

Четвертое ДЗ

Машинное обучение в продакшене

Всем привет!

Представляю вам последнее дз, оно направлено на знакомство в kubernetes. Задание, как и обещал, является логическим продолжением второго.

В каждом пункте, кроме 0 и 1, вам потребуется поднятый kubernetes кластер и утилита, которая помогает с ним взаимодействовать.

<https://kubernetes.io/docs/reference/kubectl/cheatsheet/>

Если вам понравилось, можете так же установить себе Lens -- та штука, для визуализации.

ветку назовите homework4, label -- hw4

0 Установите kubectl

1) Разверните kubernetes

Вы можете развернуть его в облаке:

- <https://cloud.google.com/kubernetes-engine>
- <https://mcs.mail.ru/containers/>
- <https://cloud.yandex.ru/services/managed-kubernetes>

Либо воспользоваться локальной инсталляцией

- <https://kind.sigs.k8s.io/docs/user/quick-start/>
- <https://minikube.sigs.k8s.io/docs/start/>

Напишите, какой способ вы выбрали.

Убедитесь, с кластер поднялся (kubectl cluster-info)

(5 баллов)

2) Напишите простой pod manifests для вашего приложения, назовите его online-inference-pod.yaml (<https://kubernetes.io/docs/concepts/workloads/pods/>)

Задеплойте приложение в кластер (kubectl apply -f online-inference-pod.yaml), убедитесь, что все поднялось (kubectl get pods)

Приложите скриншот, где видно, что все поднялось

(4 балла)

2а) Пропишите requests/limits и напишите зачем это нужно в описание PR
закоммитьте файл online-inference-pod-resources.yaml

(2 балл)

3) Модифицируйте свое приложение так, чтобы оно стартовало не сразу(с задержкой секунд 20-30) и падало спустя минуты работы.

Добавьте liveness и readiness пробы, посмотрите что будет происходить.

Напишите в описании -- чего вы этим добились.

Закоммитьте отдельный манифест online-inference-pod-probes.yaml (и изменение кода приложения)

(3 балла)

Опубликуйте ваше приложение(из ДЗ 2) с тэгом v2

4) Создайте replicaset, сделайте 3 реплики вашего приложения.

(<https://kubernetes.io/docs/concepts/workloads/controllers/replicaset/>)

Ответьте на вопрос, что будет, если сменить докер образа в манифесте и одновременно с этим

а) уменьшить число реплик

б) увеличить число реплик.

Поды с какими версиями образа будут внутри будут в кластере?

(3 балла)

Закоммитьте online-inference-replicaset.yaml

5) Опишите деплоймент для вашего приложения.

(<https://kubernetes.io/docs/concepts/workloads/controllers/deployment/>)

Играя с параметрами деплоя(maxSurge, maxUnavailable), добейтесь ситуации, когда при деплое новой версии

а) Есть момент времени, когда на кластере есть как все старые поды, так и все новые (опишите эту ситуацию) (закоммититьте файл online-inference-deployment-blue-green.yaml)

б) одновременно с поднятием новых версии, гасятся старые (закоммитите файл online-inference-deployment-rolling-update.yaml)

(3 балла)

Бонусные активности:

Установить helm и оформить helm chart, включить в состав чарта ConfigMap и Service. -- 5 баллов

DEADLINE: 21 июня

ОПРОС ПРО ЗАНЯТИЕ: <https://forms.gle/KjxPvtbnaV9VeEui7>

Если вы это читаете, то вы круты=)

Оцените ДЗ

Очень сложно

Сложно

Норм

Легко

Воздержаться

