CDS502 Big Data Storage and Management
Semester 1, 2021/2022

ASSIGNMENT – REPORT (20%)
(Group Work – **Minimum 2, Maximum 3 members per group**)

Given the challenge that lies ahead with the endless growth of data, many distributions emerge data processing field to handle the rising capacity demands of a Big Data system. The outstanding of these solutions are **HortonWorks, Cloudera, MapR, IBM Infosphere BigInsights, Pivotal and Microsoft HD Insight,** to name a few. Write a report based on your review of the literature in a topic related to big data storage and management technologies. Perform the **comparative study** on any **THREE (3)** solutions mentioned above. Below are some points that guide you in preparing the report.

1. Format of the report [Standard IEEE format]:
    - i. *Abstract*
    - ii. *Introduction (Brief background)*
    - iii. *History and evolution of the distributions/services*
    - iv. *Highlights of the distributions/services and its components*
    - v. *Comparisons among the distributions/services and its components (Analysis/Results)*
    - vi. *Discussion and recommendation (Application domain)*
    - vii. *Conclusion*
    - viii. *Clear division of group members' roles*
    - ix. *References*

2. Find at least ***five relevant papers*** that are related to the above topic. Relevant papers refer to papers published in journals, conference proceedings or technical reports [Tips: IEEE Xplore, ACM, Elsevier Science Direct and etc]. Web pages, FAQs or presentation slides are not counted as relevant papers. However, they are counted as other references besides the five required relevant papers. Cite them as your references properly. Your deliverables must include the following requirements:
    - Report format & template: IEEE paper format (Hardcopy & Softcopy)
        - o http://www.ieee.org/conferences_events/conferences/publishing/templates.html
    - Failing to comply with the format will affect your overall assignment grade
    - Recorded digital presentation of the report - video (not more than 15 minutes)
        - o Tips: Record your PowerPoint presentation
        - o You can also use other tools to prepare your video
    - Try to make the report brief yet concise (Proofread your work)
    - Expected length of the report: between 5 (min.) and 7 (max.) pages
    - Marking Scheme: refer to the rubrics posted on the e-learning course page

3. At the end of your report, in the comparison and discussion section:
    - Compare their features, strength and weakness
    - Discuss and give recommendations on application of the distributions/services on real-life big data (eg. Healthcare, Public sector, Education, Insurance Services, Industry, Transportation and Finance & Crime Detection – *Choose 3 out of 7 domains*)
    - Your opinions on the future of the distributions/services
    - You can include some experiments/results in this section (**If any**)

4. Submit the following:
    - Soft copy - (Report + 5 most relevant papers + digital presentation): e-learning

Note:
> You must write the report yourself using your own ideas / words and not simply cut and paste from your references. List all references at the end of the article. ***In the event that parts of the report are directly copied from others without references, F grade is given.***

Deadline for submission is 17th December 2021, by 6.15 pm.