MLOps:
London Fire Brigade Response Time Prediction

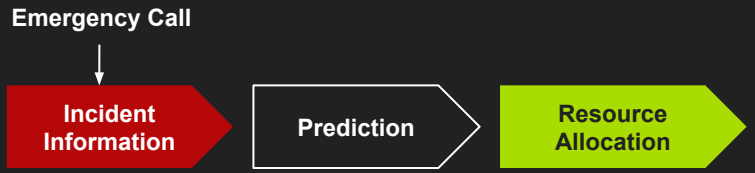
Automated model retraining, deployment, and API monitoring

René Groß
Ismarah Maier
Dr. Benjamin Schellinger

Business Case: London Fire Brigade (LFB)

Problem Statement & Goal

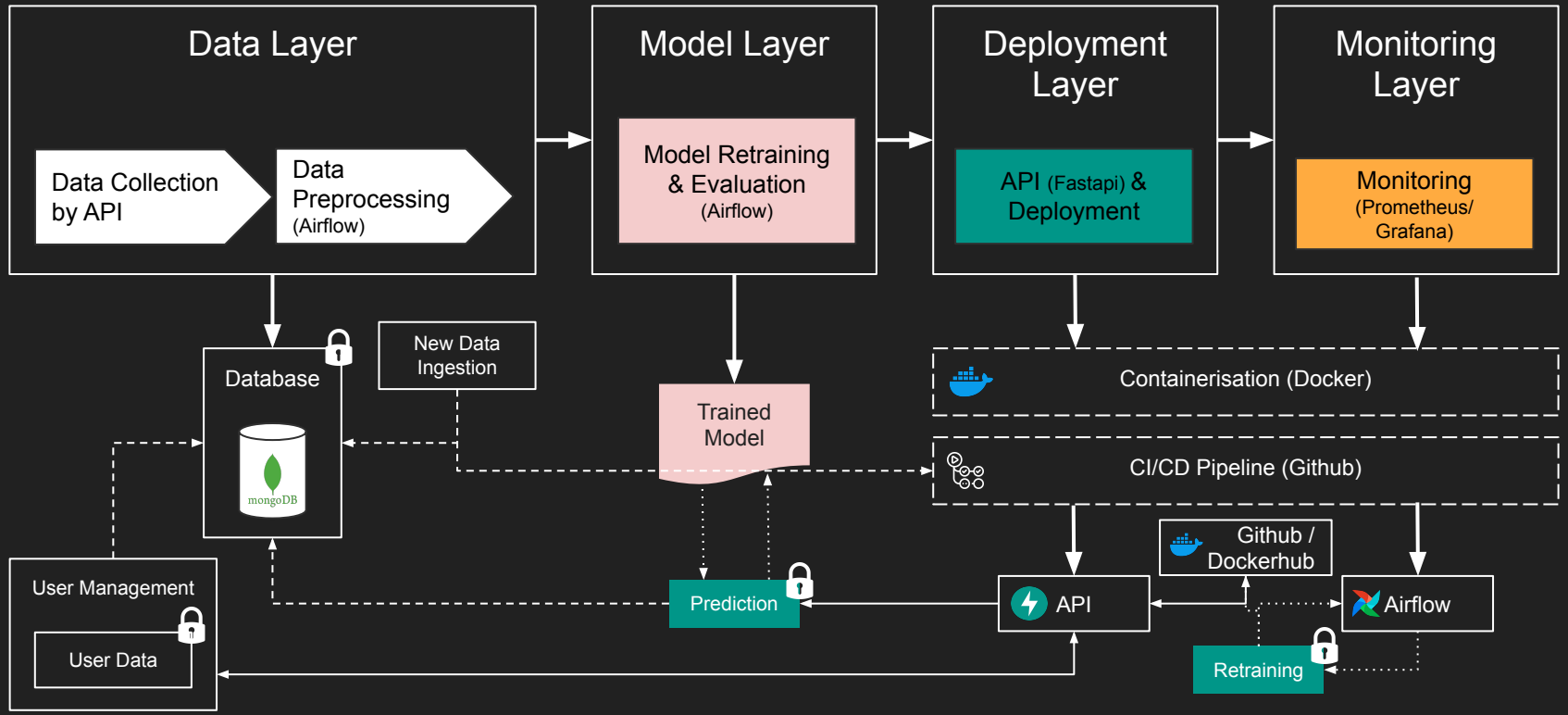
- Predict accurate response times (> or < 6 min)
- Improvement of resource allocation critical_
    - > 6 min -> additional police forces



MLOps-Project Objective

- Automate model retraining and deployment
- Ensure reliable CI/CD pipeline
- LFB app & its performance monitoring

