PROGRAM	Master of Business Administration
SEMESTER	3
COURSE TITLE	Big Data in IT and Systems
COURSE CODE	04MB0340
COURSE CREDITS	03
COURSE DURATION	42 Hrs (42 sessions of 60 minutes each)

COURSE OUTCOMES:

* Understand the concept and challenge of Big Data and why existing technology is inadequate to Analyze the Big Data;
* Integrate and Apply knowledge on Collect, Manage, Store, Query, and Analyze various form of Big Data;
* Gain hands-on experience on large-scale Analytics tools to solve some open Big Data problems
* Ability to integrate machine learning libraries and mathematical and statistical tools with modern technologies like hadoop and mapreduce.
* Understand the impact of Big Data for Business Decisions and Strategy.

COURSE CONTENTS:

Module No	Unit / Sub Unit	Sessions
I	**Introduction to Big Data Analytics and Lifecycle** Big Data Overview, Data Structures , Perspective on Data Repositories, Practices in Analytics , BI versus Data Science , Analytical Architecture , Drivers of Big Data , Big Data Ecosystem. Data Analytics Lifecycle : Phase 1 : Discovery , Phase 2 : Data Preparation , Phase 3 : Model Planning , Phase 4 : Model Building , Phase 5 : Results , Phase 6 : Operationalize	8
II	**Advanced Analytical Theory and Methods -I :** **Clustering** Overview of Clustering , K-means : Use Cases ; Overview of the Method ; Determining the Number of Clusters ; Diagnostics; Reasons to Choose and Cautions. **Association Rules** Overview of Association Rules, Apriori Algorithm, Evaluation of Candidate Rules, Applications of Association Rules, Validation and Testing, Diagnostics.	8
III	**Advanced Analytical Theory and Methods -II :** **Classification** Decision Trees: Overview of Decision Tree, The General Algorithm, Decision Tree Algorithms, Evaluating a Decision Tree. Naïve Bayes: Bayes Theorem, Naïve Bayes Classifier, Smoothing, Diagnostics. Diagnostics of Classifiers. **Text Analysis** Text Analysis Steps, Collecting Raw Text, Representing Text, Term Frequency-Inverse Document Frequency (TFIDF) , Categorizing Documents by Topics , Determining Sentiments , Gaining Insights .	8

IV	**Advanced Analytics :** **Technology and Tools - MapReduce and Hadoop** Analytics for Unstructured Data : Use Cases , MapReduce , Apache Hadoop The Hadoop Ecosystem : Pig , Hive , HBase , Mahout , NoSQL **Technology and Tools – In-Database Analytics** SQL Essentials : Joins , Set Operations , Grouping Extensions Advanced SQL : Window Functions , User-Defined Functions and Aggregates , Ordered Aggregates , MADlib	10
V	**Application of Data Mining, predictive and prescriptive analytics using big data:** Data Mining, predictive and prescriptive analytics using big data & Decision Making Predictive Analysis, Forecasting, Optimization, Simulation Gamification, Business Metrics in Action Data science in Startups Basics of Problem-Solving Design Patterns in Statistical Computing Excel for Data Science.	8

Evaluation:
The students will be evaluated on a continuous basis and broadly follow the scheme given below:

		Weight age
A	Assignment & Presentation	20% (C.E.C.)
B	Internal Assessment	30% (I.A.)
C	End-Semester Examination	50% (External Assessment)

SUGGESTED READINGS:
Text Books:

Sr. No	Name of the Book	Author /s	Publisher	Edition & Year
T-01	Big Data & Analytics	Chellappan and Acharya	Wiley	2nd edition,2019
T-03	Big Data Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses	Michael Minelli, Michele Chambers, Ambiga Dhiraj	John Wiley & Sons, Inc.	1st edition,2012

Reference Books:

Sr. No	Name of the Book	Author/s	Publisher	Edition & Year
R-01	Business Analytics: Data Analysis and Decision Making	S. Christian Albright	Cengage Learning	5th edition,2012
R-02	Analytics in a Big Data World	Bart Baesens	John Wiley & Sons	3rd edition,2014
R-03	Big Data Analytics with R & Hadoop	Vignesh Prajapati	PACKT Publishing	1st edition, 2019