

**Maharashtra State Board of Technical Education, Mumbai**  
**TEACHING PLAN (TP)**

**K-1**

**Academic Year: 2025-26**

**Date: 12/12/2025**

**Institute Name & Code: K. K. Wagh Polytechnic, Nashik-3 (0078)**

**Program & Code: Artificial Intelligence & Machine Learning (AN)**

**Course Code & Abbr.: 316318(BDA)**

**Course Name: Big Data Analytics**

**Name of Faculty: Mrs. J. S. Mahajan**

**Class: TYAN**

**Course Index: 602**

**Semester: VI<sup>rd</sup>**

**Scheme: 'K'**

**Total Hrs: 45**

**● Teaching-Learning and Assessment Scheme:**

Course Code	Course Title	Abbr	Course Category/s	Learning Scheme			Credits	Paper Duration	Assessment Scheme								Total Marks			
				Actual Contact Hrs./Week					Theory		Based on LL & TL		Practical		Based on SL					
				CL	TL	LL			SLH	NLH	FA-TH	SA-TH	Total	FA-PR	SA-PR	SLA				
				Max	Max	Max			Min	Min	Max	Max	Min	Max	Min	Max	Min			
316318	BIG DATA ANALYTICS	BDA	DSC	3	-	4	1	8	4	3	30	70	100	40	25	20	-	25 10 150		

**Abbreviations:** Abbreviations: CL- ClassRoom Learning , LL-Laboratory Learning, SLH-Self Learning Hours, NLH-Notional Learning Hours, FA - Formative Assessment, SA -Summative assessment, SLA - Self Learning Assessment

**● Course Outcomes (COs) and Theory Learning Outcomes (TLOs):**

By learning course Big Data Analytics (BDA-316318), Third Year students will be able to:

CO No.	TLO No.	Course Outcomes (COs) / Theory Learning Outcomes (TLOs)
CO602.1		<b>Illustrate different phases of big data with respect to real world application.</b>
CO602.1	1.1	Classify the given data.
	1.2	Explain the characteristics of Big Data.
	1.3	Describe different types of Big Data.
	1.4	Explain the architecture of Big Data Processing.
	1.5	Illustrate different phases of big data analytics.
	1.6	Describe real-world applications of Big Data analytics.
CO602.2		<b>Demonstrate the use of Hadoop core components for big data processing.</b>
CO602.2	2.1	Explain the feature of Hadoop framework and its ecosystem.
	2.2	Explain the functioning of Hadoop Distributed File System (HDFS) for data storage and interaction.
	2.3	Explain the processing workflow of MapReduce Framework.
	2.4	Describe the MapReduce Programming Model.
	2.5	Describe the functioning of YARN in Hadoop's execution model.
	2.6	Explain MapReduce processing, including map tasks, reduce tasks, and combiners.
CO602.3		<b>Apply NoSQL database concepts and architecture patterns to manage big data.</b>
CO602.3	3.1	Describe the purpose and importance of NoSQL in Big Data.
	3.2	Explain the CAP theorem.
	3.3	Explain the schema-less data models.
	3.4	Describe various NoSQL data architecture patterns, including key-value, document, tabular, object, and graph stores.
	3.5	Explain the use of NoSQL database for managing Big Data.
	3.6	Describe the features of MongoDB for Big Data storage and management.
CO602.4		<b>Use Hive and Pig for data processing and transformation within big data environments.</b>
	4.1	Describe the characteristics of Hive.

	4.2	Describe the architecture of Hive.
	4.3	Explain Hive data types, file formats.
	4.4	Explain Hive integration workflow.
	4.5	Write the process of applying HiveQL for data definition, manipulation, and querying.
	4.6	Compare Pig with SQL, MapReduce, and Hive.
	4.7	Describe the architecture of Pig.
	4.8	Explain the approach of Pig Scripting with Pig Latin Data Model.
<b>CO602.5</b>		<b>Use Spark to process and analyze big data in real-time or archives.</b>
	5.1	Describe the architecture of Apache Spark.
	5.2	Write a query using Spark SQL for data analysis.
	5.3	Write the purpose of various commands used with Resilient Distributed Dataset (RDDs).
	5.4	Explain the use of MLlib library for machine learning programming.
	5.5	Explain the process of composing Spark program steps for ETL.
	5.6	Explain methods for analytics, reporting, and visualization using Spark.
	5.7	Compare Spark Streaming with Structured Streaming.
	5.8	Describe Spark Streaming Architecture.
	5.9	Explain Spark Streaming characteristics including scalability, fault tolerance, and load balancing.

