23-07-2026 04:49:09 PM

BIG DATA ANALYTICS **Course Code : 316318**

Programme Name/s : **Artificial Intelligence/ Artificial Intelligence and Machine Learning/ Data Sciences**
Programme Code : **AI/ AN/ DS**
Semester : **Sixth**
Course Title : **BIG DATA ANALYTICS**
Course Code : **316318**

I. RATIONALE

Big data analytics has many applications, including in healthcare, finance, transportation, and more. It helps businesses make better decisions, improve customer experience, and reduce costs. This course on Big Data Analytics introduces learners to the fundamental concepts of handling large and diverse datasets, including data classification, architecture, and processing techniques. The course provides hands-on experience with basic Big Data tools like Hadoop, NoSQL, Hive, Pig, and Spark. By developing foundational skills, learners will understand how to manage Big Data efficiently and support data handling tasks in various contexts, preparing them for future advanced applications.

II. INDUSTRY / EMPLOYER EXPECTED OUTCOME

The aim of this course is to help the student to attain the following industry relevant outcome through various teaching learning experiences:
Apply big data analytics to manage and analyze large datasets.

III. COURSE LEVEL LEARNING OUTCOMES (COS)

Students will be able to achieve & demonstrate the following COs on completion of course based learning

- CO1 - Illustrate different phases of big data with respect to real world application.
- CO2 - Demonstrate the use of Hadoop core components for big data processing.
- CO3 - Apply NoSQL database concepts and architecture patterns to manage big data.
- CO4 - Use Hive and Pig for data processing and transformation within big data environments.
- CO5 - Use Spark to process and analyze big data in real-time or archives.

IV. TEACHING-LEARNING & ASSESSMENT SCHEME

Course Code	Course Title	Abbr	Course Category/s	Learning Scheme					Credits	Assessment Scheme												
				Actual Contact Hrs./Week			SLH	NLH		Paper Duration	Theory				Based on LL & TL				Based on SL		Total Marks
															Practical							
				CL	TL	LL					FA-TH	SA-TH	Total		FA-PR		SA-PR		SLA		
											Max	Max	Max	Min	Max	Min	Max	Min	Max	Min	
316318	BIG DATA ANALYTICS	BDA	DSC	3	-	4	1	8	4	3	30	70	100	40	25	10	-	-	25	10	150

BIG DATA ANALYTICS **Course Code : 316318**

Total IKS Hrs for Sem. : 0 Hrs
Abbreviations: CL- ClassRoom Learning , TL- Tutorial Learning, LL-Laboratory Learning, SLH-Self Learning Hours, NLH-Notional Learning Hours, FA - Formative Assessment, SA -Summative assessment, IKS - Indian Knowledge System, SLA - Self Learning Assessment
Legends: @ Internal Assessment, # External Assessment, *# On Line Examination , @$ Internal Online Examination
Note :

1. FA-TH represents average of two class tests of 30 marks each conducted during the semester.
2. If candidate is not securing minimum passing marks in FA-PR of any course then the candidate shall be declared as "Detained" in that semester.
3. If candidate is not securing minimum passing marks in SLA of any course then the candidate shall be declared as fail and will have to repeat and resubmit SLA work.
4. Notional Learning hours for the semester are (CL+LL+TL+SL)hrs.* 15 Weeks
5. 1 credit is equivalent to 30 Notional hrs.
6. * Self learning hours shall not be reflected in the Time Table.
7. * Self learning includes micro project / assignment / other activities.

