CronJob suspended (instance {{ $labels.instance }})
173                 description: "CronJob {{ $labels.namespace }}/{{ $labels.cronjob }} is suspended\n VALUE = {{ $value }}\n LABELS = {{ $labels }}"
174
175         - alert: KubernetesPersistentvolumeclaimPending
176

```

```

177     expr: kube_persistentvolumeclaim_status_phase{phase="Pending"} =
178     = 1
179     for: 2m
180     labels:
181         severity: warning
182     annotations:
183         summary: Kubernetes PersistentVolumeClaim pending (instance {{
184 $labels.instance }})
185         description: "PersistentVolumeClaim {{ $labels.namespace }}/{{
186 $labels.persistentvolumeclaim }} is pending\n  VALUE = {{ $value }}\n  LA
187 BELS = {{ $labels }}"
188
189     - alert: KubernetesVolumeOutOfDiskSpace
190     expr: kubelet_volume_stats_available_bytes / kubelet_volume_stats
191 _capacity_bytes * 100 < 10
192     for: 2m
193     labels:
194         severity: warning
195     annotations:
196         summary: Kubernetes Volume out of disk space (instance {{ $labe
197 ls.instance }})
198         description: "Volume is almost full (< 10% left)\n  VALUE = {{
199 $value }}\n  LABELS = {{ $labels }}"
200
201     - alert: KubernetesPersistentvolumeError
202     expr: kube_persistentvolume_status_phase{phase=~"Failed|Pendin
203 g", job="kube-state-metrics"} > 0
204     for: 0m
205     labels:
206         severity: critical
207     annotations:
208         summary: Kubernetes PersistentVolume error (instance {{ $label
209 s.instance }})
210         description: "Persistent volume {{ $labels.persistentvolume }}
211 is in bad state\n  VALUE = {{ $value }}\n  LABELS = {{ $labels }}"
212
213     - alert: KubernetesStatefulsetDown
214     expr: kube_statefulset_replicas != kube_statefulset_status_repl
215 ic
216 as_ready > 0
217     for: 1m
218     labels:
219         severity: critical
220     annotations:
221         summary: Kubernetes StatefulSet down (instance {{ $labels.insta
222 nce }})
223         description: "StatefulSet {{ $labels.namespace }}/{{ $labels.st
224 atefulset }} went down\n  VALUE = {{ $value }}\n  LABELS = {{ $labels }}"

```

```

212     - alert: KubernetesApiServerError
      expr: sum(rate(apiserver_request_total{job="kubernetes-apiserver",code=~"^(?:5..)$"}[1m])) / sum(rate(apiserver_request_total{job="apiserver"}[1m])) * 100 > 3
213
214     for: 2m
215     labels:
216         severity: critical
217     annotations:
218         summary: Kubernetes API server errors (instance {{ $labels.instance }})
219         description: "Kubernetes API server is experiencing high error rate\n VALUE = {{ $value }}\n LABELS = {{ $labels }}"
220
221     - alert: KubernetesClientCertificateExpiresNextWeek
      expr: apiserver_client_certificate_expiration_seconds_count{job="kubernetes-apiservers"} > 0 and histogram_quantile(0.01, sum by (job, le) (rate(apiserver_client_certificate_expiration_seconds_bucket{job="kubernetes-apiservers"}[5m]))) < 7*24*60*60
222
223     for: 0m
224     labels:
225         severity: warning
226     annotations:
227         summary: Kubernetes client certificate expires next week (instance {{ $labels.instance }})
228         description: "A client certificate used to authenticate to the apiserver is expiring next week.\n VALUE = {{ $value }}\n LABELS = {{ $labels }}"
229
230 prometheus.yml: |-
231     global:
232         scrape_interval: 5s
233         evaluation_interval: 5s
234     rule_files:
235         - /etc/prometheus/prometheus.rules
236     # 监控配置
237     alerting:
238         alertmanagers:
239         - scheme: http
240           static_configs:
241             - targets:
242               - "alertmanager.monitoring.svc:9093"
243     # 动态发现配置
244     scrape_configs:
245         - job_name: 'kubernetes-apiservers'
246           # 大致意思是说, 筛选__meta_kubernetes_namespace, __meta_kubernetes_service_name, __meta_kubernetes_endpoint_port_name
247           # 为default, kubernetes, https, 也就是说发现文件的路径为https://kubernetes.default:port/metrics

```

```

248      # 符合这个条件的只有kubernetes 这个endpoint, 所以只会发现apiserver
249      kubernetes_sd_configs:
250      - role: endpoints
251      scheme: https
252      tls_config:
253      ca_file: /var/run/secrets/kubernetes.io/serviceaccount/ca.crt
254      bearer_token_file: /var/run/secrets/kubernetes.io/serviceaccount/
255 token
256      relabel_configs:
257      - source_labels: [__meta_kubernetes_namespace, __meta_kubernetes_
258 service_name, __meta_kubernetes_endpoint_port_name]
259      action: keep
260      regex: default;kubernetes;https
261      - job_name: 'node-exporter'
262      # 同理这个匹配规则是匹配__meta_kubernetes_endpoints_name 值为node-exp
263 order的目标
264      kubernetes_sd_configs:
265      - role: endpoints
266      relabel_configs:
267      - source_labels: [__meta_kubernetes_endpoints_name]
268      regex: 'node-exporter'
269      action: keep
270      - job_name: 'kubernetes-nodes'
271      #将抓取的地地址调整为https://kubernetes.default.svc:443/api/v1/nodes/
272 ${node_name}/proxy/metrics
273      scheme: https
274      tls_config:
275      ca_file: /var/run/secrets/kubernetes.io/serviceaccount/ca.crt
276      bearer_token_file: /var/run/secrets/kubernetes.io/serviceaccount/
277 token
278      kubernetes_sd_configs:
279      - role: node
280      relabel_configs:
281      - action: labelmap
282      regex: __meta_kubernetes_node_label_(.+)
283      - target_label: __address__
284      replacement: kubernetes.default.svc:443
285      - source_labels: [__meta_kubernetes_node_name]
286      regex: (.+)
287      target_label: __metrics_path__
288      replacement: /api/v1/nodes/${1}/proxy/metrics
289      - job_name: 'kubernetes-pods'
290      # 只抓取annotation中 prometheus_io_scrape: true的pod
291      # 将抓取的路径/metircs 调整为prometheus_io_path 设置的值
292      # 定义address抓取端口
293      # labelmap的意思是将kubernetes_pod_label 所有标签都映射到prometheus t
294 arget面板

```