- Teaching Plan:**

Unit No. (Allotted Hrs.) Marks	Theory Learning Outcomes (TLO)	Title/Topic Details with CO	Plan (From -To & No. of Lectures)	Actual Execution (From-To & No. of Lectures)	Pedagogy used (Teaching Method/ Media)	Remark
<b>01</b> (08hrs) 12M	1.1	<b>Unit - I Introduction to Big Data Analytics and Data Architecture</b> 1.1 Classification of Data : Structured, Semi-structured and Unstructured 1.2 Introduction :Big Data Definitions, Need of Big Data	18/12/25 To 19/12/25 (02)		Chalk, Board, PPTs, Web References	
	1.2,1.3	1.3 Big Data Characteristics : Volume, Velocity, Variety, Veracity 1.4 Big Data Types	20/12/25 To 26/12/25 (02)			
	1.4	1.5 Big Data Processing Architecture Design	01/01/26 (01)			
	1.5	1.6 Big Data Analytics : Data analytics Definitions, Phases in Analytics	02/01/26 (01)			
	1.6	1.7 Big Data Analytics Applications : Big Data in Marketing and Sales, Big Data and Healthcare, Big Data in Medicine, Big Data in Advertising	03/01/26 To 08/01/26 (02)			
	2.1	<b>Unit - II Introduction to Hadoop and MapReduce</b> 2.1 Introduction to Hadoop 2.2 Hadoop and its Ecosystem : Hadoop Core Components, Features of Hadoop, Hadoop Ecosystem Components	09/01/26 To 10/01/26 (02)			
<b>02</b> (10hrs) 16M					Chalk, Board, PPTs, Web References	References

	2.2	2.3 Hadoop Distributed File System : HDFS data storage, HDFS Commands for interacting with files in HDFS	15/01/26 To 16/01/26 (02)			
	2.3,2.4	2.4 MapReduce Framework and Programming Model : Hadoop MapReduce Framework, MapReduce Programming Model	17/01/26 To 22/01/26 (02)			
	2.5	2.5 Hadoop Yarn : Hadoop 2 Execution Model	23/01/26 (02)			
	2.6	2.6 MapReduce : Map Tasks, Key-Value Pair, Grouping by Key, Partitioning, Combiners, Reduce Tasks, Details of MapReduce Processing Steps	23/01/26 To 24/01/26 (02)			
<b>03</b> (08hrs) 12M	3.1, 3.2, 3.3	<b>Unit - III NoSQL Databases and Big Data Management</b> 3.1 Introduction NoSQL in Big Data 3.2 NoSQL Data Store : NoSQL, CAP theorem, Schema-less Models	29/01/26 To 30/01/26 (02)		Chalk, Board, PPTs, Web References	
	3.4	3.3 NoSQL Data Architecture Patterns : Key-Value Store, Document Store, Tabular Data, Object Data Store, Graph Database	31/01/26 To 05/02/26 (02)			
	3.5	3.4 NoSQL to manage Big Data	06/02/26 (01)			
	3.6	3.5 MongoDB Database	06/02/26 To 07/02/26 (02)			
<b>04</b> (09hrs) 14M	4.1	<b>Unit - IV Hive and Pig</b> 4.1 Introduction to Hive : Hive Characteristics, Limitations	12/02/26 (01)		Chalk, Board, PPTs, Web References	
	4.2	4.2 Hive Architecture	13/02/26 (01)			
	4.3,4.4	4.3 Hive Data Types and File Formats 4.4 Hive Integration and Workflow Steps	14/02/26 To 20/02/26 (02)			
	4.5	4.5 Hive Built-in functions 4.6 HiveQL : HiveQL DDL, HiveQL DML, HiveQL for Querying the Data	21/02/26 To 26/02/26 (02)			
	4.6	4.7 Introduction to Pig : Applications of Apache Pig, Features of Pig, Compare Pig with SQL, MapReduce, and Hive	27/02/26 (01)			
	4.7	4.8 Pig Architecture	28/02/26 (01)			
	4.8	4.9 Pig Latin Data Model	05/03/26 (01)			
<b>05</b> (10hrs) 16M	5.1	<b>Unit - V Spark and Real-Time Analytics</b> 5.1 Introduction to Big Data tool Spark : Main components of Spark	06/03/26 (02)		Chalk, Board, PPTs, Web References	

