

Academic Council: 26/07/2019

Item No: 4.76

**UNIVERSITY OF MUMBAI**



**Syllabus for M.Sc. Part I**  
**(Semester I and II)**

**Programme: M.Sc.**

**Subject: Information Technology**

(Choice Based Credit System with effect from  
the academic year 2019 – 2020)

<b>Semester – I</b>		
<b>Course Code</b>	<b>Course Title</b>	<b>Credits</b>
PSIT101	Research in Computing	4
PSIT102	Data Science	4
PSIT103	Cloud Computing	4
PSIT104	Soft Computing Techniques	4
PSIT1P1	Research in Computing Practical	2
PSIT1P2	Data Science Practical	2
PSIT1P3	Cloud Computing Practical	2
PSIT1P4	Soft Computing Techniques Practical	2
<b>Total Credits</b>		<b>24</b>

<b>Semester – II</b>		
<b>Course Code</b>	<b>Course Title</b>	<b>Credits</b>
PSIT201	Big Data Analytics	4
PSIT202	Modern Networking	4
PSIT203	Microservices Architecture	4
PSIT204	Image Processing	4
PSIT2P1	Big Data Analytics Practical	2
PSIT2P2	Modern Networking Practical	2
PSIT2P3	Microservices Architecture Practical	2
PSIT2P4	Image Processing Practical	2
<b>Total Credits</b>		<b>24</b>

## **Program Specific Outcomes**

PSO1: Ability to apply the knowledge of Information Technology with recent trends aligned with research and industry.

PSO2: Ability to apply IT in the field of Computational Research, Soft Computing, Big Data Analytics, Data Science, Image Processing, Artificial Intelligence, Networking and Cloud Computing.

PSO3: Ability to provide socially acceptable technical solutions in the domains of Information Security, Machine Learning, Internet of Things and Embedded System, Infrastructure Services as specializations.

PSO4: Ability to apply the knowledge of Intellectual Property Rights, Cyber Laws and Cyber Forensics and various standards in interest of National Security and Integrity along with IT Industry.

PSO5: Ability to write effective project reports, research publications and content development and to work in multidisciplinary environment in the context of changing technologies.

# **SEMESTER II**

<b>M. Sc (Information Technology)</b>		<b>Semester – II</b>	
<b>Course Name: Big Data Analytics</b>		<b>Course Code: PSIT201</b>	
<b>Periods per week</b> <b>1 Period is 60 minutes</b>	<b>Lectures</b>	<b>4</b>	
	<b>Credits</b>	<b>4</b>	
		<b>Hours</b>	<b>Marks</b>
<b>Evaluation System</b>	<b>Theory Examination</b>	<b>2½</b>	<b>60</b>
	<b>Theory Internal</b>	<b>--</b>	<b>40</b>

<b>Objectives</b>	<ul style="list-style-type: none"> <li>To provide an overview of an exciting growing field of big data analytics.</li> <li>To introduce the tools required to manage and analyze big data like Hadoop, NoSql MapReduce.</li> <li>To teach the fundamental techniques and principles in achieving big data analytics with scalability and streaming capability.</li> <li>To enable students to have skills that will help them to solve complex real-world problems in for decision support.</li> </ul>
-------------------	--

<b>Unit</b>	<b>Details</b>	<b>Lectures</b>
<b>I</b>	Introduction to Big Data, Characteristics of Data, and Big Data Evolution of Big Data, Definition of Big Data, Challenges with big data, Why Big data? Data Warehouse environment, Traditional Business Intelligence versus Big Data. State of Practice in Analytics, Key roles for New Big Data Ecosystems, Examples of big Data Analytics. Big Data Analytics, Introduction to big data analytics, Classification of Analytics, Challenges of Big Data, Importance of Big Data, Big Data Technologies, Data Science, Responsibilities, Soft state eventual consistency. Data Analytics Life Cycle	<b>12</b>
<b>II</b>	Analytical Theory and Methods: Clustering and Associated Algorithms, Association Rules, Apriori Algorithm, Candidate Rules, Applications of Association Rules, Validation and Testing, Diagnostics, Regression, Linear Regression, Logistic Regression, Additional Regression Models.	<b>12</b>
<b>III</b>	Analytical Theory and Methods: Classification, Decision Trees, Naïve Bayes, Diagnostics of Classifiers, Additional Classification Methods, Time Series Analysis, Box Jenkins methodology, ARIMA Model, Additional methods. Text Analysis, Steps, Text Analysis Example, Collecting Raw Text, Representing Text, Term Frequency-Inverse Document Frequency (TFIDF), Categorizing Documents by Topics, Determining Sentiments	<b>12</b>
<b>IV</b>	Data Product, Building Data Products at Scale with Hadoop, Data Science Pipeline and Hadoop Ecosystem, Operating System for Big Data, Concepts, Hadoop Architecture, Working with Distributed file system, Working with Distributed Computation, Framework for Python and Hadoop Streaming, Hadoop Streaming, MapReduce with Python,	<b>12</b>

