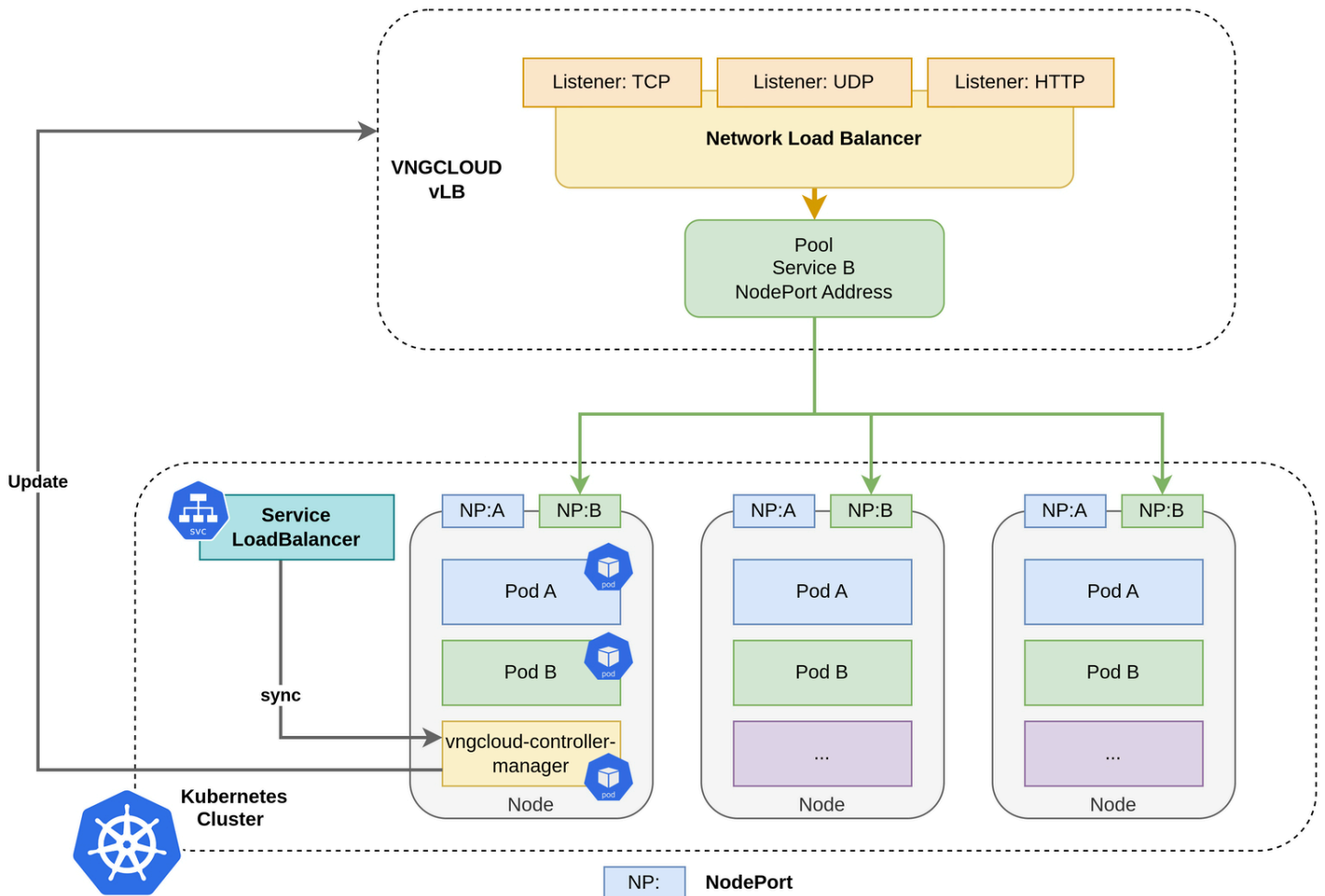


2. Tích hợp Network Load Balancer (NLB) trên VKS

2.1. Network Load Balancer (NLB) là gì?

- **Network Load Balancer** (tên viết tắt là **NLB**) là một bộ cân bằng tải được cung cấp bởi VNGCloud giúp phân phối lưu lượng truy cập mạng đến nhiều máy chủ back-end (backend servers) trong một nhóm máy tính (instance group). NLB hoạt động ở layer 4 của mô hình OSI, giúp cân bằng tải dựa trên địa chỉ IP và cổng TCP/UDP. Để biết thêm thông tin chi tiết về NLB, vui lòng tham khảo tại <https://docs.vngcloud.vn/vng-cloud-document/vn/vks/network/lam-viec-voi-network-load-balancing-nlb>

2.2. Mô hình triển khai



(Đường dẫn đến hình ảnh: [https://docs.vngcloud.vn/~gitbook/image?url=https%3A%2F%2F3672463924-files.gitbook.io%2F%2Ffiles%2Fv0%2Fb%2Fgitbook-x-prod.appspot.com%2Fo%2Fspaces%252FB0NrrdJdpYOYzRkbWp5%252Fuploads%252FVYBtJjEoUNgDi1f5J9vL%252Fimage.png%3Falt%3Dmedia%26token%3D554a2d62-320e-48d1-a884-3c7cce589071&width=768&dpr=2&quality=100&sign=4336a109&sv=2\)](https://docs.vngcloud.vn/~gitbook/image?url=https%3A%2F%2F3672463924-files.gitbook.io%2F%2Ffiles%2Fv0%2Fb%2Fgitbook-x-prod.appspot.com%2Fo%2Fspaces%252FB0NrrdJdpYOYzRkbWp5%252Fuploads%252FVYBtJjEoUNgDi1f5J9vL%252Fimage.png%3Falt%3Dmedia%26token%3D554a2d62-320e-48d1-a884-3c7cce589071&width=768&dpr=2&quality=100&sign=4336a109&sv=2)))

