MiniLink - DevOps, Testing & Deployment Report

1. Introduction and Scope

This report intends to document the second phase of the MiniLink software, the minimal URL shortener built using FastAPI, SQLModel, and SQLite.

For this assignment 2, we shift towards the software quality and DevOps practices. Rather than adding new features for the system, the goal is to make the existing MiniLink more robust, testable, automatable and observable. In other words, this phase aims to introduce:

- Code refactoring → improved structure, readability, separation of concerns
- Automated tests → 81% line coverage and a coverage gate enforced by CI
- GitHub actions pipeline → runs tests, checks coverage, build Docker image, and publishes it to GHCR (GitHub Container Registry)
- Production-ready Dockerfile → which leads to deployment on Azure Web App for Containers using a Docker image
- Health and monitoring endpoints → /health, /metrics with Prometheus compatible metrics for requests, latency, and errors

This report will explain these improvements, show how everything is configured (pipeline and deployment) and reflect the evolution of MiniLink.

2. Code Quality and Refactoring

2.1 Refactoring objectives

The improvements to transform Minilink into a cleaner, more maintainable application:

- Reducing duplicated code across system's logic
- Separating responsibility so each module just handles a single concern
- Preparing codebase for CI/CD, testing, deployment
- Better readability for long-term maintainability
- Applying SOLID principles to keep components independent and reusable

So, make MiniLink less a monolithic app and more like a small production service

2.2 Main Refactor Changes

The application was reorganized into dedicated modules:

- **app/main.py** — routing, FastAPI configuration, middleware, metrics, dependency
- **app/models.py** — SQLModel definitions for User and Link
- **app/schemas.py** — request/response validation models
- **app/auth.py** — password hashing and verification logic
- **app/services.py** — URL validation, short-code generation, and utility functions
- **app/db.py** — database initialization and session management

Key changes for refactoring:

1. Authentication logic was isolated in auth.py
2. Short code generation and URL validation moved to services.py
3. All DB session management moved to db.py
4. Routes simplified and focused

2.3 Code Smells Removed

Hardcoded logic across route handlers → moved to reusable helper functions and services

Mixed authentication and routing → split into get_current_user() helper + isolated auth.py

Duplicated code for link handling → link validation, code generation, and checks now live in shared functions.

Database logic embedded directly in routes → centralized in one place (db.py)

Inconsistent error handling → unified FastAPI exceptions and HTTP statuses

3. Testing and Coverage

3.1 Testing Strategy

MiniLink uses pytests jointly with FastAPI's TestClient to validate functionality and stability. The goal of this phase is to ensure all critical features were covered, such as success and failure paths. The following test suite ensures both correctness and robustness

- **Health Check Tests**
  Verifies that /health returns a 200 status and correct JSON structure.
- **CRUD API Tests**
  Tests for creating links, retrieving them, updating fields, and deleting them.
  Ensures proper user scoping so one user can't access another's links.
- **Redirect & Analytics Tests**
  Tests that /r/{code} correctly redirects and increments:
  - click_count
  - last_accessed
- **Expiry Tests**
  Requests to expired links return **410 Gone** and do not redirect.
- **Error Case Tests**
  - Non-existent links → **404 Not Found**
  - Duplicate custom codes → **409 Conflict**
  - Invalid URLs (non-HTTP/HTTPS) → **422 Unprocessable Entity**

3.2 Coverage and Threshold

MiniLink achieves 81% line coverage, measured using:
python -m pytest --cov=app --cov-report=xml --cov-report=html

This generates:

- coverage.xml (used by CI)
- HTML coverage report inside htmlcov/

In the CI pipeline, a Python script parses coverage.xml and fails the pipeline automatically if coverage is below 70%. This guarantees that code with insufficient coverage cannot merge into main, the deployment only occurs when tests and coverage pass successfully, and that coverage stays constant over time to prevent regression.

