NATIONAL INSTITUTE OF TECHNOLOGY, TIRUCHIRAPPALLI

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

COURSE PLAN – PART I			
Name of the programme and specialization	**B.TECH / CSE**		
Course Title	**Big Data Analytics**		
Course Code	**CSOE11**	**No. of Credits**	**3**
Course Code of Pre-requisite subject(s)		**Semester**	**V**
Session	**July 2021**	**Section (if, applicable)**	
Name of Faculty	**Dr. Chandramani Chaudhary**	**Department**	**CSE**
Official Email	**chandramani@nitt.edu**	**Telephone No.**	**-**
Name of Course Coordinator(s) (if, applicable)	**NIL**		
Official E-mail	**NIL**	**Telephone No.**	**NIL**
Course Type (please tick appropriately)	**Elective Course**		

Syllabus (approved in Senate)

UNIT I Introduction to Big Data Introduction: Big Data - Characteristics of Big Data - Big data management architecture - Examining Big Data Types - Big Data Technology Components - Big data analytics - Big data analytics examples - Web Data Overview - Web Data in Action.

UNIT II Hadoop Introduction: History of Hadoop - Hadoop Ecosystem - Analyzing data with Hadoop - Hadoop Distributed File System - Design - HDFS concepts - Hadoop filesystem - Data flow - Hadoop I / O - Data integrity - Serialization - Setting up a Hadoop cluster - Cluster specification - cluster setup and installation - YARN.

UNIT III MapReduce Introduction: Understanding MapReduce functions - Scaling out - Anatomy of a MapReduce Job Run - Failures - Shuffle and sort - MapReduce types and formats - features - counters - sorting - MapReduce Applications –Configuring and setting the environment - Unit test with MR unit - local test.

UNIT IV Spark Installing spark - Spark applications - Jobs - Stages and Tasks - Resilient Distributed databases - Anatomy of a Spark Job Run - Spark on YARN - SCALA: Introduction - Classes and objects - Basic types and operators - built-in control structures - functions and closures - inheritance.

UNIT V NoSQL Databases Introduction to NoSQL - MongoDB: Introduction - Data types - Creating - Updating and deleing documents - Querying - Introduction to indexing - Capped collections - Hbase: Concepts - Hbase Vs RDBMS - Creating records - Accessing data - Updating

and deleting data - Modifying data - exporting and importing data. USE CASES: Call detail log analysis - Credit fraud alert - Weather forecast.

TEXT BOOKS

1. EMC Education Services, "Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data", Wiley Publishers, 2015.
2. Simon Walkowiak, "Big Data Analytics with R", PackT Publishers, 2016.
3. David Loshin, "Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools, Techniques, No SQL, and Graph", Morgan Kaufmann/Elsevier Publishers, 2013.
4. Bart Baesens, "Analytics in a Big Data World: The Essential Guide to Data Science and its Applications", Wiley Publishers, 2015.
5. Kim H. Pries, Robert Dunnigan, "Big Data Analytics: A Practical Guide for Managers", CRC Press, 2015.

COURSE OBJECTIVES

- To understand the Big Data Platform and its Use cases
- To Provide an overview of Apache Hadoop
- To Provide HDFS Concepts and Interfacing with HDFS
- To understand NoSQL database

MAPPING OF COs with POs

Course Outcomes	Programme Outcomes (PO) (Enter Numbers only)
1. Understand the concepts of Scala programming	**1,2,3,4**
2. Apply Mapreduce programming model to process big data	**3**
3. Analyze Spark and its uses for big data processing	**4**
4. Design programs for big data applications using Hadoop components	**2,3,4,5**

COURSE PLAN – PART II

COURSE OVERVIEW

This course covers big data analysis techniques and tools, focusing on ways to handle large-scale data efficiently using various algorithms. It also discribe methods for how big data can be visualized.

COURSE TEACHING AND LEARNING ACTIVITIES (Add more rows)

S.No.	Week/Contact Hours	Topic	Mode of Delivery
1	02/08/2021 to 06/08/2021 3 hours	**Unit I** Introduction: Big Data - Characteristics of Big Data - Big data management architecture - Examining Big Data Types - Big Data Technology Components - Big data analytics - Big data analytics examples - Web Data Overview - Web Data in Action.	Online PPT Presentation
2	09/08/2021 to 13/08/2021 3 hours	**Unit II**- Introduction: History of Hadoop - Hadoop Ecosystem - Analyzing data with Hadoop - Hadoop Distributed File System - Design - HDFS concepts - Hadoop filesystem	Online PPT Presentation
3	16/08/2021 to 20/08/2021 3 hours	Data flow - Hadoop I / O - Data integrity - Serialization	Online PPT Presentation
4	23/08/2021 to 27/08/2021 3 hours	Setting up a Hadoop cluster - Cluster specification - cluster setup and installation - YARN	Online PPT Presentation
5	30/08/2021 to 03/09/2021 3 hours	**UNIT III MapReduce** Introduction: Understanding MapReduce functions - Scaling out - Anatomy of a MapReduce Job Run	Online PPT Presentation
6	06/09/2021 to 09/09/2021 1 hour	**Cycle Test-1**	
7	13/10/2021 to 17/10/2021 3 hours	Failures - Shuffle and sort - MapReduce types and formats - features - counters - sorting	Online PPT Presentation
8	20/10/2021 to 24/10/2021 3 hours	MapReduce Applications –Configuring and setting the environment - Unit test with MR unit - local test.	Online PPT Presentation
9	27/10/2021 to 28/10/2021 1hour	**UNIT IV Spark** Installing spark - Spark applications	Online PPT Presentation
10	6/10/2021 to 8/10/2021 1 hour	**Cycle Test- 2**	

