**Practical Take-home Assignment**

The GOAL of this take-home task is to ultimately do research and get our hands dirty by discovering how to prepare working environments to run some selected Big Data tools within an Oracle VirtualBox.

*Note:* “A mind ready to dive into technical research, with proper documentation at every stage and eagerness to fix missing dots is probably going to be your friend on this journey”.

**1. Essential Downloads**

**A]** Download Oracle VirtualBox 7.1.8 for windows hosts and appropriate extension pack(s).

**B]** Download Ubuntu 24.04.2 LTS from its official site.

**C]** Download Kaggle Chronic illness Dataset from:

<https://www.kaggle.com/datasets/flaredown/flaredown-autoimmune-symptom-tracker>

* Any other downloads you may feel are necessary for you to achieve assigned tasks.

**2. Virtual Box installation**

**Required:** Install **A]** with its components on a computer / laptop running windows 10 or 11 operating system. Check to ensure all virtual hardware are setup correctly in preparation for **B].**

**3. Ubuntu Installation**

**Required:**

* Prepare a dynamic **VDI** of 30Gb
* You are required to do your own research and install **B]** inside **2.**
* **Account / Username** and **Password** for the Ubuntu operating system installed **SHOULD** both be your student number.

**4.** **Big Data Analysis with Hadoop MapReduce**

**Required:**

* Do your own research and prepare or install HDFS environment(s) for Hadoop MapReduce operations inside **3.**
* Using the downloaded Kaggle dataset **C],** find, provide, and extract insights and knowledge from the data using Hadoop MapReduce operations.
* Provide an explanation of all insights harvested from the dataset using a ppt. (This MUST include a live video of your self explaining or describing your data insights found). Any appropriate screen recorder to achieve this can be used as long as it is documented.

**Note:**

* You MUST screenshot and save all your steps in an MsWord file.
* You MUST keep a record of all your programming source codes or command codes used at every stage where appropriate.
* Anywhere not mentioned needing a username and password, please use your student number for both.

**To be handled in on examination day:**

* A flash disk having your dynamic VDI file, all ppts, MsWord, and video as mentioned above.
* A printed signed document including your ppt and MsWord files.
* Before handing in your work, you MUST show evidence of ALL the tasks on your laptop.

===END===