Department of Management
Amrita Vishwa Vidyapeetham
Amritapuri

Term II (10 September – 09 December 2018)

Course Title:	Cloud and Big Data Analytics
Course Code:	
Credits:	3 credits
Total Sessions:	24 sessions
Course Instructor:	Dr Biplab Bhattacharjee
Contact Information:	biplab@am.amrita.edu
Course Link:	
Office:	
Office hours:	8:45 am-4:45 pm
Course contributes mostly to:	Employability/ Skill Development

Course Description

(Give overall perspective of the course)

Course Objectives (COs)

CO1: Familiarizing the learners with most important information technologies used in manipulating, storing, and analyzing big data

CO2: Learn to build and maintain reliable, scalable, distributed systems with Apache Hadoop and SPARK

CO3: Learn about different deployment models of cloud and different services offered by cloud

CO4: Understand building machine learning models using ANN and Deep learning algorithms on big data

CO5: Understand how to effectively implement machine learning models using TensorFlow in Google Cloud platforms

(Specify the skills/ benefits that students will build during the course. You can have 5 to 7 COs)

Alignment of course objectives (CO) with learning goals (LG) of Assurance of Learning

Each CO should be mapped to the five learning goals set for the MBA program. Below table is given for your reference and should be updated as per your course objectives. Please do not change LGs.

Here you have to mention in writing, how your course is going to contribute to each LG. You may specify the activities or assignments in the course that will ensure attainment of each LG (wherever relevant).

Change the values in the below table according to your Cos.

LG / CO	Critical and integrative Thinking	Effective written and oral communication	Societal and Environmental Awareness	Ethical Reasoning	Leadership
CO1	3	0	0	0	1
CO2	3	1	0	0	0
CO3	3	1	0	0	0
CO4	3	1	1	1	1
CO5	3	0	0	0	0

Key: 3 – Highly relevant; 2 –Moderately relevant; 1 – Low relevance; 0- No relevance

Unit-wise scope for outcomes and Bloom's taxonomy

Please mention how your course is designed to focus on the Bloom's learning levels mentioned in the below table.

After the write-up, update the relevance in below box also.

CO / Bloom's Levels of Learning	CO 1	CO 2	CO 3	CO 4	CO 5
Creating		X			
Evaluating	X	X	X	X	
Analyzing	X		X	X	X
Applying	X	X		X	X
Understanding	X	X	X	X	X
Remembering			X		X

Structure of the course

Pedagogy

- Lecture
- Hands-on data handling sessions
- Group Projects
- Peer-to-peer learning

.

Assessment (Grading Policy: Relative)

S. no	Assessment exercise	Description	Weight
Group assessment			
1	Project	Each group is assigned to work on Deep learning project using Google Colab.	25%
Individual Assessment			
1	Case study analysis	Case study analysis with datasets	10%
2	Assignments	Assignments on sparklyr, TensorFlow, Keras and Google Colab	30%
3	End-term exam	End-exam will consist of 3 hrs test involving both theory and lab sessions. The lab sessions will basically be handling data-intensive tasks	40%

Course Requirements

DA1, DA2 and Introduction to Business Analytics
Prior knowledge of handling spreadsheets
Basic foundation to statistics and mathematics
Knowledge of R and Python programming languages

Course Text

Core Text:

1. Big Data Analytics with R by Simon Walkowiak, Packet Publishing 2016.

Reference Texts:

1. Learning Spark by Matei Zaharia, Patrick Wendell, Andy Konwinski, Holden Karau

2. Ian Goodfellow, Yoshua Bengio, Aaron Courville. Deep Learning

3. Hands-On Machine Learning on Google Cloud Platform by Alexis Perrier, Giuseppe Ciaburro, Et al, Packet Publishing 2018.

4. Cloud Computing by A. Srinivasan, Pearson India, 2014.

Session Plan (please add rows and columns as per your course requirements)

SN NO	TOPIC	CLASS PREPARATION	POST-CLASS READING
1	Introduction to big data, distributed file system, Big Data and its importance, Drivers		Big Data Analytics with R Text Book
2	Big data analytics, Big data applications. Algorithms, Matrix-Vector, Multiplication by Map Reduce		Big Data Analytics with R Text Book
3	Apache Hadoop & Hadoop Ecosystem		Big Data Analytics with R Text Book
4	MapReduce, Data Serialization		Big Data Analytics with R Text Book
5	Introduction to NOSQL, Definition of NOSQL, History of NOSQL and Different NOSQL products, Exploring MondoDB		Big Data Analytics with R Text Book
6	Introduction to Apache Spark, Features of Apache Spark, Apache Spark Architecture		Big Data Analytics with R Text Book
7	Spark Applications, Apache Spark Components, Describe the Different Data Sources and Formats in Spark		Big Data Analytics with R Text Book
8	Spark with R, Set up and apply sparklyr; Use logical verbs in R by applying native sparklyr versions of the verbs		Codes provided in the class
9	Supervised machine learning with sparklyr, Applying sparklyr to machine learning regression and classification models		Codes provided in the class
10	Applying sparklyr to machine learning regression and classification models		Codes provided in the class
11	Introduction to ANN, Activation Functions, Illustrate Perceptron, Training a Perceptron, Important Parameters of Perceptron		
12	Understand limitations of a Single Perceptron, Understand Neural		

	Networks in Detail, Illustrate Multi-Layer Perceptron, Backpropagation – Learning Algorithm, Understand Backpropagation – Using Neural Network Example		
13	Deep Learning Basics		
14	Deep learning Basics		
15	Deep learning with H2O and sparklyr, Applying massive amounts of data to train multi-layer networks for classification		
16	Understand some of the guiding principles behind training deep networks, including		
17	The use of autoencoders, dropout, regularization, and early termination;		
18	Use sparklyr and H2O to train deep networks.		Codes provided in the class
19	Use sparklyr and H2O to train deep networks.		Codes provided in the class
20	Introduction to Cloud Computing, definition, benefits, usage scenarios, History of Cloud Computing - Cloud Architecture Types of Clouds - Business models around Clouds – Major Players in Cloud Computing -		Cloud Computing (Reference Book)
21	Types of Cloud services: Software as a Service - Platform as a Service – Infrastructure as a Service - Database as a Service - Monitoring as a Service – Communication as services. Service providers- Google, Amazon, Microsoft Azure, IBM, Sales force		Cloud Computing (Reference Book)
22	Machine learning with TensorFlow in Google Cloud platform, Introduction to TensorFlow		Codes provided in the class in Google Colab
23	TensorFlow code-basics, Graph Visualization,Constants, Placeholders, Variables, Creating a Model		Codes provided in the class in Google Colab
24	Convolutional neural networks, Deep Belief Nets, Recurrent neural networks		Codes provided in the class in Google Colab

Contribution to Placements

(Please state how your course will help the student to get placed in a good company)

It prepares students for data intensive jobs like Market Research Analyst, Business Analyst, Data Scientist, Operations Analyst, Financial Analyst, Big Data Analyst,