SCSA1603	BIG DATA ANALYTICS	L	T	P	Credits	Total Marks
		3	0	0	3	100

COURSE OBJECTIVES
- ➢ To become familiar with the fundamental concepts of Big Data.
- ➢ To be competent in identifying the challenges in handling large volumes of data.
- ➢ To propose scalable solutions.
- ➢ To understand the impact of Big Data in business intelligence, scientific discovery, and in day-to-day life.
- ➢ To learn the tools and techniques for handling large datasets.

UNIT 1   INTRODUCTION   9 Hrs.
Introduction to Big Data – Issues and Challenges in the traditional systems - Evolution of Big Data – Four V's of Big Data – Big Data Use Cases and characteristics – Intelligent Data Analysis – Data Analytic Tools – Big Data StorageStatistical Concepts: Sampling Distributions - Re-Sampling - Statistical Inference - Prediction Error – Random Sampling.

UNIT 2   BIG DATA TOOLS I   9 Hrs.
Big Data Applications using Pig and Hive – Fundamentals of HBase and ZooKeeper – IBM Infosphere Big Insights – Introduction to FLUME – KAFKA.

UNIT 3   BIG DATA TOOLS II   9 Hrs.
Introduction to NoSQL - MongoDB – Spark – Cassandra - Cassandra Data Model – Data Design – Cassandra Architecture – Read and Write Data – Clients – Integrate with Hadoop. Introduction - Importance of Effective Data Visualization - Introduction to Tableau - Choosing the Right Chart Type - Using the Color Effectively Reducing Clutter - Dashboard Creation and Formatting.

UNIT 4   HADOOP   9 Hrs.
Introduction to Hadoop – Hadoop Distributed File System – Analysing data with Hadoop – Scaling – Streaming – Clustering: Single Node and Multi Node – Working with Hadoop Commands – Working with Apache Oozie.

UNIT 5   MAP REDUCE   9 Hrs.
Algorithms using map reduce - Matrix-Vector – Multiplication – Word Count - Understanding inputs and outputs of MapReduce, Data Serialization – Introduction to YARN – MapReduce Vs YARN – YARN Architecture – Scheduling in YARN – Fair Scheduler – Capacity Scheduler.

Max. 45 Hrs.

COURSE OUTCOMES
On completion of the course, student will be able to
CO1 - Configure the tools required for setting up Big Data Ecosystem.
CO2 - Understand conceptually how Big Data is stored and organized.
CO3 - Use appropriate models of analysis, assess the quality of input, derive insight from results, and investigate potential issues.
CO4 - Interpret data findings effectively in visual formats.
CO5 - Explore the fundamentals of various big data applications.
CO6 - Implement the Algorithms for data analytics.

TEXT / REFERENCE BOOKS
1.   Joshua N. Milligan, "Learning Tableau", Packt Publishing, 2015.
2.   Chuck Lam, "Hadoop in Action", Manning Publications Co., 2018.
3.   Tom White, "Hadoop the Definitive Guide", Oreilly, 4th Edition, 2015.
4.   Eben Hewitt, "Cassandra: The Definitive Guide", O'Reilly, 2010.
5.   Anand Rajaraman and Jeffrey David Ullman, "Mining of Massive Datasets", Cambridge University Press, Edition I, ISBN-10: 1107015359 | ISBN-13: 978-1107015357, 2011.
6.   Jimmy Lin and Chris Dyer, "Data-Intensive Text Processing with MapReduce", Morgan and Claypool Publishers, 2010.
7.   Jonathan R. Owens, Brian Femiano, and Jon Lentz, "Hadoop Real World Solutions Cookbook", Packt Publishing, ISBN-10: 1849519129 | ISBN-13: 978-1849519120, 2013

<div align="center">

END SEMESTER EXAMINATION QUESTION PAPER PATTERN

</div>

Max. Marks : 100   **Exam Duration : 3 Hrs.**
PART A: 10 Questions of 2 marks each-No choice   **20 Marks**
PART B: 2 Questions from each unit with internal choice, each carrying 16 marks   **80 Marks**