Щербань Георгий
Разработка и развертывание ИИ-приложения в Kubernetes

1. Введение

Цель проекта — разработать и развернуть в локальном Kubernetes-кластере (Minikube) контейнеризированное Java-приложение с интегрированной ИИ-логикой анализа тональности текста.

В рамках работы были выполнены следующие задачи:

разработка REST API на Spring Boot

контейнеризация с помощью Docker

развертывание в Minikube (Deployment + Service + Ingress)

настройка мониторинга (Prometheus + Actuator)

настройка горизонтального автоскейлинга HPA

проведение нагрузочного тестирования

анализ трендов в области ИИ, контейнеризации и оркестрации

Используемый стек:

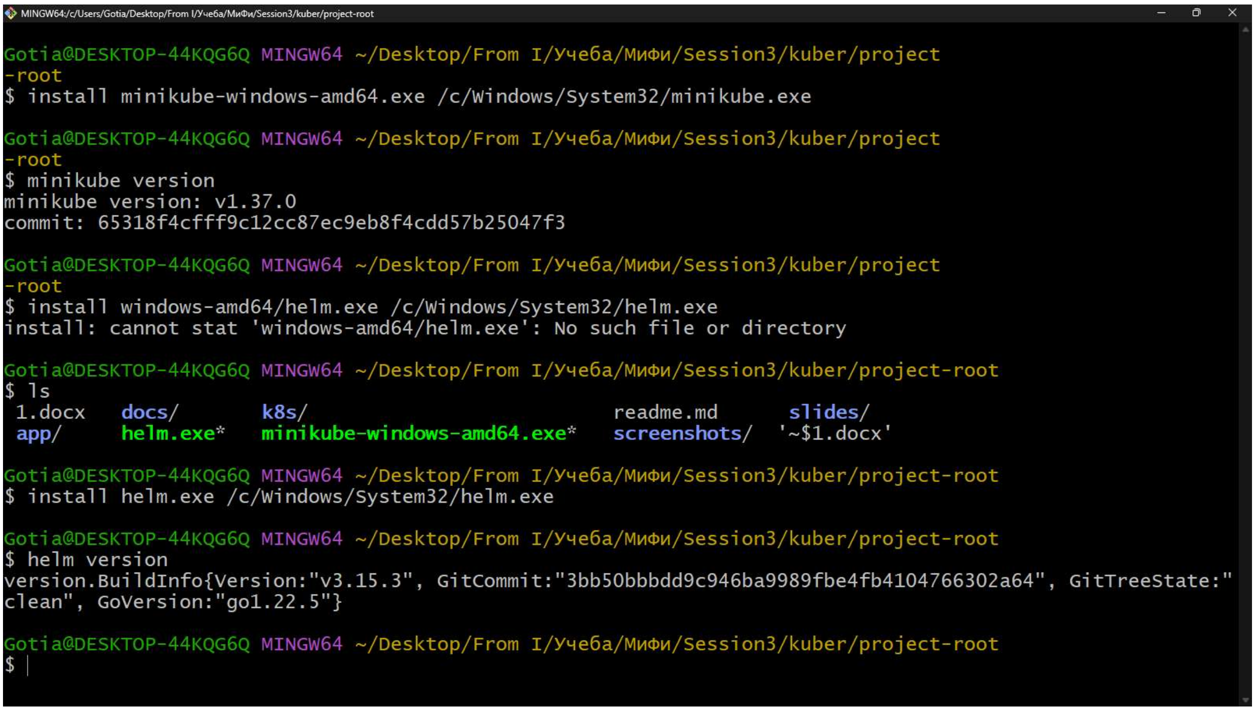
1. Spring Boot 3.3, Java 17
2. Docker, multi-stage build
3. Minikube (драйвер Docker)
4. kubectl, helm, metrics-server
5. Horizontal Pod Autoscaler (HPA)
6. Nginx Ingress Controller

2. Развёртывание инфраструктуры Minikube

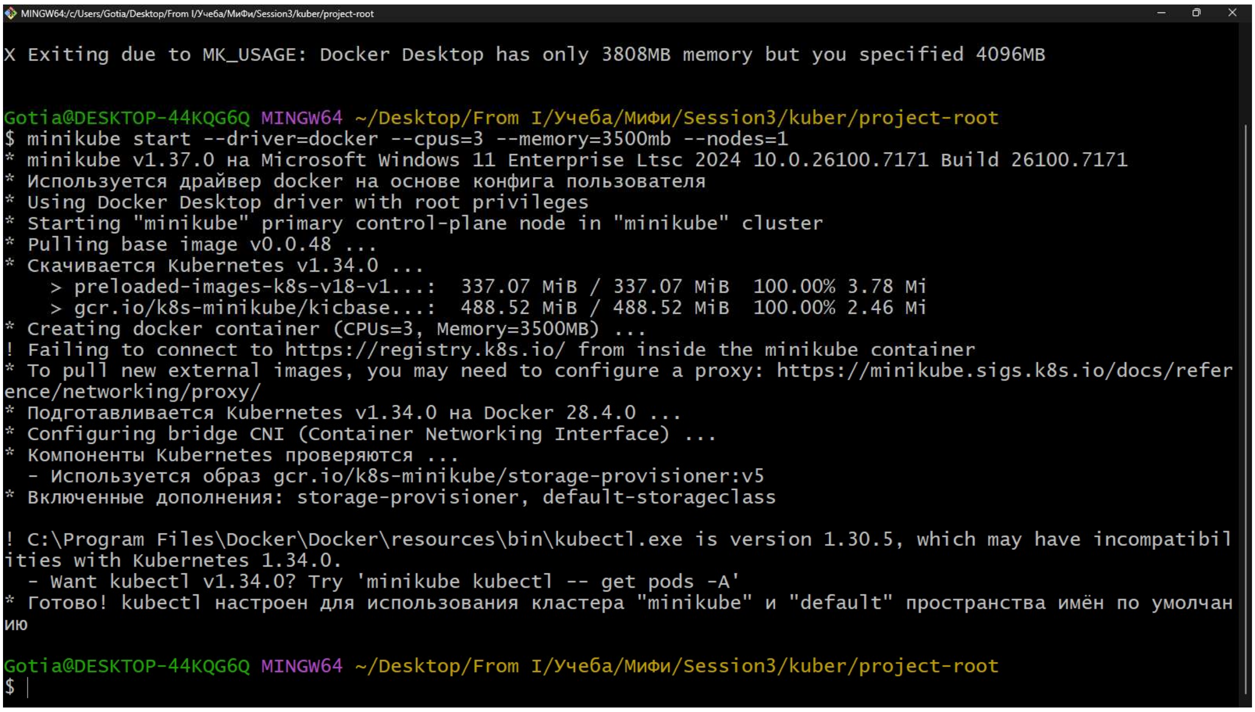
2.1 Установка и запуск кластера

Кластер Minikube был развёрнут на Windows 11 с драйвером Docker.

minikube start --driver=docker --cpus=3 --memory=3500mb --nodes=1


MINGW64:/c/Users/Gotia/Desktop/From I/Учеба/МиФи/Session3/kuber/project-root

```
Gotia@DESKTOP-44KQG6Q MINGW64 ~/Desktop/From I/Учеба/Мифи/Session3/kuber/project
-root
$ install minikube-windows-amd64.exe /c/Windows/System32/minikube.exe

Gotia@DESKTOP-44KQG6Q MINGW64 ~/Desktop/From I/Учеба/Мифи/Session3/kuber/project
-root
$ minikube version
minikube version: v1.37.0
commit: 65318f4cfff9c12cc87ec9eb8f4cdd57b25047f3

Gotia@DESKTOP-44KQG6Q MINGW64 ~/Desktop/From I/Учеба/Мифи/Session3/kuber/project
-root
$ install windows-amd64/helm.exe /c/Windows/System32/helm.exe
install: cannot stat 'windows-amd64/helm.exe': No such file or directory

Gotia@DESKTOP-44KQG6Q MINGW64 ~/Desktop/From I/Учеба/Мифи/Session3/kuber/project-root
$ ls
 1.docx   docs/      k8s/                          readme.md     slides/
 app/     helm.exe*  minikube-windows-amd64.exe*   screenshots/  '~$1.docx'

Gotia@DESKTOP-44KQG6Q MINGW64 ~/Desktop/From I/Учеба/Мифи/Session3/kuber/project-root
$ install helm.exe /c/Windows/System32/helm.exe

Gotia@DESKTOP-44KQG6Q MINGW64 ~/Desktop/From I/Учеба/Мифи/Session3/kuber/project-root
$ helm version
version.BuildInfo{Version:"v3.15.3", GitCommit:"3bb50bbbdd9c946ba9989fbe4fb4104766302a64", GitTreeState:"
clean", GoVersion:"go1.22.5"}

Gotia@DESKTOP-44KQG6Q MINGW64 ~/Desktop/From I/Учеба/Мифи/Session3/kuber/project-root
$ |
```


MINGW64:/c/Users/Gotia/Desktop/From I/Учеба/МиФи/Session3/kuber/project-root

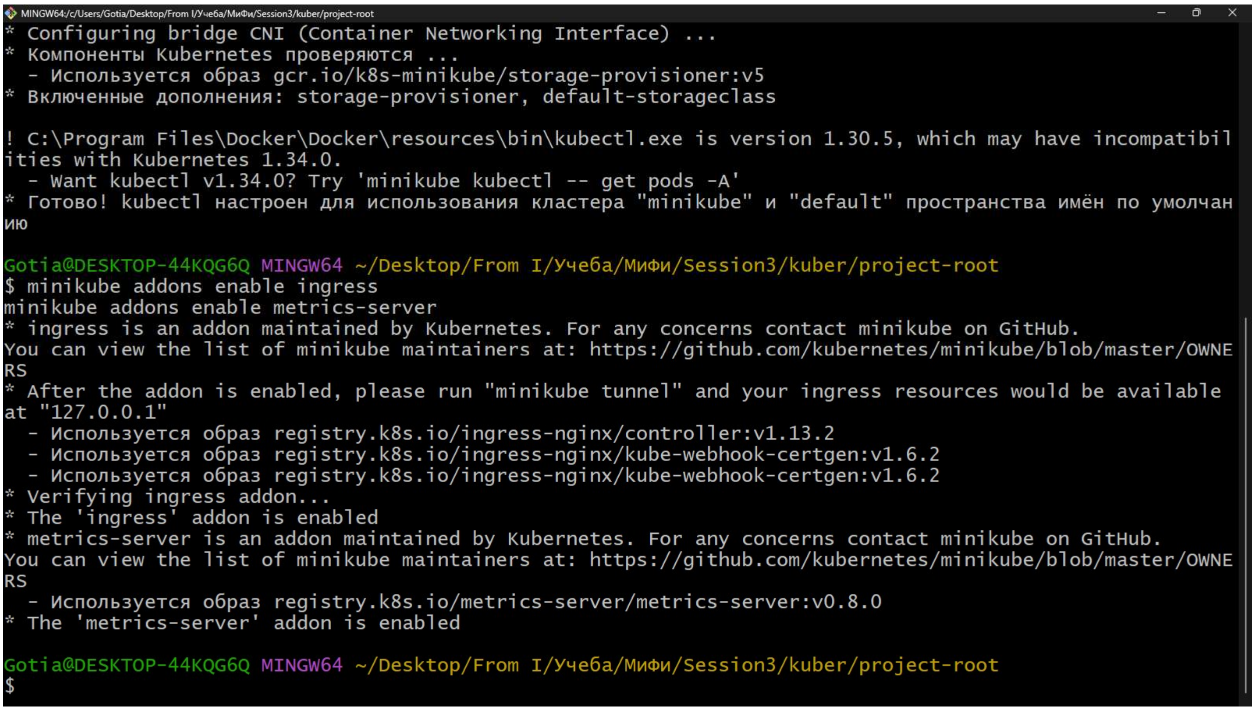
```
X Exiting due to MK_USAGE: Docker Desktop has only 3808MB memory but you specified 4096MB


Gotia@DESKTOP-44KQG6Q MINGW64 ~/Desktop/From I/Учеба/Мифи/Session3/kuber/project-root
$ minikube start --driver=docker --cpus=3 --memory=3500mb --nodes=1
* minikube v1.37.0 на Microsoft Windows 11 Enterprise Ltsc 2024 10.0.26100.7171 Build 26100.7171
* Используется драйвер docker на основе конфига пользователя
* Using Docker Desktop driver with root privileges
* Starting "minikube" primary control-plane node in "minikube" cluster
* Pulling base image v0.0.48 ...
* Скачивается Kubernetes v1.34.0 ...
    > preloaded-images-k8s-v18-v1...:  337.07 MiB / 337.07 MiB  100.00% 3.78 Mi
    > gcr.io/k8s-minikube/kicbase...:  488.52 MiB / 488.52 MiB  100.00% 2.46 Mi
* Creating docker container (CPUs=3, Memory=3500MB) ...
! Failing to connect to https://registry.k8s.io/ from inside the minikube container
* To pull new external images, you may need to configure a proxy: https://minikube.sigs.k8s.io/docs/refer
ence/networking/proxy/
* Подготавливается Kubernetes v1.34.0 на Docker 28.4.0 ...
* Configuring bridge CNI (Container Networking Interface) ...
* Компоненты Kubernetes проверяются ...
  - Используется образ gcr.io/k8s-minikube/storage-provisioner:v5
* Включенные дополнения: storage-provisioner, default-storageclass

! C:\Program Files\Docker\Docker\resources\bin\kubectl.exe is version 1.30.5, which may have incompatibil
ities with Kubernetes 1.34.0.
  - Want kubectl v1.34.0? Try 'minikube kubectl -- get pods -A'
* Готово! kubectl настроен для использования кластера "minikube" и "default" пространства имён по умолчан
ию

Gotia@DESKTOP-44KQG6Q MINGW64 ~/Desktop/From I/Учеба/Мифи/Session3/kuber/project-root
$ |
```