```

289         # 然后自定义两个label分别是kubernetes_namespace 对应值__meta_kubern
290         es_namespace
291         kubernetes_sd_configs:
292             - role: pod
293             relabel_configs:
294             - source_labels: [__meta_kubernetes_pod_annotation_prometheus_io_
295               scrape]
296               action: keep
297               regex: true
298             - source_labels: [__meta_kubernetes_pod_annotation_prometheus_io_
299               path]
300               action: replace
301               target_label: __metrics_path__
302               regex: (.+)
303             - source_labels: [__address__, __meta_kubernetes_pod_annotation_p
304               rometheus_io_port]
305               action: replace
306               regex: ([^:]+)(?::\d+)?;(\d+)
307               replacement: $1:$2
308               target_label: __address__
309             - action: labelmap
310               regex: __meta_kubernetes_pod_label_(.+)
311             - source_labels: [__meta_kubernetes_namespace]
312               action: replace
313               target_label: kubernetes_namespace
314             - source_labels: [__meta_kubernetes_pod_name]
315               action: replace
316               target_label: kubernetes_pod_name
317             - job_name: 'kube-state-metrics'
318               #直接指定的静态路径
319               static_configs:
320                 - targets: ['kube-state-metrics.kube-system.svc:8080']
321             - job_name: 'kubernetes-cadvisor'
322               # 同上
323               scheme: https
324               tls_config:
325                 ca_file: /var/run/secrets/kubernetes.io/serviceaccount/ca.crt
326                 bearer_token_file: /var/run/secrets/kubernetes.io/serviceaccount/
327               token
328               kubernetes_sd_configs:
329                 - role: node
330                 relabel_configs:
331                 - action: labelmap
332                   regex: __meta_kubernetes_node_label_(.+)
333                   target_label: __address__
334                   replacement: kubernetes.default.svc:443
335                 - source_labels: [__meta_kubernetes_node_name]
336                   regex: (.+)

```

```

333         target_label: __metrics_path__
334         replacement: /api/v1/nodes/${1}/proxy/metrics/cadvisor
335     - job_name: 'kubernetes-service-endpoints'
336       kubernetes_sd_configs:
337     - role: endpoints
338       relabel_configs:
339     - source_labels: [__meta_kubernetes_service_annotation_prometheus
340 _io_scrape]
341       action: keep
342       regex: true
343     - source_labels: [__meta_kubernetes_service_annotation_prometheus
344 _io_scheme]
345       action: replace
346       target_label: __scheme__
347       regex: (https?)
348     - source_labels: [__meta_kubernetes_service_annotation_prometheus
349 _io_path]
350       action: replace
351       target_label: __metrics_path__
352       regex: (.+)
353     - source_labels: [__address__, __meta_kubernetes_service_annotati
354 on_prometheus_io_port]
355       action: replace
356       target_label: __address__
357       regex: ([^:]+)(?::\d+)?;(\d+)
358       replacement: $1:$2
359     - action: labelmap
360       regex: __meta_kubernetes_service_label_(.+)
361     - source_labels: [__meta_kubernetes_namespace]
362       action: replace
363       target_label: kubernetes_namespace
364     - source_labels: [__meta_kubernetes_service_name]
365       action: replace
366       target_label: kubernetes_name

```

### 1.3 创建prometheus deployment

使用deployment.yaml 部署服务，需要注意的是这个deployment使用的是最新的prometheus镜像，另外并未使用持久盘，如果是生产环境部署，请提前创建持久盘，并修改deployment卷部分



Plain Text |

```
1 kubectl apply -f deployment.yaml
```

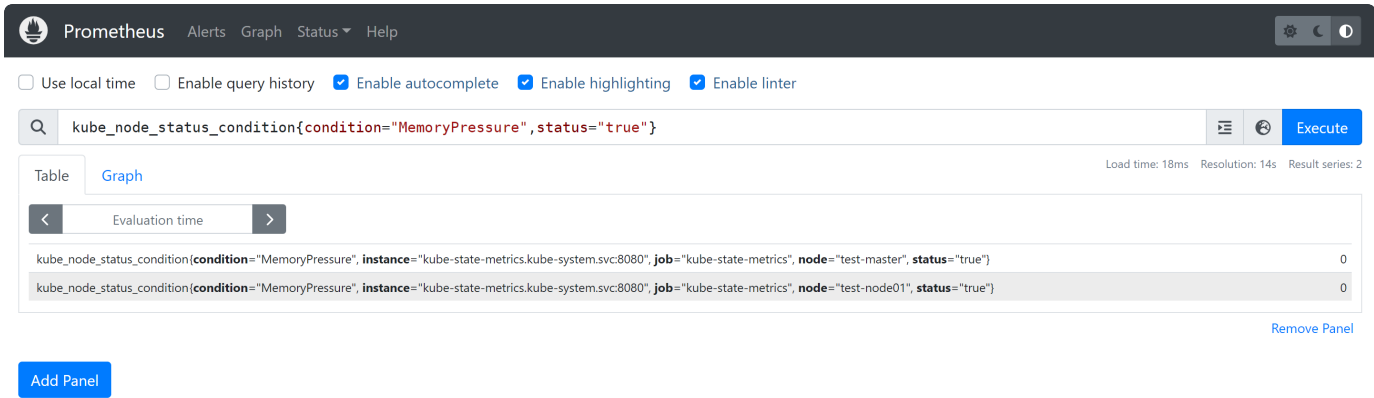
## 1.4 通过Endpoint 对外提供服务

如果需要访问prometheus的控制面板，需要通过kubernetes service服务的方式对外开放端口

▼ Plain Text |

