Nirma University

Institute of Technology

Computer Engineering Department

Course Policy Template

2CS702 **Big Data Analytics** **[2-0-2-3]**

B.Tech. (CSE)

Semester: VII , Academic Year: 2020-21, Term: Odd

Course Code & Name	**:**	**2CS702 - Big Data Analytics**
Credit Details	**:**	**3**
Course Co-ordinator	**:**	Dr. Jigna Ashish Patel
Subject Faculties	**:**	Dr. JaiPrakash Verma, Dr. Purnima Gandhi Dr. Aparna Kumari
Contact No. & Email	**:**	jignas.patel@nirmauni.ac.in
Office	**:**	N-F2
Visiting Hours	**:**	Monday to Friday - 8:45 to 4:00, Saturday (Odd) - 8:45 to 4:45
Course Blog	**:**	https://wordpress.com/stats/day/it7c4.wordpress.com
Course Site	**:**	https://sites.google.com/a/nirmauni.ac.in/bigdataanalytics/

Introduction to Course:

Big data analytics is the process of examining large and varied data sets i.e., big data to uncover hidden patterns, unknown correlations, market trends, customer preferences and other useful information that can help organizations make more-informed business decisions. This has applications in almost all domains.

Course Learning Outcomes:

After successful completion of this course, student will be able to
At the end of the course, students will be able to -
1. outline the significance and challenges of big data
2. model big data using different tools and frameworks

3. apply big data techniques for useful business analytic applications
4. design algorithms for mining the data from large volumes

Program Outcomes:

1. **Engineering knowledge**: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

2. **Problem analysis**: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

3. **Design/development of solutions**: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

4. **Conduct investigations of complex problems**: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

5. **Modern tool usage:** Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modelling to complex engineering activities with an understanding of the limitations.

6. **The engineer and society**: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

7. **Environment and sustainability**: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

8. **Ethics:** Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

9. **Individual and team work:** Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

10. **Communication:** Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

11. **Project management and finance**: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

12. **Life-long learning**: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

Program Specific Outcomes:

PSO 1: To apply the theoretical concepts of computer engineering and practical knowledge in analysis, design and development of computing systems and interdisciplinary applications

PSO 2: To work as a socially responsible professional by applying computer engineering principles and management practices

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
CLO1	2	0	0	0	2	0	0	0	0	0	0	0	1	0
CLO2	0	1	0	2	0	0	0	0	0	0	0	0	0	0
CLO3	0	0	0	0	3	0	0	0	0	1	0	1	2	1
CLO4	1	2	2	1	1	0	0	0	3	0	2	0	2	0

Syllabus

Syllabus **Teaching Hours: 30**

Unit I

Introduction to Big Data: Evolution of Big Data, Types of Digital Data, Classification of Digital Data, Structured Data, Semi-Structured Data, Unstructured Data, Definition of Big Data, Challenges of Conventional Systems, Big data platforms and data storage **04**

Unit II

Big Data Analytics: Importance of Big data analytics, Classification of Analytics, Top Challenges Facing Big Data, Technologies to meet the Challenges Posed by Big Data, Terminologies Used in Big Data Environment **04**

Unit III

Hadoop: Introducing Hadoop, comparisons of RDBMS and Hadoop, Distributed Computing Challenges, Hadoop Overview, Business Value of Hadoop, Hadoop Distributed File System, Processing Data with Hadoop, working with Map Reduce, Hadoop YARN, Hadoop in the Cloud, Applications on Big Hadoop Ecosystem, Fundamentals of Pig, Hive, HBase and ZooKeeper, Basic concepts of Apache Spark **08**

Unit IV

The Big data technology landscape: CAP Theorem - BASE Concept, NoSQL, Types of No SQL databases, Introduction to MongoDB, Data Types in MongoDB, CRUD, Apache Cassandra, Features of Cassandra, CRUD **08**

Unit V **06**

Big data analytics Algorithm: Applying Linear Regression, Clustering, Association rule mining, Decision tree on Big Data.