	Advanced MapReduce. In-Memory Computing with Spark, Spark Basics, Interactive Spark with PySpark, Writing Spark Applications,	
V	Distributed Analysis and Patterns, Computing with Keys, Design Patterns, Last-Mile Analytics, Data Mining and Warehousing, Structured Data Queries with Hive, HBase, Data Ingestion, Importing Relational data with Sqoop, Injesting stream data with flume. Analytics with higher level APIs, Pig, Spark's higher level APIs.	12

Books and References:					
Sr. No.	Title	Author/s	Publisher	Edition	Year
1.	Big Data and Analytics	Subhashini Chellappan Seema Acharya	Wiley	First	
2.	Data Analytics with Hadoop <i>An Introduction for Data Scientists</i>	<i>Benjamin Bengfort and Jenny Kim</i>	O'Reilly		2016
3.	Big Data and Hadoop	V.K Jain	Khanna Publishing	First	2018

## Evaluation Scheme

### Internal Evaluation (40 Marks)

The internal assessment marks shall be awarded as follows:

1. 30 marks (Any one of the following):
  - a. Written Test or
  - b. SWAYAM (Advanced Course) of minimum 20 hours and certification exam completed or
  - c. NPTEL (Advanced Course) of minimum 20 hours and certification exam completed or
  - d. Valid International Certifications (Prometric, Pearson, Certiport, Coursera, Udemy and the like)
  - e. One certification marks shall be awarded one course only. For four courses, the students will have to complete four certifications.

2. 10 marks

The marks given out of 40 for publishing the research paper should be divided into four course and should awarded out of 10 in each of the four course.

- i. Suggested format of Question paper of 30 marks for the written test.

Q1.	Attempt <u>any two</u> of the following:	16
a.		
b.		
c.		
d.		
Q2.	Attempt <u>any two</u> of the following:	14
a.		

b.		
c.		
d.		

- ii. **10 marks from every course coming to a total of 40 marks, shall be awarded on publishing of research paper in UGC approved Journal with plagiarism less than 10%. The marks can be awarded as per the impact factor of the journal, quality of the paper, importance of the contents published, social value.**

## External Examination: (60 marks)

	<b>All questions are compulsory</b>	
<b>Q1</b>	<b>(Based on Unit 1) Attempt <u>any two</u> of the following:</b>	<b>12</b>
a.		
b.		
c.		
d.		
<b>Q2</b>	<b>(Based on Unit 2) Attempt <u>any two</u> of the following:</b>	<b>12</b>
<b>Q3</b>	<b>(Based on Unit 3) Attempt <u>any two</u> of the following:</b>	<b>12</b>
<b>Q4</b>	<b>(Based on Unit 4) Attempt <u>any two</u> of the following:</b>	<b>12</b>
<b>Q5</b>	<b>(Based on Unit 5) Attempt <u>any two</u> of the following:</b>	<b>12</b>

## Practical Evaluation (50 marks)

A Certified copy journal is essential to appear for the practical examination.

<b>1.</b>	<b>Practical Question 1</b>	<b>20</b>
<b>2.</b>	<b>Practical Question 2</b>	<b>20</b>
<b>3.</b>	<b>Journal</b>	<b>5</b>
<b>4.</b>	<b>Viva Voce</b>	<b>5</b>

OR

<b>1.</b>	<b>Practical Question</b>	<b>40</b>
<b>2.</b>	<b>Journal</b>	<b>5</b>
<b>3.</b>	<b>Viva Voce</b>	<b>5</b>