```
1 kubectl apply -f service.yaml
```

当这个服务创建后，可以通过访问nodeip 30000端口的方式访问到，prometheus 端点



## 2 部署exporter

在安装node-exporter 和kube-state-metrics 之前有必要先说明二者的区别

node-exporter 关注的是底层资源， 主要是监控 CPU 使用率、内存使用率、磁盘空间、网络流量等

kube-state-metrics 负责的关注点是 Kubernetes 集群的状态，state-metrics通过访问 kubernetes api 获取大量的监控指标如 Pod 运行状态、副本集状态、服务状态等，然后将其加工为prometheus可以抓取的数据类型。

### 2.1 创建node-exporte配置文件

此 Daemonset 将部署在监控命名空间中。如果您希望将其部署在不同的命名空间中，请在以下 YAML 中进行更改，该目录下包含了demonset 和service 应用

▼ Plain Text |

```
1 cd nodeExporter
2 kubectl apply -f ./
```

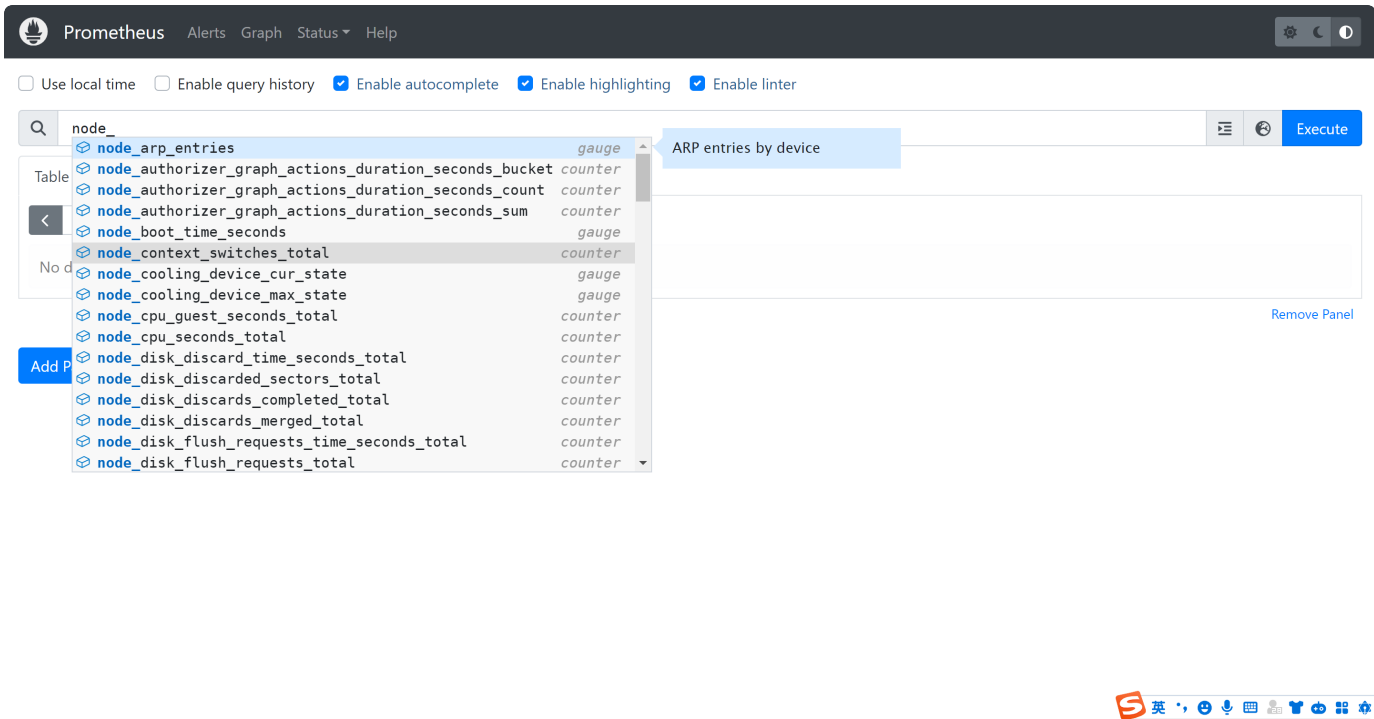
2.2 查看应用状态

▼ Plain Text

```
1 # 列出daemonset监控命名空间中的并确保其处于可用状态。
2 kubectl get daemonset -n monitoring
3 kubectl get endpoints -n monitoring
4
5 # 通过Endpoint可以看到svc已经关联到pod的相应端口上
6 #NAMESPACE      NAME                                ENDPOINTS
7 #monitoring      node-exporter                      10.244.130.102:9100
```

2.3 查看node-exporter 监控指标

来自节点导出器的所有指标都带有前缀node\_



2.4 部署kube-state-metrics

所有 Kube 静态指标都可以从 URI 上的 Kube 状态服务端点获取/metrics。

此配置可以添加为Prometheus 作业配置的一部分。您需要将以下作业配置添加到 Prometheus 配置中，以便 Prometheus 抓取所有 Kube 状态指标。



Plain Text