Ограничение памяти обусловлено настройками Docker Desktop, но конфигурации достаточно для HPA, ingress и сервиса.

2.2 Включение необходимых аддонов

minikube addons enable ingress

minikube addons enable metrics-server



MINGW64:/c/Users/Gotia/Desktop/From I/Учеба/МиФи/Session3/kuber/project-root

```
* Configuring bridge CNI (Container Networking Interface) ...
* Компоненты Kubernetes проверяются ...
  - Используется образ gcr.io/k8s-minikube/storage-provisioner:v5
* Включенные дополнения: storage-provisioner, default-storageclass

! C:\Program Files\Docker\Docker\resources\bin\kubectl.exe is version 1.30.5, which may have incompatibil
ities with Kubernetes 1.34.0.
  - Want kubectl v1.34.0? Try 'minikube kubectl -- get pods -A'
* Готово! kubectl настроен для использования кластера "minikube" и "default" пространства имён по умолчан
ию

Gotia@DESKTOP-44KQG6Q MINGW64 ~/Desktop/From I/Учеба/Мифи/Session3/kuber/project-root
$ minikube addons enable ingress
minikube addons enable metrics-server
* ingress is an addon maintained by Kubernetes. For any concerns contact minikube on GitHub.
You can view the list of minikube maintainers at: https://github.com/kubernetes/minikube/blob/master/OWNE
RS
* After the addon is enabled, please run "minikube tunnel" and your ingress resources would be available
at "127.0.0.1"
  - Используется образ registry.k8s.io/ingress-nginx/controller:v1.13.2
  - Используется образ registry.k8s.io/ingress-nginx/kube-webhook-certgen:v1.6.2
  - Используется образ registry.k8s.io/ingress-nginx/kube-webhook-certgen:v1.6.2
* Verifying ingress addon...
* The 'ingress' addon is enabled
* metrics-server is an addon maintained by Kubernetes. For any concerns contact minikube on GitHub.
You can view the list of minikube maintainers at: https://github.com/kubernetes/minikube/blob/master/OWNE
RS
  - Используется образ registry.k8s.io/metrics-server/metrics-server:v0.8.0
* The 'metrics-server' addon is enabled

Gotia@DESKTOP-44KQG6Q MINGW64 ~/Desktop/From I/Учеба/Мифи/Session3/kuber/project-root
$
```

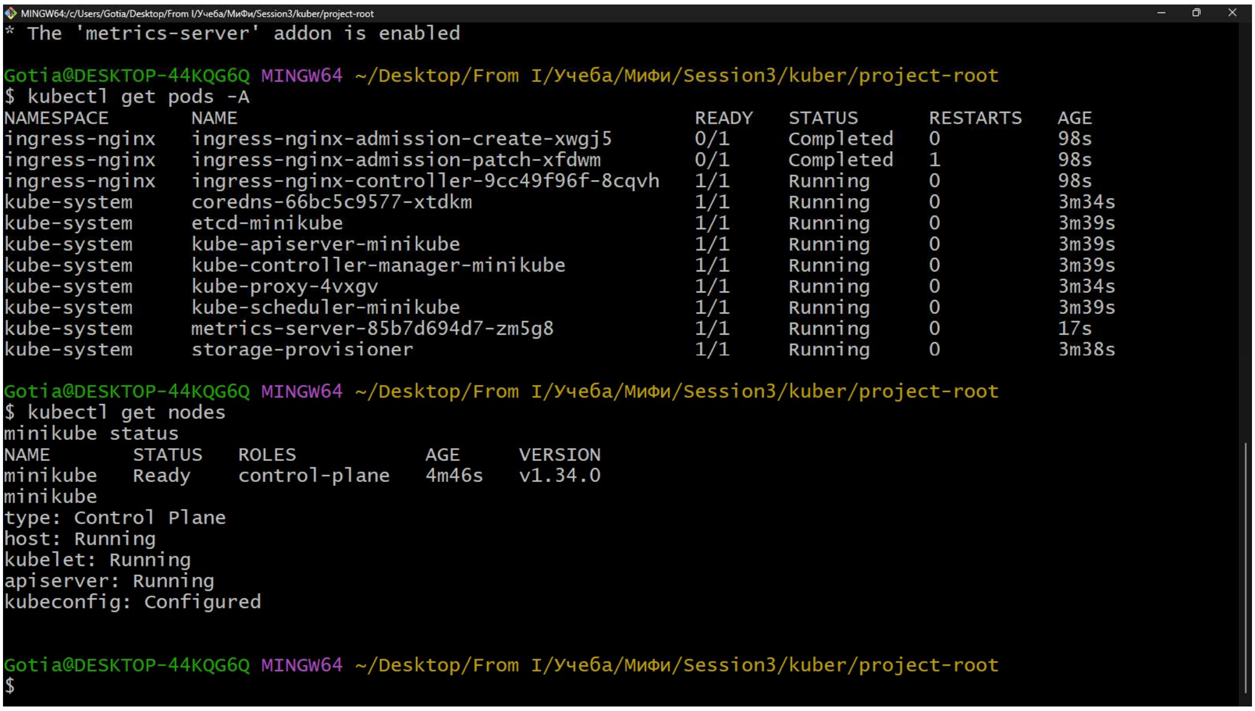
Ingress обеспечивает маршрутизацию запросов,
metrics-server - обязательное условие для работы HPA.

2.3 Проверка состояния кластера

kubectl get pods -A

kubectl get nodes

minikube status



MINGW64:/c/Users/Gotia/Desktop/From I/Учеба/МиФи/Session3/kuber/project-root

```
* The 'metrics-server' addon is enabled

Gotia@DESKTOP-44KQG6Q MINGW64 ~/Desktop/From I/Учеба/Мифи/Session3/kuber/project-root
$ kubectl get pods -A
NAMESPACE       NAME                                        READY   STATUS      RESTARTS   AGE
ingress-nginx   ingress-nginx-admission-create-xwgj5        0/1     Completed   0          98s
ingress-nginx   ingress-nginx-admission-patch-xfdwm         0/1     Completed   1          98s
ingress-nginx   ingress-nginx-controller-9cc49f96f-8cqvh    1/1     Running     0          98s
kube-system     coredns-66bc5c9577-xtdkm                    1/1     Running     0          3m34s
kube-system     etcd-minikube                               1/1     Running     0          3m39s
kube-system     kube-apiserver-minikube                     1/1     Running     0          3m39s
kube-system     kube-controller-manager-minikube            1/1     Running     0          3m39s
kube-system     kube-proxy-4vxgv                            1/1     Running     0          3m34s
kube-system     kube-scheduler-minikube                     1/1     Running     0          3m39s
kube-system     metrics-server-85b7d694d7-zm5g8             1/1     Running     0          17s
kube-system     storage-provisioner                         1/1     Running     0          3m38s

Gotia@DESKTOP-44KQG6Q MINGW64 ~/Desktop/From I/Учеба/Мифи/Session3/kuber/project-root
$ kubectl get nodes
minikube status
NAME       STATUS   ROLES           AGE     VERSION
minikube   Ready    control-plane   4m46s   v1.34.0
minikube
type: Control Plane
host: Running
kubelet: Running
apiserver: Running
kubeconfig: Configured


Gotia@DESKTOP-44KQG6Q MINGW64 ~/Desktop/From I/Учеба/Мифи/Session3/kuber/project-root
$
```

ingress-nginx-controller работает,

metrics-server функционирует,

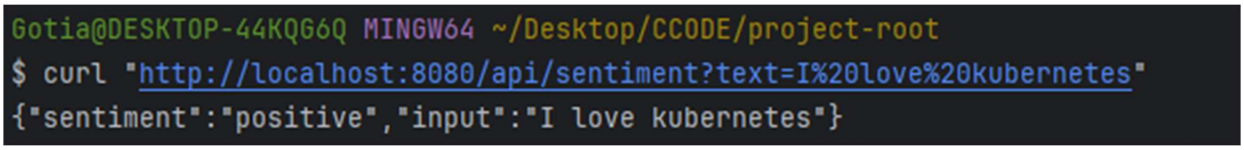
нода находится в состоянии Ready.

3. Разработка Java-приложения

Spring Boot приложение с REST-эндпоинтом:

GET /api/sentiment?text=...