- **VNGCloud LoadBalancer Controller**: là một bộ điều khiển (hay còn gọi là **controller**) chạy trên các cụm Kubernetes được triển khai trên VNGCloud. Controller này chịu trách nhiệm:
 - **Quản lý các Network Load Balancer (NLB)** cho các Service của Kubernetes có thuộc tính `serviceType` là `LoadBalancer`.
 - **Điều phối lưu lượng mạng** đến các pod trên worker node của bạn.

2.3. Cài đặt VNGCloud LoadBalancer Controller để sử dụng NLB

Để sử dụng NLB trên VKS, bạn cần cài đặt plugin **VNGCloud LoadBalancer Controller**. Plugin này đảm nhận vai trò quản lý và cấu hình NLB cho các tài nguyên Service có thuộc tính `serviceType` là `LoadBalancer`. Mặc định, một cụm Kubernetes trên VKS đã được **cài đặt sẵn** plugin **VNGCloud LoadBalancer Controller** này. Việc cài đặt sẵn này giúp bạn có thể sử dụng NLB mà không cần phải thực hiện thêm bất kỳ

thao tác cấu hình nào khác và đảm bảo luôn được cập nhật phiên bản mới nhất. Tuy nhiên người dùng hoàn toàn có thể bật/tắt plugin này thông qua thiết lập **Enable vLB Native Integration Driver** trên **VKS Console**.