```
1 cd stateMetrics/
2 kubectl apply -f ./
```

Prometheus Alerts Graph Status Help

Targets

All scrape pools All Unhealthy Collapse All Filter by endpoint or labels Unknown Unhealthy Healthy

kube-state-metrics (1/1 up) show less

Endpoint	State	Labels	Last Scrape	Scrape Duration	Error
http://kube-state-metrics.kube-system.svc:8080/metrics	UP	instance="kube-state-metrics.kube-system.svc:8080" job="kube-state-metrics"	2.404s ago	5.136ms	

kubernetes-apiservers (1/1 up) show less

Endpoint	State	Labels	Last Scrape	Scrape Duration	Error
https://192.168.40.200:6443/metrics	UP	instance="192.168.40.200:6443" job="kubernetes-apiservers"	4.368s ago	149.018ms	

kubernetes-cadvisor (2/2 up) show less

Endpoint	State	Labels	Last Scrape	Scrape Duration	Error
https://kubernetes.default.svc/api/v1/nodes/	UP	beta_kubernetes_io_arch="amd64"	2.933s ago	59.281ms	

3 部署altermanager

alter-manager 是开源的告警系统，它接收来自Prometheus的所有告警信息，然后将这些告警推送到指定的收件人列表中，收件人可以是邮箱，钉钉或者其他web-hook，安装步骤分为下列4个部分

- 1 修改配置告警收件人，告警模板
- 2 配置deployment
- 3 配置钉钉webhook-deployment
- 4 prometheus.rules 告警规则

3.1 修改配置告警收件人

在开始配置之前，请确保你的Prometheus已经部署且正运行，确保prometheus 配置文件中告警部分配置为下列内容

```
1 alerting:
2   alertmanagers:
3     - scheme: http
4     static_configs:
5       - targets:
6         - "alertmanager.monitoring.svc:9093"
```

网易的发件secret 获取

登录到网易邮箱，点击红框部分确保pop3/SMTP状态为开启，然后点击新增授权密码，可能需要短信认证



开启服务： IMAP/SMTP服务 已关闭 | [开启](#)  
POP3/SMTP服务 [已开启](#) | [关闭](#)

POP3/SMTP/IMAP服务能让你在本地客户端上收发邮件，[了解更多 >](#)

温馨提示：在第三方登录网易邮箱，可能存在邮件泄露风险，甚至危害Apple或其他平台账户安全

收取选项：  
☒ 收取最近30天邮件  
☐ 收取全部邮件

温馨提示：收取大量邮件，会耗费您更多的流量，建议您选择“收取最近30天邮件”

通知提醒：  
☐ 开启客户端删除邮件提醒  
当邮件客户端大量删除邮件时，系统会发送提醒信息

授权密码管理： 授权码是用于登录第三方邮件客户端的专用密码。  
适用于登录以下服务：您开启的服务（例如POP3/IMAP/SMTP）、Exchange/CardDAV/CalDAV服务。

使用设备	启用时间	操作
设备1	2024.1.10	<a href="#">删除</a>

[新增授权密码](#)

每个帐号最多设置5个授权密码

alterManager目录里记载了所有的alter相关配置文件

```

1 cd alertManager/
2 cat alertManagerConfigmap.yaml
3
4 kind: ConfigMap
5 apiVersion: v1
6 metadata:
7   name: alertmanager-config
8   namespace: monitoring
9 data:
10  config.yml: |-
11    global:
12      # 发件人, 和发件凭证注意是凭证password对应之前创建的secret, 不是登录密码这些都需要修改
13      smtp_smarthost: 'smtp.163.com:465'
14      smtp_from: lvtujingji@163.com
15      smtp_auth_username: lvtujingji@163.com
16      smtp_auth_password: #####
17      smtp_require_tls: false
18    templates:
19    - '/etc/alertmanager/*.tmpl'
20    route:
21      # 收件人name,group_by 是基于alertname priority 分组
22      # group_wait 组合并等待时间 repeat_interval重发送时间
23      receiver: alert-emailer
24      group_by: ['alertname', 'priority']
25      group_wait: 10s
26      repeat_interval: 1m
27      # 如果需要多联系人过滤, 就可以修改这部分, 匹配severity=slack的告警, 然后发送到收件人slack_demo
28      #routes:
29      #- receiver: slack_demo
30      # Send severity=slack alerts to slack.
31      # match:
32      #   severity: slack
33      #   group_wait: 10s
34      #   repeat_interval: 1m
35      # 收件人信息, 可以配置邮箱地址, 我需要发送到钉钉, 所以填的webhook地址
36      receivers:
37      - name: alert-emailer
38        webhook_configs:
39        - url: 'http://prometheus-webhook-dingtalk.monitoring.svc/dingtalk/webhook_ding/send'
40        send_resolved: true
41      # email address example
42      #- name: email-notifications

```

