

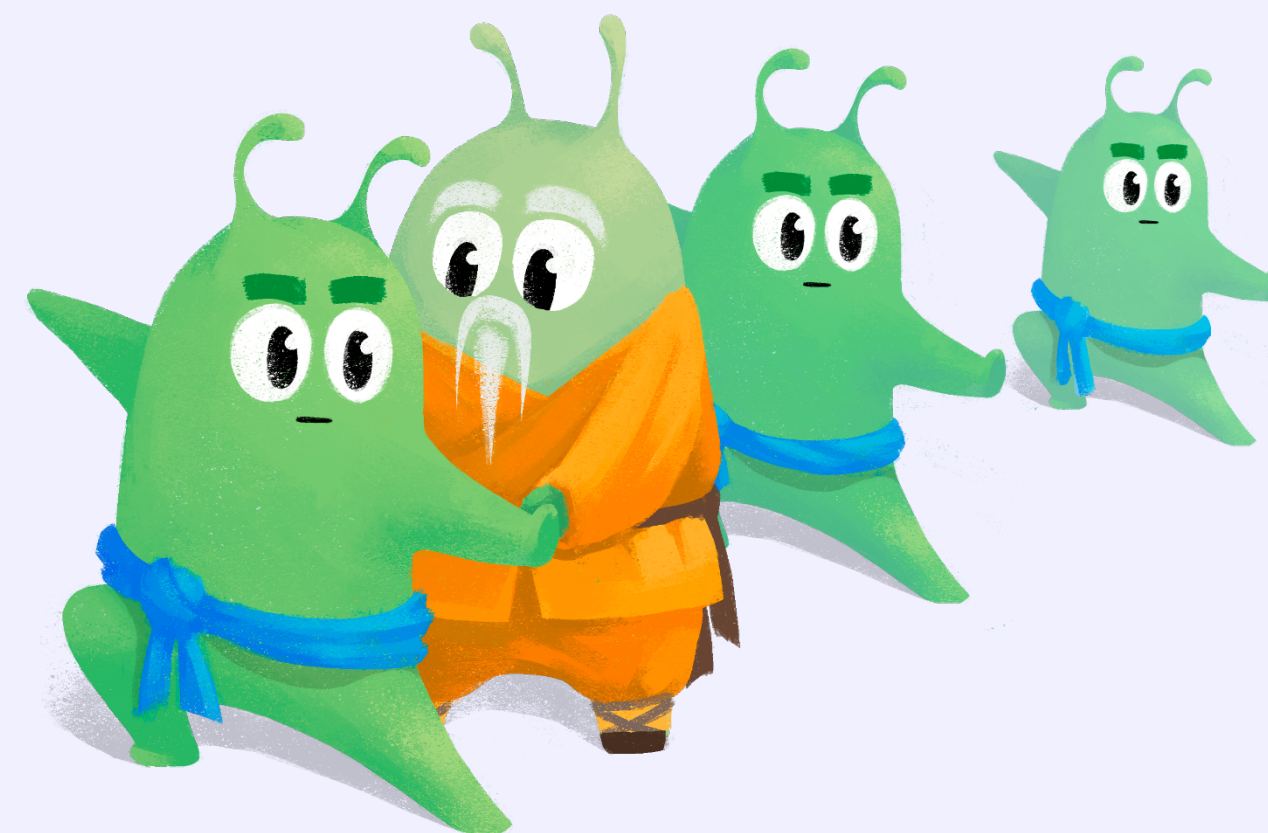


Урок 17. Horizontal Pod Autoscaler

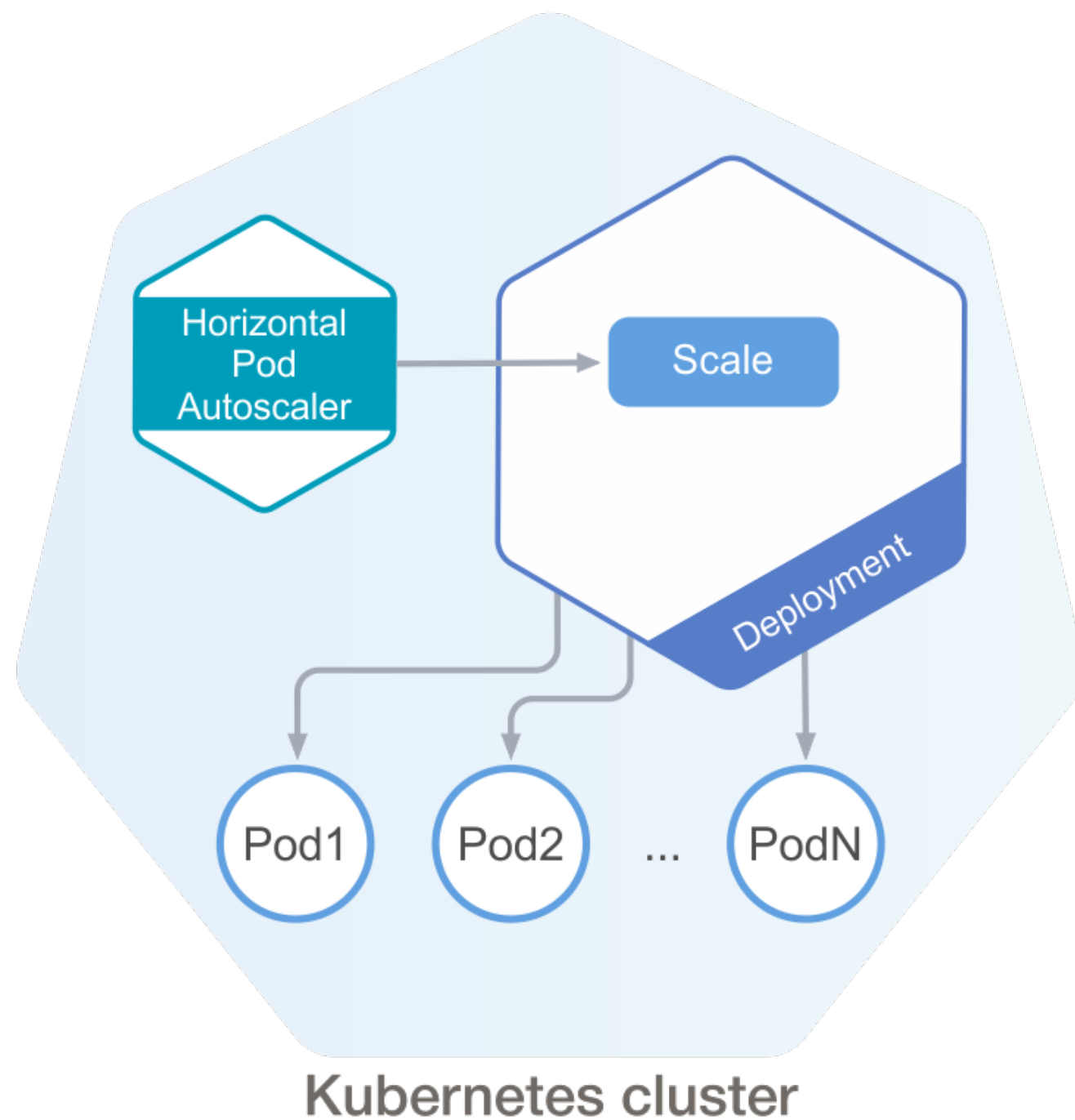
Павел Селиванов
Developer Advocate
VK Cloud Solutions