V. THEORY LEARNING OUTCOMES AND ALIGNED COURSE CONTENT

Sr.No	Theory Learning Outcomes (TLO's)aligned to CO's.	Learning content mapped with Theory Learning Outcomes (TLO's) and CO's.	Suggested Learning Pedagogies.
1	TLO 1.1 Classify the given data. TLO 1.2 Explain the characteristics of Big Data. TLO 1.3 Describe different types of Big Data. TLO 1.4 Explain the architecture of Big Data Processing. TLO 1.5 Illustrate different phases of big data analytics. TLO 1.6 Describe real-world applications of Big Data analytics.	**Unit - I Introduction to Big Data Analytics and Data Architecture** 1.1 Classification of Data : Structured, Semi-structured and Unstructured 1.2 Introduction :Big Data Definitions, Need of Big Data 1.3 Big Data Characteristics : Volume, Velocity, Variety, Veracity 1.4 Big Data Types 1.5 Big Data Processing Architecture Design 1.6 Big Data Analytics : Data analytics Definitions, Phases in Analytics 1.7 Big Data Analytics Applications : Big Data in Marketing and Sales, Big Data and Healthcare, Big Data in Medicine, Big Data in Advertising	Lecture Using Chalk-Board Presentations Video Demonstrations Case Study Sample Dataset Handling (Demonstration Only)

BIG DATA ANALYTICS **Course Code : 316318**

Sr.No	Theory Learning Outcomes (TLO's)aligned to CO's.	Learning content mapped with Theory Learning Outcomes (TLO's) and CO's.	Suggested Learning Pedagogies.
2	TLO 2.1 Explain the feature of Hadoop framework and its ecosystem. TLO 2.2 Explain the functioning of Hadoop Distributed File System (HDFS) for data storage and interaction. TLO 2.3 Explain the processing workflow of MapReduce Framework. TLO 2.4 Describe the MapReduce Programming Model. TLO 2.5 Describe the functioning of YARN in Hadoop's execution model. TLO 2.6 Explain MapReduce processing, including map tasks, reduce tasks, and combiners.	**Unit - II Introduction to Hadoop and MapReduce** 2.1 Introduction to Hadoop 2.2 Hadoop and its Ecosystem : Hadoop Core Components, Features of Hadoop, Hadoop Ecosystem Components 2.3 Hadoop Distributed File System : HDFS data storage, HDFS Commands for interacting with files in HDFS 2.4 MapReduce Framework and Programming Model : Hadoop MapReduce Framework, MapReduce Programming Model 2.5 Hadoop Yarn : Hadoop 2 Execution Model 2.6 MapReduce : Map Tasks, Key-Value Pair, Grouping by Key, Partitioning, Combiners, Reduce Tasks, Details of MapReduce Processing Steps	Lecture Using Chalk-Board Presentations Demonstration
3	TLO 3.1 Describe the purpose and importance of NoSQL in Big Data. TLO 3.2 Explain the CAP theorem. TLO 3.3 Explain the schema-less data models. TLO 3.4 Describe various NoSQL data architecture patterns, including key-value, document, tabular, object, and graph stores. TLO 3.5 Explain the use of NoSQL database for managing Big Data. TLO 3.6 Describe the features of MongoDB for Big Data storage and management.	**Unit - III NoSQL Databases and Big Data Management** 3.1 Introduction NoSQL in Big Data 3.2 NoSQL Data Store : NoSQL, CAP theorem, Schema-less Models 3.3 NoSQL Data Architecture Patterns : Key-Value Store, Document Store, Tabular Data, Object Data Store.Graph Database 3.4 NoSQL to manage Big Data 3.5 MongoDB Database	Lecture Using Chalk-Board Presentations Demonstration
4	TLO 4.1 Describe the characteristics of Hive. TLO 4.2 Describe the architecture of Hive. TLO 4.3 Explain Hive data types, file formats. TLO 4.4 Explain Hive integration workflow. TLO 4.5 Write the process of applying HiveQL for data definition, manipulation, and querying. TLO 4.6 Compare Pig with SQL, MapReduce, and Hive. TLO 4.7 Describe the architecture of Pig. TLO 4.8 Explain the approach of Pig Scripting with Pig Latin Data Model.	**Unit - IV Hive and Pig** 4.1 Introduction to Hive : Hive Characteristics, Limitations 4.2 Hive Architecture 4.3 Hive Data Types and File Formats 4.4 Hive Integration and Workflow Steps 4.5 Hive Built-in functions 4.6 HiveQL : HiveQL DDL, HiveQL DML, HiveQL for Querying the Data 4.7 Introduction to Pig : Applications of Apache Pig, Features of Pig, Compare Pig with SQL, MapReduce, and Hive 4.8 Pig Architecture 4.9 Pig Latin Data Model	Lecture Using Chalk-Board Presentations Demonstration

