**MANB 1163**

**Assignment 1: Mini Group Project**

**Data Analytics with Hadoop – Collect & Process Data and Build Applications with Hadoop (mark 25)**

**Lecturer: Saiful Adli Ismail**

This project will combine three component hadoop, apache hive, apache hcatalog and apache pig to form as one solution, i.e.:

* Deployment of Hadoop servers to develop, analyze and operate data directly access from Web servers.
* Present charts for extensive data analysis (if you can develop the graph of visualization that would be great). E.g using Ms Excel 2013, Ms Power BI

Each of these components plays a vital role in data analysis. In fact, data analysis itself largely contributes to a thorough decision means in organization. The internal servers could be hosting database application for data storage.

Your task is to install the **Hortonworks server into your virtualbox.** You are freely to set up your own experiment. Testing your setup and collect and process the data based on respective tutorial at http://hortonworks.com/get-started/develop/. In addition, **provide suggestion for the type of data analytic (diagnostic, prescriptive, predictive and** during data analysis from the cloud or Web. E.g (kaggle.com) data set.

**Prepare your report proposal, which includes your analysis of your data in form of chart.**

**Prepare your 10 minutes presentation (ppt) on 26 NOV 2016.**

Submit both files (report and ppt) to my email at saifuladli@utm.my before xx Oct 2017

Note: Don't procrastinate! An early start will help you do a timely and quality project.