

**PEDAGOGICAL CARD  
FOR  
“R PROGRAMMING AND BIG DATA ANALYTICS”**

---

**A. OVERVIEW OF THE COURSE**

In the new data era, the power of data analytics is allowing businesses to provide more value added products and services. New discoveries and trends are being identified in fields such as banking, medical, manufacturing, sales and marketing, mastering the appropriate tools to derive knowledge from data has become key. Several tools and platforms have emerged to be able to mine the Big Data and provide meaningful insights. In this course, you will discover the power of R integrated in a Big Data environment. You will first be introduced with the basics of R and Big Data before embarking on the journey to R and Big Data analytics. Through the guided activities provided, any novice user can easily embark in R and Big Data.

**Why join this course?**

During this course you will learn how to use R programming and Hadoop Framework for Big Data Analytics. It consists of the following main concepts:

- R – a programming language used for statistical computing
- Big Data – huge volumes of different types of data from varying sources
- Big Data Platform (Hadoop) - a platform to be able to manipulate Big Data. Hadoop is an open- source software for reliable, scalable and distributed computing.
- R on Hadoop – R will be used to perform statistical computing on Hadoop

Different learning materials including articles, videos, exercises and assignments have been embedded in the module to allow you to meet the learning outcomes.

**Who is the course for?**

This course has been designed for those who are interested in R and big data analytics. It is useful for both undergraduate students, postgraduate students as well as PhD students in the field of Informatics and Statistics.

### **What software and tools do you need?**

The software which be used will be based on open-source platform. Useful links and instructions will be provided to help you in downloading and installing the necessary tools and platforms. A virtual machine needs to be installed and run on the local machine. You will need a local machine with memory space of at least 16GB for this module.

### **B. COURSE AUTHORS**

Dr (Mrs) Maleika Heenaye- Mamode khan

Dr (Mrs) Baby Gobin- Rahimbux

Mrs Soulakshmee Devi Nagowah

Mr Leckraj Nagowah

Mrs Bibi Zarine Cadarsaib

### **EDUCATIONAL TECHNOLOGIST**

Mrs Shamim Ajaheb-Bahadoor

### **C. COURSE AIMS**

The aim of this module is to allow the student to acquire the basic foundations of R and Big Data and to be able to apply R in a Big Data environment (Hadoop).

### **D. LEARNING OUTCOMES**

By the end of this course, you are expected to be able to:

1. Write R codes to manipulate data.
2. Explain concepts related to Big Data which includes the Big Data Ecosystem and technologies.
3. Integrate R in a Hadoop platform.
4. Manipulate data using R on a Big Data Platform.

### **E. PROPOSED OUTLINE OF THE COURSE**

This course is divided into 6 units, namely:

1. Setting up the environment for R
2. R fundamentals

3. Data Analysis and Visualisation
4. Big Data Essentials
5. Big Data Ecosystem
6. R and Big Data Analytics

***Unit 1: Setting up the environment for R***

- Topic 1: Introduction to R
- Topic 2: Installation of R Studio
- Topic 3: Console and Script Editor
- Topic 4: Installation of R Packages
- Topic 5: R Calculator
- Topic 6: R help

***Unit 2: R fundamentals***

- Topic 1: Use the console window and the script editor.
- Topic 2: Have an overview of the arithmetic, relational and logical operators.
- Topic 3: Work with variables.
- Topic 4: Examine the different data structures that exist in R.
- Topic 5: Familiarise yourself with the two main control structures: decisions and loops.
- Topic 6: Work with in-built and user-defined functions

***Unit 3: Data Analysis and Visualisation***

- Topic 1: Reading Data Frames
- Topic 2: Manipulating Data
- Topic 3: Exporting Data
- Topic 4: Descriptive Statistic Measures
- Topic 5: Data Visualization in R

