



DEPARTMENT OF COMPUTER APPLICATIONS

COURSE PLAN – PART I			
Name of the programme and specialization	M.Sc. Computer Science		
Course Title	BIG DATA ANALYTICS		
Course Code	CAS7A1	No. of Credits	3
Course Code of Pre-requisite subject(s)			
Session	January 2020	Section (if, applicable)	
Name of Faculty	G.R.Gangadharan	Department	Computer Applications
Official Email	ganga@nitt.edu	Telephone No.	0431-2503737
Name of Course Coordinator(s) (if, applicable)	Dr. Michael Arock		
Official E-mail	michael@nitt.edu	Telephone No.	0431-2503736
Course Type (please tick appropriately)	Elective course		
<b>Syllabus (approved in BoS)</b>			
<p>INTRODUCTION TO BIG DATA : Introduction to BigData Platform – Challenges of Conventional Systems - Intelligent data analysis – Nature of Data - Analytic Processes and Tools - Analysis vs Reporting - Modern Data Analytic Tools - Statistical Concepts: Sampling Distributions - Statistical Inference - Prediction Error.</p> <p>Big Data Analytics: Classification, Clustering, and Mining, Information Extraction, Regression and Feature Selection, Reasoning: Logic and its Limits, Dealing with Uncertainty, Bayesian Inference, Forecasting, Neural Models, Introduction to Deep Learning.</p> <p>HADOOP : History of Hadoop- The Hadoop Distributed File System – Components of Hadoop-Analyzing the Data with Hadoop- Scaling Out- Hadoop Streaming- Design of HDFS-Java interfaces to HDFS Basics- Developing a Map Reduce Application-How Map Reduce Works-Anatomy of a Map Reduce Job run- Failures-Job Scheduling-Shuffle and Sort – Task execution - Map Reduce Types and Formats- Map Reduce Features</p> <p>NoSQL Databases : Evolution of Document DataBases – Design and use of NoSQL Databases – Storage and Retrieval of Unstructured Data – NoSQL Applications and query options. Types of NoSQL Databases, Graph Databases – Neo4j ; Case Studies - Real Time Sentiment Analysis, Stock Market Predictions.</p> <p>FRAMEWORKS: Applications on Big Data Using Pig and Hive – Data processing operators in Pig – Hive services – HiveQL – Querying Data in Hive - fundamentals of HBase and ZooKeeper - Visualizations -</p>			





## NATIONAL INSTITUTE OF TECHNOLOGY, TIRUCHIRAPPALLI

Visual data analysis techniques, interaction techniques; Systems and applications; Introduction to Tableau.

### REFERENCES:

1. Michael Berthold, David J. Hand, Intelligent Data Analysis, Springer, 2007.
2. Tom White — Hadoop: The Definitive Guide, Third Edition, O'reilly Media, 2012.
3. Chris Eaton, Dirk DeRoos, Tom Deutsch, George Lapis, Paul Zikopoulos, Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, McGraw Hill Publishing, 2012
4. Anand Rajaraman and Jeffrey David Ullman, —Mining of Massive Datasets, Cambridge University Press, 2012.
5. Bill Franks, Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics, JohnWiley & sons, 2012
6. Glenn J. Myatt, Making Sense of Data, John Wiley & Sons, 2007
7. PeteWarden, Big Data Glossary, O'Reilly, 2011.
8. Jiawei Han, Micheline Kamber, Data Mining Concepts and Techniques, Second Edition, Elsevier, Reprinted 2008.
9. Da Ruan, Guoqing Chen, Etienne E. Kerre, GeertWets, Intelligent Data Mining, Springer, 2007
10. Paul Zikopoulos, Dirk deRoos , Krishnan Parasuraman , Thomas Deutsch , James Giles , David Corrigan , Harness the Power of Big Data The IBM Big Data Platform, Tata McGraw Hill, 2012.

### COURSE OBJECTIVES

- To get introduced to big data analytics and to understand the importance of big data.
- To get introduced with different approaches of exploiting big data sources such as social media, mobile devices and sensors through understanding methodologies of analyzing big data.
- To acquire knowledge of handling unstructured and semi-structured data using NoSQL database.