		Architecture, Features of Spark, Spark Software Stack			
5.2	5.2 Introduction to Data Analysis with Spark : Spark SQL		07/03/26 (01)		
5.3,5.4	5.3 Programming with RDDs and Machine learning with MLlib		12/03/26 To 13/03/26 (02)		
5.5,5.6	5.4 Data ETL (Extract, Transform and Load) Process: Composing Spark Program steps for ETL 5.5 Analytics, Reporting and Visualization		14/03/26 To 20/03/26 (02)		
5.7,5.8	5.6 Apache Spark Streaming Platform: Spark Streaming Architecture, Spark streaming vs Structured streaming, Internal Working of Spark Streaming		27/03/26 To 28/03/26 (02)		
5.9	5.7 Spark streaming characteristics: Scalable, Fault Tolerance and Load Balancing		03/04/26 (01)		
<b>45 Hrs</b>		<b>Total</b>	<b>45 Hrs</b>		

- **COs – POs & PSOs MATRIX:**

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 & Environment	PO-6 Project Management	PO-7 Lifelong Learning	PSO- 1	PSO- 2
CO602.1	3	-	-	-	-	-	-	3	-
CO602.2	3	3	3	2	2	-	1	2	1
CO602.3	3	3	3	3	1	-	1	3	1
CO602.4	3	3	3	3	1	3	1	2	2
CO602.5	3	3	3	3	2	2	2	1	3

Legends :- High:03, Medium:02,Low:01, No Mapping: -

**PSO1:** Apply fundamental concepts of Artificial Intelligence and machine learning to solve technical problems.

**PSO2:** Implement the domain knowledge to achieve a successful career as an engineering professional.

- **Formative Assessment(FA-TH):**

Two offline class tests of 30 marks each will be conducted. Average of two class tests marks will be considered as Formative Assessment for Theory marks out of 30.

- **Summative Assessment(SA-TH):**

End semester assessment of 70 marks through paper based examination by MSBTE .

Total theory marks (100) will be calculated as marks of Formative Assessment (30) + marks of Summative Assessment (70)

## • Self Learning Activities

### Micro project

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

### Assignment

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

### Other

Infosys Springboard Course on Big Data Analytics

## • References:

### A. 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.

### B. Software Learning Websites :

Sr.No	Link / Portal	Description
1	<a href="https://hadoop.apache.org/">https://hadoop.apache.org/</a>	Official website for Apache Hadoop, including documentation, downloads, and tutorials.
2	<a href="https://spark.apache.org/">https://spark.apache.org/</a>	Official website for Apache Spark, providing guides, API references, and use case examples.

3	<a href="https://pig.apache.org/">https://pig.apache.org/</a>	Official site for Apache Pig, with resources for learning Pig Latin and building scripts.
4	<a href="https://hive.apache.org/">https://hive.apache.org/</a>	Official resource for Apache Hive, including installation guides and HiveQL references.
5	<a href="https://www.mongodb.com/">https://www.mongodb.com/</a>	MongoDB official site offering documentation, downloads, and free learning courses.
6	<a href="https://onlinecourses.nptel.ac.in/oc20_cs92/preview">https://onlinecourses.nptel.ac.in/oc20_cs92/preview</a>	This course provides an in-depth understanding of terminologies and the core concepts behind big data problems, applications, systems and the techniques that underlie today's big data computing technologies.
7	<a href="https://www.tutorialspoint.com/hadoop/index.htm">https://www.tutorialspoint.com/hadoop/index.htm</a>	This brief tutorial provides a quick introduction to Big Data, MapReduce algorithm, and Hadoop Distributed File System.
8	<a href="https://www.w3schools.com/mongodb/">https://www.w3schools.com/mongodb/</a>	This brief tutorial provides a quick introduction to MongoDB.

● **Tools :**

1. Google Classroom to share subject material to students.
2. MKCL ERA LMS login.
3. Infosys Springboard

Mrs. J. S. Mahajan  
**(Name & signature of Staff)**

Mrs. R. Y. Thombare  
**(Name & signature of HOD)**

Cc: Course File – BDA(316318)