Self-Study:

The self-study contents will be declared at the commencement of semester. Around 10% of the questions will be asked from self-study contents.

Laboratory Work:

Laboratory work will be based on the above syllabus with minimum 10 experiments to be incorporated.

Suggested Readings^:

1. Michael Berthold, David J. Hand, Intelligent Data Analysis, Springer
2. Tom White, Hadoop: The Definitive Guide, O'reilly Media
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
4. Anand Rajaraman and Jeffrey David Ullman, Mining of Massive Datasets, Cambridge University Press
5. Bill Franks, Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics, John Wiley & sons
6. Glenn J. Myatt, Making Sense of Data, John Wiley & Sons
7. Da Ruan, Guoquing Chen, Etienne E.Kerre, GeertWets, Intelligent Data Mining, Springer
8. 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 Publications
9. Michael Minelli, Michele Chambers, Ambiga Dhiraj, Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses, Wiley Publications
10. Zikopoulos, Paul, Chris Eaton, Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, Tata McGraw Hill Publications
11. Seema Acharya and Subhashini C, Big Data and Analytics, Wiley India

---

Nirma University
Institute of Technology
Computer Engineering Department

Practical List

2CS702 **Big Data Analytics** **[2-0-2-3]**

Sr. NO	Practical Title	Hours	CLO
1.	Study and explore various applications of big data in different domains. Choose one of it and study in detail, Also write down the report on different types of digital data generated in selected application. For eg: • Big Data in Retail • Big Data in Healthcare • Big Data in Education • Big Data in E-commerce • Big Data in Media and Entertainment • Big Data in Finance • Big Data in Travel Industry • Big Data in Telecom • Big Data in Automobile	02 Hours	3
2	Learning limitation of data analytics by applying Machine Learning Techniques on large amount of data. Write a program to read data set from any online website, excel file and CSV file and to perform a) Linear regression and logistic regression on iris dataset.	02 Hours	3

	b) K-means clustering. • Students will learn the limitation of platform and algorithm.		
3.	Setup single node Hadoop cluster and apply HDFS commands on single node Hadoop Cluster.	04 hours	3
4.	Design MapReduce algorithms to take a very large file of integers and produce as output: a) The largest integer b) The average of all the integers. c) The same set of integers, but with each integer appearing only once. * d) The count of the number of distinct integers in the input.*	04 hours	3
5	Apply MapReduce algorithms to find phrase frequency from given dataset. • Prepare a report to guide design of mapper and reducer.	02 Hours	3
6	Analyse impact of different number of mapper and reducer on same definition as practical 4. • Prepare a conclusive report on analysis.	02 Hours	3
7	Implement any one of the analytic algorithm using mapreduce by handling larger datasets in main memory. (Machine Learning application) • PCY/Multi-Hash/SON algorithm • Regression • K-means Clustering	04 Hours	3
8	Setup MongoDB environment in your system. Import Restaurant Dataset and perform CRUD operation.	02 Hours	3
9	Setup Cassandra environment in your system and apply Create, Update, Read and Delete operations.	04 Hours	3
10	Case study: Use following platforms for solving any big data analytic problem of your choice. (1) Amazon web services,(2) Microsoft Azure, (3)Google App engine	02 Hours	3
11*	Extend MongoDB functionality for MapReduce on document collection	02 Hours	3
12*	Extend Cassandra functionality for Map Reduce on restaurant dataset.	02 Hours	3