План

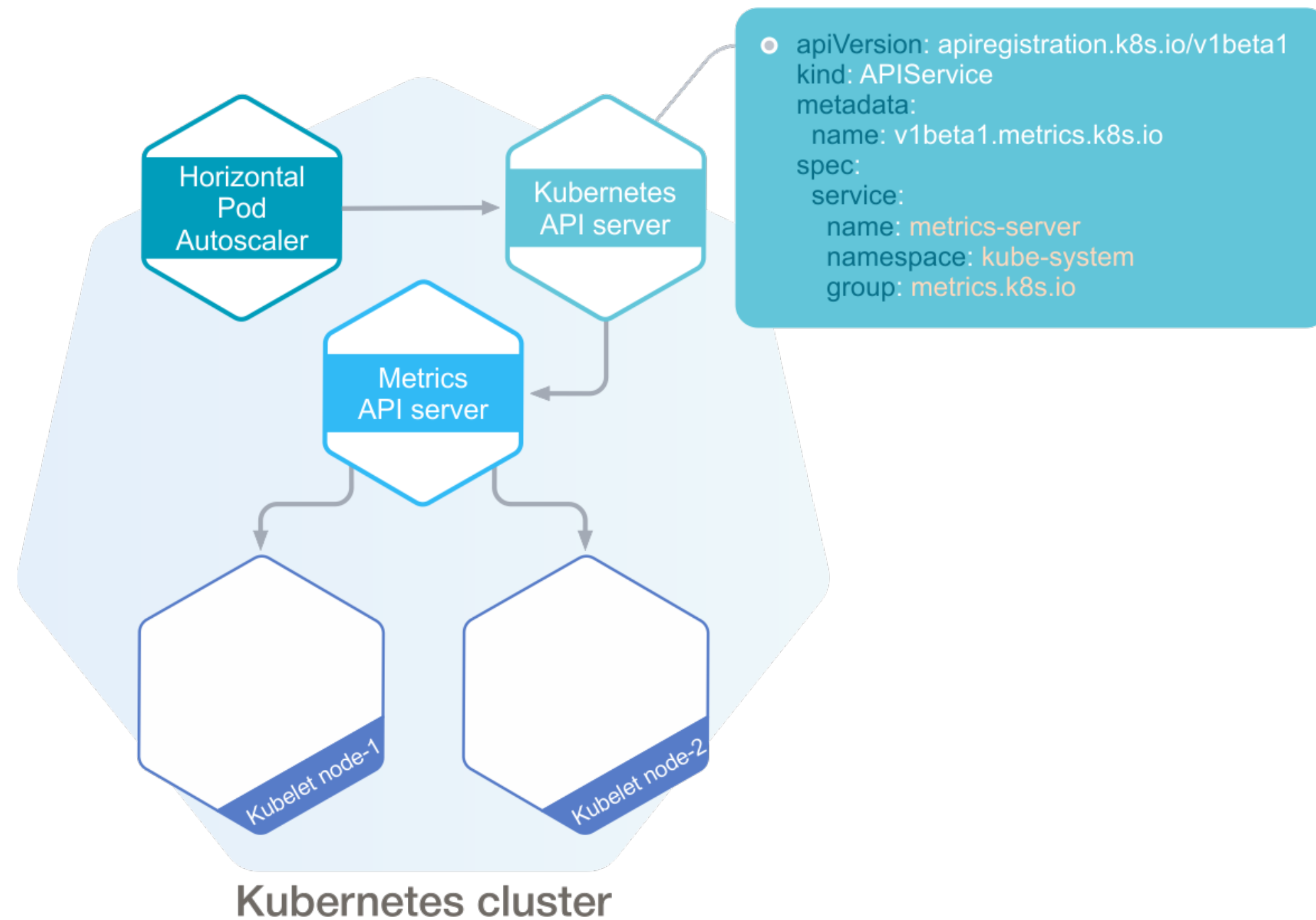
- HPA
- Metrics Server
- Дефолтные метрики
- Кастомные метрики
- Cluster Autoscaler



HPA



Metrics Server



HPA v1

spec:

maxReplicas: 5

minReplicas: 1

scaleTargetRef:

apiVersion: apps/v1

kind: Deployment

name: php-apache

targetCPUUtilizationPercentage: 50

Пределы
скейлинга

Что скейлим

При каких
условиях
скейлим

HPA v2

spec:

maxReplicas: 5

minReplicas: 1

scaleTargetRef:

apiVersion: apps/v1

kind: Deployment

name: php-apache

metrics:

– type: Resource

resource:

name: cpu

target:

type: Utilization

averageUtilization: 50

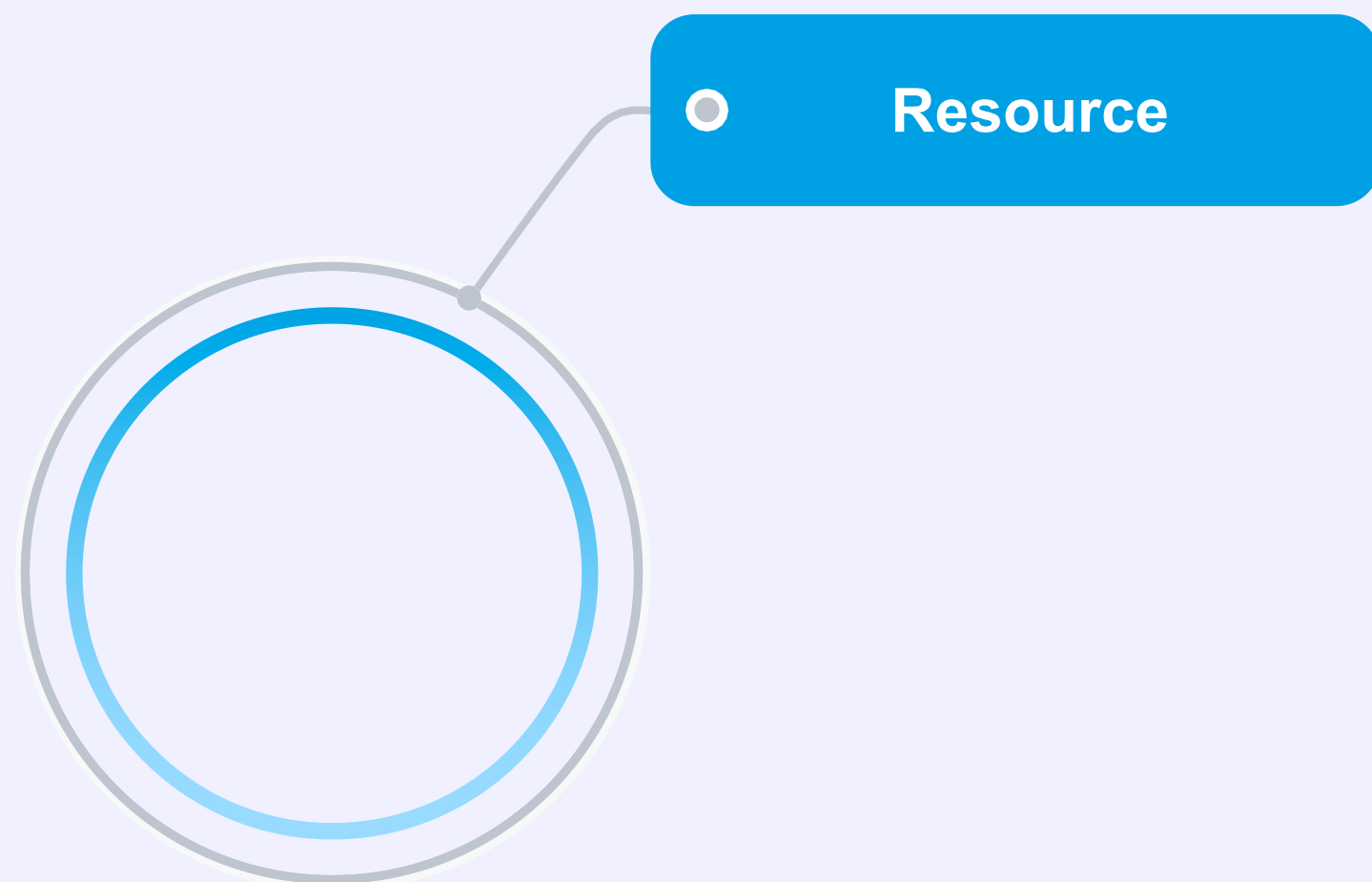
Пределы
скейлинга

Что скейлим

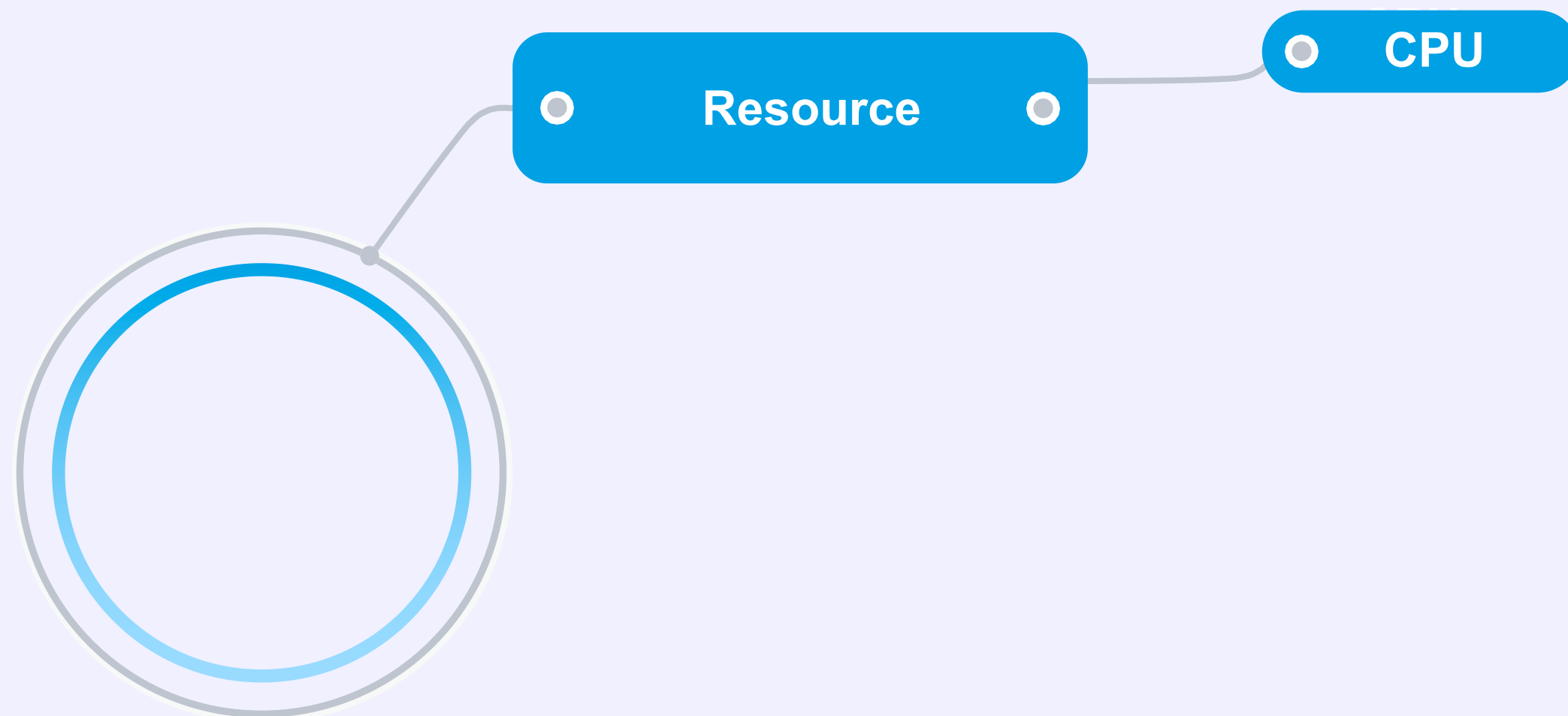
На какую
метрику
смотрим

При каких
условиях
скейлим (%)