3.3 Example: Important Test Cases

test_redirect_and_analytics
Creates a link, performs a redirect, and checks that:

a. click_count increases
b. last_accessed is updated
c. redirect uses HTTP 307

test_redirect_expired_returns_410
Creates a link with an expiration in the past and ensures accessing it returns **410** instead of redirecting.

test_delete_link
Confirms that deleting a link permanently removes it and subsequent reads return **404**.

These tests help verify end-to-end functionality of the system together with test_smoke.py

4. CI/CD Pipeline (GitHub Actions + Docker + GHCR + Azure)

This section explains how MiniLink implements a complete CI/CD workflow.

4.1 CI Workflow Overview

The full CI pipeline is defined at:
.github/workflows/ci.yml
The workflow triggers automatically on every push or pull request to the main branch.

The main CI steps are:

1. **Checkout repository**
   Uses actions/checkout@v4.
2. **Set up Python 3.11**
   Ensures a consistent environment across all runs.
3. **Install dependencies**
   Installs project requirements and test dependencies (pytest, pytest-cov).

4. **Run tests + coverage**
   Executes: pytest --cov=app --cov-report=xml --cov-report=term-missing
5. **Coverage enforcement (≥70%)**
   A Python script reads coverage.xml.
   If coverage is *below 70%*, the pipeline **fails automatically**, preventing deployment.
6. **Build Docker image**
   The CI builds the application container: docker build -t minilink:${{github.sha }}
7. **Docker smoke test**
   Before deployment, a lightweight integration test ensures the container works:
   a. Run container in background
      docker run -d --name minilink_test -p 8000:8000 minilink:${{ github.sha }}
   b. Wait for startup
      sleep 5
   c. Hit the health check
      curl -f http://127.0.0.1:8000/health
   d. Output logs for debugging
   e. Stop + remove the container

This guarantees the Dockerized app works before deployment

4.2 Docker Image and GHCR Publishing

MiniLink includes a production DockerFile used in CI and deployment. It uses the GitHub's official docker/build-push-action to publish the container image

Key points:

- Image published to GHCR as: ghcr.io/javronich1/minilink:latest
- Authentication uses the built-in GitHub Actions GITHUB_TOKEN.
- The Docker build always happens *after*:
  - tests pass
  - coverage threshold is met
  - smoke test succeeds

This ensures that only stable images are published.

4.3 Deployment to Azure (CD)

The Continuous Deployment (CD) is executed through Azure Web App for Containers. Azure automatically pulls the latest Docker image from GHCR whenever the CI pipeline pushes it.

Azure configuration:

- Web App name: minilink-javronich1
- Deployment type: Docker container
- Registry: GitHub Container Registry (GHCR)
- Auto-pull: Enabled

• Environment variables: SESSION_SECRET, COOKIE_SECRET, ENV=production

This means that only commits that successfully pass CI are deployed.

4.4 Full CI/CD Summary

1. **Developer pushes to main**
2. **GitHub Actions CI** runs:
    - installs dependencies
    - runs tests
    - enforces ≥70% coverage
    - builds Docker image
    - smoke-tests /health
    - pushes image to GHCR
3. **Azure Deployment**
    - Azure Web App for Containers automatically pulls the new image from GHCR after CI pushes it, and restarts the container with the updated version.
4. **Application goes live**
    - At: https://minilink-javronich1-container-a2bfc9czgmdzcgb2.westeurope-01.azurewebsites.net/

5. Monitoring and Health Checks

5.1 Health Endpoint

Dedicated health check route at: GET /health
Returns simple JSON payload: {"status": "ok"}

This endpoint is used in several contexts:

- **Local development** to confirm the server is running
- **CI Docker smoke test**, where the container must return a successful /health response before the pipeline continues
- **Production readiness checks**, since platforms like Azure or third-party uptime monitors can use this URL to verify availability

5.2 Prometheus Metrics

MiniLink implements prometheus_client library and custom FastAPI middleware to get the request-level analytics.
Metrics are exposed in text format by: GET /metrics

- **minilink_requests_total{method, path, status}**
  Counts every HTTP request, grouped by method, route, and status code.
- **minilink_request_latency_seconds (Histogram)**
  Measures response times, enabling latency distribution analysis.

