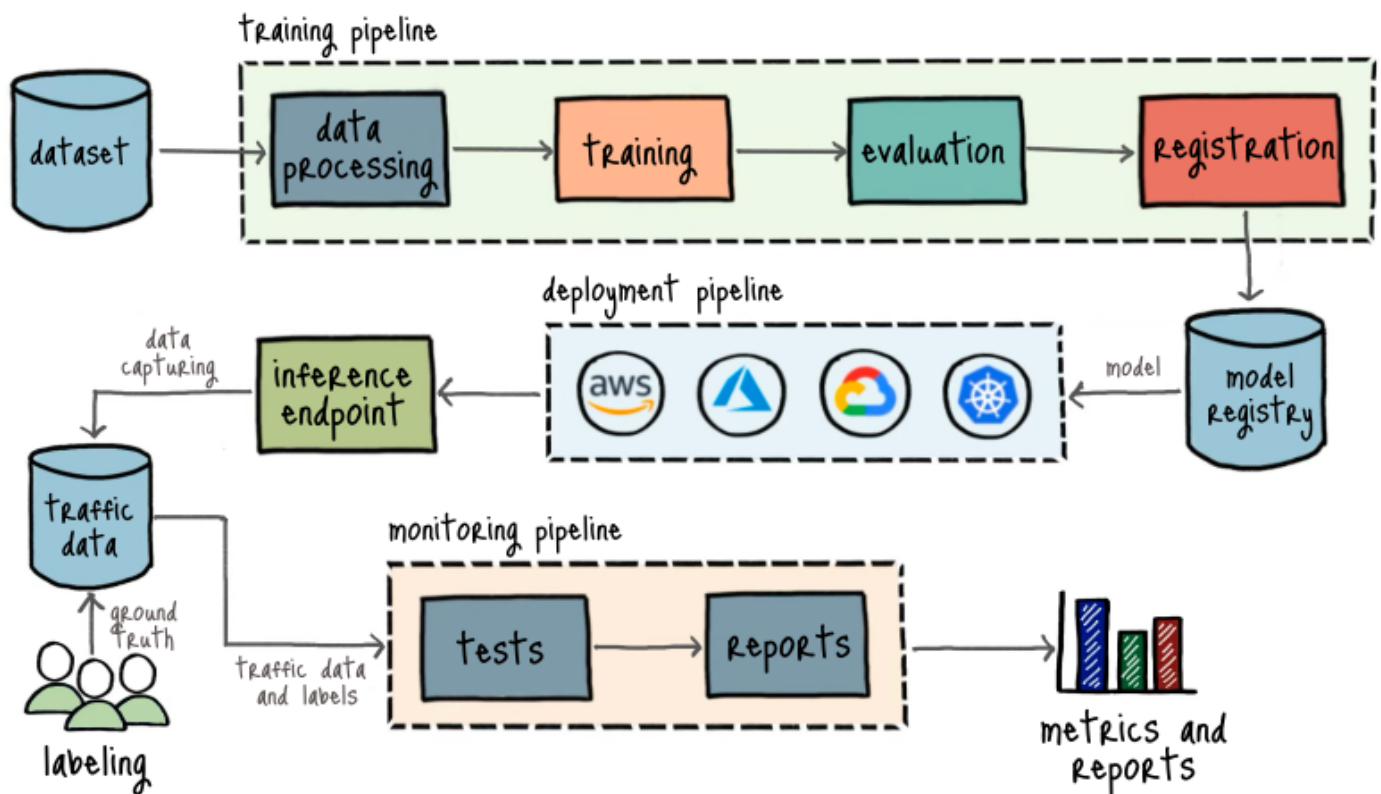


# High-level Architecture



## Mối quan hệ giữa các thành phần

1. **Training Pipeline** tạo ra mô hình → đẩy vào **Model Registry**.
2. **Deployment Pipeline** lấy mô hình từ registry → triển khai trên AWS → cung cấp API inference.
3. **Traffic Data** cập nhật **ground truth** → dùng để cải thiện mô hình qua feedback loop.

Dưới đây là giải thích chi tiết từng thành phần trong kiến trúc bạn cung cấp, được chia theo 3 pipeline chính:

## 1. Training Pipeline (Quy trình huấn luyện mô hình)

### Dataset

- **Ý nghĩa:** Tập dữ liệu đầu vào để huấn luyện mô hình AI/ML.
- **Chi tiết:**
  - Có thể bao gồm dữ liệu lịch sử giao thông (ví dụ: hình ảnh camera, cảm biến, dữ liệu GPS).
  - Thường được chia thành **train/validation/test sets**.

