

## Lecture Plan

<b>Name of Course</b>	<b>Big Data</b>
<b>AKTU Course Code</b>	<b>BCS-061</b>
<b>Branch</b>	<b>Computer Science and Engineering</b>
<b>Semester</b>	<b>6</b>
<b>Section</b>	<b>CS</b>
<b>Total Number of Students</b>	<b>72</b>
<b>Name of Faculty</b>	<b>Ram Prakash</b>
<b>Number of Lecture Proposed</b>	<b>55</b>

S.No.	Unit No	Topic	CO	No of Lectures Required	Actual Date of Completion	Suggested Reference
1	1	<b>Introduction to Big Data:</b> Type of digital data, history of Big Data innovation, Introduction to Big Data Platform	1	01	02/02/2026	4,1
2		Drivers for Big Data, Big Data architecture and characteristics, 5V <sub>s</sub> of Big Data		01	04/02/2026	4,1
3		Big data technology components, Big Data importance and applications		01	05/02/2026	4
4		Big Data features		01	07/02/2026	2
5		Big Data Analytics, Challenges of conventional systems,		01	09/02/2026	4,1
6		Intelligent data analysis		01	11/02/2026	3
7		Nature of Data, analysis vs reporting		01	12/02/2026	3
8		Modern data analytic tools.		01	13/02/2026	3
<b>No of Lectures Required to complete Unit 1</b>				<b>8</b>		
9	2	<b>Hadoop:</b> History of Hadoop, Apache Hadoop	2	01	14/02/2026	2,4
10		The Hadoop Distributed File System, components of Hadoop		01	16/02/2026	3
11		Data format, analyzing data with Hadoop		01	18/02/2026	3
12		Scaling out, Hadoop steaming, Hadoop pipes		01	19/02/2026	3
13		Hadoop Echo System		01	20/02/2026	3
14		<b>MapReduce:</b> MapReduce framework and basic		01	21/02/2026	2

15		How MapReduce works, Developing a MapReduce application		01	23/02/2026	2
16		Unit tests with MR unit, test data and local tests		01	25/02/2026	2
17		Anatomy of a MapReduce job run, failures, job scheduling, shuffle and sort, task execution		01	26/02/2026	2
18		MapReduce types, input formats, output formats		01	27/02/2026	1
19		MapReduce features, real-world MapReduce		01	28/02/2026	1
<b>No of Lectures Required to complete Unit 2</b>				<b>11</b>		
20	3	<b>HDFS (Hadoop Distributed File System):</b> Design of HDFS, HDFS concept, Benefits and challenges	3	1	06/03/2026	3,2
21		File sizes, block sizes and block abstraction inn HDFS		1	07/03/2026	3,2
22		Data replication, how does HDFS store, read, and write files		1	09/03/2026	3,2
23		Java interfaces to HDFS, command line interface, Hadoop file system interfaces		1	11/03/2026	3,2
24		Data flow data ingest with Flume and Scoop, Hadoop archives		1	17/03/2026	3,1
25		Hadoop I/O		2	19/03/2026	3,1
26		<b>Hadoop Environment:</b> setting up a Hadoop cluster, cluster specification, cluster setup and installation		1	20/03/2026	3,1
27		Hadoop configuration, security in Hadoop		1	23/03/2026	3,1
28		HDFS monitoring & maintenance, Hadoop in the cloud		1	24/03/2026	3,1
<b>No of Lectures Required to complete Unit 3</b>				<b>10</b>		
29	4	<b>Hadoop Eco System and YARN:</b> Hadoop ecosystem components, schedulers, fair and capacity	4	1	25/03/2026	2
30		Hadoop 2.0 New Features-NameNode high availability		1	28/03/2026	2
31		HDFS federation MRv2		1	30/03/2026	2,4
32		YARN		1	01/04/2026	2,3
33		Running MR1 in YARN		1	02/04/2026	2,3
34		<b>NoSQL Databases:</b> Introduction to NoSQL		1	04/04/2026	2,3
35		<b>MongoDB:</b> Introduction, data types		1	06/04/2026	4
36		Updating and deleting documents, querying		1	08/04/2026	1,2

37		Introduction to indexing, capped collections		1	09/04/2026	1
38		Spark: Installing spark, spark applications, jobs		1	10/04/2026	2,3
39		Stages and tasks, Resilient Distributed Databases, anatomy of a Spark job run, Spark on YARN		1	13/04/2026	1
40		<b>SCALA:</b> Introduction, classes and objects, basis types and operators		1	15/04/2026	1
41		Basic types and operators, built-in control structures		1	16/04/2026	1
42		Functions and closures, inheritance		1	17/04/2026	4
<b>No of Lectures Required to complete Unit 4</b>				<b>14</b>		
43	5	<b>Hadoop Eco System Frameworks:</b> Applications on Big Data using Pig, Hive, HBase	5	1	18/04/2026	1,4
44		<b>Pig-</b> Introduction to PIG, Execution Modes of Pig		1	20/04/2026	1,4
45		Comparison of Pig with Databases		1	22/04/2026	1,3
46		Grunt, Pig Lattin, UserDefined Functions		1	23/04/2026	1,4
47		<b>Hive-</b> Apache Hive architecture and installation, Hive shell, Hive services		1	25/04/2026	1,4
48		HiveQL, tables, querying data and user defined functions, sorting and aggregating		1	27/04/2026	1
49		MapReduce scripts, join & subqueries		1	29/04/2026	1
50		<b>HBase-</b> HBase concepts, clients, example		1	30/04/2026	1
51		HBase vs RDBMS, advanced usage, scheme design, advance indexing		1	01/05/2026	1
52		Zookeeper- How it helps in monitoring a cluster, how to build applications with Zookeeper		1	02/05/2026	1,4
53		IBM BigData strategy, Introduction to Infosphere		1	04/05/2026	4
54		BigInsights and Big Sheets, introduction to Big SQL		1	05/05/2026	4
<b>No of Lectures Required to complete Unit 5</b>				<b>12</b>		

<b>Text Books &amp; References</b>	
1	Thomas Erl, Wazid Khattak, Paul Buhler, Big Data Fundamentals: Concepts, Drivers and Techniques”, Prentice Hall.
2	ArshdeepBahga, Vijay Madiseti, “Big Data Science & Analytics: A HandsOn Approach”, VPT.
3	DT Editorial Services, Big-Data Black Book, Wiley.

4	Radha Shankarmani, “Big Data Analytics”, Wiley.
5	

<b>Course Outcomes (COs)</b>		
<b>At the end of this course students will demonstrate the ability to:</b>		
<b>CO 1</b>	Demonstrate knowledge of Big Data Analytics concepts and its applications in business.	
<b>CO 2</b>	Demonstrate functions and components of MapReduce Framework and HDFS.	
<b>CO 3</b>	Discuss Data Management concepts in NoSQL environment.	
<b>CO 4</b>	Explain process of developing MapReduce based distributed processing applications.	
<b>CO 5</b>	Explain process of developing applications using HBASE, Hive, Pig etc.	

**Name of Faculty with Signature:** Ram Prakash

**Lecture Plan In-charge:**

**Head of Department:**

**Dean Academics:**