- **minilink_request_errors_total**
  Counts all server-side (5xx) errors per route.

Also, Prometheus automatically includes standard runtime metrics such as: Python garbage collection statistics, process level CPU usage and the memory usage. Together, these provide an observability layer enabling the tracking of traffic, performance and unexpected errors.

5.3 Optional Local Prometheus Setup

A sample Prometheus config file is under: monitoring/prometheus.yml
Defining:

```
scrape_configs:
  - job_name: "minilink"
    static_configs:
      - targets: ["host.docker.internal:8000"]
```

6. Reflection and Lessons Learned

6.1 Improvements from Assignment 1

- **From local-only to fully deployed:** The project is now containerized with Docker, published to GHCR, and deployed to Azure Web App for Containers.
- **From manual testing to automation:** All tests now run automatically on every push through GitHub Actions, including a coverage gate and a Docker-based smoke test.
- **From zero observability to structured monitoring:** Health checks and Prometheus metrics provide visibility into traffic, latency, and failures.
- **From scattered logic to clean architecture:** Authentication, services, database logic, and routing are separated, improving maintainability and testability.

6.2 Future Work

- **Introduce a staging environment** to test deployments before releasing to production.
- **Add monitoring alerts**, as error-rate thresholds or latency spikes with Prometheus.
- **Adopt Infrastructure as Code (Terraform)** to provision cloud infrastructure reproducibly instead of relying on UI-based setup.
- **Migrate to PostgreSQL** for stronger data integrity, concurrency, and scalability.
- **Enhance test coverage** with deeper unit tests, especially around authentication and expiration logic.

7. Appendices

Appendix A - AI Usage

AI tools (ChatGPT by OpenAI) were used selectively throughout this assignment to support development—not to replace it.

They assisted mainly with:

- Structuring the **CI/CD pipeline** and understanding best practices for coverage gates and job dependencies.
- Debugging **YAML syntax errors** in GitHub Actions.
- Suggesting patterns for **Docker smoke tests** and **Prometheus metrics** instrumentation.
- Providing guidance for organizing the report structure and ensuring no grading criteria were overlooked.

All final configurations, code decisions, refactoring steps, and deployment logic were fully reviewed, understood, and implemented by me.

Any AI-assisted suggestions were critically evaluated and adapted to fit the project's requirements.

