AMITY UNIVERSITY
KOLKATA

Course Title: Big Data Analytics
Credit Units: 04

Course Level: UG

Course Code:IT424

L	T	P/S	SW/FW	No of PSDA	TOTAL CREDIT UNITS
3	0	2	-	-	4

Course Objectives:

- Approach business problems data-analytically by identifying opportunities to derive business value from data.
- Know the basics of data mining techniques and how they can be applied to extract relevant business intelligence

Pre-requisites: Understanding of basic **c**oncepts of Database Management System and Algorithms and Data Structures

Course Contents/Syllabus:

	Weightage (%)
Module I Overview of Data Mining	**15%**
Classification Techniques, K-means Clustering, Association rules, Decision Trees, Linear and Logistic Regression	
Module II Introduction to Big Data	**15%**
Introduction – distributed file system – Big Data and its importance, Four Vs, Drivers for Big data, Big data analytics, Big data applications.	
Module III Introudction to Hadoop	**20%**
Big Data – Apache Hadoop & Hadoop EcoSystem – Moving Data in and out of Hadoop – Understanding inputs and outputs of MapReduce - Data Serialization. Algorithms using map reduce, Matrix-Vector Multiplication by Map Reduce.	
Module IV Hadoop Architecture	**20%**
Hadoop Architecture, Hadoop Storage: HDFS, Common Hadoop Shell commands , Anatomy of File Write and Read., NameNode, Secondary NameNode, and DataNode, Hadoop MapReduce paradigm, Map and Reduce tasks, Job, Task trackers - Cluster Setup – SSH & Hadoop Configuration – HDFS Administering –Monitoring & Maintenance.	
Module V Big Data and Applications	**20%**

Applications and case studies- Banking and Securities, Communications, Media and Entertainment, Healthcare Providers, Education, Manufacturing and Natural Resources, Government, Insurance, Retail and Whole sale trade, Transportation and Energy and Utilities	
Module VI Case Studies	
Insufficient understanding and acceptance of big data, Complexity of managing data quality, Dangerous big data security holes, Tricky process of converting big data into valuable insights, Troubles of up scaling.	**10%**

Course Learning Outcomes:

1. Approach business problems data-analytically by identifying opportunities to derive business value from data.
2. Know the basics of data analytical techniques and how they can be applied to extract relevant business intelligence.
3. Examine the types of the data to be able to analyze and apply the analytics techniques on a variety of applications.
4. Discover interesting patterns from large amounts of data to analyze and extract patterns to solve

Pedagogy for Course Delivery:

The class will be taught using remote teaching methodology. Students' learning and assessment will be on the basis of four quadrants and flipped class method. E-content will be also provided to the students for better learning. The class will be taught using theory, practical and case-based method.

List of Experiments of Big Data Analytics using Hadoop and R

1. Write an R-program to solve roots of the quadratic equation
2. Write an R-Program to find factorial and palindrome of given number.
3. To define and install Hadoop.
4. To implement the following file management tasks in Hadoop System (HDFS): Adding files and directories, Retrieving files, Deleting files
5. To run a basic Word Count MapReduce program to understand MapReduce Paradigm: To count words in a given file.
6. To study and implement basic functions and commands in R Programming.
7. To build WordCloud, a text mining method using R for easy to understand and visualization than a table data
8. To implement Bloom Filters for filter on Stream Data in C++/java.
9. To implement clustering program using R programming.
10. Write a program to read and write on . CSV file in R.

Assessment/ Examination Scheme:

Theory L/T (%)	Lab/Practical/Studio (%)	End Term Examination
100%	**NA**	**100%**

Theory Assessment (L&T):

Continuous Assessment/Internal Assessment					End Term Examination
Components(Drop down)	**Attendance**	**Mid Term Exam**	**Home Assignment**	**Quiz**	**EE**
Weightage (%)	5	15	10	10	60

Lab/ Practical/ Studio Assessment:

Continuous Assessment/Internal Assessment					End Term Examination
Components(Dropdown)	Performance	Lab Record	Viva	Attendance	EE
Weightage(%)	15	10	10	**5**	**60**

Text & References:

- Noreen Burlingame, “The little book on Big Data”, New Street publishers, 2012.
- Norman Matloff, “The Art of R Programming: A Tour of Statistical Software Design”, No Starch Press 1 edition, 2011.