Week	Topics
1 21 Apr	Module introduction (Module overview, Learning outcomes, Assessment) **Topic 1:** Introduction to Big Data and its Application - Definition of Big Data - Big Data applications and practices - Overview of Big Data Framework **Lab 1:** Introduction to PySpark on Google Colab -Setting up PySpark - Practical: Introduction to PySpark
2 28 Apr	**Topic 2:** Fundamentals of Big Data - The 5Vs of Big Data -Big Data solution - Big Data Storage Characteristics -Clusters **Lab 2:** Introduction to Apache Spark -Overview of Spark Architecture -Practical (Running Spark and RDD)
3 5 May (HBL)	**Topic and Lab 3:** Introduction to Hadoop (Asynchronous) -Overview of Hadoop -Difference between Hadoop and Spark -Setting up an environment for Hadoop
4 12 May (PH on Monday)	**Topic 3.1**: Classification of Big Data -How to classifiy batch or streaming data **Topic 3.2:** Data Ingestion Tools -The features and characteristics of the tools -Tools for based on different data types **Lab 3**: Apache Spark Dataframes -Understanding Spark Dataframes -DataSet/DataFrame -Hands on with Spark DataFrame
5 19 May	**Topic 5:** Characteristics of Big Data Analytics -Big data pre-processing -Data transformation **Lab 5:** Apache Spark SQL -Understanding Spark SQL -DataSet/DataFrame APIs

	-Hands on with Spark SQL -Data transformation tools -Hands on analysis of big data using Spark
6 26 May	Test Topic 1 to 4 **Lab 1 to 3 only** Proctored Test (MCQ, fill in the blanks,short answer question, close book, access to documentation)
7 2 June	**Topic 6:** Preparing Big Data Dataset -Data Labelling -Data standardisation and vectorisation **Lab 6:**Obtaining insight for data -Processing of big data for insights -Hands on with processing big data using Mlib on Spark
8 9 June	**Topic 7:** Deploy data analytics applications -Data visualisation and application interpretation -Platforms for deploying application Cloud or on prem **Lab 7:** -Hands on with PySpark Classification Model
9 to 10 16 June	**Term Break**
11 30 June	**Topic 8:** Introduction to AI Piplines - Overview of AI Piplines - Pyspark -Pipeline -Data catalog **Lab 8:** -Hands on setting up a pipeline