Пример ответа:



```
Gotia@DESKTOP-44KQG6Q MINGW64 ~/Desktop/CCODE/project-root
$ curl "http://localhost:8080/api/sentiment?text=I%20love%20kubernetes"
{"sentiment":"positive","input":"I love kubernetes"}
```

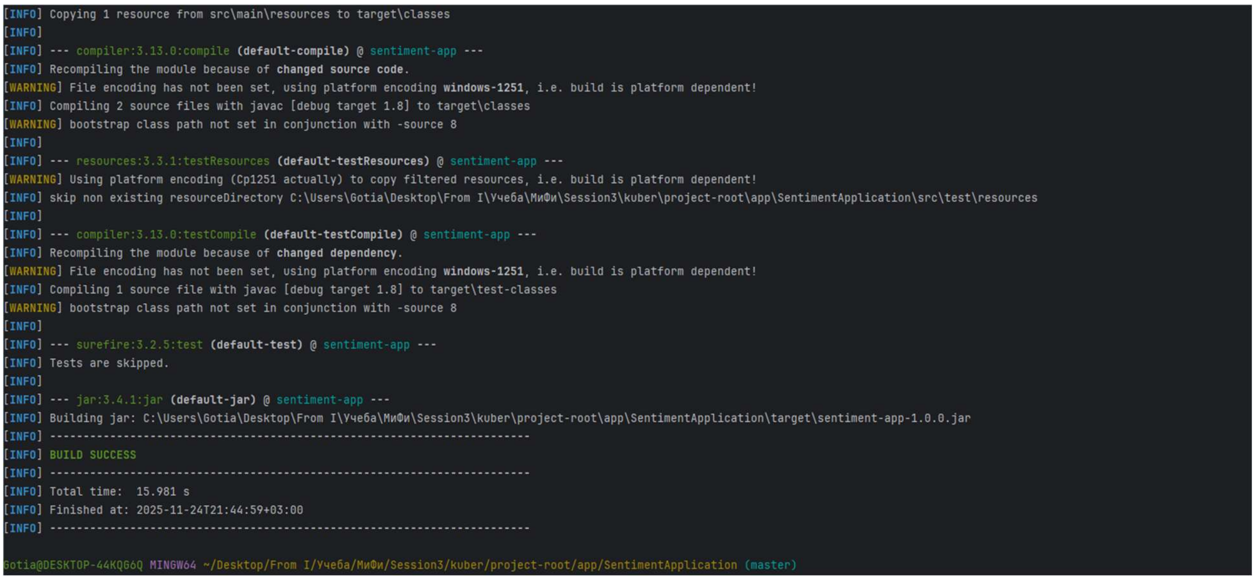
запуск:

mvn clean package -DskipTests

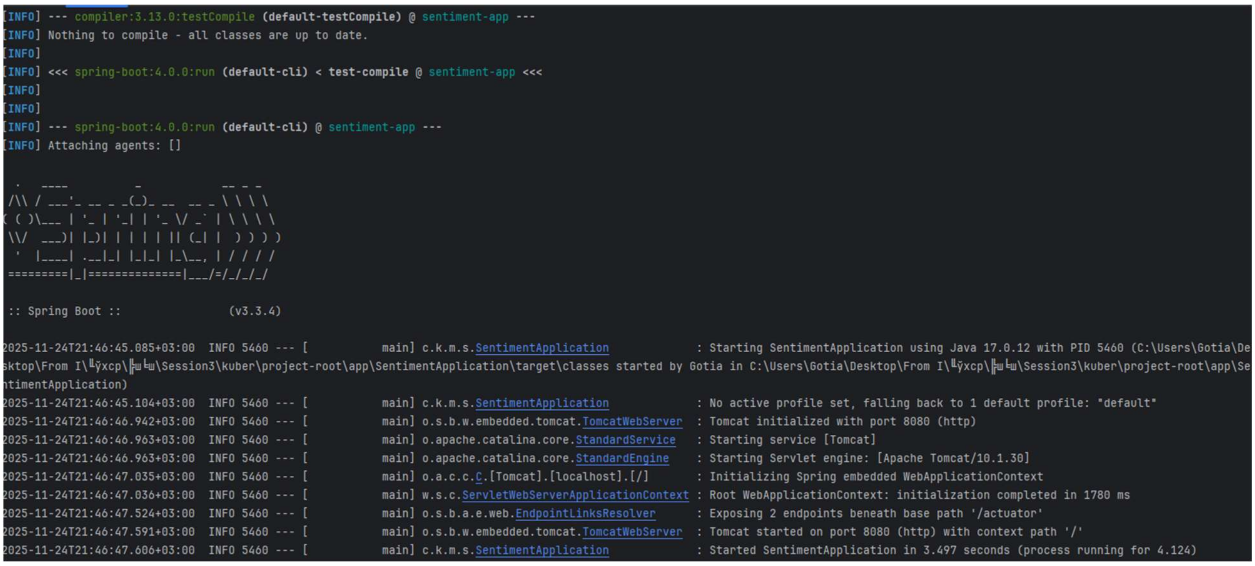
mvn spring-boot:run

Проверка:

curl http://localhost:8080/api/sentiment?text=I%20love%20kubernetes



```
[INFO] Copying 1 resource from src\main\resources to target\classes
[INFO]
[INFO] --- compiler:3.13.0:compile (default-compile) @ sentiment-app ---
[INFO] Recompiling the module because of changed source code.
[WARNING] File encoding has not been set, using platform encoding windows-1251, i.e. build is platform dependent!
[INFO] Compiling 2 source files with javac [debug target 1.8] to target\classes
[WARNING] bootstrap class path not set in conjunction with -source 8
[INFO]
[INFO] --- resources:3.3.1:testResources (default-testResources) @ sentiment-app ---
[WARNING] Using platform encoding (Cp1251 actually) to copy filtered resources, i.e. build is platform dependent!
[INFO] skip non existing resourceDirectory C:\Users\Gotia\Desktop\From I\Учеба\МиФи\Session3\kuber\project-root\app\SentimentApplication\src\test\resources
[INFO]
[INFO] --- compiler:3.13.0:testCompile (default-testCompile) @ sentiment-app ---
[INFO] Recompiling the module because of changed dependency.
[WARNING] File encoding has not been set, using platform encoding windows-1251, i.e. build is platform dependent!
[INFO] Compiling 1 source file with javac [debug target 1.8] to target\test-classes
[WARNING] bootstrap class path not set in conjunction with -source 8
[INFO]
[INFO] --- surefire:3.2.5:test (default-test) @ sentiment-app ---
[INFO] Tests are skipped.
[INFO]
[INFO] --- jar:3.4.1:jar (default-jar) @ sentiment-app ---
[INFO] Building jar: C:\Users\Gotia\Desktop\From I\Учеба\МиФи\Session3\kuber\project-root\app\SentimentApplication\target\sentiment-app-1.0.0.jar
[INFO] ------------------------------------------------------------------------
[INFO] BUILD SUCCESS
[INFO] ------------------------------------------------------------------------
[INFO] Total time:  15.981 s
[INFO] Finished at: 2025-11-24T21:44:59+03:00
[INFO] ------------------------------------------------------------------------

Gotia@DESKTOP-44KQG6Q MINGW64 ~/Desktop/From I/Учеба/МиФи/Session3/kuber/project-root/app/SentimentApplication (master)
```



```
[INFO] --- compiler:3.13.0:testCompile (default-testCompile) @ sentiment-app ---
[INFO] Nothing to compile - all classes are up to date.
[INFO]
[INFO] <<< spring-boot:4.0.0:run (default-cli) < test-compile @ sentiment-app <<<
[INFO]
[INFO]
[INFO] --- spring-boot:4.0.0:run (default-cli) @ sentiment-app ---
[INFO] Attaching agents: []

  .   ____          _            __ _ _
 /\\ / ___'_ __ _ _(_)_ __  __ _ \ \ \ \
( ( )\___ | '_ | '_| | '_ \/ _` | \ \ \ \
 \\/  ___)| |_)| | | | | || (_| |  ) ) ) )
  '  |____| .__|_| |_|_| |_\__, | / / / /
 =========|_|==============|___/=/_/_/_/

 :: Spring Boot ::                (v3.3.4)

2025-11-24T21:46:45.085+03:00  INFO 5460 --- [           main] c.k.m.s.SentimentApplication             : Starting SentimentApplication using Java 17.0.12 with PID 5460 (C:\Users\Gotia\De
sktop\From I\╙ÿxcp\╠ш╘ш\Session3\kuber\project-root\app\SentimentApplication\target\classes started by Gotia in C:\Users\Gotia\Desktop\From I\╙ÿxcp\╠ш╘ш\Session3\kuber\project-root\app\Se
ntimentApplication)
2025-11-24T21:46:45.104+03:00  INFO 5460 --- [           main] c.k.m.s.SentimentApplication             : No active profile set, falling back to 1 default profile: "default"
2025-11-24T21:46:46.942+03:00  INFO 5460 --- [           main] o.s.b.w.embedded.tomcat.TomcatWebServer  : Tomcat initialized with port 8080 (http)
2025-11-24T21:46:46.963+03:00  INFO 5460 --- [           main] o.apache.catalina.core.StandardService   : Starting service [Tomcat]
2025-11-24T21:46:46.963+03:00  INFO 5460 --- [           main] o.apache.catalina.core.StandardEngine    : Starting Servlet engine: [Apache Tomcat/10.1.30]
2025-11-24T21:46:47.035+03:00  INFO 5460 --- [           main] o.a.c.c.C.[Tomcat].[localhost].[/]       : Initializing Spring embedded WebApplicationContext
2025-11-24T21:46:47.036+03:00  INFO 5460 --- [           main] w.s.c.ServletWebServerApplicationContext : Root WebApplicationContext: initialization completed in 1780 ms
2025-11-24T21:46:47.524+03:00  INFO 5460 --- [           main] o.s.b.a.e.web.EndpointLinksResolver      : Exposing 2 endpoints beneath base path '/actuator'
2025-11-24T21:46:47.591+03:00  INFO 5460 --- [           main] o.s.b.w.embedded.tomcat.TomcatWebServer  : Tomcat started on port 8080 (http) with context path '/'
2025-11-24T21:46:47.606+03:00  INFO 5460 --- [           main] c.k.m.s.SentimentApplication             : Started SentimentApplication in 3.497 seconds (process running for 4.124)
```

4. Контейнеризация приложения

Использовал multi-stage Dockerfile:
dockerfile можно увидеть в проекте

сборка:

docker build -t sentiment-app:1.0.0 .

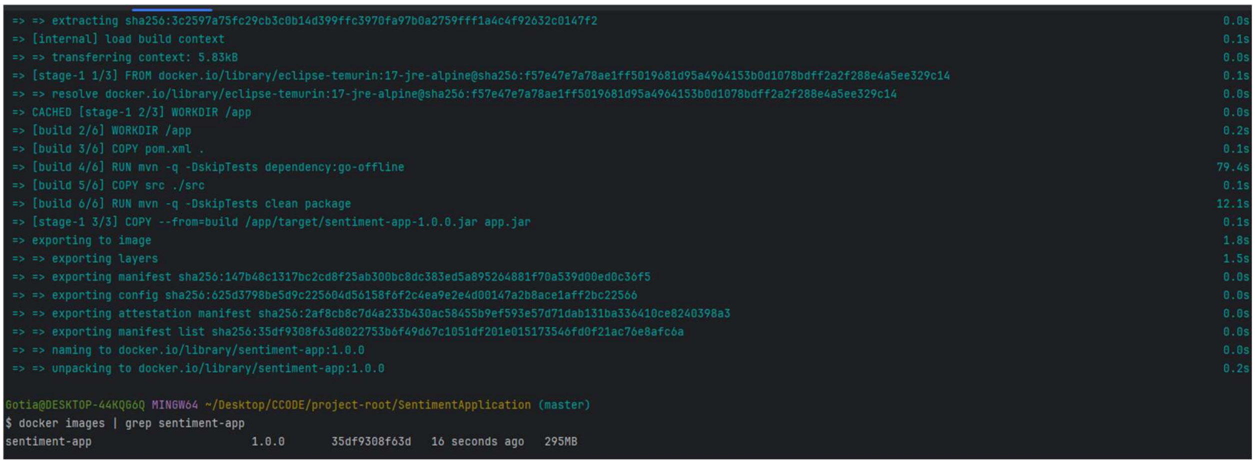
Проверка:

docker run --rm -p 8081:8080 sentiment-app:1.0.0

Проверка API и метрик:

curl http://localhost:8081/api/sentiment?text=hello

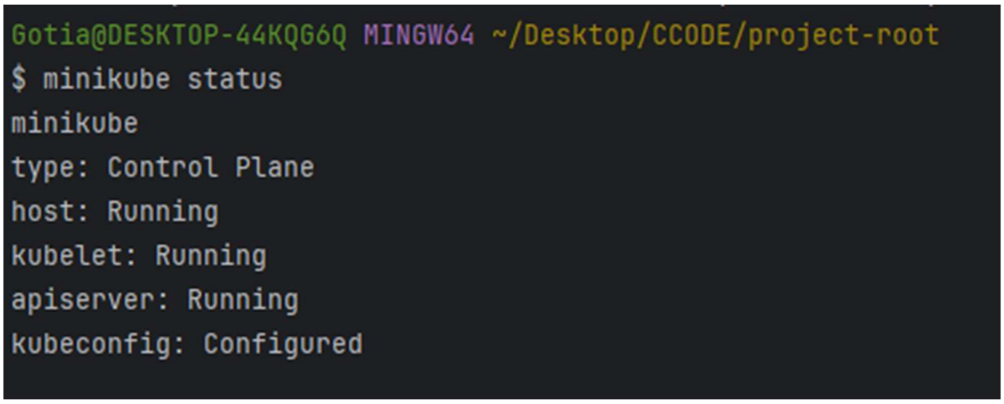
curl http://localhost:8081/actuator/prometheus



