**COMP1204 Data Management**

**Coursework-2 LINUX**

**Group Members:**

|  |  |  |
| --- | --- | --- |
| **No.** | **Name** | **University Email** |
| 1. | Brandon Ting Wee Kang | bwkt1n22@soton.ac.uk |
| 2. | Pang Jia Hui | jhp1n22@soton.ac.uk |

**Submitted Documents in ZIP File:**

|  |  |  |
| --- | --- | --- |
| **No.** | **Documents** | **File Name** |
| 1. | Report |  |
| 2. | Scraper/Tracker Script: |  |
| 3. | Database |  |
| 4. | Github Repository |  |

**Setup Development Environment**

Text editor: VSCode (write .sh file)

Run .sh file: Ubuntu WSL

Database: XAMPP

Version Control: Github

Website URL:

<https://www.lazada.com.my/>

Products URL:

<https://www.lazada.com.my/products/smartphone-galaxy-s22-ultra-sale-original-big-sale-2022-65inch-12gb-ram-512gb-rom-android-cellphone-on-sale-5000mah-wifi-bluetooth-5g-smartphone-online-learning-google-game-phone-legal-cell-cellphone-lowest-price-mobile-phone-free-shipping-cod-i3177845655-s19763201563.html>

My own refer: (don’t delete 😊)

cd /mnt/c/Users/Hp/Desktop/OneDrive\ -\ University\ of\ Southampton/Projects/Others/OnlinePromotionStoreTracker

# 1.0 Introduction

Blah blah blah. Halo help me to write introdcution thanksssssssssssssssssssss

# 2.0 Table of Content

# 3.0 Timeline

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Phase** | **Process** | **Week1** | | | | | | | **Week2** | | | | | | | **Week3** |
| **Date** | **29-Apr** | **30-Apr** | **1-May** | **2-May** | **3-May** | **4-May** | **5-May** | **6-May** | **7-May** | **8-May** | **9-May** | **10-May** | **11-May** | **12-May** | **13-May** |
| **Database Design** | Design the Schema |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Create Database |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Scraper/ Tracker Script** | Fetching Web Data |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Parsing Data |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Data Manipulation |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Insert into Database |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Crontab Setup** | Scheduling |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Plotting Script** | Create script |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Plot Types |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Report Writing** | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Github Repository** | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Report Submission** | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

# 4.0 Database Design

* + **Design the Schema:** Define the database schema in MySQL with necessary tables and relationships to store product tracking data, ensuring proper normalization.
  + **Create the Database:** Write and execute the SQL script to create the database and tables.

# 5.0 Scraper/Tracker Script

* + **Fetching Web Data:** Write a shell script that uses **curl** or similar tools to fetch the HTML content of the product pages.
  + **Parsing Data:** Use tools like **grep** and **awk** to extract relevant information (price, stock, rating) from the HTML content.
  + **Data Manipulation:** Process and convert extracted data into appropriate formats (e.g., numeric, timestamp).
  + **Insert into Database:** Add MySQL commands to insert collected data into the database, ensuring error handling for database interactions.

# 6.0 Contab Setup

* + **Scheduling:** Set up a crontab entry to run the scraper script at regular intervals, such as daily or hourly, to ensure continuous data collection.

# 7.0 Plotting Script

* + **Create the Script:** Write a shell script to generate at least 10 different plots using GNU Plot or similar tools.
  + **Plot Types:** Implement plotting functions that show trends such as:
    - Price changes over time.
    - Stock changes over time.
    - Rating changes over time.
  + **Script Organization:** Ensure the plotting script is organized into easily callable functions.

# 8.0 Conclusion

Blah blah. Help me to writeeeeeeee thankssssssss

# 9.0 Appendices

Blah blah