Nirma University

Institute of Technology

Computer Engineering Department

Lesson Planning

2CS702 **Big Data Analytics** **[2-0-2-3]**

Lecture No.	Topic	Mapped CLO
1	**Introduction to Data Analytics, teaching scheme , Evaluation** methodology and overall instructions	-
2	Nature of Data, Types of Digital Data, Classification of Digital Data, Structured Data, Semi-Structured Data , Unstructured Data, Characteristics of Data, need for data analytics	**CLO1**
3	Introduction to Big Data, Evolution of Big Data , Definition of Big Data Challenges of Conventional Systems	CLO2
4	Intelligent Data Analysis, Challenges of Big Data Analytic Processes and Tools, Analysis vs Reporting, ,	**CLO3**
5	Statistical Concepts	**CLO4**
6	Sampling Distributions, Re-Sampling, Statistical Inference - Prediction Error	**CLO1**
7	importance of Big data analytics, Sudden Hype Around Big Data Analytics, ,	CLO2
8	Classification of Analytics	**CLO2**
9	Top Challenges Facing Big Data, Kind of Technologies to meet the Challenges Posed by Big Data	**CLO2**
10	Data Science, Role of data scientist, Terminologies Used in Big Data Environment	**CLO1**
11	Hadoop: Introducing Hadoop, comparisons of RDBMS and Hadoop, ,	CLO3
12	Distributed Computing Challenges, A Brief History of Hadoop, Hadoop Overview	**CLO2**
13	Business Value of Hadoop, Hadoop Distributors, Map Reduce	**CLO3**
14	Hadoop Distributed File System	**CLO3**
15	Processing Data with Hadoop ,	**CLO4**
16	Applications in Hadoop	**CLO4**
17	**I**ntroduction to Map reduce, working of Map reduce, , ,	**CLO3**
18	Hadoop YARN , Hadoop Ecosystem	**CLO3**
19	HDFS, Hadoop in the Cloud	**CLO2**
20	The Big data technology landscape, NoSQL Vs SQL	**CLO4**
21	NoSQL, Types of No SQL databases	**CLO4**
22	why No SQL for big data Analytics	**CLO1**
23	**NoSQL databases**	CLO3
24	Introduction to MongoDB, Introduction to MongoDB	CLO4

25	Apache Cassandra, Features of Cassandra,	CLO4
26	Applications on Big Data Using Pig and Hive,	CLO4
27	Data Processing Operators in Pig, Pig Scripting language	**CLO4**
28	Hive Services, HiveQL , Querying Data in Hive,	**CLO4**
29	Fundamentals of HBase	**CLO4**
30	Fundamentals of Zookeeper	**CLO4**
	Total Hours:	**30 hrs**

Component wise Continuous Evaluation & Semester End Examination weightage:

Examination Scheme

	CE	SEE	LPW
Exam Duration	Continuous Evaluation	3.0 Hrs	Continuous Evaluation + 2 hrs Semester End LPW Exam
Component Weightage	0.4	0.4	0.2

Breakup of CE

	Unit 1	Unit 2	Unit 3
	Quizzes(Quiz 1+Quiz 2)	Sessional Exam	Term Paper/Innovative Assignment
Inter Component Weightage	0.35	0.35	Term Paper (Project) 0.30

Teaching-learning methodology: (Mention the proposed)

- Lectures: Use of Black board, PPT, Discussion, Case Studies
- Laboratory: Experiment performance, programming, co-relation of Theory & Practical, Application of Experiment in Industry/Society, Scale-up, Modification, Upgradation etc.

Active learning techniques (Mention the proposed)

- Flip Classroom
- Collaborative Teaching

Types of Special/Innovative Assignments, Term Papers, mini Projects etc.

- **A soil Moisture project for all the stages of BDA**
- **Visualization in Tubule**
- **Presentation on Block Chain Technology**

Course Material: (In the website)

- Course Policy
- PPTs, Notes, other Material
- Assignments, Tutorials, Lab Manuals
- Question bank
- Web-links, Blogs, Video Lectures, Journals
- Animations /Simulations, Software
- Advanced topics
- Industries/Organizations

Course Outcome Attainment:

- Use of formal evaluation components of continuous evaluation, tutorials, laboratory work, semester end examination
- Informal feedback during course conduction