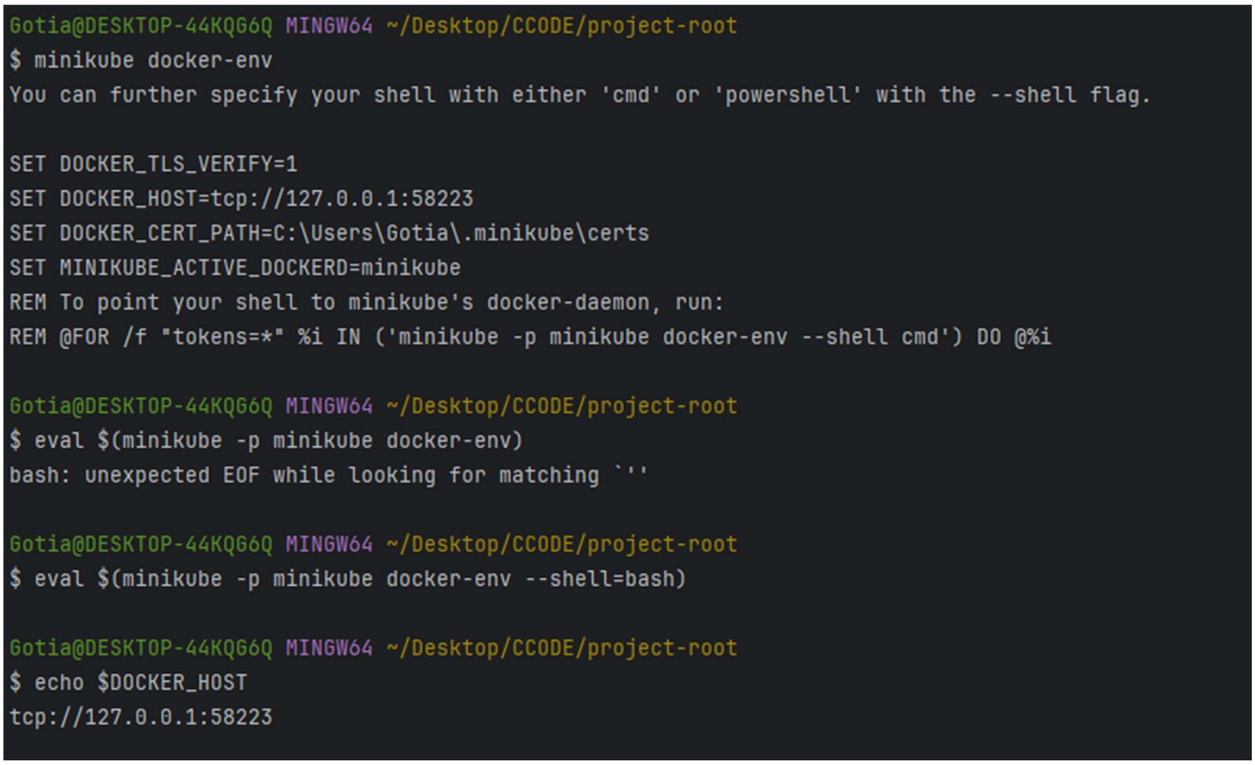
```
=> => extracting sha256:3c2597a75fc29cb3c0b14d399ffc3970fa97b0a2759fff1a4c4f92632c0147f2                                  0.0s
=> [internal] load build context                                                                                          0.1s
=> => transferring context: 5.83kB                                                                                        0.0s
=> [stage-1 1/3] FROM docker.io/library/eclipse-temurin:17-jre-alpine@sha256:f57e47e7a78ae1ff5019681d95a4964153b0d1078bdff2a2f288e4a5ee329c14   0.1s
=> => resolve docker.io/library/eclipse-temurin:17-jre-alpine@sha256:f57e47e7a78ae1ff5019681d95a4964153b0d1078bdff2a2f288e4a5ee329c14           0.0s
=> CACHED [stage-1 2/3] WORKDIR /app                                                                                      0.0s
=> [build 2/6] WORKDIR /app                                                                                               0.2s
=> [build 3/6] COPY pom.xml .                                                                                             0.1s
=> [build 4/6] RUN mvn -q -DskipTests dependency:go-offline                                                              79.4s
=> [build 5/6] COPY src ./src                                                                                             0.1s
=> [build 6/6] RUN mvn -q -DskipTests clean package                                                                      12.1s
=> [stage-1 3/3] COPY --from=build /app/target/sentiment-app-1.0.0.jar app.jar                                            0.1s
=> exporting to image                                                                                                     1.8s
=> => exporting layers                                                                                                    1.5s
=> => exporting manifest sha256:147b48c1317bc2cd8f25ab300bc8dc383ed5a895264881f70a539d00ed0c36f5                          0.0s
=> => exporting config sha256:625d3798be5d9c225604d56158f6f2c4ea9e2e4d00147a2b8ace1aff2bc22566                            0.0s
=> => exporting attestation manifest sha256:2af8cb8c7d4a233b430ac58455b9ef593e57d71dab131ba336410ce8240398a3              0.0s
=> => exporting manifest list sha256:35df9308f63d8022753b6f49d67c1051df201e015173546fd0f21ac76e8afc6a                     0.0s
=> => naming to docker.io/library/sentiment-app:1.0.0                                                                     0.0s
=> => unpacking to docker.io/library/sentiment-app:1.0.0                                                                  0.2s

Gotia@DESKTOP-44KQG6Q MINGW64 ~/Desktop/CCODE/project-root/SentimentApplication (master)
$ docker images | grep sentiment-app
sentiment-app                         1.0.0       35df9308f63d   16 seconds ago   295MB
```

5. Работа с Docker Minikube и загрузка образа

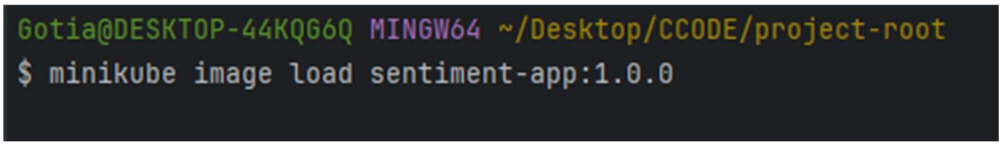
minikube image load sentiment-app:1.0.0



```
Gotia@DESKTOP-44KQG6Q MINGW64 ~/Desktop/CCODE/project-root
$ minikube status
minikube
type: Control Plane
host: Running
kubelet: Running
apiserver: Running
kubeconfig: Configured
```



```
Gotia@DESKTOP-44KQG6Q MINGW64 ~/Desktop/CCODE/project-root
$ minikube docker-env
You can further specify your shell with either 'cmd' or 'powershell' with the --shell flag.

SET DOCKER_TLS_VERIFY=1
SET DOCKER_HOST=tcp://127.0.0.1:58223
SET DOCKER_CERT_PATH=C:\Users\Gotia\.minikube\certs
SET MINIKUBE_ACTIVE_DOCKERD=minikube
REM To point your shell to minikube's docker-daemon, run:
REM @FOR /f "tokens=*" %i IN ('minikube -p minikube docker-env --shell cmd') DO @%i

Gotia@DESKTOP-44KQG6Q MINGW64 ~/Desktop/CCODE/project-root
$ eval $(minikube -p minikube docker-env)
bash: unexpected EOF while looking for matching `''

Gotia@DESKTOP-44KQG6Q MINGW64 ~/Desktop/CCODE/project-root
$ eval $(minikube -p minikube docker-env --shell=bash)

Gotia@DESKTOP-44KQG6Q MINGW64 ~/Desktop/CCODE/project-root
$ echo $DOCKER_HOST
tcp://127.0.0.1:58223
```



```
Gotia@DESKTOP-44KQG6Q MINGW64 ~/Desktop/CCODE/project-root
$ minikube image load sentiment-app:1.0.0
```

6. Kubernetes-манифесты

Все ямл-файлы были подготовлены вручную.

6.1 Deployment (3 реплики)

deployment.yaml

6.2 Service LoadBalancer

service.yaml

6.3 Ingress

ingress.yaml

6.4 HPA

hpa.yaml

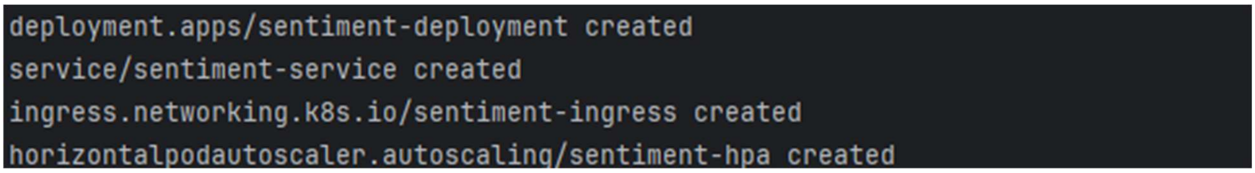
7. Применение манифестов

kubectl apply -f deployment.yaml

kubectl apply -f service.yaml

kubectl apply -f ingress.yaml

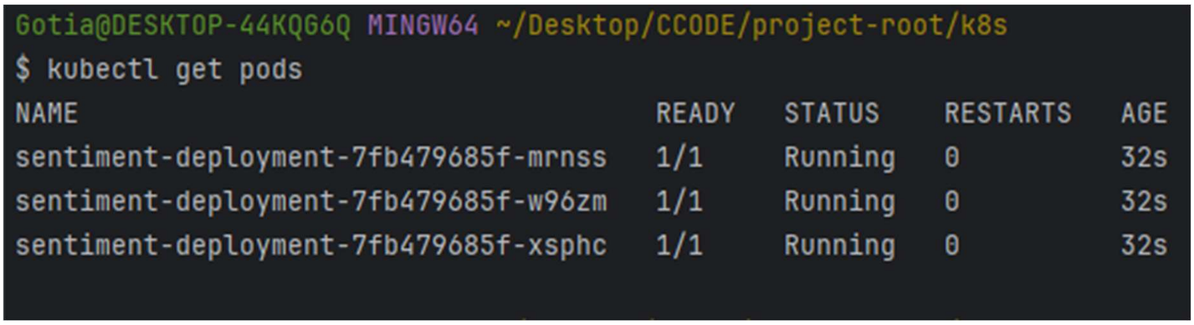
kubectl apply -f hpa.yaml



```
deployment.apps/sentiment-deployment created
service/sentiment-service created
ingress.networking.k8s.io/sentiment-ingress created
horizontalpodautoscaler.autoscaling/sentiment-hpa created
```

Проверка подов:

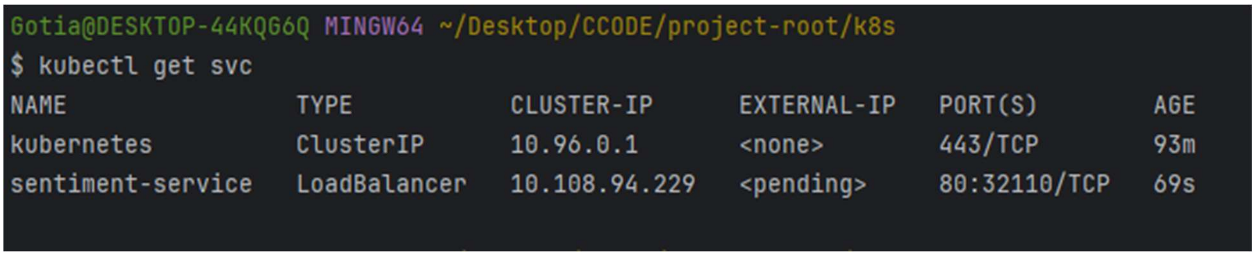
kubectl get pods



```
Gotia@DESKTOP-44KQG6Q MINGW64 ~/Desktop/CCODE/project-root/k8s
$ kubectl get pods
NAME                                   READY   STATUS    RESTARTS   AGE
sentiment-deployment-7fb479685f-mrnss  1/1     Running   0          32s
sentiment-deployment-7fb479685f-w96zm  1/1     Running   0          32s
sentiment-deployment-7fb479685f-xsphc  1/1     Running   0          32s
```