***Unit 4: Big Data Essentials***

- Topic 1: Big Data Overview
- Topic 2: Characteristics of Big Data

Topic 3: Challenges

Topic 4: Application Domains

Topic 5: Big Data Tools

Topic 6: Setting up the environment for Big Data

**Unit 5: Big Data Ecosystem**

Topic 1: Big Data Ecosystem

Topic 2: Apache Hadoop Core Components

Topic 3: The HDFS architecture

Topic 4: Understanding the MapReduce architecture

Topic 5: The MapReduce Programming Model

Topic 6: Other Components of Hadoop

Topic 7: RHadoop and MapReduce Practicals

**Unit 6: R and Big Data Analytics**

Topic 1: Introduction to Big Data Analytics

Topic 2: Big Data Analytics Lifecycle

Topic 3: Big Data Analytics Problems

Topic 4: Big Data Analytics using Machine Learning Techniques

Topic 5: Setting up the environment for Big Data Analytics using SparkR

Topic 6: Applying supervised Machine Learning Techniques

Topic 7: Applying unsupervised Machine Learning Techniques

Unit	Author/s	Indicative Learning outcomes	Indicative Activities time frame & purpose - weightage for CA	Media/ Video	Timeframe
1	Z Cadessaib	<ul style="list-style-type: none"><li>Explain the R environment</li><li>Install and use R Studio</li></ul>	<ul style="list-style-type: none"><li>Self-assessment activities</li><li>Installing RStudio</li></ul>	Overview	1 week
2	L Nagowah	<ul style="list-style-type: none"><li>Use the console window and the script editor</li></ul>	<ul style="list-style-type: none"><li>Activities</li></ul>	Overview	3 weeks

		<ul style="list-style-type: none"> <li>• Have an overview of the arithmetic, relational and logical operators</li> <li>• Work with variables</li> <li>• Examine the different data structures that exist in R</li> <li>• Familiarise yourself with the two main control structures: decisions and loops</li> <li>• Work with in-built and user-defined functions</li> </ul>	<ul style="list-style-type: none"> <li>• Writing R scripts and programs</li> </ul>		
3	<b>B Gobin-Rahimbux</b>	<ul style="list-style-type: none"> <li>• Identify datasets and explain how they are organised.</li> <li>• Manipulate data in a dataframe.</li> <li>• Import and export data in RStudio.</li> <li>• Use R functions for data visualization in RStudio</li> </ul>	<ul style="list-style-type: none"> <li>• Activities.</li> <li>• Using RStudio for Visualisation</li> </ul>	Overview	2 weeks
<b>Assignment 1</b>					
4	<b>S Nagowah, L Nagowah and Z Cadarsaib</b>	<ul style="list-style-type: none"> <li>• Understand and explain Big Data concepts, challenges, application domains and tools</li> <li>• Set up a Big Data environment</li> </ul>	<ul style="list-style-type: none"> <li>• Activities.</li> <li>• Installing a Virtual Machine and a Big Data platform</li> </ul>	Overview	2 weeks
5	<b>M Mamode Khan and L Nagowah</b>	<ul style="list-style-type: none"> <li>• Describe and explain the Hadoop Ecosystem</li> <li>• Understand the concepts of HDFS and MapReduce</li> <li>• Describe other related tools in the Hadoop Ecosystem</li> <li>• Run MapReduce programs</li> </ul>	<ul style="list-style-type: none"> <li>• Activities.</li> <li>• MCQs.</li> <li>• Hands-on on RHadoop and MapReduce</li> </ul>	Overview	2 weeks
6	<b>M Mamode Khan and S Nagowah</b>	<ul style="list-style-type: none"> <li>• Understand techniques for Big Data Analytics</li> <li>• Understand phases of data analytics project life cycle</li> <li>• Insight on Big Data analytics problems</li> </ul>	<ul style="list-style-type: none"> <li>• Activities.</li> <li>• Hands-on on SparkR</li> </ul>	Overview	3 weeks

		<ul style="list-style-type: none"> <li>• Identify tools for Big Data analytics</li> <li>• Understand the importance of Machine Learning</li> <li>• Differentiate between supervised and unsupervised machine learning algorithms</li> <li>• Apply supervised and unsupervised algorithms using SparkR</li> </ul>			
<b>Assignment 2</b>					

## F. PROPOSED MODULE MAP

Unit	Hr(s)	Theme(s)	Topics Covered	Resource Persons
1	3	<ul style="list-style-type: none"> <li>• Introducing R</li> <li>• Setting up RStudio</li> </ul>	<ul style="list-style-type: none"> <li>• Introduction to R</li> <li>• Installation of R Studio</li> <li>• Console and Script Editor</li> <li>• Installation of R Packages</li> <li>• R Calculator</li> <li>• R help</li> </ul>	Z Cadessaib
2	1	<ul style="list-style-type: none"> <li>• Basics of R</li> </ul>	<ul style="list-style-type: none"> <li>• R Syntax</li> <li>• R Operators</li> <li>• Variables</li> </ul>	L Nagowah
	4	<ul style="list-style-type: none"> <li>• Data structures</li> </ul>	<ul style="list-style-type: none"> <li>• Vectors including Scalars</li> <li>• Matrices</li> <li>• Arrays</li> <li>• Data frames, and</li> <li>• Lists</li> </ul>	L Nagowah
	4	<ul style="list-style-type: none"> <li>• Control Structures</li> <li>• Functions</li> </ul>	<ul style="list-style-type: none"> <li>• Decision structures</li> <li>• Loops</li> <li>• User defined Functions</li> <li>• Functions and Arguments</li> <li>• Functions and Return Types</li> </ul>	L Nagowah
3	6	<ul style="list-style-type: none"> <li>• Data Analysis and Visualisation in RStudio</li> </ul>	<ul style="list-style-type: none"> <li>• Reading Data Frames</li> <li>• Manipulating Data</li> <li>• Exporting Data</li> <li>• Descriptive Statistic Measures</li> </ul>	B Gobin-Rahimbux