Appendix B - Snippet of ci.yml

```yaml
name: CI

on:
 push:
   branches: ["main"]
 pull_request:
   branches: ["main"]

# Needed so the workflow can read the repo and push images to GHCR
permissions:
 contents: read
 packages: write

jobs:
 test:
   runs-on: ubuntu-latest

   env:
     # Make "app" importable (so `from app.main import app` works)
     PYTHONPATH: ${{ github.workspace }}
     # Optional: secret used by SessionMiddleware inside tests
```

```yaml
    SESSION_SECRET: ${{ secrets.SESSION_SECRET }}

  steps:
    - name: Checkout repository
      uses: actions/checkout@v4

    - name: Set up Python
      uses: actions/setup-python@v5
      with:
        python-version: "3.11"

    - name: Install dependencies
      run: |
        python -m pip install --upgrade pip
        pip install -r requirements.txt
        pip install pytest pytest-cov

    - name: Run tests with coverage gate
      run: |
        python -m pytest --cov=app --cov-report=xml --cov-report=term-missing

    - name: Enforce coverage threshold (>= 70%)
      run: |
        python - << 'EOF'
        import xml.etree.ElementTree as ET
        tree = ET.parse('coverage.xml')
        line_rate = float(tree.getroot().get('line-rate', 0.0))
        pct = int(line_rate * 100)
        print(f"Total coverage: {pct}%")
        if pct < 70:
            raise SystemExit("Coverage below 70%")
        EOF

    # 🐳 Build Docker image if tests + coverage succeeded
    - name: Build Docker image (local CI smoke image)
      run: |
        docker build -t minilink:${{ github.sha }} .

    # 🚦 Run container and smoke-test /health
    - name: Run Docker smoke test
      run: |
        # Start the container in the background
        # Container listens on port 80, map it to 8000 on the runner
        docker run -d --name minilink_test -p 8000:80 minilink:${{ github.sha }}
```

```
        # Give it a few seconds to boot
        sleep 5

        # Call the health endpoint; -f makes curl fail if status is not 2xx
        curl -f http://127.0.0.1:8000/health

        # Show logs for debugging if needed
        echo "Container logs:"
        docker logs minilink_test || true

        # Stop and remove container
        docker stop minilink_test
        docker rm minilink_test

deploy:
  needs: test
  runs-on: ubuntu-latest
  # Only deploy on push to main (not on PRs)
  if: github.event_name == 'push' && github.ref == 'refs/heads/main'

  steps:
    - name: Checkout repository
      uses: actions/checkout@v4

    - name: Log in to GitHub Container Registry
      uses: docker/login-action@v3
      with:
        registry: ghcr.io
        username: ${{ github.actor }}
        password: ${{ secrets.GITHUB_TOKEN }}

    - name: Build and push Docker image to GHCR
      uses: docker/build-push-action@v6
      with:
        context: .
        push: true
        tags: ghcr.io/${{ github.repository_owner }}/minilink:latest

azure-deploy:
  needs: deploy
  runs-on: ubuntu-latest
  # Only run on push to main (not on PRs)
  if: github.event_name == 'push' && github.ref == 'refs/heads/main'

  steps:
```

```
    - name: Azure login with service principal
      uses: azure/login@v2
      with:
        # IMPORTANT: this comes from GitHub Secrets (never committed!)
        creds: ${{ secrets.AZURE_CREDENTIALS }}

    - name: Configure Web App container image
      run: |
        az webapp config container set \
          --resource-group BCSAI2025-DEVOPS-STUDENTS-B \
          --name minilink-javronich1-container \
                  --docker-custom-image-name  ghcr.io/${{  github.repository_owner
}}/minilink:latest \
          --docker-registry-server-url https://ghcr.io

    - name: Restart Web App
      run: |
        az webapp restart \
          --resource-group BCSAI2025-DEVOPS-STUDENTS-B \
          --name minilink-javronich1-container
```