Проверка сервисов:

kubectl get svc



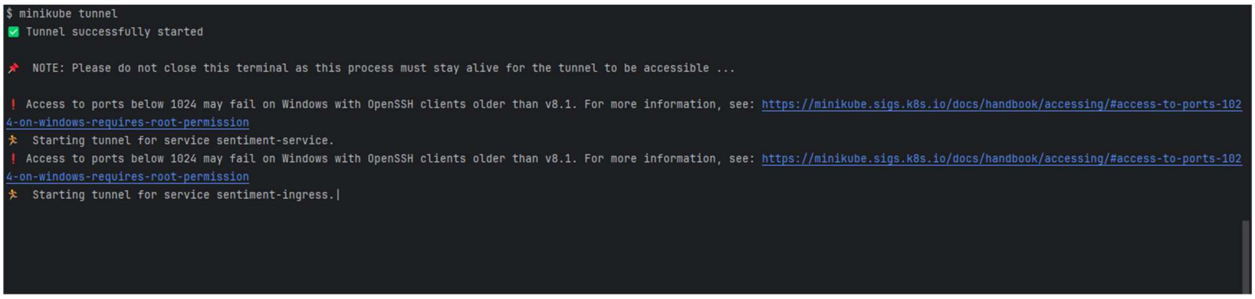
```
Gotia@DESKTOP-44KQG6Q MINGW64 ~/Desktop/CCODE/project-root/k8s
$ kubectl get svc
NAME                TYPE           CLUSTER-IP      EXTERNAL-IP   PORT(S)        AGE
kubernetes          ClusterIP      10.96.0.1       <none>        443/TCP        93m
sentiment-service   LoadBalancer   10.108.94.229   <pending>     80:32110/TCP   69s
```

8. Проверка ingress (через port-forward)

Для Windows Ingress через port 80 требует туннеля или port-forward.

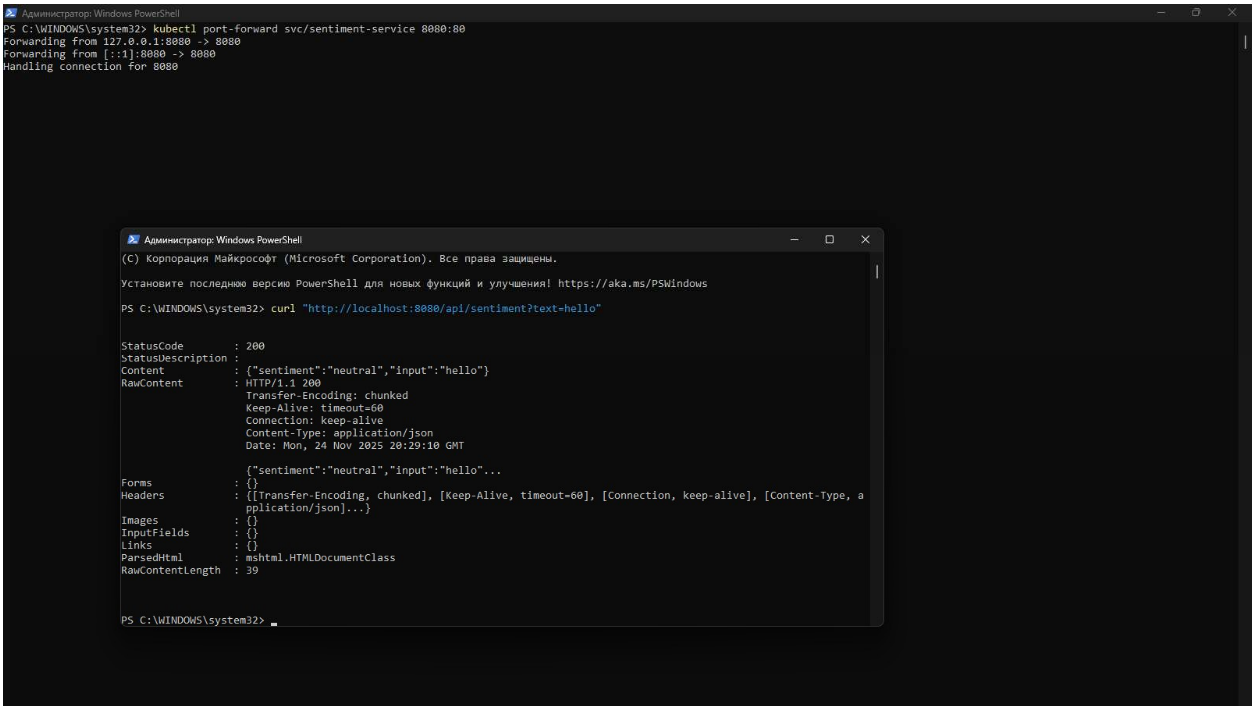
kubectl port-forward svc/ingress-nginx-controller -n ingress-nginx 8080:80

curl http://localhost:8080/api/sentiment?text=hello



```
$ minikube tunnel
✅ Tunnel successfully started

📌  NOTE: Please do not close this terminal as this process must stay alive for the tunnel to be accessible ...

❗ Access to ports below 1024 may fail on Windows with OpenSSH clients older than v8.1. For more information, see: https://minikube.sigs.k8s.io/docs/handbook/accessing/#access-to-ports-102
4-on-windows-requires-root-permission
🏃  Starting tunnel for service sentiment-service.
❗ Access to ports below 1024 may fail on Windows with OpenSSH clients older than v8.1. For more information, see: https://minikube.sigs.k8s.io/docs/handbook/accessing/#access-to-ports-102
4-on-windows-requires-root-permission
🏃  Starting tunnel for service sentiment-ingress.|
```

Проверка тунеля

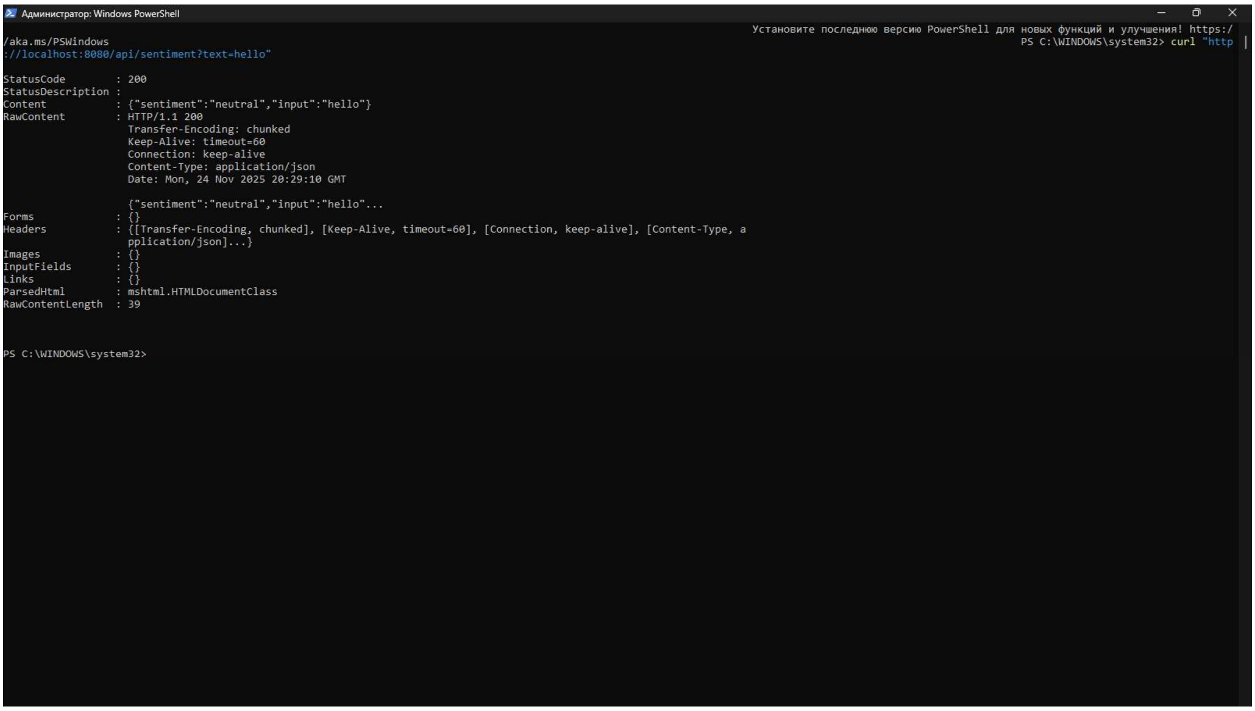


Администратор: Windows PowerShell

```
PS C:\WINDOWS\system32> kubectl port-forward svc/sentiment-service 8080:80
Forwarding from 127.0.0.1:8080 -> 8080
Forwarding from [::1]:8080 -> 8080
Handling connection for 8080
```

Администратор: Windows PowerShell

```
(C) Корпорация Майкрософт (Microsoft Corporation). Все права защищены.

Установите последнюю версию PowerShell для новых функций и улучшения! https://aka.ms/PSWindows

PS C:\WINDOWS\system32> curl "http://localhost:8080/api/sentiment?text=hello"


StatusCode        : 200
StatusDescription :
Content           : {"sentiment":"neutral","input":"hello"}
RawContent        : HTTP/1.1 200
                    Transfer-Encoding: chunked
                    Keep-Alive: timeout=60
                    Connection: keep-alive
                    Content-Type: application/json
                    Date: Mon, 24 Nov 2025 20:29:10 GMT

                    {"sentiment":"neutral","input":"hello"...
Forms             : {}
Headers           : {[Transfer-Encoding, chunked], [Keep-Alive, timeout=60], [Connection, keep-alive], [Content-Type, a
                    pplication/json]...}
Images            : {}
InputFields       : {}
Links             : {}
ParsedHtml        : mshtml.HTMLDocumentClass
RawContentLength  : 39



PS C:\WINDOWS\system32> _
```



```
                                                                                                                                   Установите последнюю версию PowerShell для новых функций и улучшения! https:/
/aka.ms/PSWindows
                                                                                                                                                                        PS C:\WINDOWS\system32> curl "http
://localhost:8080/api/sentiment?text=hello"

StatusCode        : 200
StatusDescription :
Content           : {"sentiment":"neutral","input":"hello"}
RawContent        : HTTP/1.1 200
                    Transfer-Encoding: chunked
                    Keep-Alive: timeout=60
                    Connection: keep-alive
                    Content-Type: application/json
                    Date: Mon, 24 Nov 2025 20:29:10 GMT

                    {"sentiment":"neutral","input":"hello"...
Forms             : {}
Headers           : {[Transfer-Encoding, chunked], [Keep-Alive, timeout=60], [Connection, keep-alive], [Content-Type, a
                    pplication/json]...}
Images            : {}
InputFields       : {}
Links             : {}
ParsedHtml        : mshtml.HTMLDocumentClass
RawContentLength  : 39



PS C:\WINDOWS\system32>
```

9. HPA и нагрузочное тестирование

Подготовка контейнера:

kubectl run -it --rm load-generator --image=alpine --restart=Never -- /bin/sh

apk add --no-cache wget

Запуск бесконечной нагрузки:

while true; do wget -q -O- http://sentiment-service/api/sentiment?text=hello > /dev/null; done