BIG DATA ANALYTICS **Course Code : 316318**

Sr.No	Theory Learning Outcomes (TLO's)aligned to CO's.	Learning content mapped with Theory Learning Outcomes (TLO's) and CO's.	Suggested Learning Pedagogies.
5	TLO 5.1 Describe the architecture of Apache Spark. TLO 5.2 Write a query using Spark SQL for data analysis. TLO 5.3 Write the purpose of various commands used with Resilient Distributed Dataset (RDDs). TLO 5.4 Explain the use of MLib library for machine learning programming. TLO 5.5 Explain the process of composing Spark program steps for ETL. TLO 5.6 Explain methods for analytics, reporting, and visualization using Spark. TLO 5.7 Compare Spark Streaming with Structured Streaming. TLO 5.8 Describe Spark Streaming Architecture. TLO 5.9 Explain Spark Streaming characteristics including scalability, fault tolerance, and load balancing.	**Unit - V Spark and Real-Time Analytics** 5.1 Introduction to Big Data tool Spark : Main components of Spark Architecture, Features of Spark, Spark Software Stack 5.2 Introduction to Data Analysis with Spark : Spark SQL 5.3 Programming with RDDs and Machine learning with MLib 5.4 Data ETL (Extract, Transform and Load) Process: Composing Spark Program steps for ETL 5.5 Analytics, Reporting and Visualization 5.6 Apache Spark Streaming Platform: Spark Streaming Architecture, Spark streaming vs Structured streaming, Internal Working of Spark Streaming 5.7 Spark streaming characteristics: Scalable, Fault Tolerance and Load Balancing	Lecture Using Chalk-Board Presentations Demonstration

VI. LABORATORY LEARNING OUTCOME AND ALIGNED PRACTICAL / TUTORIAL EXPERIENCES.

Practical / Tutorial / Laboratory Learning Outcome (LLO)	Sr No	Laboratory Experiment / Practical Titles / Tutorial Titles	Number of hrs.	Relevant COs
LLO 1.1 Identify Big Data use cases and explain the analytics process applied.	1	*Conduct a Case Study on Big Data and Big Data Analysis	2	CO1
LLO 2.1 Setup a Hadoop ecosystem on a local cluster.	2	*Install Hadoop Ecosystem on Local Cluster	2	CO2
LLO 3.1 Configure Hadoop Ecosystem on Local Cluster.	3	*Configure Hadoop Ecosystem on Local Cluster 1. Configure core-site.xml, hdfs-site.xml, mapred-site.xml 2. Start HDFS daemons 3. Upload any dataset to HDFS	2	CO2
LLO 4.1 Setup multi-node hadoop cluster.	4	Install Hadoop on multiple nodes	2	CO2
LLO 5.1 Configure multi-node hadoop cluster.	5	Configure Multi-Node Hadoop Cluster 1. Configure core-site.xml and hdfs-site.xml for multi-node HDFS setup 2. Start NameNode, DataNode daemons across nodes 3. Upload any dataset to HDFS and verify distributed storage	2	CO2