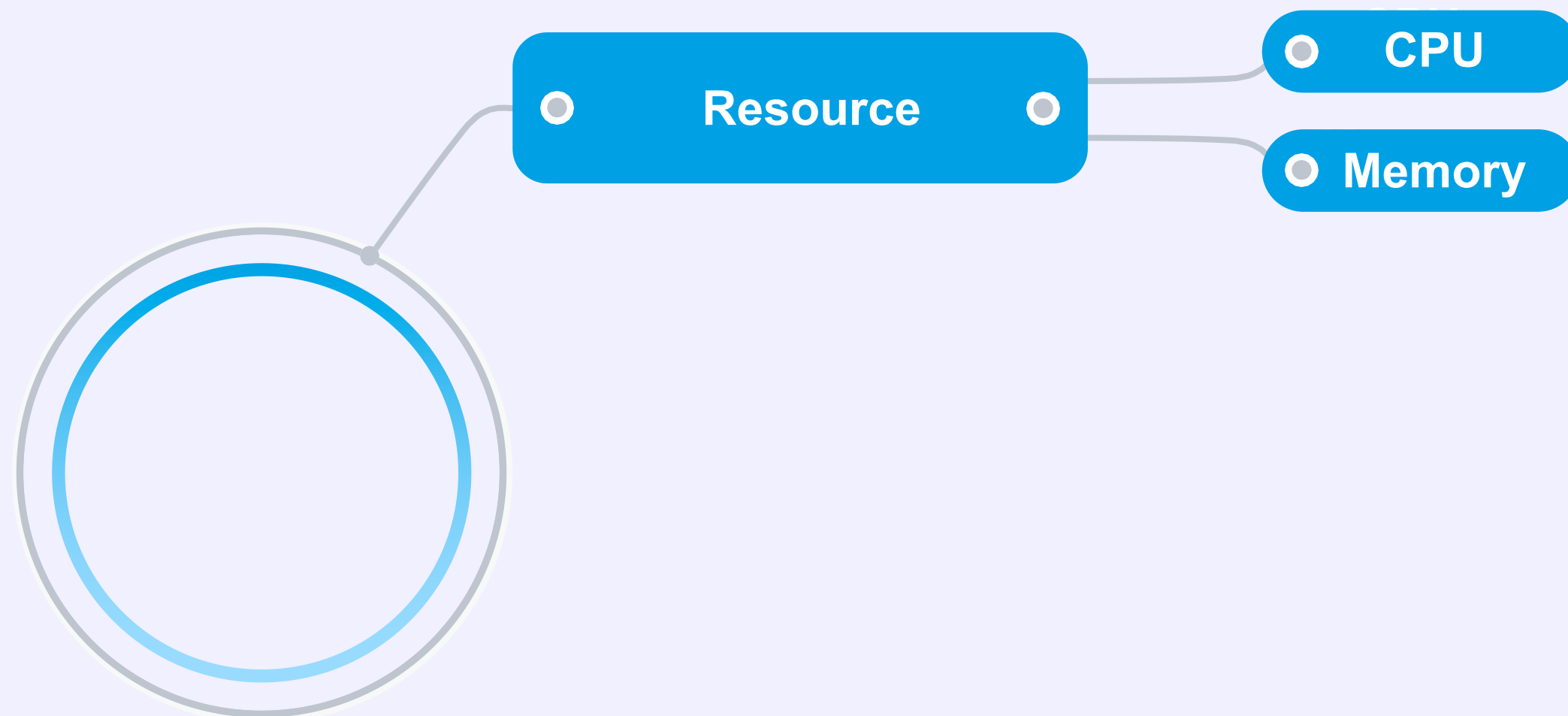
HPA v2 объекты метрик



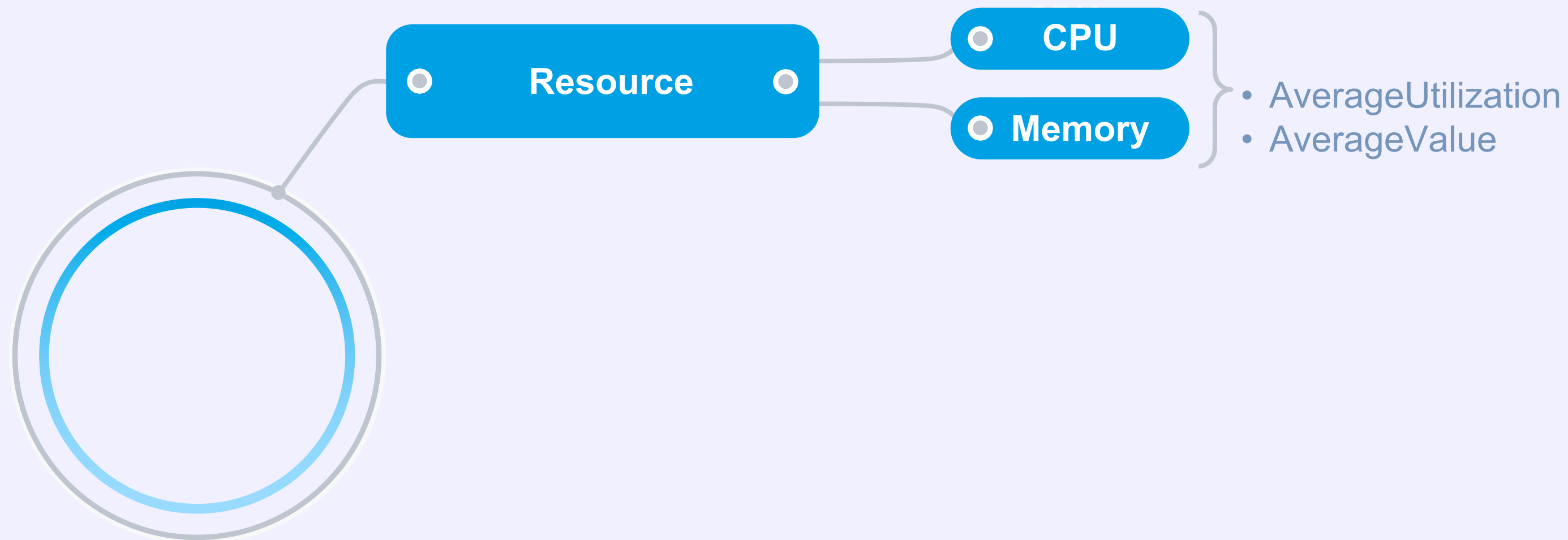
HPA v2 объекты метрик



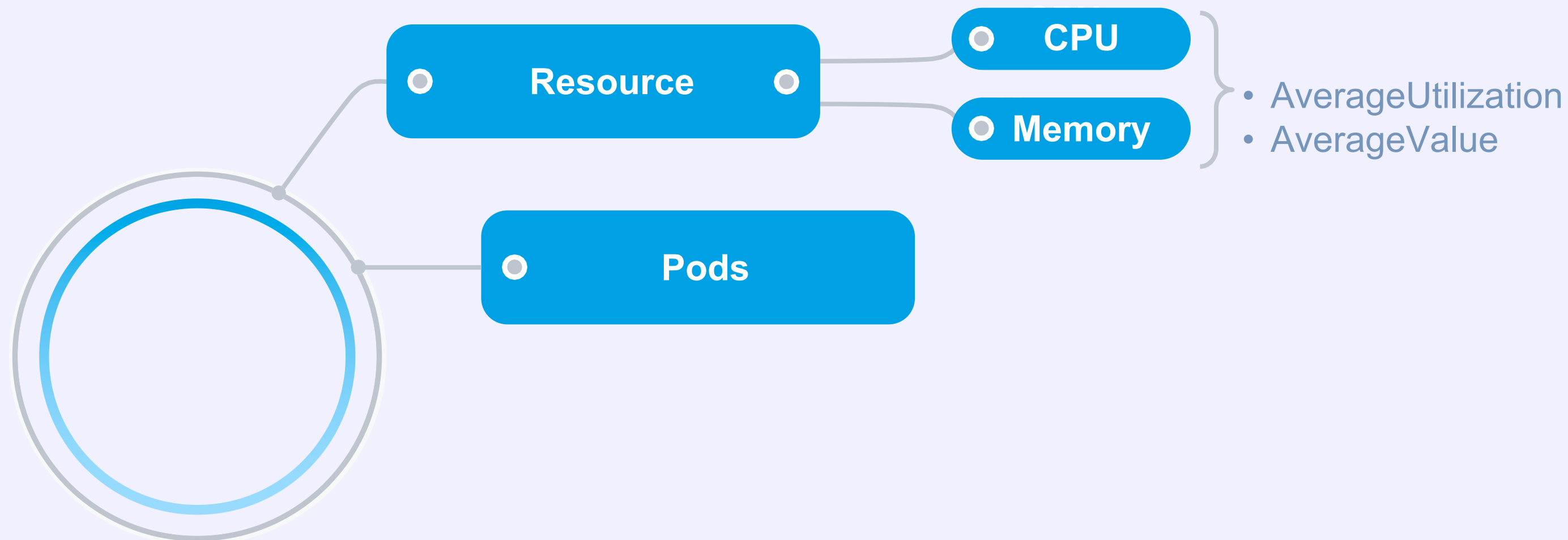
HPA v2 объекты метрик



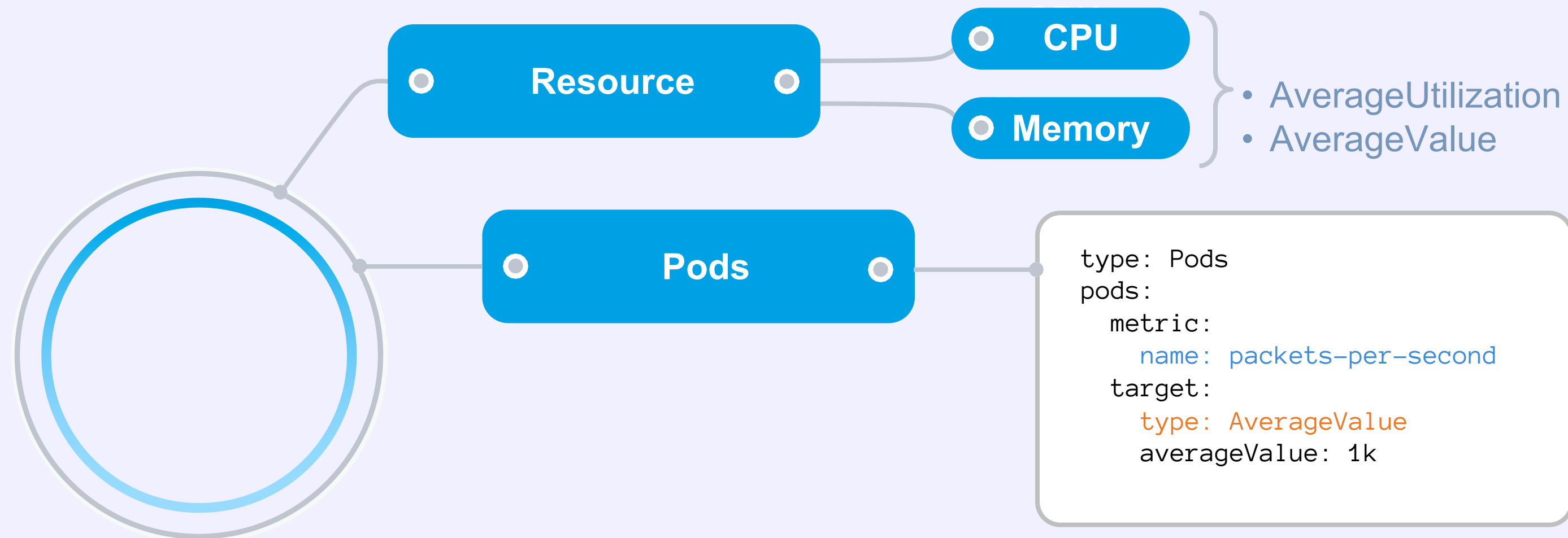