11	11/10/2021 to 14/10/2021 3 hours	Jobs - Stages and Tasks - Resilient Distributed databases	Online PPT Presentation
12	18/10/2021 to 22/10/2021 3 hours	Anatomy of a Spark Job Run - Spark on YARN - SCALA: Introduction - Classes and objects	Online PPT Presentation
13	25/10/2021 to 29/10/2021 3 hours	Basic types and operators - built-in control structures - functions and closures - inheritance.	Online PPT Presentation
14	1/11/2021 to 5/11/2021 3 hours	**UNIT V NoSQL Databases** Introduction to NoSQL - MongoDB: Introduction - Data types - Creating - Updating and deleing documents	Online PPT Presentation
15	8/11/2021 to 11/11/2021 2 hours	Querying - Introduction to indexing - Capped collections	Online PPT Presentation
16	12/11/2021 to 16/11/2021 1 hour	**Cycle Test-3**	
17	17/11/2021 to 18/11/2021 1 hour	Hbase: Concepts - Hbase Vs RDBMS - Creating records - Accessing data	Online PPT Presentation
18	22/11/2021 to 25/11/2021 2 hours	Updating and deleting data - Modifying data - exporting and importing data USE CASES: Call detail log analysis - Credit fraud alert - Weather forecast.	Online PPT Presentation

COURSE ASSESSMENT METHODS (shall range from 4 to 6**)**

S.No.	Mode of Assessment	Week/Date	Duration	% Weightage
1	Cycle Test 1	06/09/2021 to 08/09/2021	1 hour	15
2	Cycle Test 2	06/10/2021 to 08/10/2021	1 hour	15
3	Assignment 1	12/11/2021 to 16/11/2021	1 hours	15
4	Assignment 2	22/10/2021 to 25/10/2021	1.5 hours	15
5	Assignment 3	22/11/2021 to 25/11/2021	1.5 hours	10

CPA	Compensation Assessment*	As per academic schedule	1 hour	15
6	Final Assessment *	As per academic schedule	2 hours	30

***mandatory; refer to guidelines on page 4**

COURSE EXIT SURVEY (mention the ways in which the feedback about the course shall be assessed)

1. **Students' feedback through PAC meetings**
2. **Feedbacks are collected before final examination through MIS or any other standard format followed by the institute**
3. **Students, through their Class Representatives, may give their feedback at any time to the course faculty which will be duly addressed.**

COURSE POLICY (including compensation assessment to be specified**)**

MODE OF CORRESPONDENCE (email/ phone etc)

Email and Phone

COMPENSATION ASSESSMENT POLICY

1. **One compensation assessment will be given after completion of Cycle Test 1 and 2 for the students those who are absent for any assessment due to genuine reason.**
2. **Compensatory assessments would cover the syllabus of Cycle tests 1 & 2**
3. **The prior permission and required documents must be submitted for absence signed by HoD/CSE.**

ATTENDANCE POLICY (A uniform attendance policy as specified below shall be followed)

- At least 75% attendance in each course is mandatory.
- A maximum of 10% shall be allowed under On Duty (OD) category.
- Students with less than 65% of attendance shall be prevented from writing the final assessment and shall be awarded 'V' grade.

ACADEMIC DISHONESTY & PLAGIARISM

- Possessing a mobile phone, carrying bits of paper, talking to other students, copying from others during an assessment will be treated as punishable dishonesty.
- Zero mark to be awarded for the offenders. For copying from another student, both students get the same penalty of zero mark.
- The departmental disciplinary committee including the course faculty member, PAC chairperson and the HoD, as members shall verify the facts of the malpractice and award the punishment if the student is found guilty. The report shall be submitted to the Academic office.

- The above policy against academic dishonesty shall be applicable for all the programmes.

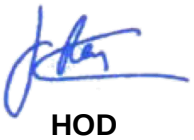
ADDITIONAL INFORMATION , IF ANY

1. **The Course Coordinator is available for consultation during the time intimated to the students**
2. **Relative grading adhering to the instructions from the office of the dean (Academic) will be adopted for the course.**

For Approval

Course Faculty

CC-Chairperson



HOD

Guidelines

a) The number of assessments for any theory course shall range from 4 to 6.

b) Every theory course shall have a final assessment on the entire syllabus with at least 30% weightage.

c) One compensation assessment for absentees in assessments (other than final assessment) is mandatory. Only genuine cases of absence shall be considered.

d) The passing minimum shall be as per the regulations.

B.Tech. Admitted in				P.G.
2018	2017	2016	2015	
35% or (Class average/2) whichever is greater.		(Peak/3) or (Class Average/2) whichever is lower		40%

e) Attendance policy and the policy on academic dishonesty & plagiarism by students are uniform for all the courses.

f) Absolute grading policy shall be incorporated if the number of students per course is less than 10.

g) Necessary care shall be taken to ensure that the course plan is reasonable and is objective.