



# NATIONAL INSTITUTE OF TECHNOLOGY, TIRUCHIRAPPALLI

## DEPARTMENT OF CSE

COURSE PLAN – PART I			
<b>Name of the programme and specialization</b>	B.TECH All Branches		
<b>Course Title</b>	Big Data Analytics		
<b>Course Code</b>	CSOE17	<b>No. of Credits</b>	3
<b>Course Code of Pre-requisite subject(s)</b>	-	<b>Semester</b>	VII
<b>Session</b>	July 2019	<b>Section (if, applicable)</b>	-
<b>Name of Faculty</b>	Dr. S.SARASWATHI	<b>Department</b>	CSE
<b>Official Email</b>	saras@nitt.edu	<b>Telephone No.</b>	
<b>Name of Course Coordinator(s) (if, applicable)</b>	---		
<b>Official E-mail</b>		<b>Telephone No.</b>	
<b>Course Type (please tick appropriately)</b>	<b>Elective course</b>		
<b>Syllabus (approved in BoS)</b>			
<p>UNIT – I Introduction to Big Data: Analytics – Nuances of big data – Value – Issues – Case for Big data – Big data options Team challenge – Big data sources – Acquisition – Features of Big Data - Security, Compliance, auditing and protection - Evolution of Big data – Best Practices for Big data Analytics - Big data characteristics - Volume, Veracity, Velocity, Variety – Big Data Architecture – Big Data and Cloud.*</p> <p>UNIT – II Data Analysis: Evolution of analytic scalability – Convergence – parallel processing systems – Cloud computing – grid computing – Map reduce Basics – Map Reduce Algorithm Design - enterprise analytic sand box – analytic data sets – Analytic methods – analytic tools – Cognos – Microstrategy – Pentaho - Regression Modeling - Multivariate Analysis - Bayesian Modeling - Inference and Bayesian Networks - Support Vector and Kernel Methods.*</p> <p>UNIT – III Stream Computing : Introduction to Streams Concepts – Stream data model and architecture - Stream Computing, Sampling data in a stream – Filtering streams – Counting distinct elements in a stream – Estimating moments – Counting oneness in a window – Decaying window - Realtime Analytics Platform(RTAP) applications IBM Infosphere – Big data at rest – Infosphere streams – Data stage – Statistical analysis – Intelligent scheduler – Infosphere Streams.*</p> <p>UNIT – IV Predictive Analytics and Visualization: Predictive Analytics – Supervised – Unsupervised learning – Neural networks – Kohonen models – Normal – Deviations from normal patterns – Normal behaviours – Expert options – Variable entry - Mining Frequent</p>			



## NATIONAL INSTITUTE OF TECHNOLOGY, TIRUCHIRAPPALLI

itemsets - Market based model – Apriori Algorithm – Handling large data sets in Main memory – Limited Pass algorithm – Counting frequent itemsets in a stream – Clustering Techniques – Hierarchical – K- Means – Clustering high dimensional data Visualizations - Visual data analysis techniques, interaction techniques; Systems and applications.\*

UNIT – V Frameworks and Applications: IBM for Big Data – Map Reduce Framework - Hadoop – Hive – Sharding - MongoDB – NoSQL Databases - S3 - Hadoop Distributed file systems – Hbase – Impala – Analyzing big data with Twitter and Facebook – Big data for E-commerce – Big data for blogs.\*

\*Programming assignments are mandatory.

### COURSE OBJECTIVES

- To understand the financial value of big data analytics
- To explore tools and practices for working with big data
- To understand how big data analytics can leverage into a key component

### MAPPING OF COs with POs

Course Outcomes	Programme Outcomes (PO) (Enter Numbers only)
1. Ability to apply the concepts of big data analytics for a domain	PO1, PO2,PO7,PO8
2. Ability to design and develop Hadoop and Map Reduce Framework	PO3,PO4,PO5,PO6,
3. Ability to contextually integrate and correlate large amounts of information	PO1,PO2,PO3,PO4, PO5,PO6

### COURSE PLAN – PART II

#### COURSE OVERVIEW

Big data Analytics deals with large amount of data storage, process, analytics and visualization. Nuances of big data and big data characteristics are the major elements of the big data technology. Hadoop processing tools with map reduce programming are included and various data analytic methods such as regression, bayesian, Support vector machines are also included. stream computing and its architectures are highlighted. Various predictive analytics and visualization techniques are discussed. Analyzing big data with twitter and facebook with Framework and applications.