BIG DATA ANALYTICS **Course Code : 316318**

Practical / Tutorial / Laboratory Learning Outcome (LLO)	Sr No	Laboratory Experiment / Practical Titles / Tutorial Titles	Number of hrs.	Relevant COs
LLO 6.1 Use resource utilization and performance tools in Hadoop Cluster.	6	Use Monitoring tools to Observe Cluster Resources 1. Access Hadoop monitoring interfaces (CLI or web UI) 2. Locate resource usage indicators (CPU, memory, disk) 3. Run basic monitoring commands 4. Record resource values during cluster operation	2	CO2
LLO 7.1 Execute file operations using HDFS commands.	7	*Perform Basic File Operation using HDFS 1. Create directories and Files in HDFS 2. Perform read, write, update and delete operations 3. Set directory permissions 4. Verify file replication	2	CO2
LLO 8.1 Execute basic backup and restore operations for Big Data stored in HDFS.	8	Perform Backup and Restore of Datasets in HDFS 1. Load any datasets in HDFS 2. Copy datasets to a backup directory 3. Simulate accidental data removal 4. Restore the dataset from the backup	2	CO2
LLO 9.1 Develop a MapReduce program.	9	*Execute WordCount MapReduce Program 1. Load any text file into HDFS 2. Develop a Word-Count program in Java 3. Compile, execute, and validate output	2	CO2
LLO 10.1 Execute a data processing workflow with MapReduce for CSV files.	10	*Process Large CSV Dataset Using MapReduce 1. Load a CSV dataset into HDFS 2. Run a MapReduce job to extract specific fields 3. Aggregate data using the Reducer	2	CO2
LLO 11.1 Setup a MongoDB NoSQL database with collections.	11	*Install NoSQL Database (MongoDB) and Create Collections	2	CO3
LLO 12.1 Create a schema for unstructured data in MongoDB using any dataset.	12	*Create a schema for unstructured data in MongoDB 1. Select any unstructured dataset in JSON format (e.g., reviews, logs, or profiles) 2. Identify common fields and structure 3. Define a suitable document schema for storing the data in MongoDB	2	CO3

BIG DATA ANALYTICS **Course Code : 316318**

Practical / Tutorial / Laboratory Learning Outcome (LLO)	Sr No	Laboratory Experiment / Practical Titles / Tutorial Titles	Number of hrs.	Relevant COs
LLO 13.1 Run basic aggregation queries on Unstructured dataset in MongoDB.	13	*Run basic aggregation queries on Unstructured dataset in MongoDB 1. Import the dataset created in the previous experiment into a MongoDB collection 2. Run basic aggregation queries to validate the schema (such as counting records, grouping by a field, or calculating averages)	2	CO3
LLO 14.1 Apply basic operations in MongoDB to observe the behavior of CAP theorem properties.	14	Perform Basic MongoDB Operations Demonstrating CAP Theorem Behavior 1. Install and set up MongoDB with replica sets 2. Perform basic write and read operations across replica sets 3. Simulate a simple network delay or temporary disconnection between nodes (using basic server control commands) 4. Observe whether read and write operations continue or pause during the disruption 5. Record the status of the system (whether operations succeed or wait) during normal and disrupted states 6. Save sample outputs (screenshots or logs) showing system behavior	2	CO3
LLO 15.1 Install Hive within a Hadoop environment.	15	*Install and Configure Hive 1. Install Hive on Hadoop environment 2. Load a dataset into Hive 3. Run basic HiveQL queries (select, insert, delete)	2	CO4
LLO 16.1 Execute after writing complex data queries using HiveQL.	16	*Execute Data Queries Using HiveQL 1. Load any dataset into Hive 2. Execute joins, group by, and aggregate queries	2	CO4
LLO 17.1 Run after writing Pig scripts for data transformation.	17	Run Pig Scripts for data transformation 1. Load any dataset into Pig 2. Apply filtering and grouping operations	2	CO4