			<ul style="list-style-type: none"> <li>• Data Visualization in R</li> </ul>	
	3	<ul style="list-style-type: none"> <li>• Assignment 1</li> </ul>	<ul style="list-style-type: none"> <li>• R Studio</li> </ul>	L Nagowah B Gobin-Rahimbux
4	3	<ul style="list-style-type: none"> <li>• Big Data Concepts</li> </ul>	<ul style="list-style-type: none"> <li>• Big Data Overview</li> <li>• Characteristics of Big Data</li> <li>• Challenges</li> <li>• Application Domains</li> <li>• Big Data Tools</li> </ul>	S Nagowah
	3	<ul style="list-style-type: none"> <li>• Setting up a Big Data platform</li> </ul>	<ul style="list-style-type: none"> <li>• Virtual Machine</li> <li>• Big Data Platform</li> </ul>	S Nagowah
5	3	<ul style="list-style-type: none"> <li>• Big Data Ecosystem</li> <li>• Hadoop</li> </ul>	<ul style="list-style-type: none"> <li>• Big Data Ecosystem</li> <li>• Apache Hadoop Core Components</li> <li>• The HDFS architecture</li> <li>• Understanding the MapReduce architecture</li> <li>• The MapReduce Programming Model</li> <li>• Other Components of Hadoop</li> </ul>	M Mamode Khan
	3	<ul style="list-style-type: none"> <li>• RHadoop and MapReduce</li> </ul>	<ul style="list-style-type: none"> <li>• RHadoop and MapReduce Practicals</li> </ul>	M Mamode Khan
6	5	<ul style="list-style-type: none"> <li>• Big Data Analytics</li> </ul>	<ul style="list-style-type: none"> <li>• Introduction to Big Data Analytics</li> <li>• Big Data Analytics Lifecycle</li> <li>• Big Data Analytics Problems</li> <li>• Big Data Analytics using Machine Learning Techniques</li> </ul>	S Nagowah M Mamode Khan
	4	<ul style="list-style-type: none"> <li>• SparkR</li> </ul>	<ul style="list-style-type: none"> <li>• Setting up the environment for Big Data Analytics using SparkR</li> <li>• Applying supervised Machine Learning techniques using Spark</li> <li>• Applying unsupervised Machine Learning Techniques</li> </ul>	S Nagowah M Mamode Khan
	3	<ul style="list-style-type: none"> <li>• Assignment 2</li> </ul>	<ul style="list-style-type: none"> <li>• RHadoop/ SparkR</li> </ul>	S Nagowah M Mamode Khan

## G. ASSESSMENT

The course is fully assessed on coursework as specified below:

	<b>Activity</b>	<b>% Weightage</b>	<b>Submission Due</b>
1	Activities for Unit 2	10	Week 6
2	Activities for Unit 3	5	Week 7
3	Assignment 1	30	Week 8
4	Activities for Unit 4	5	Week 10
5	Activities for Unit 5	5	Week 12
6	Activities for Unit 6	5	Week 14
7	Assignment 2	40	Week 16

Activities: 30%, Assignment 1: 30 % and Assignment 2: 40%

#### **H. RECOMMENDED READINGS**

1. Golemund, G. and Wickham, H., 2017. R for Data Science. O'Reilly, January 2017 First Edition.
2. John Walker, S., 2014. Big data: A revolution that will transform how we live, work, and think.
3. Kabacoff, R.I., 2010. R in Action. Manning.
4. Prajapati, V., 2013. Big data analytics with R and Hadoop. Packt Publishing Ltd.
5. Ryza, S., Laserson, U., Owen, S. and Wills, J., 2017. Advanced analytics with spark: patterns for learning from data at scale. " O'Reilly Media, Inc.".
6. Zikopoulos, P. and Eaton, C., 2011. Understanding big data: Analytics for enterprise class hadoop and streaming data. McGraw-Hill Osborne Media.