#### COURSE TEACHING AND LEARNING ACTIVITIES

( Add more rows)

S.No.	Week/Contact Hours	Topic	Mode of Delivery
1	Week I	Unit1: Introduction to Big Data Analytics , Nuances of big data , Value, Issues, Case for Big data, Big data options, Team challenge	ppt
2	Week II	Big data sources , Acquisition, Features of Big Data, Security, Compliance, auditing and protection , Evolution of Big data – Best Practices for Big data Analytics	ppt



## NATIONAL INSTITUTE OF TECHNOLOGY, TIRUCHIRAPPALLI

3	Week III	Big data characteristics , Volume, Veracity, Velocity, Variety, Big Data Architecture, Big Data and Cloud.	ppt
4	Week IV	UNIT – II Data Analysis, Evolution of analytic scalability, Convergence, parallel processing systems, Cloud computing, grid computing	ppt
5	Week V	Map reduce Basics, Map Reduce Algorithm Design, enterprise analytic sand box, analytic data sets, Analytic methods, analytic tools, Cognos , Microstrategy, Pentaho	ppt
6	Week VI	Regression Modeling, Multivariate Analysis, Bayesian Modeling, Inference and Bayesian Networks, Support Vector and Kernel Methods	ppt
7	Week VII	UNIT – III Stream Computing Introduction to Streams Concepts, Stream data model and architecture, Stream Computing, Sampling data in a stream, Filtering streams, Counting distinct elements in a stream, Estimating moments, Counting oneness in a window	ppt
8	Week VIII	Decaying window, Realtime Analytics Platform(RTAP) applications IBM Infosphere , Big data at rest , Infosphere streams, Data stage, Statistical analysis, Intelligent scheduler, Infosphere Streams.*	ppt
9	Week IX	UNIT – IV Predictive Analytics and Visualization, Predictive Analytics, Supervised, Unsupervised learning, Neural networks, Kohonen models, Normal, Deviations from normal patterns, Normal behaviours, Expert options, Variable entry, Mining Frequent itemsets, Market based model , Apriori Algorithm	ppt
10	Week X	Handling large data sets in Main memory, Limited Pass algorithm, Counting frequent itemsets in a stream, Clustering Techniques, Hierarchical – K- Means – Clustering high dimensional data Visualizations , Visual data analysis techniques, interaction techniques, Systems and applications	ppt



# NATIONAL INSTITUTE OF TECHNOLOGY, TIRUCHIRAPPALLI

11	Week XI	UNIT – V Frameworks and Applications, IBM for Big Data, Map Reduce Framework, Hadoop, Hive, Sharding, MongoDB, NoSQL Databases, S3	ppt
12	Week XII	Hadoop Distributed file systems, Hbase , Impala, Analyzing big data with Twitter and Facebook, Big data for E-commerce, Big data for blogs	ppt

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

S.No.	Mode of Assessment	Week/Date	Duration	% Weightage
1	Cycle Test I	September Second Week	1 hour	20
2	Cycle Test II	October Third week	1 hour	20
3	Programming Assignment	Periodically	Bi-monthly	10
CPA	Compensation Assessment*	After completion of Cycle Test 2	1 hour	20
4	Final Assessment *	November Third week	3 hours	50

\*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 class committee 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)

Mobile phone, Email, in-person – after 4.00 pm.

#### ASSESSMENT POLICY

1. One compensation assessment will be given after completion of Cycle Test 1 and 2 for the students those who **COMPENSATION** 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 document must be submitted for absence.

#### 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.



## NATIONAL INSTITUTE OF TECHNOLOGY, TIRUCHIRAPPALLI


- Students with less than 65% of attendance shall be prevented from writing the final assessment and shall be awarded "V" grade.

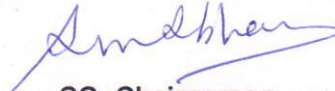
### ACADEMIC DISHONESTY & PLAGIARISM

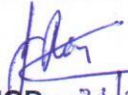
- Possessing a mobile phone, carrying bits of paper, talking to other students, copying from others during an assessment will be treated as punishable dishonesty.
- Zero mark to be awarded for the offenders. For copying from another student, both students get the same penalty of zero mark.
- The departmental 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

### FOR APPROVAL

  
Course Faculty  
Dr. S. Saraswathi

  
CC- Chairperson  
Dr. S. Mary Saira Bhanu

  
HOD 31/9/15  
Dr. Rajeswari Sridhar



## NATIONAL INSTITUTE OF TECHNOLOGY, TIRUCHIRAPPALLI

### 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.