## Data processing

- **Ý nghĩa:** Giai đoạn tiền xử lý dữ liệu thô.
- **Chi tiết:**
  - **Làm sạch dữ liệu:** Loại bỏ nhiễu, giá trị thiếu, dữ liệu trùng lặp.
  - **Chuẩn hóa:** Đưa dữ liệu về cùng định dạng (ví dụ: resize ảnh, scaling số liệu).
  - **Feature engineering:** Tạo đặc trưng mới (ví dụ: thời gian cao điểm, thời tiết).

## Training

- **Ý nghĩa:** Huấn luyện mô hình học máy từ dữ liệu đã xử lý.
- **Chi tiết:**
  - Sử dụng các thuật toán như Neural Networks, Random Forest, SVM...
  - Mục tiêu: Học các patterns từ dữ liệu giao thông (ví dụ: dự đoán ùn tắc).

## Evaluation

- **Ý nghĩa:** Đánh giá hiệu suất mô hình.
- **Chi tiết:**
  - Dùng metrics như **Accuracy, Precision, Recall, F1-score** (với bài toán phân loại) hoặc **MAE, RMSE** (với bài toán hồi quy).
  - Kiểm tra overfitting/underfitting bằng validation set.

## Registration

- **Ý nghĩa:** Lưu trữ mô hình đã huấn luyện vào kho quản lý.
- **Chi tiết:**
  - **Model Registry** (ví dụ: MLflow, AWS SageMaker Model Registry) giúp versioning, tracking.
  - Cho phép triển khai lại mô hình cũ nếu mô hình mới hoạt động kém.

## 2. Deployment Pipeline (Quy trình triển khai mô hình)

### Data capturing

- **Ý nghĩa:** Thu thập dữ liệu mới từ hệ thống thực tế.
- **Chi tiết:**
  - Ví dụ: Ảnh từ camera giao thông, dữ liệu cảm biến IoT.
  - Có thể sử dụng **Kafka** hoặc **AWS Kinesis** để xử lý real-time.

### Inference endpoint

- **Ý nghĩa:** Điểm kết nối để gọi mô hình dự đoán.
- **Chi tiết:**

- Triển khai dưới dạng **API** (REST/gRPC) hoặc **serverless function** (AWS Lambda).
- Nhận input (ví dụ: ảnh giao thông) → trả kết quả (ví dụ: mức độ ùn tắc).

## AWS

- **Ý nghĩa:** Nền tảng cloud để triển khai hệ thống.
- **Chi tiết:**
  - **SageMaker:** Huấn luyện và triển khai mô hình.
  - **EC2/Lambda:** Chạy inference.
  - **S3:** Lưu trữ dữ liệu.

## Model

- **Ý nghĩa:** Mô hình đã được huấn luyện sẵn sàng triển khai.
- **Chi tiết:**
  - Được lấy từ **Model Registry** trong Training Pipeline.
  - Đóng gói thành container (Docker) để deploy.

## Model Registry

- **Ý nghĩa:** Kho lưu trữ các phiên bản mô hình.
- **Chi tiết:**
  - Theo dõi metadata: Hiệu suất, ngày huấn luyện, người tạo.
  - Hỗ trợ rollback nếu cần.

## 3. Traffic Data (Dữ liệu giao thông)

### Ground truth

- **Ý nghĩa:** Dữ liệu thực tế được gán nhãn chính xác.
- **Chi tiết:**
  - Ví dụ: Ảnh giao thông được cảnh sát gán nhãn "kẹt xe" hoặc "thông thoáng".
  - Dùng để so sánh với kết quả dự đoán của mô hình.

### Traffic data and labels

- **Ý nghĩa:** Dữ liệu thô + nhãn tương ứng.
- **Chi tiết:**
  - **Dữ liệu:** Lưu lượng xe, tốc độ di chuyển, thời gian.
  - **Nhãn:** Mức độ ùn tắc (0-10), loại sự cố (tai nạn, đường hư hỏng).

## Testing

- **Ý nghĩa:** Kiểm thử mô hình trên dữ liệu mới.
- **Chi tiết:**
  - **A/B testing:** So sánh mô hình cũ vs mới.

- **Canary deployment:** Triển khai thử nghiệm trên một phần hệ thống.

## Reports & Metrics and reports

- **Ý nghĩa:** Đo lường hiệu quả và báo cáo.
- **Chi tiết:**
  - **Metrics:** Precision/Recall của mô hình, độ trễ inference.
  - **Reports:** Xuất file PDF/CSV hoặc dashboard (Power BI, Grafana).

**Tác giả: Đỗ Ngọc Tú**  
**Công Ty Phần Mềm VHTSoft**

---

Phiên bản #1

Được tạo 19 tháng 4 2025 14:23:02 bởi Đỗ Ngọc Tú

Được cập nhật 22 tháng 4 2025 17:41:28 bởi Đỗ Ngọc Tú