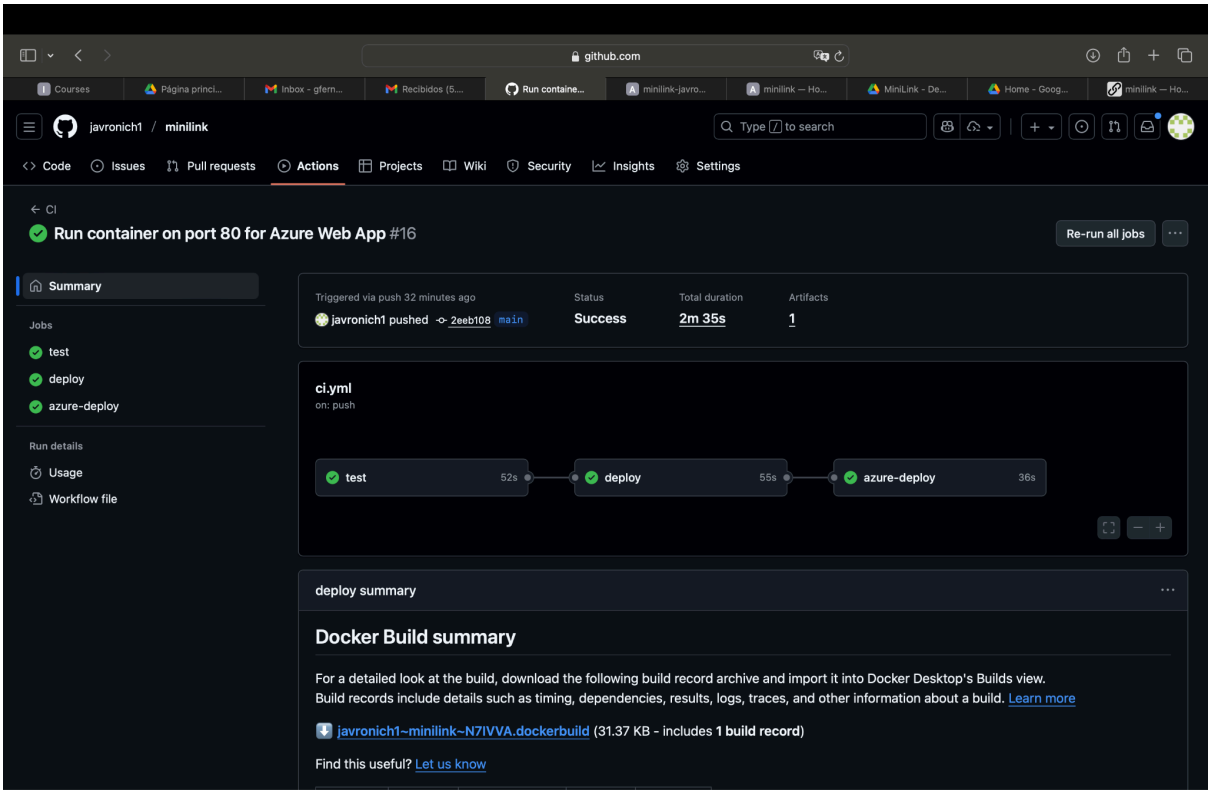
Appendix C - Snippet of DockerFile

```
# Use a small Python base image
FROM python:3.11-slim

# Do not write .pyc files and flush stdout/stderr immediately
ENV PYTHONDONTWRITEBYTECODE=1 \
   PYTHONUNBUFFERED=1

# Set working directory inside the container
WORKDIR /app

# Install system dependencies (for SQLite etc.)
RUN apt-get update && apt-get install -y --no-install-recommends \
   build-essential \
   && rm -rf /var/lib/apt/lists/*