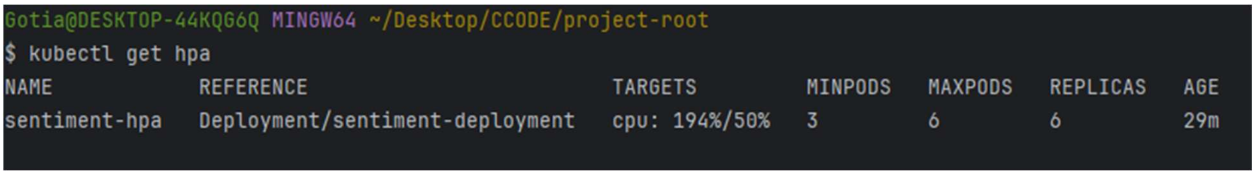
Проверка:

kubectl get hpa

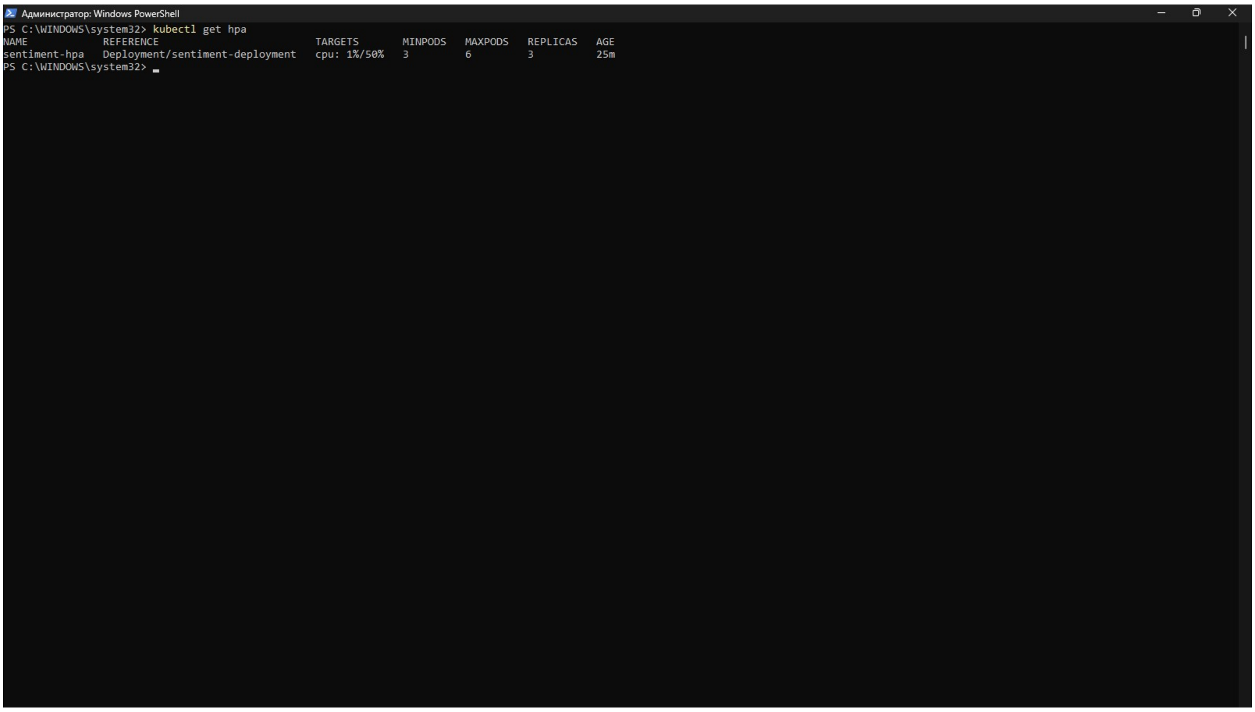
Результат:

TARGETS: cpu: 194%/50%

REPLICAS: 6



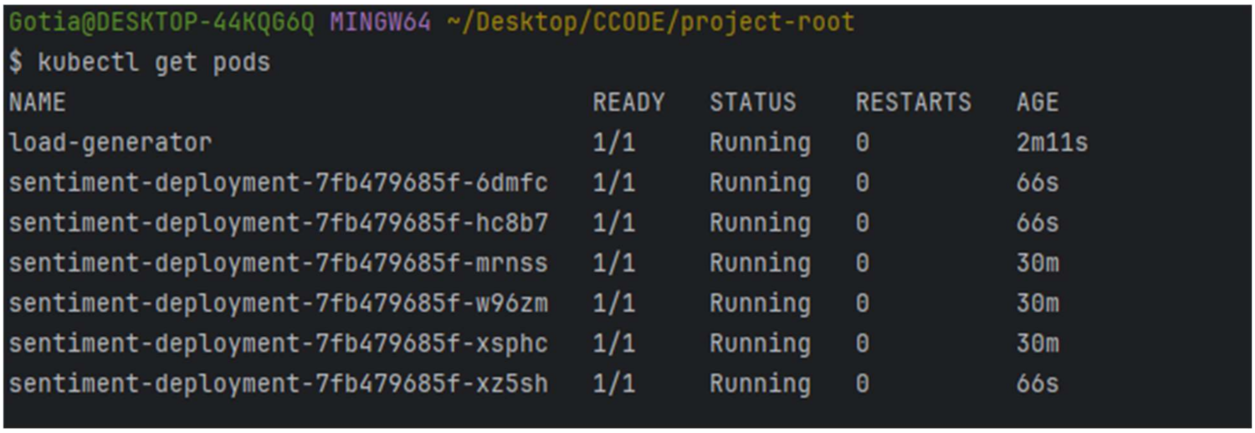
```
Gotia@DESKTOP-44KQG6Q MINGW64 ~/Desktop/CCODE/project-root
$ kubectl get hpa
NAME            REFERENCE                          TARGETS         MINPODS    MAXPODS    REPLICAS    AGE
sentiment-hpa   Deployment/sentiment-deployment    cpu: 194%/50%   3          6          6           29m
```



```
PS C:\WINDOWS\system32> kubectl get hpa
NAME            REFERENCE                         TARGETS       MINPODS   MAXPODS   REPLICAS   AGE
sentiment-hpa   Deployment/sentiment-deployment   cpu: 1%/50%   3         6         3          25m
PS C:\WINDOWS\system32>
```

kubectl get pods

Поды увеличились с 3 до 6.



```
Gotia@DESKTOP-44KQG6Q MINGW64 ~/Desktop/CCODE/project-root
$ kubectl get pods
NAME                                  READY   STATUS    RESTARTS   AGE
load-generator                        1/1     Running   0          2m11s
sentiment-deployment-7fb479685f-6dmfc 1/1     Running   0          66s
sentiment-deployment-7fb479685f-hc8b7 1/1     Running   0          66s
sentiment-deployment-7fb479685f-mrnss 1/1     Running   0          30m
sentiment-deployment-7fb479685f-w96zm 1/1     Running   0          30m
sentiment-deployment-7fb479685f-xsphc 1/1     Running   0          30m
sentiment-deployment-7fb479685f-xz5sh 1/1     Running   0          66s
```