BIG DATA ANALYTICS **Course Code : 316318**

Practical / Tutorial / Laboratory Learning Outcome (LLO)	Sr No	Laboratory Experiment / Practical Titles / Tutorial Titles	Number of hrs.	Relevant COs
LLO 18.1 Use after writing User Defined Functions (UDF) in Pig.	18	*Execute User Defined Functions (UDF) in Pig for Data Normalization 1. Load any numerical dataset into Pig 2. Develop a custom UDF to normalize numerical readings 3. Register the UDF Script Apply the UDF for normalization	2	CO4
LLO 19.1 Install Spark on a cluster and verify installation.	19	*Install Spark on Cluster environment and verify installation with any dataset 1. Install Spark on the cluster 2. Start spark shell 3. Load a sample dataset 4. Verify basic operations	2	CO5
LLO 20.1 Perform data transformations with RDDs in Spark.	20	*Perform Data Transformations with RDDs and apply map, filter, reduce operations 1. Load any dataset into an RDD 2. Apply map, filter and reduce operations	2	CO5
LLO 21.1 Execute an end-to-end ETL process using Spark.	21	*Perform ETL Process on Big Data Using Spark 1. Load dataset from local storage or HDFS 2. Apply basic transformations (such as filtering and aggregation) on the dataset 3. Store the transformed data back into local storage, or HDFS	2	CO5
LLO 22.1 Load structured data and create temporary views in Spark SQL.	22	Load Structured Dataset and Create Temporary Views in Spark SQL 1. Select and load any structured dataset (such as CSV or JSON) into Spark 2. Inspect the dataset to understand its structure (columns and data types) 3. Create a temporary view from the loaded dataset 4. Display basic records from the view to verify successful creation	2	CO5
LLO 23.1 Run after writing basic SQL queries on datasets using Spark SQL.	23	Perform Select, Join, and Group By Queries in Spark SQL 1. Use the temporary view created from a structured dataset in a previous experiment 2. Execute SELECT queries to retrieve specific columns 3. Perform JOIN operations between two views or datasets (if applicable) 4. Apply GROUP BY queries to aggregate data (such as sums or averages) 5. Display and save the query results	2	CO5

BIG DATA ANALYTICS **Course Code : 316318**

Practical / Tutorial / Laboratory Learning Outcome (LLO)	Sr No	Laboratory Experiment / Practical Titles / Tutorial Titles	Number of hrs.	Relevant COs
LLO 24.1 Stream real- time text and count words using Spark streaming.	24	Perform Real-Time Word Count Using Spark Streaming 1. Connect to socket or stream text data 2. Process incoming lines 3. Count word frequency in real-time 4. Display live results	2	CO5
LLO 25.1 Save after extracting streaming data into persistent storage using Spark Streaming.	25	Store Real-Time Streamed Data from Spark streaming into HDFS 1. Connect to a text stream (such as socket input or generated logs) 2. Capture real-time data using Spark Streaming 3. Write the incoming stream to HDFS in regular intervals	2	CO5
LLO 26.1 Run after creating a regression model on a large dataset in Spark.	26	*Apply Simple Regression on Large Dataset in Spark 1. Load any numerical dataset 2. Build and train a simple regression model 3. Train and apply the model 4. Predict target values	2	CO5
LLO 27.1 Apply the K-Means clustering algorithm in Spark MLlib.	27	*Apply K-Means Clustering on any Dataset using Spark MLlib 1. Select any structured dataset suitable for clustering (e.g., customer data, product features) 2. Load the dataset into Spark and prepare it for clustering (select relevant numerical features) 3. Apply the K-Means clustering algorithm with a defined number of clusters (K) 4. Output the cluster assignments for each data point 5. Save the clustering result for further use	2	CO5
LLO 28.1 Visualize clustering results of a dataset processed with Spark MLlib.	28	*Visualize K-Means Clustering Results Using Spark MLlib 1. Load the cluster assignment results from the previous experiment 2. Use visualization tools (such as Matplotlib, Seaborn, or any Spark-compatible library) 3. Plot the clusters on a 2D graph based on two key features 4. Color-code data points according to cluster labels 5. Export the visualization as an image or report	2	CO5

BIG DATA ANALYTICS **Course Code : 316318**

Practical / Tutorial / Laboratory Learning Outcome (LLO)	Sr No	Laboratory Experiment / Practical Titles / Tutorial Titles	Number of hrs.	Relevant COs
LLO 29.1 Classify a dataset using the Decision Tree algorithm in Spark MLlib.	29	Apply Decision Tree Classification on any Dataset using Spark MLlib 1. Select any labeled dataset suitable for classification (e.g., spam detection, loan approval) 2. Load the dataset into Spark and preprocess it (select features and label) 3. Apply the Decision Tree classification algorithm to train the model 4. Generate predictions for the dataset 5. Save the prediction results for further processing	2	CO5
LLO 30.1 Display classification outputs from Decision Tree results.	30	Display Classification Results from Decision Tree Model in Spark MLlib 1. Load the saved classification results from the previous experiment 2. Display the predicted class labels alongside the original data 3. Use simple metrics (such as counts of predicted classes) to observe results 4. Save the final classified dataset to a file 5. Generate a basic summary of the output	2	CO5

