NATIONAL INSTITUTE OF TECHNOLOGY, TIRUCHIRAPPALLI

DEPARTMENT OF COMPUTER APPLICATIONS

COURSE PLAN – PART I				
Name of the programme and specialization	M.Tech. Data Analytics			
Course Title	Big Data Analytics			
Course Code	CA603		**No. of Credits**	3
Course Code of Pre-requisite subject(s)				
Session	July 2019	**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. P.J.A. Alphonse			
Official E-mail	alphonse@nitt.edu	**Telephone No.**		0431-2503742
Course Type (please tick appropriately)	Core course			

Syllabus (approved in BoS)

Introduction to big data : Introduction to Big Data Platform – Challenges of Conventional Systems - Intelligent data analysis – Nature of Data - Analytic Processes and Tools - Analysis vs Reporting.

Mining data streams : 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 - Real time Analytics Platform (RTAP) Applications – Case Studies - Real Time Sentiment Analysis- Stock Market Predictions.

Hadoop: History of Hadoop- the Hadoop Distributed File System – Components of Hadoop Analysing 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- Hadoop environment.

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 - IBM InfoSphere BigInsights and Streams.

Predictive Analytics- Simple linear regression- Multiple linear regression- Interpretation of regression coefficients. Visualizations - Visual data analysis techniques- interaction techniques - Systems and applications.
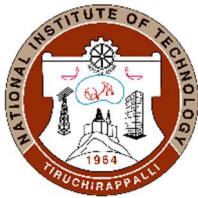
References:
1. Michael Berthold, David J. Hand, "Intelligent Data Analysis", Springer, 2007.
2. Tom White "Hadoop: The Definitive Guide" Third Edition, O'reilly Media, 2012.
3. Chris Eaton, Dirk DeRoos, Tom Deutsch, George Lapis, Paul Zikopoulos, "Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data", McGrawHill Publishing, 2012.
4. Anand Rajaraman and Jeffrey David Ullman, "Mining of Massive Datasets", CUP, 2012.
5. Bill Franks, "Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics", John Wiley& sons, 2012.
6. Glenn J. Myatt, "Making Sense of Data", John Wiley & Sons, 2007.
7. Pete Warden, "Big Data Glossary", O'Reilly, 2011.
8. Jiawei Han, Micheline Kamber "Data Mining Concepts and Techniques", 2nd Edition, Elsevier, 2008.
9. Da Ruan, Guoquing Chen, Etienne E.Kerre, Geert Wets, "Intelligent Data Mining", Springer, 2007.
10. Paul Zikopoulos, Dirkde Roos, Krishnan Parasuraman, Thomas Deutsch, James Giles , David Corrigan, "Harness the Power of Big Data The IBM Big Data Platform", Tata McGraw Hill, 2012.
11. Arshdeep Bahga, Vijay Madisetti, "Big Data Science & Analytics: A Hands-On Approach", VPT, 2016
12. Bart Baesens "Analytics in a Big Data World: The Essential Guide to Data Science and its Applications (WILEY Big Data Series)", John Wiley & Sons, 2014.

COURSE OBJECTIVES

- To optimize business decisions and create competitive advantage with Big Data
- analytics
- To explore the fundamental concepts of big data analytics.
- To learn to analyze the big data using intelligent techniques.
- To understand the various search methods and visualization techniques.
- To learn to use various techniques for mining data stream.
- To understand the applications using Map Reduce Concepts.
- To introduce programming tools PIG & HIVE in Hadoop echo system.

MAPPING OF COs with POs

| Course Outcomes	Programme Outcomes (PO) (Enter Numbers only)
1. Work with big data platform and explore the big data analytics techniques business applications.	PO I, II, III, IV
2. Design efficient algorithms for mining the data from large volumes.	PO II, III, IV, V
3. Analyze the HADOOP and Map Reduce technologies associated with big data analytics.	PO III, IV
4. Explore on Big Data applications Using Pig and Hive.	PO III, IV
5. Understand the fundamentals of various big data analytics techniques.	PO I, II

COURSE PLAN – PART II
COURSE OVERVIEW
COURSE TEACHING AND LEARNING ACTIVITIES
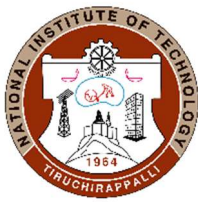
NATIONAL INSTITUTE OF TECHNOLOGY, TIRUCHIRAPPALLI

S.No.	Week/Contact Hours	Topic	Mode of Delivery
1	Week 1 (3 Classes)	Introduction to Big Data Platform – Challenges of Conventional Systems	PPT
2	Week 2 (3 Classes)	Nature of Data - Analytic Processes and Tools - Analysis vs Reporting	PPT
3	Week 3 (3 Classes)	Introduction To Streams Concepts – Stream Data Model and Architecture - Stream Computing - Sampling Data in a Stream	PPT
4	Week4 (3 Classes)	Filtering Streams – Counting Distinct Elements in a Stream – Estimating Moments – Counting Oneness in a Window – Decaying Window	PPT
5	Week 5 (3 Classes)	Real time Analytics Platform (RTAP) Applications – Case Studies - Real Time Sentiment Analysis- Stock Market Predictions	PPT
6	Week 6 (3 Classes)	History of Hadoop- the Hadoop Distributed File System – Components of Hadoop - Analysing the Data with Hadoop- Scaling Out- Hadoop Streaming	PPT
7	Week 7 (3 Classes)	Design of HDFS-Java interfaces to HDFS Basics- Developing a Map Reduce Application	PPT
8	Week 8 (3 Classes)	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-Hadoop environment	PPT
9	Week 9 (3 Classes)	Analyzing data with Pig- Hive services – HiveQL - Querying Data in Hive	PPT
10	Week 10 (3 Classes)	Fundamentals of HBase and ZooKeeper - IBM InfoSphere BigInsights and Streams	PPT
11	Week 11 (3 Classes)	Predictive Analytics- Simple linear regression- Multiple linear regression- Interpretation of regression coefficients	PPT
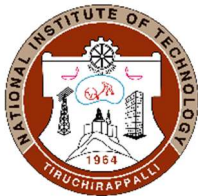
12	Week 12 (3 Classes)	Visualizations - Visual data analysis techniques- interaction techniques - Systems and applications	PPT

COURSE ASSESSMENT METHODS

S.No.	Mode of Assessment	Week/Date	Duration	% Weightage
1	Cycle Test 1	6th Week	60 Minutes	20
2	Cycle Test 2	10th Week	60 Minutes	20
3	Assignment & Seminar	7th week	-	30
CPA	Compensation Assessment			
4	Final Assessment	-	180 Minutes	50

COURSE EXIT SURVEY

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

MODE OF CORRESPONDENCE

By Email: 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.
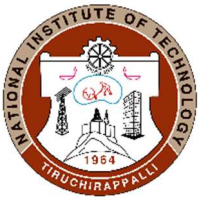
> ➢ Zero mark to be awarded for the offenders. For copying from another student, both students get the same penalty of zero mark.

> ➢ 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 _____

S.R.Balasundaram
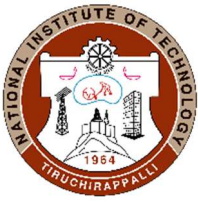
NATIONAL INSTITUTE OF TECHNOLOGY, TIRUCHIRAPPALLI

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.