Ngoài ra, người dùng hoàn toàn có thể tự cài đặt plugin **VNGCloud LoadBalancer Controller** bằng cách thực hiện các bước sau:

- **Bước 1:** Tạo một Kubernetes cluster trên VKS, hoặc sử dụng một cluster đã có. Tải file `kubeconfig` của cluster về thiết bị của bạn. Ngoài ra, bạn nên kiểm tra file `kubeconfig` này đã có thể kết nối đến cluster thành công hay chưa thông qua `kubectl` CLI.
- **Bước 2:** Khởi tạo hoặc sử dụng một **IAM Service Account** đã có trên **IAM Console** và thiết lập hai policy `vLBFullAccess` và `vServerFullAccess` cho **Service Account** này. Bạn có thể tham khảo cách tạo **Service Account** tại <https://hcm-3.console.vngcloud.vn/iam/service-accounts>. Sau đó, bạn cần lưu lại **Client ID** và **Client Secret** của **Service Account** này.
- **Bước 3:** Bạn cần cài đặt **Helm 3.0** trên thiết bị của bạn, bạn có thể tham khảo cách cài đặt Helm tại <https://helm.sh/docs/intro/install>. Sau khi `helm` CLI đã được cài đặt thành công, bạn đã có thể dùng `helm` CLI để cài đặt plugin, lưu ý rằng bạn cần thay thế các giá trị **Client ID** và **Client Secret** của **Service Account** vào lệnh dưới đây:

```
helm install vngcloud-load-balancer-controller oci://vcr.vngcloud.vn/81-vks-public/vks-helm-charts/vngcloud-load-balance
```

- **Bước 4:** Sau khi Helm hoàn tất việc cài đặt **VNGCloud LoadBalancer Controller**, bạn có thể kiểm tra xem plugin đã hoạt động hay chưa bằng lệnh:

```
kubectl -n kube-system get pod -l app.kubernetes.io/name=vngcloud-load-balancer-controller
```

Dưới đây là ví dụ kết quả trả về khi plugin đã hoạt động:

NAME	READY	STATUS	RESTARTS	AGE
vngcloud-load-balancer-controller-1736217866-manager-77599vrxpz	1/1	Running	0	4h24m

2.4. Sử dụng Network Load Balancer (NLB) trên VKS

Sau khi đã cài đặt plugin **VNGCloud LoadBalancer Controller** thành công, bạn đã có thể sử dụng NLB thông qua tài nguyên `Service` trên Kubernetes. Dưới đây là một vài tình huống sử dụng NLB trên VKS:

Ví dụ 1: Công khai một dịch vụ Nginx trên cổng 80 với NLB đến Internet:

```
apiVersion: apps/v1
kind: Deployment
metadata:
  name: nginx-app
spec:
  selector:
    matchLabels:
      app: nginx
  replicas: 1
  template:
    metadata:
      labels:
        app: nginx
    spec:
      containers:
        - name: nginx
          image: nginx:1.19.1
          ports:
            - containerPort: 80
```

```
---
apiVersion: v1
kind: Service
metadata:
  name: nginx-service
spec:
  selector:
    app: nginx
  type: LoadBalancer
  ports:
    - protocol: TCP
      port: 80
      targetPort: 80
```

Ví dụ 2: Triển khai dịch vụ Apache HTTP và chỉ cho phép truy cập nội bộ qua port 8080:

```

apiVersion: apps/v1
kind: Deployment
metadata:
  name: internal-http-apache2-deployment
spec:
  replicas: 2
  selector:
    matchLabels:
      app: apache2
  template:
    metadata:
      labels:
        app: apache2
    spec:
      containers:
        - name: apache2
          image: httpd
          ports:
            - containerPort: 80
---
apiVersion: v1
kind: Service
metadata:
  name: internal-http-apache2-service
  annotations:
    vks.vngcloud.vn/scheme: "internal" # Thiết lập annotation để tạo NLB nội bộ
spec:
  selector:
    app: apache2
  type: LoadBalancer
  ports:
    - name: http
      protocol: TCP
      port: 8080
      targetPort: 80