### MAPPING OF COs with POs

Course Outcomes	Programme Outcomes (PO) (Enter Numbers only)
1. Comprehend the concepts of big data analytics.	I,II,III,IV,V
2. Build web-intelligence applications exploiting big data using new big data platforms based on the 'map-reduce' parallel programming framework.	I,II,III,IV,V

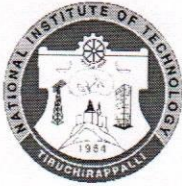




3. Effectively use NoSQL database for storage and retrieval of big data	I,II,III,IV,V
---	---------------

COURSE PLAN – PART II			
COURSE OVERVIEW			
COURSE TEACHING AND LEARNING ACTIVITIES			( Add more rows)
S.No.	Week/Contact Hours	Topic	Mode of Delivery
1	Week 1 (3 Classes)	Introduction to BigData Platform – Challenges of Conventional Systems - Intelligent data analysis – Nature of Data - Analytic Processes and Tools	Chalk and Talk, PPT
2	Week 2 (3 Classes)	Analysis vs Reporting - Modern Data Analytic Tools - Statistical Concepts: Sampling Distributions - Statistical Inference - Prediction Error	Chalk and Talk, PPT
3	Week 3 (3 Classes)	Classification, Clustering, and Mining, Information Extraction, Regression and Feature Selection	Chalk and Talk, PPT
4	Week 4 (3 Classes)	Reasoning: Logic and its Limits, Dealing with Uncertainty, Bayesian Inference, Forecasting, Neural Models, Introduction to Deep Learning	Chalk and Talk, PPT
5	Week 5 (3 Classes)	History of Hadoop- The Hadoop Distributed File System – Components of Hadoop - Analyzing the Data with Hadoop- Scaling Out- Hadoop Streaming-	Chalk and Talk, PPT
6	Week 6 (3 Classes)	Design of HDFS-Java interfaces to HDFS Basics	Chalk and Talk, PPT
7	Week 7 (3 Classes)	Developing a Map Reduce Application-How Map Reduce Works- Anatomy of a Map Reduce Job run- Failures	Chalk and Talk, PPT
8	Week 8 (3 Classes)	Job Scheduling-Shuffle and Sort – Task execution - Map Reduce Types and Formats- Map Reduce Features	Chalk and Talk, PPT
9	Week 9 (3 Classes)	Evolution of Document DataBases – Design and use of NoSQL Databases – Storage and Retrieval of Unstructured Data – NoSQL Applications and query options	Chalk and Talk, PPT





## NATIONAL INSTITUTE OF TECHNOLOGY, TIRUCHIRAPPALLI

10	Week 10 (3 Classes)	Types of NoSQL Databases, Graph Databases – Neo4j	Chalk and Talk, PPT
11	Week 11 (3 Classes)	Case Studies - Real Time Sentiment Analysis, Stock Market Predictions	Chalk and Talk, PPT
12	Week 12 (3 Classes)	Applications on Big Data Using Pig and Hive – Data processing operators in Pig – Hive services – HiveQL – Querying Data in Hive	Chalk and Talk, PPT
13	Week 13 (3 Classes)	Fundamentals of HBase and ZooKeeper	Chalk and Talk, PPT
14	Week 14 (3 Classes)	Visualizations - Visual data analysis techniques, interaction techniques- Systems and applications - Introduction to Tableau	Chalk and Talk, PPT

### **COURSE ASSESSMENT METHODS**

S.No.	Mode of Assessment	Week/Date	Duration	% Weightage
1	Cycle Test 1	6 <sup>th</sup> Week	60 Minutes	20
2	Cycle Test 2	10 <sup>th</sup> Week	60 Minutes	20
3	Assignment	9 <sup>th</sup> week	-	10
CPA	Compensation Assessment*			
4	Final Assessment *		180 Minutes	50

\*mandatory

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

- The students through the class representative may give their feedback at any time to the course coordinator which will be duly addressed.
- The students may give their feedback during class committee meetings.

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

### **MODE OF CORRESPONDENCE**

By Email: [ganga@nitt.edu](mailto:ganga@nitt.edu)





**COMPENSATION ASSESSMENT POLICY**

Compensation assessment will be conducted for absentees in cycle test I or cycle test II only after the submission of medical or On-Duty certificates signed by competent authority.

**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**

**FOR APPROVAL**

Course Faculty

CC- Chairperson

HOD





## NATIONAL INSTITUTE OF TECHNOLOGY, TIRUCHIRAPPALLI

### Guidelines

- The number of assessments for any theory course shall range from 4 to 6.
- Every theory course shall have a final assessment on the entire syllabus with at least 30% weightage.
- One compensation assessment for absentees in assessments (other than final assessment) is mandatory. Only genuine cases of absence shall be considered.
- 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%

- Attendance policy and the policy on academic dishonesty & plagiarism by students are uniform for all the courses.
- Absolute grading policy shall be incorporated if the number of students per course is less than 10.
- Necessary care shall be taken to ensure that the course plan is reasonable and is objective.