НРА v2 объекты метрик



НРА v2 объекты метрик



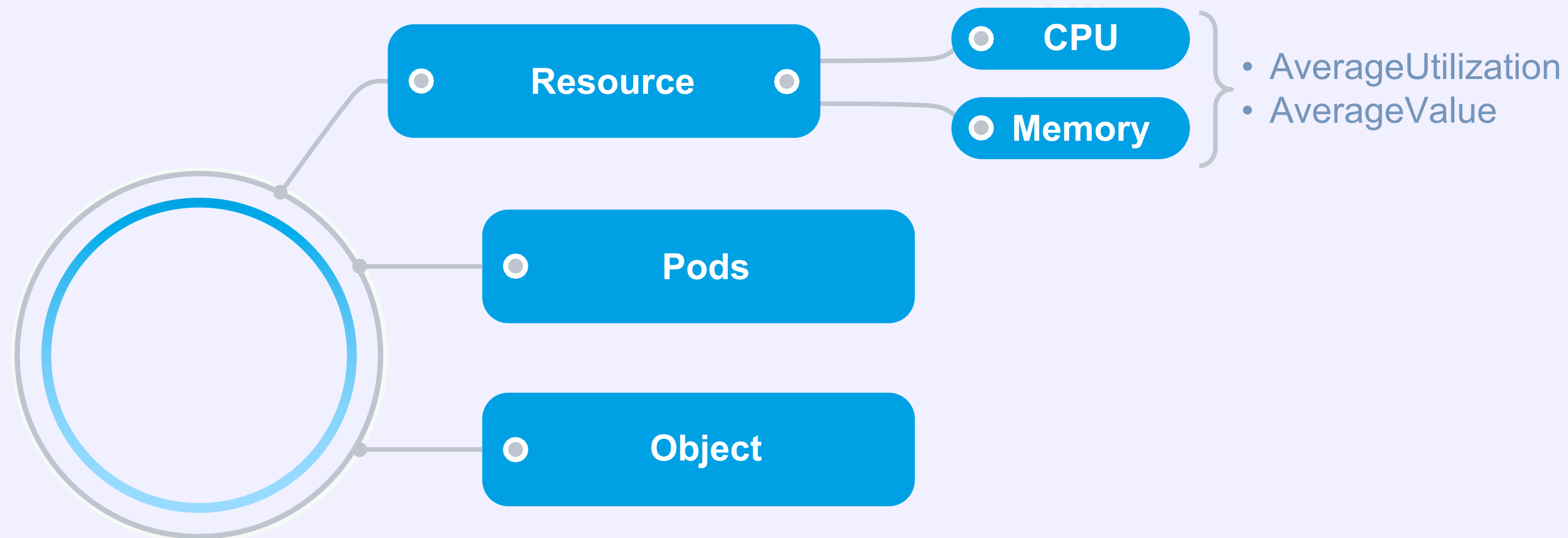
НРА v2 объекты метрик



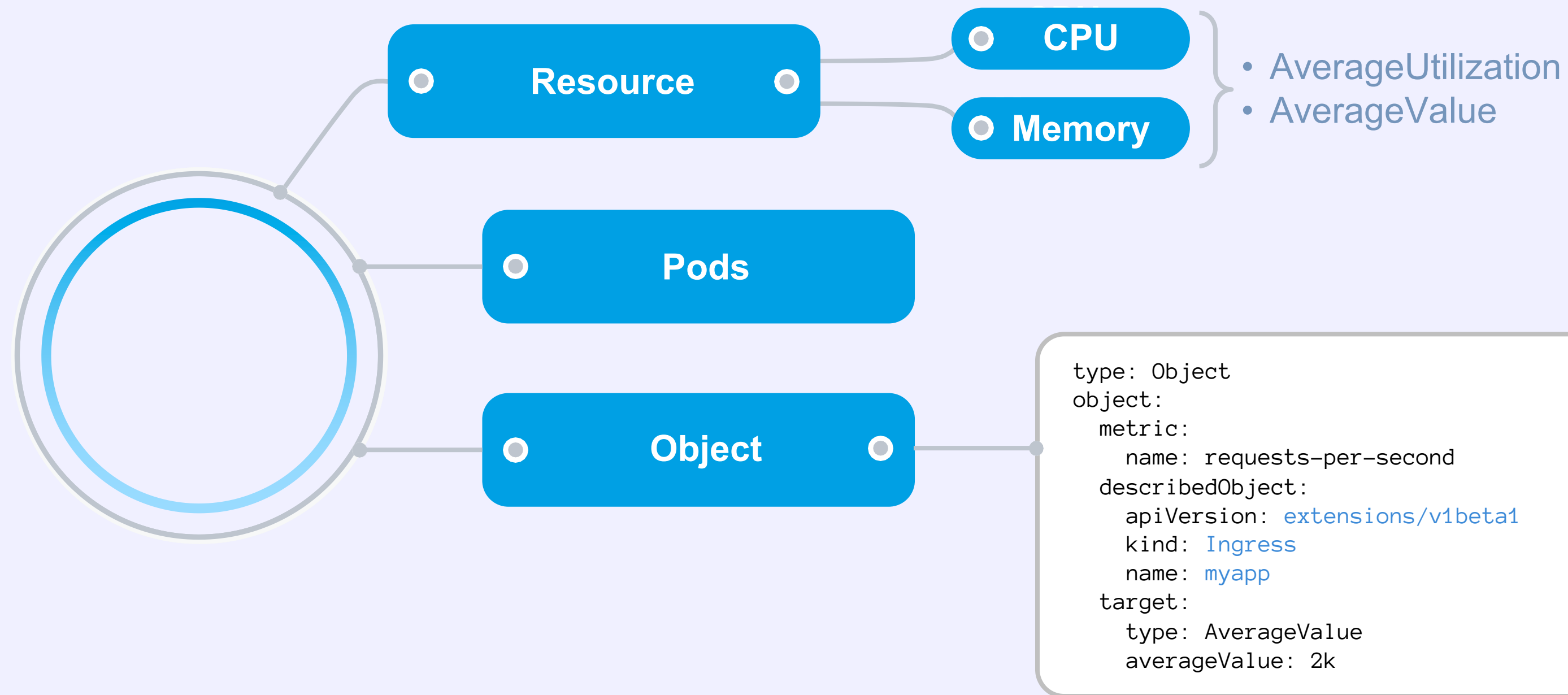
13



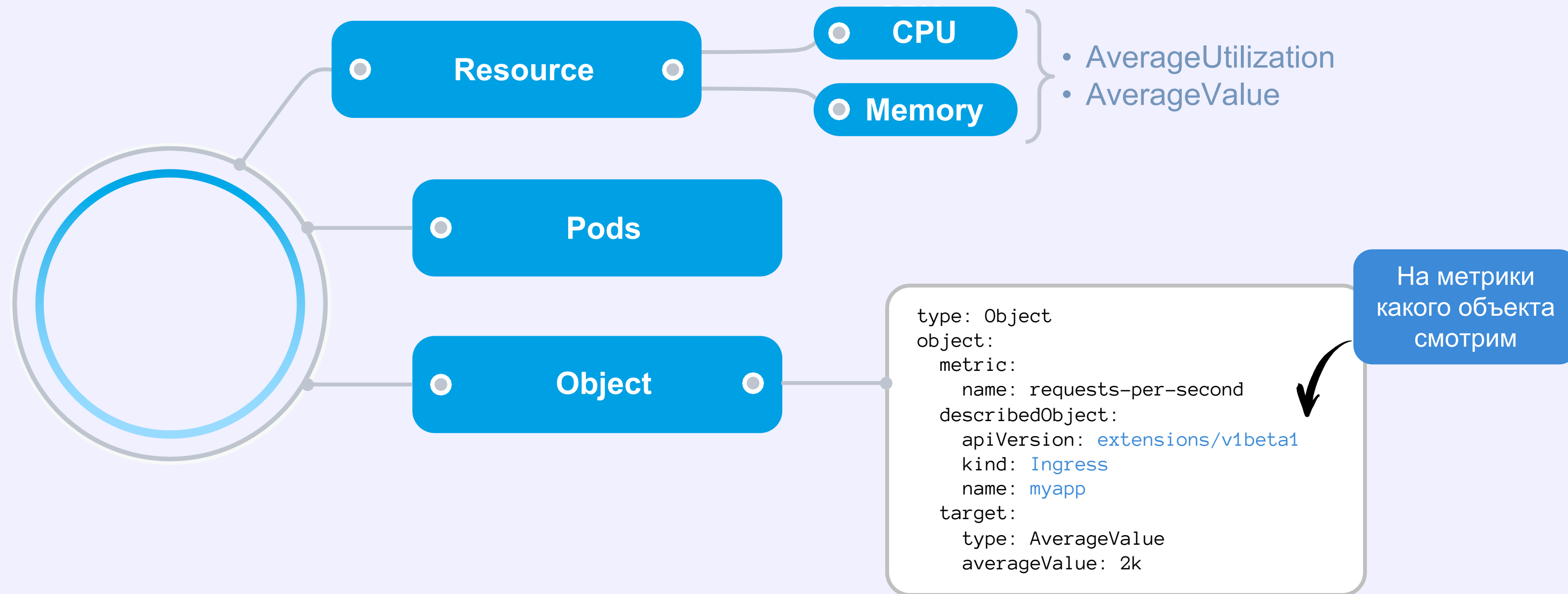
НРА v2 объекты метрик



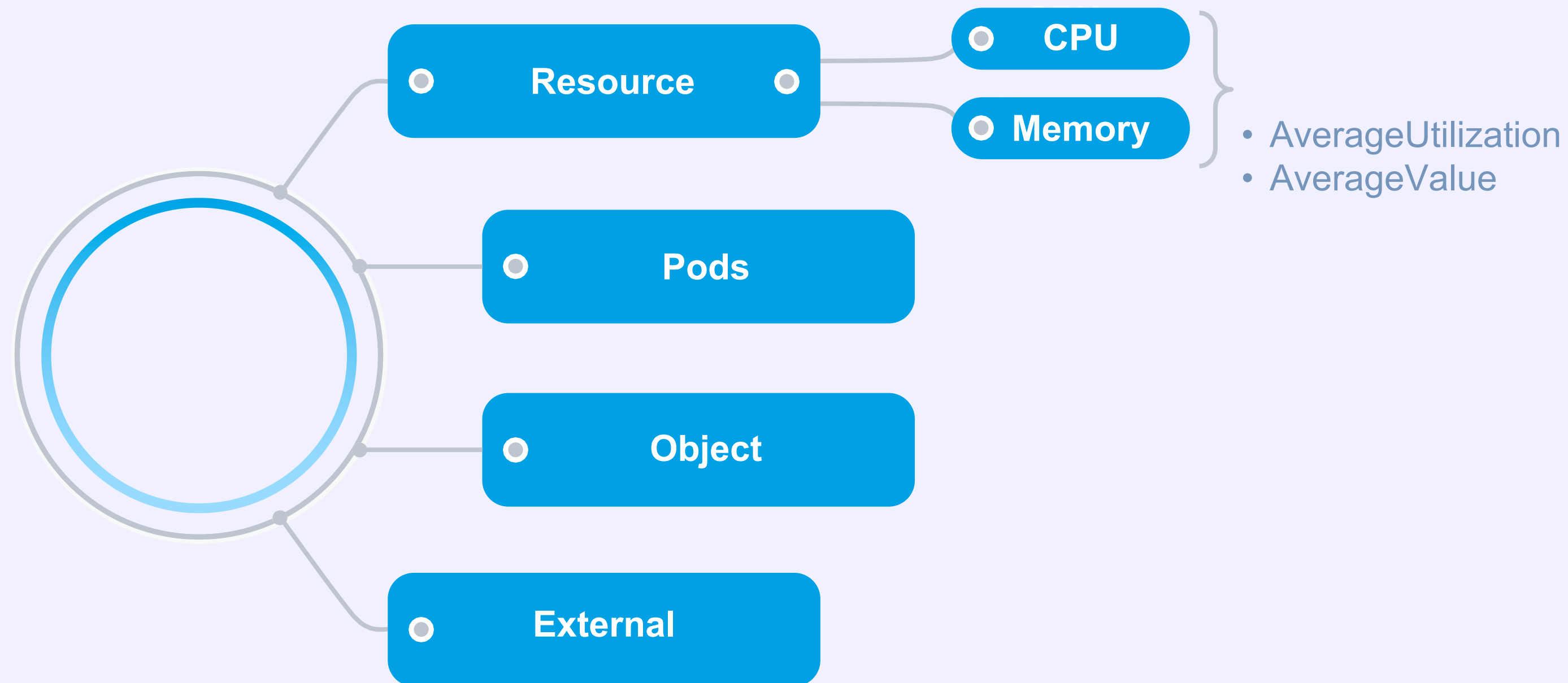