```

Ví dụ 3: Triển khai một dịch vụ sử dụng protocol UDP:

```

apiVersion: apps/v1
kind: Deployment
metadata:
  name: udp-server-deployment
spec:
  selector:
    matchLabels:
      name: udp-server
  replicas: 5
  template:
    metadata:
      labels:
        name: udp-server
    spec:
      containers:
        - name: udp-server
          image: vcr.vngcloud.vn/udp-server
          imagePullPolicy: Always
          ports:
            - containerPort: 10001
              protocol: UDP
---
apiVersion: v1
kind: Service
metadata:
  name: udp-server-service
  annotations:
    vks.vngcloud.vn/pool-algorithm: "source-ip" # Thiết lập annotation để chọn thuật toán cân bằng tải
  labels:
    app: udp-server
spec:
  type: LoadBalancer
  sessionAffinity: ClientIP
  ports:
    - port: 10001
      protocol: UDP
  selector:
    name: udp-server

```

Ví dụ 4: Sử dụng lại một NLB đã có trên hệ thống vLB. Để sử dụng lại một NLB đã có, bạn cần nhập thông tin **Load Balancer ID** vào annotation `vks.vngcloud.vn/load-balancer-id` của Service:

```

apiVersion: apps/v1
kind: Deployment
metadata:
  name: external-http-nginx-deployment
spec:
  replicas: 2
  selector:
    matchLabels:
      app: nginx
  template:
    metadata:
      labels:
        app: nginx
    spec:
      containers:
        - name: nginx
          image: nginx
          ports:
            - containerPort: 80
---
kind: Service
apiVersion: v1
metadata:
  name: external-http-nginx-service
  annotations:
    vks.vngcloud.vn/load-balancer-id: "__PUT_YOUR_LOAD_BALANCER_ID__"
spec:
  selector:
    app: nginx
  type: LoadBalancer
  ports:
    - name: http
      port: 80
      targetPort: 80

```

Ngoài ra, bạn cũng có thể tìm hiểu thêm về các annotation khác mà bạn có thể sử dụng để tùy chỉnh NLB phù hợp với nhu cầu của bạn tại đường dẫn <https://docs.vngcloud.vn/vng-cloud-document/vn/vks/network/lam-viec-voi-network-load-balancing-nlb/cau-hinh-cho-mot-network-load-balancer>.

2.5. Cấu hình cho một Network Load Balancer

Bạn có thể tùy chỉnh hoặc cấu hình NLB phù hợp với nhu cầu của bạn thông qua các annotation dưới đây:

- `vks.vngcloud.vn/load-balancer-id` :
 - Bắt buộc phải có trong spec của tài nguyên `Service` : Không bắt buộc
 - Ý nghĩa: Dùng để sử dụng lại một NLB sẵn có trên hệ thống vLB. Nếu annotation này không được thiết lập, một NLB mới sẽ được tạo tự động.
- `vks.vngcloud.vn/load-balancer-name` :
 - Bắt buộc phải có trong spec của tài nguyên `Service` : Không bắt buộc
 - Ý nghĩa: Giá trị của annotation này sẽ dùng làm tên cho NLB của bạn. Nếu annotation này không được thiết lập, **VNGCloud LoadBalancer Controller** sẽ tự động tạo một tên cho NLB của bạn. Annotation này sẽ không có tác dụng nếu một trong các annotation sau được thiết lập: `vks.vngcloud.vn/load-balancer-id`.
- `vks.vngcloud.vn/package-id` :
 - Bắt buộc phải có trong spec của tài nguyên `Service` : Không bắt buộc
 - Ý nghĩa: Dùng để thiết lập package của NLB. Nếu không được thiết lập, mặc định sẽ sử dụng cấu hình **NLB Small**. Ngoài ra, bạn cũng có thể dùng annotation này để cấu hình lại package của NLB hiện tại.
- `vks.vngcloud.vn/tags` :
 - Bắt buộc phải có trong spec của tài nguyên `Service` : Không bắt buộc
 - Ý nghĩa: Tiến hành gắn thẻ cho NLB của bạn. Thẻ này bao gồm các cặp **key-value** nhằm giúp bạn dễ dàng quản lí các NLB trên **VNGCloud Portal**.
- `vks.vngcloud.vn/scheme` :