Note : Out of above suggestive LLOs -

- '*' Marked Practicals (LLOs) Are mandatory.
- Minimum 80% of above list of lab experiment are to be performed.
- Judicial mix of LLOs are to be performed to achieve desired outcomes.

VII. SUGGESTED MICRO PROJECT / ASSIGNMENT/ ACTIVITIES FOR SPECIFIC LEARNING / SKILLS DEVELOPMENT (SELF LEARNING)

Micro project

- Store and query weather data to create visualization of temperature patterns.
- Analyze customer purchase data to identify buying patterns by collecting e-commerce dataset. Identify top selling products, frequently purchased items and trends.
- Analyze real-time tweets to determine public sentiment on a topic and display sentiment analysis results.
- Analyze student records using a NoSQL database to summarize insights of student performance.
- Process and transform sales data to generate a report showing sales trends.

Assignment

- Collect and classify real-world datasets (structured, semi-structured, unstructured) from publicly available sources and create a comparison table.
- Prepare a poster or infographic explaining the 4 Vs (Volume, Velocity, Variety, Veracity) using real-life examples.
- Compare features of different NoSQL models (Key-Value, Document, Graph, Tabular) in a table with real-life use cases.

Other

- Swayam NPTEL Course on Big Data Computing

BIG DATA ANALYTICS **Course Code : 316318**

Note :

- Above is just a suggestive list of microprojects and assignments; faculty must prepare their own bank of microprojects, assignments, and activities in a similar way.
- The faculty must allocate judicial mix of tasks, considering the weaknesses and / strengths of the student in acquiring the desired skills.
- If a microproject is assigned, it is expected to be completed as a group activity.
- SLA marks shall be awarded as per the continuous assessment record.
- For courses with no SLA component the list of suggestive microprojects / assignments/ activities are optional, faculty may encourage students to perform these tasks for enhanced learning experiences.
- If the course does not have associated SLA component, above suggestive listings is applicable to Tutorials and maybe considered for FA-PR evaluations.

VIII. LABORATORY EQUIPMENT / INSTRUMENTS / TOOLS / SOFTWARE REQUIRED

Sr.No	Equipment Name with Broad Specifications	Relevant LLO Number
1	MongoDB Community Server (version 6.x)	10,11,12,13
2	Apache Hive (version 3.x)	14,15,16
3	Apache Pig (version 0.17.x)	17,18
4	Apache Spark (version 3.x)	19,20,21,22,23,24,25,26,27,28,29,30
5	Apache Hadoop (version 3.x)	2,3,4,5,6,7,8,9
6	Personal Computer (i5 or higher), Minimum 8GB RAM (16 GB Recommended), 500 GB HDD/SSD	All
7	Operating System:Windows 10 or higher/Ubuntu 20.04 LTS or higher.	All

IX. SUGGESTED WEIGHTAGE TO LEARNING EFFORTS & ASSESSMENT PURPOSE (Specification Table)

Sr.No	Unit	Unit Title	Aligned COs	Learning Hours	R-Level	U-Level	A-Level	Total Marks
1	I	Introduction to Big Data Analytics and Data Architecture	CO1	8	4	8	0	12
2	II	Introduction to Hadoop and MapReduce	CO2	10	4	6	6	16
3	III	NoSQL Databases and Big Data Management	CO3	8	4	4	4	12
4	IV	Hive and Pig	CO4	9	4	6	4	14
5	V	Spark and Real-Time Analytics	CO5	10	4	6	6	16
Grand Total				**45**	**20**	**30**	**20**	**70**

X. ASSESSMENT METHODOLOGIES/TOOLS

Formative assessment (Assessment for Learning)

- For theory two offline unit tests of 30 marks and average of two unit test marks will be considered for out of 30 marks.
- Each practical will be assessed considering 60% weightage to process, 40% weightage to product.

