

Vai trò	Sprint 1: Nền tảng & Thiết kế	Sprint 2: Phát triển Độc lập	Sprint 3: Tích hợp & Chạy thử	Sprint 4: Phân tích & Hoàn thiện
Mục tiêu Chính	Hoàn tất thiết kế & Cài đặt hạ tầng.	Hoàn thiện các thành phần riêng lẻ.	Chạy luồng End-to-End (E2E) với 10k ảnh.	Xử lý dữ liệu tăng cường (5k ảnh) & Hoàn thiện báo cáo.
M1. Lại Trí Dũng (Link-up & Spark Architect)	<ul style="list-style-type: none"> * Hoàn thiện kiến trúc dự án. * Data contract: Định nghĩa schema JSON cho Kafka & MongoDB. * Thiết lập Git repo. 	<ul style="list-style-type: none"> * Viết khung PySpark Streaming App. * Code logic đọc từ Kafka và ghi vào MongoDB. * Chạy thử Kafka -> Spark -> Mongo với dữ liệu giả lập. 	<ul style="list-style-type: none"> * Tích hợp UDFs vào Spark App. * Thêm logic fetch LMDB từ MinIO. * Triển khai Spark App hoàn chỉnh lên K8s (với P3). * Chủ trì buổi test E2E. 	<ul style="list-style-type: none"> * Đảm bảo hệ thống xử lý tăng cường 5k ảnh mới. * Tối ưu hiệu năng (nếu cần). * Biên soạn (Compile) báo cáo cuối cùng từ các thành viên.
M2. Nguyễn Trọng Tâm (Data Engineer)	<ul style="list-style-type: none"> * Phân tích ảnh .jpg & label .txt. * Viết Spark Batch Job (local) để đọc dữ liệu thô và đóng gói thành 1 file LMDB. 	<ul style="list-style-type: none"> * Hoàn thiện script LMDB và upload file LMDB (10k ảnh) lên MinIO. * Viết Kafka Producer (Python) để đọc metadata từ LMDB và đẩy 10k bản ghi vào Kafka. 	<ul style="list-style-type: none"> * Chạy Kafka Producer để đẩy 10k metadata. * Theo dõi log và xác nhận đẩy thành công. 	<ul style="list-style-type: none"> * Xử lý 5k ảnh mới vào LMDB và Upload MinIO. * Chạy Producer với 5k metadata mới. * Viết tài liệu phần Ingestion & Pre-processing.
M3. Vũ Hữu An (DevOps Engineer)	<ul style="list-style-type: none"> * Cài đặt Kubernetes (K8s). * Triển khai các dịch vụ nền tảng: MinIO, Kafka, MongoDB. * Cung cấp "Connection Guide" (endpoints, ports). 	<ul style="list-style-type: none"> * Viết Dockerfiles cho app của P1 và P2. * Viết file YAML (.yaml) để triển khai ứng dụng Spark (P1) và Producer (P2) lên K8s. 	<ul style="list-style-type: none"> * Hỗ trợ P1 triển khai Spark App. * Theo dõi tài nguyên K8s (CPU/RAM) khi chạy E2E. * Gỡ lỗi (troubleshoot) kết nối mạng. 	<ul style="list-style-type: none"> * Hoàn thiện toàn bộ file YAML & Dockerfiles. * Viết tài liệu Hướng dẫn triển khai (Deployment Strategy) & phần Hạ tầng.
M4. Lưu Thiện Việt Cường (Data Scientist)	<ul style="list-style-type: none"> * Viết các hàm Python local (dùng OpenCV) để phân tích label .txt, tính mật độ và diện tích trống. * Lựa chọn mô hình. 	<ul style="list-style-type: none"> * Chuyển đổi các hàm Python (tính mật độ) thành Pandas UDFs cho Spark. * Đóng gói UDFs thành module Python (.py) và chuyển giao cho P1. 	<ul style="list-style-type: none"> * Kiểm tra tính đúng đắn (Validate) của kết quả trong MongoDB. * Sửa lỗi logic UDFs nếu cần. 	<ul style="list-style-type: none"> * Validate kết quả của 5k ảnh mới. * Viết tài liệu phần Thuật toán & Logic AI (UDFs).
M5. Đàm Quang Đức (BI & Data Analyst)	<ul style="list-style-type: none"> * Cài đặt Grafana. * Kết nối Grafana với MongoDB (do P3 cung cấp). * Phác thảo (Mockup) dashboard dựa trên Data Contract. 	<ul style="list-style-type: none"> * Xây dựng Dashboard Grafana hoàn chỉnh (biểu đồ, bảng) dựa trên dummy data. * Tinh chỉnh các câu truy vấn MongoDB. 	<ul style="list-style-type: none"> * Theo dõi dashboard khi P2 chạy Producer. * Xác nhận dữ liệu "live" hiển thị trên Grafana. * Tinh chỉnh lại dashboard . 	<ul style="list-style-type: none"> * Dựa trên 15k ảnh, rút ra các insight (giờ cao điểm, điểm nóng...). * Viết báo cáo phần Problem Definition & Results Analysis.