

# GUJARAT TECHNOLOGICAL UNIVERSITY

**Subject Name: BigData Analytic(Elective)**

**Subject Code: 3725505**

## Semester II

**Type of course:** ME - Computer Engineering (HIGH PERFORMANCE COMPUTING [HPC])

**Prerequisite:**

1. Big Data
2. Cloud Computing

**Rationale:**

**Teaching and Examination Scheme:**

Teaching Scheme			Credits C	Examination Marks				Total Marks
L	T	P		Theory Marks		Practical Marks		
				ESE (E)	PA (M)	ESE (V)	PA (I)	
3	2#	0	4	70	30	30	20	150

L- Lectures; T- Tutorial/Teacher Guided Student Activity; P- Practical; C- Credit; ESE- End Semester Examination; PA- Progressive Assessment;

**Content:**

Sr. No.	Content	Total Hrs	% Weightage
1	Fundamentals of Big-data analytics, Overview & analytics life cycle, Need, Structured and multi-structured data analysis, Big-data analytics major components, Analytical models and approaches, Relational and non-relational Databases, Application areas, Design and analysis of Analytics model-Analytics design steps, Understanding different data processing models, Statistical models, Predictive models, Descriptive models, Introduction to Data analysis using R- Basics of R Language, Statistical models in R, Statistical techniques applied using R, Graphical analysis techniques, Practical examples/case studies.	10	20
2	Analytics Using MapReduce, MapReduce design patterns, Clustered Hadoop environment, Advanced HDFS, Graph Algorithms, Searching and Indexing approaches, MapReduce Applications, Introduction to Pig and HIVE- Programming Pig: Engine for executing data flows in parallel on Hadoop, Programming with Hive: Data warehouse system for Hadoop	10	25
3	Case Studies -Real time stream processing using MapReduce and R and Text Analytics based on different documents for identifying interesting patterns and correlations	10	25

**Reference Books:**

1. <http://www.amazon.com/Big-Data-Analytics-For-Beginners/dp/1495387348>
2. [http://www.amazon.com/Handbook-Statistical-Analyses-Second-Edition/dp/1420079336#reader\\_1420079336](http://www.amazon.com/Handbook-Statistical-Analyses-Second-Edition/dp/1420079336#reader_1420079336)
3. <http://www.packtpub.com/hadoop-mapreduce-cookbook/book>

4. <http://www.goodreads.com/book/show/14514285-mapreduce-design-patterns>
5. <http://www.goodreads.com/book/show/17568036-real-time-big-data-analytics>
6. <http://shop.oreilly.com/product/0636920023555.do>

**Course Outcome:**

After learning the course the students should be able to:

1. Demonstrates the basic operation in bigdata
2. Awareness of Architectures use in bigdata
3. Study about Big Data processing language Hadoop
4. Describes about the modern databases
5. Case studies in big data are analysed

**Review Presentation (RP):** The concerned faculty member shall provide the list of peer reviewed Journals and Tier-I and Tier-II Conferences relating to the subject (or relating to the area of thesis for seminar) to the students in the beginning of the semester. The same list will be uploaded on GTU website during the first two weeks of the start of the semester. Every student or a group of students shall critically study 2 papers, integrate the details and make presentation in the last two weeks of the semester. The GTU marks entry portal will allow entry of marks only after uploading of the best 3 presentations. A unique id number will be generated only after uploading the presentations. Thereafter the entry of marks will be allowed. The best 3 presentations of each college will be uploaded on GTU website