

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