Summative Assessment (Assessment of Learning)

- End Semester Examination, Viva-Voce.

XI. SUGGESTED COS - POS MATRIX FORM

BIG DATA ANALYTICS **Course Code : 316318**

Course Outcomes (COs)	Programme Outcomes (POs)							Programme Specific Outcomes* (PSOs)		
	PO-1 Basic and Discipline Specific Knowledge	PO-2 Problem Analysis	PO-3 Design/ Development of Solutions	PO-4 Engineering Tools	PO-5 Engineering Practices for Society, Sustainability and Environment	PO-6 Project Management	PO-7 Life Long Learning	PSO-1	PSO-2	PSO-3
CO1	3									
CO2	3	3	3	2	2		1			
CO3	3	3	3	3	1		1			
CO4	3	3	3	3	1	3	1			
CO5	3	3	3	3	2	2	2			

Legends :- High:03, Medium:02,Low:01, No Mapping: -
*PSOs are to be formulated at institute level

XII. SUGGESTED LEARNING MATERIALS / BOOKS

Sr.No	Author	Title	Publisher with ISBN Number
1	Raj Kamal, Preeti Saxena	Big Data Analytics: Introduction to Hadoop, Spark, and Machine-Learning	McGraw Hill Education, New Delhi. ISBN: 9789353164962
2	Seema Acharya, Subhashini Chellappan	Big Data and Analytics	Wiley India Pvt. Ltd., ISBN: 9788126554782
3	M. Vijayalakshmi, Radha Shankarmani	Big Data Analytics	Publication details: Wiley c2017, 2022 N. Delhi Edition: 2nd ed. c2017,ISBN: 9788126565757
4	Holden Karau, Andy Konwinski, Patrick Wendell, Matei Zaharia	Learning Spark: Lightning-Fast Data Analytics	O'Reilly Media Publication Date: January 28, 2015 ISBN-10: 1449358624 ISBN-13: 978-1449358624
5	Pramod J. Sadalage, Martin Fowler	NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence	Addison-Wesley August 10, 2012 ISBN: 978-0321826626
6	Tom White	Hadoop: The Definitive Guide	4th Edition, Released April 2015, Publisher(s): O'Reilly Media, Inc. ISBN: 9781491901632.

XIII . LEARNING WEBSITES & PORTALS

Sr.No	Link / Portal	Description
1	https://hadoop.apache.org/	Official website for Apache Hadoop, including documentation, downloads, and tutorials.
2	https://spark.apache.org/	Official website for Apache Spark, providing guides, API references, and use case examples.
3	https://pig.apache.org/	Official site for Apache Pig, with resources for learning Pig Latin and building scripts.
4	https://hive.apache.org/	Official resource for Apache Hive, including installation guides and HiveQL references.
5	https://www.mongodb.com/	MongoDB official site offering documentation, downloads, and free learning courses.

23-07-2026 04:49:09 PM

BIG DATA ANALYTICS **Course Code : 316318**

Sr.No	Link / Portal	Description
6	https://onlinecourses.nptel.ac.in/noc20_cs92/preview	This course provides an in-depth understanding of terminologies and the core concepts behind big data problems, applications, systems and the techniques, that underlie todays big data computing technologies. It provides an introduction to some of the most common frameworks such as Apache Spark, Hadoop, MapReduce.
7	https://www.tutorialspoint.com/hadoop/index.htm	This brief tutorial provides a quick introduction to Big Data, MapReduce algorithm, and Hadoop Distributed File System.
8	https://www.w3schools.com/mongodb/	This brief tutorial provides a quick introduction to MongoDB.

Note :

- Teachers are requested to check the creative common license status/financial implications of the suggested online educational resources before use by the students

MSBTE Approval Dt. 04/09/2025 **Semester - 6, K Scheme**