Title: Building data pipeline to collect Data from website

Data source: https://goodreads.com/list/show/19.Best\_for\_Book\_Clubs

Industrial university of Ho Chi Minh city

for Data Engineer

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Student ID | Fullname | Task | Contribution rate (%) | Score (lecturer) |
|  | (\*) Leader information on first row |  |  |  |
| 22652711 | Võ Tất Thiên | Website survey, project progress management | 15% |  |
| 22724461 | Đỗ Công Danh | Crawl Web | 13% |  |
| 22653661 | Võ Nhật Trường | Putting data in MongoDB | 13.5% |  |
| 21107481 | Võ Nguyễn Hoành Hợp | Build a Postgres database | 12.5% |  |
| 22633701 | Trần Quang Vinh | Building Tables | 8% |  |
| 22653331 | Nguyễn Thành Kiệt | Data Cleaning by pyspark, Postgres Inclusion, Virtual Machine Demo | 30% |  |
| 22639591 | Nguyễn Đình Huy Hùng | ERD Construction | 8% |  |

Trello URL: <https://trello.com/b/0FpryE2G/crawldata> Established Date: 25/08/2024

Github URL: <https://github.com/VoTatThien/CrawlWeb_Nhom> Established Date: 23/08/2024

Data Pipeline System Architecture DesignA diagram of a diagram

Description automatically generated

A diagram of a diagram

Description automatically generated

A screenshot of a computer

Description automatically generated