```
43     # email_configs:
44     #   - to: your-email@example.com
45
```

### 3.2 配置deployment

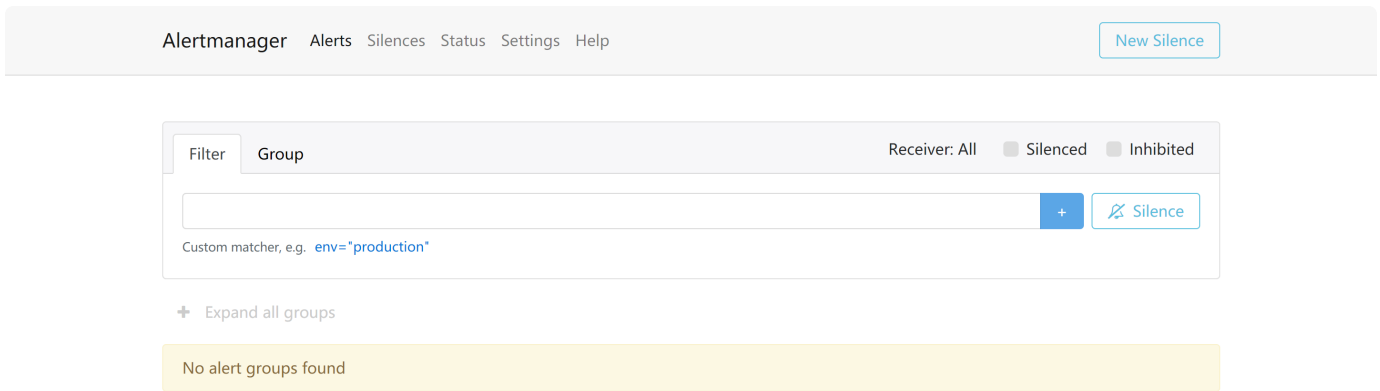
执行下列代码之前确保当前路径在alertManager/

```
▼ Plain Text |
1  kubectl apply -f ./
```

apply 会在monitor名称空间下创建altermanager deployment和serivce，对外的开放的nodeport是31000，通过get命令获取到service状态信息

```
▼ Plain Text |
1  kubectl get -n monitoring svc
2  NAME                                TYPE                CLUSTER-IP          EXTERNAL-IP          P0
   RT(S)          AGE
3  alertmanager    NodePort            10.110.212.137      <none>                90
   93:31000/TCP    4d6h
4  node-exporter   ClusterIP           10.102.77.171       <none>                91
   00/TCP          4d6h
5  prometheus-service NodePort            10.102.142.208      <none>                80
   80:30000/TCP    4d7h
6
```

输入ip+31000端口可以进入altermanager控制面板，可以看到没有任何告警信息



### 3.3 配置钉钉webhook-deployment(可选)

如果要发送告警信息到钉钉，就需要配置这部分，其他webhook应用原理大概都一致，都是创建个告警机器人，然后记录告警地址，修改alertWebhook/configMap.yaml

测试的话可以自己创建个群聊，然后点击机器人



记录webhook和加签里的secret

Webhook:  复制 重置

\* 请保管好此 Webhook 地址，不要公布在外部网站上，泄露有安全风险

使用 Webhook 地址，向钉钉群推送消息 [查看文档](#)

---

\* 安全设置 ? ☐ 自定义关键词 [说明文档](#)

☒ 加签

重置 复制

取消 完成

```
1  apiVersion: v1
2  kind: ConfigMap
3  metadata:
4    name: prometheus-webhook-dingtalk
5    namespace: monitoring
6  data:
7    config.yml: |
8      targets:
9        # 将 URL 替换成你钉钉告警机器的人的地址
10       # 将secre替换成你钉钉的secret
11       webhook_ding:
12         url: https://oapi.dingtalk.com/robot/send?access_token=1dfa9f2d576
13         49ace8ecd5917814a8c8d43201419f421ba5155f02936d8aec51b # 配置机器人的webhook
14         _url)
15         secret: SECd2239c2db306ad97e80fbab4fdc9b1c8df58d368c0eb8fda1f7f3e
16         81e3312a6
17       message:
18         title: '{{ template "default.title" . }}'
19         text: '{{ template "default.content" . }}'
```

修改完毕之后直接应用就行

```
1 kubectl apply -f ./
2 [root@test-master alertWebhook]# kubectl get svc -n monitoring
3 NAME                                TYPE                CLUSTER-IP          EXTERNAL-IP    P0
  RT(S)          AGE
4 alertmanager                NodePort            10.110.212.137      <none>         90
  93:31000/TCP    4d6h
5 node-exporter                ClusterIP           10.102.77.171       <none>         91
  00/TCP          4d6h
6 prometheus-service          NodePort            10.102.142.208      <none>         80
  80:30000/TCP    4d7h
7 prometheus-webhook-dingtalk  ClusterIP           10.96.200.178       <none>         8
  0/TCP           4d6h
8
```

### 3.4 prometheus.rules 告警规则

这部分是关于Prometheus监控告警规则的大致介绍，大部分都是直接从下面那个地址获取的，我截取了我目前需要的一部分，主要分为basic和kubernetes两个部分的监控告警，也可以根据你们的应用情况，去截取修改

<https://samber.github.io/awesome-prometheus-alerts/>



```

1  apiVersion: v1
2  kind: ConfigMap
3  metadata:
4    name: prometheus-config
5    labels:
6      name: prometheus-config
7    namespace: monitoring
8  data:
9    prometheus.rules: |-
10     groups:
11       # 可以基于group的name进行监控项分类，然后进行告警收件人分类
12       - name: basic
13         rules:
14           # 这个监控的意思是当获取up{kubernetes_name="prometheus-service"} 抓取
15           # 不到任何监控时
16           # 判断这个Prometheus-service已经出现问题了，然后触发告警
17           - alert: PrometheusJobMissing
18             expr: absent(up{kubernetes_name="prometheus-service"})
19             for: 0m
20             labels:
21               severity: warning
22             annotations:
23               summary: Prometheus job missing (instance {{ $labels.instance
24               value }})
25               description: "A Prometheus job has disappeared\n  VALUE = {{ $v
26               alue }}\n  LABELS = {{ $labels }}"
27             # 如果出现自动发现的目标断开就会触发告警
28           - alert: PrometheusTargetMissing
29             expr: up == 0
30             for: 0m
31             labels:
32               severity: critical
33             annotations:
34               summary: Prometheus target missing (instance {{ $labels.instanc
35               e }})
36               description: "A Prometheus target has disappeared. An exporter
37               might be crashed.\n  VALUE = {{ $value }}\n  LABELS = {{ $labels }}"
38             #统计Prometheus关键组件有没有出现频繁重启现象
39           - alert: PrometheusTooManyRestarts
40             expr: changes(process_start_time_seconds{kubernetes_name=~"promet
41             heus-service|alertmanager|node-exporter"}[15m]) > 2
42             for: 0m
43             labels:
44               severity: warning
45             annotations:

```