# Install Python dependencies
# 1) Copy only requirements first (better for Docker layer caching)
COPY requirements.txt .

RUN pip install --upgrade pip && \
   pip install --no-cache-dir -r requirements.txt
```

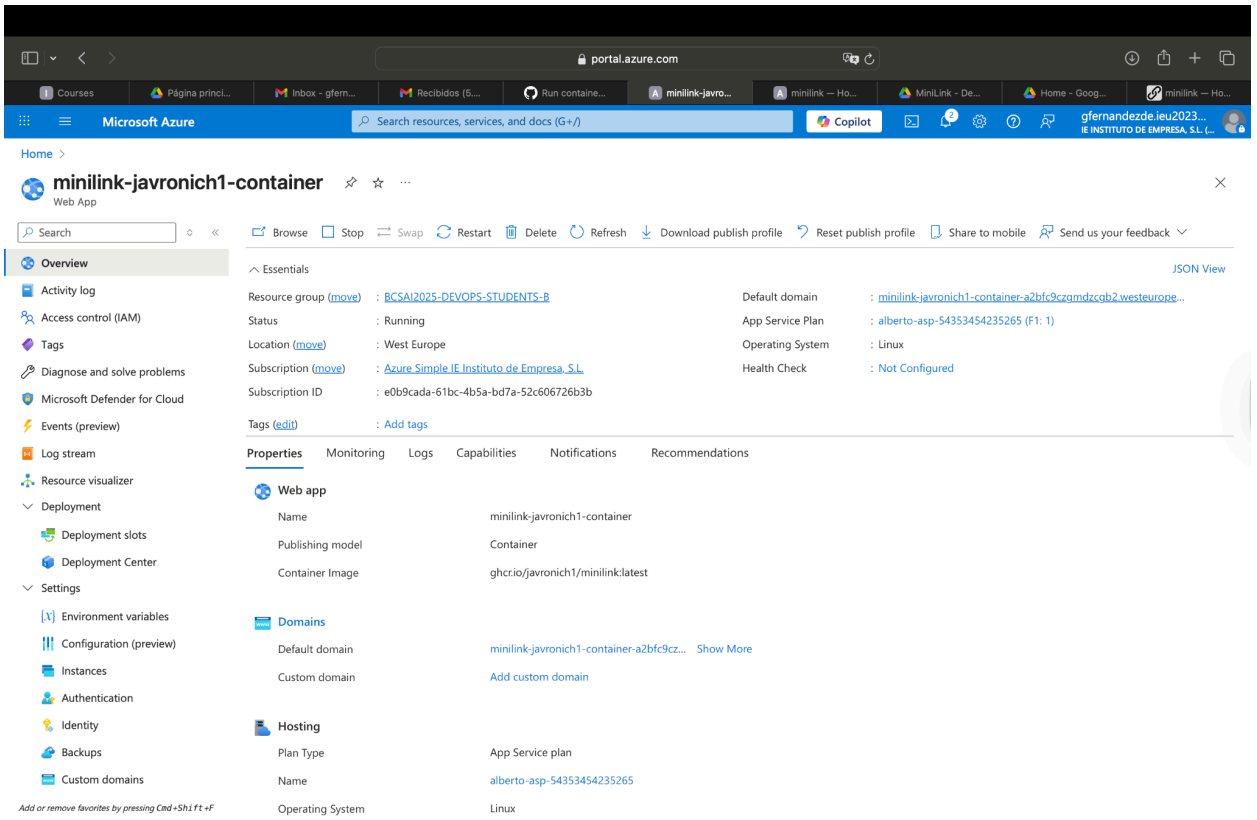
```
# 2) Now copy the actual application code
COPY app ./app
COPY tests ./tests

# Expose the application port (Azure expects 80)
EXPOSE 80

# Default command: run uvicorn on port 80
CMD ["uvicorn", "app.main:app", "--host", "0.0.0.0", "--port", "80"]
```