- Bắt buộc phải có trong spec của tài nguyên Service : Không bắt buộc
- Ý nghĩa: Bao gồm 2 giá trị là `internet-facing` và `internal` . Nếu giá trị của annotation này là `internet-facing` , NLB sẽ được cấu hình để truy cập từ Internet. Nếu giá trị của annotation này là `internal` , NLB sẽ được cấu hình để truy cập từ mạng nội bộ.
- `vks.vngcloud.vn/security-groups` :
 - Bắt buộc phải có trong spec của tài nguyên Service : Không bắt buộc
 - Ý nghĩa: Mặc định, **VNGCloud LoadBalancer Controller** sẽ tự động tạo một **Security Group** mới cho NLB của bạn. Tuy nhiên, nếu bạn muốn sử dụng một **Security Group** đã có, bạn có thể sử dụng annotation này để thiết lập **Security Group** cho NLB của bạn. Ví dụ: `vks.vngcloud.vn/security-groups: secg-303d9c91-f4b2-400d-8e81-12039167da0c` . Ngoài ra, bạn cũng có thể cấu hình nhiều **Security Group** cho NLB của bạn bằng cách sử dụng dấu phẩy , để ngăn cách giữa các **Security Group**. Ví dụ: `vks.vngcloud.vn/security-groups: secg-303d9c91-f4b2-400d-8e81-12039167da0c, secg-303d9c91-f4b2-400d-8e81-12039167da0d` .
- `vks.vngcloud.vn/inbound-cidrs` :
 - Bắt buộc phải có trong spec của tài nguyên Service : Không bắt buộc
 - Ý nghĩa: Mặc định, NLB sẽ được cấu hình để truy cập từ mọi địa chỉ IP. Tuy nhiên, nếu bạn muốn cấu hình NLB để chỉ cho phép truy cập từ một số địa chỉ IP cụ thể, bạn có thể sử dụng annotation này. Ví dụ: `vks.vngcloud.vn/inbound-cidrs: 183.23.23.13` . Ngoài ra, bạn cũng có thể cấu hình nhiều địa chỉ IP cho NLB của bạn bằng cách sử dụng dấu phẩy , để ngăn cách giữa các địa chỉ IP. Ví dụ: `vks.vngcloud.vn/inbound-cidrs: 183.23.23.13, 183.23.23.23` .
- `vks.vngcloud.vn/healthy-threshold-count` :
 - Bắt buộc phải có trong spec của tài nguyên Service : Không bắt buộc
 - Ý nghĩa: Là ngưỡng số lần liên tiếp mà một máy chủ back-end phải trả về trạng thái `healthy` để được coi là **HEALTHY**. Mặc định, giá trị của annotation này là `3` .
- `vks.vngcloud.vn/unhealthy-threshold-count` :
 - Bắt buộc phải có trong spec của tài nguyên Service : Không bắt buộc
 - Ý nghĩa: Là ngưỡng số lần liên tiếp mà một máy chủ back-end phải trả về trạng thái `unhealthy` để được coi là **UNHEALTHY**. Mặc định, giá trị của annotation này là `3` .
- `vks.vngcloud.vn/healthcheck-interval-seconds` :
 - Bắt buộc phải có trong spec của tài nguyên Service : Không bắt buộc
 - Ý nghĩa: Là khoảng thời gian giữa các lần kiểm tra sức khỏe của máy chủ back-end. Mặc định, giá trị của annotation này là `30` giây.
- `vks.vngcloud.vn/healthcheck-timeout-seconds` :