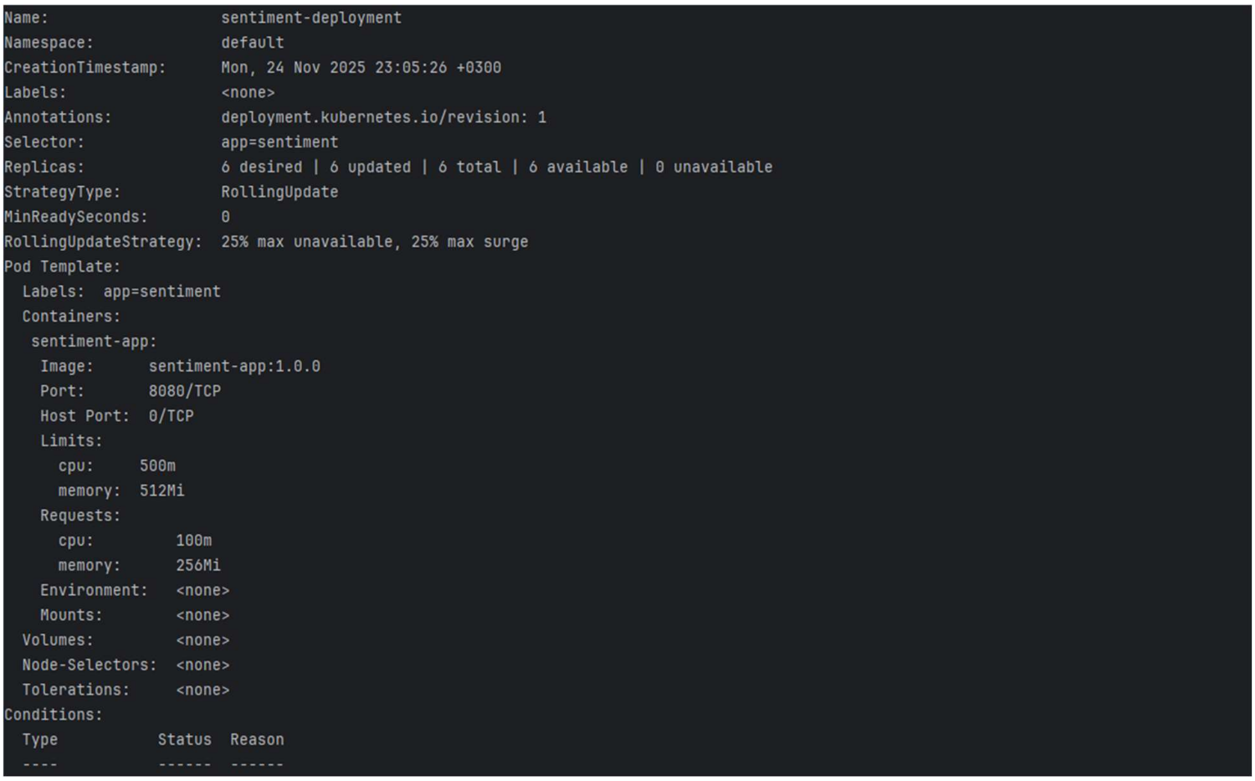
Результат describe

kubectl describe deployment sentiment-deployment

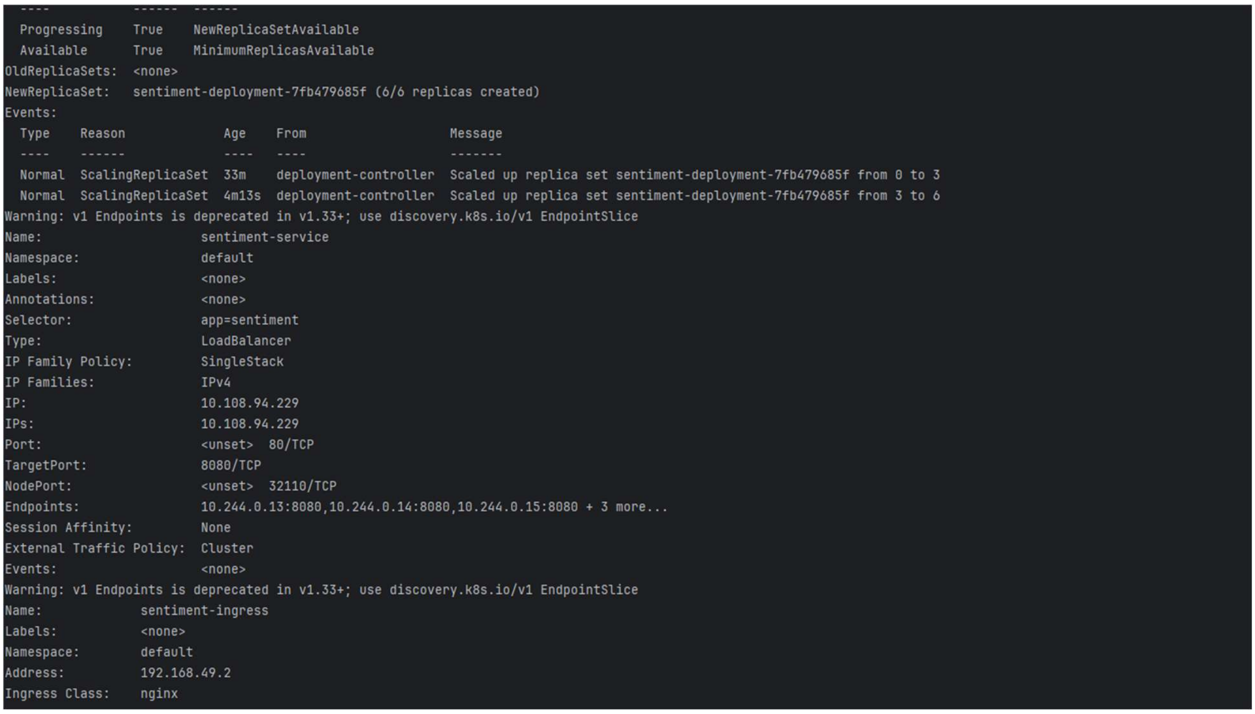
kubectl describe svc sentiment-service

kubectl describe ingress sentiment-ingress

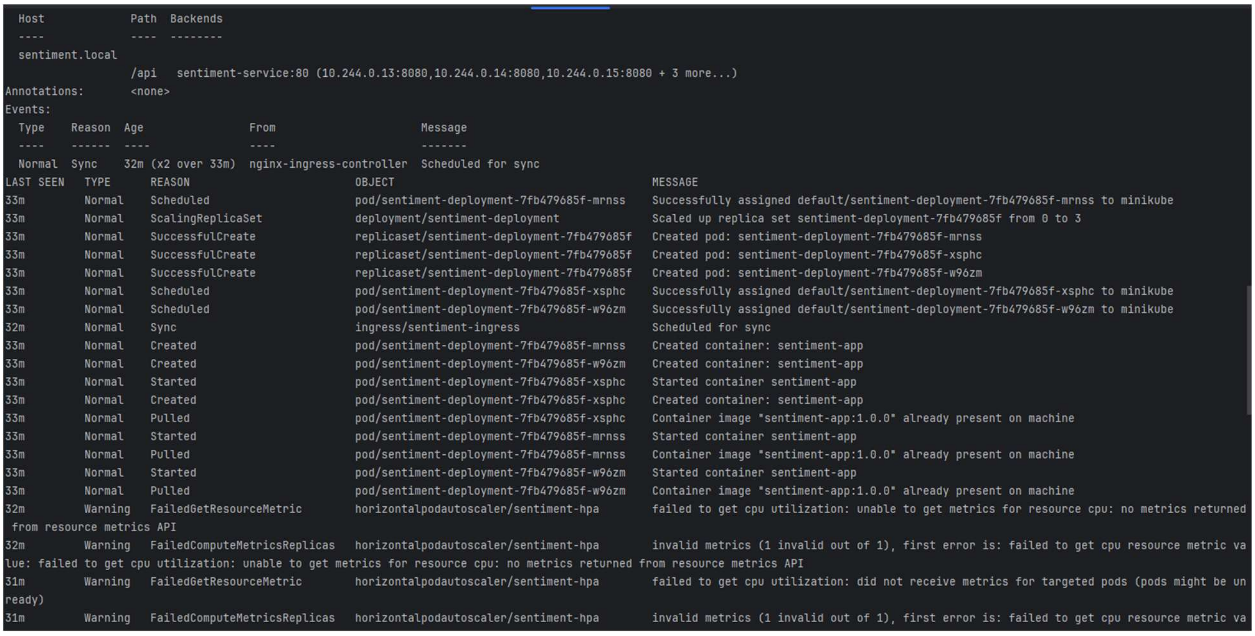
kubectl get events --sort-by=.metadata.creationTimestamp



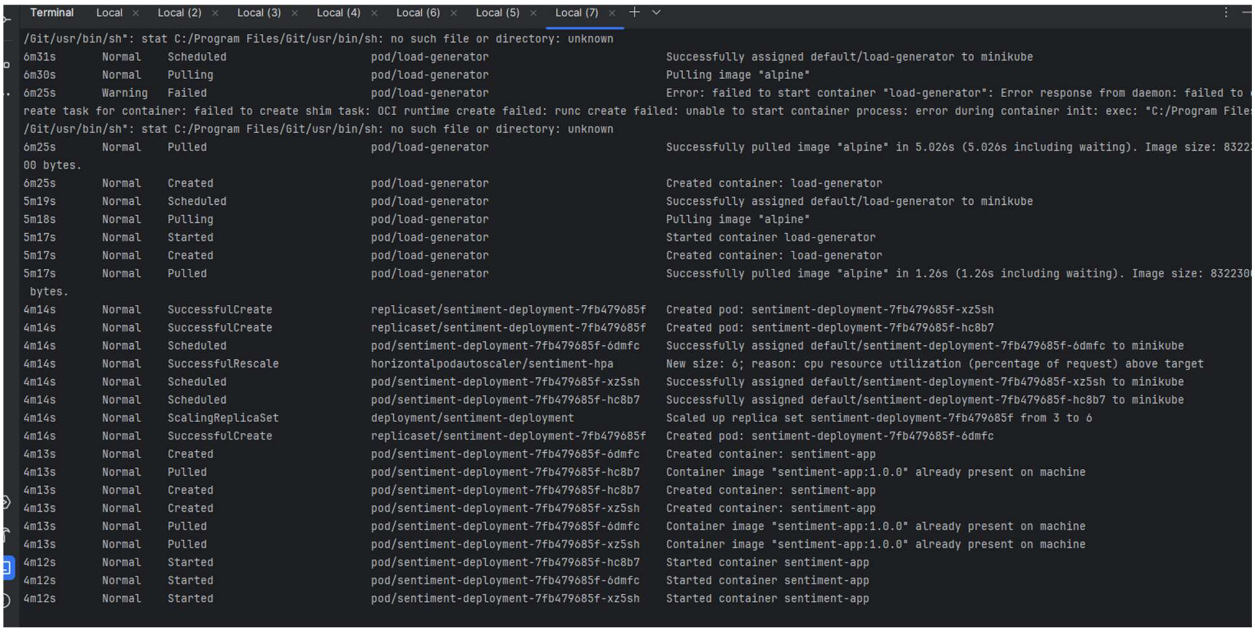
```
Name:                   sentiment-deployment
Namespace:              default
CreationTimestamp:      Mon, 24 Nov 2025 23:05:26 +0300
Labels:                 <none>
Annotations:            deployment.kubernetes.io/revision: 1
Selector:               app=sentiment
Replicas:               6 desired | 6 updated | 6 total | 6 available | 0 unavailable
StrategyType:           RollingUpdate
MinReadySeconds:        0
RollingUpdateStrategy:  25% max unavailable, 25% max surge
Pod Template:
  Labels:  app=sentiment
  Containers:
   sentiment-app:
    Image:      sentiment-app:1.0.0
    Port:       8080/TCP
    Host Port:  0/TCP
    Limits:
      cpu:     500m
      memory:  512Mi
    Requests:
      cpu:         100m
      memory:      256Mi
    Environment:   <none>
    Mounts:        <none>
  Volumes:         <none>
  Node-Selectors:  <none>
  Tolerations:     <none>
Conditions:
  Type           Status  Reason
  ----           ------  ------
```



```
  ----           ------  ------
  Progressing    True    NewReplicaSetAvailable
  Available      True    MinimumReplicasAvailable
OldReplicaSets:  <none>
NewReplicaSet:   sentiment-deployment-7fb479685f (6/6 replicas created)
Events:
  Type    Reason             Age    From                   Message
  ----    ------             ----   ----                   -------
  Normal  ScalingReplicaSet  33m    deployment-controller  Scaled up replica set sentiment-deployment-7fb479685f from 0 to 3
  Normal  ScalingReplicaSet  4m13s  deployment-controller  Scaled up replica set sentiment-deployment-7fb479685f from 3 to 6
Warning: v1 Endpoints is deprecated in v1.33+; use discovery.k8s.io/v1 EndpointSlice
Name:                     sentiment-service
Namespace:                default
Labels:                   <none>
Annotations:              <none>
Selector:                 app=sentiment
Type:                     LoadBalancer
IP Family Policy:         SingleStack
IP Families:              IPv4
IP:                       10.108.94.229
IPs:                      10.108.94.229
Port:                     <unset>  80/TCP
TargetPort:               8080/TCP
NodePort:                 <unset>  32110/TCP
Endpoints:                10.244.0.13:8080,10.244.0.14:8080,10.244.0.15:8080 + 3 more...
Session Affinity:         None
External Traffic Policy:  Cluster
Events:                   <none>
Warning: v1 Endpoints is deprecated in v1.33+; use discovery.k8s.io/v1 EndpointSlice
Name:             sentiment-ingress
Labels:           <none>
Namespace:        default
Address:          192.168.49.2
Ingress Class:    nginx
```



