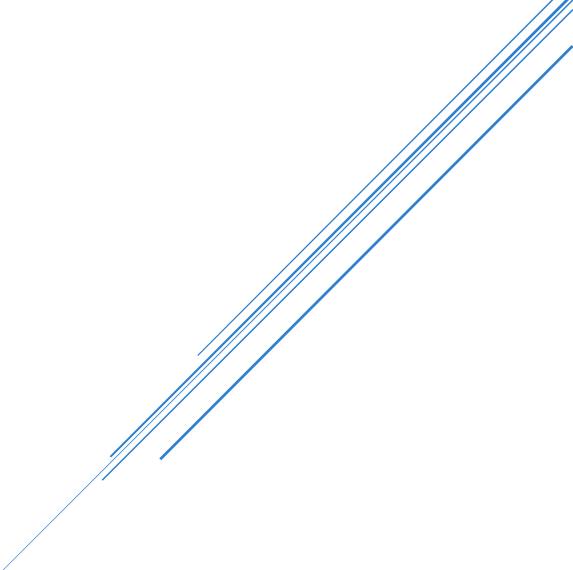
TÊN ĐỀ TÀI: XÂY DỰNG LUÒNG THU THẬP DỮ LIỆU TỪ WEBSITE BÁN KHOÁ HỌC

Nguồn dữ liệu: https://unica.vn/



Trường Đại Học Công Nghiệp TP.HCM Phân tích định lượng

Mã số sinh viên	Họ và tên	Phân công	Tỉ lệ đóng	Điểm số
			góp (%)	(giảng viên)
22695761	Nguyễn Thành Đạt	 Crawl dữ liệu(công nghệ, thiết kế xây dựng) Kiểm soát mã nguồn Xử lý và làm sạch dữ liệu 	36%	
22665311	Thái Văn Sơn	Thiết kế crawl dữ liệuCrawl dữ liệu(marketing)	14%	
22691511	Đỗ Viết Thanh Bằng	 Crawl dữ liệu (ngoại ngữ) Gộp dữ liệu đã crawl (loại bỏ trùng lặp) Đẩy dữ liệu lên mongodb 	12%	
22693501	Trần Tuấn Sang	Crawl dữ liệu(Tài chính kế toán)Thiết kê mô hình ERD	13%	
22696961	Mai Hoàng Tuấn	 Crawl dữ liệu(sức khoẻ-làm đẹp) Tạo bảng và đưa dữ liệu đã làm sạch vào postgres 	13%	
22690271	Đỗ Tiến Đạt	 Crawl dữ liệu(tin học văn phòng) Triển khai hệ thống lên trên máy ảo ubuntu server dựa trên docker test. 	12%	

Trello:

https://trello.com/invite/b/6707e985a3e2c825354f3699/ATTI6405a2570e6ed98c25ad2b7d87a12 3e9BB4A1DAA/unicaproject Ngày thiết lập: 09/09/2024

Github: https://github.com/3002tad/Unica_Project Ngày thiết lập: 09/09/2024

Sơ đồ thiết kế hệ thống

Sơ đồ Data Pipeline