 - Bắt buộc phải có trong spec của tài nguyên Service : Không bắt buộc
 - Ý nghĩa: Là thời gian tối đa mà một máy chủ back-end phải trả về trạng thái `healthy` hoặc `unhealthy` . Mặc định, giá trị của annotation này là `5` giây.
- `vks.vngcloud.vn/healthcheck-protocol` :
 - Bắt buộc phải có trong spec của tài nguyên Service : Không bắt buộc
 - Ý nghĩa: Là giao thức sử dụng để kiểm tra sức khỏe của máy chủ back-end. Mặc định, giá trị của annotation này là `TCP` . Ngoài ra, một vài giá trị khác mà bạn có thể sử dụng là `PING-UDP` , `HTTP` , `HTTPS` .
- `vks.vngcloud.vn/healthcheck-http-method` :
 - Bắt buộc phải có trong spec của tài nguyên Service : Không bắt buộc
 - Ý nghĩa: Là phương thức HTTP sử dụng để kiểm tra sức khỏe của máy chủ back-end. Mặc định, giá trị của annotation này là `GET` . Ngoài ra, một vài giá trị khác mà bạn có thể sử dụng là `PUT` , `POST` . Lưu ý rằng annotation này chỉ có tác dụng khi `vks.vngcloud.vn/healthcheck-protocol` được thiết lập là `HTTP` hoặc `HTTPS` .
- `vks.vngcloud.vn/healthcheck-path` :
 - Bắt buộc phải có trong spec của tài nguyên Service : Không bắt buộc
 - Ý nghĩa: Là đường dẫn sử dụng để kiểm tra sức khỏe của máy chủ back-end. Mặc định, giá trị của annotation này là `/` . Lưu ý rằng annotation này chỉ có tác dụng khi `vks.vngcloud.vn/healthcheck-protocol` được thiết lập là `HTTP` hoặc `HTTPS` .
- `vks.vngcloud.vn/healthcheck-http-version` :
 - Bắt buộc phải có trong spec của tài nguyên Service : Không bắt buộc
 - Ý nghĩa: Là phiên bản HTTP sử dụng để kiểm tra sức khỏe của máy chủ back-end. Mặc định, giá trị của annotation này là `1.0` . Ngoài ra bạn cũng có thể sử dụng giá trị `1.1` khi cần thiết. Lưu ý rằng annotation này chỉ có tác dụng khi `vks.vngcloud.vn/healthcheck-protocol` được thiết lập là `HTTP` hoặc `HTTPS` .
- `vks.vngcloud.vn/healthcheck-http-domain-name` :
 - Bắt buộc phải có trong spec của tài nguyên Service : Không bắt buộc
 - Ý nghĩa: Là tên miền sử dụng để kiểm tra sức khỏe của máy chủ back-end. Mặc định, giá trị của annotation này là **RỎNG**. Lưu ý rằng annotation này chỉ có tác dụng khi `vks.vngcloud.vn/healthcheck-protocol` được thiết lập là `HTTP` hoặc `HTTPS` .
- `vks.vngcloud.vn/healthcheck-port` :
 - Bắt buộc phải có trong spec của tài nguyên Service : Không bắt buộc
 - Ý nghĩa: Là cổng sử dụng để kiểm tra sức khỏe của máy chủ back-end. Mặc định, giá trị của annotation này sẽ lấy theo trường `port` của tài nguyên Service cần dùng NLB này. Lưu ý rằng annotation này chỉ có tác dụng khi