```

40         summary: Prometheus too many restarts (instance {{ $labels.instance }})
41         description: "Prometheus has restarted more than twice in the last 15 minutes. It might be crashlooping.\n VALUE = {{ $value }}\n LABELS = {{ $labels }}"
42         # 内存的可用率小于10%
43         - alert: HostOutOfMemory
44           expr: node_memory_MemAvailable_bytes / node_memory_MemTotal_bytes * 100 < 10
45           for: 2m
46           labels:
47             severity: warning
48           annotations:
49             summary: Host out of memory (instance {{ $labels.instance }})
50             description: "Node memory is filling up (< 10% left)\n VALUE = {{ $value }}\n LABELS = {{ $labels }}"
51             # 网络传入流量超过100M/s
52           - alert: HostUnusualNetworkThroughputIn
53             expr: sum by (instance) (rate(node_network_receive_bytes_total[2m])) / 1024 / 1024 > 100
54             for: 5m
55             labels:
56               severity: warning
57             annotations:
58               summary: Host unusual network throughput in (instance {{ $labels.instance }})
59               description: "Host network interfaces are probably receiving too much data (> 100 MB/s)\n VALUE = {{ $value }}\n LABELS = {{ $labels }}"
60             #网络传出流量超过100M/s
61           - alert: HostUnusualNetworkThroughputOut
62             expr: sum by (instance) (rate(node_network_transmit_bytes_total[2m])) / 1024 / 1024 > 100
63             for: 5m
64             labels:
65               severity: warning
66             annotations:
67               summary: Host unusual network throughput out (instance {{ $labels.instance }})
68               description: "Host network interfaces are probably sending too much data (> 100 MB/s)\n VALUE = {{ $value }}\n LABELS = {{ $labels }}"
69             # 磁盘读取速率超出50MB/s
70           - alert: HostUnusualDiskReadRate
71             expr: (sum by (instance) (rate(node_disk_read_bytes_total[2m])) / 1024 / 1024 > 50) * on(instance) group_left(nodename) node_uname_info {nodename=~".+"}
72             for: 5m
73             labels:

```

```

74         severity: warning
75     annotations:
76         summary: Host unusual disk read rate (instance {{ $labels.instance }})
77         description: "Disk is probably reading too much data (> 50 MB/s)\n VALUE = {{ $value }}\n LABELS = {{ $labels }}"
78         # 磁盘写入速率超出50MB/s
79         - alert: HostUnusualDiskWriteRate
80             expr: (sum by (instance) (rate(node_disk_written_bytes_total[2m])) / 1024 / 1024 > 50) * on(instance) group_left (nodename) node_uname_info{nodename=~".+"}
81             for: 2m
82             labels:
83                 severity: warning
84             annotations:
85                 summary: Host unusual disk write rate (instance {{ $labels.instance }})
86                 description: "Disk is probably writing too much data (> 50 MB/s)\n VALUE = {{ $value }}\n LABELS = {{ $labels }}"
87                 # 磁盘剩余空间小于10%
88                 - alert: HostOutOfDiskSpace
89                     expr: ((node_filesystem_avail_bytes * 100) / node_filesystem_size_bytes < 10 and ON (instance, device, mountpoint) node_filesystem_readonly == 0) * on(instance) group_left (nodename) node_uname_info{nodename=~".+"}
90                     for: 2m
91                     labels:
92                         severity: warning
93                     annotations:
94                         summary: Host out of disk space (instance {{ $labels.instance }})
95                         description: "Disk is almost full (< 10% left)\n VALUE = {{ $value }}\n LABELS = {{ $labels }}"
96                         # CPU 10分钟的负载超过80%
97                         - alert: HostHighCpuLoad
98                             expr: (sum by (instance) (avg by (mode, instance) (rate(node_cpu_seconds_total{mode!="idle"}[2m])))) > 0.85) * on(instance) group_left (nodename) node_uname_info{nodename=~".+"}
99                             for: 10m
100                             labels:
101                                 severity: warning
102                             annotations:
103                                 summary: Host high CPU load (instance {{ $labels.instance }})
104                                 description: "CPU load is > 80%\n VALUE = {{ $value }}\n LABELS = {{ $labels }}"
105                                 # 主机内出现了OOM
106                                 - alert: HostOomKillDetected
107

```