High-Level Architecture



API Endpoints & Securitization

Endpoint	Method	Roles	Description
/token	**POST**	**All Users**	**Generate JWT token using OAuth2 protocol**
/user 🔒	GET / POST / DELETE	**Admin**	Manage user accounts
/predict 🔒	POST	User / **Admin**	Sends request to model to predict response time
/evaluate 🔒	POST	User / **Admin**	Retrain & evaluate Model
/database 🔒	GET / POST / DELETE	**Admin**	Query database, add new & delete old data

Secret Management:

- .env files
- GitHub Secrets

Isolation via Containerization

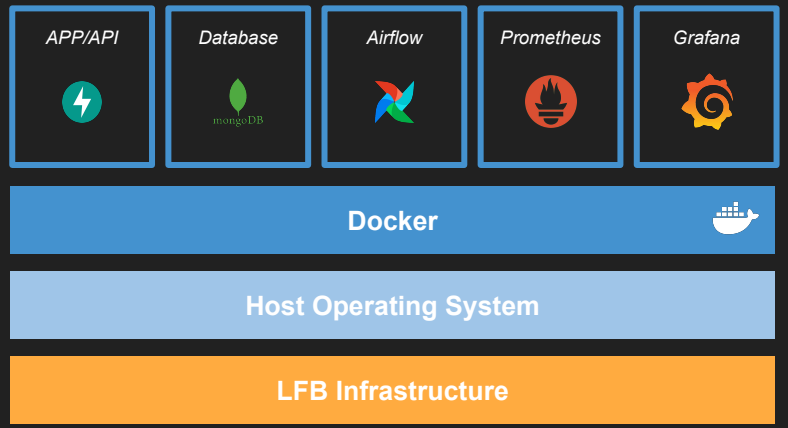
docker-compose.yml

```yaml
version: '3.13'

services:
  api:
    build:
      context: .
      dockerfile: Dockerfile
    container_name: lfb_api
    ports:
      - "8000:8000"
    environment:
      SECRET_KEY: ${SECRET_KEY}
      ALGORITHM: ${ALGORITHM}
      PYTHONPATH: ${PYTHONPATH}
      MONGO_URI: ${MONGO_URI}
    depends_on:
      - mongodb
    networks:
      - lfb_network

  mongodb:
    image: mongo:4.4
    container_name: mongodb
    restart: always
    environment:
      MONGO_INITDB_ROOT_USERNAME: ${MONGO_INITDB_ROOT_USERNAME}
      MONGO_INITDB_ROOT_PASSWORD: ${MONGO_INITDB_ROOT_PASSWORD}
    ports:
      - "27017:27017"
    networks:
      - lfb_network
    command: ["mongod", "--auth", "--bind_ip_all", "--port", "27017"]
```

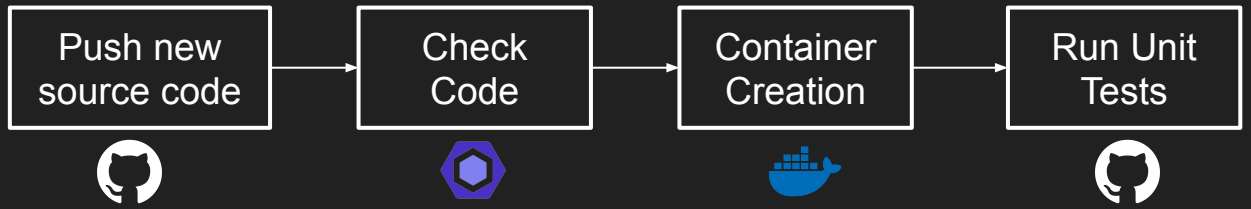
Containerized Applications



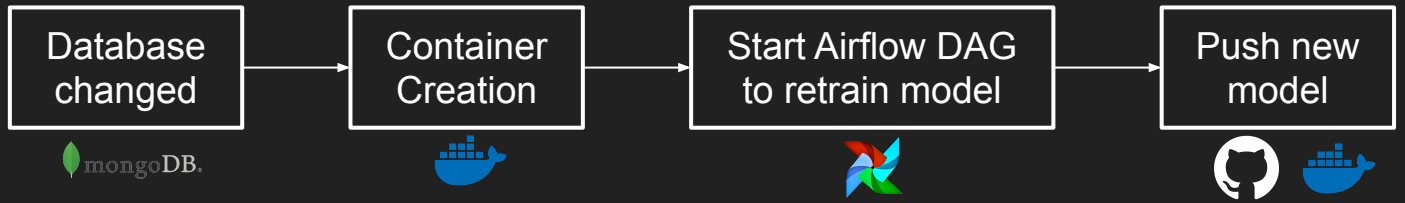
- ➔ Scalability
- ➔ Reproducibility
- ➔ Environment Parity

Automation & CI/CD Pipeline

CI Pipeline



CD Pipeline



Monitoring & Logging using Prometheus & Grafana