`vks.vngcloud.vn/healthcheck-protocol` được thiết lập là HTTP hoặc HTTPS .

- `vks.vngcloud.vn/success-codes` :
 - Bắt buộc phải có trong spec của tài nguyên Service : Không bắt buộc
 - Ý nghĩa: Là mã trạng thái HTTP sẽ được coi là thành công khi kiểm tra sức khỏe của máy chủ back-end. Mặc định, giá trị của annotation này là `200` . Lưu ý rằng annotation này chỉ có tác dụng khi `vks.vngcloud.vn/healthcheck-protocol` được thiết lập là HTTP hoặc HTTPS .
- `vks.vngcloud.vn/idle-timeout-client` :
 - Bắt buộc phải có trong spec của tài nguyên Service : Không bắt buộc
 - Ý nghĩa: Là thời gian tối đa mà một kết nối từ client có thể không hoạt động trước khi bị đóng. Mặc định, giá trị của annotation này là `50` giây.
- `vks.vngcloud.vn/idle-timeout-member` :
 - Bắt buộc phải có trong spec của tài nguyên Service : Không bắt buộc
 - Ý nghĩa: Là thời gian tối đa mà một kết nối từ máy chủ back-end có thể không hoạt động trước khi bị đóng. Mặc định, giá trị của annotation này là `50` giây.
- `vks.vngcloud.vn/idle-timeout-connection` :
 - Bắt buộc phải có trong spec của tài nguyên Service : Không bắt buộc
 - Ý nghĩa: Là thời gian tối đa mà một kết nối giữa NLB và máy chủ back-end có thể không hoạt động trước khi bị đóng. Mặc định, giá trị của annotation này là `5` giây.
- `vks.vngcloud.vn/pool-algorithm` :
 - Bắt buộc phải có trong spec của tài nguyên Service : Không bắt buộc
 - Ý nghĩa: Là thuật toán cân bằng tải sẽ được sử dụng cho NLB. Mặc định, giá trị của annotation này là `ROUND_ROBIN` . Ngoài ra, một vài giá trị khác mà bạn có thể sử dụng là `SOURCE_IP` , `LEAST_CONNECTIONS` .
- `vks.vngcloud.vn/target-node-labels` :
 - Bắt buộc phải có trong spec của tài nguyên Service : Không bắt buộc
 - Ý nghĩa: Mặc định **VNGCloud LoadBalancer Controller** sẽ chọn toàn bộ các node trong cluster để thêm vào pool của NLB. Tuy nhiên, nếu bạn muốn chỉ chọn một số node cụ thể, bạn có thể sử dụng annotation này. Ví dụ:
`vks.vngcloud.vn/target-node-labels: node-role.kubernetes.io/worker=true` . Lưu ý rằng bạn cũng có thể cấu hình nhiều node cho NLB của bạn bằng cách sử dụng dấu phẩy , để ngăn cách giữa các node. Ví dụ:
`vks.vngcloud.vn/target-node-labels: node-role.kubernetes.io/worker=true,node-role.kubernetes.io/worker=false` .
- `vks.vngcloud.vn/enable-proxy-protocol` :
 - Bắt buộc phải có trong spec của tài nguyên Service : Không bắt buộc
 - Ý nghĩa: Mặc định, giá trị của annotation này là **RÕNG**. Bạn cần chỉ định danh sách các **Service Port Name** (trong tài nguyên Service , bạn có thể chỉ định nhiều Service Port, mỗi Service Port này bạn cần cấu hình thuộc tính *name* cho chúng để sử dụng tính năng này) mà Proxy Protocol sẽ được kích hoạt. Ví dụ:
`vks.vngcloud.vn/enable-proxy-protocol: service_port_name_1,service_port_name_2` . Hoặc nếu bạn muốn kích hoạt Proxy Protocol cho tất cả các Service Port, bạn có thể sử dụng giá trị `*` . Ví dụ: `vks.vngcloud.vn/enable-proxy-protocol: "*"` .

2.6. Những hạn chế của NLB trên VKS

- Một NLB có thể được sử dụng chung cho nhiều cluster nhưng phải đảm bảo các cluster này có chung một **VPC**.
- Một NLB sẽ bị giới hạn về số lượng pools, listeners, members. Chi tiết tại <https://docs.vngcloud.vn/pages/viewpage.action?pageId=59802094>.

2.7. Tài liệu tham khảo

- Bạn có thể tham khảo tài liệu trên website của VNGCloud tại <https://docs.vngcloud.vn/vng-cloud-document/vn/vks/network/lam-viec-voi-network-load-balancing-nlb>.