```

108         expr: (increase(node_vmstat_oom_kill[1m]) > 0) * on(instance) gro
up_left (nodename) node_uname_info{nodename=~".+"}
109         for: 0m
110         labels:
111             severity: warning
112         annotations:
113             summary: Host OOM kill detected (instance {{ $labels.instance
}})
114             description: "OOM kill detected\n  VALUE = {{ $value }}\n  LABE
LS = {{ $labels }}"
115             # node状态notREADY
116             - alert: KubernetesNodeNotReady
117                 expr: kube_node_status_condition{condition="Ready",status="tru
e"} == 0
118                 for: 10m
119                 labels:
120                     severity: critical
121                 annotations:
122                     summary: Kubernetes Node not ready (instance {{ $labels.instanc
e }})
123                     description: "Node {{ $labels.node }} has been unready for a lo
ng time\n  VALUE = {{ $value }}\n  LABELS = {{ $labels }}"
124                     # node面临内存压力
125                     - alert: KubernetesNodeMemoryPressure
126                         expr: kube_node_status_condition{condition="MemoryPressure",statu
s="true"} == 1
127                         for: 2m
128                         labels:
129                             severity: critical
130                         annotations:
131                             summary: Kubernetes Node memory pressure (instance {{ $labels.i
nstance }})
132                             description: "Node {{ $labels.node }} has MemoryPressure condit
ion\n  VALUE = {{ $value }}\n  LABELS = {{ $labels }}"
133                             # node网络出现不可用现象
134                             - alert: KubernetesNodeNetworkUnavailable
135                                 expr: kube_node_status_condition{condition="NetworkUnavailable",s
tatus="true"} == 1
136                                 for: 2m
137                                 labels:
138                                     severity: critical
139                                 annotations:
140                                     summary: Kubernetes Node network unavailable (instance {{ $labe
ls.instance }})
141                                     description: "Node {{ $labels.node }} has NetworkUnavailable co
ndition\n  VALUE = {{ $value }}\n  LABELS = {{ $labels }}"
142                                     # nodekill 容器, kill原因内存不足或超出
143                                     - alert: KubernetesContainerOomKiller

```

```

    expr: (kube_pod_container_status_restarts_total - kube_pod_contai
ner_status_restarts_total offset 10m >= 1) and ignoring (reason) min_over
144 _time(kube_pod_container_status_last_terminated_reason{reason="OOMKille
145 d"}[10m]) == 1
146     for: 0m
147     labels:
148         severity: warning
149     annotations:
150         summary: Kubernetes Container oom killer (instance {{ $labels.i
151 nstance }})
152         description: "Container {{ $labels.container }} in pod {{ $labe
153 ls.namespace }}/{{ $labels.pod }} has been OOMKilled {{ $value }} times i
154 n the last 10 minutes.\n  VALUE = {{ $value }}\n  LABELS = {{ $labels }}"
155     # Job任务执行failed
156     - alert: KubernetesJobFailed
157     expr: kube_job_status_failed > 0
158     for: 0m
159     labels:
160         severity: warning
161     annotations:
162         summary: Kubernetes Job failed (instance {{ $labels.instance
163 }})
164         description: "Job {{ $labels.namespace }}/{{ $labels.job_name
165 }} failed to complete\n  VALUE = {{ $value }}\n  LABELS = {{ $labels }}"
166     # 计划任务暂停
167     - alert: KubernetesCronjobSuspended
168     expr: kube_cronjob_spec_suspend != 0
169     for: 0m
170     labels:
171         severity: warning
172     annotations:
173         summary: Kubernetes CronJob suspended (instance {{ $labels.inst
174 ance }})
175         description: "CronJob {{ $labels.namespace }}/{{ $labels.cronjo
176 b }} is suspended\n  VALUE = {{ $value }}\n  LABELS = {{ $labels }}"
177     # 磁盘PVC未绑定PV
178     - alert: KubernetesPersistentvolumeclaimPending
179     expr: kube_persistentvolumeclaim_status_phase{phase="Pending"} =
180 = 1
181     for: 2m
182     labels:
183         severity: warning
184     annotations:
185         summary: Kubernetes PersistentVolumeClaim pending (instance {{
186 $labels.instance }})
187         description: "PersistentVolumeClaim {{ $labels.namespace }}/{{
188 $labels.persistentvolumeclaim }} is pending\n  VALUE = {{ $value }}\n  LA
189 BELS = {{ $labels }}"

```

