Báo cáo cuối kỳ Nhập Môn BigData

👥 Phân công nhiệm vụ cho 3 thành viên

Thành viên Trách nhiệm chính Kỹ năng cần thiết

Người 1 Thu thập dữ liệu (Data Crawling)

– Viết crawler Tiki.vn (API)

– Viết crawler VnExpress (RSS/XML)

– Chạy định kỳ để lấy dữ liệu lớn Python, requests, JSON, RSS, HTML

Người 2 Xử lý, lưu trữ & Hệ quản trị dữ liệu (DBMS)

– Thiết kế CSDL (MySQL / MongoDB)

– Làm sạch dữ liệu (null, trùng, sai kiểu)

– Đưa dữ liệu vào HDFS, MongoDB hoặc MySQL MySQL, MongoDB, Pandas, Hadoop

Người 3 Hadoop Ecosystem & Phân tích

– Cài đặt Hadoop trên Docker / Windows

– Thực hiện MapReduce xử lý dữ liệu

– (Optional) Xây dựng giao diện đơn giản + trực quan hóa dữ liệu Hadoop, Hive, HDFS, MapReduce