НРА v2 объекты метрик



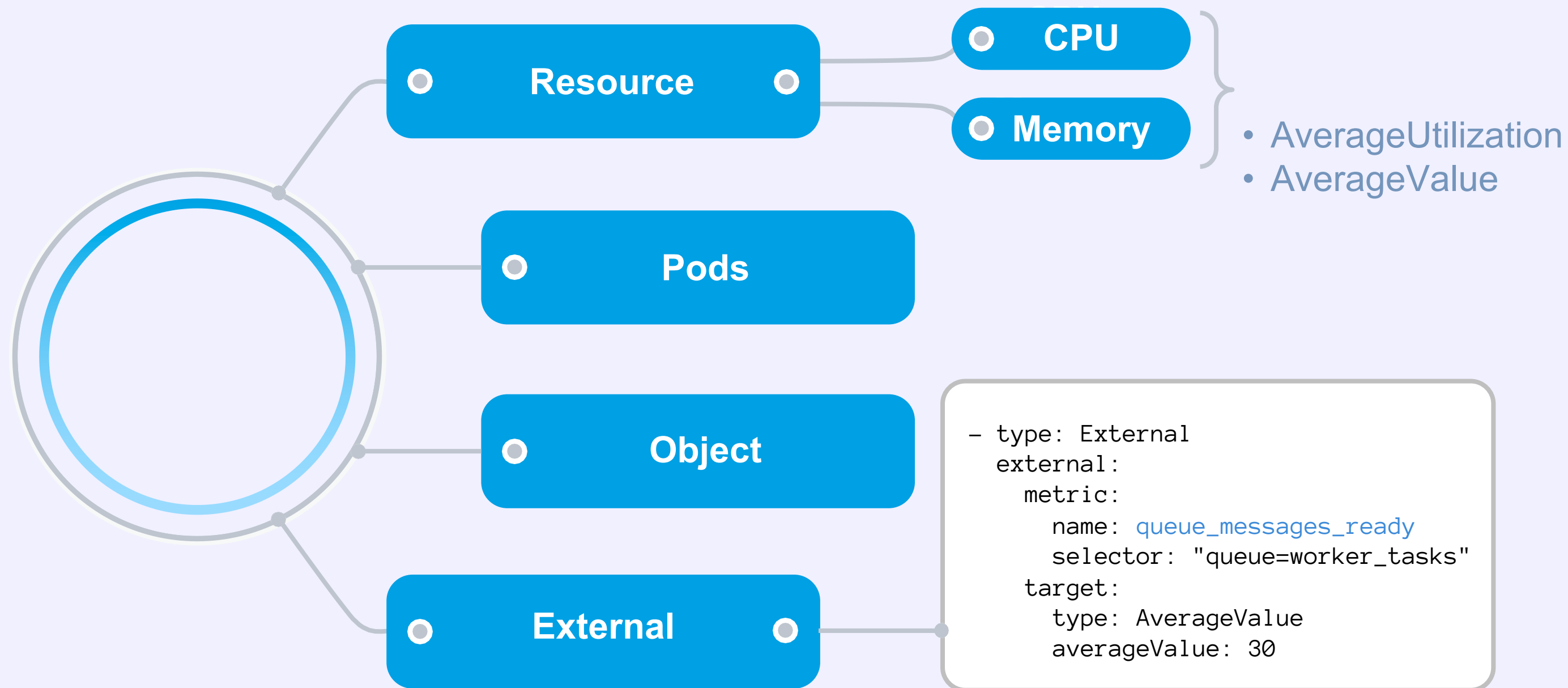
НРА v2 объекты метрик



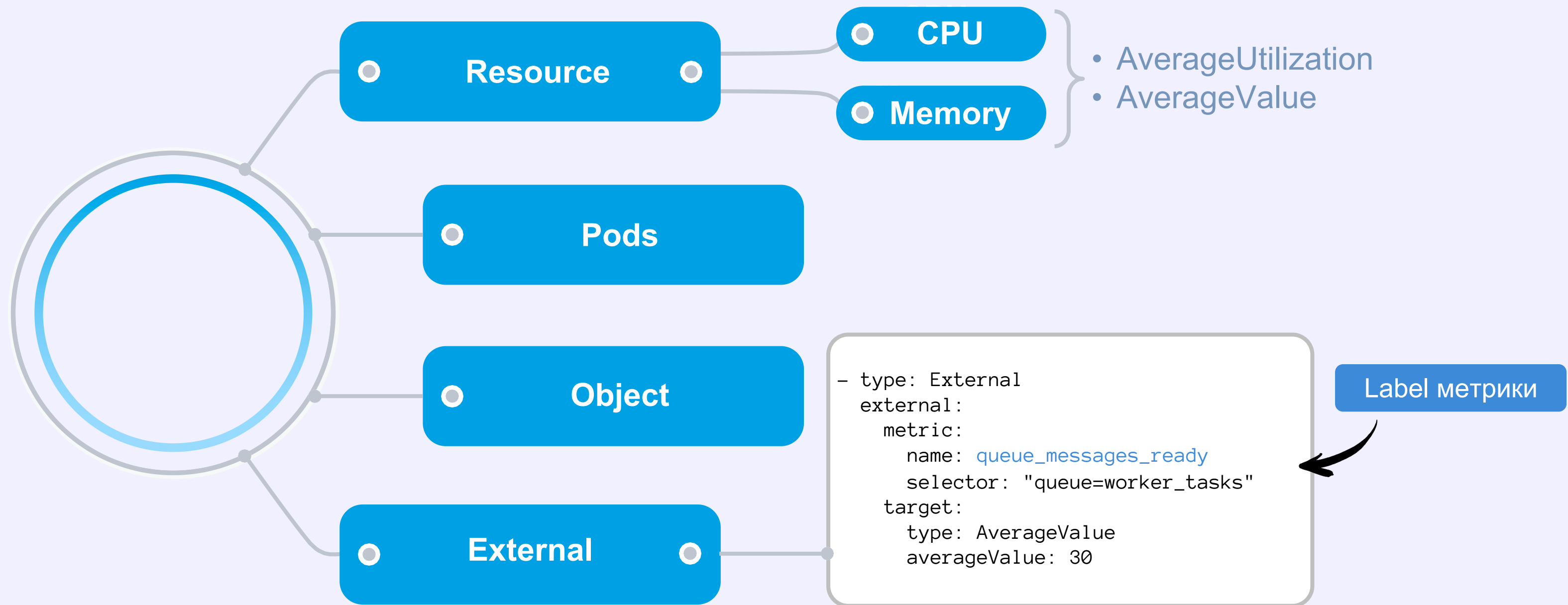
НРА v2 объекты метрик



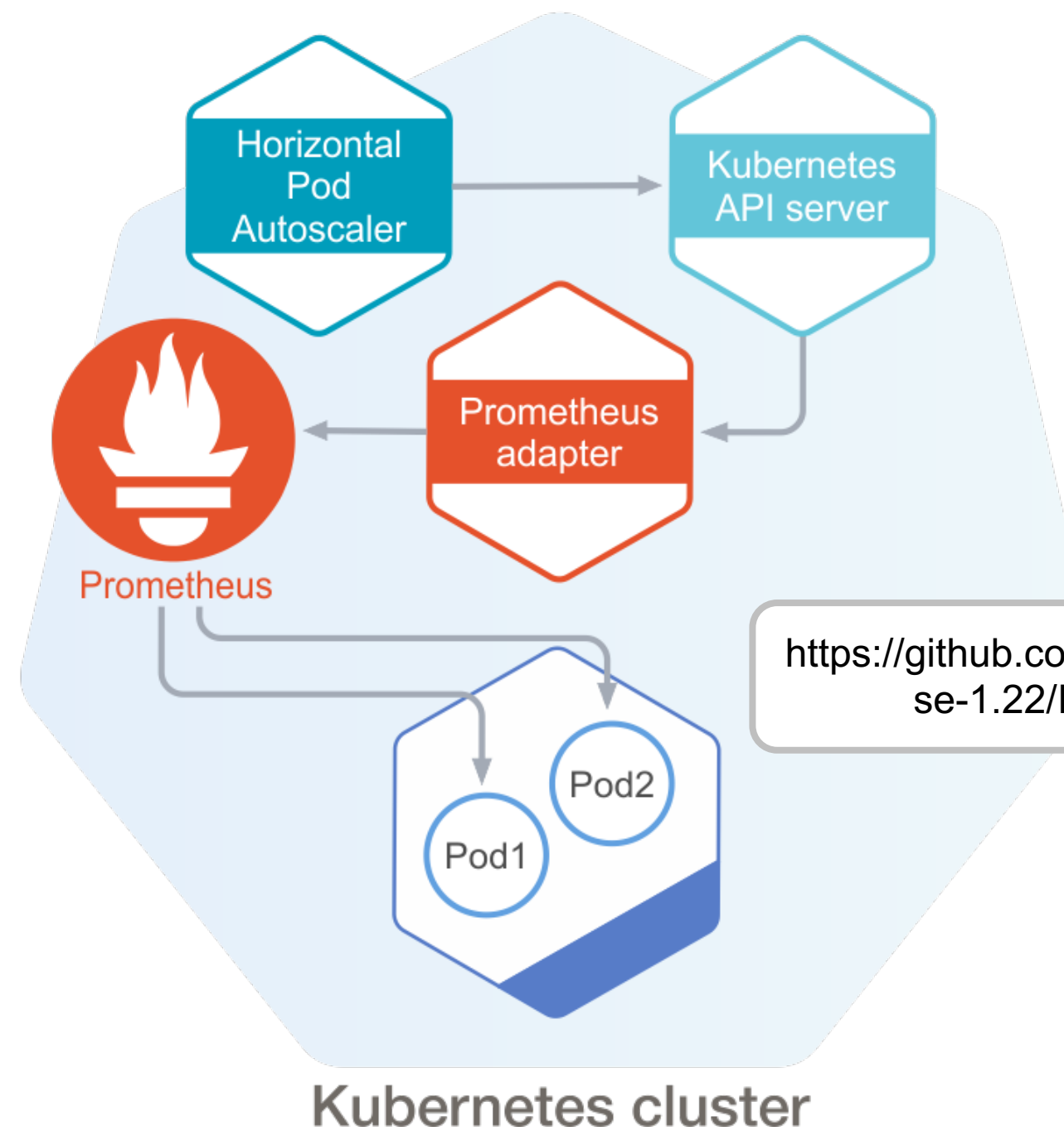
НРА v2 объекты метрик



НРА v2 объекты метрик



Реализации metrics server



Cluster Autoscaler

- Автоматически добавляет новые ноды в кластер
- Реализуется на стороне провайдера
- Может быть использован не только совместно с
HРА

- Динамические раннеры CI/CD
- Динамические стэнды



Экономия средств



VK Cloud Solutions

Почитать статью:

«Три уровня
автомасштабирования
в Kubernetes: как их
эффективно
использовать»

Подписаться на канал t.me/k8s_vk



Пообщаться в чате

с экспертами
VK Cloud Solutions
по K8s и не только:
t.me/vk_cs_chat

