

# TÊN ĐỀ TÀI: XÂY DỰNG LUỒNG 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 góp (%)	Điểm số (giảng viên)
22695761	Nguyễn Thành Đạt	<ul style="list-style-type: none"><li>Crawl dữ liệu(công nghệ, thiết kế xây dựng)</li><li>Kiểm soát mã nguồn</li><li>Xử lý và làm sạch dữ liệu</li></ul>	36%	
22665311	Thái Văn Sơn	<ul style="list-style-type: none"><li>Thiết kế crawl dữ liệu</li><li>Crawl dữ liệu(marketing)</li></ul>	14%	
22691511	Đỗ Viết Thanh Bằng	<ul style="list-style-type: none"><li>Crawl dữ liệu (ngoại ngữ)</li><li>Gộp dữ liệu đã crawl (loại bỏ trùng lặp)</li><li>Đẩy dữ liệu lên mongodb</li></ul>	12%	
22693501	Trần Tuấn Sang	<ul style="list-style-type: none"><li>Crawl dữ liệu(Tài chính kế toán)</li><li>Thiết kế mô hình ERD</li></ul>	13%	
22696961	Mai Hoàng Tuấn	<ul style="list-style-type: none"><li>Crawl dữ liệu(sức khỏe-làm đẹp)</li><li>Tạo bảng và đưa dữ liệu đã làm sạch vào postgres</li></ul>	13%	
22690271	Đỗ Tiến Đạt	<ul style="list-style-type: none"><li>Crawl dữ liệu(tin học văn phòng)</li><li>Triển khai hệ thống lên trên máy ảo ubuntu server dựa trên docker test.</li></ul>	12%	

Trello:  
<https://trello.com/invite/b/6707e985a3e2c825354f3699/ATTI6405a2570e6ed98c25ad2b7d87a123e9BB4A1DAA/unicaproject>  
Ngày thiết lập: 09/09/2024

Github: [https://github.com/3002tad/Unica\\_Project](https://github.com/3002tad/Unica_Project)  
Ngày thiết lập: 09/09/2024

Sơ đồ thiết kế hệ thống

Sơ đồ Data Pipeline