```

178     # 挂载卷的存储不足
179     - alert: KubernetesVolumeOutOfDiskSpace
      expr: kubelet_volume_stats_available_bytes / kubelet_volume_stats
180     _capacity_bytes * 100 < 10
181     for: 2m
182     labels:
183     severity: warning
184     annotations:
      summary: Kubernetes Volume out of disk space (instance {{ $labels.instance }})
185     description: "Volume is almost full (< 10% left)\n  VALUE = {{
186     $value }}\n  LABELS = {{ $labels }}"
187     # pv出现异常
188     - alert: KubernetesPersistentvolumeError
      expr: kube_persistentvolume_status_phase{phase=~"Failed|Pending", job="kube-state-metrics"} > 0
189     for: 0m
190     labels:
191     severity: critical
192     annotations:
193     summary: Kubernetes PersistentVolume error (instance {{ $labels.instance }})
194     description: "Persistent volume {{ $labels.persistentvolume }}
195     is in bad state\n  VALUE = {{ $value }}\n  LABELS = {{ $labels }}"
196     # statefulset控制器出现异常
197     - alert: KubernetesStatefulsetDown
      expr: kube_statefulset_replicas != kube_statefulset_status_replicas_ready > 0
198     for: 1m
199     labels:
200     severity: critical
201     annotations:
202     summary: Kubernetes StatefulSet down (instance {{ $labels.instance }})
203     description: "StatefulSet {{ $labels.namespace }}/{{ $labels.statefulset }} went down\n  VALUE = {{ $value }}\n  LABELS = {{ $labels }}"
204     # 集群api接口出现异常
205     - alert: KubernetesApiServerErrors
      expr: sum(rate(apiserver_request_total{job="kubernetes-apiservers",code=~"^(?:5..)$"}[1m])) / sum(rate(apiserver_request_total{job="apiserver"}[1m])) * 100 > 3
206     for: 2m
207     labels:
208     severity: critical
209     annotations:
210     summary: Kubernetes API server errors (instance {{ $labels.instance }})
211
212

```

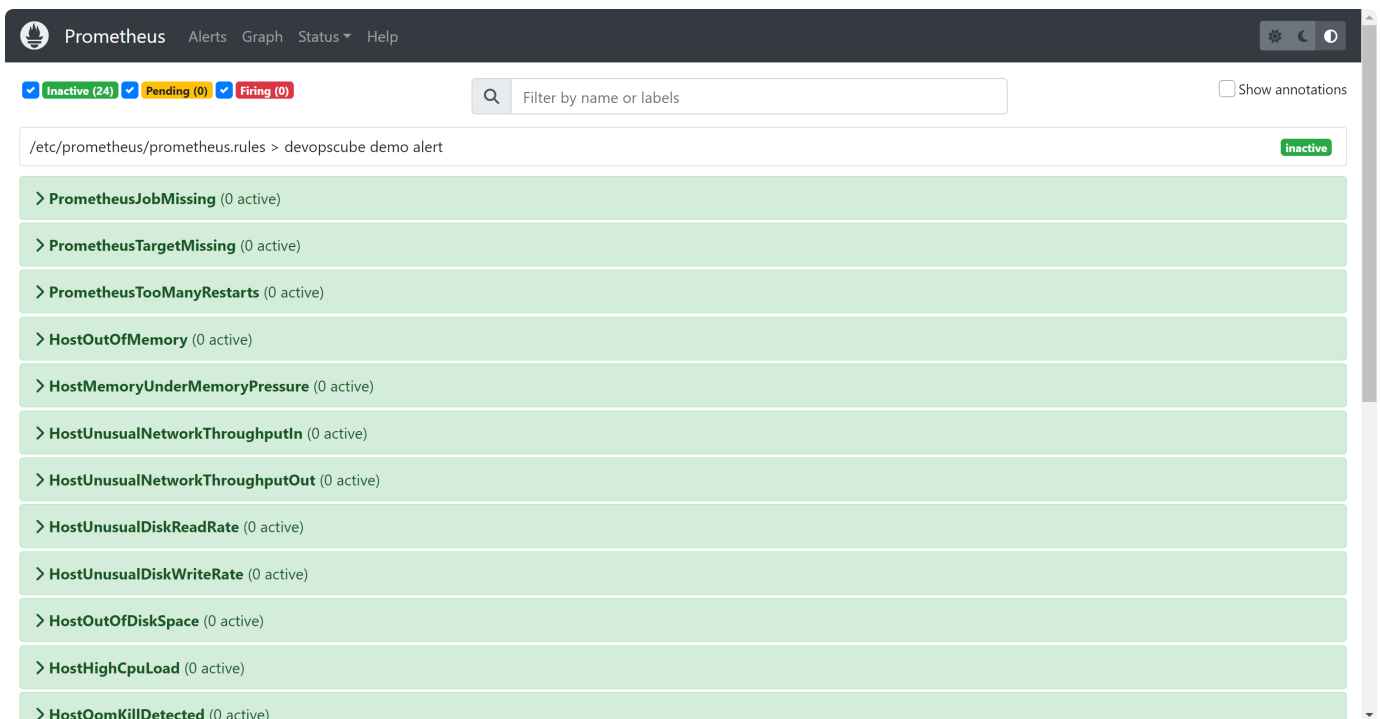
```

213         description: "Kubernetes API server is experiencing high error
214 rate\n VALUE = {{ $value }}\n LABELS = {{ $labels }}"
215         # 证书即将过期
        - alert: KubernetesClientCertificateExpiresNextWeek
          expr: apiserver_client_certificate_expiration_seconds_count{job
216 ="kubernetes-apiservers"} > 0 and histogram_quantile(0.01, sum by (job, l
217 e) (rate(apiserver_client_certificate_expiration_seconds_bucket{job="kub
218 netes-apiservers"}[5m]))) < 7*24*60*60
          for: 0m
          labels:
219             severity: warning
          annotations:
221             summary: Kubernetes client certificate expires next week (insta
nce {{ $labels.instance }})
          description: "A client certificate used to authenticate to the
222 apiserver is expiring next week.\n VALUE = {{ $value }}\n LABELS = {{
$labels }}"

```

## 结尾

如果一切顺利的话，可以在Prometheus面板查看所有的告警项，和集群里的发现目标，有些配置可能需要基于你的集群配置去调整，手册里附带了关于grafana的yaml文件，如果需要可以直接在grafana目录中apply，还附带了所有关于ingress入口的yaml文件，如果你希望通过域名访问Prometheus或者grafana，可以参考ingress目录里的配置进行调整，如果在部署过程中有其他问题，也可以邮箱联系作者 lvtujingji@163.com




## Targets

All scrape pools ▾

AllUnhealthyExpand All

 Filter by endpoint or labels

 Unknown Unhealthy Healthy

kube-state-metrics (1/1 up) [show more](#)

kubernetes-apiservers (1/1 up) [show more](#)

kubernetes-cadvisor (2/2 up) [show more](#)

kubernetes-nodes (2/2 up) [show more](#)

kubernetes-service-endpoints (5/5 up) [show more](#)

node-exporter (1/1 up) [show more](#)

kubernetes\_pod\_name:nginx-pod]

Graph: 

Details:

- alertname: PrometheusTargetMissing
- instance: 10.244.130.110:6000
- job: kubernetes-pods
- kubernetes\_namespace: default
- kubernetes\_pod\_name: nginx-pod



机器人 机器人 1月26日 09:13

**[RESOLVED] PrometheusTargetMissing**

**Alerts Resolved**

**[CRITICAL] Prometheus target missing  
(instance 10.244.130.110:6000)**

**Description:** A Prometheus target has disappeared. An exporter might be crashed.

VALUE = 0 LABELS = map[name:up  
instance:10.244.130.110:6000 job:kubernetes-  
pods kubernetes\_namespace:default  
kubernetes\_pod\_name:nginx-pod]

Graph: 

Details:

- alertname: PrometheusTargetMissing
- instance: 10.244.130.110:6000