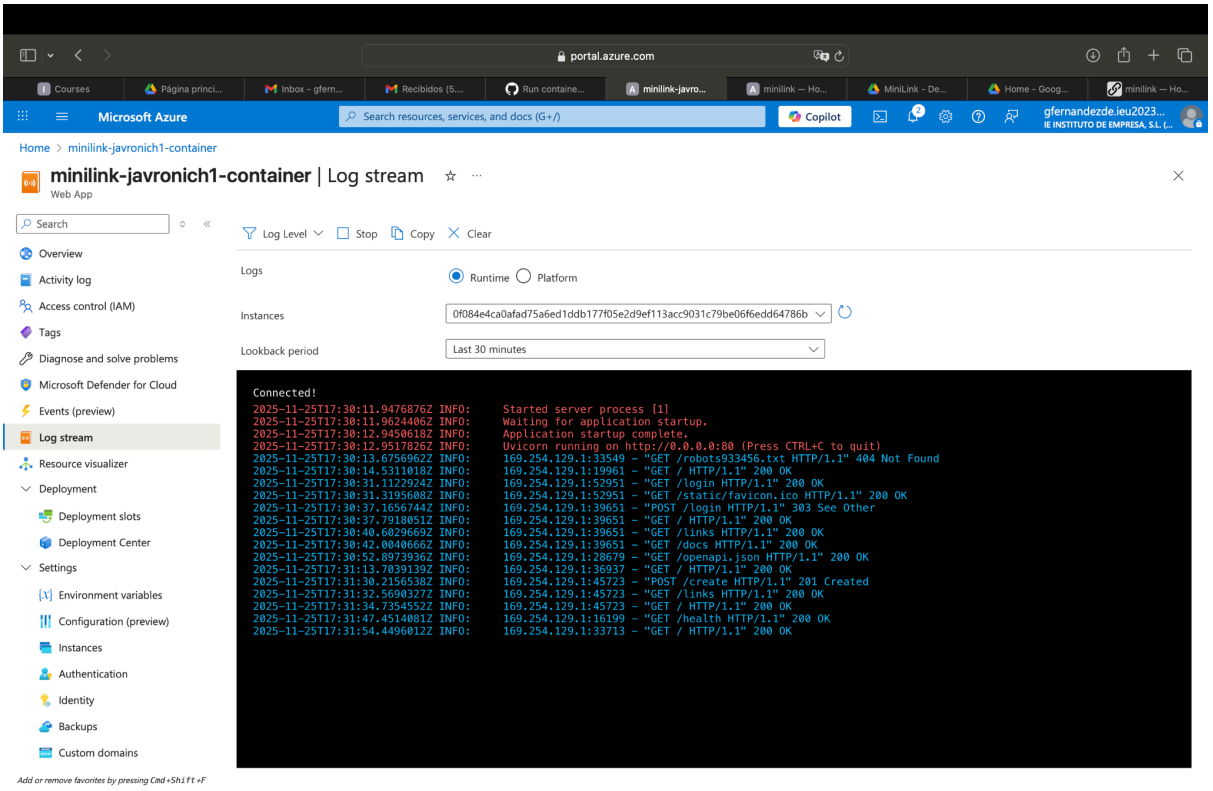
Appendix D - Screenshots

C.1 GitHub Actions (Successful run)



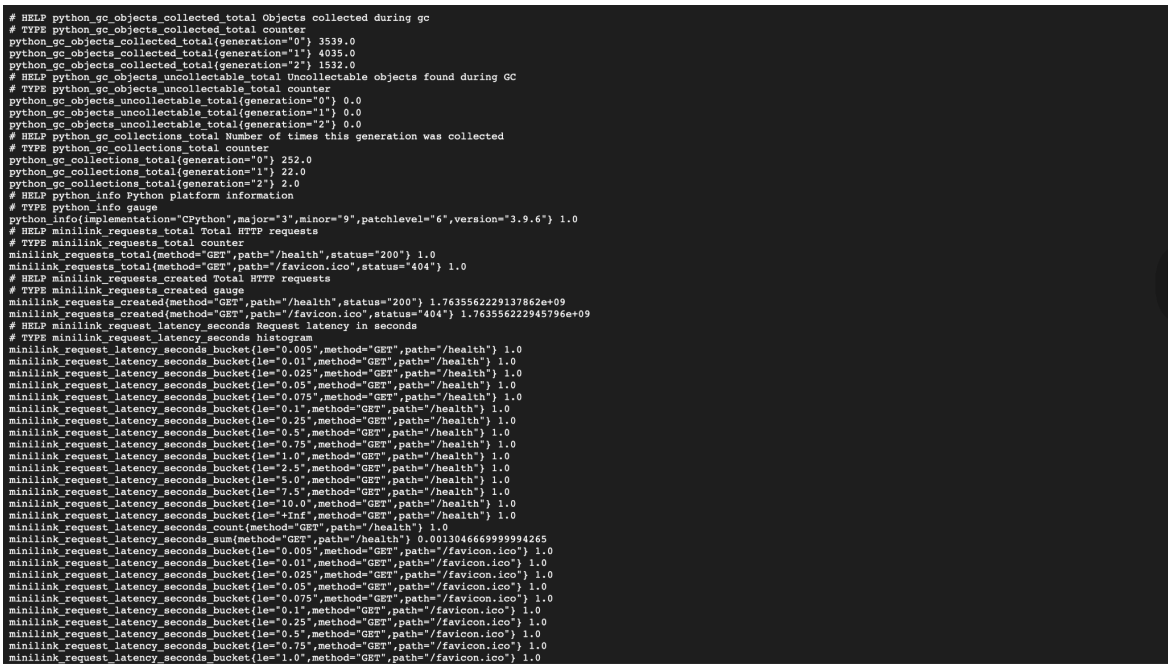
C2. Azure Web App Deployment Dashboard





Link:

C3. Metrics Output



```
# HELP python_gc_objects_collected_total Objects collected during gc
# TYPE python_gc_objects_collected_total counter
python_gc_objects_collected_total{generation="0"} 3539.0
python_gc_objects_collected_total{generation="1"} 4035.0
python_gc_objects_collected_total{generation="2"} 1532.0
# HELP python_gc_objects_uncollectable_total Uncollectable objects found during GC
# TYPE python_gc_objects_uncollectable_total counter
python_gc_objects_uncollectable_total{generation="0"} 0.0
python_gc_objects_uncollectable_total{generation="1"} 0.0
python_gc_objects_uncollectable_total{generation="2"} 0.0
# HELP python_gc_collections_total Number of times this generation was collected
# TYPE python_gc_collections_total counter
python_gc_collections_total{generation="0"} 252.0
python_gc_collections_total{generation="1"} 22.0
python_gc_collections_total{generation="2"} 2.0
# HELP python_info Python platform information
# TYPE python_info gauge
python_info{implementation="CPython",major="3",minor="9",patchlevel="6",version="3.9.6"} 1.0
# HELP minilink_requests_total Total HTTP requests
# TYPE minilink_requests_total counter
minilink_requests_total{method="GET",path="/health",status="200"} 1.0
minilink_requests_total{method="GET",path="/favicon.ico",status="404"} 1.0
# HELP minilink_requests_created Total HTTP requests
# TYPE minilink_requests_created gauge
minilink_requests_created{method="GET",path="/health",status="200"} 1.7635562229137862e+09
minilink_requests_created{method="GET",path="/favicon.ico",status="404"} 1.763556222945796e+09
# HELP minilink_request_latency_seconds Request latency in seconds
# TYPE minilink_request_latency_seconds histogram
minilink_request_latency_seconds_bucket{le="0.005",method="GET",path="/health"} 1.0
minilink_request_latency_seconds_bucket{le="0.01",method="GET",path="/health"} 1.0
minilink_request_latency_seconds_bucket{le="0.025",method="GET",path="/health"} 1.0
minilink_request_latency_seconds_bucket{le="0.05",method="GET",path="/health"} 1.0
minilink_request_latency_seconds_bucket{le="0.075",method="GET",path="/health"} 1.0
minilink_request_latency_seconds_bucket{le="0.1",method="GET",path="/health"} 1.0
minilink_request_latency_seconds_bucket{le="0.25",method="GET",path="/health"} 1.0
minilink_request_latency_seconds_bucket{le="0.5",method="GET",path="/health"} 1.0
minilink_request_latency_seconds_bucket{le="0.75",method="GET",path="/health"} 1.0
minilink_request_latency_seconds_bucket{le="1.0",method="GET",path="/health"} 1.0
minilink_request_latency_seconds_bucket{le="2.5",method="GET",path="/health"} 1.0
minilink_request_latency_seconds_bucket{le="5.0",method="GET",path="/health"} 1.0
minilink_request_latency_seconds_bucket{le="7.5",method="GET",path="/health"} 1.0
minilink_request_latency_seconds_bucket{le="10.0",method="GET",path="/health"} 1.0
minilink_request_latency_seconds_bucket{le="+Inf",method="GET",path="/health"} 1.0
minilink_request_latency_seconds_count{method="GET",path="/health"} 1.0
minilink_request_latency_seconds_sum{method="GET",path="/health"} 0.00130466699999994265
minilink_request_latency_seconds_bucket{le="0.005",method="GET",path="/favicon.ico"} 1.0
minilink_request_latency_seconds_bucket{le="0.01",method="GET",path="/favicon.ico"} 1.0
minilink_request_latency_seconds_bucket{le="0.025",method="GET",path="/favicon.ico"} 1.0
minilink_request_latency_seconds_bucket{le="0.05",method="GET",path="/favicon.ico"} 1.0
minilink_request_latency_seconds_bucket{le="0.075",method="GET",path="/favicon.ico"} 1.0
minilink_request_latency_seconds_bucket{le="0.1",method="GET",path="/favicon.ico"} 1.0
minilink_request_latency_seconds_bucket{le="0.25",method="GET",path="/favicon.ico"} 1.0
minilink_request_latency_seconds_bucket{le="0.5",method="GET",path="/favicon.ico"} 1.0
minilink_request_latency_seconds_bucket{le="0.75",method="GET",path="/favicon.ico"} 1.0
minilink_request_latency_seconds_bucket{le="1.0",method="GET",path="/favicon.ico"} 1.0
```