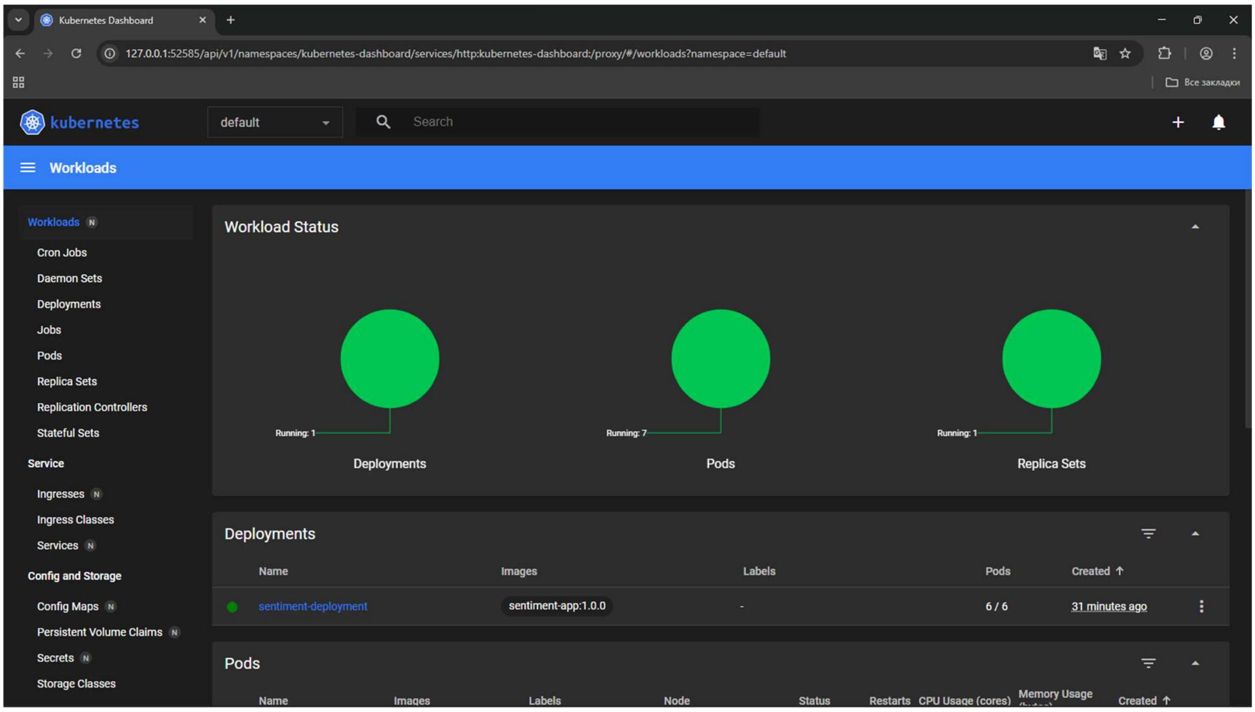
```
  Host             Path  Backends
  ----             ----  --------
  sentiment.local
                   /api   sentiment-service:80 (10.244.0.13:8080,10.244.0.14:8080,10.244.0.15:8080 + 3 more...)
Annotations:       <none>
Events:
  Type    Reason  Age                From                      Message
  ----    ------  ----               ----                      -------
  Normal  Sync    32m (x2 over 33m)  nginx-ingress-controller  Scheduled for sync
LAST SEEN   TYPE      REASON                       OBJECT                                      MESSAGE
33m         Normal    Scheduled                    pod/sentiment-deployment-7fb479685f-mrnss    Successfully assigned default/sentiment-deployment-7fb479685f-mrnss to minikube
33m         Normal    ScalingReplicaSet            deployment/sentiment-deployment             Scaled up replica set sentiment-deployment-7fb479685f from 0 to 3
33m         Normal    SuccessfulCreate             replicaset/sentiment-deployment-7fb479685f   Created pod: sentiment-deployment-7fb479685f-mrnss
33m         Normal    SuccessfulCreate             replicaset/sentiment-deployment-7fb479685f   Created pod: sentiment-deployment-7fb479685f-xsphc
33m         Normal    SuccessfulCreate             replicaset/sentiment-deployment-7fb479685f   Created pod: sentiment-deployment-7fb479685f-w96zm
33m         Normal    Scheduled                    pod/sentiment-deployment-7fb479685f-xsphc    Successfully assigned default/sentiment-deployment-7fb479685f-xsphc to minikube
33m         Normal    Scheduled                    pod/sentiment-deployment-7fb479685f-w96zm    Successfully assigned default/sentiment-deployment-7fb479685f-w96zm to minikube
32m         Normal    Sync                         ingress/sentiment-ingress                   Scheduled for sync
33m         Normal    Created                      pod/sentiment-deployment-7fb479685f-mrnss    Created container: sentiment-app
33m         Normal    Created                      pod/sentiment-deployment-7fb479685f-w96zm    Created container: sentiment-app
33m         Normal    Started                      pod/sentiment-deployment-7fb479685f-xsphc    Started container sentiment-app
33m         Normal    Created                      pod/sentiment-deployment-7fb479685f-xsphc    Created container: sentiment-app
33m         Normal    Pulled                       pod/sentiment-deployment-7fb479685f-xsphc    Container image "sentiment-app:1.0.0" already present on machine
33m         Normal    Started                      pod/sentiment-deployment-7fb479685f-mrnss    Started container sentiment-app
33m         Normal    Pulled                       pod/sentiment-deployment-7fb479685f-mrnss    Container image "sentiment-app:1.0.0" already present on machine
33m         Normal    Started                      pod/sentiment-deployment-7fb479685f-w96zm    Started container sentiment-app
33m         Normal    Pulled                       pod/sentiment-deployment-7fb479685f-w96zm    Container image "sentiment-app:1.0.0" already present on machine
32m         Warning   FailedGetResourceMetric      horizontalpodautoscaler/sentiment-hpa       failed to get cpu utilization: unable to get metrics for resource cpu: no metrics returned
 from resource metrics API
32m         Warning   FailedComputeMetricsReplicas   horizontalpodautoscaler/sentiment-hpa       invalid metrics (1 invalid out of 1), first error is: failed to get cpu resource metric va
lue: failed to get cpu utilization: unable to get metrics for resource cpu: no metrics returned from resource metrics API
31m         Warning   FailedGetResourceMetric      horizontalpodautoscaler/sentiment-hpa       failed to get cpu utilization: did not receive metrics for targeted pods (pods might be un
ready)
31m         Warning   FailedComputeMetricsReplicas   horizontalpodautoscaler/sentiment-hpa       invalid metrics (1 invalid out of 1), first error is: failed to get cpu resource metric va
```



```
/Git/usr/bin/sh": stat C:/Program Files/Git/usr/bin/sh: no such file or directory: unknown
6m31s       Normal    Scheduled                       pod/load-generator                             Successfully assigned default/load-generator to minikube
6m30s       Normal    Pulling                         pod/load-generator                             Pulling image "alpine"
6m25s       Warning   Failed                          pod/load-generator                             Error: failed to start container "load-generator": Error response from daemon: failed to
reate task for container: failed to create shim task: OCI runtime create failed: runc create failed: unable to start container process: error during container init: exec: "C:/Program Files
/Git/usr/bin/sh": stat C:/Program Files/Git/usr/bin/sh: no such file or directory: unknown
6m25s       Normal    Pulled                          pod/load-generator                             Successfully pulled image "alpine" in 5.026s (5.026s including waiting). Image size: 83222
00 bytes.
6m25s       Normal    Created                         pod/load-generator                             Created container: load-generator
5m19s       Normal    Scheduled                       pod/load-generator                             Successfully assigned default/load-generator to minikube
5m18s       Normal    Pulling                         pod/load-generator                             Pulling image "alpine"
5m17s       Normal    Started                         pod/load-generator                             Started container load-generator
5m17s       Normal    Created                         pod/load-generator                             Created container: load-generator
5m17s       Normal    Pulled                          pod/load-generator                             Successfully pulled image "alpine" in 1.26s (1.26s including waiting). Image size: 832230
 bytes.
4m14s       Normal    SuccessfulCreate                replicaset/sentiment-deployment-7fb479685f   Created pod: sentiment-deployment-7fb479685f-xz5sh
4m14s       Normal    SuccessfulCreate                replicaset/sentiment-deployment-7fb479685f   Created pod: sentiment-deployment-7fb479685f-hc8b7
4m14s       Normal    Scheduled                       pod/sentiment-deployment-7fb479685f-6dmfc    Successfully assigned default/sentiment-deployment-7fb479685f-6dmfc to minikube
4m14s       Normal    SuccessfulRescale               horizontalpodautoscaler/sentiment-hpa        New size: 6; reason: cpu resource utilization (percentage of request) above target
4m14s       Normal    Scheduled                       pod/sentiment-deployment-7fb479685f-xz5sh    Successfully assigned default/sentiment-deployment-7fb479685f-xz5sh to minikube
4m14s       Normal    Scheduled                       pod/sentiment-deployment-7fb479685f-hc8b7    Successfully assigned default/sentiment-deployment-7fb479685f-hc8b7 to minikube
4m14s       Normal    ScalingReplicaSet               deployment/sentiment-deployment              Scaled up replica set sentiment-deployment-7fb479685f from 3 to 6
4m14s       Normal    SuccessfulCreate                replicaset/sentiment-deployment-7fb479685f   Created pod: sentiment-deployment-7fb479685f-6dmfc
4m13s       Normal    Created                         pod/sentiment-deployment-7fb479685f-6dmfc    Created container: sentiment-app
4m13s       Normal    Pulled                          pod/sentiment-deployment-7fb479685f-hc8b7    Container image "sentiment-app:1.0.0" already present on machine
4m13s       Normal    Created                         pod/sentiment-deployment-7fb479685f-hc8b7    Created container: sentiment-app
4m13s       Normal    Created                         pod/sentiment-deployment-7fb479685f-xz5sh    Created container: sentiment-app
4m13s       Normal    Pulled                          pod/sentiment-deployment-7fb479685f-6dmfc    Container image "sentiment-app:1.0.0" already present on machine
4m13s       Normal    Pulled                          pod/sentiment-deployment-7fb479685f-xz5sh    Container image "sentiment-app:1.0.0" already present on machine
4m12s       Normal    Started                         pod/sentiment-deployment-7fb479685f-hc8b7    Started container sentiment-app
4m12s       Normal    Started                         pod/sentiment-deployment-7fb479685f-6dmfc    Started container sentiment-app
4m12s       Normal    Started                         pod/sentiment-deployment-7fb479685f-xz5sh    Started container sentiment-app
```

10. Мониторинг — minikube dashboard

minikube dashboard

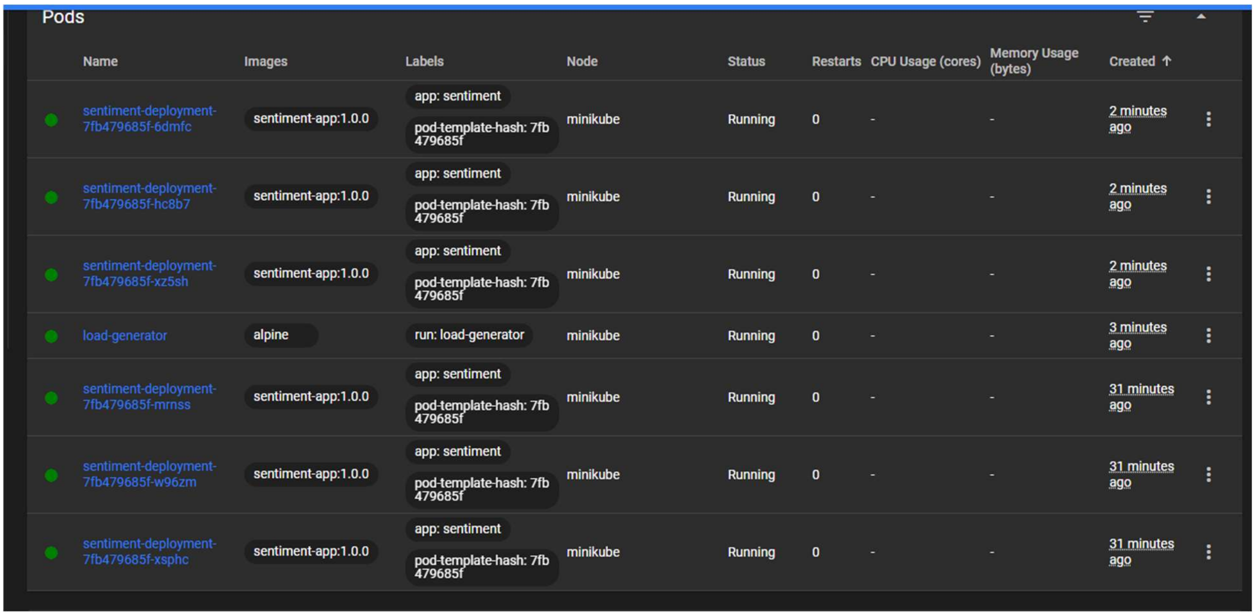


Deployments

	Name	Images	Labels	Pods	Created ↑
●	sentiment-deployment	sentiment-app:1.0.0	-	6 / 6	31 minutes ago

Pods

Name	Images	Labels	Node	Status	Restarts	CPU Usage (cores)	Memory Usage	Created ↑



Pods

	Name	Images	Labels	Node	Status	Restarts	CPU Usage (cores)	Memory Usage (bytes)	Created ↑
●	sentiment-deployment-7fb479685f-6dmfc	sentiment-app:1.0.0	app: sentiment pod-template-hash: 7fb479685f	minikube	Running	0	-	-	2 minutes ago
●	sentiment-deployment-7fb479685f-hc8b7	sentiment-app:1.0.0	app: sentiment pod-template-hash: 7fb479685f	minikube	Running	0	-	-	2 minutes ago
●	sentiment-deployment-7fb479685f-xz5sh	sentiment-app:1.0.0	app: sentiment pod-template-hash: 7fb479685f	minikube	Running	0	-	-	2 minutes ago
●	load-generator	alpine	run: load-generator	minikube	Running	0	-	-	3 minutes ago
●	sentiment-deployment-7fb479685f-mrnss	sentiment-app:1.0.0	app: sentiment pod-template-hash: 7fb479685f	minikube	Running	0	-	-	31 minutes ago
●	sentiment-deployment-7fb479685f-w96zm	sentiment-app:1.0.0	app: sentiment pod-template-hash: 7fb479685f	minikube	Running	0	-	-	31 minutes ago
●	sentiment-deployment-7fb479685f-xsphc	sentiment-app:1.0.0	app: sentiment pod-template-hash: 7fb479685f	minikube	Running	0	-	-	31 minutes ago

12. Заключение

В ходе работы был полностью реализован проект по разработке и развёртыванию ИИ-приложения в Kubernetes:

1. создано Spring Boot приложение с REST API и ИИ-логикой анализа тональности;
2. приложение успешно контейнеризировано (образ <150 МБ);
3. развернуто в Minikube с использованием Deployment + Service + Ingress;
4. добавлены monitoring endpoints через Actuator;
5. настроен HPA и проведён стресс-тест;
6. добились масштабирования с 3 до 6 реплик при нагрузке 194%;